Studying human neuropsychiatric disease in *DLG2* deficient neurons

By

Bret Sanders

A thesis submitted to Cardiff University for the degree of Doctor of Philosophy

August 2020

Abstract

Neuropsychiatric diseases such as schizophrenia can have profound impacts both on affected individuals and society more widely yet the neurobiology underlying these conditions is incompletely understood. Various environmental and genetic risk factors have been identified for these diseases which include rare deletions of the *DLG2* gene, identified in multiple schizophrenia patients. *DLG2* is known to encode a structural scaffolding protein within mature synapses; however, research presented here indicates protein expression earlier than previously appreciated and prior to synaptogenesis. To investigate a hypothesised role for DLG2 in early cortical neurogenesis and determine potential disease relevance, *DLG2*$^{-/-}$ and isogenic wild-type (WT) sister human embryonic stem cell (hESC) lines were differentiated into cortical excitatory neurons and characterised at multiple time points through various techniques, including RNA sequencing. These revealed novel roles for DLG2 during cortical development including the regulation of neural precursor cell (NPC) proliferation and adhesion to the extracellular matrix (ECM) as well as several biological progresses during neurogenesis responsible for the generation of neuronal subtype identity. *DLG2*$^{-/-}$ cultures produce more immature neurons characterised by deficits in morphology, migration, electrical properties and gene expression, including the cortical layer V marker CTIP2. Genes downregulated at day 30 of cortical differentiation, a time point corresponding to early neurogenesis, were enriched for schizophrenia common risk variants, linking DLG2 regulated biological processes in corticoneurogenesis with neuropsychiatric disease and potentially providing insight into the neurodevelopmental component of schizophrenia pathophysiology.

Acknowledgments

I would like to sincerely thank my supervisors Dr Jenny Shin and Dr Andrew Pocklington, their tireless support and guidance throughout my PhD that has made me the scientist I am today. Jenny may have taken a risk in accepting a mature student with no cell culture experience as her PhD student, but I hope the completion of this thesis vindicates that decision and I will be forever grateful for the opportunity she provided me. I particularly thank Andrew for his excellent explanations and assistance in regard to the bioinformatics components of this research, I feel I have progressed greatly in this area thanks to his aid and I am pleased to report the letter R no longer brings me out in cold sweats. I also thank by funders, as the research presented here would not have been possible without the generous support of the Welcome Trust Strategic Award (100202/Z/12/Z), Waterloo Foundation 'Changing Minds' Programme and start-up funding from the Neuroscience and Mental Health Research Institute, Cardiff University.

I would also like to acknowledge and thank all my colleagues and friends within the Hadyn Ellis Building both for their support over the last four years of my PhD and for making my time in the lab so enjoyable. Although the full list of those who provided some measure of assistance to me is too long to write here I must particularly thank Dr Claudia Tamburini, Dr Zoe Noakes and Dr Daniel Cabezas de la Fuente for their assistance in the initial stages of my PhD as well as more recently Dr Daniel D'Andrea for his advice with bioinformatics analyses. Two students who joined us briefly in Shin Lab and contributed to my research must also be acknowledged here: Sophie Pocklington who made neuronal morphology tracings and Sophia Wilson who assisted in attempts to create a DLG2-FLAG knock-in cell line, both impressed me with their work and have the capacity to become excellent scientists. Very little research would have been performed in the lab without the support of Emma Dalton, Trudy Workman and other members of the NMHRI technical team and thanks also go to Joanne Morgan and the MRC Centre Core Team for their assistance with RNA sequencing. No PhD is without both ups and downs and I must particularly thank two members of the 'TC2 Old Guard' for their invaluable moral support in getting me to where I am today: Dr Mouhamed Alsaqati who kindly kept is lab bench so messy over four years that mine always appeared immaculate by comparison and Dr Gareth Chapman who always had an answer to every question, even when he really didn't.

I would like to acknowledge the unfailing support provided to me by my family. It looks like their 'perpetual student' may finally be coming to the end of his formal education and I could not have made it to this point without them. In particular I would like to thank my two sisters Ava and Nell, although a gulf of years separates us they have been an inspiration to me through the dark days of the thesis write-up and I hope in some small way I may inspire them in their future academic

endeavours. Finally, I would like to dedicate this thesis to my late and dearly missed mother, Christine Sanders, who always dreamed that one day I would become a doctor.

Glossary of Abbreviations

ACC	Anterior cingulate cortex
ADHD	Attention deficient hypoactivity disorder
AIS	Axon initial segment
ASD	Autism spectrum disorder
AMPA	α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic
AP	Action potential
BMP	Bone morphogenic protein
CAM	Cell adhesion molecule
CARD	Caspase activation and recruitment domain
CASK	Calcium/calmodulin-dependent serine protein kinase
CNS	Central nervous system
CNV	Copy number variant
CP	Cortical plate
crRNA	CRISPR RNA
DEGs	Differentially expressed genes
DLPFC	Dorsolateral prefrontal cortex
DMEM/F12	Dulbecco's modified Eagle medium: nutrient mixture F-12
DSM-5	Diagnostic and statistical manual of mental disorders 5
ECM	Extracellular matrix
EM	Electron microscopy
EMSA	Electrophoretic mobility shift assay
FACS	Fluorescence activated cell sorting
FISH	Fluorescence in situ hybridisation
GABA	γ-aminobutyric acid
GFR	Growth factor reduced
GK	Guanylate kinase
GO	Gene ontology
gRNA	Guide RNA
G-S	Glycine-Serine
GWAS	Genome wide association study
HDR	Homology directed repair
HEK	Human embryonic kidney
hESC	Human embryonic stem cell
hiPSC	Human induced pluripotent stem cell

HITI	Homology independent targeted integration
hPSC	Human pluripotent stem cell
ICC	Immunocytochemistry
ICD	International classification of disease
ID	Inhibitor of DNA binding
Indels	Insertions and deletions
IPC	Intermediate precursor cell
KO	Knockout
Kv	Voltage gated potassium channel
L27	LIN2-LIN27
LoFi	Loss of function intolerant
LTD	Long-term depression
LTP	Long-term potentiation
MAGI	Membrane associated guanylate kinase inverted
MAGMA	Multi-marker Analysis of GenoMic Annotation
MAGUK	Membrane associated guanylate kinase
MEA	Multi electrode array
mEPSC	Miniature excitatory postsynaptic current
MHC	Major histocompatibility complex
MP	Mammalian phenotype
MZ	Marginal zone
nAChR	Nicotinic acetylcholine receptor
NDD	Neurodevelopmental delay
NMDA	N-methyl-D-aspartate
NPC	Neural precursor cell
NSC	Neural stem cell/neuroepithelial stem cell
OR	Odds ratio
oRG	Outer radial glia
OSVZ	Outer subventricular zone
PAM	Protospacer adjacent motif
PBS	Phosphate buffered saline
PCA	Principle component analysis
PDM	PDZ domain binding motif
PDL	Poly-d-lysine hydrobromide
PDZ	PSD95-DLG-ZO1

PFA	Paraformaldehyde
PSD	Postsynaptic density
PSG	PSD-SH3-GKs
QC	Quality control
RA	Retinoic acid
RG	Radial glia
RIN	RNA integrity number
RNAi	RNA interference
RNAseq	RNA sequencing
RNP	Ribonucleoprotein
sEPSC	Spontaneous excitatory postsynaptic current
SFK	Src family protein kinase
SH3	SRC homology 3
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SP	Subplate
ssHDR	Single stranded HDR
SVZ	Subventricular zone
TALEN	Transcription activator-like effector nuclease
TARP	Transmembrane AMPA receptor regulating protein
tracrRNA	Transactivating crRNA
tRG	Truncated radial glia
vRG	Ventricular radial glia
VZ	Ventricular zone
WPC	Weeks posts conception
WT	Wild type
ZFN	Zinc finger nuclease
ZO	zonula occludens

Contents

1. General introduction 1
   1.1. Schizophrenia is a complex neuropsychiatric disease 1
   1.2. Environmental and genetic risk factors in schizophrenia aetiology 4
   1.3. *DLG2* and human neuropsychiatric disease 7
   1.4. *Dlg2* and rodent neuropsychiatric disease models 10
   1.5. DLG2 and related MAGUK proteins 13
   1.6. DLG2 in the central nervous system 19
   1.7. Cortical development in rodents and primates 22
   1.8. Modelling cortical projection neuron development using hPSCs 27
   1.9. hPSC models of schizophrenia 29
   1.10. Hypothesis 33
2. Materials & methods 34
   2.1. Cell maintenance 34
   2.2. Cortical differentiation 34
   2.3. Western blotting 35
   2.4. Preparation of DLG2-FLAG isoforms 36
   2.5. CRISPR/Cas9 genome editing to generate tagged *DLG2* 36
   2.6. PCR & agarose gel electrophoresis 37
   2.7. TOPO cloning 38
   2.8. Sanger sequencing 38
   2.9. *DLG2* knockout hESC line generation 38
   2.10. Genetic validation of *DLG2* knockout hESC lines 38
   2.11. Immunocytochemistry (ICC) 39
   2.12. Peptide affinity pulldowns & mass spectrometry 39
   2.13. Morphometric analysis 40
   2.14. RNA sequencing 40
   2.15. Adhesion assay 42
   2.16. Proliferation assay 42
   2.17. Migration assay 43
   2.18. Electrophysiology 43
   2.19. MAGMA gene set enrichment 43
   2.20. Statistical analysis and data presentation 44
   2.21. Key resources 45
3. The spatial and temporal pattern of *DLG2* expression 49
   3.1. Introduction 49

3.2. Experimental procedure - outline 53
3.3. Western blotting to determine DLG2 expression during cortical differentiation 53
3.4. Multiplex CRISPR/Cas9 to generate tagged *DLG2* 56
3.5. iCas9 cell line & synthesised gRNAs to generate tagged *DLG2* 62
3.6. Cas9 nickase & synthesised gRNAs to generate tagged *DLG2* 67
3.7. Discussion 73
3.7.1 *DLG2* expression during early neurodevelopment 73
3.7.2 FLAG-tag knock-in to *DLG2* using CRISPR/Cas9 approaches was not successful 74
3.7.3 Alternatives to CRISPR/Cas9 to investigate the pattern of *DLG2* expression 75
3.7.4 Possible refinements to CRISPR/Cas9 for knock-in generation 76
3.7.5 Summary 77
4. Characterisation of the *DLG2*$^{-/-}$ phenotype in human NPCs and neurons 79
4.1. Introduction 79
4.2. Experimental procedure – outline 80
4.3. Generation of *DLG2*$^{-/-}$ hESC lines 81
4.4. Mass spectrometry confirms WT DLG2 expression at day 30 and full KO in *DLG2*$^{-/-}$ lines 83
4.5. *DLG2*$^{-/-}$ lines show no off-target effects and retain pluripotency 84
4.6. *DLG2*$^{-/-}$ NPCs show an increasing trend in PAX6 and KI67 expression 87
4.7. *DLG2* deficiency has no effect on production of post-mitotic neurons but significantly reduces expression of layer V identity 90
4.8. *DLG2*$^{-/-}$ neurons display immature morphology 91
4.9. Discussion 95
4.9.1. DLG2 *is expressed from day 30 in WT cells and absent from knock-outs* 95
4.9.2. *DLG2*$^{-/-}$ hESCs were successfully validated for off-target effects and pluripotency marker expression 96
4.9.3. Directed differentiation of *DLG2*$^{-/-}$ hESCs can be used to model human cortical neurogenesis 96
4.9.4. *DLG2* deficiency has a significant effect on the expression of neuronal identity, if not on the rate of neurogenesis 98
4.9.5. *An altered state within DLG2*$^{-/-}$ *NPCs may explain subsequent neuronal phenotypes* 99
4.9.6. Summary 100
5. Underlying mechanisms of the *DLG2*$^{-/-}$ phenotype 101
5.1. Introduction 101
5.2. Experimental procedure - outline 103
5.3. Large numbers of genes were differentially expressed in *DLG2*$^{-/-}$ cells during cortical differentiation 104
5.4. DEGs suggest neurodevelopmental deficits in *DLG2*$^{-/-}$ NSCs 111
5.5. *DLG2*$^{-/-}$ NPCs display increased proliferation and adhesion to the ECM 114
5.6. *DLG2*$^{-/-}$ neurons display altered morphology, migration and electrical properties 119

5.7. Aspects of the *DLG2*$^{-/-}$ phenotype show recovery at later time points 125
5.8. Discussion 127
5.8.1. RNAseq supports a novel role for *DLG2 in cortical neurogenesis* 127
5.8.2. DLG2 mutations act at early stages of cortical differentiation in NSCs/NPCs 127
5.8.3. Multiple *DLG2*$^{-/-}$ phenotypes indicated a delayed expression of cell-type identity 129
5.8.4. RNAseq indicates mitochondrial abnormalities in *DLG2*$^{-/-}$ NPCs and neurons 131
5.8.5. DLG2 mediated signalling may regulate multiple signalling pathways during early cortical differentiation 132
5.8.6. Possible DLG2-scaffolded signalling complexes in early neurodevelopment 133
5.8.7. Summary 134
6. The *DLG2*$^{-/-}$ phenotype and neuropsychiatric disease 135
6.1. Introduction 135
6.2. Experimental procedure 136
6.3. Enrichment for schizophrenia common risk variants 137
6.4. Enrichment for Alzheimer's common risk variants as a negative control 140
6.5. Relevance of specific processes and phenotypes to disease 141
6.6. Discussion 143
6.6.1. Genes down-regulated in *DLG2*$^{-/-}$ cells at day 30 of cortical differentiation are enriched for schizophrenia common risk variants 143
6.6.2. *DLG2* knockout reveals multiple biological processes disrupted in schizophrenia 144
6.6.3. Summary 145
7. General discussion 147
7.1. Research Summary 147
7.2. DLG2-mediated signalling in early cortical differentiation 147
7.3. DLG2-mediated signalling regulates schizophrenia relevant aspects of corticoneurogenesis 150
7.4. Stochastic signalling due to an absent DLG2 scaffold may explain the *DLG2*$^{-/-}$ phenotype 152
7.5. Future studies 153
7.6. Conclusions 155
8. References 157
9. Supplementary Information 194

1. General introduction

1.1. Schizophrenia is a complex neuropsychiatric disease

The term neuropsychiatric disease covers a broad range of complex and incompletely understood conditions in which nervous system dysfunction affects behaviour and cognition (Taber et al., 2010), such as bipolar disorder (Grande et al., 2016), autism spectrum disorder (ASD) (Lord et al., 2018) and schizophrenia (Owen et al., 2016). Like many other neuropsychiatric diseases, schizophrenia can be considered a severe condition both for affected individuals and society at large (Picchioni and Murray, 2007). Prevalence is relatively high, for many years believed to be around 1% of the population worldwide regardless of country, cultural group or sex (Mueser and McGurk, 2004). More recent meta-analyses have challenged this view and shown a prevalence at roughly 0.3-0.66% when a definitive schizophrenia diagnosis is considered, with variation depending on the precise diagnostic criteria used such as the minimum duration of illness, age of patients and symptoms (McGrath et al., 2008). When the diagnostic criteria are broadened further to included delusional disorder, brief psychotic disorder and psychotic disorders not otherwise specified the lifetime prevalence of Schizophrenia and related disorders rises to 2.3% (van Os and Kapur, 2009). These meta-analyses have also shown an increased prevalence of schizophrenia in men compared to women as well as variations in populations that cannot be explained by differences in study methodology or diagnostic criteria, indicating exposure to underlying risk factors is not uniform (McGrath et al., 2008; van Os and Kapur, 2009; Owen et al., 2016).

The symptoms of schizophrenia can be separated into three broad classes, positive symptoms, negative symptoms and cognitive impairment (Liddle, 1987; Mueser and McGurk, 2004). The positive symptoms (or psychotic symptoms) involve a loss of contact with reality and include hallucinations and delusions, these tend to be episodic and undergo periods of remission and relapse. The negative symptoms involve deficits in basic emotional and behaviour processes and include apathy, anhedonia (lack of pleasure) and alogia (poverty of speech), unlike the generally episodic nature of the positive symptoms negative symptoms are typically chronic. Cognitive impairments also tend to be chronic and include deficits in learning and memory, attention and concentration as well as executive function (problem solving). Although a reduction in cognitive ability is present in most patients with schizophrenia there is large degree of individual variability and some patients function at normal levels (Joyce and Roiser, 2007). As there are currently no diagnostic tests or biomarkers available for Schizophrenia, diagnosis is made clinically according to criteria established in the WHO International Classification of Disease (ICD) (WHO, 1992) or Diagnostic and Statistical Manual of Mental Disorders (DSM-5) (American Psychiatric Association, 2013) which include the presence of at least 2 our of 5 characteristic symptoms (delusions, hallucinations,

disorganised speech, grossly disorganised or catatonic behaviour and negative symptoms) for a significant portion of time during a 1 month period. This diagnosis generally occurs in late adolescence or early adulthood (16-30 years) after the first episode of psychosis, although in some cases negative symptoms and cognitive impairment are present prior to this (Lieberman et al., 2001; Mueser and McGurk, 2004; Addington and Heinssen, 2012). Various anti-psychotic drugs can be used for the effective long-term treatment of the positive symptoms of schizophrenia, although undesirable side effects often lead to poor adherence (Leucht et al., 2012). There is currently no effective pharmacological treatment for the negative symptoms or cognitive dysfunction associated with schizophrenia, which necessitates management through a framework of psychological and social support (Owen et al., 2016).

Although many of those diagnosed with schizophrenia show long-term improvement, over 50% have recurrent symptoms potentially required repeated hospitalisation while roughly 20% have chronic symptoms and disability (Barbato, 1998; Owen et al., 2016; Vita and Barlati, 2018). The profound impact of schizophrenia on affected individuals can been seen in unemployment rates of between 80-90% (Marwaha and Johnson, 2004; Kooyman et al., 2007), a roughly 5% increase in the rate of suicide and an overall decrease in life expectancy of between 10-20 years (Chesney et al., 2014). Schizophrenia is therefore one of the most disabling neuropsychiatric disorders and as such occupies a disproportionately large amount of mental health care resources, accounting for up to 50% of psychiatric hospital admission and 25% of beds (Geller, 1992; Terkelsen and Menikoff, 1995). In England alone Schizophrenia costs society £11.8 billion pounds per a year, a third of which is for direct expenditure on health and social care (Schizophrenia Commission, 2012). Due to the severe impact of schizophrenia and the resulting social and economic costs the World Health Organisation has placed it among the top ten causes of disability adjusted life years (years of life lost to premature death and severe disability) worldwide (Murray et al., 1996).

Many aspects of schizophrenia pathophysiology are incompletely understood; however, multiple studies have shown an enlargement of the ventricular system in the brain of schizophrenic patients along with a reduction in both overall brain volume and cortical grey matter volume in addition to various white matter abnormalities, although some inconsistency does exist between studies in regards to white matter changes, particularly within the lateral ventricles (Olabi et al., 2011; Haijma et al., 2013). In particular grey matter reduction in the temporal lobe appears to progress with the duration of schizophrenia, with the effect being exacerbated by antipsychotic treatment (Vita et al., 2012). Dysfunction of the prefrontal cortex has also been implicated in several cognitive defects found in schizophrenia patients including working memory and executive function (Barch and Ceaser, 2012; Lewis, 2012). Although imaging studies have therefore revealed several brain structural features consistently associated with Schizophrenia these do not currently have the

sensitivity to qualify as a diagnostic marker (Linden, 2012), possibly due to the complex and varied symptoms associated with schizophrenia along with a degree of overlap with other neuropsychiatric diseases. When the underlying neurobiology is considered, various brain imaging and pharmacological studies have implicated a dysregulation in neurotransmission in schizophrenia patients (Foss-Feig et al., 2017). This includes hyperactive dopaminergic transmission which has be linked to the positive symptoms of schizophrenia such as hallucinations and delusions (the dopamine hypothesis), with the acute psychotic state associated with an increase in resting state dopamine concentrations, dopamine synthesis and dopamine release (Guillin et al., 2007). Although potentially explain the positive symptoms, which are shared with other neuropsychiatric disorders (Howes and Murray, 2014), a dysregulation of dopaminergic signalling alone cannot explain the wide range of symptoms associated with schizophrenia (Owen et al., 2016). Hypoactive glutamate neurotransmission via N-methyl-D-aspartate (NMDA) receptors was also indicated by brain imaging (Bressan et al., 2005; Stone et al., 2008), pharmacological and later genetic studies (see section 1.2) and potentially explains certain negative symptoms and cognitive impairments associated with schizophrenia (the glutamate hypothesis) (Kantrowitz and Javitt, 2010; Moghaddam and Javitt, 2012). Studies have also shown a role for oxidative damage in schizophrenia pathophysiology with parvalbumin positive inhibitory interneurons appearing particularly vulnerable both to oxidative stress (Owen et al., 2016) as well as disruptions to NMDA receptor mediated signalling (Lewis, 2012).

Despite the onset on schizophrenia, with associated aberrant neurotransmission, occurring during adolescence or early adulthood, there is evidence to suggests that it is a neurodevelopmental disease with a prenatal component (Murray and Lewis, 1987; Weinberger, 1987; Harrison, 1997, 1999). Lateral ventricle enlargement in the brains of schizophrenia patients manifests prior to the first psychotic episode and can also be found in first degree relatives, suggesting they are not the result of psychosis but inherited neurodevelopmental risk factors (McDonald et al., 2002). A prenatal origin is also supported by the increased incidence of various morphological abnormalities of early brain development in schizophrenia patients such as the absence or reversal of cerebral structural asymmetries (Sharma et al., 1999). Although caveats exist regarding small sample size post-mortem histological studies have also shown defects in the laminar organisation of the cortical plate with neurons displaced towards the inner surface in schizophrenic brains compared to controls (Akbarian et al., 1993). As the cortex forms in an inside-out fashion (see section 1.7) this indicates a defect in the normal process of neuronal migration to the outer surface of the cortical plate, a process that occurs during the second trimester of gestation and hence clearly prenatal in nature (Mueser and McGurk, 2004). Given the indicated neurodevelopmental nature of schizophrenia the exact mechanism underlying the delayed onset of psychosis is unclear; however, progression can

be triggered by various postnatal environmental exposures that are likely modified by prenatal environmental and genetic risk factors (Owen et al., 2016).

1.2. Environmental and genetic risk factors in schizophrenia aetiology

Although the aetiology of schizophrenia is incompletely understood there are well established environmental risk factors, both biological and psychosocial, associated with the disease (Mueser and McGurk, 2004; van Os and Kapur, 2009; Owen et al., 2016). In line with the hypothesised neurodevelopmental nature of schizophrenia many of these environmental risk factors act prenatally during key periods of early neurodevelopment and include maternal malnutrition (Susser and Lin, 1992; McGrath et al., 2010), stress (Khashan et al., 2008) and infection (Brown, 2012; Khandaker et al., 2013) as well as various obstetric complications, particularly those associated with hypoxia (Thomas et al., 2001; Cannon et al., 2002; Brown, 2011). Other environmental risk factors act during childhood and early adolescence and include growth in an urbanised environment (Vassos et al., 2012), an immigrant background (Cantor-Graae and Selten, 2005; Cantor-Graae and Pedersen, 2013), childhood adversity (Varese et al., 2012), head injury (Orlovska et al., 2014), epilepsy (Clancy et al., 2014), infections (Khandaker et al., 2012), autoimmune diseases (Benros et al., 2014) and cannabis use (Radhakrishnan et al., 2014; Gage et al., 2017). Parental age is also a risk factor with children born to relatively young parents (<20 years) or old fathers (>40 years) being associated with an increased risk of schizophrenia (Miller et al., 2011; McGrath et al., 2014), as are children born in the late winter or early spring (Davies et al., 2003).

Although various studies have identified environmental risk factors which significantly contribute to schizophrenia aetiology, alone these cannot fully explain vulnerability to the disease. For many years schizophrenia aetiology has also been known to possess a strong genetic component; with twin studies (studies of identical or fraternal twins to determine the contribution of genetic and environmental factors to phenotypes) showing a heritability estimate for the disease of roughly 80%, in comparison to 50% for osteoarthritis of the hip or 30-50% for hypotension (Cardno and Gottesman, 2000; Sullivan et al., 2003; van Os and Kapur, 2009; Hilker et al., 2018). Despite this strong genetic association, any specific molecular genetic variation underlying schizophrenia remained unclear due to the highly polygenic nature of the disease and the inherent limitation of genetic epidemiology studies (Owen et al., 2016). It is only in recent years that advancements in both genome sequencing technology and analysis techniques have enabled large-scale studies of schizophrenia patient cohorts to identify genetic variation compared to control groups (van de Leemput et al., 2016). As a result, genome wide association studies (GWAS) have now identified single nucleotide polymorphisms (SNPs) at 145 distinct genetic loci that are associated with schizophrenia (Pardiñas et al., 2018). These common risk variants occur at varying population

frequencies but all have relatively small individual effects, it is the combined effect of many such SNPs that contributes to overall disease risk (Sullivan et al., 2012; Ripke et al., 2014; Pardiñas et al., 2018). Although SNPs collectively account for up to 50% of schizophrenia liability (Purcell et al., 2014), large scale deleted or duplicated regions of the genome termed copy number variants (CNVs) have also been associated with schizophrenia risk through genomic studies (Kirov et al., 2014). These CNVs individually carry a relatively high risk for schizophrenia with odds ratios between 2-60 (Marshall et al., 2017) but are rare compared to SNPs, 12 recurrent CNVs along with various *de novo* (newly occurring) CNVs associated with disease risk have currently been identified (Xu et al., 2008; Kirov et al., 2012a; Malhotra and Sebat, 2012; Rees et al., 2014; Warland et al., 2020). Other rare variants implicated in schizophrenia by whole exome sequencing studies include insertions and deletions (indels) as well as rare *de novo* and inherited single nucleotide variants (SNVs) (Fromer et al., 2014; Purcell et al., 2014).

The full range of biological processes regulated by schizophrenia risk variants is yet to be determined. However, rare variants including CNVs, SNVs and indels have been found in genes relating to mature synaptic signalling including those encoding voltage-dependent calcium channels, NMDA receptor and ARC complexes, PSD-95 complexes , $GABA_A$ receptor complexes and FMRP targets (Kirov et al., 2012a; Fromer et al., 2014; Purcell et al., 2014; Szatkiewicz et al., 2014; Hall et al., 2015; Pocklington et al., 2015; Genovese et al., 2016); strongly implicating a dysregulation of neuronal signalling in schizophrenia aetiology, in agreement with brain imaging and pharmacological studies. The biological processes regulated by schizophrenia common risk variants are less clear, GWAS studies implicate components of the major histocompatibility complex (MHC) that encodes proteins essential for the adaptive immune system (Ripke et al., 2014), a finding possibly supported by clinical and epidemiological studies implicating immune and inflammatory processes in schizophrenia (Smyth and Lawrie, 2013; Benros et al., 2014); however, enrichment is limited compared to diseases of known immune origin and there is no GWAS enrichment of immune loci outside of the MHC region (Pouget et al., 2016). Within the MHC locus alleles of the complement component 4 (C4) genes appear central to the GWAS signal (Sekar et al., 2016), these encode neuronal localised proteins that may be involved in synaptic pruning (the elimination of synapse during postnatal development) potentially contributing to the reduced dendritic spine density observed in the cortical pyramidal neurons of schizophrenic brains (Garey et al., 1998; Glantz and Lewis, 2000; Glausier and Lewis, 2013). Variation in C4 genes also have a greater effect on schizophrenia risk in men than women, offering an explanation to male-biased disease vulnerability (Kamitaki et al., 2020). The NMDA receptor and ARC complexes strongly implicated by rare variant studies show little enrichment in GWAS and although broader synapse related genes are enriched for common risk variants these still account for only a small

proportion of identified SNPs (Pardiñas et al., 2018). Genes identified from GWAS studies do include those encoding FMRP targets, voltage-dependent calcium channels, glutamate receptors and dopamine receptor (D2), the main target of antipsychotic drugs (Ripke et al., 2014; Pardiñas et al., 2018). This GWAS signal also consistently maps onto specific brain cell types including pyramidal neurons (glutamatergic cells) and cortical interneurons (GABAergic cells) due to specific sets of genes expressed in these cells (Skene et al., 2018). Although clearly not the only mechanisms involved, both rare and common variant studies indicating disruption to glutamatergic, GABAergic and dopaminergic signalling provides some clue as to the neurobiology underlying schizophrenia pathophysiology.

Studies have also shown significant overlap between schizophrenia risk variants and those for other neuropsychiatric and neurodevelopmental disorders; providing genetic support for a neurodevelopmental component in schizophrenia aetiology, in addition to the previously described brain anatomy and cytoarchitectural changes as well as prenatal environmental risk factors. There is significant sharing of schizophrenia common risk variants with bipolar disorder, major depressive disorder, attention deficit hyperactivity disorder (ADHD) and ASD (Hamshere et al., 2013; Lee et al., 2013). A similar pattern of shared genetic risk is also seen in schizophrenia rare variants including CNVs, SNVs and indels, which show overlap with a range of established neurodevelopmental disorders including ASD, ADHD, intellectually disability and severe neurodevelopmental delay (NDD) (Malhotra and Sebat, 2012; Owen, 2012; Fromer et al., 2014; Kirov et al., 2014). Additionally, it has been shown that common schizophrenia risk variants are strongly enriched for loss of function intolerant genes (LoFi) which capture roughly 50% of genic SNP-based heritability for the disorder; as LoFi genes are under strong selective pressure they are believed to play key roles in normal development (Pardiñas et al., 2018). Furthermore LoFi genes are also enriched for rare risk variants associated with ASD, intellectually disability/NDD and schizophrenia (Genovese et al., 2016; Kosmicki et al., 2017; Singh et al., 2017; Rees et al., 2020). This therefore suggests that the disruption of neurodevelopmental pathways may explain a significant proportion of schizophrenia genetic risk, as well as indicating that schizophrenia may lie on a neurodevelopmental continuum with other neurodevelopmental disorders such as intellectually disability, ASD and ADHD, with the severity of symptoms governed by the timing, severity and pattern of disrupted brain development (Owen and O'Donovan, 2017). However, despite the strong genetic, anatomical and epidemiological evidence for a neurodevelopmental component to schizophrenia aetiology, little is known regarding the neurodevelopmental processes disrupted in the disease other than they are likely to occur early in the second trimester of pregnancy (Harrison, 1997; Hill and Bray, 2012; Clifton et al., 2019), with many common risk variants impacting gene expression in the mid-foetal brain, particularly in cortical excitatory neurons and their progenitors

(Hill and Bray, 2012; O'Brien et al., 2018; Clifton et al., 2019; Polioudakis et al., 2019; Walker et al., 2019).

1.3. *DLG2* and human neuropsychiatric disease

The discs large MAGUK scaffold protein 2 or discs large homologue 2 (*DLG2*) gene is a human protein coding gene located at 11q14.1 which multiple studies have linked to schizophrenia and neuropsychiatric disease more widely. *In situ* hybridisation and western blotting studies have shown significant differences in *DLG2* expression within the anterior cingulate cortex (ACC) of schizophrenia brains compared to controls (Kristiansen et al., 2006). In these studies, post-mortem brain tissue containing the dorsolateral prefrontal cortex (DLPFC) and ACC were obtained from 24 elderly well-diagnosed schizophrenia patients and 16 matched controls, none of which showed signs of neurodegenerative disease. A combination of *in situ* hybridisation and western blotting showed no changes in *DLG2* expression within the DLPFC; however, there was a significant reduction in DLG2 protein expression within the ACC of the schizophrenia samples compared to controls. Interesting these schizophrenia samples also showed a matching increase in the number of *DLG2* transcripts within the ACC, indicating that the loss of protein product may trigger a compensatory response at the transcriptional level.

Through the integration of schizophrenia and subcortical brain volume GWAS data, variants at the *DLG2* locus have also been associated with abnormal putamen volume in schizophrenia patients (Smeland et al., 2018). To determine this GWAS data on schizophrenia was obtained from the PGC study (http://pgc.unc.edu/) while GWAS data on MRI volumetric measures of brain regions (amygdala, caudate nucleus, hippocampus, nucleus acumens, pallidum, putamen and thalamus as well as intracranial volume) were obtained from the ENIGMA study (http://enigma.ini.usc.edu/). After performing appropriate controls on the data sets a conditional false discovery rate (FDR) approach was used to investigate genetic overlap between schizophrenia and brain structure volumes, identifying 6 shared genetic loci. *DLG2* was one such loci and was found to be associated with an increase in putamen volume, a feature which has been observed in some neuroimaging studies of schizophrenia patients (Okada et al., 2016). While an increased putamen volume has also been liked to certain functional disturbances in schizophrenia patients including verbal learning, executive functioning and working memory (Hartberg et al., 2011).

Genealogical studies have also indicated a for *DLG2* mutations in schizophrenia risk such as one recently published for a large multiplex 22q11.2 deletion syndrome family (Michaelovsky et al., 2019). The 22q11.2 microdeletion is one of the greatest genetic risk for schizophrenia (Marshall et al., 2017) with roughly 25% of individuals this mutation developing schizophrenia in adulthood (Schneider et al., 2014; Van et al., 2017). As not all individuals with 22q11.2 deletions develop

schizophrenia the authors of this study (Michaelovsky et al., 2019) sort to investigate genetic risk factors outside of this region potentially contributing to the disease. A multiplex 22q11.2 deletion syndrome family being selected as the risk variants are expected to be enriched in schizophrenia affected individuals compared to non-affected family members, with a genealogical analysis eliminating the need for population stratification as required in a case-control study. Although a total of 15 CNVs were detected in this family the one member with a confirmed schizophrenia diagnosis was the only individual to contain a *de novo* 11q14.1 deletion, the CNV containing the *DLG2* gene.

As described previously CNVs although rare in the population can confer significant schizophrenia risk (see section 1.2), although limited samples sizes and hence statistical power have previously hampered efforts to robustly confirm the contribution of CNVs to genetic risk. To address this an analysis of CNV contribution to schizophrenia risk was performed on large schizophrenia cohort (21,094 cases and 20,277 controls) using data from the PGC study (Marshall et al., 2017). The results of this study confirmed a global increase of CNV burden in a schizophrenia cohort compared to controls even when loci implicated in previous studies were excluded, furthermore this CNV burden was particularly enriched for genes associated with synaptic function, which included *DLG2*. Relating to its function within the synapse DLG2 protein is known requirement for the correct assembly of NMDA receptor supercomplexes (See section 1.5) (Frank et al., 2016), such complexes are themselves enriched for rare mutations in schizophrenia cases providing an additional link to disease (Kirov et al., 2012a; Fromer et al., 2014; Purcell et al., 2014; Szatkiewicz et al., 2014; Pocklington et al., 2015; Genovese et al., 2016).

Schizophrenia in common with other neurodevelopmental phenotypes is associated with reduced reproductive output, with patients possessing a fertility of 40% or lower than the general population (Laursen and Munk-Olsen, 2010; Bundy et al., 2011). As a number of rare CNVs are known to substantially increase schizophrenia risk it is likely that such CNVs are removed from the population through natural selection and hence must occur frequently *as de novo* mutations. To investigate this a study of rare CNV mutations in 662 schizophrenia proband-parent trios (an affected individual with both parents) was performed (Kirov et al., 2012a) which showed that *de novo* CNV mutation were more frequent in cases (5.1% of all cases or 5.5% of cases with no history of schizophrenia) compared to 2623 controls (2.2%), indicating the rate of de novo CNVs in schizophrenia patients is roughly double that of the general population. In addition to confirming the involvement of *de novo* CNVs in the pathogenesis of schizophrenia this study also implicated specific abnormalities of postsynaptic signalling complexes in the disease which included multiple *de novo* deletions within the *DLG2* gene.

Providing further evidence of the role of *DLG2* CNVs in schizophrenia a separate study identified two *DLG2* deletions and one *DLG2* duplication within a 1,115 strong schizophrenia cohort, this study also finding one *DLG2* deletion in an unaffected offspring of a deletion patient although none were found in the 978 strong control cohort (Nithianantharajah et al., 2013). These four individuals with identified mutations in the coding region of *DLG2* were also assessed for cognitive function using the Cambridge Neuropsychological Test Automated Battery (CANTAB), particularly focusing on 3 tasks which were comparable to a rodent touchscreen battery previously used to examine *Dlg2*$^{-/-}$ mice (see section 1.4); these were intra-extradimensional set-shifting to assess visual discrimination acquisition and cognitive flexibility, paired associates learning to assess visuo-spatial learning and memory and rapid visual information processing to assess sustained attention. The results of these tests showing that cognitive function was disrupted in humans with *DLG2* mutations, who made significantly more errors in all tests compared to healthy controls from the general population.

In addition to schizophrenia both *de novo* and inherited deletions of the *DLG2* promoter and first coding exons have been linked to intellectually disability, with a study identifying 29 patients with deletions in the exon 7-9 region of *DLG2* (Reggiani et al., 2017). An analysis of whole genome sequencing data from a cohort of 2,308 individuals from 493 multiplex ASD families also identifying a recurrent 2.5kb deletion in the *DLG2* promoter in 3 families while no deletions overlapping this region were found in controls, linking inherited mutations of the *DLG2* promoter to ASD (Ruzzo et al., 2019).

Multiple studies have therefore implicated *DLG2* mutations, in particular *DLG2* deletions, in schizophrenia and related disorders although do date there is no published data on the penetrance of deletions at this locus, that is the probability that an individual will develop schizophrenia given they carry a *DLG2* deletion. This has recently been addressed in unpublished work by Dr Andrew Pocklington (MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University) (Supplementary Document 1) using an established approach to estimate penetrance developed by (Kirov et al., 2014). This used data from a CNV study of 11,355 schizophrenia cases and 16,416 matched controls, in which 14 cases were found to have a deletion overlapping the *DLG2* gene compared to 4 in controls (Pocklington et al., 2015). In order to better estimate the probability that an individual from the general population carriers a *DLG2* deletion additional data was needed, as individuals suffering from other disorders (not schizophrenia) associated with *DLG2* deletions are inevitably underrepresented in the 'healthy control' group from a schizophrenia case-control study. (Kirov et al., 2014). This data on other disorders was taken from a large CNV of study of 38,799 individuals diagnosed with a diverse range of neurological deficits and subsequently referred for genetic testing, which identified partial *DLG2* deletions in 5 individuals (Sahoo et al., 2011).

Calculating from these values give the penetrance of *DLG2* deletions in schizophrenia at 4.95% and in other more severe developmental disorders at 2.07% (Supplementary Document 1). Interestingly this schizophrenia penetrance is comparable to that found at more widely studied loci such as 1q21.1 deletions/duplications and duplications at the Angelman/Prader-Willi-syndrome locus (Kirov et al., 2014).

1.4. *Dlg2* and rodent neuropsychiatric disease models

In addition to human studies various rodent models have been used to investigate the link between *Dlg2* and neuropsychiatric disease as well as identify phenotypes associated with *Dlg2* deficiency. As described previously hypoactive glutamate neurotransmission via NMDA receptors may explain some of the neurobiology underlying schizophrenia (see section 1.1) (Kantrowitz and Javitt, 2010; Moghaddam and Javitt, 2012) with competitive antagonists of NMDA receptors such as ketamine and phencyclidine shown to increase the severity of schizophrenia symptoms in patients as well as induce these symptoms (positive, negative and cognitive impairments) in healthy individuals (Javitt and Zukin, 1991; Krystal et al., 1994, 2000). In support of this rodent models of NMDA receptor hypofunction have also been produced which show both cellular and behavioural changes resembling those of schizophrenia patients including a reduction in the numbers of parvalbumin positive GABAergic interneurons and cognitive deficits (Rujescu et al., 2006). In one such rodent model of NMDA receptor hypofunction, produced by chronically treating 35 day old rats with low doses of the highly selective non-competitive NMDA receptor antagonist MK-801 for 20 days, RNAseq was performed on isolated hippocampal tissue to determine gene expression compared to control animals (Ingason et al., 2015). This analysis showed that *Dlg2* was among the top 20 most differentially expressed genes, showing a significant reduction within the treated animals compared to controls. When the human orthologues of *Dlg2* and the other 20 most differently expressed genes were tested for schizophrenia association in the PGC GWAS data *DLG2* reached gene-wide significance providing further support for *DLG2* as a schizophrenia risk gene. It should be noted that although *DLG2* reached significance when tested as one of 20 genes in an hypothesis-driven study it has not yet reached genome-wide significance as a canonical schizophrenia risk gene due to the high threshold for association in GWAS studies (roughly $p=5x10^{-8}$) required due to the larger number or markers tested.

Dlg2$^{-/-}$ mouse models have also been generated as part of a wider study into the role of the four *Dlg* paralogues in cognitive function (Nithianantharajah et al., 2013). For this research attempts were made to generate homozygous knockouts of *Dlg*1-4 through the breeding of heterozygous mice, this was successful for *Dlg2-4* although *Dlg1*$^{-/-}$ mice were embryonic lethal. Homozygous knockouts of *dlg* in *Drosophila* (Woods et al., 1996) and *dlg-1 in Caenorhabditis elegans*

(Bossinger et al., 2001) have also been shown to be embryonic lethal suggesting that *Dlg1* has retained the more developmentally critical functions of the invertebrate gene compared to the other paralogues, including *Dlg2*. Once generated the *Dlg2*$^{-/-}$, *Dlg3*$^{-/-}$ and *Dlg4*$^{-/-}$ mice as well as the *Dlg1*$^{+/-}$ mice (as *Dlg1*$^{-/-}$ were not viable) were assessed through a battery of touchscreen tasks of increasing complexity designed to assess cognitive function, in all cases the *Dlg1*$^{+/-}$ mice responding similar to WT with the focus of this research therefore being the differential roles of *Dlg2-4*. The cognitive tasks in the touchscreen battery were designed around the response of nose-poking a stimulus on the touchscreen to obtain a food reward, with the first element being training the mice though several sessions with the touchscreens to acquire this simple conditioned response. Both *Dlg2*$^{-/-}$ and *Dlg3*$^{-/-}$ completed this training at a similar rate to WT controls while *Dlg4*$^{-/-}$ mice showed a marked deficit in the acquisition of this conditioned response. Similarly, *Dlg4*$^{-/-}$ mice showed deficits in simple associative learning, including a test of Pavlovian conditioning whereby a stimulus appearing on the touchscreen in the form of a white vertical rectangle signalled a delivery of a food reward. Overall, this initial data indicates differences in function for the *Dlg* paralogues and shows that unlike *Dlg4*, *Dlg2 a*nd *Dlg3* are not required for simple forms of conditioning and associate learning.

To investigate the role of *Dlg2* and *Dlg3* in more complex cognitive tasks (Nithianantharajah et al., 2013) next presented mice with two stimuli on the touchscreens and required them to learn which one was rewarded when selected. In this task *Dlg2*$^{-/-}$ mice showed no differences from WT while interestingly *Dlg3*$^{-/-}$ mice showed an increased learning rate, indicating that neither *Dlg2* or *Dlg3* are essential for basic perceptual processing and in fact that *Dlg3* appears to restrict or modulate this specific aspect of cognition. When complexity was increased further in the form of an object-location pared associates task (i.e. through the incorporation of spatial information) *Dlg3*$^{-/-}$ mice no longer showed an increased learning rate and behaved similar to WT while *Dlg2*$^{-/-}$ mice now showed significant impairment, behaving a roughly 50% of WT levels; indicating that visual and spatial learning are separately regulated by the *Dlg* paralogues. Additional cognitive tasks further demonstrated that *Dlg2* and *Dlg3* have opposing actions, in tests of reversal learning (switching which of the two stimuli on screen produced a reward) *Dlg3*$^{-/-}$ mice behaved similar to WT while *Dlg2*$^{-/-}$ showed impairments and took significantly longer to learn the new reward contingency; suggesting that while *Dlg3* regulates the acquisition of learned information *Dlg2* regulates its flexibility. Differences between these 2 paralogues were also seen in extinction learning tasks (no longer rewarding the conditioned response after it has been learned) in which compared to WT *Dlg3*$^{-/-}$ mice showed faster extinction and *Dlg2*$^{-/-}$ slower. Also, in a test of sustained divided attention, the five-choice serial reaction time task (requiring a rapid response at one of 5 random location to receive a reward), *Dlg3*$^{-/-}$ mice acquired a stable level of performance at a similar rate to

WT while *Dlg2*$^{-/-}$ mice took significantly longer. These studies of mice models have therefore indicated a role for *Dlg2* in complex learning and cognitive flexibility, with *Dlg2*$^{-/-}$ mice showing deficits in reversal learning, object-location paired-associates learning, extinction and attention; validating these results similar impairments were also found in schizophrenia patients with *DLG2* mutations (see section 1.3) (Figure 1.1).

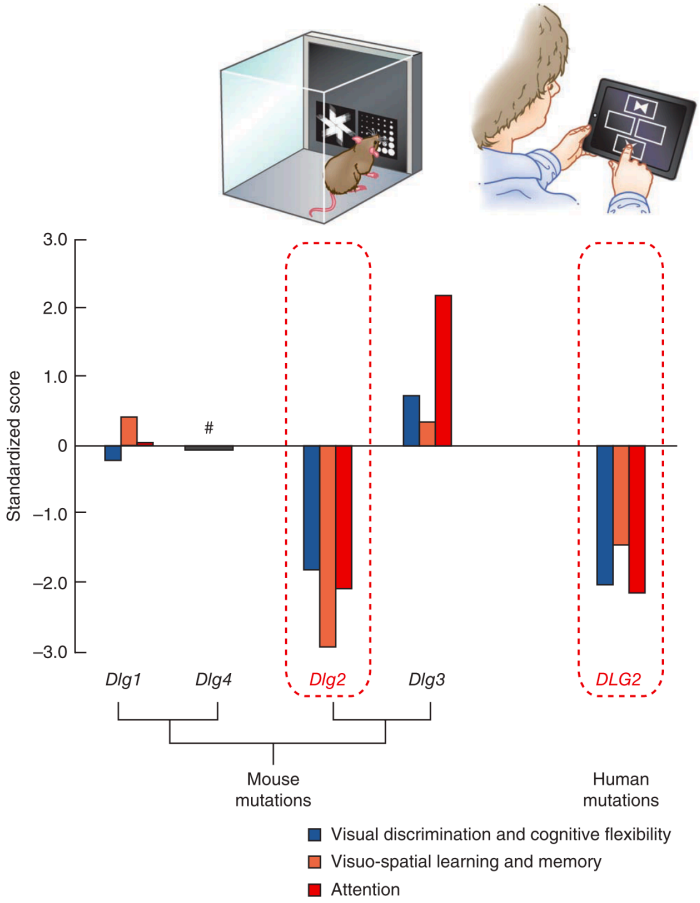


Figure 1.1 Conservation of *Dlg2* functions in mice and humans.
Conservation of *Dlg2* functions in mice and humans. Comparison of performance in touchscreen tasks for mice carrying mutations in *Dlg1, Dlg2, Dlg3* or *Dlg4* with humans carrying mutations in *DLG2*. The Cambridge Neuropsychological Test Automated Battery (CANTAB) was used to asses four people with mutations in *DLG2* on three tasks comparable to those in the rodent touchscreen battery. The intra-extradimensional set-shifting task assessed discrimination acquisition and cognitive flexibility, the paired associates learning task assessed visuo-spatial learning and memory and the rapid visual information processing task assessed sustained attention. A standardized performance score compared to WT littermates or healthy human subjects from the general population is shown, where a negative score indicates poorer than average performance. #Bar denotes the lack of data for comparison owing to the inability to test *Dlg4* mutant mice on any of the three tasks represented. From (Nithianantharajah *et al.*, 2013).

Both *Dlg2*$^{-/-}$ and *Dlg2*$^{+/-}$ mice have also been studied using a battery of tests to investigate their processing of social stimuli and social behaviour; these tests including social interaction in pairs, resident-intruder, ultrasonic vocalization, behaviour in the home cage and exploratory activity in the open field (Winkler et al., 2018). The results of this study showing that while *Dlg2*$^{+/-}$ mice behaved similarly to WT, *Dlg2*$^{-/-}$ mice showed highly increased sociability despite a motor impairment phenotype which prevented their evaluation on certain tests (such as those involving

swimming or moving backwards out of a tube). A relatively similar hypersocial phenotype was observed in *Dlg4*$^{-/+}$ mice and interestingly these mice also showed elevated levels of DLG2 protein in hippocampal synapse preparations compared to WT controls, indicating a certain degree of shared function and a possible compensatory mechanism between DLG2 and DLG4. Despite this research showing hypersocial behaviour in *Dlg2*$^{-/-}$ mice (Winkler et al., 2018) a separate study appears to contradict this and has shown that DLG2 deficiency in mice leads to reduced sociability as well as an increase in repetitive behaviour (Yoo et al., 2020). In this research *Dlg2*$^{-/-}$ mice showed both hypoactivity in a novel environment as well as decreased social approach compared to WT controls; furthermore they exhibited strong self-grooming both in novel environments and in their home cages. Although it appears further work may be required to clarify the behavioural phenotypes associated with *DLG2* deficiency both studies do agree on the aberrant locomotor responses in *Dlg2*$^{-/-}$ mice (Winkler et al., 2018; Yoo et al., 2020).

1.5. DLG2 and related MAGUK proteins

The protein encoded by the *DLG2* gene (DLG2 or PSD93) is a member of the membrane-associated guanylate kinase (MAGUK) superfamily (Zheng et al., 2011). MAGUKs are a family of structural scaffolding proteins that are predominantly located on cell membranes, at the sites of cell-cell junctions, and appear particularly enriched within the postsynaptic density (PSD) of excitatory synapses (Zhu et al., 2016). These MAGUK scaffolds have been shown to play a key role in both synapse maturation and synaptic plasticity (Funke et al., 2005; Oliva et al., 2012), with various nervous system disorders resulting from mutation in genes encoding these proteins (Grant, 2012). The PSD was initially identified from electron microscopy (EM) studies as an electron-dense thickening located below the postsynaptic membrane (PALAY, 1956); subsequent biochemical analyses have shown the PSD to be composed of densely packed proteins with a wide range of abundances. These include neurotransmitter receptors, scaffolding proteins, signalling enzymes, components of the cytoskeleton, adhesion molecules, membrane trafficking proteins and others (Feng and Zhang, 2009). MAGUKs within the PSD act as scaffolding proteins linking cell surface receptors or ion channels with signalling enzymes or components of the cytoskeleton located deeper within the PSD (Figure 1.2). MAGUK scaffolds do not hold the proteins of the PSD rigidly in place bur rather facilitate a dynamic environment of assembly and disassembly in response to neuronal stimuli (Choquet and Triller, 2013), a process linked to synaptic plasticity including long-term potentiation (LTP) and long-term depression (LTD) (Caroni et al., 2012; Huganir and Nicoll, 2013; Bosch et al., 2014). Although sharing many similar features MAGUK proteins can be divided into various subfamilies based on domain structure (Figure 1.3 A). These subfamilies include the disc large homologues (DLGs), calcium/calmodulin-dependent serine protein kinase (CASK),

membrane-associated guanylate kinase inverted (MAGI), zonula occludens (ZO) and caspase activation and recruitment domain (CARD)-containing MAGUK proteins (CARMAs) (Feng and Zhang, 2009; Danielson et al., 2012). As indicated by their name all MAGUK contain a guanylate kinase (GK) domain, although this has been shown to be catalytically inactive (Zhu et al., 2011). With the exception of the MAGI subfamily members all MAGUKs also contain a single SRC homology 3 (SH3) and one or more PSD95-DLG-ZO1 (PDZ) domains, with a PDZ-SH3-GK (PSG) tandem at the protein's C-terminal end; along with this basic architecture some MAGUKs contain additional domains such as CaM kinase, WW and LIN2-LIN27 (L27) (Feng and Zhang, 2009). With the exception of the CaM kinase domain found in CASK, which can function as a Mg^{2+} independent neurexin kinase (Mukherjee et al., 2008), no MAGUK domain have catalytic activity; however, all are known to function as protein-protein interaction modules (Zhu et al., 2016). DLG2 is a member of the DLG subfamily which contains four closely related proteins; DLG1 (SAP97), DLG2 (PSD93), DLG3 (SAP102) and DLG3 (PSD95). All 4 DLG proteins contain the characteristic C-terminal PSG tandem with two further PDZ domains, while DLG1 also contains an additional N-terminal L27 domain (Feng and Zhang, 2009; Zhu et al., 2016). The genes encoding the DLG proteins are highly conserved between human and mice and are paralogues of invertebrate *Dlg* with which they share a relatively high degree of sequence similarity. Two pairs of vertebrate *DLG* genes (*DLG1-DLG4* and *DLG2-DLG3*) were generated by two rounds of whole genome duplication roughly 550 million years ago (Nithianantharajah et al., 2013) (Figure 1.4). Although initially believed to be a DLG protein, DLG5 was subsequently shown to contain both a CARD domain and a distinct phylogenetic origin to DLG1-4, identifying it instead as a member of the CARMA protein subfamily (Friedrichs et al., 2008).

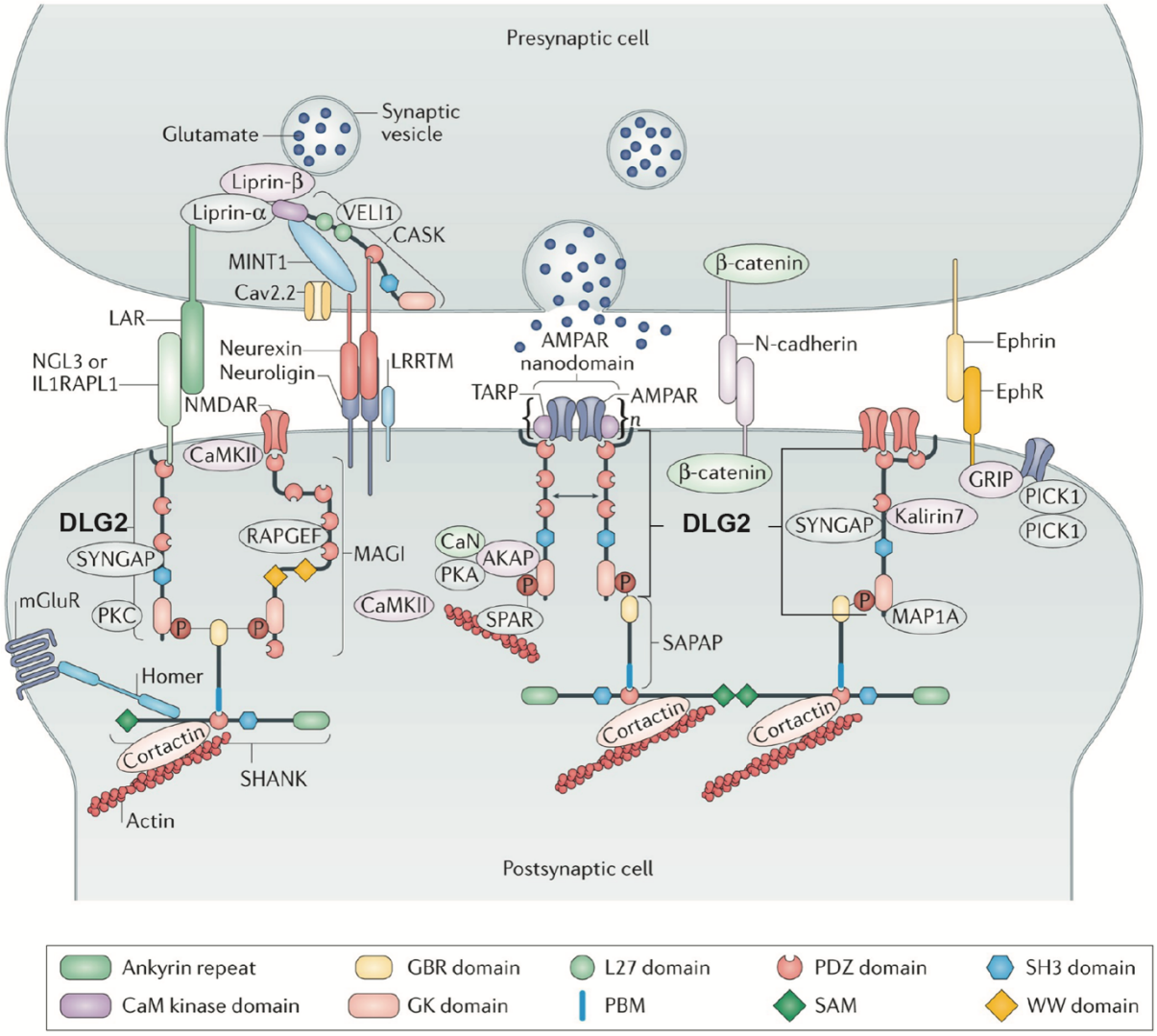


Figure 1.2 DLG2 scaffold mediated protein complexes in excitatory synapses
Protein complexes within the postsynaptic density and presynaptic active zone. DLG2 in common with other MAGUK proteins contains 1 GK domain, 1SH3 domain and 3 PDZ domains and acts as scaffold linking cell surface receptors with downstream signalling molecules. Key: AKAP, A-kinase anchor protein; CaMKII, calcium/calmodulin-dependent protein kinase II; CaN, calcineurin; Cav2.2, voltage-gated calcium channel subunit-α Cav2.2; EphR, ephrin receptor; GBR, GK-binding region; GK, guanylate kinase-like; GRIP, glutamate receptor-interacting protein; L27, LIN2–LIN7; LAR, leukocyte common antigen related; LRRTM, leucine-rich repeat transmembrane protein; MAGI, membrane-associated guanylate kinase inverted; MAP1A, microtubule-associated protein 1A; mGluR, metabotropic glutamate receptor; NGL-3, netrin-G3 ligand; NMDAR, NMDA-type glutamate receptor; PBM, PDZ-binding motif; PDZ, PSD95–DLG1–Zonula occludens 1; PKA, protein kinase A; PKC, protein kinase C; PICK1, protein interacting with C kinase 1; SAM, sterile α-motif; TARP, transmembrane AMPAR regulatory protein; VELI1, vertebrate lin-7 homologue 1. From (Zhu *et al.*, 2016).

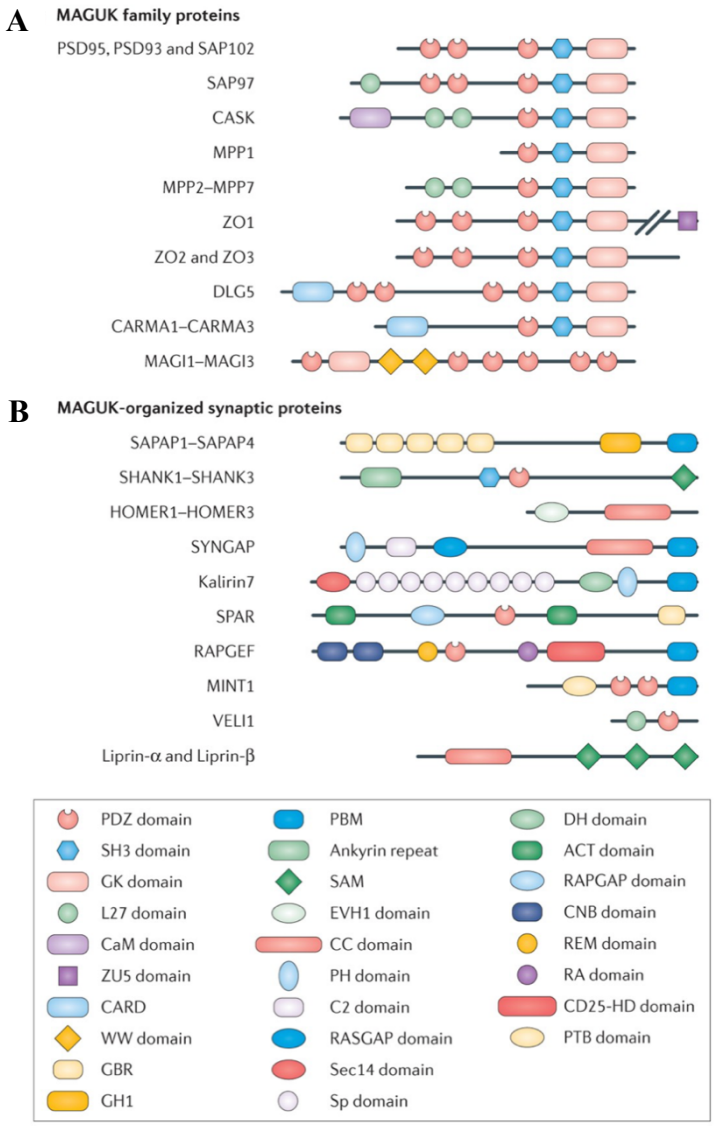


Figure 1.3 Domain organisation of MAGUKs and their related synaptic proteins
(A) Schematic diagrams of the domain organization of the membrane-associated guanylate kinase (MAGUK) family proteins. **(B)** Domain organization of selected proteins forming MAGUK-assembled synaptic complexes. Key: CaMK, calcium/calmodulin-dependent protein kinase; CARD, caspase activation and recruitment domain; CARMA, CARD-containing MAGUK protein; CASK, calcium/calmodulin-dependent serine protein kinase; CC, coiled-coil; CDC25-HD, RASGEF domain; CNB, cyclic nucleotide binding; DH, DBL homology; DLG, Discs large homologue; EVH1, Enabled/VASP homology 1; GBR, GK-binding region; GH1, GKAP homology domain 1; L27, LIN2–LIN7; PBM, PDZ-binding motif; PH, pleckstrin homology; PSD, postsynaptic density protein; PTB, phosphotyrosine-binding; RA, RAS-associating; RAPGAP, RAP GTPase-activating protein;RAPGEF, RAP guanine nucleotide exchange factor; RASGAP, RAS GTPase-activating protein; REM, RAS-GEF-like; SAM, sterile α-motif; SAP, synapses-associated protein; SAPAP, SAP90/PSD95-associated protein; Sec14, Sec14p-like; SH3, SRC homology 3; SHANK, SH3 and multiple ankyrin repeat domains protein; Sp, spectrin-like; SPAR, spine-associated RAP GTPase activating protein; VELI1, vertebrate lin-7 homologue 1; ZO, Zonula occludens; ZU5, ZO1–UNC5. From (Zhu *et al.*, 2016).

Due to the architecture of DLG2 and the MAGUK family more generally, with all domains functioning as protein-protein interaction modules, these proteins can interact with a wide variety of targets (Figure 1.3 B) and are ideally suited to act as a scaffold for the assembly of varied signalling complexes. Details of specific interaction targets and biological processes regulated by DLG2 are discussed in the following section (see section 1.5); however, when considered more

generally many transmembrane proteins including NMDA receptors, α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic (AMPA) receptors, ion channels and adhesion molecules contain PDZ domain binding motifs (PBMs) on their cytoplasmic C-terminal tails and are known targets of MAGUK proteins (Zhu et al., 2016; Won et al., 2017). All MAGUK proteins carry one or more PDZ domain and interaction between these domains and PBMs of interacting proteins has been shown to play a key role in the clustering, targeting and degradation of transmembrane proteins within the PSD (Feng and Zhang, 2009; Xu, 2011; Huganir and Nicoll, 2013; Chen et al., 2015). In addition many proteins also bind the GK domain of MAGUKs either independently or as part of an integrated structural unit with the SH3 domain, the SH3-GK supramodule (Zhu et al., 2012, 2016), although catalytically inactive the GK domain does conserve the residues necessary to bind the phosphate group of GMP (McGee et al., 2001) and this enables MAGUKs to specifically recognise target proteins in a phosphorylation dependent manner (Zhu et al., 2011).

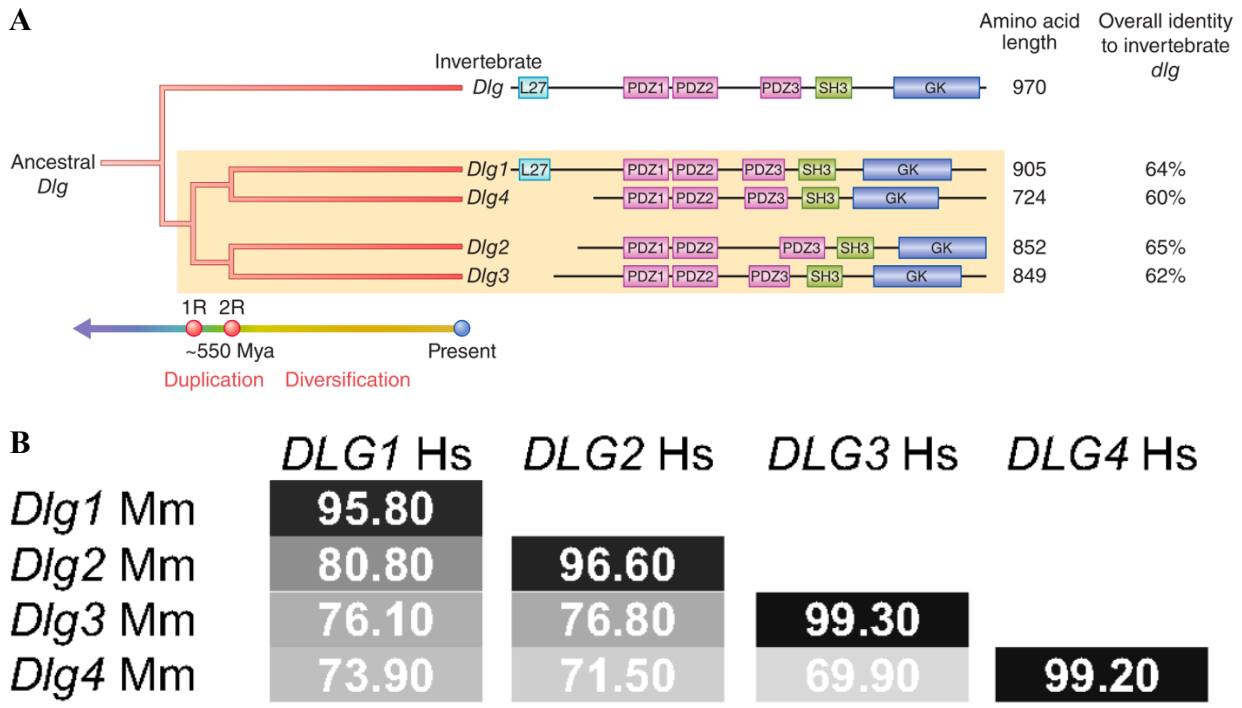


Figure 1.4 *Dlg* paralogues
(A). Comparison of invertebrate *Dlg* and four vertebrate paralogs (*Dlg1–Dlg4*) from the mouse. Two pairs of vertebrate *Dlg* genes can be identified, reflecting their evolutionary origins in the two rounds (1R, 2R) of whole genome duplication at the base of the chordates ~550 Mya. Yellow box highlights the four vertebrate Dlg proteins, showing high conservation of domain architecture. L27: domain in receptor targeting proteins Lin-2 and Lin-7; PDZ: PSD-95, Dlg, ZO-1/2 domain; SH3: Src-homology type 3 domain; GK, guanylate kinase–like domain. **(B)** Sequence comparison of mouse *Dlgs* and human *DLGs*. Percentage identity of human (Hs) and Mouse (Mm) aligned *Dlg* genes**.** From (Nithianantharajah *et al.*, 2013).

As examples of proteins that interact with and the processes regulated by MAGUKs, the first and second PDZ domains (PDZ1 and PDZ2) of DLG4, one of the more studied MAGUKs, are known to directly bind the PBMs in the NR2 subunit of the NMDA receptor (Kornau et al., 1995). DLG4 can also interact indirectly with AMPA receptors by binding the PBMs of transmembrane AMPA receptor regulating proteins (TARPs) such as stargazin (Schnell et al., 2002; Hafner et al., 2015). These interactions are required to cluster and anchor NMDA and AMPA receptors within the PSD thereby playing a key role in synaptic signalling and maturation (Chen et al., 2015; Levy et al., 2015; Won et al., 2017). Interactions between MAGUK proteins and cell adhesion molecules including neurexins and neuroligins, which link the presynaptic and postsynaptic membranes by bridging the synaptic cleft, may provide a direct link between neurotransmitter release at the presynaptic active zone and neurotransmitter receptor clustering at the PSD (Yamagata et al., 2003; Benson and Huntley, 2012; Thalhammer and Cingolani, 2014). The PDZ domains of both DLG4 and CASK can bind the PBMs of neurexins and neuroligins (Irie et al., 1997; Iida et al., 2004; Li et al., 2014b; Reissner and Missler, 2014) while the CaM kinase domain of CASK also binds liprin-α, a scaffolding protein in the presynaptic active zone required for the assembly of signalling complexes regulating neurotransmitter release (Wei et al., 2011; Südhof, 2012). MAGUKs have also been shown to regulate adhesion more generally, particularly at tight junctions where they anchor adhesion proteins to maintain strong cell-cell contact (Funke et al., 2005) as well as being the target of proteases that cleave MAGUKs prior to cell detachment (Ivanova et al., 2011).

Numerous MAGUK target proteins have been identified outside of the membrane region including several with enzymatic activity (Figure 1.3 B) (Feng and Zhang, 2009). In the PSD these includes the MAGUK interacting enzymes SYNGAP, SPAR, RAPGEF and Kalirin7 which are small GTPase regulating factors, thus providing a link between synaptic transmission and downstream G-protein coupled signalling events (Kim et al., 1998b; Boeckers, 2006; Zhu et al., 2016). Other non-membrane targets include scaffold proteins without enzymatic activity many of which provide links to the actin cytoskeleton, a dynamic structure that both stabilises the PSD while being actively remodelled in response to neuronal activity, in what is believed to be the basis of synaptic plasticity (Hotulainen and Hoogenraad, 2010; Caroni et al., 2012; Bosch et al., 2014; Zhu et al., 2016). MAGUKs may also play key roles in various signalling pathways regulating cell polarity, cell-cell adhesion, proliferation and migration through interaction with the Scribble cell polarity module; in invertebrates Dlg is known to bind Scribbled (Scrib) with this role believed to occupied in humans by DLG1 binding the scribbled homologue (SCRIB) (Humbert et al., 2008; Massimi et al., 2008; Stephens et al., 2018). Finally the MAGUK CASK is also able to act as a transcriptional regulator, regulating cell proliferation through an interaction with Inhibitor Of DNA binding 1 (ID1) (Qi et al., 2005), axon outgrowth and branching through interaction with the transcription factor CTIP1

(Kuo et al., 2010) and multiple aspects of neuronal development through interaction with the key neuronal transcription factor TBR1 (Hsueh et al., 2000; Wang et al., 2004; Huang and Hsueh, 2017).

1.6. DLG2 in the central nervous system

Following its initial isolation and characterisation DLG2 protein expression has been detected in excitatory synapses throughout multiple adult brain regions including the cortex, hippocampus, striatum and cerebellum (Brenman et al., 1996, 1998; Kim et al., 1996; Scannevin and Huganir, 2000; del Rio and Feller, 2006; Chen et al., 2015). At the mRNA level *in situ* hybridisation has shown *DLG2* expression throughout mouse brains at various stages of development from embryonic day 18 to postnatal day 56, interestingly *DLG2* expression appears elevated during early development up to postnatal day 7, in contrast to relatively low levels in adult brains (Yoo et al., 2020). Outside of the central nervous system (CNS) research into DLG2 has focused on its role as a tumour suppressor (Shao et al., 2019; Keane et al., 2020); however, within the CNS DLG2 along with other MAGUK proteins is known to act as a scaffold to regulate neurotransmitter receptor and ion-channel clustering as well as facilitate intracellular trafficking through associations both with adaptor proteins and the actin cytoskeleton (Feng and Zhang, 2009; Zhu et al., 2016).

DLG2 has been shown to regulate the targeting and clustering of AMPA receptors within the PSD to control the strength of excitatory synaptic transmission (Elias et al., 2006). This has been linked to a form of homeostatic plasticity termed synaptic scaling, a mechanism whereby the amplitude of miniature excitatory postsynaptic currents (mEPSC) is bidirectionally moderated to stabilise neuronal firing (Turrigiano, 2008); with DLG2 shown to regulate the recruitment of AMPA receptors to newly formed but functionally silent dendritic spines, enabling the generation of functional synapses (Levy et al., 2015). Although structurally similar to other members of the DLG subfamily of MAGUK proteins, studies have shown that DLG2 does play a distinct role during synaptic scaling, unlike DLG4 whose abundance appears to regulate the bidirectional accumulation or loss of AMPA receptors during synaptic scaling DLG2 regulates synaptic up-scaling only (Sun and Turrigiano, 2011). A recent study also indicating that DLG2 and DLG4 play opposing roles in regulating silent synapse maturation, inhibiting and promoting respectively, with the correct balance between these two proteins required for the fine-tuning or cortical networks (Favaro et al., 2018). DLG2 is also the most variable paralogue of the DLG proteins existing in at least six different N-terminal isoforms, with evidence to suggest that these isoforms differentially regulate the strength of AMPA receptor mediated transmission providing an additional mechanism for the fine-tuning of synaptic signalling (Kruger et al., 2013).

As well as regulating homeostatic synaptic plasticity through the AMPA receptor composition of dendritic spines, DLG2 has key roles in the assembly, localisation and function of NMDA receptors.

DLG2 acting in tandem with DLG4 is required for the assembly of NMDA receptor supercomplexes in the adult brain, triggered by activity dependent synapse maturation (Frank et al., 2016; Zeng et al., 2018). DLG2 has also been shown to regulate the localisation of NMDA receptors to specific membrane subdomains of the PSD (Tao et al., 2003; Delint-Ramirez et al., 2010) as well as regulate NMDA receptor function by mediating their tyrosine phosphorylation by Src family protein kinases (SFKs) such as Fyn, a process known to upregulate NMDA receptor function (Sato et al., 2008). Although research into the role of DLG2 in excitatory synapses has focused on AMPA and NMDA receptor containing glutamatergic synapses, it is also expressed in cholinergic synapses where it colocalises with neuronal nicotinic acetylcholine receptors (nAChRs) and plays a key role in synaptic stability (Parker et al., 2004). Despite its extensive expression in excitatory synapses it is possible DLG2 may not be required for normal synapse formation and function in all brain regions; one study has shown that although expressed in the cerebellum DLG2 knock-out rodent models show no structural or function abnormalities in parallel fibre synapses (McGee et al., 2001).

In addition to synaptic scaling DLG2 has also been shown to regulate another aspect of synaptic plasticity, LTP (Carlisle et al., 2008), which involves the strengthening of synapses due to high levels of stimulation resulting in a persistent increase in signal transmission between two neurons (Bliim et al., 2016). LTP has been shown to involve NMDA receptor mediated signalling (Kerchner and Nicoll, 2008), the recruitment and removal of AMPA receptors from the PSD (Malinow and Malenka, 2002) and the reorganisation of the actin cytoskeleton (Cingolani and Goda, 2008), all processes with known DLG2 involvement as previously described (Elias et al., 2006; Sato et al., 2008; Feng and Zhang, 2009; Delint-Ramirez et al., 2010; Sun and Turrigiano, 2011; Levy et al., 2015; Frank et al., 2016; Zhu et al., 2016; Ye et al., 2018). DLG2 has been shown to facilitate LTP induction in rodent models while having no effect on long LTD, in what appears to be a distinct role for DLG2; in contrast DLG4 facilitates LTD and disrupts LTP (Carlisle et al., 2008). The processes required for the induction and maintenance of LTP involve the remodelling of existing synaptic proteins, protein synthesis from existing mRNA transcripts and the synthesis of new mRNA transcripts via the activation of various downstream signalling pathways regulating transcription factors (Figure 1.5) (Bliim et al., 2016, 2019).

Given its role in neurotransmitter receptor recruitment within the PSD it is perhaps unsurprising that *DLG2* knockdown using RNA interference (RNAi) has been shown to disrupt neuronal networks *in vitro* (MacLaren et al., 2011). These studies performed using mouse hippocampal neurons cultured on multi-electrode arrays (MEAs) which allow the pattern of spontaneously generated action potentials to be recorded. Results showed that while DLG2 deficient neurons demonstrated hyperactive signalling compared to controls as determined by an increase both the

frequency of synaptic transmission and the duration of synaptic depolarisation the overall network was more disordered with a significant reduction in the synchronicity of neuronal firing.

Outside of the synapse DLG2 expression has also been detected in the axon initial segment (AIS), the site of action potential (AP) initiation, where it colocalises with the cell adhesion molecules TAG-1 and Caspr2 as well as several voltage gated potassium channel (Kv) subunits and has been shown *in vitro* to mediate the localisation of Kv1 channels to the AIS (Ogawa et al., 2008). However, *In vivo* studies using *Dlg2*$^{-/-}$ mouse models have shown that DLG2 is not required for Kv1 channel localisation with the motor neurons AIS, suggesting a possible compensatory mechanism exists (Duflocq et al., 2011). Dlg2 similarly colocalises with Kv channels in the juxtaparanodes of myelinated axons and is recruited to these location by a disintegrin and metalloproteinase 22 (ADAM22), although once again mouse models demonstrate that DLG2 is not required for Kv channel clustering within juxtaparanodes (Horresh et al., 2008; Ogawa et al., 2010; Duflocq et al., 2011).

As described previously (section 1.4) although DLG2 contains no catalytically active domains it does possess multiple sites for protein-protein interaction making it ideally suited to act as a scaffold for the assembly of larger protein complexes, these include the 3 PDZ domains as well as the GK and SH3 domains, which are believed to directly interact to form a SH3-GK tandem that function as a distinct supramodule (Feng and Zhang, 2009). Such interactions in include the binding of GluN2 subunits of NMDA receptors to the PDZ domains of DLG2, although this does not appear to be via the canonical PDZ-ligand of GluN2 as previously believed (Frank et al., 2016). Other membrane bound protein complexes share similar C termini to NMDA receptor subunits and also bind to DLG2 via their PDZ domains, such as Kv channels and Ca^{2+} ATPases (Kim et al., 1998a; Ogawa et al., 2008), while AMPA receptors bind to DLG2 PDZ domains via stargazin or related transmembrane AMPA receptor regulatory proteins (TARPs) (Dakoji et al., 2003). The GK domain of DLG2 (or SH3-GK supramodule) is known to interact with components of the cytoskeleton (Brenman et al., 1998; Zhu et al., 2016) and through similarity with other DLG subfamily members and its invertebrate orthologue potentially other signalling complexes, such as the SCRIB cell polarity module (Mathew et al., 2002; Zhu et al., 2014). Conformational changes that occur within the SH3-GK supramodule upon ligand binding to the PDZ3 domain can also promote the formation of DLG2 multimers, which include both DLG2 homomultimers and DLG2-DLG4 heteromultimers, suggesting a mechanism for the cooperative assembly of large signalling complexes.

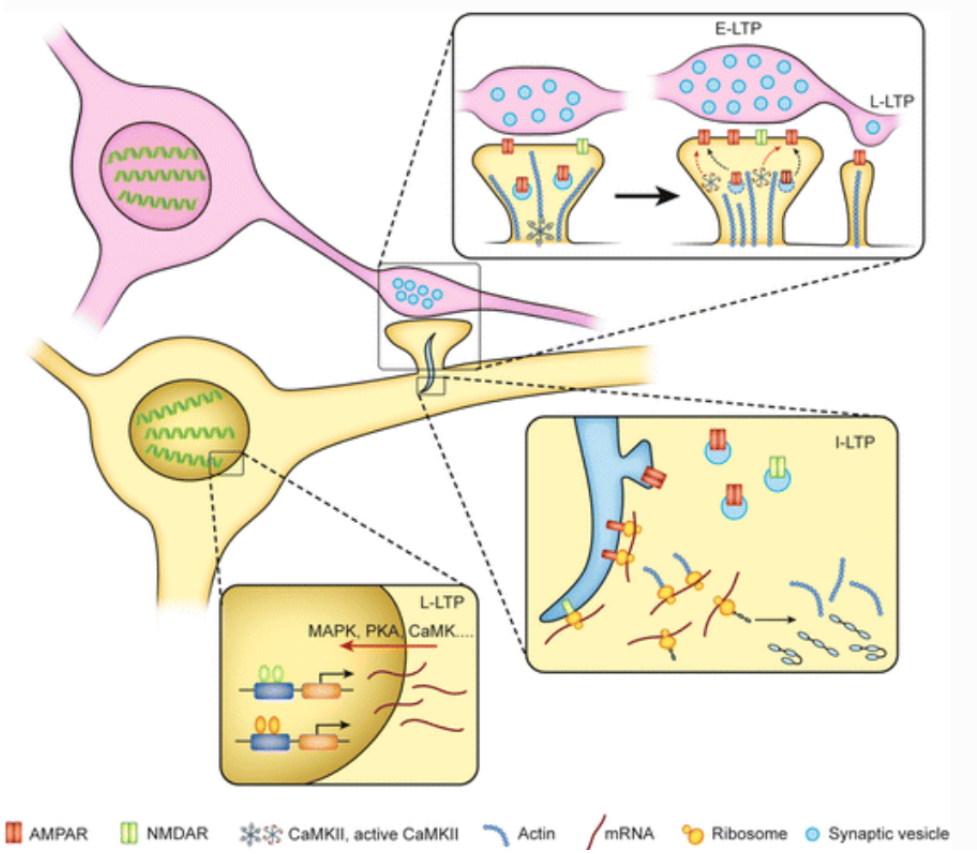


Figure 1.5 Mechanisms of long-term potentiation
A schematic representation of the processes required for the induction and maintenance of the different forms of LTP. Only some examples are shown for simplicity. Early phase LTP (E-LTP) depends on remodelling of the pre-existing proteins, including post-synaptic accumulation and activation of CaMKIIα, insertion of AMPA receptors to the post-synaptic membrane, and their phosphorylation by CaMKIIα. Intermediate phase LTP (I-LTP) depends on the protein synthesis from mRNA transcripts pre-existing at synapses, and may be associated with specialized organelles, such as spine apparatus. Late phase LTP (L-LTP) depends on the synthesis of new mRNA transcripts in response to activation of a number of enzymes and downstream transcription factors. From (Bliim *et al.*, 2016).

1.7. Cortical development in rodents and primates

The cerebral cortex is the outermost layer of neural tissue in the brain and is formed primary of the neocortex, a region of tightly compacted cells which form a highly organised six-layered structure, with layer I closest to the outer (pial) surface while layer VI has the deepest location (Gilmore and Herrup, 1997). A diverse assortment of neuronal subtypes populate the laminar structure of the neocortex which form distinct populations with unique morphological features, electrophysiological properties and gene expression profiles (Migliore and Shepherd, 2005; Molyneaux et al., 2007; Kumamoto and Hanashima, 2014; Lodato and Arlotta, 2015). Despite the clear differences in volume between rodent and primate (including human) brains they do share numerous similarities in structure and function which includes both the layered structure of the cortex, its patterns of neuronal connectivity and division into functionally distinct areas (Mason and Price, 2016).

Many aspects of cortical development are also conserved between mammalian species including the initial formation of the cortex as a thin layer of pioneer neurons beneath the pial surface termed the preplate; the migration of future layer VI neurons into the preplate causing it to split into an upper marginal zone (MZ) and lower subplate (SP) being first step in the formation of the laminar structure (Olson, 2014). As subsequent populations of neurons are born they migrate though the subplate and earlier born neurons coming to rest under the marginal zone, resulting in a layered structure with neurons stratified by birthday; the oldest neurons those in layer I (derived from the marginal zone) and the deep portion of layer VI (derived from the subplate), with the remaining cortical layers being generated in a sequential 'inside-out' fashion, with layer VI containing the oldest cells and layer II the youngest (Gilmore and Herrup, 1997). Once neurons arrive at their final position within the cortex they undergo terminal differentiation, including the extension of axons and dendritic arborisation, to form connections and generate the cortical circuity (Mason and Price, 2016). Within these cortical layers rodent and primate neurons appear to share similar patterns of gene expression, connectivity and function (Stiles and Jernigan, 2010; Hodge et al., 2019), providing further evidence that fundamental developmental mechanism are conserved between species.

Despite several conserved features in cortical development between rodents and primates there are notable dissimilarities between species in addition to the obvious differences in scale, such as the timings of cortical neurogenesis. In mice cortical neurons are produced from around 10 days post conception with the process continuing for roughly 8 days (Levers et al., 2001). In contrast primate cortical neurogenesis is greatly protracted, beginning at roughly 35 days post conception and continuing over a period of roughly 3 months (Bayatti et al., 2008; Bystron et al., 2008). In addition to the overall increase in the cortical neurogenesis period in primates the cell cycle length of their progenitor cells is also significantly increased and can be as much as 5 times longer than rodents at a comparable developmental stage (Breunig et al., 2011).

In both rodents and primates all projection neurons are derived from proliferating neuroepithelial stem cells (NSCs or NESCs) within the dorsal telencephalon at the anterior portion of the neural tube. After initial proliferative cell divisions to expand the NSC pool these cells give rise to more committed neural progenitor/precursor cells (NPCs) including radial glia (RG), long known to produce processes than can span the neuroepithelium and provide guidance for migrating neurons (Rakic, 1972, 1988). RG themselves proliferate through multiple rounds of cell division in the ventricular zone (VZ) lining the ventricle and as such are often referred to as ventricular RG (vRG) or apical progenitors (Lodato and Arlotta, 2015). These vRG undergo a process termed interkinetic nuclear migration in which their nucleus moves radially within the cytoplasm in phase with cell-cycle progression, with s-phase occurring at the basal edge of the VZ while mitosis occurs at the

apical surface (Florio and Huttner, 2014). As cortical development proceeds an increasing proportion of vRG divide asymmetrically to produce other cells types including cortical projection neurons, although only 10-20% of such neurons are directly generated from vRG by this process (Kowalczyk et al., 2009). Alternatively, vRGs can produce another type of NPCs which lose their apical attachment to the surface of the VZ and proliferate in a region superficial to it called the subventricular zone (SVZ), these progenitors being termed intermediate precursor cells (IPCs) (Haubensak et al., 2004; Miyata et al., 2004).

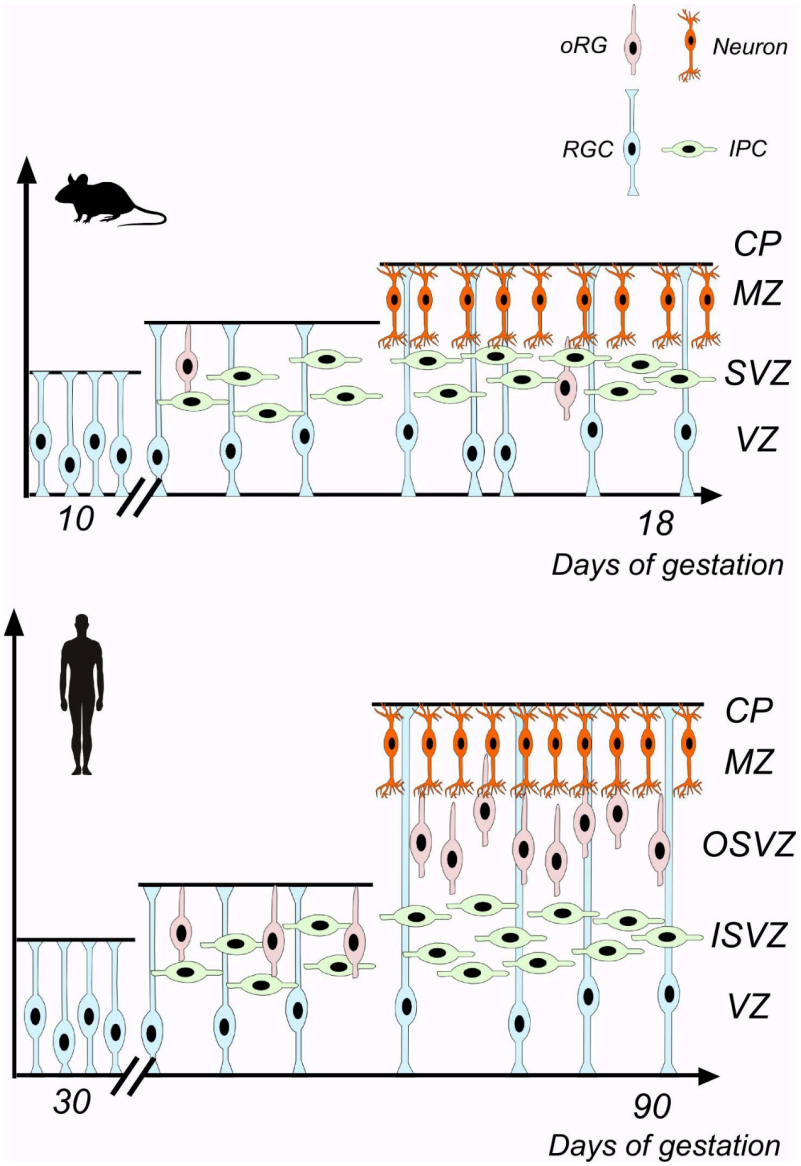


Figure 1.6 Comparison of key differences in mouse and human cortical development
Radial glial cells (RGCs) proliferate and create subpopulations of progenitors including IPCs and oRGs, before differentiating into neurons which migrate to the cortical plate. In humans there is a greater abundance of oRGs which begin to populate a distinct layer, the OSVZ, which is not present in mice. Key: VZ, ventricular zone; SVZ, subventricular zone; ISVZ, inner subventricular zone; OSVZ, outer subventricular zone; IZ, intermediate zone; CP, cortical plate; RGC, radial glia cell; IPC, intermediate progenitor cell; oRG, outer radial glial cell. From (Marshall & Mason, 2019).

Following the generation of IPCs differences begin to emerge between cortical development in rodents in rodents and primates. In rodents, IPCs within the SVZ undergo limited proliferative divisions and predominantly divide symmetrically to produce two neurons which subsequently migrate along radial glia fibres to their final location within the cortical layers (Farkas and Huttner, 2008). In contrast primate IPCs undergo multiple proliferative divisions resulting in a greatly expanded SVZ compared to rodents, which is further divided into an inner SVZ (ISVZ) and outer (OSVZ) (Figure 1.6) (Dehay et al., 2015). It is the primate ISVZ that is most comparable to the rodent SVZ and is comprised mainly of IPCs which are broadly comparable between species; however, the OSVZ contains a different progenitor type with similar properties to vRG but lacking connections to the apical surface which have been termed outer RG (oRG) (Hansen et al., 2010; Florio and Huttner, 2014; Mason and Price, 2016). Although small numbers of oRG have been detected in the rodent SVZ this cell type makes up roughly half of the progenitor pool within the primate OSVZ, here they undergo both proliferative divisions and asymmetric divisions that produce one daughter oRG and one IPC (Dehay et al., 2015). It is the OSVZ that from mid stage of cortical development onwards contains the majority of progenitors and is the major source of neurons for the supragranular (upper) cortical layers, which also show a proportionally greater enlargement in primates than the infragranular (lower) cortical layers (Hansen et al., 2010; Betizeau et al., 2013; Dehay et al., 2015; Mason and Price, 2016).

As briefly mentioned previously during early neurogenesis vRG extended fibres which attach to the pial surface and form a continuous scaffold across the cortex, used by neurons following differentiation to migrate to their final destination at the cortical plate (CP) (Rakic, 1972, 1988; Tan and Shi, 2013). Although this continuous scaffold plays a key role in the formation of the infragranular layers of the cortex it has been demonstrated that as human neurodevelopment progresses vRG lose their basal attachment to the pial surface to become cells with a distinct morphology and gene expression profile termed truncated RG (tRG) (Nowakowski et al., 2016). As tRG fibres terminate abruptly in the OSVZ, formation of the supragranular layers is not supported by a continuous scaffold produced by vRG but rather a discontinuous scaffold produced by the oRG within the OSVZ, which extend fibres to connect with the pia but have no contact with the VZ (Figure 1.7) (Nowakowski et al., 2016).

The neurons with the cortex can be broadly separated into two major classes, interneurons and projection neurons (Migliore and Shepherd, 2005; Molyneaux et al., 2007; Kumamoto and Hanashima, 2014; Lodato and Arlotta, 2015). Projection neurons form more distant connections, projecting axons to different brain regions or more caudal targets such as the spinal cord, are excitatory, use glutamate as neurotransmitter and have a typical pyramidal morphology. Interneurons are primarily inhibitory in nature, use γ-aminobutyric acid (GABA) as a

neurotransmitter and make more local connections within the cortex. The broad similarities in cortical development between rodents and primates previously described reflect the generation of cortical projection neurons with the situation for interneurons being less well resolved. In rodents interneurons are not generated within the cortex itself but rather in distant subcortical domains, mainly the caudal and medial ganglionic eminences, and subsequently migrate tangentially to infiltrate the developing cortex while maintaining a laminar organisation (Gelman and Marín, 2010; Silva et al., 2019). In primates the situation is less clear as several studies have indicated that while a proportion of neurons are generated in subcortical domains in a similar manner to rodents others are produced within the SVZ of the cortex itself (Zecevic et al., 2005; Radonjić et al., 2014); however, these finding are not consistent with some studies showing no evidence of interneuron production within the primate cortex (Hansen et al., 2013).

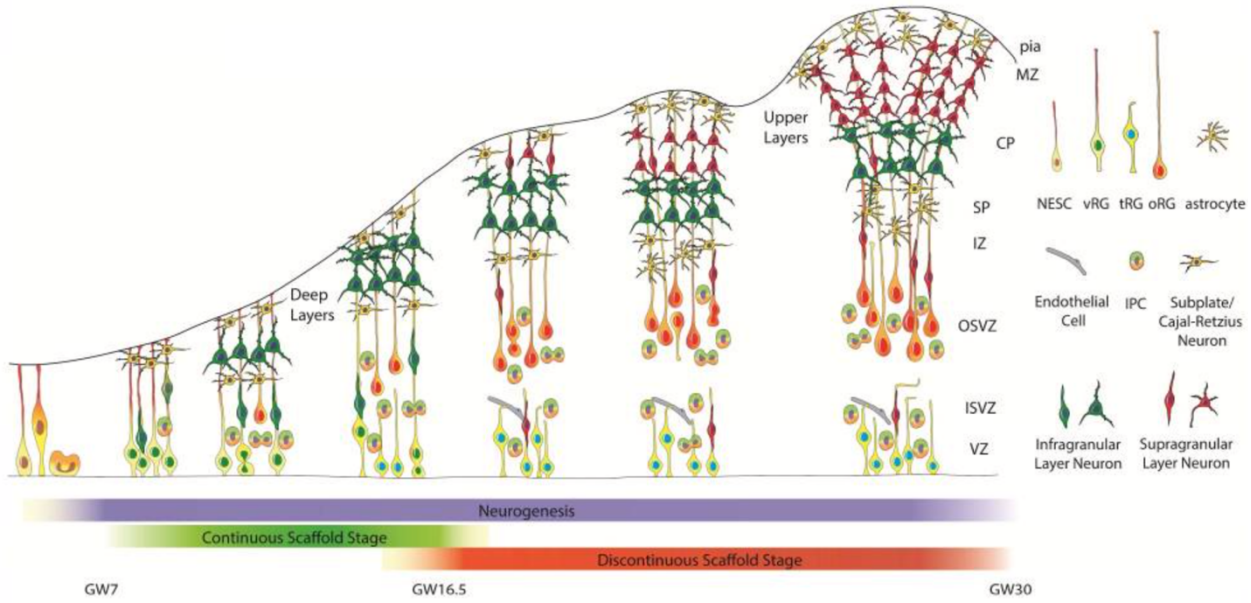


Figure 1.7 Expansion of supragranular layers is supported by a discontinuous scaffold

During early neurogenesis (left), basal fibres of vRG contact the pial surface and new-born neurons migrate along ventricular as well as oRG glia fibres. During late neurogenesis (right), vRG detached from the pial surface to form tRG and new-born neurons reach the cortical plate only along outer radial glia fibres. VZ – ventricular zone, NESC – neuroepithelial stem cell, ISVZ – inner subventricular zone, OSVZ – outer subventricular zone, IZ – intermediate zone, SP – subplate, CP – cortical plate, MZ – marginal zone, IPC – intermediate progenitor cell, vRG – ventricular radial glia, tRG – truncated radial glia, oRG – outer radial glia, GW – gestation week. From (Nowakowski *et al.*, 2016).

1.8. Modelling cortical projection neuron development using hPSCs

Evidence suggests a potential role for the perturbation of normal cortical development in the aetiology of schizophrenia, although little is known regarding the specific neurodevelopmental processes disrupted during the disease. Modelling of cortical neurodevelopment could therefore provide insights into the neurobiology underlying schizophrenia; this includes cortical projection neuron development, as these excitatory neurons are a key component of cortical function and circuit formation and many schizophrenia common genetic risk variants known to impact gene expression in these cells and their progenitors. Several systems exist for modelling cortical projection neuron development, of which human pluripotent stem cells (hPSCs) offer some advantages over animal model systems such as the mouse; beyond the fact that they are modelling human rather than murine cortical development, a process that although broadly conserved between species has certain dissimilarities as previously described (Zhu and Huangfu, 2013; Zhao and Bhattacharyya, 2018). hPSCs are relatively cheap to maintain and easy to handle; they can be reliably differentiated into various cell types including cortical projection neurons using well established protocols (Chambers et al., 2009; Cambray et al., 2012); during differentiation cells of interest are readily accessible at all times including the earliest developmental time points; and of course their use is consistent with the 3R approach (Replacement, Reduction and Refinement) to animal research (https://www.nc3rs.org.uk/the-3rs).

hPSCs can be split broadly into two categories, human embryonic stem cells (hESCs) are derived from the inner cell mass of a blastocyst while human induced pluripotent stem cells (hiPSCs) are derived from somatic cells that have been reprogrammed into an embryonic-like pluripotent state. Patient derived iPSCs are primarily of use as a disease model (Rowe and Daley, 2019); however, as these likely contain multiple genetic risk factors distributed throughout the genome, as found in patients with neuropsychiatric diseases such as schizophrenia (Bray and O'Donovan, 2018), comparison to controls and the identification of processes regulated by a single gene is problematic. In contrast hESCs offer an isogenic background for control purposes and with the use of genome editing to introduce a mutation into an hESC line, processes regulated by a specific gene of interest can be determined without additional confounding factors.

As described previously hPSCs can be differentiated towards cortical projection neurons using various well-established protocols which enables samples to be taken and phenotypes characterised from multiple cell types across the neurodevelopmental time course. Broadly speaking two approaches exist for neuronal differentiation from hPSCs *in vitro*, developmental patterning is a slower method in which cells are exposed to various morphogens that broadly replicate the conditions of *in vivo* cortical development (Chambers et al., 2009; Cambray et al., 2012) and reprogramming, where various neurogenic transcription factors are expressed within hPSCs causing

a more rapid neuronal differentiation (Zhang et al., 2013; Mertens et al., 2016; Nehme et al., 2018). Most *in vitro* developmental patterning protocols in use are based on the dual SMAD inhibition method developed by Chambers et al., in which Noggin (or LDN193189) and SB431542 are used to inhibit bone morphogenic protein (BMP) signalling via the canonical SMAD dependent pathway (Chambers et al., 2009, 2012), similar to *in vivo* development where the inhibition of BMP signalling differentiates neuroectoderm from ectoderm (Bond et al., 2012). After a period of neural induction the SMAD inhibitors are removed from culture, as once a neural fate has been specified BMP signalling is required for neurogenesis (Tozer et al., 2013) a process that can be induced by the addition of retinoic acid (RA) to cultures, again mimicking the processes of *in vivo* neurodevelopment where RA signalling promotes neuronal differentiation and maturation (Maden, 2007). Although developmental patterning is a relatively slow method it does recapitulate key stages of neurodevelopment, such as the NPC stage, that are greatly accelerated or entirely bypassed with a reprogramming approach and as such can identify key developmental and disease relevant phenotypes that would otherwise be lost (Schafer et al., 2019)

The well-established protocols for the generation of cortical projection neurons previously described were designed for monolayer (2D) culture system. 2D systems have many advantages including time, cost and accessibility although the primary advantage is that with appropriate protocols relatively homogenous populations containing predominately if not exclusively a given cell type of interest can be generated for study; such systems have been successfully used to investigate gene expression, cell morphology, electrophysiological properties and synapse formation among others (Quadrato et al., 2016; Hoffmann et al., 2019; Rowe and Daley, 2019). Recently developed 3D culture systems such as organoids hold the potential to more closely model neurodevelopment *in vivo* with anatomically relevant spatial information and cellular connections and such models have been shown to broadly recapitulate features of the developing brain and have been used to model several developmental and neuropsychiatric disorders (Quadrato et al., 2016; Hoffmann et al., 2019; Rowe and Daley, 2019) As such systems are designed to broadly model aspects of neurodevelopment *in vivo* (such as the developing forebrain) rather than produce a single cell type of interest (such as cortical projection neurons) they are by their nature more heterogenous and due to the greater degree of self-organisation in a organoids compared to 2D cultures, less replicable; organoids from a neural lineage will also lack vascularisation, which in such large 3D structures can lead to cell death within the interior (Quadrato et al., 2016; Hoffmann et al., 2019; Rowe and Daley, 2019). Therefore, although 3D systems have great potential for modelling human neurodevelopment and disease *in vitro,* 2D systems currently possess several advantages for modelling the development of a specific cell type.

While the neurons produced through any *in vitro* protocol are by no means identical to those of the adult brain, they share developmental trajectories and recapitulate key features of early human neuronal development; they possess many of the key components found in mature neurons and are able to form functional synapses and neuronal networks in a certain degree (Varrault et al., 2019). While the functionality of neurons generated *in vitro* may not be as developed as that found in adult human brains, these cells provide a reductionist model to assess the effects of mutations, drugs and other insults on the development and function of human neurons.

1.9. hPSC models of schizophrenia

Although the symptoms of schizophrenia do not typically manifest until early adulthood various lines of evidence from genetic, anatomical and epidemiological studies suggest that the disease has a neurodevelopmental component with its potential origins in the perturbation of early brain development by genetic and/or environmental risk factors (Owen et al., 2016). Despite the evidence for their existence the cellular and molecular mechanisms underlying disrupted neurodevelopment in schizophrenia are largely unknown and are the focus on ongoing study through multiple techniques: analysis of post-mortem tissue, brain imaging studies, animal models and cellular models (Moslem et al., 2019). While all of these approaches are complimentary each has particular advantages for the study of schizophrenia neurobiology; considering that many schizophrenia risk factors have been shown to act prenatally hPSC models are of particular use as they enable biological pathways acting during the earlier stages of human neurodevelopment and disrupted in schizophrenia to be more readily studied (Falk et al., 2016). Such models include schizophrenia patient derived hiPSC which can be differentiated towards disease relevant cell types such as neurons and glia, enabling the phenotypes of these cells to be characterised and the mechanisms underlying these phenotypes determined at various stages of neurodevelopment (Millan et al., 2016). Alternatively, to investigate the effect of a single identified genetic risk factor on schizophrenia neurobiology genetic engineering of hESC can be used to generate an isogenic cell line containing a mutation in the gene of interest, which can subsequent be differentiated and studied in a similar manner (Pak et al., 2015).

The use of hiPSC models of early neurodevelopment in schizophrenia has identified several cellular phenotypes in both NPCs which include defects in proliferation, adhesion and cytoskeletal organisation as well as neurons which show altered morphology, electrophysiological properties and synaptic connectivity (Moslem et al., 2019). Initial studies were performed using hiPSC generated from patients with idiopathic schizophrenia which were differentiated to projection neurons (Brennand et al., 2011), these showed reduced neurite branching in agreement with the reduced dendritic arborisation observed in post-mortem schizophrenic brains and animal models

(Selemon and Goldman-Rakic, 1999; Jaaro-Peled et al., 2010), reduced synaptic connections as well as reduced levels of the synaptic protein DLG4. More recently studies have used hiPSC from genetically defined schizophrenia cohorts with appropriate controls to further investigate the disease phenotype in various cells types at different stages of neurodevelopment (Hoffman et al., 2019). These include the early stages of differentiation prior to neurogenesis where NPCs derived from schizophrenia patients have been shown to have aberrant proliferation, migration, apical polarity and canonical Wnt signalling. At later stages of differentiation patient derived neurons, either excitatory projection neurons or inhibitory interneurons, generally display a less mature phenotype than controls particularly characterised by decreased synaptic connectivity and neurotransmitter release (Falk et al., 2016; Moslem et al., 2019).

Although SNPs collected account for up to 50% of schizophrenia genetic liability their individual effect size is small (Sullivan et al., 2012; Ripke et al., 2014; Pardiñas et al., 2018) making any dissection of their role in cell biology problematic. In contrast, CNVs although rare individually carry a relatively high risk of disease (Kirov et al., 2014; Marshall et al., 2017) and due to this greater effect size can be more readily studied using a model system, as such the majority of studies using hiPSC models of schizophrenia are focused on the role of CNVs in disease neurobiology (Figure 1.8). These include the 22q11.2 deletion that was one of the first CNVs associated with schizophrenia and remains the diseases strongest known molecular genetic risk factor, with roughly 25% of individuals with this deletion developing schizophrenia in later life (Van et al., 2017). Although the size of the 22q11.2 deletion can vary it is large and usually encompasses at least 30 genes, several of these have been identified as disease relevant although the size of CNV makes the contribution of any one gene to schizophrenia pathophysiology difficult to ascertain (Van et al., 2017). The study of 22q11.2 deletion hiPSCs has identified one gene of particular interest within the CNV, *DGCR8* which encodes a regulator of microRNA processing (Toyoshima et al., 2016). Within 22q11.2 deletion cells the expression of several microRNAs, which function in the post-transcriptional regulation of gene expression, were found to be significantly downregulated in agreement with their perturbed expression in tissue samples from post-mortem schizophrenic brains; the genes targeted by these microRNAs including those with associated with neuron development and maturation as well as synaptic transmission (Zhao et al., 2015). An RNAseq analysis of these cells also found significant differential gene expression as a result of the 22q11.2 deletion with gene related to cell cycle progression, cytoskeletal organisation and glutamate metabolism being significant down-regulated while genes related to apoptosis and the immune response were upregulated (Lin et al., 2016). This increase in apoptosis is believed to results from a disruption the MAPK signalling pathway and may potentially explain the reduced synaptic

numbers and dendritic arborisation seen in schizophrenia patients as well as the general decrease in neuronal and glial viability (Lin et al., 2016).

The 15q11.2 deletion is the CNV most frequently found in schizophrenia patients (Rees et al., 2014) and although it shows variable expressivity is generally associated with mild symptoms. This CNV covers 4 genes *NIPA1*, *NIPA2, CYFIP1* and *TUBGCP5* of which *CYFIP1* deletion likely has the greatest contribution to disease given the known interaction between CYFIP1 protein and fragile X mental retardation protein (FMRP), patients with fragile X syndrome and schizophrenia also sharing certain behavioural and cognitive symptoms (Farrell et al., 2020). hiPSC models of 15q11.2 deletion show significant differential gene expression in NPCs compared to controls and again the biological processes these genes appear to regulate are cell proliferation and cytoskeletal organisation (Nebel et al., 2016). At the protein level a role of CYFIP1 in the regulation of cytoskeletal dynamics is further supported by 15q11.2 deletion NPCs showing significant reduction in the levels of wave regulatory complex (WRC) components including WAVE1 and WAVE2, which have established function in the regulation of Arp2/3-mediated actin polymerisation and membrane protrusion formation (Yoon et al., 2014). During later differentiation 15q11.2 deletion neurons showed reduced DLG4 protein levels as well as qualitative differences in dendrite morphology compared to controls (Das et al., 2015).

Multiple studies have demonstrated a link between mutations in the *DISC1* gene and schizophrenia (Dahoun et al., 2017) while rodent models of *DISC1* mutation show several behavioural phenotypes associated with the disease including reduced social interactions and depressive like behaviour in addition to defects in neuronal morphology and connectivity (Johnstone et al., 2011; Tomoda et al., 2016). hiPSC models containing frameshift mutations in *DISC1* have shown a dysregulation of gene in NSCs known to regulate proliferation as well as a microRNA regulator of oligodendrocyte differentiation (Murai et al., 2016). *DISC1* mutant NPCs show decreased expression of key neurodevelopmental transcription factors such as FOXG1 and TBR2 as well as increased proliferation resulting from aberrant Wnt signalling (Srikanth et al., 2015), a finding supported my mouse models which show DISC1 regulates neural progenitor proliferation via modulation of the downstream effectors of canonical Wnt signalling GSK3β/β-catenin (Mao et al., 2009). During later differentiation *DISC1* mutant neurons show synaptic dysregulation, including differential expression of synapse associated genes, reduced numbers of synaptic boutons and reduced neurotransmitter release form synaptic vesicles (Wen et al., 2014).

Genes encoding members of the neurexin family of proteins which function as cell adhesion molecules and receptors of in the vertebrate nervous system (Reissner et al., 2013) have also been both associated with schizophrenia and instigated though hPSC models. These include hESC models of NRXN1 deletion, which when differentiated to neurons show normal synapse formation

but reduced neurotransmitter release with associated electrophysiological defects in synaptic transmission (Pak et al., 2015). hiPSC models carrying a deletion within another neurexin gene, *CNTNAP2,* also shown synaptic defects following neuronal differentiation including a dysregulation of synapse related genes and aberrant network activity (Lee et al., 2015; Flaherty et al., 2017). Overall hPSC models of schizophrenia have provided further support for the perturbation of early neurodevelopment in disease aetiology and has enabled several biological processes regulated by schizophrenia risk genes in both NPCs and neurons to be identified. However, challenges remain in understanding how these biological processes identified from cellular phenotypes *in vitro* correlate to the development of schizophrenia in patients.

CNV loci	Genetic mutation	Sample details	Reprograming method	Cell type	Phenotype	Reference
15q11.2	382 kb deletion at 15q11.2	SAD proband and unaffected mother carrying 15q11.2 deletion; 1 unrelated control	Sendai viral	NPC and neurons	Altered dendritic morphology; reduction in CYFIP1 and PSD-95 protein levels	(Das et al., 2015)
	15q11.2 microdeletion	3 SCZ patients carrying 15q11.2 deletion; 3 unrelated controls and 2 controls carrying DISC1 mutation	Episomal vectors or sendai viral	iPSC-derived NPC	NPC exhibit defects in maintenance of adherens junctions, apical polarity and WAVE complex stability	(Yoon et al., 2014)
		3 hiPSC-derived NPC lines with lentiviral shRNA targeting *CYFIP1*		*CYFIP1*^{KD} NPC	Differential expression of FMRP targets and postsynaptic density genes in *CYFIP1*^{KD} lines	(Nebel et al., 2016)
22q11.2	22q11 deletion	3 SCZ patients (1 carrying 22q11 deletion, 2 sporadic cases); 2 unrelated controls	Retroviral LTR	Neurons	Significant delay in reduction of endogenous OCT4 and NANOG expression during differentiation of iPSC-derived neurons	(Pedrosa et al., 2011)
	22q11.2 microdeletion	3 SCZ patients (with 2 COS, 1 VCFS); 3 SAD patients (with VCFS); 6 controls	Plasmids	Neurons	Multiple miRNAs differentially expressed between patient and control neurons, with targets of the differentially expressed miRNAs linked to SCZ	(Zhao et al., 2015)
	22q11.2 microdeletion	5 SCZ patients (with 4 COS, 1 VCFS); 3 SAD patients (with VCFS); 7 controls	Plasmids	Early differentiating neurons	Differentially expressed genes (DEGs) involved in MAPK signalling, cell cycle and apoptosis; 22q11.2 genes *CDC45* and *PRODH* linked to expression networks for many of the DEGs	(Lin et al., 2016)
	2.6-Mb hemizygous deletion at 22q11.2	2SCZ patients carrying 22q11.2 deletion; 2 controls	Retroviral	Neurons	Increased L1 retrotransposon copy number in SCZ patient neurons; increased insertion of L1 in genes related to synapses and schizophrenia	(Bundo et al., 2014)
	2.6-Mb hemizygous deletion at 22q11.2	2 SCZ patients carrying 22q11.2 deletion; 4 controls	Retroviral	Neurons and neurospheres	Reduced neurosphere size, neurite outgrowth and neural differentiation efficiency in patient-derived cells; reduced expression of miRNAs in miR-17/92 cluster (involved in cell proliferation and survival) and miR-106a/b (represses p38α, a regulator of gliogenic differentiation)	(Toyoshima et al., 2016)
DISC1	4 base-pair frameshift mutation	1 SCZ patient and 1 MD patient, both carrying DISC1 mutation; 2 related controls and 1 unrelated control; 1 TALEN-corrected *DISC1* line and 2 TALEN-introduced *DISC1* mutation line	Episomal vectors	Forebrain neurons	Dysregulation of genes associated with synapses, DISC1-interacting proteins and psychiatric-disorder-associated proteins; synaptic vesicle release deficits and reduction of synaptic boutons in patient neurons; isogenic lines show *DISC1* mutation is necessary and sufficient for the observed synaptic defects	(Wen et al., 2014)
	Frameshift mutation in *DISC1* exon 2 or exon 8	2 TALEN-induced *DISC1* frameshift mutation in exon 8 and 2 Cas9-induced *DISC1* frameshift mutation in exon 2; 1 healthy control	Lentiviral	NPC and dorsal neurons	Frameshift results in expression of a truncated DISC1 protein and loss of full-length protein; increased WNT signalling and altered transcriptional profile in NPC and neurons; decreased expression of NPC fate markers including *Foxg1* and *Tbr2*	(Srikanth et al., 2015)
	4 base pair frameshift mutation	2 SCZ patients carrying DISC1 mutation; 2 isogenic iPSC lines with DISC mutation; 3 healthy controls	Episomal vectors	NSCs	Increased expression of *miR-219* (promotor of oligodendrocyte differentiation) and decreased expression of *TLX* (regulator of NSC proliferation and self-processing) in DISC1-mutant cells	(Murai et al., 2016)
NRXN1	Heterozygous *NRXN1* deletion	2 H1 hESCs lines with mutations to *NRXN1* gene (1 with conditional KO of *NRXN1*, 1 with conditional truncation of *NRXN1*)	N/A	*Ngn2*-induced neurons	Decrease in presynaptic neurotransmitter release Decreased frequency of spontaneous mEPSCs and reduced amplitude of EPSCs in *NRXN1*-mutant neurons, but normal synapse number	(Pak et al., 2015)
CNTNAP2	293.3 kb heterozygous deletion in CNTNAP2 gene (exons 14–15)	1 SCZ patient carrying CNTNAP2 deletion; healthy father carrying CNTNAP2 deletion; 1 related control and 1 non-related control	Sendai viral	NPC, glutamatergic neurons and forebrain neurons	Differential expression of genes involved in synaptic transmission; altered neuronal activity	(Flaherty et al., 2017; I.S. Lee et al., 2015)

Figure 1.8 iPSC models of schizophrenia cases with a known genetic cause.
Key: CNV, copy number variation; DISC1, disrupted-in-schizophrenia 1; NRXN1, neurexin-1; SAD, schizoaffective disorder; SCZ, schizophrenia, NPC, neural progenitor cells; CYFIP1, cytoplasmic FMR1interacting protein 1; VCFS, velocardiofacial syndrome; MD, major depression; TALEN, transcription activator-like effector nucleases; iPSC, induced pluripotent stem cells; KO, knockout; LTR, long terminal repeats; NSCs, neural stem cells, ESC, embryonic stem cell.From (Moslemi *et al.*, 2019).

1.10. Hypothesis

It was hypothesised that in addition to DLG2s established role as a scaffolding protein with the PSD, DLG2-mediated signalling may regulate schizophrenia relevant aspects of early human cortical neurogenesis. There were several lines of reasoning behind this including multiple studies linking mutation in the *DLG2* gene to schizophrenia (Kirov et al., 2012b; Nithianantharajah et al., 2013; Pocklington et al., 2015) and related disorders including ID (Reggiani et al., 2017) and ASD (Ruzzo et al., 2019), with the penetrance of *DLG2* deletions in schizophrenia at 4.95% and in more severe developmental disorders at 2.07% (Supplementary Document 1). Schizophrenia has an established neurodevelopmental competent to its aetiology and although the mechanisms underlying this are not fully understood many genetic risk factors for the disease impact gene expression in the mid-foetal brain, particularly in NPCs and cortical excitatory neurons (Hill and Bray, 2012; O'Brien et al., 2018; Clifton et al., 2019; Polioudakis et al., 2019; Walker et al., 2019). Many hPSC models have shown that mutations at schizophrenia risk loci do impact early cortical neurogenesis resulting in abnormal NPC and neuronal phenotypes (Falk et al., 2016; Moslem et al., 2019); however, *DLG2* has yet been studied in this context with the focus of published research being rodent and human cognitive studies (Nithianantharajah et al., 2013) as well as network activity *in vitro* (MacLaren et al., 2011). Although a role for DLG2 in early cortical neurogenesis would be novel *in situ* hybridisation data has shown expression in mouse brains from embryonic day 18 (Yoo et al., 2020) suggestive of a role in early neurodevelopment while both the invertebrate orthologue of *DLG2* and well as mammalian MAGUK proteins with similar domain structures are known to regulate key developmental signalling pathways (Humbert et al., 2008; Zhu et al., 2016; Stephens et al., 2018).

To test this hypothesis, a WT control and two DLG2 knockout (*DLG2*$^{-/-}$) hESC lines were differentiated towards cortical excitatory neurons using a 2D developmental patterning protocol. The phenotype of cells were characterised at multiple stages of cortical differentiation using a range of experimental techniques including western blotting, immunocytochemistry, mass spectrometry, RNA sequencing and various phenotypic assays to determine both if DLG2 is expressed during early cortical neurogenesis and which biological processes DLG2-mediated signalling may regulate during this period. To determine the potential disease relevance of any such processes genes differentially expressed in *DLG2*$^{-/-}$ cells at 4 timepoints of cortical differentiation were tested for schizophrenia common variant enrichment.

2. Materials & methods

2.1. Cell maintenance

All hESC lines were maintained at 37°C and 5% CO_2 in 6 well cell culture plates (Greiner) coated for at least 1 hour at 37°C 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 EDTA 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 using the MycoAlert PLUS Mycoplasma Detection Kit (Lonza).

Human embryonic kidney 293 (HEK) cells were maintained at 37°C and 5% CO_2 in uncoated T75 tissue culture flasks (Sarstedt). Cells were fed on alternate days with complete media consisting of: 10% foetal bovine serum (BioSera) in G-MEM BHK-21 (Thermo Fisher scientific), 1x Pen Step Glutamine (Thermo Fisher Scientific), 1X MEM Non-essential amino acids (Thermo Fisher Scientific), 50 µM 2-Mercaptoethanol (Thermo Fisher Scientific) and 1mM sodium pyruvate (Thermo Fisher Scientific). Cells were passaged at 80% confluency using TrypLE Express (Thermo Fisher Scientific) for 3-5 minutes at 37°C followed by gentle manual dissociation. All cells were kept below passage 25 and confirmed as negative for mycoplasma infection.

2.2. Cortical differentiation

Differentiation to cortical projection neurons 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 by 1 hour 37°C incubation 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 by at least 2 hour 37°C

incubation 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 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 by at least 2 hour 37°C incubation with 10 µg/ml poly-d-lysine hydrobromide (PDL, Sigma) and 2 hour to overnight 37°C incubation with 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 coated plates 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.

2.3. Western blotting

Total protein was extracted from dissociated cultured cells by incubating in 1x RIPA buffer (New England Biolabs) with added 1x MS-SAFE Protease and Phosphatase Inhibitor (Sigma), 150ul per culture plate well, 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) with standard curves produced using Pierce Bovine Serum Albumin standards (Thermo Fisher Scientific) and a minimum of three replicates per sample. 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), run on ice at 120V for 1 hour 45 minutes using ice cold Bolt Transfer Buffer (Thermo Fisher Scientific). Transfer was confirmed by visualising protein bands with 0.1% (w/v) Ponceau S (Sigma) in 5% (v/v) acetic acid (Sigma) in water, 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 (see section 2.21 Key resources), 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).

2.4. Preparation of DLG2-FLAG isoforms

Six cDNA clones corresponding to protein coding DLG2 transcript variants within the pCMV6-Entry vector were obtained (Cambridge Bioscience, see section 2.21 Key resources – Recombinant DNA) , expression of these plasmids within mammalian cells resulting in C-terminal FLAG-tagged DLG2 isoforms. 5-alpha competent *E. coli* (New England Biolabs) were transformed with these 6 plasmids and empty vector negative controls. Transformed bacterial were plated into 90mm petri dishes (Scientific Laboratory Supplies) coated with LB Agar (Formedium) with 100 µg/ml ampicillin (VWR). Following incubation bacterial colonies were picked and grown in LB Broth (Formedium) with 100 µg/ml ampicillin (VWR) after which plasmids were extracted and purified using the PureYield Plasmid Maxiprep System (Promega). To produce DLG2-FLAG isoforms HEK cells were transfected with 4 µg of these plasmids in 64µl of Lipofectamine 2000 (Invitrogen), made up to 200µl with Opti-MEM reduced serum media (Thermo Fisher Scientific), following 72 hours incubation total protein was extracted and quantified as described in section *2.3 Western blotting*. Western blotting for both DLG2 protein and the FLAG tag was used to confirm successful expression of DLG2-FLAG isoforms.

2.5. CRISPR/Cas9 genome editing to generate tagged *DLG2*

A multiplex CRISPR/Cas9 vector containing 2 guide RNA (gRNA) expression cassettes and a Cas9 nuclease expression cassette was constructed essentially as described (Sakuma et al., 2014). Two oligonucleotides corresponding to gRNAs targeting the *DLG2* stop codon were designed and synthesised (Sigma) prior to annealing into either plasmids px330S[2] or px330S[3]. These vectors along with px330A-1x[2] were used in a "golden gate" assembly reaction to produce the multiplex CRISPR/Cas9 vector. Transfection of this vector along with a synthesised homology directed repair (HDR) template (GeneWiz), designed to insert a C-terminal FLAG tag to DLG2 and containing a puromycin resistance gene, into H9 hESCs was either with the 4D-Nucleofector system (Lonza), using P3 solution and CB150 programme, or Lipofectamine 2000 (Invitrogen). Selection for successfully edited cells began at 48 hours post transection using between 0.1 and 0.5 µg/ml puromycin over a period of 5 days.

In a second approach synthetic CRISPR RNA (crRNA) + transactivating crRNA (tracrRNA) were produced (sygRNA system – Sigma) and combined in equimolar quantities to produce gRNAs targeting the *DLG2* stop codon. Lipofectamine Stem Transfection Reagent (Invitrogen) was used to transfect these gRNAs into the Cas9 nuclease expressing iCas9 cell line (González et al., 2014)

following doxycycline induction. The Guide-It Long ssDNA Production System (Takara) was used to generate a single strand HDR template which was also transfected along with the gRNAs. This template being generated from a slightly modified version of the original double stranded HDR template, containing a blasticidin as opposed to puromycin resistance gene, due to the iCas9 cell line already possessing puromycin resistance. Selection for successfully edited cells began at 48 hours post transection using 0.5 µg/ml blasticidin (Thermo Fisher Scientific) over a period of 5 days.

In a final approach pairs of gRNAs, intended to function with a Cas9 D10A nickase, targeting either side of the DLG2 stop codon in a protospacer adjacent motif (PAM)-out configuration were designed synthesised as cRNA (IDT). cRNAs were annealed to tracrRNA (IDT) to generate gRNAs which were then complexed with Cas9 D10A nickase (IDT). Transfection of the generated ribonucleoprotein (RNP) complexes, a single strand HDR template and electroporation enhancer (IDT) into H9 hESCs was with using the 4D-Nucleofector system (Lonza), using P3 solution and CB150 programme. Selection for successfully edited cells began at 48 hours post transection using 0.5 µg/ml blasticidin over a period of 5 days. Further details of all approaches can be found in Chapter 3.

2.6. PCR & agarose gel electrophoresis

PCR primers were either obtained from papers as indicated or designed using the NCBI Primer-BLAST tool (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). Primers were ordered as lyophilised DNA oligonucleotides (Sigma) and resuspended at 100µM in 1X Tris-EDTA buffer solution (Sigma). PCR was performed using either MyTaq DNA Polymerase (Bioline) or PrimeStar Max DNA Polymerase (Takara) essential according to the manufacturer's protocols, with optimisation to ensure the resulting PCR products produce a clear single band of the expected size following agarose gel electrophoresis.

Agarose gels were prepared using 1% UltraPure Agarose (Invitrogen) in TAE buffer (Formedium), unless otherwise stated, with 1X SYBR Safe DNA Gel Stain (Invitrogen) to enable DNA visualisation. DNA samples were loaded with Gel Loading Dye Purple 6x (New England Biolabs) and run alongside Quick-Load Purple DNA ladder (New England Biolabs) of a size appropriate to the loaded sample, as indicated in figure legends. Gels were run in TAE buffer at 100V until the dye front had migrated through at last ¾ of the gel and DNA bands visualised and imaged using the iBright CL1500 Imaging System (Thermo Fisher Scientific).

2.7. TOPO cloning

Genomic regions of interest were amplified using PCR with appropriately designed primers prior to the separation of PCR fragments by agarose gel electrophoresis. Gel bands containing fragments of interest were excised and DNA extracted using the Nucleospin Gel and PCR Clean-up kit (Takara). Blunt-end cloning of purified PCR fragments was achieved using the Zero Blunt TOPO PCR Cloning Kit for Sequencing (Invitrogen) according to the manufacturers protocol. 5-alpha competent *E. coli* (New England Biolabs) were transformed with the assembled TOPO vectors then plated into 90mm petri dishes (Scientific Laboratory Supplies) coated with LB Agar (Formedium) with 100 µg/ml ampicillin (VWR). Following incubation bacterial colonies were picked and grown in LB Broth (Formedium) with 100 µg/ml ampicillin (VWR) after which plasmids were extracted and purified using the PureYield Plasmid Miniprep System (Promega). The resulting purified TOPO vectors containing inserted PCR fragments were sent for Sanger sequencing.

2.8. Sanger sequencing

Sanger sequencing of plasmids or PCR fragments was using either the GATC SUPREMERUN tube service (https://www.eurofinsgenomics.eu/en/custom-dna-sequencing/gatc-services/supremerun-tube/) or the LGC Ready2 Run service (https://www.biosearchtech.com/services/sequencing/sanger-sequencing/single-sample). Sample preparation and primer design was according to these companies' specifications.

2.9. *DLG2* knockout hESC line generation

Two *DLG2*$^{-/-}$ knockout hESC lines (KO1 and KO2) were generated by Dr Eunju Jenny Shin (Neuroscience and Mental Health Research Institute, Cardiff University) as described in (Sanders et al., 2020) and provided for this research along with a genetically WT sister line.

2.10. Genetic validation of *DLG2* knockout hESC lines

The gRNA pair used in *DLG2* knockout hESC line generation had zero predicted off-target nickase sites when used with a Cas9 D10A nickase (Figure 4.2). Even though a wild-type Cas9 nuclease (where only a single gRNA is required to create a double-stranded break) was not used, the predicted genic off-target sites for each individual gRNA were checked to thoroughly validate the lines. Following manual dissociation of WT and *DLG2* KO hESC into DPBS, genomic DNA was extracted using the ISOLATE II Genomic DNA kit (Bioline) and potential off target-sites investigated by PCR amplification and Sanger sequencing (GATC & LGC). Out of 30 sites identified, 14 were randomly selected (7 for each gRNA) for validation.

Genotyping using the Illumina PsychArray v1.1 was a performed by Alexandra Evans and Ngoc-Nga Vin (MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University) For this *DLG2* KO and WT DNA samples were amplified and fragmentated according to the associated Illumina HTS assay protocol and 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).

2.11. Immunocytochemistry (ICC)

Cells were fixed in 4% paraformaldehyde (PFA, Sigma) in PBS for 20 minutes at 4°C followed by a 1-hour room temperature incubation in blocking solution of 5% donkey serum (Biosera) in 0.3% Triton-X-100 (Sigma) in PBS (0.3% PBST). Primary antibodies, used at an assay-dependent concentration (see section 2.21 Key resources), 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. Following a final PBS wash stained cells were 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.

2.12. Peptide affinity pulldowns & mass spectrometry

PDZ domain containing proteins were enriched from total protein extracts by peptide affinity purification. NMDA receptor subunit 2 C-terminal peptide "SIESDV" (Husi and Grant, 2001) 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 at days 30 and 60 of cortical differentiation (4 samples per line at each time point from 2

independent differentiations) 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). Peptide affinity pulldown samples were prepared for mass-spectrometry and analysed by Dr Mark Collins (Department of Biomedical Science, University of Sheffield) as described in (Sanders et al., 2020).

2.13. Morphometric 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 500 ng 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.

2.14. RNA sequencing

WT and *DLG2* KO cells were cultured to days 15, 20, 30 and 60 of cortical differentiation as described in section *2.2 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). Quality control (QC) to ensure a high RNA integrity number (RIN) for all samples was with the RNA 6000 Nano kit analysed using the 2100 Bioanalyzer Eukaryote Total RNA Nano assay (Agilent), performed by Dr Amanda Redfern (Central Biotechnology Services, Cardiff University). 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) according to the manufacturer's protocol. The concentration of generated cDNA libraries was determined by the Qubit 1x dsDNA HS Assay kit (Thermo Fisher Scientific) and QC for library fragment size by High Sensitivity DNA kit analysed using the 2100 Bioanalyzer High

Sensitivity DNA assay (Agilent). Sequencing was performed by Joanne Morgan (MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University) 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, for approximately 30 million reads per sample.

RNA sequencing (RNAseq) Reads were analysed as described in Chapter 5, Underlying mechanisms of the *DLG2*$^{-/-}$ phenotype. Briefly, the fastq files from the Illumina HiSeq400 were trimmed to remove adapter sequences and reads with a low base call quality (Phred score < 30) using *Trimmomatic* (Bolger et al., 2014) and the quality of both the original and trimmed reads assessed using *FastQC* (Andrews, 2010). *STAR* (Dobin et al., 2013) was then used to index the GRCh38.p12 assembly of the human reference genome (Zerbino et al., 2018) and map the generated trimmed reads. *Picard* (Broad Institue, 2019) and *BamTools* (Barnett et al., 2011) were used to mark duplicates within the mapped trimmed reads and compute the percentage of duplication. This revealed a high level of duplicate reads in day 30 KO2 samples (~72% compared to an average of 23% for other samples) and as such these samples were removed prior to further analyses. The BAM files produced by *Star* were sorted using samtools (Li et al., 2009) and raw read counts per gene generated with f*eatureCounts* (Liao et al., 2014), using the comprehensive gene annotations on the primary assembly from Gencode (release 29). All software was run using the default parameters.

Differential expression analysis for protein coding genes was performed using the *DESeq2* package (Love et al., 2014) and differentially expressed genes were considered significant if their p value after Bonferroni correction was < 0.05. When analysing differential gene expression in *DLG2*$^{-/-}$ relative to WT, samples from KO1 and KO2 lines were combined so for each timepoint a single differential gene expression analysis was performed, comparing expression in KO1 & KO2 against WT.

For principal component analysis (PCA) and hierarchical cluster analysis gene counts were imported and log2-normalized by using the DESeq2 package (Love et al., 2014). The genes with expression >0 in at least 3 samples were considered for the further analyses. Dimensionality reduction was performed using the PCA implemented in DESDeq2. Hierarchical clustering of the samples was performed using the Pearson's correlation matrix of the Z-score-scaled and log2-normalized values, using Ward's linkage criterion and 1-r as the distance function, where r is the pairwise Pearson's correlation coefficient. Reported plots were designed by using the dendextend R package.

Using a script kindly provided by Dr Andrew Pocklington (MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University) the degree of overlap between the identified

differentially expressed genes and other gene sets was evaluated by a Fisher's Exact test, where the background set consisted of all genes expressed across the KO1, KO2 and WT cell lines. This approach was used to test enrichment for both a strictly annotated (with good experimental evidence) set of gene ontology (GO) terms (Ashburner et al., 2000; Carbon et al., 2019) and mammalian phenotype (MP) data from the mouse genome informatics (MGI) resource (Smith and Eppig, 2009). GO annotations were filtered to exclude genes with the evidence codes: NAS (Non-traceable Author Statement), IEA (Inferred from Electronic Annotation), and RCA (inferred from Reviewed Computational Analysis). The MP data was restricted to phenotypes reported in single-gene functional studies and where the mouse gene had a 1:1 mapping onto a human gene. GO or MP terms containing fewer than 10 genes were then excluded.

The results of these initial functional enrichment tests underwent a subsequent iterative refinement step to identify semi-independent subsets of enriched annotations. This involved selecting the gene set with the largest odds ratio following the initial enrichment and removing all genes found in this set from all other enriched annotations. These reduced gene-sets were re-tested for enrichment and only retained if $P < 0.05$ after Bonferroni correction. This process of odds ratio refinement was repeated, with each iteration resulting in the successive depletion of genes, until no significantly enriched (Bonferroni corrected $P < 0.05$) gene sets remained.

2.15. 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 a 96 well-plate pre-coated in various extracellular matrix (ECM) substrates (Collagen I, Collagen II, Collagen IV, Fibronectin, Laminin, Tenascin, Vitronectin) and a BSA coated negative control, a component of the ECM540 Cell Adhesion Array Kit (EMD Millipore) which was subsequently used to assess cell adhesion. After a 2 hour incubation at $37^{O}C$ and 3 washes in Assay Buffer (EMD Millipore) to remove non-adherent cells, all adherent cells within a well were stained using Cell Stain Solution (EMD Millipore). Following repeated H_2O washes to remove excess, cells were lysed to release bound stain using Extraction Buffer (EMD Millipore) and 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).

2.16. 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 or 10 day period. The relative number of cells present within a well at a given time point was quantified colorimetrically at 1 or 2 day intervals using the Cell Stain Solution and Extraction Buffer components of the ECM540 Cell Adhesion Array Kit (EMD Millipore) described in section 2.7. The absorbance at 560 nm of the Extraction Buffer, being relative to the total number of cells within the stained well at that time point, was quantified using a CLARIOstar microplate reader (BMG Labtech).

2.17. 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.

2.18. Electrophysiology

WT and *DLG2* KO cells were cultured to day 20 of cortical differentiation as described in section *2.2 Cortical differentiation*. At day 20 cells were prepared for electrophysiology by Dr Eunju Jenny Shin (Neuroscience and Mental Health Research Institute, Cardiff University). Electrophysiology experiments were conducted by Dr Daniel Whitcombe and Mr Tom Steward (Bristol Medical School, University of Bristol) and Dr Ying Zhu (Neuroscience and Mental Health Research Institute, Cardiff University) as described in (Sanders et al., 2020).

2.19. MAGMA gene set enrichment

Gene set enrichment was performed as described in Chapter 6, The *DLG2*$^{-/-}$ phenotype and neuropsychiatric disease. Briefly, common variant gene-set enrichment analysis was performed on sets of genes differentially expressed in *DLG2*$^{-/-}$ cells at 4 experimental time points (days 15, 20, 30 and 60) using a one-sided competitive gene-set enrichment test implemented in Multi-marker Analysis of GenoMic Annotation (MAGMA) version 1.07 (de Leeuw et al., 2015). Conditioning was on either all genes expressed in KO1, KO2 and WT cells at a given experimental time point, or alternatively across all timepoints, using the *condition-residualize* function of MAGMA. To

determine whether common variant enrichment differed between two gene sets their regression coefficients (β) and standard error (SE) were compared as described in (Altman and Bland, 2003) by calculating z and comparing to a standard normal distribution, where $z = d/SE(d)$, $d = \beta1 - \beta2$ and $SE(d) = \sqrt{[SE(\beta1)^2 + SE(\beta2)^2]}$.

To test whether specific GO terms enriched in day 30 down-regulated genes captured more or less of the schizophrenia association than expected, a two-sided enrichment test was performed on GO term target genes within the gene set, conditioning on both all expressed genes across time points and all day 30 down-regulated genes. Gene-level association statistics for schizophrenia (Pardiñas et al., 2018) and Alzheimer's disease (Lambert et al., 2013) were calculated using the MAGMA multi model, with a fixed 20,000 permutations for each gene.

2.20. 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<0.001; ****P<0.0001 vs. WT control or the first experimental time point as indicated in figure legends. All data presented as mean ± SEM. For phenotypic assays including ICC, western blotting, adhesion and proliferation n indicates the number of replicates used, where applicable this is from at least two independent differentiations unless otherwise stated in figure legends. For ICC quantification individual replicates are a mean value for a cell culture well taken from randomly chosen microscope fields, 6 for validation of hESC lines (Figure 4.5) and a minimum of 25 for characterisation of the *DLG2*$^{-/-}$ phenotype (Figure 4.6 + 4.8). All western blots are presented with expression normalised to GAPDH, those in Figure 4.9 are further normalised to mean WT expression at the earliest time point investigated (usually day 30).

2.21. Key resources

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-DLG2 (used at 1:1000)	Alamone	Cat# APZ-002
Rat monoclonal anti-CTIP2 (used at 1:500 ICC)	Abcam	Cat#ab18465
Rabbit polyclonal anti-CTIP2 (used at 1:1000 WB)	Abcam	Cat#ab70453
Rabbit polyclonal anti-FOXG1 (used at 1:250)	Abcam	Cat#ab18259
Mouse monoclonal anti-FLAG (used at 1:1000)	Sigma	Cast#F1804
Rabbit polyclonal anti-GAPDH (used at 1:5000)	Abcam	Cat#ab9485
Mouse monoclonal anti-KI67 (used at 1:150 ICC, 1:1000 WB)	BD Pharmingen	Cat#550609
Mouse monoclonal anti-MAP2 (used at 1:250 ICC)	Sigma-Aldrich	Cat#M1406
Rabbit polyclonal anti-MAP2 (used at 1:1000 WB)	Merck Millipore	Cat#AB5622
Rabbit monoclonal anti-NANOG (used at 1:400)	Cell Signaling Technology	Cat#4903
Rabbit polyclonal anti-NEUN (used at 1:500)	Merck Millipore	Cat#ABN78
Rabbit monoclonal anti-OCT4 (used at 1:400)	Cell Signaling Technology	Cat#2840
Rabbit monoclonal anti-PAX6 (used at 1:500)	Abcam	Cat#ab195045
Rat monoclonal anti-PH3 (used at 1:1000)	BioLegend	Cat#641002
Mouse monoclonal anti-SATB2 (used at 1:25 ICC, 1:100 WB)	Abcam	Cat#51502
Rabbit monoclonal anti-SOX2 (used at 1:400 ICC)	Cell Signaling Technology	Cat#3579
Goat polyclonal anti-SOX2 (used at 1:200 WB)	Santa Cruz	Cat#sc-17320
Rabbit polyclonal anti-TBR1 (used at 1:500 ICC, 1:1000 WB)	Abcam	Cat#ab31940
Donkey anti-Rabbit IgG (H+L) secondary antibody, Alexa Fluor 488 (used at 1:500)	Thermo Fisher Scientific	Cat#A-21206
Donkey anti-Rat IgG (H+L) secondary antibody, Alexa Fluor 488 (used at 1:500)	Thermo Fisher Scientific	Cat#A-21208
Donkey anti-Mouse IgG (H+L) secondary antibody, Alexa Fluor 594 (used at 1:500)	Thermo Fisher Scientific	Cat#A-21203
Donkey anti-Rabbit IgG (H+L) secondary antibody, Alexa Fluor 594 (used at 1:500)	Thermo Fisher Scientific	Cat#A-21207
IRDye 680RD goat anti-Rabbit (used at 1:15000)	Li-Cor	Cat#926-68071
IRDye 800CW goat anti-Mouse (used at 1:15000)	Li-Cor	Cat#926-32210
IRDye 800CW donkey anti-Goat (used at 1:15000)	Li-Cor	Cat#925-68074
Chemicals, Peptides, and Recombinant Proteins		
NMDA receptor subunit 2 peptide, SIESDV	Pepceuticals	Custom
E8	Thermo Fisher Scientific	Cat# A1517001
DMEM/F-12	Thermo Fisher Scientific	Cat#21331020
Neurobasal	Thermo Fisher Scientific	Cat#21103049
N2 supplement	Thermo Fisher Scientific	Cat#17502001
B27 supplement	Thermo Fisher Scientific	Cat#17504001
B27 supplement minus vitamin A	Thermo Fisher Scientific	Cat#12587001
Pen Strep Glutamine	Thermo Fisher Scientific	Cat#10378016
2-Mercaptoethanol	Thermo Fisher Scientific	Cat#31350010
Human ES-certified Matrigel	Corning	Cat#354277
Growth Factor-reduced Matrigel	Corning	Cat#354230
LDN193189	Cambridge Biosciences	Cat#SM23-5
SB431542	Stratech Scientific	Cat#S1067
Accutase	Sigma-Aldrich	Cat#A6964
Versene	Thermo Fisher Scientific	Cat#15040033
Y27632 Dihydrochloride	Stratech Scientific	Cat#S1049-SEL
Human plasma fibronectin	Merck Millipore	Cat#FC010
Poly-D-Lysine	Sigma-Aldrich	Cat#P7280

Laminin	Sigma-Aldrich	Cat#L2020
Opti-MEM reduced serum media	Thermo Fisher Scientific	Cat#31985070
Syn-PER	Thermo Fisher Scientific	Cat#87793
MS-SAFE Protease and Phosphatase inhibitor	Sigma-Aldrich	Cat#MSSAFE
DC Protein assay	BioRad	Cas#5000111
Bolt LDS sample buffer	Thermo Fisher Scientific	Cat#B0008
Bolt sample reducing agent	Thermo Fisher Scientific	Cat#B0009
Bolt 4-12% Bis-Tris Plus gels	Thermo Fisher Scientific	Cat#NW04125BOX
Bolt MES SDS running buffer	Thermo Fisher Scientific	Cat#B000202
Bolt Transfer buffer	Thermo Fisher Scientific	Cat#BT00061
Pierce Bovine Serum Albumin standards	Thermo Fisher Scientific	Cat#23208
Ponceau S	Sigma-Aldrich	Cat#141194
Affi-Gel 10 resin	Bio-Rad	Cat#1536099
Lipofectamine 3000	Thermo Fisher Scientific	Cat#L3000001
Foetal bovine serum	BioSera	Cat#FB-1285
Donkey serum	BioSera	Cat#AS-228
GMEM BHK-21	Thermo Fisher Scientific	Cat#11710035
MEM Non-essential amino acids	Thermo Fisher Scientific	Cat#11140035
Sodium pyruvate	Thermo Fisher Scientific	Cat#11360070
TrypLE Express	Thermo Fisher Scientific	Cat#12605028
Ampicillin	VWR	Cat#0339-EU-25G
Lipofectamine 2000	Thermo Fisher Scientific	Cat#11668030
Lipofectamine 3000	Thermo Fisher Scientific	Cat#L3000015
Lipofectamine stem transfection reagent	Thermo Fisher Scientific	Cat#STEM00008
Blasticidin	Thermo Fisher Scientific	Cat#R21001
Puromycin	Thermo Fisher Scientific	Cat#A1113803
Alt-R® S.p. Cas9 D10A Nickase	IDT	Cat#1091062
MyTaq DNA polymerase	Bioline	Cat#BIO-21105
PrimeStar Max DNA Polymerase	Takara	Cat#R045A
SYBR Safe DNA Gel Stain	Thermo Fisher Scientific	Cat#S33102
Paraformaldehyde	Sigma-Aldrich	Cat#6148
DAPI nucleic acid stain	Thermo Fisher Scientific	Cat#62248
TRIzol reagent	Thermo Fisher Scientific	Cat#15596026
HEPES pH7	Sigma-Aldrich	Cat#H0887
Sodium deoxycholate	Sigma-Aldrich	Cat#30970
Kits		
ECM540 Cell Adhesion array kit	Millipore	Cat#ECM540
DC protein assay	Bio-Rad	Cat#5000116
ISOLATE II Genomic DNA kit	Bioline	Cat#BIO52066
PureLink RNA Mini kit	Thermo Fisher Scientific	Cat#12183020
RNA 6000 Nano Kit	Agilent	Cat#5067-1511
Infinium PsychArray v1.1 BeadChip	Illumina	Cat#20024692
KAPA mRNA HyperPrep kit for Illumina platforms	KAPA Biosystems	Cat#KK8580
KAPA single-indexed adapter set A+B	KAPA Biosystems	Cat#KK8700
Qubit 1X dsDNA HS Assay Kit	Thermo Fisher Scientific	Cat#Q33230
High Sensitivity DNA kit	Agilent	Cat#5067-4626
P3 Primary Cell 4D-Nucleofector X Kit L	Lonza	Cat#V4XP-3024

MycoAlert PLUS Mycoplasma Detection Kit	Lonza	Cat#LT07-518
Guide-It Long ssDNA Production System	Takara	Cat#632666
Nucleospin Gel and PCR clean-up kit	Fisher Scientific	Cat#12303368
Zero Blunt TOPO PCR Cloning Kit for Sequencing	Thermo Fisher Scientific	Cat#K2875J10
PureYield Plasmid Miniprep System	Promega	Cat#A1223
Deposited Data		
GRCh38.p12 human reference genome FASTA file and annotation (Gencode release 29) GTF file	Ensembl (Zerbino et al., 2018)	http://www.ensembl.org/info/data/ftp/index.html
GO terms	(Ashburner et al., 2000; Carbon et al., 2019)	http://geneontology.org
MP terms	(Smith and Eppig, 2009)	http://www.informatics.jax.org/
MAGMA auxiliary files (Gene locations build 38 & European SNP locations)	(de Leeuw et al., 2015)	https://ctg.cncr.nl/software/magma
Schizophrenia GWAS data (PGC2 + CLOZUK)	(Pardiñas et al., 2018)	https://walters.psycm.cf.ac.uk
Alzheimer's disease GWAS data (IGAP)	(Lambert et al., 2013)	http://web.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php
Human Brain Transcriptome	(Kang et al., 2011)	https://hbatlas.org
Cortecon	(van de Leemput et al., 2014)	http://cortecon.neuralsci.org
Experimental Models: Cell Lines		
H7 human embryonic stem cells (hESCs)	Wi Cell	WA07 (H7)
H9 human embryonic stem cells (hESCs)	Wi Cell	WA09 (H9)
HEK 293	In house	N/A
5-alpha competent *E. coli*	NEB	Cat#C2987H
JSD4 *DLG2*$^{+/+}$ hESCs (WT)	In house	N/A
JSD2 *DLG2*$^{-/-}$ hESCs (KO1)	In house	N/A
JSD21 *DLG2*$^{-/-}$ hESCs (KO2)	In house	N/A
Oligonucleotides		
SygRNA crRNA (details in Chapter 3)	Sigma-Aldrich	Custom
SygRNA tracrRNA (details in Chapter 3)	Sigma-Aldrich	TRACRRNA05N
Alt-R® CRISPR-Cas9 crRNA (details in Chapter 3)	IDT	Custom
Alt-R® CRISPR-Cas9 tracrRNA (details in Chapter 3)	IDT	1072532
PCR primers (details in main text and supplementary information)	Sigma-Aldrich	VC00021 (custom)
Recombinant DNA		
pmaxGFP plasmid	Lonza	Component of Cat# V4XP-4024
DLG2 (Myc-DDK tagged) transcript variant 1	Cambridge Bioscience	Cat#RC227175
DLG2 (Myc-DDK tagged) transcript variant 2	Cambridge Bioscience	Cat#RC223915
DLG2 (Myc-DDK tagged) transcript variant 3	Cambridge Bioscience	Cat#RC227260
DLG2 (Myc-DDK tagged) transcript variant 4	Cambridge Bioscience	Cat#RC226876
DLG2 (Myc-DDK tagged) transcript variant 5	Cambridge Bioscience	Cat#RC233167
DLG2 (Myc-DDK tagged) transcript variant 6	Cambridge Bioscience	Cat#RC239685
Multiplex CRISPR/Cas9 Assembly System Kit	AddGene (Sakuma et al., 2014)	Kit#1000000055
Synthesized HDR vectors (details in Chapter 3)	GeneWiz	Gene Synthesis (custom)
Software and Algorithms		

R	N/A	https://www.r-project.org/ RRID:SCR_001905
Trimmomatic	(Bolger et al., 2014)	http://www.usadellab.org/cms/?page=trimmomatic
FastQC	(Andrews, 2010)	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ RRID:SCR_014583
STAR	(Dobin et al., 2013)	https://github.com/alexdobin/STAR/ RRID:SCR_015899
Samtools	(Li et al., 2009)	http://www.htslib.org/ RRID:SCR_002105
Picard	(Broad Institue, 2019)	https://broadinstitute.github.io/picard/
BamTools	(Barnett et al., 2011)	https://github.com/pezmaster31/bamtools
featureCounts	(Liao et al., 2014)	https://bioinformaticshome.com/tools/rna-seq/descriptions/FeatureCounts.html
DESeq2	(Love et al., 2014)	https://bioconductor.org/packages/release/bioc/html/DESeq2.html RRID:SCR_015687
MAGMA	(de Leeuw et al., 2015)	https://ctg.cncr.nl/software/magma
gRNA design web-based tool	Zhang Lab	Crispr.mit.edu
HCS Studio Cell Analysis Software	Thermo Fisher Scientific	N/A
GraphPad Prism	GraphPad	N/A
Genome Studio	Illumina	https://emea.support.illumina.com/array/array_software/genomestudio/downloads.html
PennCNV	(Wang et al., 2007)	http://penncnv.openbioinformatics.org/en/latest/user-guide/download/
ImageJ	(Schneider et al., 2012)	https://imagej.nih.gov/ij/
StackReg plugin	(Thévenaz, 2011)	http://bigwww.epfl.ch/thevenaz/stackreg/
Neurolucida	MBF Bioscience	N/A
NCBI Primer-BLAST	NCBI	https://www.ncbi.nlm.nih.gov/tools/primer-blast/

3. The spatial and temporal pattern of *DLG2* expression

3.1. Introduction

Mutations in the *DLG2* gene have been linked to neuropsychiatric diseases including schizophrenia (Kirov et al., 2012a; Marshall et al., 2017; Reggiani et al., 2017; Ruzzo et al., 2019) although research into DLG2s role in the underlying neurobiology of disease has been limited, focusing on synaptic transmission and neural network physiology during relatively late stages of neurodevelopment (MacLaren et al., 2011). Given the long understood neurodevelopmental component of schizophrenia aetiology (Murray and Lewis, 1987; Weinberger, 1987; Harrison, 1997, 1999), the evidence that disruption to neurodevelopmental processes in disease occur relatively early during pregnancy (Murray and Lewis, 1987; Weinberger, 1987; Harrison, 1997, 1999) and the established role *DLG2s* invertebrate orthologue in key developmental signalling pathways (Humbert et al., 2008; Massimi et al., 2008; Stephens et al., 2018) it was hypothesised that DLG2 may have a currently unreported role in the regulation of early neurodevelopment, in addition to its established function within the postsynaptic density of mature neurons.

Research into the broad pattern of DLG2 protein expression during the course of neurodevelopment, as opposed to a narrow focus on the synapse of mature neurons, would be a logical first step in investigating a potential novel neurodevelopmental role for the gene. This includes the temporal pattern of expression for DLG2 during neurodevelopment, to determine if it is restricted solely to later time points containing more mature cell types. Clearly DLG2 expression at early time points of neurodevelopment and prior to synaptogenesis would be indicative of an additional role beyond synaptic transmission. The spatial pattern of DLG2 expression would also provide key insights, that is both the cell types it is expressed as well as subcellular localisation. This could confirm whether DLG2 is expressed solely in more mature synapse containing neurons or if it is also expressed in NPCs and new-born immature neurons. Additionally, the subcellular localisation could provide evidence of function, in mature neurons DLG2 is known to be localised beneath the postsynaptic membrane within the PSD, if DLG2 is expressed in other cell types and in a similar location it would be indicative of a similar scaffolding role in signal transduction complexes. However, if DLG2 was found to be expressed in the nucleus during early neurodevelopment this may indicate an alternative role potentially as a transcription factor, as is the case for proteins with a similar domain structure to DLG2 such as CASK (Hsueh et al., 2000; Huang and Hsueh, 2017).

At the mRNA level there is some published work relating to *DLG2* expression during early human neurodevelopment in the form of transcriptome data available in several publicly accessible databases. One such database is the Human Brain Transcriptome (Kang et al., 2011) (https://hbatlas.org) which provides genome-wide exon-level transcriptome data for the developing and adult human brain. Samples being taken from 16 regions of post-mortem human brains, as well

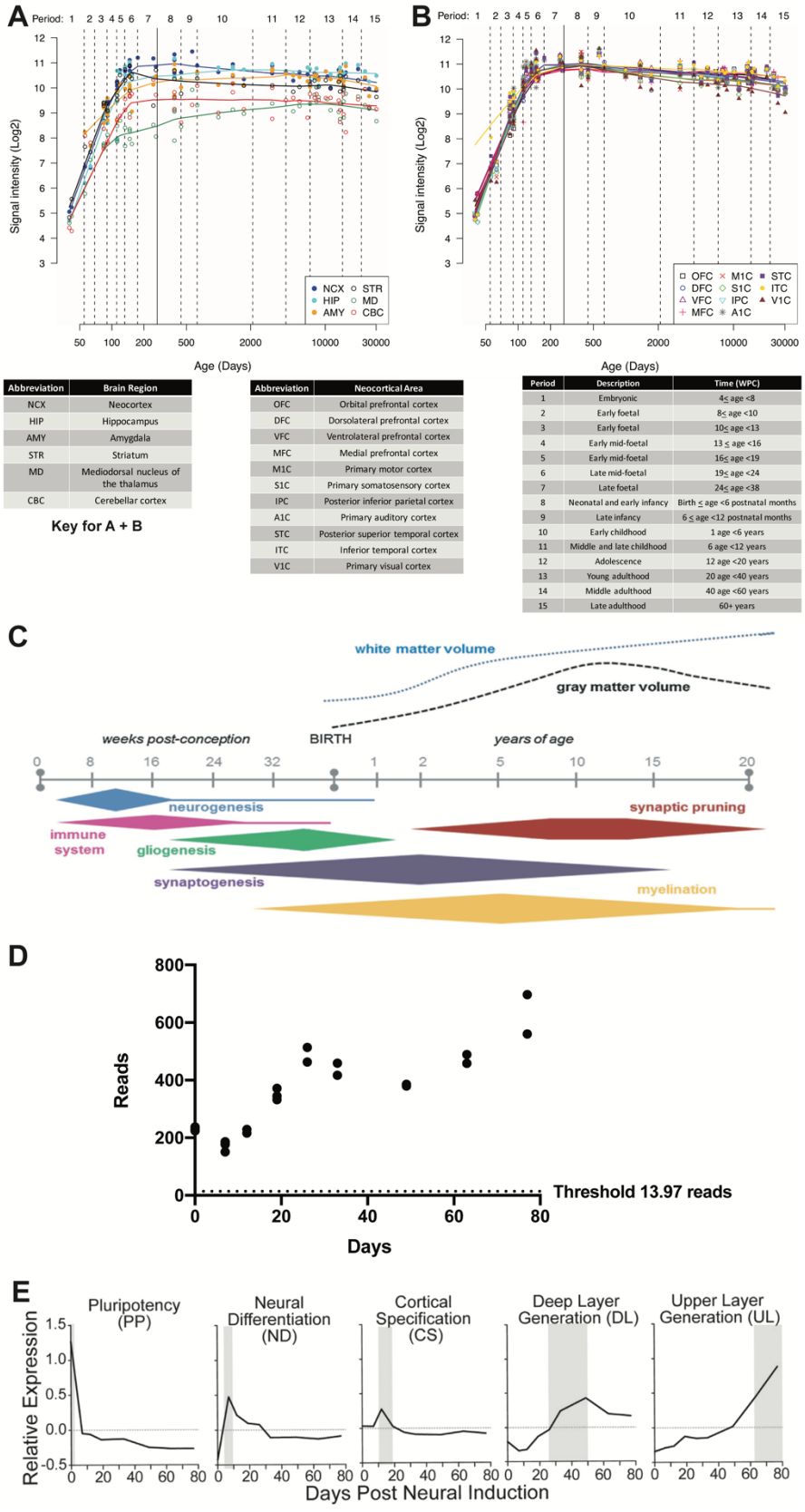
as more specifically 11 regions of the neocortex, across a wide range of developmental periods. The data for *DLG2* both across brain regions (Figure 3.1 A) and within the neocortex (Figure 3.1 B) show expression throughout the developing and adult brain at all periods investigated. Due to the limitations of the technique, expression during the earliest developmental period (period 1), representing less than 8 weeks post conception (WPC), is of marginal significance; however, there is clearly significant *DLG2* expression in all sampled brain regions during the subsequent 8-10 WPC period (period 2). 8-10 WPC lies within the neurogenesis period of human development and is prior to the initiation of synaptogenesis, which begins around 20 WPC (Semple et al., 2013) (Figure 3.1 C), indicative of a role for *DLG2* beyond synaptic transmission.

A second such database is Cortecon (van de Leemput et al., 2014) (http://cortecon.neuralsci.org), a temporal transcriptome analysis of *in vitro* human cortical neurogenesis, using RNAseq performed at various time points as WA-09 hESCs are differentiated to cortical projection neurons via a dual SMAD inhibition protocol, adapted from that developed by Chambers et al. (Chambers et al., 2009). In agreement with the previous *in vivo* data Cortecon shows significant *DLG2* expression during early neurodevelopment and at all time points of *in vitro* cortical differentiation investigated, including hESCs at day 0 (Figure 3.1 D). It should be noted that using this hESC line and differentiation protocol deep layer neurons (cortical layer VI) start to be generated from day 26 of differentiation (Figure 3.1E), indicating that *DLG2* is expressed prior to neurogenesis. Indeed, the level of *DLG2* at the latest timepoint in culture containing the most mature neurons is only 2-3 times that found during earlier periods of neural differentiation (around day 7) and cortical specification (around day 12). This is again indicative of role for *DLG2* beyond that of signal transmission in mature neuronal synapses and suggests a potential role in early cortical neurogenesis.

Unlike the transcriptome data there are no available databases or published literature with comparable data on DLG2 protein expression during early human neurodevelopment, requiring experimental investigation. The standard techniques to determine the expression pattern of a protein of interest are western blotting and ICC, both of which require specific antibodies to the target protein. However, preliminary experiments determined these were not available for DLG2, various commercially available as well as custom made antibodies all showed a degree of non-specific binding indicated by additional bands or smears of protein on western blots. In order to positively confirm the identity of DLG2 bands in western blots, DLG2 protein isoforms fused to a FLAG tag (Einhauer and Jungbauer, 2001), commercially available as cDNA clones, were exogenously produced in HEK cells and run alongside endogenous protein samples as a positive control. Such a technique can obviously not be used in conjunction with ICC to investigate the spatial pattern of DLG2 expression, as such the decision was taken to use genome editing to knock-in a FLAG-tag to the *DLG2* gene to enable endogenous expression to be accurately determined. Various strategies

are available for targeted genome editing including zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) (Wood et al., 2011); however, the CIRSPR/Cas9 system was selected due to its ease of customisation and higher targeting efficiency (Ran et al., 2013a).

The aim of the research presented in this chapter was to determine the spatiotemporal pattern of DLG2 expression during neurodevelopment. To this end the level of DLG2 protein expression during the course of cortical differentiation was also determined experimentally using western blotting with appropriate positive controls, although due to issues with anti-DLG2 antibody specificity ICC could not be used to determine the spatial pattern of DLG2 expression. In order to allow spatial identification of endogenous DLG2, efforts were made to knock-in a FLAG-tag to the stop codon region of the *DLG2* gene, using CRISPR/Cas9 technology to target cutting of the genome at this location and a HDR template to insert the tag.



Abbreviation	Brain Region
NCX	Neocortex
HIP	Hippocampus
AMY	Amygdala
STR	Striatum
MD	Mediodorsal nucleus of the thalamus
CBC	Cerebellar cortex

Abbreviation	Neocortical Area
OFC	Orbital prefrontal cortex
DFC	Dorsolateral prefrontal cortex
VFC	Ventrolateral prefrontal cortex
MFC	Medial prefrontal cortex
M1C	Primary motor cortex
S1C	Primary somatosensory cortex
IPC	Posterior inferior parietal cortex
A1C	Primary auditory cortex
STC	Posterior superior temporal cortex
ITC	Inferior temporal cortex
V1C	Primary visual cortex

Period	Description	Time (WPC)
1	Embryonic	4≤ age <8
2	Early foetal	8≤ age <10
3	Early foetal	10≤ age <13
4	Early mid-foetal	13 ≤ age <16
5	Early mid-foetal	16≤ age <19
6	Late mid-foetal	19≤ age <24
7	Late foetal	24≤ age <38
8	Neonatal and early infancy	Birth ≤ age <6 postnatal months
9	Late infancy	6 ≤ age <12 postnatal months
10	Early childhood	1 age <6 years
11	Middle and late childhood	6 age <12 years
12	Adolescence	12 age <20 years
13	Young adulthood	20 age <40 years
14	Middle adulthood	40 age <60 years
15	Late adulthood	60+ years

Figure 3.1 *DLG2* mRNA expression during neurodevelopment
(A+B) *In vivo DLG2* mRNA expression for 6 regions of the human brain (A) and 11 neocortical areas (B) from post-mortem samples across 15 developmental periods. A solid vertical line following period 7 indicates birth and Log2 signal intensities above 5 indicate significant results. WPC = weeks post conception. Adapted from (https://hbatlas.org). **(C)** The human neurodevelopmental time course. From (Semple *et al.*, 2013). **(D)** *In vitro DLG2* mRNA expression during cortical differentiation from hESCs. From (http://cortecon.neuralsci.org). **(E)** The developmental stages during cortical differentiation from hESCs, as determined by gene expression profiles. From (van de Leemput *et al.*, 2014).

3.2. Experimental procedure - outline

To investigate the temporal pattern of DLG2 expression, protein samples were extracted in triplicate from a WT cell line (H7) at six time points during differentiation from hESCs to cortical projection neurons (days 15, 20, 30, 40, 50, 60). Western blotting was used to analyse these protein samples which were run alongside FLAG-tagged DLG2 isoforms as a positive control to identify DLG2 bands. FLAG-tagged isoforms being exogenously produced in HEK cells following transfection, via lipofection, with commercially available tagged cDNA clones for the DLG2 transcript variants. Various different CRISPR/Cas9 genome editing approaches were used in an attempt to knock-in a FLAG-tag to the vicinity of the *DLG2* stop codon. Initially a multiplex CRISPR/Cas9 plasmid-based approach, containing multiple gRNAs and a Cas9 expression cassette, was attempted in H7 hESC using 2 gRNAs to target a Cas9 nuclease mediated double strand break, along with a HDR vector (either circular or linear) for FLAG-tag insertion. In a second approach, 2 synthesised gRNAs were used in conjunction with the Cas9 nuclease expressing iCas9 cell line, with a single stranded HDR template for FLAG-tag insertion. Finally, 4 synthesised gRNAs (2 pairs) were used with a Cas9 nickase to create a pair of single stranded breaks again with a single stranded HDR template for FLAG-tag insertion. Transfection of CRISPR/Cas9 editing components into hESCs was either via nucleofection or lipofection with various different concentrations and ratios of gRNA/Cas9/HDR template tested. Correct assembly of vectors for CRISPR/Cas9 editing as well as confirmation of results was by PCR followed by agarose gel electrophoresis or sanger sequencing. Detailed protocols can be found in Chapter 2, Sections *2.1 Cell maintenance, 2.2 Cortical differentiation, 2.3 Western blotting, 2.4 Preparation of DLG2-FLAG isoforms, 2.5 CRISPR/Cas9 genome editing to generate tagged DLG2, 2.6 PCR & agarose gel electrophoresis, 2.7 TOPO cloning, 2.8 Sanger sequencing*. Further details of CRISPR/Cas9 genome editing follow within this chapter.

3.3. Western blotting to determine DLG2 expression during cortical differentiation

Although data on *DLG2* mRNA expression during development provides useful insights, it is important to validate this at the protein level, as mRNA and protein expression often do not show good correlation (Liu et al., 2016). Due to a lack of published data relating to DLG2 protein expression during early development, this was investigated experimentally *in vitro* during the course of cortical differentiation. H7 hESCs were differentiated to cortical projection neurons using an adaptation of the Chambers et al. dual SMAD inhibition protocol (Chambers et al., 2009), although with modifications both from the original protocol and that used by to generate the Cortecon dataset (Figure 3.1) previously described (see section *2.2 Cortical differentiation*). Protein was extracted at various time points of differentiation and western blotting used to analyse the

temporal pattern of DLG2 expression during cortical neurogenesis (Figure 3.2). Due to preliminary data that suggested available anti-DLG2 antibodies, both commercial and custom made, showed a degree of non-specific binding, DLG2 isoforms tagged with a FLAG epitope were also run in western blots as a positive control. Staining of these controls using an anti-FLAG antibody allowed clear identification of bands corresponding to all isoforms with the exception of isoform 4 which was absent in all replicates, potentially due to an error with the corresponding cDNA clone or experimental error in the protein production and extraction protocol. Additional staining of these positive controls and a comparison to the anti-FLAG blots showed that the anti-DLG2 antibody used in these experiments was able to detect isoforms 1, 2, 5 and 6 but not isoform 3, although in all cases there was a degree of non-specific binding as indicated by additional bands or smears on the anti-DLG2 blots that are absent from the anti-FLAG. By running these positive controls alongside protein samples extracted from 6 time points of WT cortical differentiation it was therefore possible to identify DLG2 expression in these endogenous samples despite the deficiencies with the antibody, by comparison the predominant DLG2 isoforms in these samples appearing be 1 and 2/5, which have a similar molecular weight (Figure 3.2A).

Unlike the published RNAseq data which shows *DLG2* mRNA expression at all time points of cortical differentiation from hESCs (Figure 3.1D) the expression of DLG2 protein as determined experimentally by western blotting is less clear. Faint DLG2 bands are visible at days 15, 20 and 30 although due to issues with antibody quality and the resulting 'noise' on western blots it is not possible to determine if there is any genuine protein expression at these time points using this approach. From day 40 of differentiation there is a trend upwards in expression and by days 50 and 60 clear protein bands are visible which when quantified (Figure 3.2A) show significantly increased intensity compared to all earlier time points. It would be expected for a PSD protein such as DLG2 to have high levels of expression at the later timepoints of cortical differentiation, including days 50 and 60. At this stage of the differentiation, cultures were observed to have large proportions of cells with a clear neuronal morphology and the western blot results for the postmitotic neuron marker MAP2 support this (Figure 3.2 B). Prior to day 30 there is little detectable MAP2 expression and no neuronal morphology was observed in the cultures with the predominant cell type at days 15 and 20 being NPCs, as confirmed by western results for the general NPC marker SOX2 (Figure 3.2 C) and the dorsal telencephalic NPC marker PAX6 (Figure 3.2 D). However, it appears that western blotting using currently available anti-DLG2 antibodies is unable to confirm DLG2 protein expression at these early, pre-neuronal, time points of cortical differentiation.

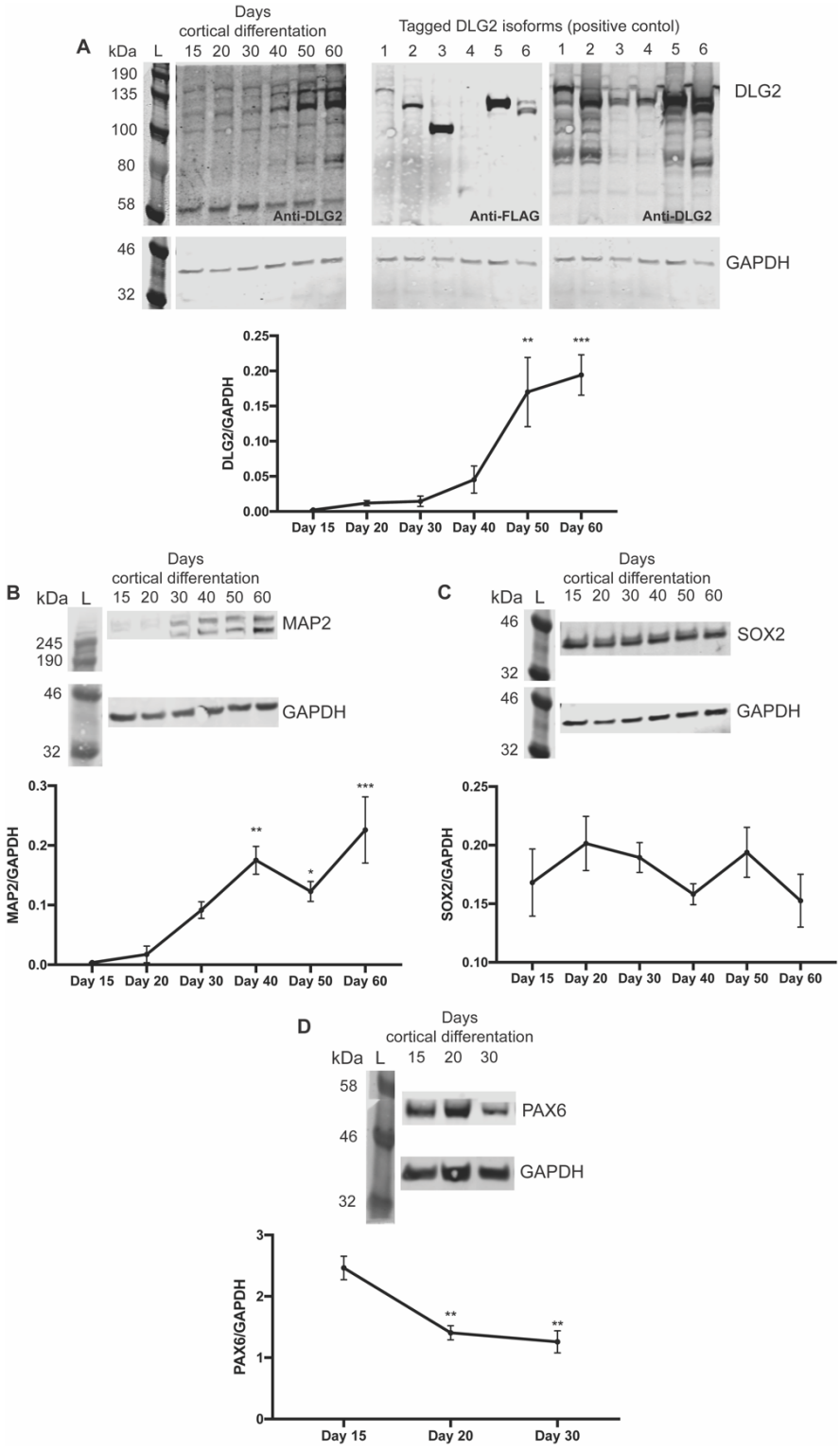


Figure 3.2 Protein expression during WT cortical differentiation from hESCs
(A) DLG2 western blot protein bands and graph of expression normalised to GAPDH. Exogenously produced DLG2 isoforms were run as a positive control, detected by FLAG-tag expression. Bands corresponding to isoforms 1 ,2 and 5 as identified from positive controls were quantified for all 6 experimental time points with combined results shown. Significance was determined by one-way ANOVA ($F(5,12) = 12.05$, $p = 0.0002$) **(B)** MAP2 western blot protein bands and graph of expression normalised to GAPDH. Significance was determined by one-way ANOVA ($F(5,12) = 10.65$, $p = 0.0004$) **(C)** SOX2 western blot protein bands and graph of expression normalised to GAPDH. Significance was determined by one-way ANOVA ($F(5,12) = 0.9693$, $p = 0.4741$) **(D)** PAX6 western blot protein bands and graph of expression normalised to GAPDH. Significance was determined by one-way ANOVA ($F(5,12) = 15.70$, $p = 0.0041$). For all graphs stars above points indicate Bonferroni corrected post hoc tests comparing to day 15 expression, *P<0.05; **P<0.01; ***P<0.001; ****P<0.0001.

3.4. Multiplex CRISPR/Cas9 to generate tagged *DLG2*

Due to issues with anti-DLG2 antibody specificity efforts were made to knock-in a FLAG-tag to the *DLG2* gene using CRISPR/Cas9 technology so the temporal and spatial pattern of endogenously produced DLG2 protein could be accurately determined. The initial method selected for this was a plasmid based multiplex CRISPR/Cas9 approach (Sakuma et al., 2014). This involves the construction of an all-in-one vector with DNA sequences encoding multiple gRNAs, each gRNA containing a variable crRNA sequence for binding to target DNA and a common tracrRNA scaffold for Cas9 protein binding, along with a Cas9 nuclease expression cassette and required promoters (Figure 3.3). The first step in using multiplex CRISPR/Cas9 to generate tagged *DLG2* is to identify a suitable location on the gene to knock-in a FLAG-tag as well as the most appropriate gRNAs to target this location. Although the relatively small size of a FLAG-tag (24 base pairs) means it is less likely to impair the function of the tagged protein than a larger insert, such as a fluorescent protein, the decision was made to target the insertion to the gene's stop codon in exon 28 to avoid disruption to any protein coding domains. gRNA sequences were then designed to target as close as possible to the *DLG2* stop codon, dependent on the presence of a NGG PAM sequence immediately 3' to the guide, using the Zhang Lab CRISPR Design Tool (https://zlab.bio/guide-design-resources). This identified multiple potential gRNAs in the vicinity of the *DLG2* stop codon and assigned them a score from 0 to 100 based on the predicted efficiency of cutting at that location, with 100 indicating the best possible sequence. Based on previous reports that the efficiency of knock-in for FLAG-tags at stop codons decreased as the distance of the CRISPR/Cas9 cut from the stop codon increased (Paquet et al., 2016) the decision was taken to use two gRNAs targeting as close as possible to the stop codon (-14 and +19 base pairs), even though these had slightly reduced scores (60 and 55 respectively) than more distant gRNAs (Figure 3.4 A). The selected gRNAs were named nuclease gRNA 1 & 2 and both sense and antisense DNA oligos were designed and ordered for each (Figure 3.4 B). These oligos included a 20 nucleotide crRNA targeting sequence with the addition of a G/C to allow for guanine as the first base in the transcript, as this increases efficiency when using a U6 promoter (Ran et al., 2013a) as in the multiplex CRISPR vector. No tracrRNA sequence was added to these oligos as a gRNA scaffold sequence is already included in the multiplex CRISPR vector; however, overhangs were added to allow ligation into plasmid BbsI sites (Figures 3.3 & 3.5 A).

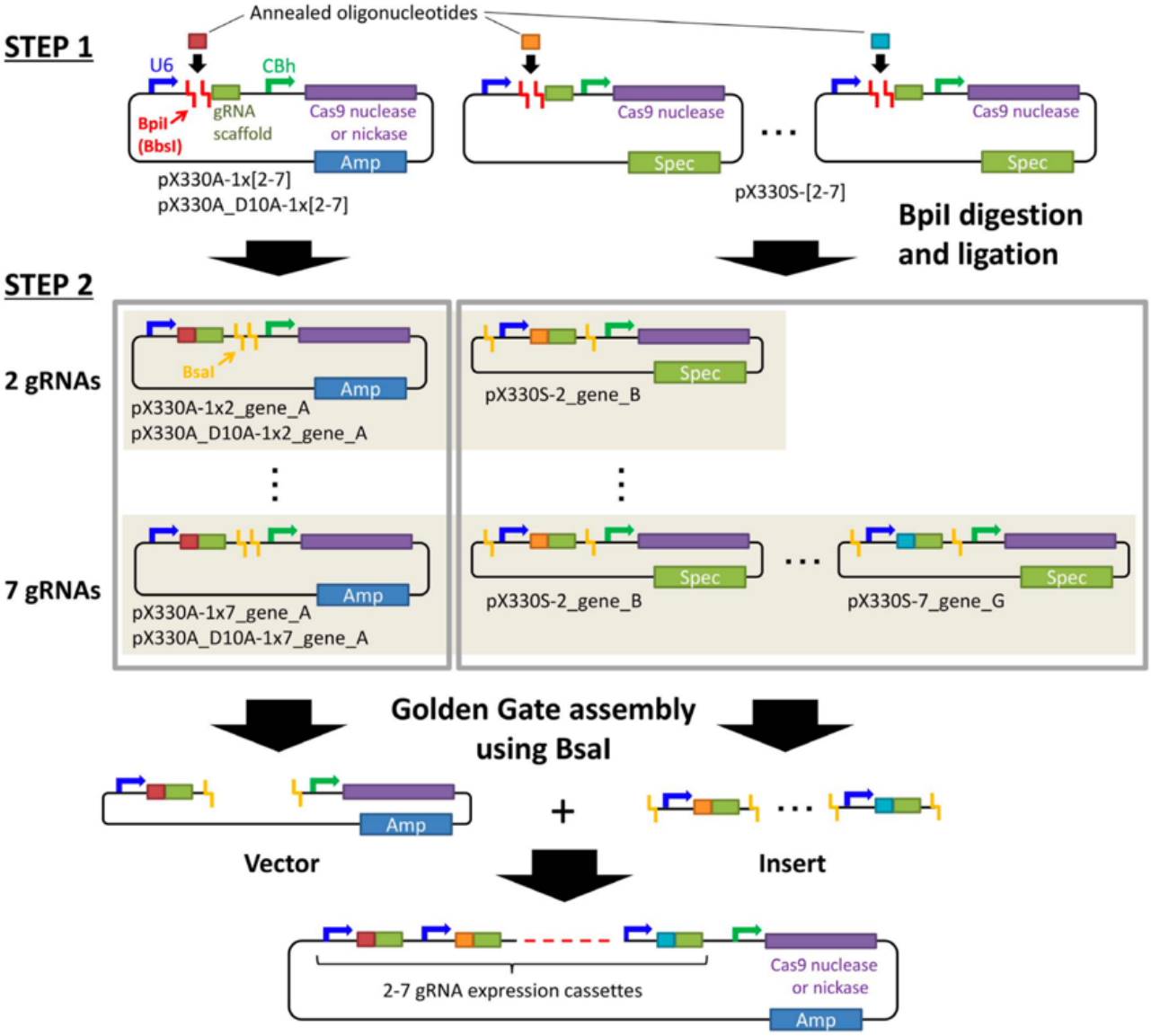


Figure 3.3 Schematic overview of the multiplex CRISPR/Cas9 system
Between 2 and 7 annealed DNA oligonucleotides encoding crRNA sequences targeting desired Cas9 nuclease cut sites are inserted into BpiI-digested pX330A or pX330S vectors adjacent to a common tracrRNA scaffold (gRNA scaffold) (STEP 1). The constructed vectors containing single gRNA expression cassettes are then assembled into an all-in-one vector containing multiple gRNA cassettes along with a Cas9 nuclease cassette using Golden Gate assembly (STEP 2). *Amp, ampicillin; Spec, spectinomycin; U6, human U6 promoter; CBh, chicken beta-actin short promoter.* From (Sakuma *et al.*, 2014)

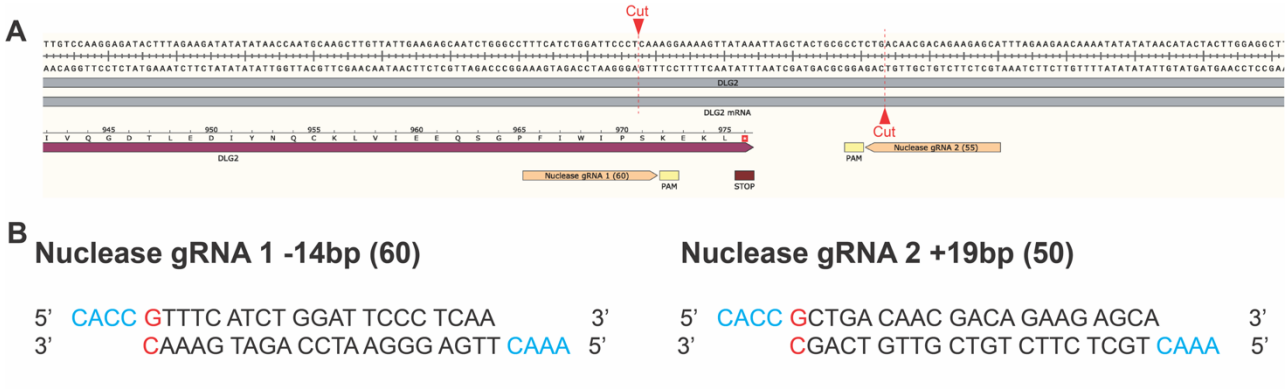


Figure 3.4 gRNAs for multiplex CRISPR/Cas9 genome editing targeting the *DLG2* stop codon
(A) Location of the 2 gRNAs targets and corresponding Cas9 cut sites selected for multiplex CRISPR/Cas9 genome editing within exon 28 of *DLG2*. The translated DLG2 sequence for this region is also given using single-letter amino acid code. **(B)** Sense and antisense DNA oligonucleotides corresponding to the designed gRNAs, with location of the target relative to the DLG2 stop codon indicated in base pairs along with the predicted editing efficiency (out of 100) indicated in brackets. Added G/C bases to increases U6 promoter efficiency are highlighted in red and BbsI sites to allow ligation into vectors are highlighted in blue.

The sense and antisense oligonucleotides for nuclease gRNA 1 & 2 were annealed prior to ligation into px330A-1 and px330S-2 respectively immediate 5' of a tracrRNA scaffold (Figures 3.3 & 3.5 A). To confirm insertion of the annealed oligonucleotides into these plasmids a diagnostic digest was performed using EcoRI and BbsI, this would result in two fragments without insertion but only 1 with, where the BbsI site is disrupted. The results showed a single band for almost all the samples tested (Figure 3.5 C + D) confirming oligonucleotide insertion, with the exception of two samples for px330S-2 which showed multiple bands, although in a different pattern to the negative control, and as such these were discarded. Px330A and px330S containing individual gRNAs were then used to assemble a multiplex CRISPR vector (Figure 3.3 – STEP 2) containing both gRNAs along with a Cas9 nuclease and other components required for CRISPR/Cas9 genome editing. As the assembled multiplex CRISPR/Cas9 vector is only slightly larger than the separate plasmids that were used in its construction, with only the addition of a crRNA oligonucleotide, tracrRNA scaffold and U6 promoter, confirmation of correct assembly with gel electrophoresis would be difficult. Sanger sequencing was therefore performed on the plasmids using primers surrounding the gRNA site (CRISPR-step2-F & CRISPR-step2-R Figure 3.5 A) which showed insertion of both gRNA oligonucleotides in all cases (Figure 3.5 E) and confirming correct multiplex CRISPR/Cas9 vector assembly.

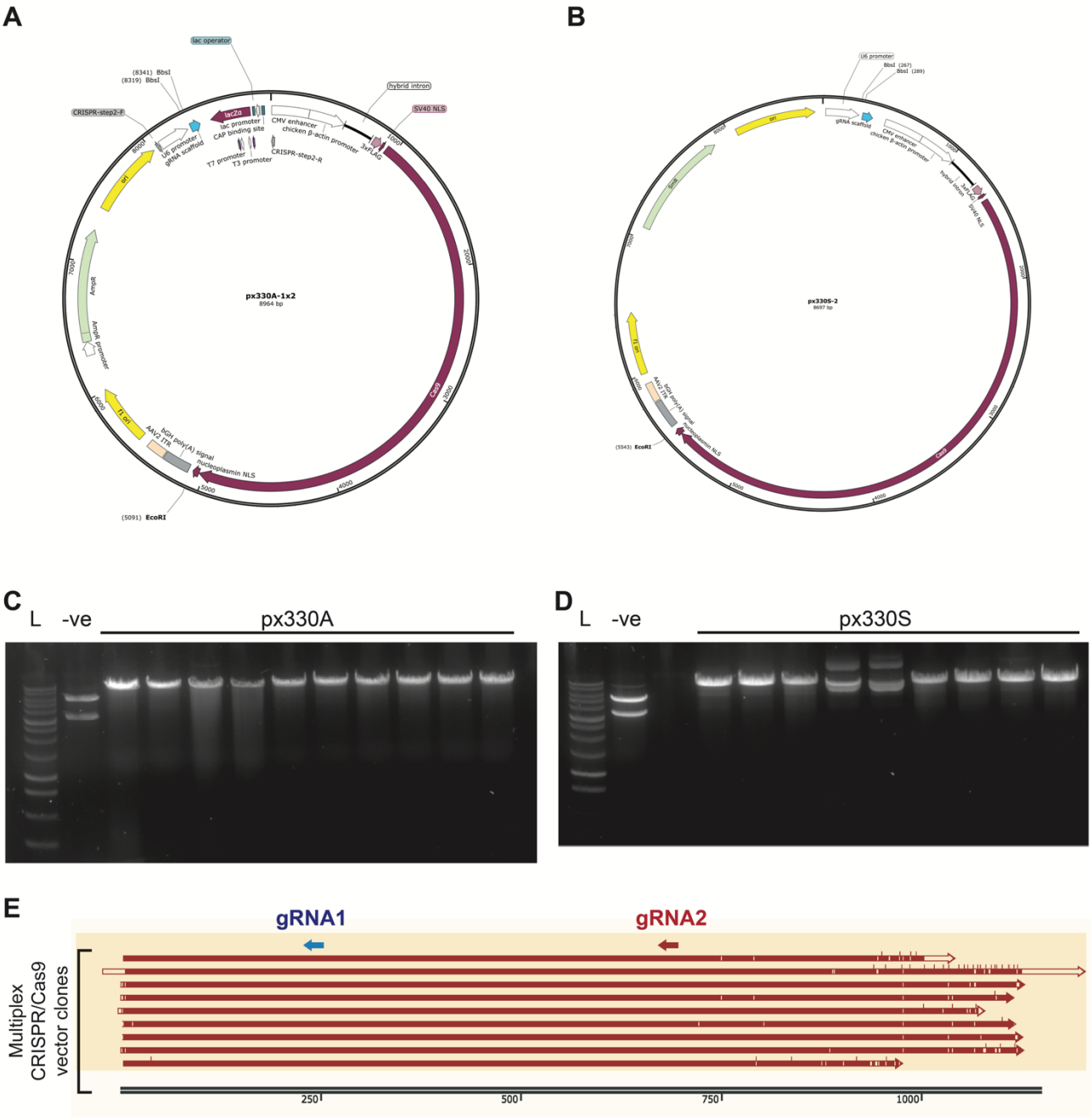


Figure 3.5 Multiplex CRISPR/Cas9 vector assembly
(A) Map of px330A, the vector for Nuclease gRNA1. **(B)** Map of px330S, the vector for Nuclease gRNA2. **(C)** Diagnostic digests of px330A following gRNA1 ligation, with an unligated negative control. **(D)** Diagnostic digests of px330S following gRNA2 ligation, with an unligated negative control. **(E)** Alignments of Nuclease gRNA1 (blue arrow) and Nuclease gRNA2 (red arrow) DNA sequences with 9 multiplex CRISPR/Cas9 clones following sanger sequencing, full alignment confirms correct vector assembly.

In order to test whether the designed gRNAs to cut in the vicinity of the DLG2 stop codon HEK cells were transfected with the multiplex CRISPR/Cas9 vector using lipofection. Multiple colonies survived transfection so genomic DNA was extracted and PCR performed using primers designed to flank the expected CRISPR/Cas9 cut sites (Figure 3.6 A). Gel electrophoresis using 4% agarose was then performed on these PCR fragments to look for indels, caused by double strand break repair following CRISPR/Cas9 cutting. This clearly showed additional bands in many samples tested (Figure 3.6 B), both larger and small than negative controls and indicative of insertions and deletions, suggesting the Cas9 driven by gRNAs are capable of cutting at this location.

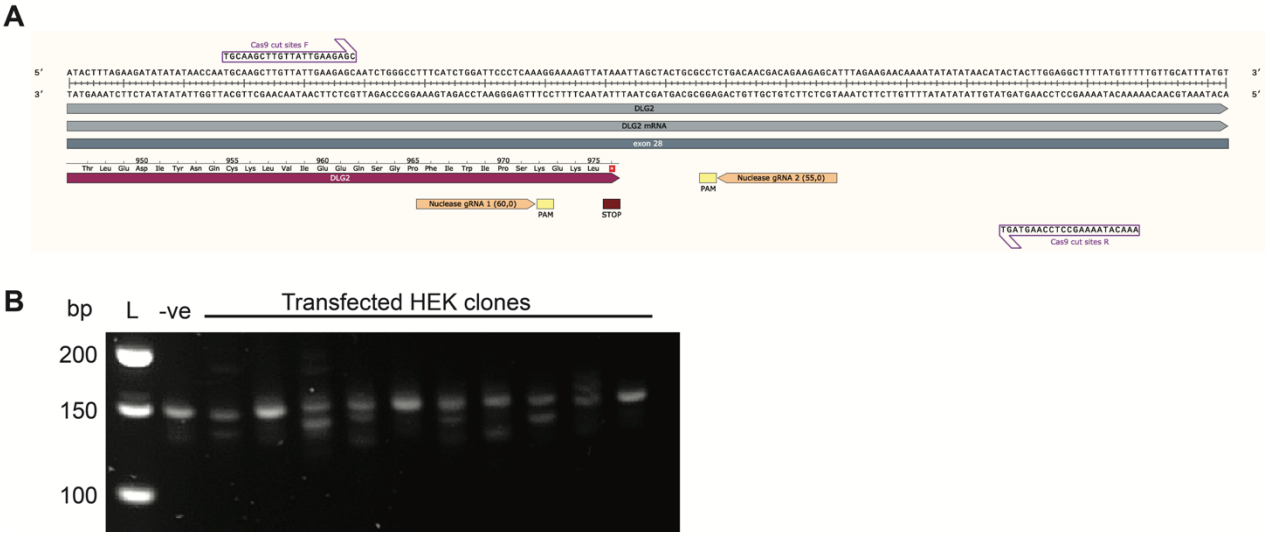


Figure 3.6 Multiplex CRISPR/Cas9 vector testing in HEK
(A) Location of primers designed to amplify the targeted Cas9 cut sites **(B)** PCR amplification of HEK genomic DNA using Cas9 cut site primers following transfection of multiplex CRISPR/Cas9 vector, additional bands over control indicated presence of indels. 4% agarose was used to better resolve bands.

Once double strands breaks had been created in the vicinity of the *DLG2* stop codon, using the multiplex CRISPR/Cas9 vector previously described, a FLAG-tag would be inserted using an HDR vector. This HDR vector was designed (Figure 3.7 A) to have two 700 base pair homology arms complementary to the DNA sequence flanking the *DLG2* stop codon to enable the vector to bind to this region and act as template to repair the breaks caused by the multiplex CRISPR/Cas9 vector and in the process insert a FLAG-tag and other elements. In addition to the homology arms the vector contained three FLAG-tags, allowing increased binding to anti-FLAG antibodies in later experiments, which were connected to 3' region of the *DLG2* gene (minus stop codon) using a glycine-serine (G-S) linker and were followed by a SV40 poly(A) signal for termination of transcription and mRNA stability (Proudfoot et al., 2002). Also present in the HDR vector was a puromycin resistance gene to enable selection of cells that had taken up the vector along with an associated PGK promoter for the gene and a bGH poly(A) signal. Finally, two loxP sites were added flanking the PGH promoter, puromycin resistance gene and bGH poly(A) signal so that these

elements could be excised later if required. The HDR vector was synthesised commercially in the pUC57 plasmid (Figure 3.7 B) and both a restriction digest (Figure 3.7 C) and sequencing (Figure 3.7 D) were performed to confirm correct assembly.

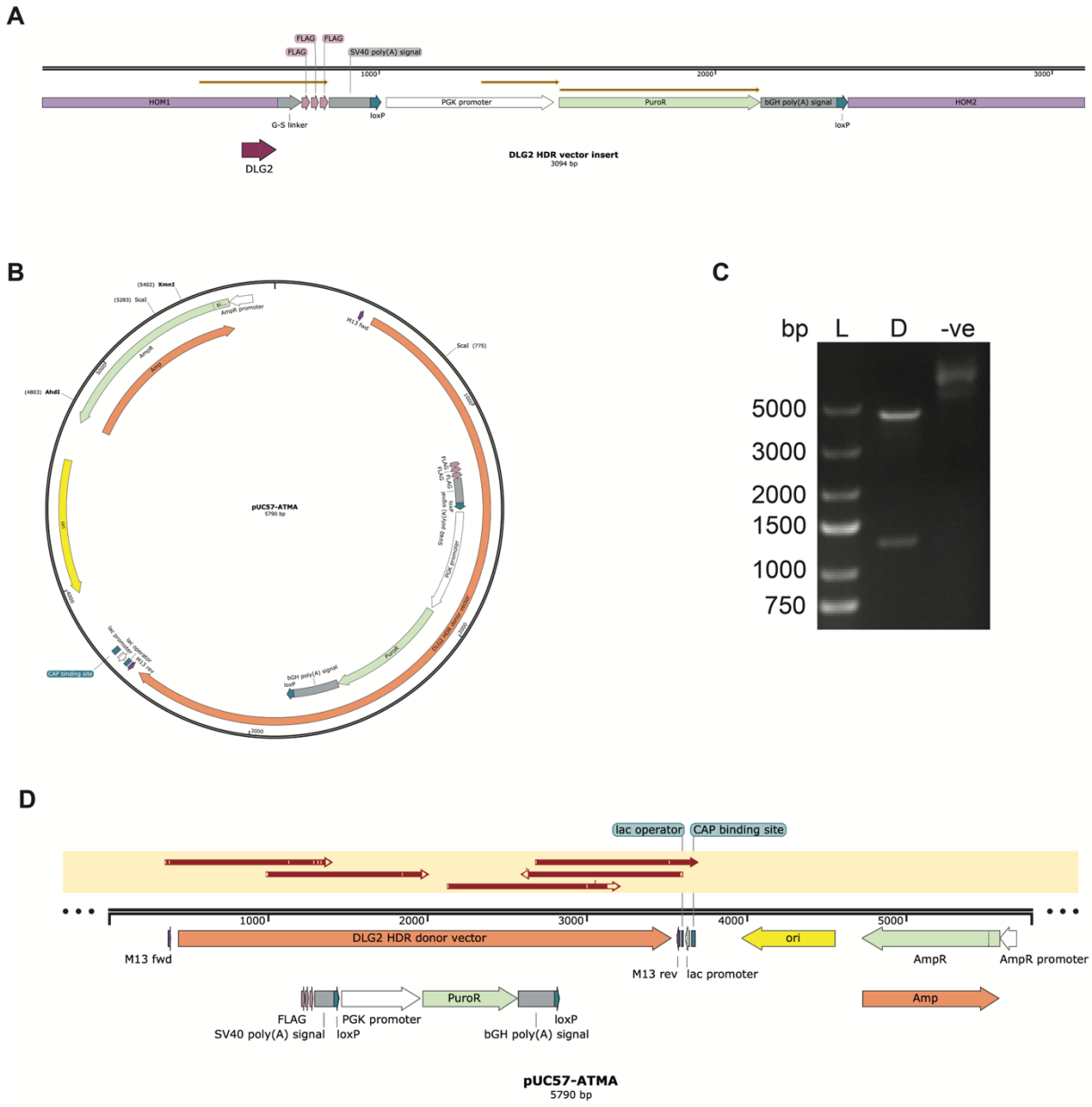


Figure 3.7 HDR vector design and synthesis for multiplex CRISPR/Cas9 genome editing
(A) Map of the HDR template for FLAG-tag insertion into *DLG2*. **(B)** Map of the HDR vector following HDR template insertion into pUC57-ATMA. **(C)** ScaI diagnostic digest confirming correct assembly of HDR vector., L = ladder, D = digested, -ve = undigested control. **(D)** Alignments (in red) of assembled HDR vector with designed template following sanger sequencing, confirming correct assembly.

With correct assembly of a multiplex CRISPR/Cas9 and HDR vector confirmed, H9 hESCs were transfected with both by nucleofection. Although multiple cells survived the transfection process and grew into small colonies, none survived selection with 0.5 µg/ml puromycin, a concentration the puromycin resistance gene should have conferred resistance to if present (Paatero et al., 2008). Multiple variations of the experiment were attempted (Figure 3.8 A) in which the ratio of CRISPR/Cas9 vector to HDR vector were adjusted, puromycin concentration varied and linearised as opposed to circular HDR vector used (Figure 3.8 B); however, in all cases no colonies survived more than 5 days puromycin selection.

A

Experiment No	Total DNA transfected (µg)	Ratio multiplex CRISPR/Cas9 vector : HDR template	[Puromycin] for selection (µg/ml)	Selection commencing (hours post transfection)
1	5	1:1	0.5	48
2	5	1:1 1:3 1:5	0.1 – 0.5 (increasing over 5 days)	48
3	5	1:3 1:5	0.1 – 0.5 (increasing over 5 days)	48
4	5	1:3 1:5	0.1 – 0.5 (increasing over 5 days)	96
5	6	1:6 (linear HDR template)	0.5	96

B

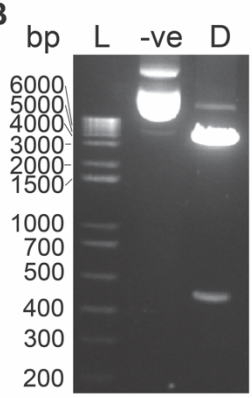


Figure 3.8 Tested conditions for multiplex CRISPR/Cas9 genome editing
(A) Summary of conditions tested for multiplex CRISPR/Cas9 and HDR vector mediated knock-in experiments **(B)** Digests with XmnI and AhdI to produce a linear HDR vector**.** L = ladder, -ve = undigested control, D = digested.

3.5. iCas9 cell line & synthesised gRNAs to generate tagged *DLG2*

As a plasmid based multiplex CRISPR/Cas9 editing approach was not successful in generating FLAG-tagged *DLG2*, an alternative technique using directly synthesised gRNAs (rather than transcribed from plasmid DNA) along with Cas9 protein was devised. As previously shown the gRNAs designed to target the DLG2 stop codon did appear to cut in HEK cells (Figure 3.8), as HEK cells are relatively easy to transfect and edit compared to hPSCs (Xu et al., 2018) this raised the possibility that hESCs may be similarly edited using a more efficient technique. The technique of choice used the iCas9 cell line (González et al., 2014) which contains a doxycycline inducible Cas9 nuclease expression cassette, enabling the protein to be expressed transiently during the course of an experiment where it can form RNP complexes with transfected gRNAs (Figure 3.9 A). The transfected gRNAs were produced using the SygRNA system from SigmaAldrich, in this approach rather than having gRNAs as a single large oligonucleotide they are separately synthesised as crRNA and tracrRNA (Figure 3.9 B). The crRNA contains a 20 base pair sequence, complementary to the region on the genome being targeted for Cas9 cutting, along with an added adapter sequence

to enable annealing to a common tracrRNA sequence to form a complete gRNA. gRNA 1 + 2 that were designed for use in multiplex CRISPR/Cas9 technique previously described (Figure 3.4) were modified for this approach. In addition to being directly synthesised as RNA rather than as DNA oligonucleotides, the synthesised crRNAs did not have an additional G/C or overhangs for BbsI ligation, while they did have an added adapter sequence for annealing to tracrRNA (Figure 3.10 A). As the iCas9 cell lines contains a puromycin resistance gene (Figure 3.10 B) the existing HDR vector which also contained a puromycin resistance gene (Figure 3.6) could not be used as selection for cells that had been edited successfully. As such second HDR vector was designed containing all the elements of the first except for a blasticidin, rather than puromycin, resistance gene (Figure 3.10 C). Again, once designed, the HDR vector was synthesised commercially in the pUC57 plasmid (Figure 3.10 D) and both a restriction digest (Figure 3.10 E) and sequencing (Figure 3.6 F) performed to confirm correct assembly.

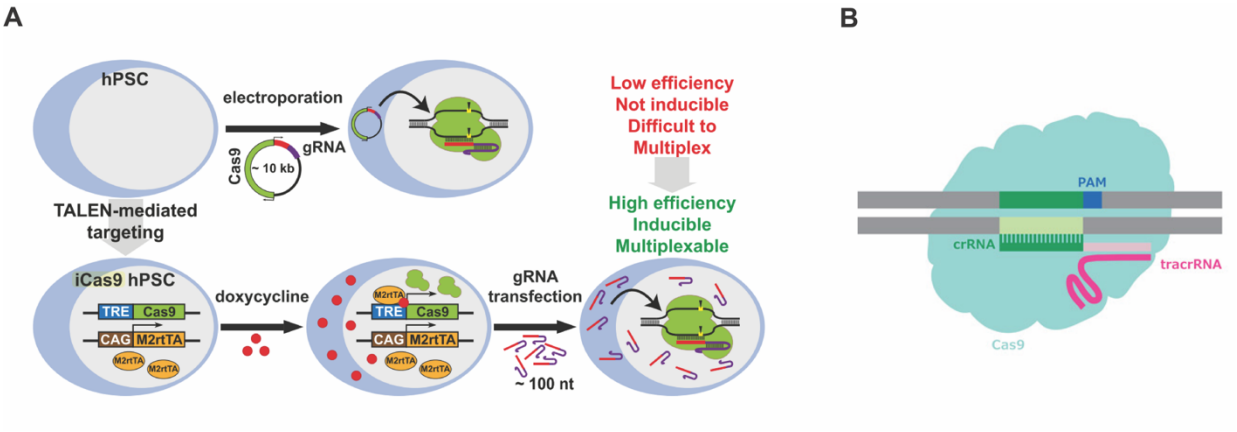


Figure 3.9 Schematic overview of the iCa9 cell line and SygRNA
(A) Overview of the iCas9 cell line, containing a doxycycline inducible Cas9 nuclease cassette. From (González et al., 2014) **(B)** Overview of the SygRNA system in which gRNAs are constructed of crRNA elements specific to genomic targets with an added common adaptor sequence. crRNAs are annealed to a standard tracrRNA scaffold sequence to produce a complete gRNA prior to transfection. From (https://www.sigmaaldrich.com/technical-documents/articles/biology/cas9-ribonucleoprotein-complexes.html).

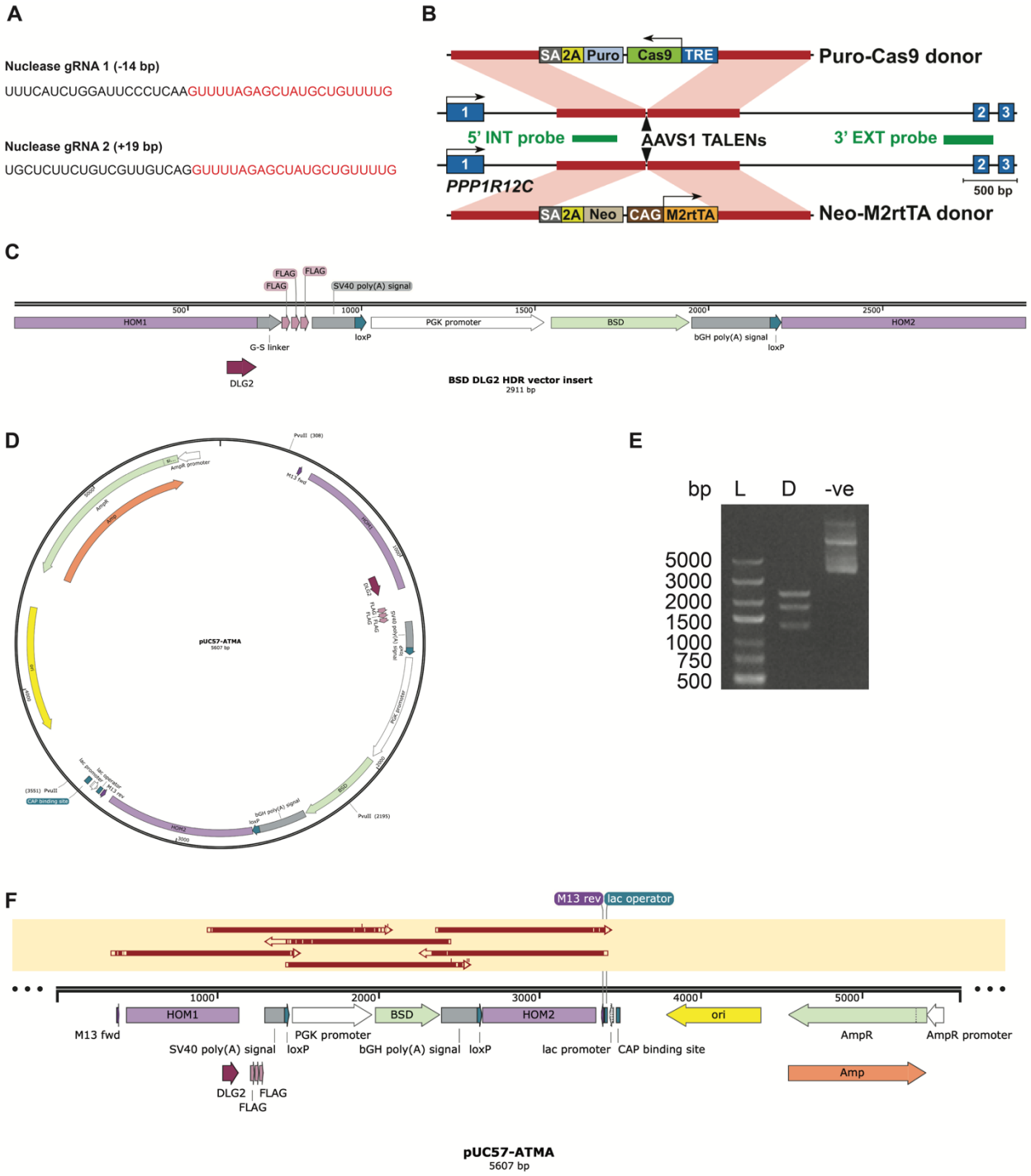


Figure 3.10 gRNA and HDR vector design for iCas9 and SygRNA genome editing
(A) Designed crRNA targeting exon 28 of DLG2 with added adaptor sequences for tracRNA binding shown in red. **(B)** Template used to generate iCAS9 cell line, containing a puromycin resistance gene. From (González et al., 2014) **(C)** Map of the new HDR template for FLAG-tag insertion into *DLG2,* containing a blasticidin resistance gene. **(D)** Map of the BSD HDR vector following HDR template insertion into pUC57-ATMA. **(E)** PvuII diagnostic digest confirming correct assembly of BSD HDR vector. L = ladder, D = digested, -ve = undigested control. **(F)** Alignments (in red) of assembled BSD HDR vector with designed template following sanger sequencing, confirm correct assembly.

Rather than directly using the synthesised HDR vector, either circular or linear, as a template to insert the FLAG-tags and other elements at a CRISPR/Cas9 mediated double stranded breaks, a single stranded HDR (ssHDR) template was produced. ssHDR templates have several advantages over traditional double stranded templates, notably improving editing efficiency and less cell toxicity following transfection (Hornung and Latz, 2010; Miura et al., 2015; Yoshimi et al., 2016; Roth et al., 2018). These templates can be generated by performing PCR to amplify the insert region, including homology arms from an HDR vector before digesting one of the strands from the PCR product to produce a ssHDR template (Figure 3.11 A). Primers were designed to amplify the desired template region from the HDR vector (Figure 3.11 B) both in full and with a slightly reduced length, in case a slight variation in primer position could increase in the efficiency of either the PCR or subsequent digestion reaction to produce a ssHDR template. Additionally, primers were synthesised in both standard and in 5' phosphorylated forms (Figure 3.11 C) as in order to mark a strand of PCR product for digestion by strandase enzymes it must be 5' phosphorylated. These primers were used in various combinations to produce double stranded DNA templates (with one strand 5' phosphorylated) but all in all cases ssHDR templates were produced successfully following strandase digestion with roughly similar efficiency (Figure 3.11 D).

Following synthesis of two gRNAs (consisting of crRNA annealed to tracrRNA) and a ssHDR template to insert FLAG-tags to the gene, iCas9 cells were treated with doxycycline to induce Cas9 nuclease production prior to transfection with both components by lipofection. After successfully using lipofection to transfect multiplex CRISPR/Cas9 vectors, as previously reported (Figure 3.8 B), this approach was also selected for these experiments in which various combinations of synthesised gRNAs and sense/antisense ssHDR templates were tested (Figure 3.12 A). Multiple colonies survived lipofection and several of these were continuing to grow following selection with 0.5 µg/ml blasticidin for 5 days, unlike negative controls which died within 2-3 days. However, despite these transfected cells displaying blasticidin resistance, following PCR using primers flanking the Cas9 cut sites (Figure 3.8 A) all were found to be WT at this location (Figure 3.12 B), indicating the ssHDR templates had not been successfully inserted to *DLG2* locus, but integrated into other genomic locus.

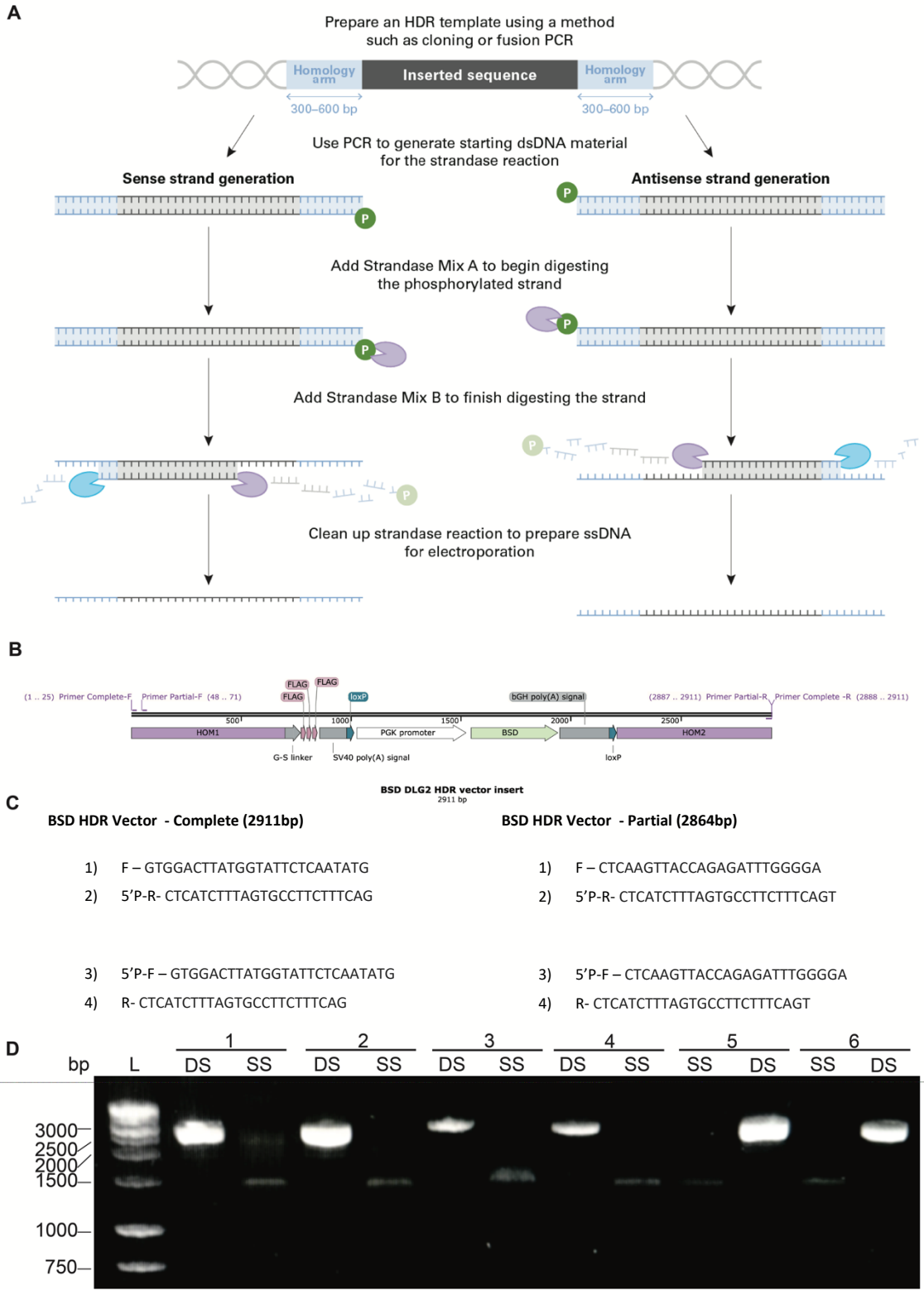


Figure 3.11 Single strand HDR template synthesis
(A) Schematic overview of ssHDR synthesis from a dsHDR vector. From (https://www.takarabio.com/products/gene-function/gene-editing/crispr-cas9/long-ssdna-for-knockins). **(B)** Location of primers for ssHDR template generation, within the BSD HDR vector. **(C)** Standard and 5' phosphorylated primers for ssHDR template generation **(D)** Gel confirming successful synthesis of ssHDR template in 6 separate reactions. L = Ladder DS = double stranded HDR template, SS = single stranded HDR template. ssDNA runs roughly twice as fast as comparable ds DNA and binds DNA dyes poorly, appearing faint upon UV visualisation.

A

Exp No	Total HDR transfected (µg)	Total gRNA transfected (µg)	HDR templates	[Blasticidin] for selection (µg/ml)	Selection commencing (hours post transfection)
1	0.5	0.5	Full sense Full antisense Partial sense Partial antisense Full sense + Full antisense Partial sense + partial antisense	2	72
2	1	1	Full sense Full antisense Full sense + Full antisense Partial sense + partial antisense	2	72

B

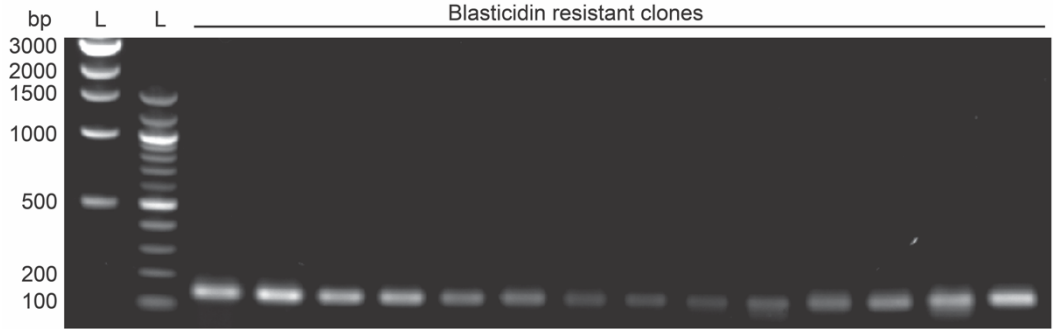


Figure 3.12 Tested conditions and experimental results for iCas9 and SygRNA genome editing
(A) Summary of conditions tested for iCas9, SygRNA and ssHDR template mediated knock-in experiments **(B)** PCR amplification of the Cas9 cut site region showed colonies that survived blasticidin selection were WT at this location (WT at DLG2 locus = 152bp, Edited at DLG2 locus = 1846 bp)

3.6. Cas9 nickase & synthesised gRNAs to generate tagged *DLG2*

In conjunction with Sophia Wilson, an undergraduate student in the lab, a further modification was made to the genome editing method in an attempt to knock-in a FLAG-tag to *DLG2,* as well as reduce the possibility of cutting at unintended sites in the genome. These off-target effects were potentially indicated by colonies surviving antibiotic selection but appearing as WT at *DLG2* stop codon (Figure 3.12 B). To this end the Cas9 nuclease previously used to cause double stranded breaks was replaced with a Cas9 nickase, as this causes single stranded breaks any off-target effects are more likely to be repaired without indels; additionally over hangs created by nickases have been shown to have an increased efficiency for HDR mediated knock-ins compared to the blunt ends generated by nucleases (Ran et al., 2013b). Using this nickase approach therefore required two separate gRNAs are to produce a cut, both creating nicks either 5' or 3' of the desired cut site (Figure 3.13 A). The IDT Alt-R CRISPR-Cas9 system was selected for this purpose, this involved annealing synthesised gRNAs with a Cas9 nickase protein to form an RNP complex, prior to transfection into cells (Figure 3.13 B). This system has several advantages, RNP transfection has improved editing

efficiency compared to plasmid-based methods (Kim et al., 2014) (Figure 3.13 C), IDT provides length optimised crRNA that both increase editing efficiency and are chemically modified to increase nuclease resistance and reduce the innate immune response from cells following transfection (Yin et al., 2017; Schubert et al., 2018). The gRNAs are also separated into two components, crRNA and tracrRNA (Figure 3.13 D), in a similar manner to the Sigma SygRNA system with the added advantage that the tracrRNA contains a fluorescent tag which can be used to verify successful transfection of RNP complexes. New gRNAs were designed for use with the Cas9 nickase utilising evidence showing that, compared to nucleases, nickase cuts can be targeted further

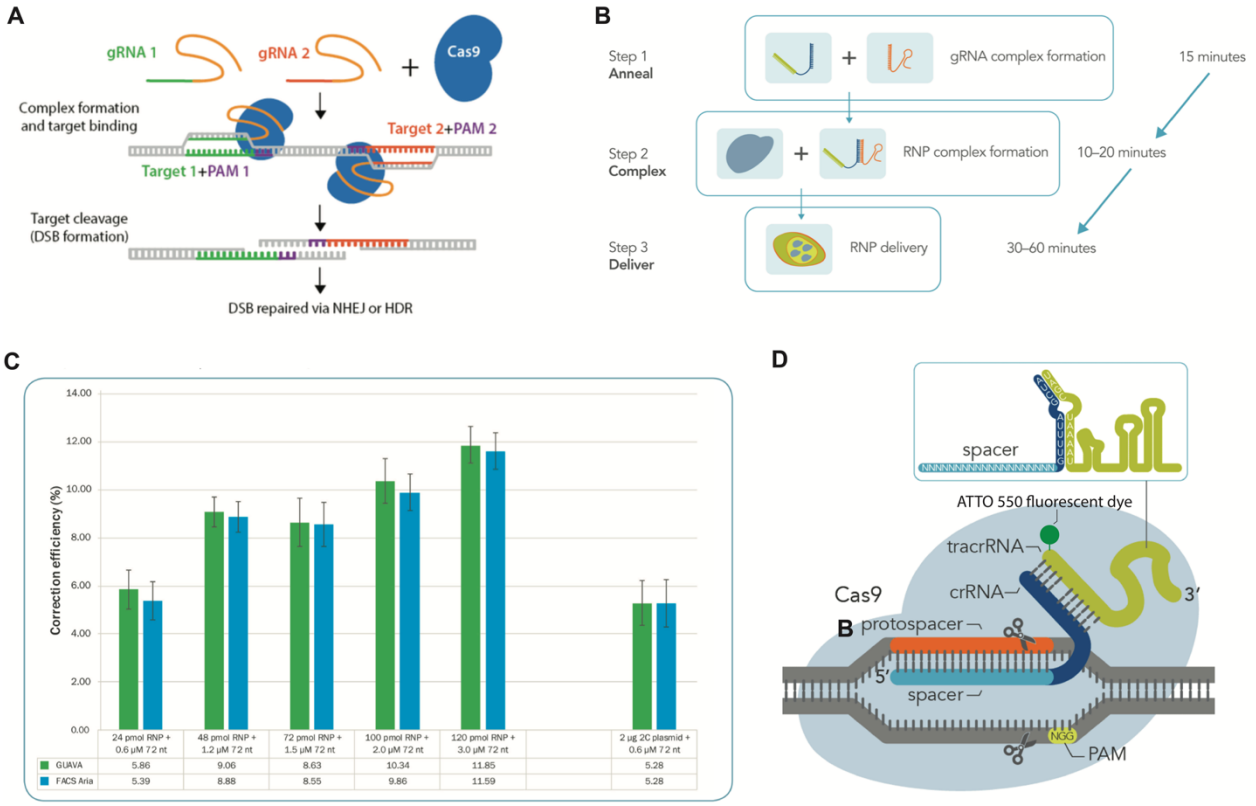
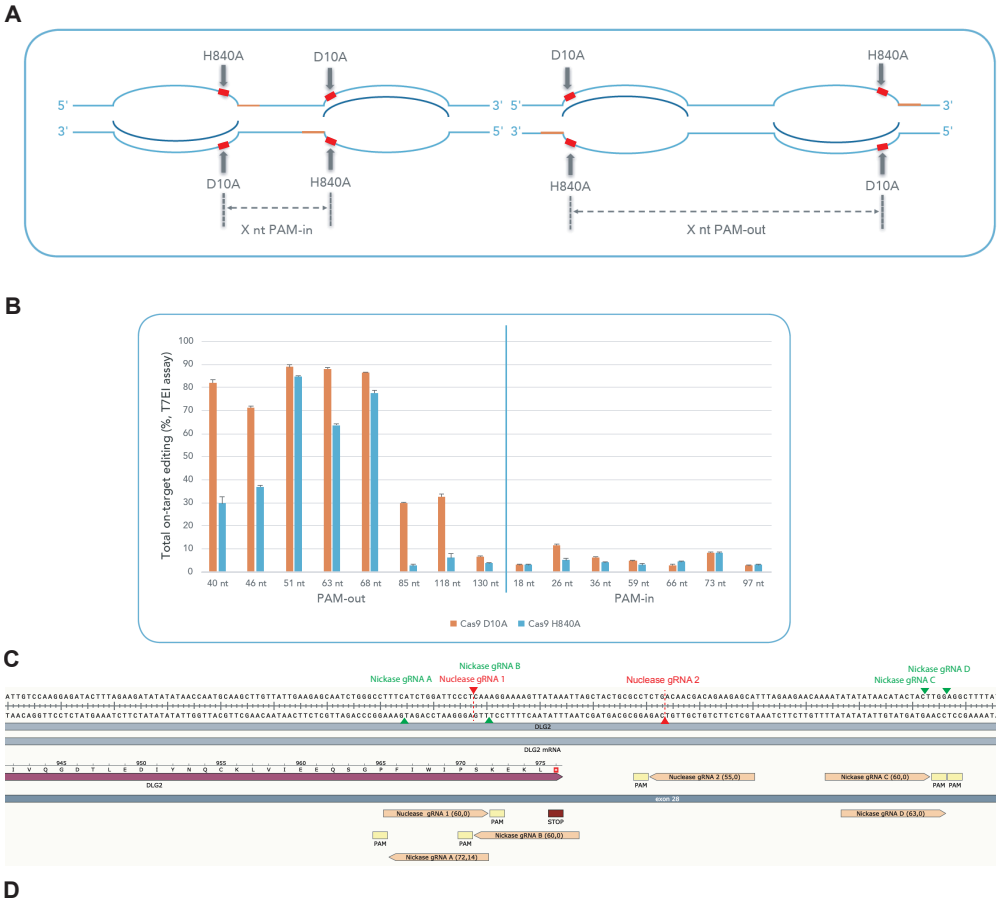


Figure 3.13 Genome editing using Cas9 nickase and the Alt-R CRISPR-Cas9 system
(A) Overview of Cas9 nickase action. From (https://www.addgene.org/crispr/nick/). **(B)** Overview of RNP complex formation using the IDT Alt-R CRISPR-Cas9 system. **(C)** RNP based CRISPR/Cas9 editing shows increased correction efficiency compared to plasmid-based approaches. **(D)** Overview of gRNA components in the Alt-R CRISPR-Cas9 system. (B-D) adapted from (https://www.idtdna.com/pages/products/crispr-genome-editing/alt-r-crispr-cas9-system).

from the desired editing site without a substantial loss of efficiency (Satomura et al., 2017), making more distant gRNAs with higher on-target score feasible. Additionally, although previous guides were designed with a PAM-in configuration, that is PAM sites facing towards the editing site (Figure 3.14 A), there is some evidence to suggests that a PAM-out configuration results in a greater editing efficiency (Yan et al., 2017) (Figure 3.14 B) . As such four PAM-out gRNAs were designed (Figure 3.14 C) and synthesised (Figure 3.14 D), with added adapters and chemical modifications, two 5' and two 3' of the intended editing site at the *DLG2* stop codon. All four had similar on-

target scores or better than those used previously, indicating that cutting should occur at these locations, although off-target scores were low meaning cutting at unintended sites may occur. Off-target effects should be minimised through the use of a nickase, as previously described, and gRNA choice are limited by the location of the desired editing site.



D

Nickase gRNA A = TTTGAGGGAATCCAGATGAA*GUUUUAGAGCUAUGCU*

Nickase gRNA B = AATTTATAACTTTTCCTTTG*GUUUUAGAGCUAUGCU*

Nickase gRNA C = AATATATATAACATACTACT*GUUUUAGAGCUAUGCU*

Nickase gRNA D = ATATATAACATACTACTTGG*GUUUUAGAGCUAUGCU*

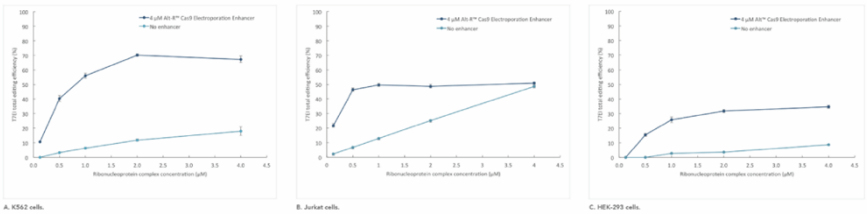
Figure 3.14 gRNAs for genome editing using Cas9 nickase and the Alt-R CRISPR-Cas9 system
(A) Overview of PAM-in versus PAM-out gRNAs, PAM sites in orange, gRNAs in dark blue and nickase cleavage sites in red. PAM-out gRNAs have PAM sites distal to the cut site **(B)** PAM-out gRNAs show increased on-target editing efficiency **(C)** Location of Nickase gRNAs in PAM-out orientation (on and off target scores provided in brackets after names), original PAM-in Nuclease gRNAs shown for comparison. Green triangles indicate nickase cut sites and red triangles nuclease cur sites. **(D)** Designed Alt-R crRNA targeting exon 28 of DLG2 with added adaptor sequences for tracRNA binding shown in red. (A+B) from (https://www.idtdna.com/pages/products/crispr-genome-editing/alt-r-crispr-cas9-system).

Following synthesis of four cRNA these were separately annealed to tracrRNA before complexing with Cas9-D10A nickase. H9 hESCs were transfected with RNP complexes containing various combinations of gRNAs pairs (Figure 3.15 A) and ssHDR template unchanged from that described previously (section 3.5). Nucleofection was once again chosen due to evidence that it triggers a reduced innate immune response in cells compared to lipofection (Wienert et al., 2018), leading to increased cell survival, and the availability of Alt-R Cas9 electroporation enhancer, a Cas9 specific single stranded DNA nucleotides that acts as a carrier to increase transfection, and hence editing, efficiency (Figure 3.15 B). Multiple transfected cells survived nucleofection and in one condition colonies survived selection with 0.5 ug/ml blasticidin for 5 days, unlike all negative controls. PCR was performed on genomic DNA extracted from these colonies using previously described primers flanking the Cas9 cut sites (Figure 3.6 A), while the majority were WT, DNA from two clones produced a larger band of a size consistent with template insertion (Figure 3.15 C), indicating the possibility that *DLG2* had been successfully tagged. However, once the PCR was repeated using primer pairs bracketing the *DLG2* stop codon but more distant, and outside of the homology arms of the ssHDR template (Figure 3.15 D), the two clones containing the insert were both found to be WT at the DLG2 stop codon (Figures 3.15 E + F). This suggests that for at least two clones the entire HDR template inserted into the genome not at the *DLG2* stop codon but at an unknown site. To confirm the initial PCR results (produced by Sophia Wilson) PCR was again performed on genomic DNA extracted from these colonies using primers flanking the Cas9 cut sites (Figure 3.8 A), despite consistent conditions these PCR results showed all clones were WT *DLG2* (Figure 3.17 A) However, when an alternative high-fidelity DNA polymerase was used larger bands of the correct size for the insert were detected (Figure 3.17 B), although not consistently for the two potentially positive clones previously identified (clone 7 + 8) and surprising also in WT cell line controls and human placental DNA. In an attempt to identity these large bands the PCR was repeated once more using a high-fidelity DNA polymerase for clones 7 and 8 as well as WT cell line and placental DNA. Gel electrophoresis showed all PCR samples contained both a smaller WT *DLG2* band and the larger unknown band (Figure 3.17 C), in varying amounts, so TOPO cloning was used to insert PCR fragments from all samples into plasmids which were sent for sanger sequencing. The sequencing results (Figure 3.17 D) showed all clone 7 samples correspond to the insert, clone 8 samples are a mixture of insert and WT *DLG2* while all WT and placental DNA samples were WT *DLG2*. These results therefore confirmed the presence of the insert in two of the clones, as stated previously this does not appear to be at the *DLG2* stop codon, although not central to this work the nature of a similar size band to the insert in WT samples remains unclear.

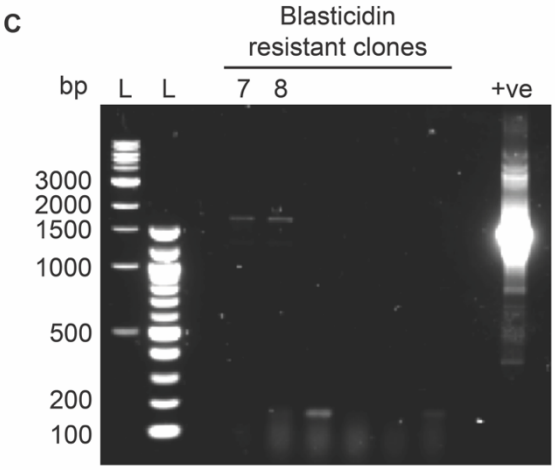
A

Condition	Total HDR transfected (µg)	Total gRNA transfected (µg)	HDR templates	Nickase gRNAs	[Blasticidin] for selection (µg/ml)	Selection commencing (hours post transfection)
1	4	4	Full antisense	A + C	2	72
2	3	3	Full antisense	B + D	2	72
3	3	3	Full sense	A + D	2	72
4	2.4	2.4	Full antisense	B + C	2	72

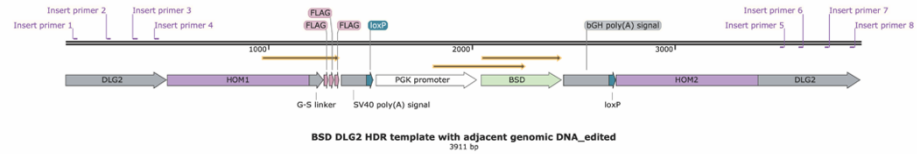
B



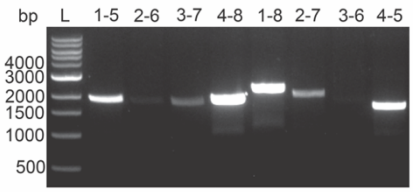
C



D



E



F

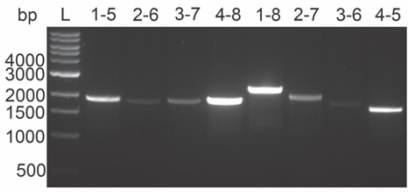


Figure 3.15 Tested conditions and experimental results for Cas9 nickase RNP genome editing
(A) Summary of conditions tested for nickase RNP and ssHDR template mediated knock-in experiments **(B)** Use of electroporation enhancer increases editing efficiency in a range of cell types, from (https://www.idtdna.com/pages/products/crispr-genome-editing/alt-r-crispr-cas9-system). **(C)** PCR amplification using previously designed Cas9 cut site primers showed 2 colonies that survived blasticidin selection may contain the insert (clones 7 and 8) while other appeared WT (WT at DLG2 locus = 152bp, Edited at DLG2 locus = 1846 bp). L = ladder, 7 = clone 7, 8 = clone 8, +ve = positive control (dsHDR vector). **(D)** Primers designed outside of homology arms of ssHDR template to check for on/off-target insertion. **(E)** PCR of clone 7 using Insert primer pairs 1-8, all WT bands of approx. 2000bp (insert – 3500bp) **(F)** PCR of clone 8 using Insert primer pairs 1-8, all WT bands of approx. 2000bp (insert – 3500bp).

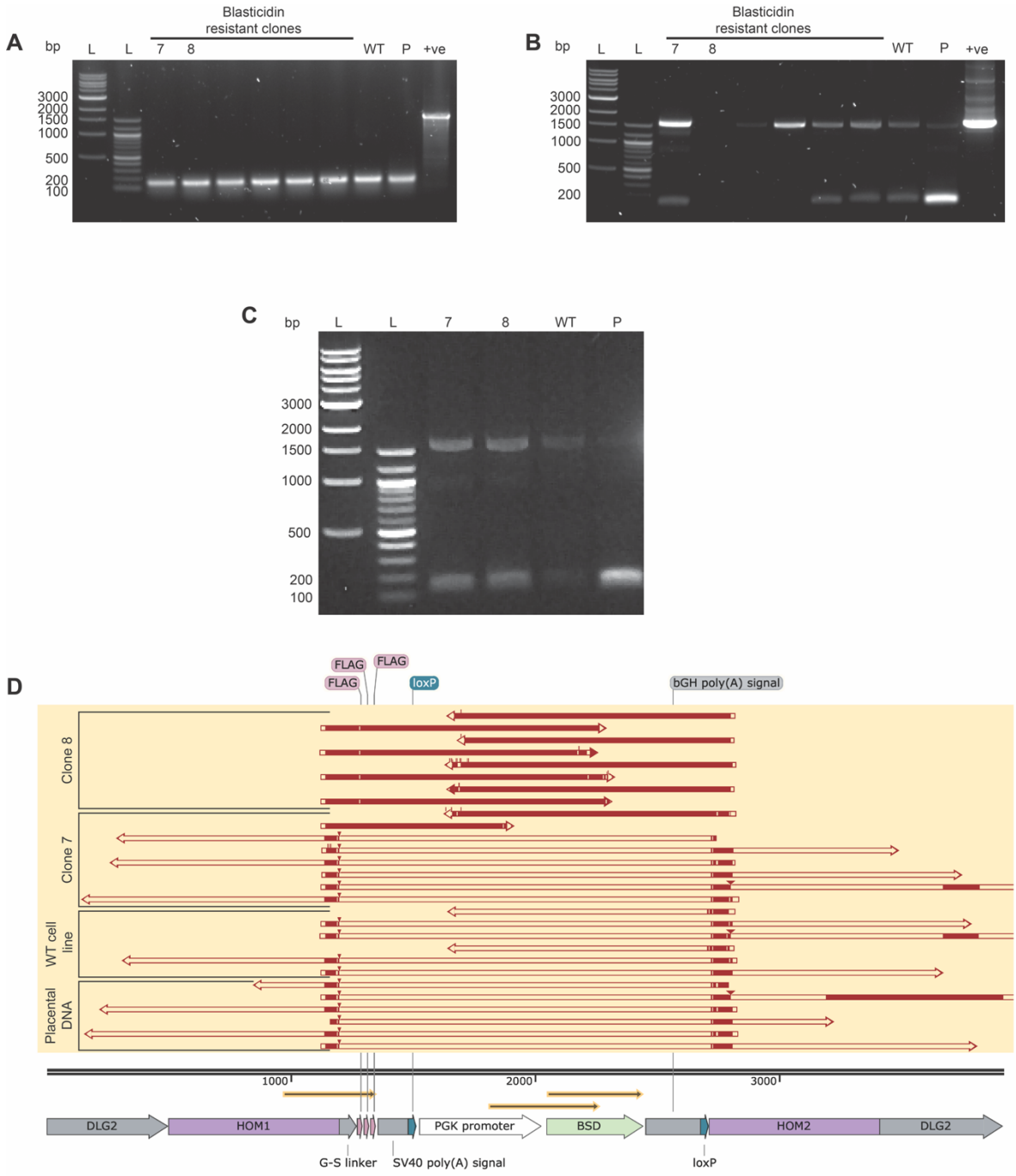


Figure 3.16 Confirming PCR fragment identify following Cas9 nickase RNP genome editing
(A) Repeating PCR amplification using original Cas9 cut site primers and MyTaq polymerase no longer showed any bands corresponding to the insert, all appear WT (WT at DLG2 locus = 152bp, Edited at DLG2 locus = 1846 bp). L = ladder, 7 = clone 7, 8 = clone 8, WT = WT cell line, P = human placental DNA, +ve = positive control (dsHDR vector). **(B)** Repeating PCR amplification using original Cas9 cut site primers and PrimeStar polymerase. Clone 7 showed insert band, as did other samples including WT and placental DNA (WT at DLG2 locus = 152bp, Edited at DLG2 locus = 1846 bp). Key as in (A). **(C)** Repeat of PrimeStar PCR for TOPO cloning all samples show both an inset and WT band (WT at DLG2 locus = 152bp, Edited at DLG2 locus = 1846 bp). **(D)** Alignments (in red) to HDR insert following sanger sequencing of TOPO cloned samples, the sequence in HOM arms also present in WT. Clone 8 are all insert (confirming off target-insertion), clone 7 a mix of insert and WT while the WT cell line and placental DNA are confirmed as WT.

3.7. Discussion

3.7.1 *DLG2* expression during early neurodevelopment

An analysis of published transcriptome data shows that although *DLG2* mRNA is strongly expressed at later development time points, in agreement with its known role in synaptic signalling, it is also expressed at significant levels throughout the course of neurodevelopment. This includes *in vivo* transcriptome data showing *DLG2* expression from 8-10 WPC throughout multiple brain regions, which corresponds to the neurogenesis period of neurodevelopment rather than the later synaptogenesis period which begins around 20 WPC. *In vitro* transcriptome data agrees with the *in vivo* data, showing significant *DLG2* expression during early neurodevelopment, specifically the NPC containing neural differentiation and cortical specification phases, as well as higher levels at later time points of differentiation, corresponding to the generation of more mature neurons.

Western blotting was used in an attempt to confirm that *in vitro* DLG2 protein expression follows as similar pattern to that found in the transcriptome analysis. In agreement with the published RNAseq data DLG2 expression was detected at later time points of cortical differentiation containing postmitotic neurons, potentially from day 40 although with significantly increased levels at days 50 and 60. Through comparison with tagged positive controls this DLG2 expression appears to be mainly isoforms 1 and either 2 or 5. Unlike the transcriptome data western blotting did not show clear DLG2 expression at earlier time points of differentiation. Faint protein bands of the appropriate size for DLG2, as determined by positive controls, were visible at days 15, 20 and 30 although due to issues with antibody quality and the resulting noise on the western blots these results cannot be considered conclusive. Expression of DLG2 at these earlier time points, in NPCs and hence prior to synaptogenesis, would be suggestive of an unknown function for the protein beyond synaptic signal transduction although further experimentation is required to determine if this is the case. In the absence of more specific anti-DLG2 antibody to increase the sensitivity of this approach a FLAG-tag knock-in to *DLG2* remains one possible technique to enable the pattern of endogenous DLG2 expression to be accurately determined, either through western blotting or ICC, although such a knock-in has not been successful to date. Although western blotting only reliably detected DLG2 protein expression from day 50 of cortical differentiation mass spectrometry was performed to better determine DLG2 protein expression which confirmed expression at days 30 and 60, the only 2 time points studied studied using this technique (see Chapter 4 for full details).

3.7.2 FLAG-tag knock-in to *DLG2* using CRISPR/Cas9 approaches was not successful

The lack of high quality and specific antibodies impedes any study into the spatiotemporal pattern of endogenous DLG2 expression. As described earlier this problem may be countered using exogenously produced FLAG-tagged DLG2 which can be run as a positive control in western blotting, enabling the pattern of DLG2 expression during differentiation from hESCs to cortical projection neurons to be determined. Unfortunately, such exogenous controls cannot be used to investigate the spatial pattern of DLG2, hence efforts were made to tag endogenous DLG2 so that its subcellular localisation could be determined. To this end CIRSPR/Cas9 technology was used to attempt to knock-in a FLAG-tag at the *DLG2* stop codon in hESCs.

Various different CRISPR/Cas9 methods were attempted including a plasmid-based approach, a Cas9 expressing cell line transfected with synthesised gRNAs and direct transfection of RNPs, however no attempt was successful in knocking-in a FLAG-tag at the correct location. In several instances colonies did survive antibiotic selection following transfection, indicating the presence of the FLAG-tag containing HDR template, although following PCR and gel electrophoresis the region around the *DLG2* stop codon was shown to be WT. This antibiotic resistance, which was not found in negative controls, but failure to insert the template at the targeted location suggests that either the whole HDR template or at least the antibiotic resistance cassette inserted at another location within the genome. When PCR was performed using primers located within the homology arms of the HDR template, DNA from certain clones did produce a band larger than WT and of the expected size for the entire template; as these were shown to WT at the DLG2 stop codon this implies that the entire HDR template including the resistance gene was inserted at another location in the genome. DNA from other clones did not produce this large band corresponding to the HDR template, only a smaller WT band, yet were still showed antibiotic resistance. In this case, it is possible that part of the HDR template was partially inserted at an off-target site, including at least the resistance gene but not the sections of homology arms containing the primer binding sites. Another possible explanation for antibiotic resistance is that rather than random insertion in the genome the HDR template, in whole or part, is persisting in cells. This is unlikely however as selection was over at least five days, involving several rounds of cell division, meaning the gene would need to be inserted into the genome and replicated for daughter cells and colonies as a whole to have resistance. Also, when a ssDNA template is being used to provide the resistance gene, no expression from a non-integrated template should be possible.

The identity of this larger PCR fragment band found in certain antibiotic resistant clones was confirmed to be the complete HDR template by TOPO cloning and Sanger sequencing. Similarly, samples from untransfected WT cells and placental DNA were sequenced as some showed both a small *WT DLG2* band and an unexpected larger but fainter band of a similar size to the HDR

template. However, all sequenced WT samples were shown to be the WT *DLG2* stop codon region, possibly as the unknown larger fragment was of lower abundance and did not insert during TOPO cloning; although its identity is unknown this larger fragment may well be due to primers binding weakly to an unintended region of the genome with the resulting fragment having a similar size to the HDR template by chance.

Clearly the HDR template can insert a FLAG-tag in the genome as following both PCR and gel electrophoresis as well as sanger sequencing, this appears to have occurred in several clones. However, in all cases this insertion appears to be at an off-target site rather than the DLG2 stop codon as intended. Off-targets effects are always a concern and indeed the gRNAs used to direct Cas9 cutting did possess low off-target scores (indicating more possible off-targets), although it was believed that the use of a Cas9 nickase would reduce the risk of unintended double strand breaks. Furthermore, ssDNA HDR templates have also been shown to have a decreased risk of off-target insertion when compared to dsDNA (Chen et al., 2011; Roth et al., 2018), although this clearly did occur in these experiments. Perhaps most interesting is that while unintended insertion of the HDR template was more frequent than predicted, presumably at sites where off-target Cas9 cuts have occurred, in no cases was the intended site at the *DLG2* stop codon edited. It is possible that the DNA at this site is tightly bound to histones and particularly inaccessible (Horlbeck et al., 2016; Yarrington et al., 2018), indeed no similar knock-ins to DLG2 have been reported in cell lines or animal models. gRNAs targeting this region of DLG2 were tested in HEK cells and did appear to produce indels, indicative of successful cutting, although the precise gRNAs designed for Cas9 nickase experiments were not tested in this manner, raising the possibility they were not effective at targeting the desired locus. It should also be noted that hESCs are considered harder to edit than immortalised cell lines such as HEK 293T (Byrne et al., 2014) so any testing performed in them should be viewed as the ideal condition. As a FLAG-tag was not successfully knocked into *DLG2,* this approached could not be used to definitively identify the protein's subcellular localisation and as such the spatial pattern of DLG2 protein expression during cortical differentiation remains unclear.

3.7.3 Alternatives to CRISPR/Cas9 to investigate the pattern of *DLG2* expression

Although knocking-in a FLAG-tag to *DLG2* has many advantages and would enable the precise subcellular localisation of the protein to be determined, other techniques could be used to investigate the spatiotemporal pattern of *DLG2* expression. To determine if DLG2 protein is expressed In NPCs during early development fluorescence activated cell sorting (FACS) could be used to isolate this cell type prior to protein extraction and western blotting as previously described. This sorting could be achieved either by staining for appropriate NPC cell-surface markers such as CD184(+)/CD271(-

)/CD44(-)/CD24(+) (Yuan et al., 2011) or by transfecting cells with a nestin promoter-GFP-nestin enhancer lentiviral vector (Eberhardt et al., 2006), which would enable nestin expressing NPCs to be detected by their fluorescence. To gain a broad understanding of the subcellular localisation in DLG2 rather than extracting whole protein, nuclear and cytoplasmic fractions could be isolated from all cell types, as well as a synaptosome fraction from neurons (Bai and Witzmann, 2007). This approach would enable comparison between these 3 subcellular compartments in different cell types and stages of neurodevelopment, such analysis could be done by western blotting with appropriate positive control or via peptide pull-down and mass spectrometry (as described in Chapter 4). Yet another method to gain information on the subcellular localisation of *DLG2* would be fluorescence *in situ* hybridisation (FISH) in which *DLG2* mRNA would be tagged with a specific fluorescence probe (Speicher and Carter, 2005), however mRNA localisation may not correlate well to protein localisation.

3.7.4 Possible refinements to CRISPR/Cas9 for knock-in generation

Three distinct CRISPR/Cas9 genome editing techniques were used in an unsuccessful attempt to knock-in a FLAG tag to the DLG2 gene, however many variations on these techniques are possible and could be used to continue this line of research. The two approaches described utilising Cas9 nuclease to create double stranded break could be repeated using gRNAs designed with a PAM-out configuration as this has been shown to increase editing efficiency (Yan et al., 2017). Different gRNAs could be selected with high on and off-target scores, meaning both increased editing efficiency and lower off-target effects, at the cost of increased distance from the desired editing site of the *DLG2* stop codon. The editing site itself could also be changed from the stop codon to the start codon which may enable the use of more specific guides, other locations within the gene could be attempted but would run the risk of an inserted FLAG-tag disrupting the localisation or function of the protein. Variations on the HDR or ssHDR template could also be used, particularly varying the length and symmetry of the homology arms, which asymmetric homology arms showing increased editing efficiency at some genomic locations (Okamoto et al., 2019). Based on recent publications perhaps the most promising technique would be move from a homology based approach to generate the knock-in to homology independent targeted integration (HITI) (Suzuki et al., 2016; Suzuki and Izpisua Belmonte, 2018) (Figure 3.17). This method generates a knock-in using the non-homologous end joining rather than HDR machinery following a double strand break; with a vector containing the same gRNA sequence and PAM that is targeted on the genome to direct integration; and has shown to have greatly improved efficiency particularly in difficult to edit cell types (Suzuki and Izpisua Belmonte, 2018). A HITI based approach has recently been used to

successfully knock-in fluorescent markers to several neural genes, including *DLG4* (Willems et al., 2019), clearly indicating its suitability for a FLAG-tag knock in to *DLG2*.

3.7.5 Summary

An analysis of transcriptome data shows *DLG2* mRNA expression, both *in vivo* within various regions of post-mortem human brain and *in vitro* during cortical differentiation from hESCs, at early neurodevelopmental time points and prior to synaptogenesis. Although western blotting confirmed DLG2 expression *in vitro* at later time points of neuronal differentiation, when postmitotic neurons are present in culture, no definitive protein expression was detected at earlier time points potentially due to a degree of non-specificity in the anti-DLG2 antibody rendering this approach insufficiently sensitive to detect low levels of expression. An effort was made to knock-in a FLAG tag to the *DLG2* stop codon which would enable expression of endogenously produced protein to be accurately determined in the absence of reliable anti-DLG2 antibodies. Several CRISPR/Cas9 methods including a plasmid-based approach, a Cas9 expressing cell line transfected with synthesised gRNAs and direct transfection of RNPs all failed to produce a knock-in at the desired locus. Various refinements on the CRISPR/Cas9 gene editing approaches have been identified as well as alternative to investigate the pattern of DLG2 expression which could be the focus of future experiments.

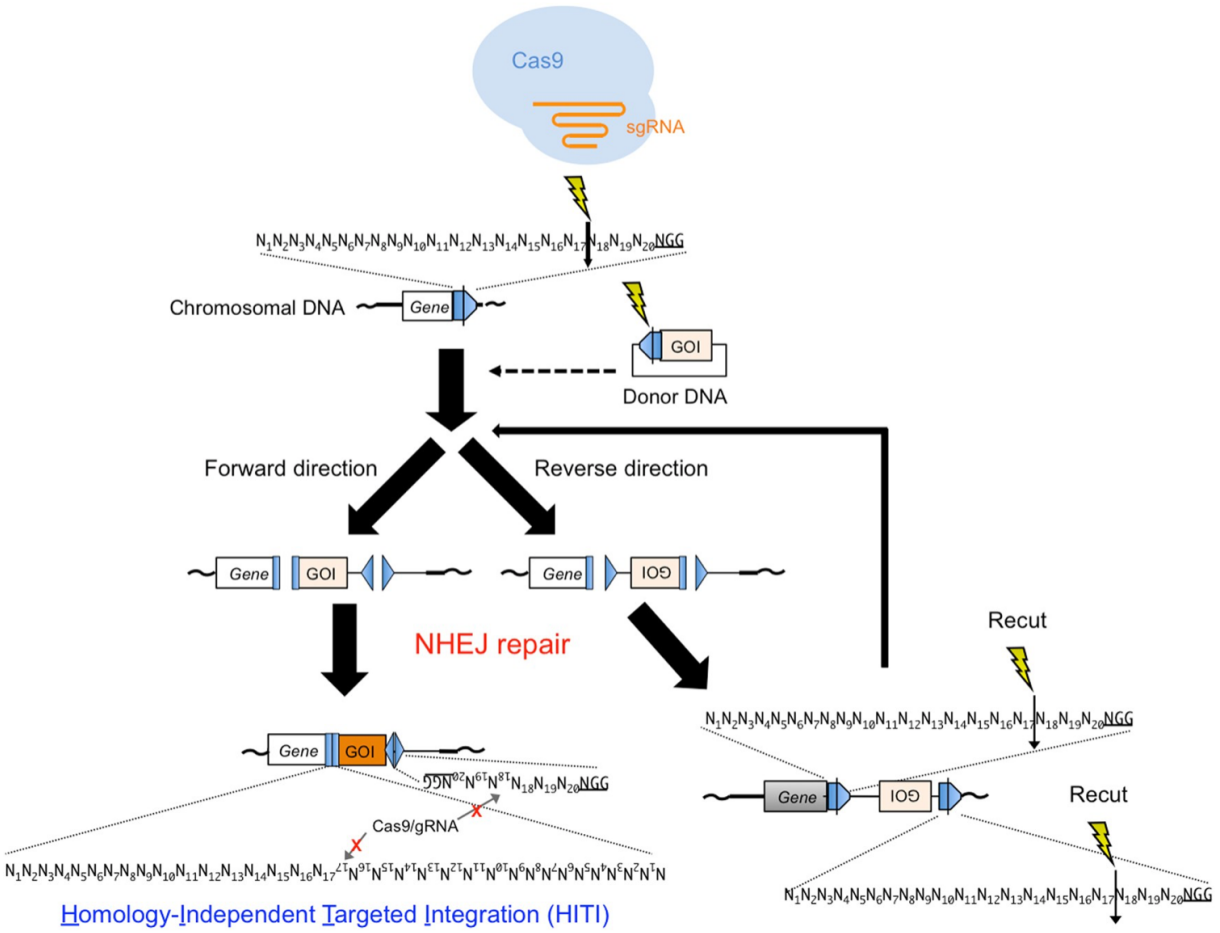


Figure 3.17 Homology independent targeted integration (HITI)
Schematic of HITI using SpCas9. The SpCas9/sgRNA complex introduces DSBs 3 base pairs upstream of the PAM sequence in the genomic target sequence, resulting in two blunt ends. The exact same SpCas9/sgRNA target sequence is loaded onto the donor DNA in the reverse direction. Targeted genomic loci, as well as the donor DNA, are cleaved by Cas9/gRNA and the linearized donor DNAs are integrated into target sites via the NHEJ DSB repair pathway. If donor DNA is integrated in the correct orientation, junction sequences are protected from further cleavage by Cas9/gRNA. If donor DNA integrates in the reverse orientation, Cas9/gRNA will excise the integrated donor DNA due to the presence of intact Cas9/gRNA target sites. Blue pentagon, Cas9/gRNA target sequence. Black line within blue pentagon, SpCas9 cleavage site. GOI, gene of interest. From (Suzuki and Izpisua Belmonte, 2018).

4. Characterisation of the *DLG2*$^{-/-}$ phenotype in human NPCs and neurons

4.1. Introduction

The established pattern of expression for DLG2 protein is within the neuronal PSD, where it acts as a scaffold protein to regulate receptor and ion channel clustering and facilitate synaptic signal transmission through interaction with other adaptor proteins (Zheng et al., 2011). However, synaptogenesis is a relatively late stage of human neurodevelopment (Semple et al., 2013) and data presented in the previous chapter indicated that *DLG2* mRNA is expressed both *in vitro* and *in vivo* substantially prior to this, during the initial stages of neurogenesis (see Chapter 3). Attempts to investigate the spatiotemporal pattern of DLG2 protein expression in WT cells during cortical differentiation were hampered by both issues with antibody specificity and the inability to knock-in a FLAG tag to the gene but did show expression *in vitro* at later time points containing postmitotic neurons, at earlier time points the pattern of expression was less clear requiring further investigation. One possible approach to both indirectly confirm DLG2 expression during early neurodevelopment and gain insight into its function is to analyse any resulting phenotypes when expression levels are perturbed. In a wider context this may be done by studying human patients in which the gene of interest is mutated but this comes with practical challenges for a gene such as *DLG2* studied in relation to neuropsychiatric disease, where affected individuals often possess *de novo* mutations (Kirov et al., 2012a) and with diagnosis typically in adolescence (Thapar and Riglin, 2020) studying the gene's effects on early human neurodevelopment are not possible, even if the obvious ethical concerns were to be ignored. A suitable model system in which the levels of *DLG2* have been manipulated is therefore the most suitable approach for studying the gene's role in early human neurodevelopment. Various approaches exist to manipulate the expression levels for a gene of interest: overexpression in an endogenous cell type or heterologous host (Prelich, 2012), a knockdown to reduce expression (Boettcher and McManus, 2015) or a knockout to eliminate expression (Ran et al., 2013a). Out of these options the clearest way to determine the function of a gene, rather than to model a particular disease, is to perform a knockout so this approach was selected for *DLG2*. If a gene knockout is embryonically lethal or if the function of a gene at a specific point in development is being studied, knockouts can be made conditional, using techniques such as Cre-Lox recombination to trigger the knockout at the desired time point (Song and Palmiter, 2018); however, as unconditional *DLG2* knockout mouse models have previously been successfully generated (Nithianantharajah et al., 2013) and the purpose of this research was to study the effect of *DLG2* on neurodevelopment in general rather than at a specific time point, such an approach was not used.

hESCs were selected as an appropriate model system to study the role of *DLG2* in early human neurodevelopment and the CIRSPR/Cas9 system selected to edit the *DLG2* gene due to its relatively

low cost, ease of customisation and higher targeting efficiency compared to other approaches. As the focus of this research was to determine the role of *DLG2* in neurodevelopment rather than to produce a direct disease model a homozygous knockout line (*DLG2*$^{-/-}$) was produced. As a hESC line was edited no additional SNPs were present to act as confounding factors, thus ensuring the observed phenotypes were the result of *DLG2* deficiency. Given the focus of this research on the role of DLG2 in early neurodevelopment, rather than in mature functional neurons, a developmental patterning approach which does not accelerate or bypass the NPCs stage (Schafer et al., 2019) was selected to differentiate hESCs towards cortical projection neurons.

The aim of the research presented in this chapter was therefore to characterise the *DLG2*$^{-/-}$ phenotype in human NPCs and neurons to determine if there may be currently unreported role for DLG2 in addition to its established function within the PSD. To this end two *DLG2*$^{-/-}$ hESC lines previously generated were validated through various approaches, this included the use of mass spectrometry to verify full knockout at the protein level and also clarify the pattern of DLG2 expression in WT cells during cortical differentiation. Following differentiation of the *DLG2*$^{-/-}$ lines and WT controls towards cortical projection neurons ICC was used to quantify neuronal morphology while a combination of ICC and western blotting used to characterise the expression of various well-established NPC and neuronal makers at various stages of neurodevelopment.

4.2. Experimental procedure – outline

Two *DLG2*$^{-/-}$ hESC lines (KO1 and KO2) that had been previously generated by Dr Eunju Jenny Shin (NMHRI, Cardiff University) (Sanders et al., 2020) using CRISPR/Cas9-D10A nickase genome editing (Chiang et al., 2016) were validated by multiple approaches. Peptide affinity pulldown followed by mass spectrometry was used both to confirm full DLG2 knockout at the protein level as well as investigate WT DLG2 expression during early cortical differentiation, mass spectrometry being performed by Dr Mark Collins (Department of Biomedical Science, University of Sheffield). To further validate these *DLG2*$^{-/-}$ hESC lines PCR amplification of predicated off-target sites and sanger sequencing along with CNV analysis were used to confirm no off-target effects as well as ICC used to analyse OCT4, SOX2 and NANOG expression to confirm pluripotency (Kashyap et al., 2009).

Following validation, both *DLG2*$^{-/-}$ lines along with a genetically WT sister line were differentiated from hESCs to cortical projection neurons using a developmental patterning protocol. To characterise the *DLG2*$^{-/-}$ phenotype at 5 time points during cortical differentiation (days 20, 30, 40, 50, 60), the expression of various well-established NPC and neuronal markers, identified from studies of neurodevelopment *in vivo*, were investigated through a combination of western blotting and ICC analysed by high content imaging. These markers included SOX2 (Hutton and Pevny,

2011), FOXG1 (Qi et al., 2017) and PAX6 (Englund et al., 2005) which were used to characterise NPCs and confirm a dorsal telencephalic fate. The markers NEUN (Mullen et al., 1992; Duan et al., 2016) and MAP2 (Soltani et al., 2005) were used to general postmitotic neuronal markers, TBR1 and CTIP2 for deep player cortical neurons of layers VI and V respectively (Molyneaux et al., 2007) and SATB2 (Britanova et al., 2008) for upper layer cortical neurons. A morphometric analysis was also performed on days 30 and 70 *DLG2*$^{-/-}$ and WT neurons following transfection with a GFP plasmid to enable the cell boundaries to be clearly identified

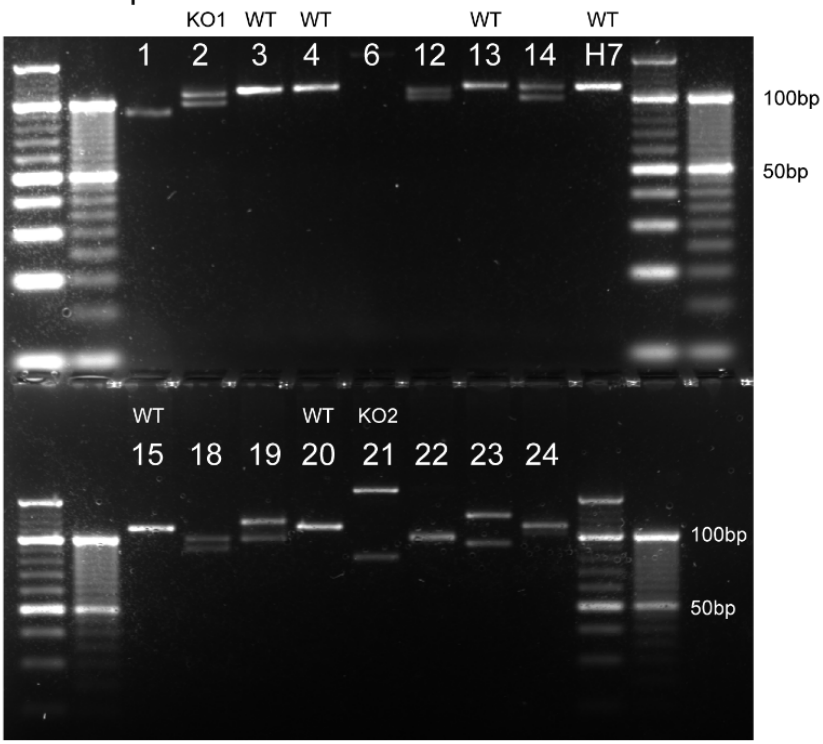
As described in Section *2.20 Statistical analysis and data presentation* all ICC and western blot characterisation results were from a minimum of two independent differentiations, unless otherwise stated in figure legends, while morphometric analysis results were from a single differentiation. Detailed protocols can be found in Chapter 2*,* Sections *2.1 Cell maintenance, 2.2 Cortical differentiation, 2.3 Western blotting, 2.6 PCR & agarose gel electrophoresis, 2.8 Sanger sequencing, 2.9 DLG2 knockout hESC generation, 2.10 Genetic validation of DLG2 knockout hESC lines, 2.11 Immunocytochemistry, 2.12 Peptide affinity pulldown & mass spectrometry and 2.13 Morphometric analysis*.

4.3. Generation of *DLG2*$^{-/-}$ hESC lines

Two previously generated homozygous *DLG2* knockout hESC lines were made available for this research (Sanders et al., 2020). These *DLG2*$^{-/-}$ lines were produced through genome editing of H7 hESCs using the CRISPR/Cas9-D10A nickase system, with two gRNAs designed to direct Cas9-D10A nickase mediated cutting within exon 22 of *DLG2,* which encodes part of the protein's first PDZ domain (Figure 4.1 A). To confirm editing, multiple hESC clones were screened by PCR amplification of the region surrounding the targeted Cas9 cut sites followed by agarose gel electrophoresis. Unlike the single band of H7 parental lines, several edited clones showed two bands, indicative of the presence of indels and hence successful editing (Figure 4.1 B). To investigate if the editing generated KOs, PCR amplicons for all clones were sent for sanger sequencing which showed two clones to be homozygous knockouts, these were named KO1 and KO2 (Figure 4.1 C). For both KO1 and KO2, the indels resulting from CRISPR/Cas9 genome editing generated premature stop codons within exon 22 of DLG2, meaning the resulting mRNA transcripts are likely to be degraded by nonsense-mediated decay (Figure 4.1 D). One genetically WT sister line, that is a line where CRISPR/Cas9 editing was attempted but failed leaving it WT as confirmed by sanger sequencing, was retained for use as a control.

A 1st gRNA sequence - 5' GGCTTCCACCGCTTTACTGT 3'
2nd gRNA sequence - 5' AGGAAGCAGGGTCTATCGTT 3'

B Forward primer: 5' ATCTTGCGGGTGAATGAGGTTG
Reverse primer: 5' GTCGTCTTCTACGCACATACAG



C

Cut by Cas9 nickase

WT 5' GAGGTTGATGTGTCAGAGGTTTCCCACAGTAAAGCGGTGGAAGCCCTGAAGGAAGCAGGGTCTATCGTTCGG
CTC CAACTACACAGTCTCCAAAGGGTGTCATTTCGCCACCT TCGGGACT TCCT TCGTCCCAGATAGCAAGCC 5'

Cut by Cas9 nickase

KO1 allele 1 5' GAGGTTGATGTGTCAGAGGTTTCCCACAGTAAAGCG~~GTGG~~AAGCCCTGAAGGAAGCAGGGTCTATCGTTCGG

KO1 allele 2 5' GAGGTTGATGTGTCAGAGGTTTCCCACAGTAAAGC~~GGTGGAAGCCCTG~~AAGGAAGCAGGGTCTATCGTTCGG

KO2 allele 1 5' GAGGTTGATGTGTGGGAAACCTCTGAACAGGATTATCTGAGGTTCTCCTTGAACTTTGTAAATCTGAACTTCAG
TATTTATAATATTATAAGATGAATCGAGGTGGAG~~CAGAGGTTTCCCACAGTAAAGCGGTGGAAGCCCTGAAGGA~~
AGCAGGGTCTATCGTTCGG

KO2 allele 2 GAGGTTGATGTGTCAGAGGTTTCCCACAG~~TAAAGCGGTGGAAGCCCTGAAGGAAGCAG~~GGTCTATCGTTCGG

D

WT	V D V S E V S H S K A V E A L K E A G S I V R L Y V R R R R P I L E T V V E I K L F K G P K G L G
KO1 allele 1	V D V S E V S H S K A K P *
KO1 allele 2	V D V S E V S H S K A R K Q G L S F G C M C V E D D L F W R P L W K S N C S K A L K V *
KO2 allele 1	V D V W E T S E Q D Y L R F S L N F V N L N F S I Y N I I R *
KO2 allele 2	V D V S E V S H R V Y R S A V C A *

Figure 4.1 Generation of *DLG2*$^{-/-}$ hESC lines
(A) Two guide RNA sequences used to target exon 22 of *DLG2* which is translated into part of the 1st PDZ domain. **(B)** A primer set used for screening successful indels on targeted DNA. PCR and gel electrophoresis show several hESC clones with indels. Parental H7 lines show a single 109bp band whereas several clones, such as 2 and 21, show two different sized bands suggesting generation of indels on the targeted exon of both copies of the gene. **(C)** PCR amplicons were sequenced and clones no.2 and 21 found to be homozygous knockouts (named KO1 and KO2). Red letters indicate deleted bases while blue indicate inserted DNA. **(D)** Indels in KOs generated premature stop codons on the targeted exon. Green indicates different amino acid sequences from WT. From (Sanders et al., 2020).

4.4. Mass spectrometry confirms WT DLG2 expression at day 30 and full KO in *DLG2*$^{-/-}$ lines

To determine if KO1 and KO2 were genuine DLG2 knockout cell lines mass spectrometry was performed to investigate the levels of unique DLG2 peptides. Western blotting and ICC were initially considered as techniques to investigate DLG2 protein expression and confirm the knockout; however, given the known issues with DLG2 antibody specificity (see Chapter 3) these approaches were not feasible. To this end KO1, KO2 and WT hESCs were differentiated towards cortical projection neurons and protein samples for mass spectrometry collected at two time points of differentiation. Given the established role of DLG2 within the PSD day 60 was chosen as a later time point as this should contain more mature neurons with synapses and hence DLG2 expression in WT cells should be relatively high, as was also indicated by western blotting (Figure 3.2 A). Day 30 was selected as an additional time point to determine whether DLG2 is also expressed in cultures containing new-born neurons, as western blotting results for DLG2 expression during the early stages of cortical differentiation *in vitro* were unclear. On the principle that protein extracts would be heterogenous with a relatively low level of DLG2 compared to other proteins, potentially below the technical limits of detection by mass spectrometry, the decision was taken to enrich for PDZ domain containing proteins which include the majority of DLG2s isoforms. This enrichment was achieved by peptide-affinity pulldown using a known PDZ domain interactor, the NMDA receptor NR2 subunit (Husi and Grant, 2001), with the resulting day 30 and day 60 PDZ domain enriched protein samples being sent for quantitative mass spectrometry. An analysis of these results both for unique DLG2 peptides and label free quantification (LFQ) intensity for DLG2, a measure of intensity normalised across samples, showed that DLG2 is not only expressed in day 60 samples as expected but also at a lower level earlier in cortical differentiation at day 30 (figure 4.2). This analysis also indicate that both KO1 and KO2 were full DLG2 knock-outs which were absent unique DLG2 peptides at both days 30 and 60 while showing a significant reduction in LFQ intensity for DLG2 (Figure 4.2).

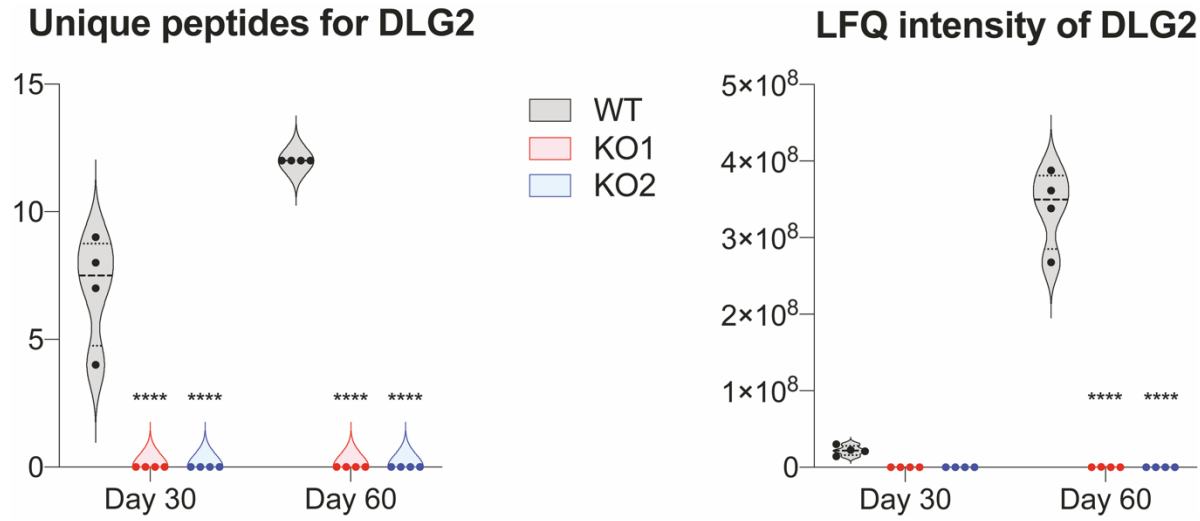


Figure 4.2 Mass-spec quantification of DLG2 expression in WT and *DLG2*$^{-/-}$ cell lines
Quantification of DLG2 expression in PDZ-ligand (NR2 C-terminus) affinity pulldowns samples from WT and *DLG2*$^{-/-}$ cells lines (KO1 & KO2) at days 30 and 60 of cortical differentiation using LC-MS/MS analysis. DLG2 expression is present in all WT samples at both time points and significantly reduced *DLG2*$^{-/-}$ lines, as determined by both the number of unique DLG2 peptides and label free quantification (LFQ) intensity. Analysis was by two-way ANOVA: unique peptides genotype ($F_{2,18}$= 309.4; P<0.0001; n=4), time ($F_{1,18}$= 21.43; P=0.0002; n=4), interaction ($F_{2,18}$= 21.43; P<0.0001; n=4); LFQ intensity genotype ($F_{2,18}$= 193.0; P<0.0001; n=4), time ($F_{1,18}$= 149.0; P<0.0001; n=4), interaction ($F_{2,18}$= 148.8; P<0.0001; n=4).

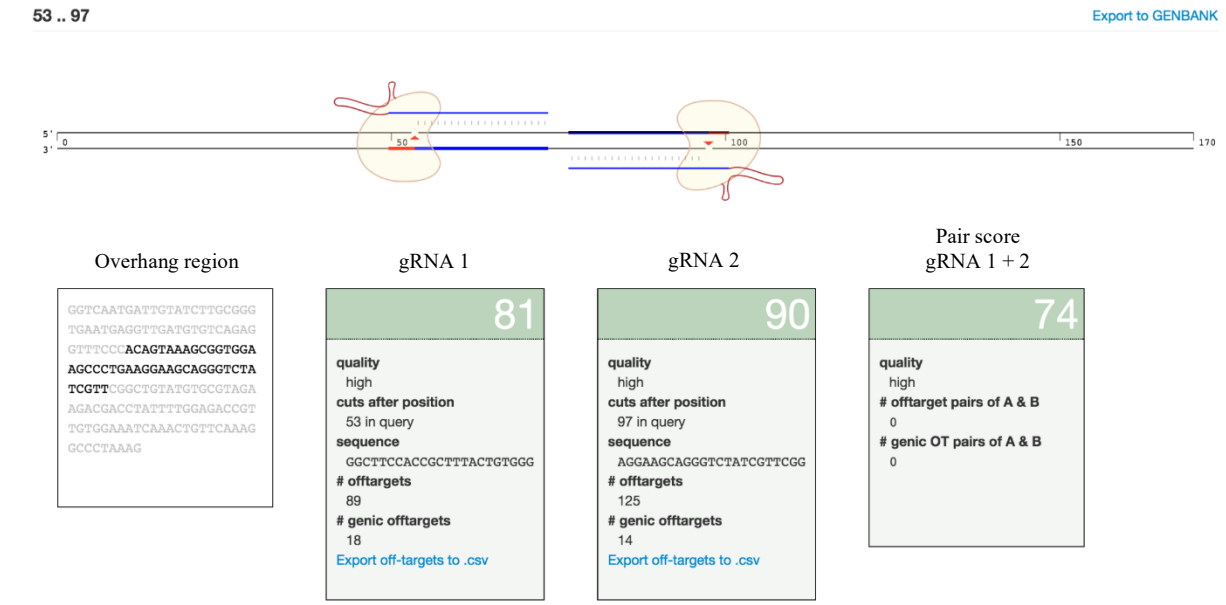
4.5. *DLG2*$^{-/-}$ lines show no off-target effects and retain pluripotency

Following confirmation that KO1 and KO2 were full DLG2 knock-out lines potential off-target effects of the Cas9/D10-A nickase were investigated to ensure editing had only occurred at the desired location within exon 22 of *DLG2* and not at other locations within the genome. The two gRNAs used during the generation of KO1 and KO2 both had several predicted off-target effects when used individually with a Cas9 nuclease to cause double stranded breaks (Supplementary Table 1); however, as the alternative Cas9/D10-A nickase approach was used to generate these hESC lines, with both gRNAs targeting the creation of a separate single-strand break, the combined gRNA pair had no predicted off target effects (Figure 4.3 A). To confirm this prediction of no off-target effects when using a Cas9 nickase, a total of fourteen randomly selected genic off-target sites identified from the Cas9 nuclease off-target predictions were selected for investigation. Genomic DNA was extracted from KO1, KO2 and WT hESCs and primers designed flanking the predicted off-target sites, which were amplified by PCR; sanger sequencing of the resulting amplicons showed no difference between either KO lines and WT at any site (Figure 4.3 B, C).

As an additional validation stem for the lines genotyping was performed to check for any deleted or duplicated regions with the genomes of KO1 and KO2. The results of this genotyping and subsequent CNV analysis did highlight two pre-existing, non-pathogenic, deleted and duplicated

regions in all three hESC lines compared to the reference genome; however, there was once again no difference between either KO lines and WT (Figure 4.4). Finally In order to confirm that KO1 and KO2 were indeed genuine hESC lines ICC was performed to determine the expression of three key markers of pluripotency, OCT4, SOX2 and NANOG, all of which were expressed in the *DLG2*$^{-/-}$ lines at 100% of WT levels (Figure 4.5). Taken together these validation data therefore confirm that KO1 and KO2 are homozygous DLG2 knockout hESCs lines, with no detectable off target effects or other genetic abnormalities compared to a WT sister line.

A



B

gRNA 1

on target	n_mismatches	score	chromosome	strand	position	gene	reference sequence	WT cell line sequence	KO1 cell line sequence	KO2 cell line sequence
False	4	0.450877754	chr7	-1	31150952	NM_001199636	GGCTTACTTAGCTTTACTGTGGG	GGCTTACTTAGCTTTACTGTGGG	GGCTTACTTAGCTTTACTGTGGG	GGCTTACTTAGCTTTACTGTGGG
False	3	0.200806759	chr1	1	90459736	NR_002830	GGATTCCACCGCCTTCCTGTAGG	GGATTCCGCCGCCTTCCTGTAGG	GGATTCCGCCGCCTTCCTGTAGG	GGATTCCGCCGCCTTCCTGTAGG
False	4	0.17599752	chr9	-1	103340717	NM_001018116	GGCTTTCACATCTTTAATGTGAG	GGCTTTCACATCTTTAATGTGAG	GGCTTTCACATCTTTAATGTGAG	GGCTTTCACATCTTTAATGTGAG
False	4	0.092421904	chrX	1	14038524	NM_001042479	GGATTCCACGGCCTTCCTGTAGG	GGATTCCACGGCCTTCCTGTAGG	GGATTCCACGGCCTTCCTGTAGG	GGATTCCACGGCCTTCCTGTAGG
False	4	0.079007985	chr12	1	57923866	NM_001261412	GGCCTCCACTGCTTTTCTATGGG	GGCCTCCACTGCTTTTCTATGGG	GGCCTCCACTGCTTTTCTATGGG	GGCCTCCACTGCTTTTCTATGGG
False	4	0.046785459	chr16	1	71894409	NM_001201556	GGCTTCCAGCAGTTTACGGTAGG	GGCTTCCAGCAGTTTACGGTAGG	GGCTTCCAGCAGTTTACGGTAGG	GGCTTCCAGCAGTTTACGGTAGG
False	4	0.012046771	chr6	1	47662977	NM_153839	GGCTTCCATCTCTATGCTGTTGG	GGCTTCCATCTCTATGCTGTTGG	GGCTTCCATCTCTATGCTGTTGG	GGCTTCCATCTCTATGCTGTTGG

C

gRNA 2

on target	n_mismatches	score	chromosome	strand	position	gene	reference sequence	WT cell line sequence	KO1 cell line sequence	KO2 cell line sequence
False	4	0.23362076	chr8	1	126958054	NR_038446	AGGAAGATTGGTCTATGGTTAGG	AGGAAGATTGGTCTATGGTTAGG	AGGAAGATTGGTCTATGGTTAGG	AGGAAGATTGGTCTATGGTTAGG
False	4	0.158578648	chr1	-1	22310277	NM_007352	ATGTAGCAGGGTGTCTCGTTGGG	ATGTAGCAGGGTGTCTCGTTGGG	ATGTAGCAGGGTGTCTCGTTGGG	ATGTAGCAGGGTGTCTCGTTGGG
False	4	0.157225945	chr15	1	33059133	NM_001103184	AGTAAACAGGGTCTATGGTCAAG	AGTAAACAGGGTCTATGGCCAAG	AGTAAACAGGGTCTATGGCCAAG	AGTAAACAGGGTCTATGGCCAAG
False	4	0.099818052	chr12	1	106457104	NM_014840	AAGAAGCAAGGTCTCTGGTTCAG	AAGAAGCAAGGTCTCTGGTTCAG	AAGAAGCAAGGTCTCTGGTTCAG	AAGAAGCAAGGTCTCTGGTTCAG
False	4	0.058898835	chr14	1	24807080	NM_006871	TGGAAGGAGGGTCTGTCCTTGGG	TGGAAGGAGGGTCTGTCCTTGGG	TGGAAGGAGGGTCTGTCCTTGGG	TGGAAGGAGGGTCTGTCCTTGGG
False	4	0.031081415	chr1	1	151008981	NM_138278	AGGAGGCAGGGTTTTTCTTTTGG	AGGAGGCAGGGTTTTTCTTTTGG	AGGAGGCAGGGTTTTTCTTTTGG	AGGAGGCAGGGTTTTTCTTTTGG
False	4	0.017491315	chr6	-1	37344619	NM_003958	AGGGAGCAGGGTCAATTATTTAG	AGGGAGCAGGGTCAATTATTTAG	AGGGAGCAGGGTCAATTATTTAG	AGGGAGCAGGGTCAATTATTTAG

Figure 4.3 gRNA off-target prediction and validation
(A) Quality of the two gRNAs used in the generation of *DLG2*$^{-/-}$ hESC lines. Separately the gRNAs had predicted off-targets of 89 and 125 loci; however, when used as a pair in conjunction with a Cas9 nickase there were no predicted off-targets. Captured from crisp.mit.edu. Based on these predictions 7 genic off-target loci were selected at random for both gRNA 1 **(B)** and gRNA 2 **(C)** for validation by PCR amplification and sanger sequencing, in all cases this showed the sequence in knockout lines to be identical to WT at these locations. Adapted from (Sanders et al., 2020).

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

Figure 4.4 CNV analysis of WT and *DLG2*$^{-/-}$ lines
WT, KO1 and KO2 genomic DNA was analysed using the Illumina PsychArray v1.1, with data analysed using PennCNV. CNVs smaller than 100Kb and containing less than 10 SNPs were filtered out during PennCNV QC. No additional CNVs are identified in wither KO lines in comparison to WT controls. From (Sanders et al., 2020).

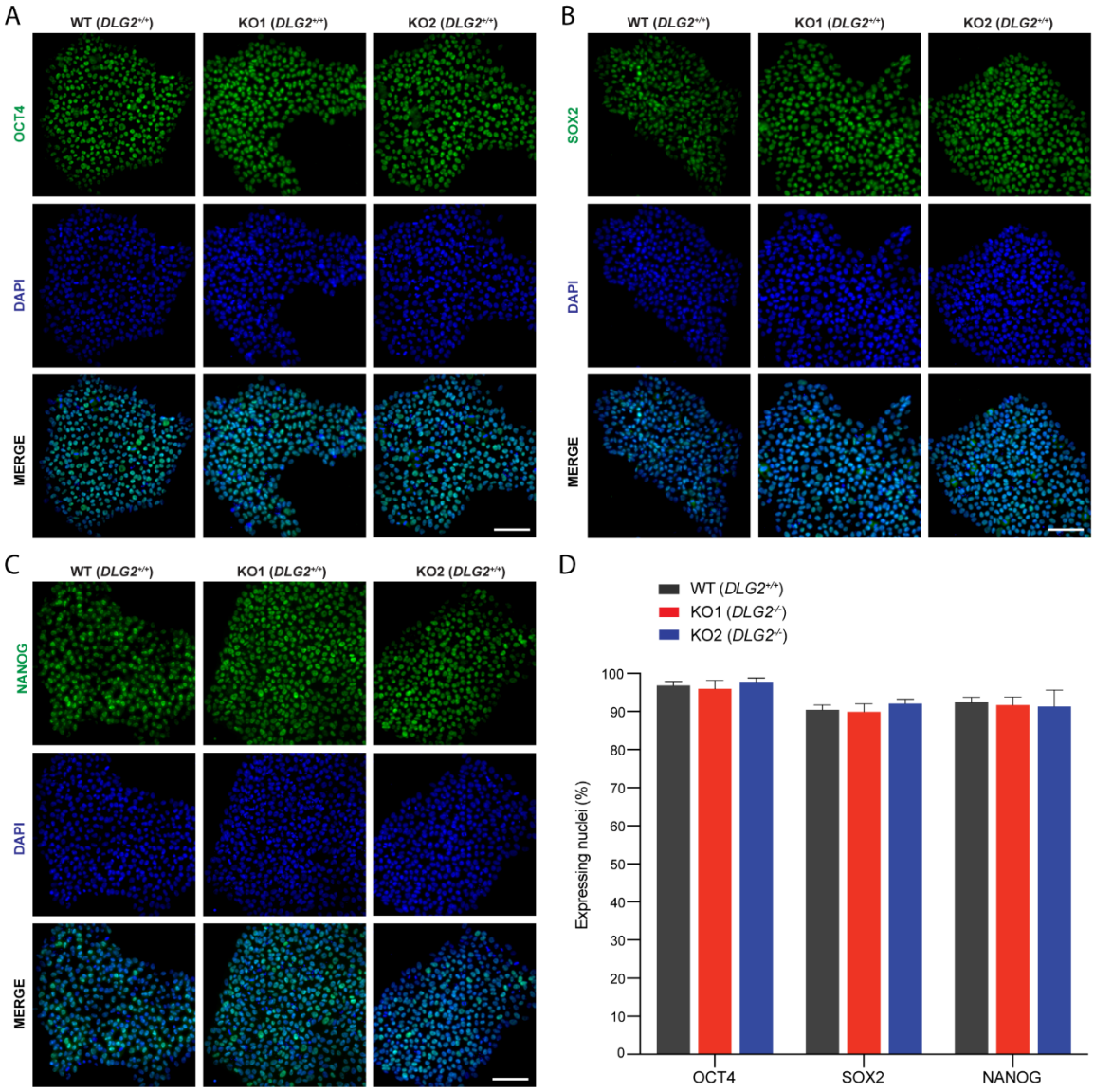


Figure 4.5 Pluripotency marker expression in WT and *DLG2*$^{-/-}$ hESCs
Representative ICC images of OCT4 **(A)**, SOX2 **(B)** and NANOG **(C)** expression with DAPI nuclear counterstain for 2 *DLG2*$^{-/-}$ hESC lines and WT controls. **(D)** ICC quantification of nuclei expressing either OCT4, SOX2 or NANOG in 2 *DLG2*$^{-/-}$ hESC lines and WT controls, the genotype had no significant effect on the expression of these markers (OCT4: $F(2,15)=0.3780$; $P=0.9616$; $n=6$, SOX2: $F(2,15)=0.5383$; $P=0.5946$; $n=6$, NANOG: $F(2,15)=0.03433$; $P=0.9663$; $n=6$). Analysis was by one-way ANOVA for each pluripotency marker. All scale bars are 100 µm. From (Sanders et al., 2020).

4.6. *DLG2*$^{-/-}$ NPCs show an increasing trend in PAX6 and KI67 expression

Following validation of the generated hESC lines, *DLG2*$^{-/-}$ and WT sister cells were differentiated towards cortical projection neurons and a combination of ICC and western blotting used to characterise their phenotypes at multiple time points. All analyses were performed on KO1, KO2 and WT cells; however, as the results for both *DLG2*$^{-/-}$ lines were broadly similar (Figure 4.6 A, Supplementary Figure 1) and the focus of this research was changes resulting from the loss of the *DLG2* gene, rather than any *DLG2*$^{-/-}$ cell line specific differences, these results were combined with

analyses presented as WT against *DLG2*$^{-/-}$ (KO1 + KO2). ICC results at differentiation day 20 (Figure 4.6) show cells from all lines are predominately NPCs expressing characteristic markers such as PAX6, FOXG1 and SOX2. In contrast there is very low expression of neuronal markers including the postotic neuron marker NEUN and the cortical layer markers TBR1 (cortical layer IV) and CTIP2 (cortical layer V). Also supporting an NPC, rather than neuronal, identity at this time point is the expression of the general proliferation marker KI67 in the majority of day 20 cells, combined with expression of the mitosis marker PH3 in a reduced number of cells (those in the M phase of the cell cycle), neither of which would be found in postmitotic neurons. In addition to indicating an NPC identity for both WT and *DLG2*$^{-/-}$ cells at day 20 of differentiation, the expression of PAX6 indicates a dorsal telencephalic fate for cells and confirms differentiation towards the intended cortical projection neuronal identity.

For all the markers investigated by ICC at day 20 of differentiation (PAX6, FOXG1, SOX2, PH3, KI67, NEUN, TBR1, CTIP2) there were no obvious trends or statistically significant differences between the cell lines (Figure 4.6). This indicates that were a similar percentage of cells expressing these given markers in the WT and *DLG2*$^{-/-}$ lines at this time point; however, as this may not reflect the amount of protein present in these cells western blotting was also performed to assess this. This western blotting was run for many of the markers investigated by ICC including SOX2, PAX6, NEUN & KI67 (Figure 4.7). The expression of TBR1 and CTIP2 was also investigated by western blotting, but unlike later time points (see section 4.7), could not be detected at day 20 of differentiation; presumably due to their low abundance, as indicated by the results of ICC analysis (Figure 4.6,) combined with the size of polyacrylamide gel wells limiting the total amount of protein that can be run for subsequent blotting.

Unlike with ICC results, which showed no difference between any markers tested, western blotting results showed increased expression (although this failed to reach significance) of PAX6 and KI67 in the *DLG2*$^{-/-}$ cell lines at day 20 of cortical differentiation (Figure 4.7). A lack of statistical significance potentially due to a degree of variability within the samples means this is only a trending increase in expression; however, this may have changed if time and resources had permitted the analysis of additional replicates. Although not as prominent there was also a decreasing trend in SOX2 expression in the *DLG2*$^{-/-}$ cell lines compared to WT, while NEUN expression was similar. PAX6 and SOX2 are key neurodevelopmental transcription factors (Englund et al., 2005; Kashyap et al., 2009; Sansom et al., 2009; Hutton and Pevny, 2011) so any perturbation to their expression in *DLG2*$^{-/-}$ cells at day 20 of cortical differentiation would both suggest that DLG2 is expressed a this relatively early time point and that is involved in developmental signalling pathways at this stage. Additionally KI67 is an established marker of cell proliferation, a process known to be regulated by invertebrate Dlg (Elsum et al., 2012; Stephens et

al., 2018), providing further evidence both for DLG2 expression and function at this stage of cortical development. As ICC results showed no change for these markers this would suggest the increased expression (possibly decreased for SOX2) relates to the amount of protein within individual cells rather than the percentage of cells expressing these proteins within the whole population. It should also be noted that although analysis of these markers is relevant and informative, due to their limited number cannot represent and extensive characterisation of *DLG2*$^{-/-}$ NPCs which would require either testing additional markers in a similar manner or an alternative approach such as RNAseq.

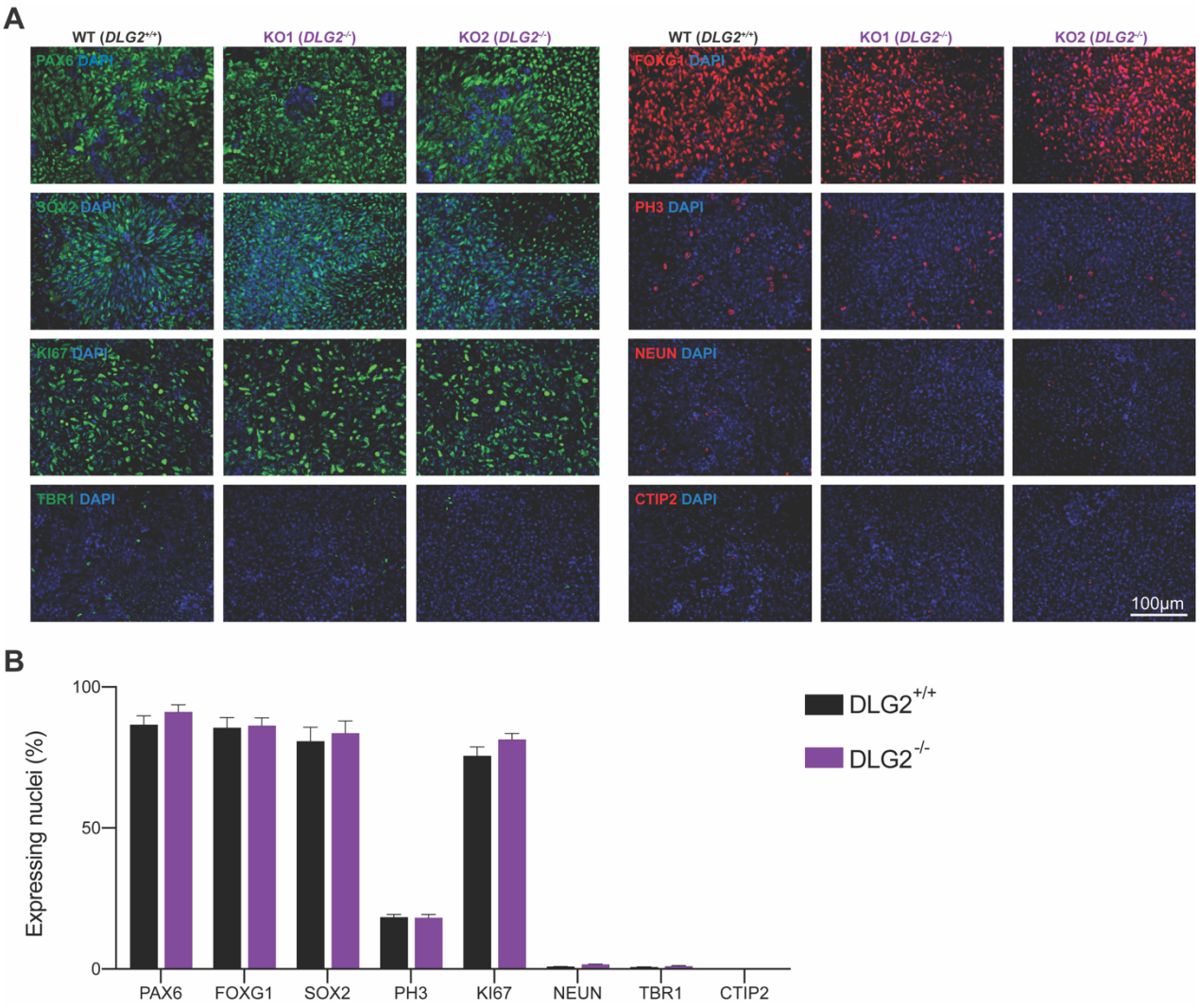


Figure 4.6 ICC characterisation of the *DLG2*$^{-/-}$ phenotype at day 20 of cortical differentiation
(A) Representative ICC images of PAX6, SOX2, KI67, TBR1, FOXG1, PH3, NEUN and CTIP2 expression with DAPI nuclear counterstain for 2 *DLG2*$^{-/-}$ hESC lines and WT controls at day 20 of cortical differentiation. **(B)** ICC quantification of nuclei expressing either PAX6, SOX2, KI67, TBR1, FOXG1, PH3, NEUN in *DLG2*$^{-/-}$ (KO1 + KO2) and WT cells at day 20 of cortical differentiation. Marker identity did have a significant effect on expression levels as determined by two-way ANOVA ($F_{7,131}$= 451.9; P<0.0001; n≥6) but genotype did not ($F_{1,131}$=1.873; P=0.1734; n≥6).

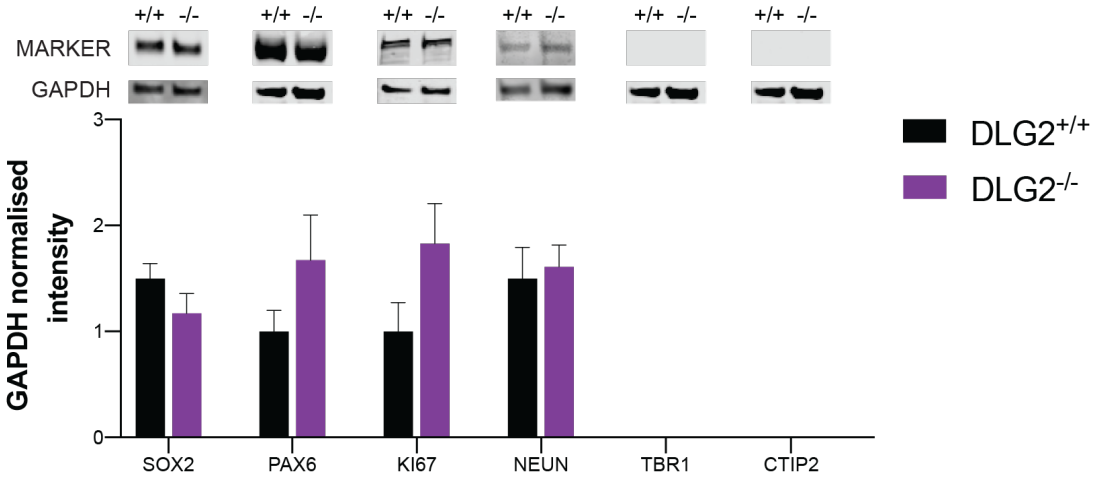


Figure 4.7 Western blot characterisation of the *DLG2*$^{-/-}$ phenotype at day 20 of cortical differentiation
SOX2, PAX6, KI67, NEUN, TBR1 and CTIP2 western blot protein bands and histogram of intensity normalised to GAPDH for *DLG2*$^{-/-}$ (KO1 + KO2) and WT cells at day 20 of cortical differentiation. Unlike subsequent experimental time points, no TBR1 or CTIP2 expression was detectable by western blot at day 20. Neither marker identity ($F_{3,76}$= 0.1525; P=0.9278; n≥6) nor genotype ($F_{1,76}$=1.631; P=0.2054; n≥6) had a significant effect on expression levels as determined by two-way ANOVA.

4.7. *DLG2* deficiency has no effect on production of post-mitotic neurons but significantly reduces expression of layer V identity

Following characterisation of *DLG2*$^{-/-}$ cells at day 20 the next step was to investigate later time points of cortical differentiation, where we would expect NPCs to exit the cell cycle to become postmitotic neurons. For this the expression of cell-type specific neuronal markers was analysed in WT and *DLG2*$^{-/-}$ lines from days 30 - 60 of cortical differentiation via ICC (Figure 4.8) and Western blotting (Figure 4.9). Using both these techniques it was clear that WT and *DLG2*$^{-/-}$ cells are able to differentiate from NPCs and produce postmitotic neurons expressing characteristic neuronal markers such as NEUN (Figures 4.8 B & 4.9 A) and MAP2 (Figure 4.9 D), plus cortical deep layer markers TBR1 and CTIP2 (Figures 4.8 C + D & 4.9 B + C). Additionally, preliminary ICC analysis (from1 round of differentiation only) also showed all lines were capable of producing neurons expressing SATB2 (Figure 4.8 F), a result validated by western blotting using cells from two independent differentiations (Figure 4.9 E). Although known to be highly expressed in upper layer neurons (Britanova et al., 2008) more recent studies have shown that *in vivo* SATB2 is also expressed in populations of deep layer neurons during cortical development, in particular being expressed for a long period within the prefrontal cortex (Nowakowski et al., 2017).

Both western blotting for NEUN and MAP2 and quantification of NEUN^{+} cells following ICC revealed no significant difference in the proportion of postmitotic neurons produced by *DLG2*$^{-/-}$ cultures compared to WT; in agreement with the similar proportion of NPCs found in all lines at days 30 to 60, as indicated by similar levels of KI67, PAX6 and SOX2 (Figures 4.8 E & 4.9 F - H).

However, despite all cultures producing similar percentage of neurons an analysis of the layer markers TBR1, CTIP2 and SATB2 revealed a significant decrease in CTIP2^{+} cells in *DLG2*$^{-/-}$ lines compared to WT, although the proportion TBR1^{+} and SATB2^{+} neurons were comparable for all timepoints investigated. The decrease of CTIP2 expressing cells found in *DLG2*$^{-/-}$ cultures was most noticeable at day 30 of cortical differentiation where levels were only 15% of the WT level, although this recovered somewhat to roughly 50% at day 60 (Figure 4.8 H). Although there was some variation between *DLG2*$^{-/-}$ lines (Supplementary Figure 1), at all time points investigated the level was significantly reduced when compared to WT. Therefore, *DLG2* deficiency does not appear to affect cell cycle exit and the rate at which neurons are produced from NPCs but does significantly impair the expression of cortical layer V identity in new-born deep layer neurons.

4.8. *DLG2*$^{-/-}$ neurons display immature morphology

After establishing a clear *DLG2*$^{-/-}$ phenotype during corticoneurogenesis, that being an impaired expression in layer V identity as determined by the proportion of cells expressing CTIP2 compared to WT, this was further investigated by performing a morphometric analysis of both *DLG2*$^{-/-}$ and WT neurons. This involved tracing both immature (day 30) and more mature (day 70) WT and *DLG2*$^{-/-}$ neurons following transfection with a GFP expressing plasmid, which enabled all major neuronal structures to be clearly identified (Figure 4.10 E). At both timepoints there was no significant difference in the soma area for *DLG2*$^{-/-}$ neurons compared to WT or either the total neurite length or the number of primary neurites projecting from the soma (Figure 4.10 A, C + D). However, *DLG2*$^{-/-}$ neurons showed greatly reduced branching with a significant reduction in the number of secondary neurites (Figure 4.10 B), those being neurites that project from primary neurites rather than directly from the soma. As the total neurite length did not differ for *DLG2*$^{-/-}$ neurons, this leads to a clear phenotype of longer and relatively unbranched primary neurites when compared to WT cells (Figure 4.10 E). Therefore, although *DLG2*$^{-/-}$ cells are capable of forming neurons at a similar rate to WT these have both a delayed expression of layer identity and neurite morphology, with both features persisting from the new-born immature stage to a later more mature stage of neuronal development.

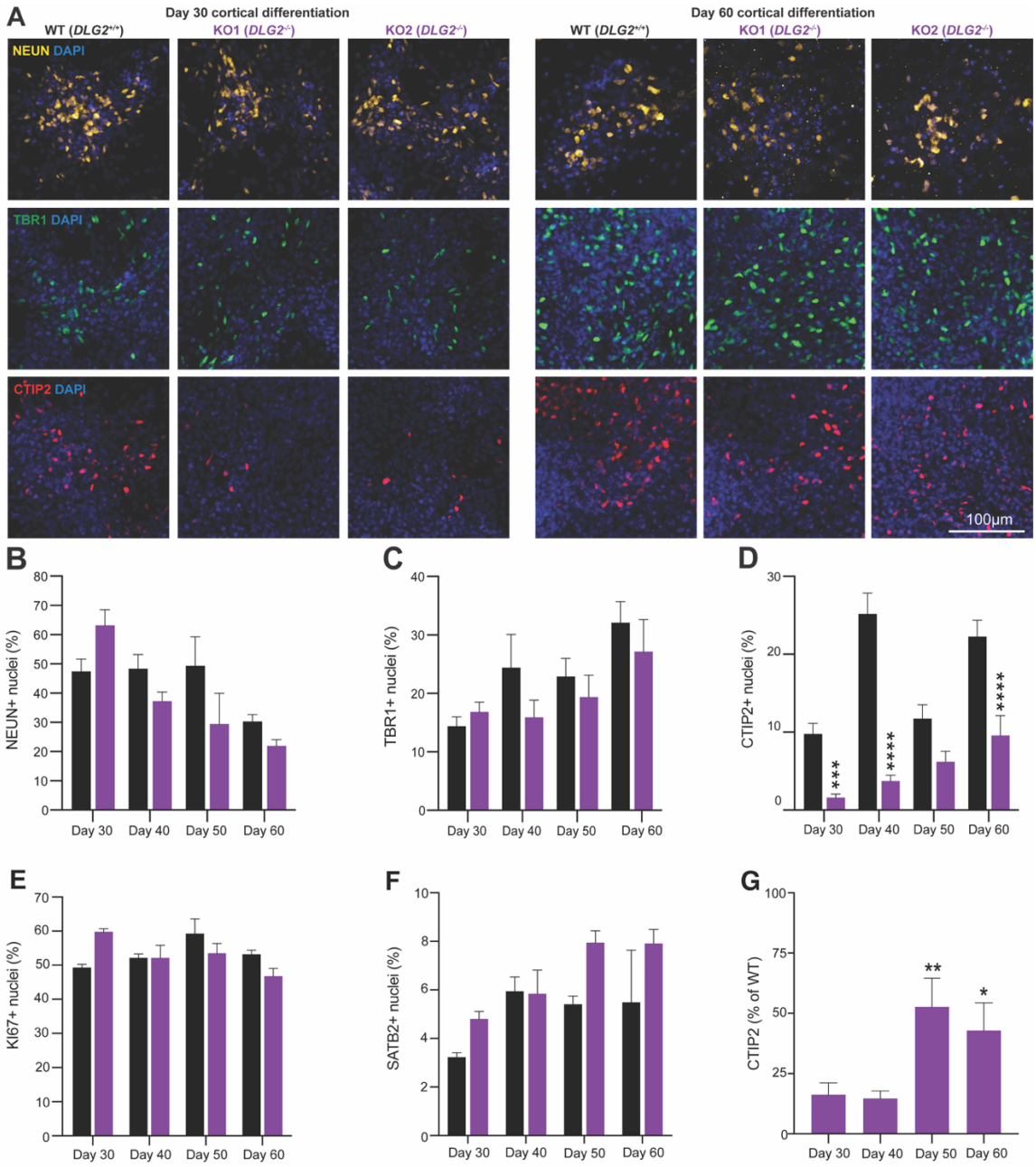


Figure 4.8 ICC characterisation of the *DLG2*$^{-/-}$ phenotype at days 30 – 60 of cortical differentiation
(A) Representative ICC images of NEUN, TBR1 and CTIP2 expression with DAPI nuclear counterstain for 2 *DLG2*$^{-/-}$ lines and WT controls at days 30 and 60 of cortical differentiation. **(B - F)** ICC quantification of expressing nuclei for *DLG2*$^{-/-}$ and WT cells at 4 time points of cortical differentiation. **(B)** Time ($F_{3,52}$=7.018, P=0.0005; n≥6) did have a significant effect on NEUN expression, while genotype ($F_{1,52}$=1.687; P=0.1998; n≥6) did not, there was also a significant interaction between these factors ($F_{3,52}$=2.842; P=0.0466; n≥6). **(C)** 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. **(D)** Both time ($F_{3,67}$=18.93, P<0.0001; n≥6) and genotype ($F_{1,67}$=101.8; P<0.0001; n≥6) had significant effects on CTIP2, there was also a significant interaction between these factors ($F_{3,67}$=7.862; P=0.0001; n≥6). **(E)** Neither time ($F_{3,27}$=1.754; P=0.1798; n≥3) nor genotype ($F_{1,27}$=0.03483; P=0.8533; n≥3) had a significant effect on KI67 expression (from 1 differentiation only). **(F)** Both genotype ($F_{4,28}$=7.373; P=0.0112; n≥3) and time ($F_{3,28}$=4.520; P=0.0105; n≥3) had significant effects on SATB2 expression (from 1 differentiation only). **(G)** ICC quantification of the proportion of CTIP2 expressing nuclei in both *DLG2*$^{-/-}$ lines relative to WT levels at 4 time points of cortical differentiation. The timepoint had a significant effect on the proportion of *DLG2*$^{-/-}$ nuclei expressing CTIP2 ($F_{3,42}$=5.391; P=0.0031; n≥6). Data sets (B-F) were analysed by two-way ANOVA and (G) by one-way ANOVA. Stars above bars in each graph represents Bonferroni-corrected post hoc tests, *P<0.05; **P<0.01; ***P<0.001; ****P<0.0001 vs. WT controls (B-F) or day 30 (H). Adapted from (Sanders et al., 2020).

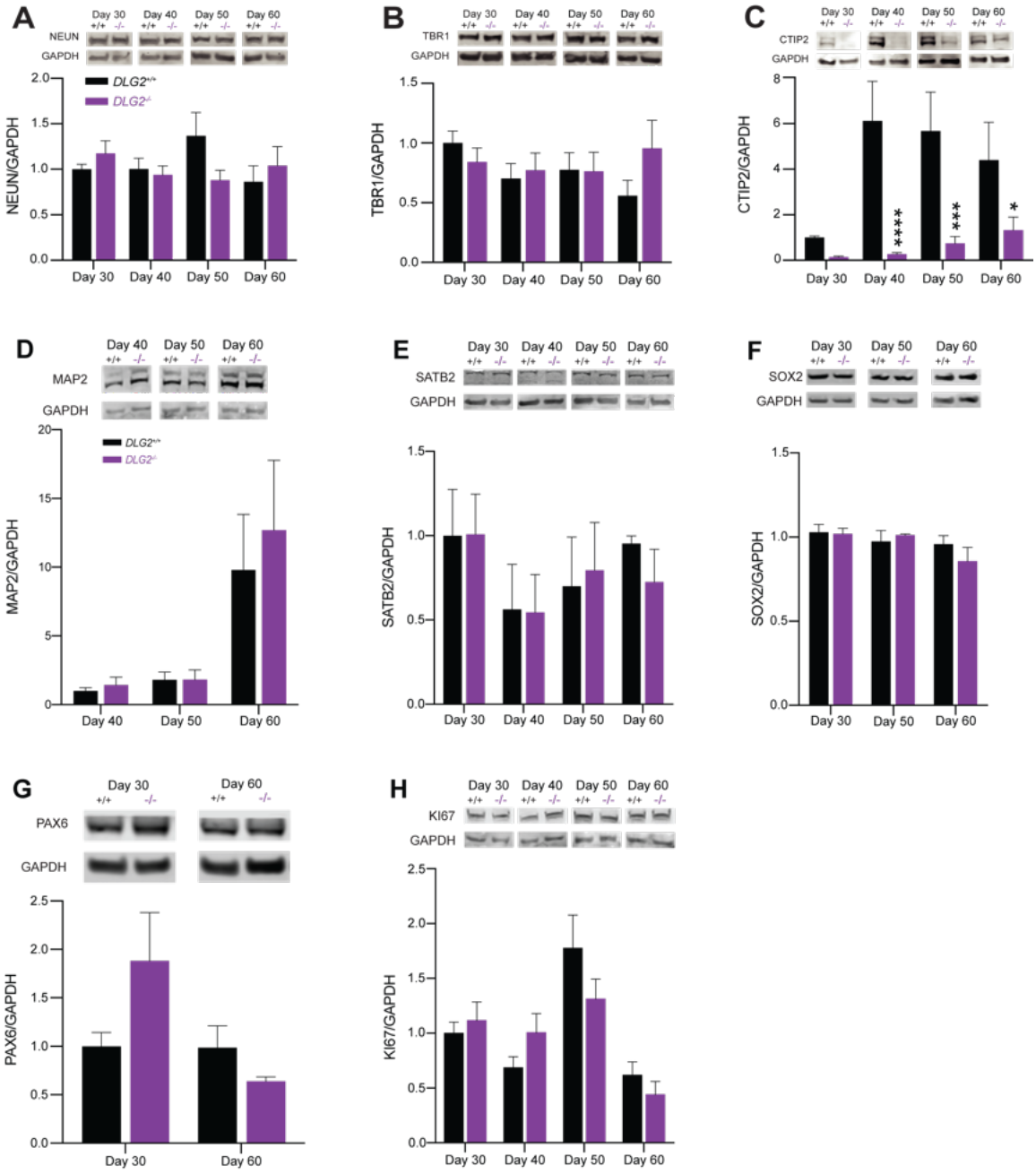


Figure 4.9 Western blot characterisation of the *DLG2*$^{-/-}$ phenotype at days 30 – 60 of cortical differentiation
(A - H) Western blot protein bands and histograms of expression normalised to GAPDH for *DLG2*$^{-/-}$ and WT cells at 4 time points of cortical differentiation. **(A)** Neither time ($F_{3,90}$=0.5382; P=0.6573; n≥7) nor genotype ($F_{1,90}$=0.1852; P=0.6680; n≥7) had a significant effect on NEUN expression. **(B)** Neither time ($F_{3,95}$=0.5052; P=0.6796; n≥9) nor genotype ($F_{1,95}$=0.3899; P=0.5338; n≥9) had a significant effect on TBR1 expression. **(C)** Both time ($F_{3,86}$=5.2262; P=0.0022; n≥7) and genotype ($F_{1,86}$=39.89; P<0.0001; n≥7) had a significant effect on CTIP2 expression, there was also a significant interaction between these factors ($F_{3,86}$=3.677; P=0.0152; n≥7). **(D)** Time ($F_{2,30}$=8.721; P=0.0010; n≥5) did have a significant effect on MAP2 expression, while genotype ($F_{1,30}$=0.2673; P=0.6157; n≥5) did not. **(E)** Neither time ($F_{3,31}$=1.028; P=0.3937; n≥2) nor genotype ($F_{1,31}$=0.0.02854; P=0.8669; n≥2) had a significant effect on SATB2 expression. **(F)**. Neither time ($F_{2,25}$=0.9226; P=0.4106; n≥3) nor genotype ($F_{1,25}$=0.09364; P=0.7621; n≥3) had a significant effect on SOX2 expression (from 1 differentiation only). **(G)** Both time ($F_{3,25}$=4.856; P=0.0085; n≥3) and genotype ($F_{1,25}$=6.248; P=0.0194; n≥3) had a significant effect on PAX6 expression (from 1 differentiation only). **(H)** Time ($F_{3,111}$=15.89; P<0.0001; n≥9) did have a significant effect on KI67 expression, while genotype ($F_{1,111}$=1.766; P=0.1866; n≥9) did not. All data sets were analysed by two-way ANOVA. Stars above bars in each graph represents Bonferroni-corrected post hoc tests, *P<0.05; **P<0.01; ***P<0.001; ****P<0.0001 vs. WT controls. Adapted from (Sanders et al., 2020).

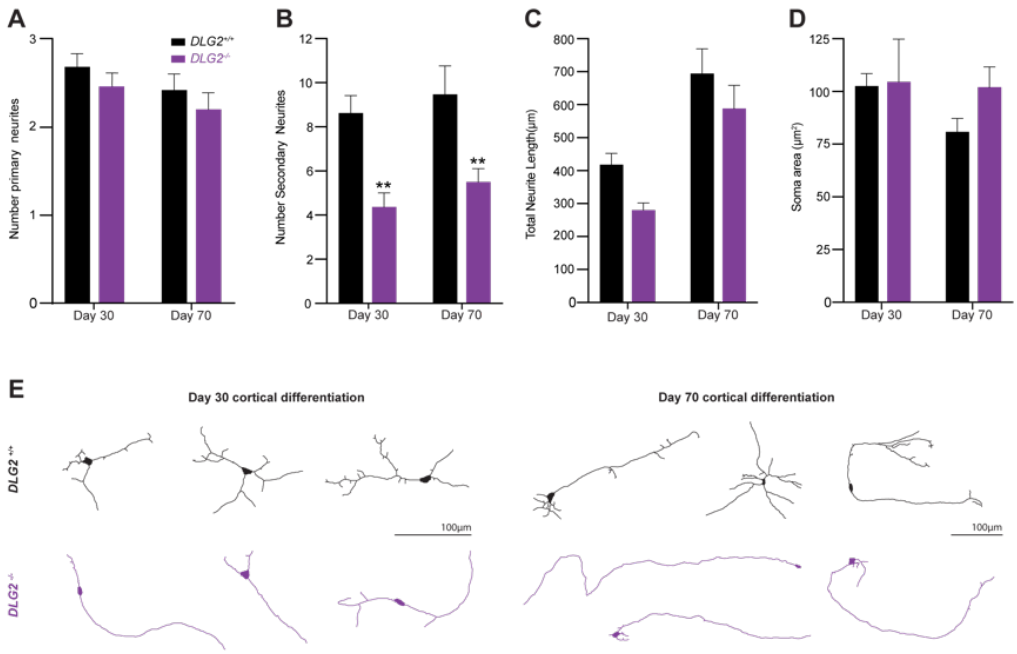


Figure 4.10 The morphology of WT and *DLG2*$^{-/-}$ neurons
(A - D) Morphological quantification of *DLG2*$^{-/-}$ and WT neurons at days 30 and 70 of cortical differentiation. **(A)** Neither genotype ($F_{1,126}$=1.591; P=0.2095; n≥28) nor time ($F_{1,126}$=2.278; P=0.1337; n≥28) had significant effects the numbers of primary neurites (those projection from the soma). **(B)** Genotype ($F_{1,126}$=18.78, P<0.0001; n≥28) did have a significant effect on the number of secondary neurites (those projecting from primary neurites), while time ($F_{1,126}$=1.082, P=0.3003; n≥28) did not. **(C)** 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 the total neurite length. However, post hoc analysis showed no significant differences at individual timepoints. **(D)** Neither genotype ($F_{1,136}$=0.9170; P=0.340; n≥28) nor time ($F_{1,136}$=1.399; P=0.2390; n≥28) had a significant effect on soma area. **(E)** Representative neuronal traces. Data sets were analysed by two-way ANOVA. Stars above bars in each graph represents Bonferroni-corrected post hoc tests, *P<0.05; **P<0.01; ***P<0.001; ****P<0.0001 vs. WT controls. From (Sanders et al., 2020).

4.9. Discussion

4.9.1. DLG2 *is expressed from day 30 in WT cells and absent from knock-outs*

Two *DLG2*$^{-/-}$ lines (KO1 & KO2) had been generated previously by using CRISPR/Cas9-D10a nickase genome editing at exon 22 gene, which is translated into part of the first PDZ domain of the DLG2 protein. The knockout was confirmed at the genetic levels by sanger sequencing which shows indels resulting in premature stop codons in both knockout lines within the targeted region of exon 22. To ensure that both KO1 and KO2 were full knockouts and unable to produce DLG2 protein mass spectrometry was performed, following a peptide affinity pulldown of PDZ domain containing proteins, on both lines as well as WT cells at days 30 and 60 of cortical differentiation. Mass spectrometry analysis not only confirmed the absence of DLG2 protein in the knock-outs but showed DLG2 expression in WT cells at day 30 of cortical differentiation, although levels are reduced compared to day 60. Previous efforts to investigate the pattern of DLG2 expression *in vitro* by western blotting were hampered by issues with antibody specificity (see Chapter 3) so that clear expression could only be seen from day 50 (Figure 3.2), these mass spectrometry data therefore bring the time point at which DLG2 expression has been detected forward by 20 days. As day 30 represents a time of early neurogenesis in culture, being the first time point that neuronal markers are detected by western blotting, it is unlikely that significant synaptogenesis has occurred and DLG2 expression may indicate a role outside of the PSD. Also of interest is that the faint bands visible in DLG2 western blots from day 30 samples are of a similar intensity to those at days 15 and 20 (Figure 3.2), although none of these are definitive from western blotting alone due to the issues previously described, the fact that mass spectrometry confirms DLG2 expression at day 30 suggests the possibility that it is expressed at similar levels at these even earlier time points.

These results show that mass spectrometry analysis of WT samples from earlier time points of cortical differentiation, including days 15 and 20, would be beneficial to determine a more accurate picture of the temporal pattern of DLG2 expression *in vitro* and this will certainly be included in any future studies. It should be noted that while the described enrichment for PDZ domain containing proteins, performed prior to mass spectrometry, would capture the majority of the DLG2 protein coding isoforms it would not capture those truncated isoforms without a PDZ domain such as isoform 4, as such these may still be produced in knock-outs. However, as the function of these truncated isoforms is not known and the focus of this research was canonical DLG2 with 3 PDZ, 1 SH3 and 1GK domains enrichment using a PDZ domain interactor such as the peptide “SIESDV” remains valid. If enrichment of all DLG2 isoforms, including those without PDZ domains, was required then other approaches could be used such as isolating protein from synaptosomes rather than whole cells.

4.9.2. *DLG2*$^{-/-}$ hESCs were successfully validated for off-target effects and pluripotency marker expression

Further validation of these generated *DLG2*$^{-/-}$ lines was undertaken prior to their use for this research. An initial validation step was to confirm that the CRISPR/Cas9 mediated genome editing had not produced any off-target effects, necessary as off-target predictions made by computer algorithms have been shown to be incorrect in some cases (Zhang et al., 2015). As predictions showed no off-target effects for the two gRNAs used as a pair for CRISPR/Cas9-D10a nickase editing but did show multiple potential off-targets for each gRNA if used with a CRISPR/Cas9 nuclease, which could cause unwanted double stranded DNA breaks. Even though the nickase was employed hence no strong possibility of indel generation, investigating these nuclease off-target sites was the logical first step. PCR primers were designed to amplify 14 of these predicted off-target sites, seven for each gRNA selected at random, and following sanger sequencing all potential off-target sites examined were confirmed as WT. Although not conclusively proving that no off-target effects occurred as a result of the CIRSPR/Cas9 genome editing, which would require whole genome sequencing based approaches (Tsai et al., 2015), these sequencing results combined with the results of the prediction algorithms are strongly indicative of no off-target effects.

Sequencing and CNV analysis was performed to detect any deleted or duplicated regions. The results of the CNV analysis run for both *DLG2*$^{-/-}$ lines and a WT sister line showed all lines had two deletions and two duplications compared to the reference genome. This may be the result of passaging to maintain the original H7 hESC line which has been implicated in the introduction of CNVs (Liu et al., 2014), although at least one of the detected CNVs, the 20q11.21 duplication, appears to be feature of the cell line and potentially stems from its genesis (Laurent et al., 2011). However, as both generated *DLG2*$^{-/-}$ were identical to WT with no additional CNVs this further suggests successful genome editing with no off-target effects. Following this genetic validation of the two *DLG2*$^{-/-}$ lines an analysis of pluripotency marker expression confirmed their identity as hESCs and making them suitable for subsequent differentiation.

4.9.3. Directed differentiation of *DLG2*$^{-/-}$ hESCs can be used to model human cortical neurogenesis

Both KO1 and KO2 along with WT controls underwent several directed differentiations from hESCs to cortical projection neurons using a developmental patterning approach adapted from that developed by Chambers *et al*. (Chambers et al., 2009). The initial purpose of this was to confirm that the generated *DLG2*$^{-/-}$ hESC lines were capable of successfully undergoing neuronal differentiation and hence be used as an *in vitro* model for human cortical neurogenesis. As the developmental patterning method for cortical differentiation broadly recapitulates *in vitro* the

developmental process *in vivo*, we expected cells of both *DLG2*$^{-/-}$ lines along with a WT control to undergo a progressive specialisation from pluripotent hESCs to NPCs, immature postmitotic neurons and finally to more mature postmitotic neurons. There was no reason to believe that *DLG2*$^{-/-}$ hESC would not be able to successfully undergo neuronal differentiation, the process is relatively well established for hESCs and both *DLG2*$^{-/-}$ lines were validated and shown to express the expected pluripotency markers as previously described; furthermore, *DLG2*$^{-/-}$ mouse models have been previously generated and are not embryonic lethal, although adults do display various cognitive deficits (Nithianantharajah et al., 2013).

Both *DLG2*$^{-/-}$ lines were indeed able to differentiate successfully; initially from hESCs expressing the pluripotency markers OCT4, SOX2 and NANOG at day 0 of the cortical differentiation to NPCs expressing SOX2, FOXG1 and PAX6 by day 20. PAX6 in conjunction with FOXG1 and SOX2 expression being of particular significance as it confirms that the NPCs have a dorsal telencephalic identity and hence are precursors to cortical projection neurons, confirming differentiation to the correct lineage. In addition to these more specific markers the cells at day 20 also express markers common to proliferative cells, which include NPCs, such as KI67 and PH3. By day 30 new-born immature neurons are successfully produced in culture expressing the postmitotic neuronal makers NEUN and MAP2 as well as the cortical deep layer markers TBR1 (initially all new-born neurons and cortical layer VI) and CTIP2 (cortical layer V). CTIP2 levels in particular increase at later timepoints of the differentiation compared to day 30, in agreement with the normal process of cortical development where cortical excitatory neurons are born in an inside-out fashion (from deep to upper layers) beginning with TBR1 expressing all new-born and layer VI neurons before progressing to CTIP2 expressing layer V neurons. Cultures also produce neurons expressing the marker SATB2 at all time points investigated, although known to be highly expressed in upper layer neurons (Britanova et al., 2008) it is unlikely that a 2D directed differentiation system of cortical development, such as the one used in this research, could produce significant numbers of these neurons from hESCs within 60 days. However, research has shown that in human foetal brains SATB2 is also expressed in populations of deep layer neurons during cortical development, in particular being expressed for a long period within the prefrontal cortex (Nowakowski et al., 2017). It is therefore likely that SATB2 expression observed in culture from days 30-60 comes from deep layer neurons meaning that the effect of DLG2 deficiency on the generation of upper layer neurons cannot be confidently concluded. *DLG2*$^{-/-}$ hESC lines can therefore be successfully differentiated to dorsal telencephalic NPCs and subsequently to deep layer cortical projection neurons that express the characteristic markers of these cell types, confirming these lines can be used to successfully model human cortical neurogenesis.

4.9.4. *DLG2* deficiency has a significant effect on the expression of neuronal identity, if not on the rate of neurogenesis

DLG2 deficiency also does not appear to affect the rate of neurogenesis, that is the production of postmitotic neurons from NPCs. From day 30 of cortical differentiation to day 60 there is no significant difference in the expression of the postmitotic neuronal markers NEUN and MAP2 in *DLG2*$^{-/-}$ cultures compared to WT. This is also supported by similar levels of TBR1 expression in both WT and *DLG2*$^{-/-}$ cultures, a marker expressed at basal levels in all deep layer neurons, but most highly expressed in new-born neurons (Hevner et al., 2001; Englund et al., 2005). A similar proportion of NPCs found in all lines at days 30 to 60 providing further evidence as indicated by the levels of KI67, PAX6 and SOX2 which show no significant difference. NPCs can be found in all time points of differentiation from day 20 onwards as a population of these cells continue to proliferate, maintaining the progenitor pool, while others cease proliferation and differentiate to produce immature postmitotic neurons. Mitotic inhibitors can be added to cultures to force NPC cell cycle exit (Crawford and Roelink, 2007; Kemp et al., 2016) but this was not done in this research to ensure that no developmental phenotypes were artificially suppressed.

Although *DLG2*$^{-/-}$ cultures may produce neurons at a similar rate to WT they express significantly less of the layer V marker CTIP2 compared to WT from days 30 to 60 of cortical differentiation, while neurons display a more immature morphology characterised by significantly reduced neurite branching. This could potentially be explained by an altered identity for *DLG2*$^{-/-}$ neurons compared to WT, meaning the proportion of neurons that do not express CTIP2 are not layer V neurons and never will be, with an alternate but currently undefined neuronal fate. If this were the case then the change would be permanent, with no recovery or catch up of the observed phenotype. A second explanation is that cortical layer identity is delayed in *DLG2*$^{-/-}$ neurons meaning that neurons initially not expressing CTIP2 will express it at later time points. This delay in expressing layer V identity seems better supported by the experimental data, although the percentage of CTIP2^{+} cells is significantly reduced in *DLG2*$^{-/-}$ cultures at all time points, there is also a significant amount of recovery over the course of the differentiation, from roughly 15% of WT levels at day 30 to roughly 50% by day 50. The immature morphology of *DLG2*$^{-/-}$ neurons could also be seen as a delay in expressing their latent neuronal identity, although this feature did not show recovery between the two time points investigated such large-scale morphological changes may well take longer to catch up to WT than levels of protein expression. It therefore possible that two broad processes may be occurring as NPCs differentiate into cortical projection neurons; neurogenesis itself which appears to be a *DLG2* independent process broadly not affected by *DLG2* deficiency and the expression neuronal subtype identity which is a *DLG2* dependent process which is significantly delayed in *DLG2*$^{-/-}$ cells.

4.9.5. *An altered state within DLG2$^{-/-}$ NPCs may explain subsequent neuronal phenotypes*

Data from ICC, western blotting and morphometric assays indicate a role for *DLG2* in the acquisition of neuronal subtype identity, with *DLG2*$^{-/-}$ neurons showing a reduced and likely delayed expression of cortical layer V identity. Although this data does not provide a clear mechanism underlying this neuronal phenotype, it likely has its origins at an earlier stage of differentiation in *DLG2*$^{-/-}$ NPCs. At day 20 of cortical differentiation, a time point where NPCs are the predominant cell type in culture *DLG2*$^{-/-}$ cells an increasing trend in the expression of the transcription factor PAX6 as well as the marker of cellular proliferation KI67, although these were not statistically significant with the current number of replicates.

The exact mechanism by which specific neuronal subtypes are generated from NPCs is not completely understood, with some evidence suggesting that specific subtypes of NPCs are generated which in turn give rise to specific neuronal populations (Mi et al., 2018). More widely accepted is that NPCs form a common pool with their internal state immediately prior to cell-cycle exit determining the eventual identity of generated neurons, this state changing over the course of multiple cell divisions leading to the progressive generation of layers VI-II neurons (Telley et al., 2019). Although neuronal subtype identity may of course be a combination of specific NPCs determining which layer neurons may be generated, or alternatively not be generated, in conjunction with the internal state of the NPCs.

This also suggests two possibilities for the role of *DLG2* in the acquiring neuronal subtype identity; it may either act to alter the ratio of specific NPC pools in culture or perhaps more likely change the internal state of NPCs prior to neurogenesis in a process linked to cell-cycle regulation. Given the indicated disruption to a key regulator of neurodevelopment such as PAX6 (Englund et al., 2005; Sansom et al., 2009) in *DLG2*$^{-/-}$ NPCs along with increased cellular proliferation indicated by levels of KI67, disrupted developmental signalling pathways within NPCs including those regulating the cell-cycle may indeed contribute to the subsequently observed neuronal phenotype. Although western blotting results are not statistically significant such as disruption to the cell-cycle in particular would be supported by the established association between *DLG2* mutations and cancer as well as the role of invertebrate *Dlg* the regulation of proliferation (Elsum et al., 2012; Stephens et al., 2018). If this is correct it would of course provide evidence for both DLG2 expression earlier in neurodevelopment than currently appreciated and an additional role for the protein outside of the PSD, although further experiments are required to clarify this.

4.9.6. Summary

Two *DLG2*$^{-/-}$ hESC lines (KO1 & KO2) previously generated using CRISPR/Cas9-D10a nickase genome editing were validated and shown to have no detectable off-target effects, express pluripotency markers at 100% of WT levels and be full DLG2 knockouts at the protein level by mass spectrometry, a technique that also showed WT DLG2 expression relatively early in cortical differentiation from day 30. These knockout hESC lines successfully underwent cortical differentiation producing NPCs, deep layer and upper layer cortical neurons. *DLG2* deficiency had no effect on the rate of neurogenesis with *DLG2*$^{-/-}$ cells showing similar numbers of NPCs and postmitotic neurons compared to WT at all time points investigated. *DLG2*$^{-/-}$ neurons did however show a delay in the expression of cortical layer V identity compared to WT, as evidence by a significant reduction in the percentage of CTIP2^{+} cells that recovered from ~15% of WT levels at day 30 to ~50% by day 50. In agreement with a delayed expression of mature characteristics (such as layer marker expression) *DLG2*$^{-/-}$ neurons also display an immature morphology compared to WT, with significantly reduced neurite branching. This neuronal phenotype potentially having its origin in aberrant developmental signalling pathways in *DLG2*$^{-/-}$ NPCs although this requires further study. As a *DLG2*$^{-/-}$ phenotype exists during early cortical neurogenesis this suggests a novel role for *DLG2* in neurodevelopment in addition to its established function in synaptic transmission, although additional work is required to validate this and determine a mechanism of action.

5. Underlying mechanisms of the *DLG2*$^{-/-}$ phenotype

5.1. Introduction

In the previous chapter, through the use of ICC and western blotting to compare *DLG2*$^{-/-}$ and WT cells at various time points throughout cortical differentiation, it was shown that a clear phenotype exists in new-born *DLG2*$^{-/-}$ neurons whereby they exhibit a significant delay in the expression of neuronal subtype identity (See Chapter 4). This was characterised by a significant reduction in the expression of the neurodevelopmental transcription factor and cortical layer V marker CTIP2 in *DLG2*$^{-/-}$ cells compared to WT at all time points investigated, although relative expression levels did increase over the course of the differentiation they did not fully recover to WT levels. Supporting delayed maturation for *DLG2*$^{-/-}$ neurons a morphometric analysis also showed these cells possess a simpler structure when compared to WT, with long but relatively unbranched neurites. Interestingly *DLG2* knockout did not affect the rate of neurogenesis itself with cells expressing the postmitotic neuron markers including NEUN as well as cortical layer VI marker TBR1 in similar levels to WT, indicating the phenotype was restricted to the maturation of neuronal subtypes. Although not statistically significant there were also trending increase in the expression of PAX6 and KI67 in *DLG2*$^{-/-}$ NPCs, suggesting a potential role for DLG2 prior to neurogenesis during the early stages of cortical differentiation. Taken together these data indicated an unreported role for DLG2 in corticoneurogenesis; however, due to the inherent limitations in marker selection for ICC and WB the characterisation of the *DLG2*$^{-/-}$ phenotype during cortical differentiation was far from complete, with the possibility that many biological processes regulated by DLG2 in NPCs and neurons remained to be identified.

Numerous possible approaches are available to further study the *DLG2*$^{-/-}$ phenotype in NPCs and neurons, perhaps the simplest of which would be to perform ICC and western blotting with additional markers, particularly alternative cortical layer markers such as FOG2 for layer VI (Wiegreffe et al., 2015) or ER81 for layer V (Molyneaux et al., 2007). Phenotypic characterisation could also be extended to later stages of neuronal differentiation past day 60 to determine if the deficiencies observed in marker expression and morphology for *DLG2*$^{-/-}$ neurons do recover to WT levels given sufficient time in culture. As maintaining neurons in culture for long term can be challenging and even at their most mature *in vitro* status is less mature than those found *in vivo*, the transplantation of human *DLG2*$^{-/-}$ neurons into mouse brains could be used both to maintain neurons and promote their maturation allowing *DLG2*$^{-/-}$ phenotypes to be studied at later time points. An alternative or complementary approach would be the use of a *DLG2*$^{-/-}$ mouse model to confirm the observed phenotypes *in vivo* during early embryonic development and determine if they recover during later embryonic development or postnatally

Rescue experiments could also be performed both to further investigate the *DLG2*$^{-/-}$ phenotype *in vitro* and gain insight into the potential mechanism of action for DLG2 during early cortical neuron development. One such experiment would be to co-culture both WT and *DLG2*$^{-/-}$ hESC during the course of differentiation and repeat the various western blotting, ICC and morphometric analysis previously described; subsequently comparing the co-culture results to when *DLG2*$^{-/-}$ cells are differentiated in isolation. If the identified *DLG2*$^{-/-}$ phenotype is still present in co-cultures this would indicate that it is the result of solely intrinsic cellular processes. If no such phenotype was present in co-cultures this would indicate that external factors such as secreted molecules or proper cell-cell connections that are required for normal development are absent in exclusively *DLG2*$^{-/-}$ cultures resulting in the observed phenotype. An additional rescue experiment would be to reintroduce *DLG2* into the *DLG2*$^{-/-}$ hESC cells lines through the use of a tetracycline inducible lentiviral vector (Pluta et al., 2005) containing *DLG2* cDNA, followed by differentiation towards cortical projection neurons and subsequent phenotypic characterisation. Using this system if *DLG2* expression through the course of differentiation rescued the *DLG2*$^{-/-}$ phenotype and resulted in NPCs and neurons similar to WT this would provide both further validation of *DLG2*$^{-/-}$ lines and a role for DLG2 in the regulation of cortical neurogenesis. The inducible nature of this system would also allow *DLG2* expression only at specific time points of differentiation and could provide additional insights into both the *DLG2*$^{-/-}$ phenotype and the mechanism of action for DLG2. Expressing *DLG2* only at later time points of differentiation and determining whether or not the observed *DLG2*$^{-/-}$ deficits in CTIP2 expression and neuronal morphology show increased recovery would help confirm DLG2s function in the expression of neuronal subtype identity and clarify its role in the development of more mature neurons. In contrast, expression of *DLG2* only at earlier time points could be used to investigate the role of DLG2 in NPCs, if *DLG2*$^{-/-}$ neurons displayed a WT phenotype following DLG2 expression solely during cortical differentiation this could establish the role of DLG2 in regulating the internal state of NPCs as a crucial factor in subsequent neuronal identity.

As the work present here was performed using *DLG2*$^{-/-}$ hESC lines it could also be repeated using patient-derived *DLG2*$^{+/-}$ cell lines; although the focus of this work was to understand the role of *DLG2* in early neurodevelopment, for which homozygous cell lines are of greatest utility, comparison to a heterozygous line could provide additional insights into possible disease relevant phenotypes. Given the success of peptide affinity pulldowns coupled with mass spectrometry to both confirm WT DLG2 expression at day 30 of cortical differentiation and validate the *DLG2* knockout, a similar technique could be used to investigate the proteome of *DLG2*$^{-/-}$ and WT cells at multiple stages of differentiation in a more complete fashion that is not feasible using ICC and western blotting. A DLG2-FLAG knock in line such as described previously (see Chapter 3) would

enable specific DLG2 interacting proteins to be pulled-down and investigated but various enrichment methods could be used to isolate different protein subpopulations, such as synaptosomes, nuclear protein fractions or membrane bound protein isolations.

Although these additional experiments would have been of benefit to this research, time and resources prohibited their execution; however, an approach was required to further clarify the *DLG2*$^{-/-}$ phenotype and develop a hypothesis for the mechanism underlying it, ideally in a more systematic and high-throughput manner than is possible with ICC and western blotting. To this end RNAseq (Emrich et al., 2007; Wang et al., 2009; Stark et al., 2019) was used to quantify gene expression in genetically WT and *DLG2*$^{-/-}$ cells at various timepoints throughout differentiation from hESCs to cortical projection neurons. Analyses were performed to identify genes differentially expressed between WT and *DLG2*$^{-/-}$ lines at each time-point and to test whether the set of genes differentially expressed at each time-point showed enrichment for particular biological processes or phenotypes. Finally, experimental assays were performed to validate a number of predicted phenotypes arising from these analyses.

5.2. Experimental procedure - outline

RNA samples for RNAseq were extracted in triplicate from WT and *DLG2*$^{-/-}$ (KO1 & KO2) cell cultures at 4 time points during cortical differentiation from hESCs (days 15, 20, 30 and 60 for a total of 36 samples). The time points were selected to cover a range of developmental stages and cell types (Figure 5.1). These were weighted towards early development to better investigate the role of *DLG2* prior to PSD formation and synaptic transmission. After sample validation and library construction, 75bp paired-end sequencing at a depth of 20 million reads was performed using an Illumina HiSeq4000. After QC the resulting reads were mapped to an indexed reference genome to produce counts of gene expression. Gene counts were normalised and sets of differentially expressed genes were created between *DLG2*$^{-/-}$ and WT cells for all time points of the experiment. Functional enrichment analyses were then performed on these sets of differentially expressed genes using gene ontology and mammalian phenotype data (Ashburner et al., 2000; Smith and Eppig, 2009; Carbon et al., 2019), to identify particular biological processes significantly over or under-represented in *DLG2*$^{-/-}$ cells. Finally, experimental assays were used to determine whether any such biological processes identified from the functional enrichment analysis show evidence of dysregulation in cultured WT and *DLG2*$^{-/-}$ cells, thereby validating the bioinformatic analyses. Detailed protocols can be found in Chapter 2*,* Sections *2.2 Cortical differentiation, 2.14 RNA sequencing, 2.15 Adhesion assay, 2.16 proliferation assay, 2.17 Migration assay, 2.18 Electrophysiology;* details of the RNAseq analysis steps follow below.

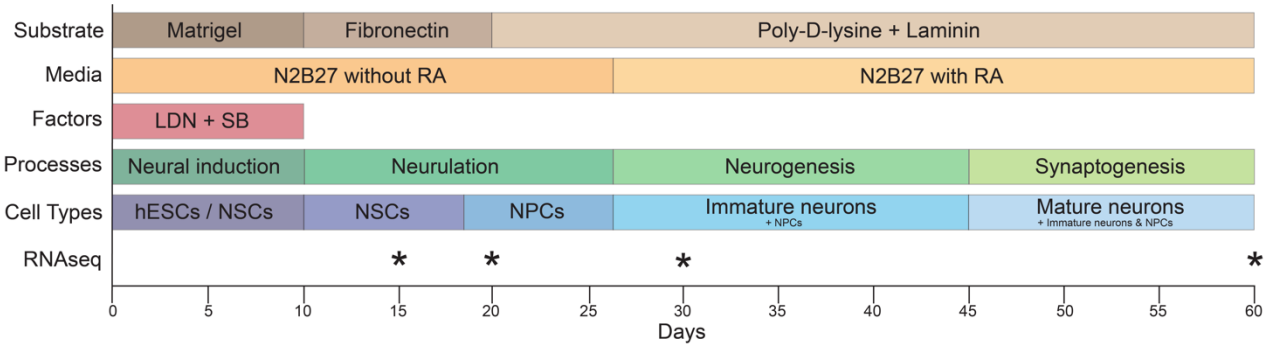


Figure 5.1. Overview of cortical differentiation protocol
Overview of cortical differentiation protocol with approximate timings of key developmental processes and predominant cell types present in culture. Asterisks indicate time point selected for RNA sequencing. The coating substrates, culture medium and small molecule factors used during differentiation are fully described in Chapter 2, section *2.2 Cortical differentiation.*

5.3. Large numbers of genes were differentially expressed in *DLG2*$^{-/-}$ cells during cortical differentiation

The integrity of the RNA samples extracted from WT, KO1 & KO2 (*DLG2*$^{-/-}$) at days 15, 20, 30 and 60 of cortical differentiation was determined using an RNA integrity number (RIN) algorithm following capillary gel electrophoresis, which assigns a value of 1 to 10 based on the amount of RNA degradation. A high RIN was obtained for all 36 samples with all ≥ 8.5 and most ≥ 9.5 (Figure 5.2, Supplementary Figure 2) indicating they were of good quality with low amounts of RNA degradation, so these were subsequently used for the construction of stranded cDNA libraries for RNAseq. These libraries were quantified and verified to contain a sufficient cDNA of an appropriate size for efficient sequencing for the majority of samples, the exception being triplicate repeats of day 30 KO2 samples. These 3 day 30 samples were of lower concentrations than expected and did not clearly show a peak DNA fragment size within the optimal range for sequencing, 200 – 300 base pairs, additional they showed a small peak at roughly 100bp potentially indicating a PCR amplification of library adapter sequences over cDNA libraries (Figure 5.3). These results were noted for further investigation in later analyses; however, the day 30 KO2 samples were included when producing equimolar pools of the constructed libraries which were subsequently sequenced with the help from Joanne Morgan in the division sequencing core team**.**

Cell Line	Day 15			Day 20			Day 30			Day 60		
	A	B	C	A	B	C	A	B	C	A	B	C
WT	9.8	9.9	9.8	9.7	9.6	9.7	9.5	9.5	9.5	9.2	9	9
KO1	9.9	9.8	9.9	10	9.9	10	9.3	9.4	9.2	9.5	9.5	9.5
KO2	10	10	10	9.6	9.6	9.6	9.3	8.5	9.2	9.5	9.4	9.3

Figure 5.2. Quality of RNA extracted from WT and *DLG2*$^{-/-}$ cultures as determine by RIN
RNA integrity numbers (RIN) for triplicate RNA samples (A, B, C) extracted from WT and DLG2$^{-/-}$ cultures (KO1 & KO2) and 4 time points of cortical differentiation from hESCs. Determined using the Agilent RNA 6000 Nano kit analysed using the 2100 Bioanalyzer Eukaryote Total RNA Nano assay. RIN numbers are on a scale of 1 (low quality RNA) to 10 (high quality RNA).

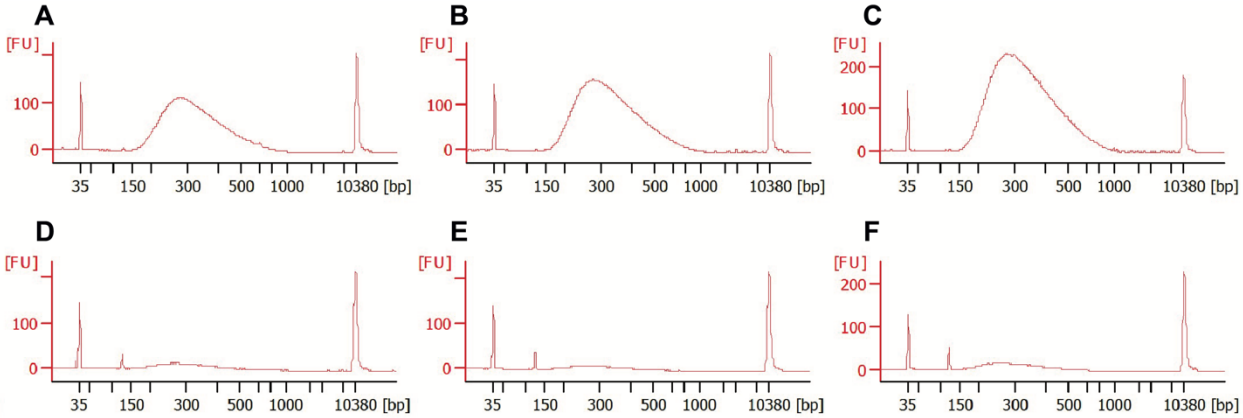


Figure 5.3. RNAseq library quantification
Quantification of fragment size in cDNA libraries for RNA sequencing, determined using the Agilent Bioanalyzer high sensitivity DNA assay. X axis gives size of fragments in base pairs, the y axis frequency. (A), (B) & (C) are triplicate repeats of day 60 KO2 samples and are representative of the majority of the constructed libraries, showing two sharp ladder peaks at 35 bp and 10000 bp in between which is a defined peak of cDNA fragments at 200-300 bp, the required size for sequencing. (D), (E), & (F) are triplicate repeats of day 30 KO2 samples and show a greatly reduced 200-300 bp peak along with an additional peak at ~100bp likely corresponding to amplified adapter sequences.

Following Illumina sequencing the *Trimmomatic* software (Bolger et al., 2014) was used with default parameters to trim the fastq files from the sequencer prior to mapping to a genome index. This trimming process ensures that any residual adapter sequences, incorporated during cDNA library construction, are removed from the final reads along with any low-quality bases as determined by PHRED score (PHRED score < 30). A quality check of both the original and trimmed reads was performed using the *FastQC* software (Andrews, 2010) which showed good statistics including per base quality, per sequence quality, overexpressed sequences, adapter content and others (Supplementary File 1). The *STAR* software (Dobin et al., 2013) was then used to index the GRCh38.p12 assembly of the human reference genome (Zerbino et al., 2018) and map the

generated trimmed reads to the resulting genome index. As an additional QC step the *Picard* software (Broad Institue, 2019) and *BamTools* software (Barnett et al., 2011) was used to mark duplicates within the mapped trimmed reads for each library and compute the percentage of duplication. Sequence duplication occurs when multiple reads map to the same locus due to the over-amplification of a single piece of cDNA. Although this inevitably occurs to some degree during RNAseq, excessive duplication can be the result of low-quality libraries or incorrect pooling of libraries prior to sequencing. For the majority of the mapped reads sequence duplication was relatively low, in the range of 15 – 25%, with the exception of the previously highlighted day 30 KO2 samples which showed an increased level of sequence duplication of approximately 70% (Supplementary File 2). Finally, gene expression of the 36 samples was quantified from the mapped trimmed reads using the *featureCounts* software (Liao et al., 2014) by computing the number of raw reads aligning to exons within a gene.

The *DESEQ2* package (Love et al., 2014) was used to normalise the raw read counts per gene generated from the sequencing data and analyse them for evidence of systematic changes between the cell lines i.e. determine differential gene expression between *DLG2*$^{-/-}$ and WT cells. Analyses of RNA sequencing data to determine differential gene expression can suffer from a lack of statistical power due to relatively low replicate numbers per condition, large dynamic range, the presence of outliers and other factors. To overcome these factors *DESEQ2* utilises the assumption that genes of a similar average expression level have a similar dispersion, a statistical model of variability within a group, allowing information to be pooled across genes and increasing statistical power. *DESEQ2* was used to analyse differential gene expression between *DLG2*$^{-/-}$ cells and genetically WT cells at each of the four time points of cortical differentiation investigated. This was performed both separately for KO1 and KO2 compared to WT, (3 replicates KO vs 3 replicates WT) (Supplementary Tables 2b, 2c) and for the combined data from DLG2$^{-/-}$ lines compared to the WT (6 replicates KO vs 3 replicates WT) (Supplementary Table 2a).

To investigate variation between the samples principal component analysis (PCA) and hierarchical clustering was performed using the genes expressed (defined as expression > 0 in at least 3 samples) in KO1, KO2 and WT at each time point (Figure 5.4). The resulting plots of principal components 1 and 2 (Figure 5.4 A) show generally good clustering between replicates. Those showing the greatest variability are notably the day 30 KO2 samples, previously found to have a high degree of sequence duplication potentially due to a low-quality initial cDNA library. As PCA did in some cases indicate a degree of biological variability both within replicates and between the two *DLG2*$^{-/-}$ cell lines, hierarchical clustering was performed to better determine the similarity of samples. Unlike PCA, which seeks to identify factors that maximise apparent differences between samples, hierarchical clustering groups samples according to their similarity taking into account all variation

in expression. The results of this clustering (Figure 5.4B) both show the similarity of triplicate samples, which cluster together in all cases, as well as the similarity between the KO1 and KO2 cell lines, which again in all cases cluster together at a given time point. These results also indicate that the *DLG2*$^{-/-}$ and WT cells lines are most similar at the earliest time point investigated (day 15), where both KO1 and KO2 cluster with WT samples (Figure 5.4 B), a point also emphasised by the tight PCA grouping of samples at this time point (Figure 5.4 A). Interestingly, later in differentiation (D20, D30 & D60) KO1 and KO2 cluster with WT samples not from their current time point but rather the one immediately prior (Figure 5.4 B), e.g. KO1 & KO2 D30 samples cluster with WT D20. This is consistent with the delayed developmental changes in gene expression seen in *DLG2*$^{-/-}$ cells in the previous chapter.

PCA and hierarchical cluster analysis of RNAseq data therefore indicate that despite some biological variability KO1 and KO2 are broadly similar, this is in agreement with ICC and western blotting data which shows a similar trend between the lines but not identical results through all analyses (Supplementary Figure 1). To increase statistical power, minimise the effects of biological variability and focus on the core phenotypes resulting from *DLG2*$^{-/-}$ deficiency the decision was taken to use the differential gene expression data from the two combined *DLG2*$^{-/-}$ lines compared to WT, which is presented here. However, the further analysis of RNAseq data presented in this chapter was also performed for KO1 and KO2 separately which is available in supplementary materials (Supplementary Tables 3b, 3c, 4b, 4c), again these analyses show broadly similar results between the two cells lines as well as those using the combined data.

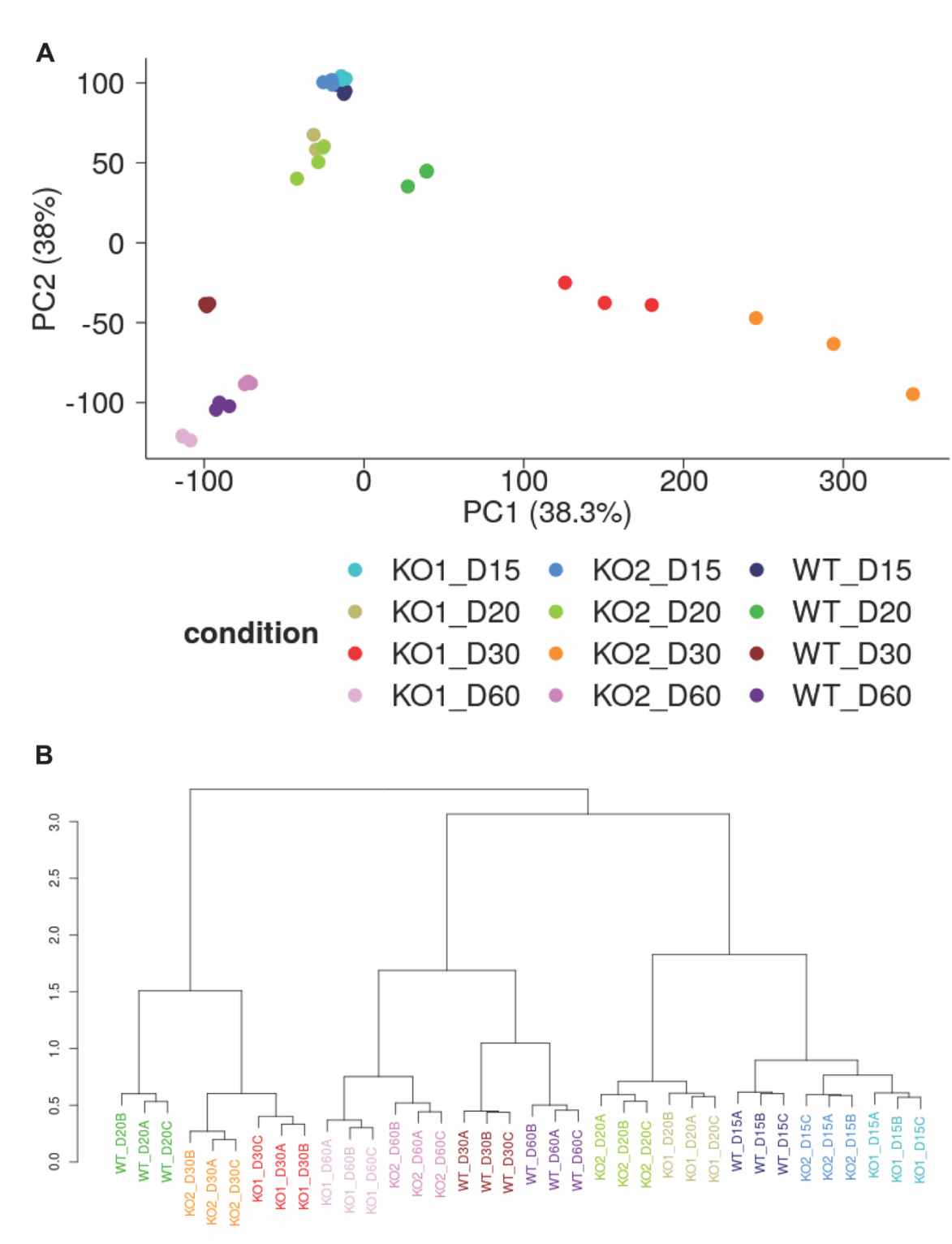


Figure 5.4. PCA and hierarchical clustering of genes expressed in WT and *DLG2*$^{-/-}$ cultures
Principle component analyses (**A**) and hierarchical clustering (**B**) of genes expressed (expression > 0 in at least 3 samples) between triplicate samples (A, B, C) of *DLG2*$^{-/-}$ (KO1 & KO2) and genetically WT cultures on four separate days (D15, D20, D30, D60) of cortical differentiation from hESC. Triplicate samples are plotted for the PCA in call cases also although all 3 points may not be visible due to overlap between replicates.

Prior to an analysis of differential gene expression for the combined KO against WT data sets the *GENEsis* (Parker and Clifton, 2018) gene conversion program was kindly provided by Dr Nicholas Clifton (NMHRI, Cardiff University) to convert the Ensembl IDs used to identify genes throughout the RNAseq analysis up to this point into Entrez Gene IDs. Genes were also separated into up-regulated (positive log2 fold change) and down-regulated (negative log 2 fold change) groups and non-protein coding genes removed. Histograms of -log10 p-value distribution were produced for the up-regulated and down-regulated protein coding genes at each time point, in all cases a positive -log10 p-value indicates up-regulated genes and a negative -log10 p-value down-regulated genes (Figure 5.5). The histograms for days 15, 20 and 60 showed the expected pattern with a clear peak frequency near 0, indicating that most genes show no significant differential expression, that decreases relatively smoothly as the -log10 p-value increases or decreases and with a roughly even distribution of up-regulated and down-regulated genes. However, the histogram for day 30 showed an unexpected pattern with a second peak present in the up-regulated differentially expressed genes, which disappeared when the day 30 KO2 results were removed (Figure 5.5).

The unusual -log10 p-value distribution identified in day 30 KO2 samples (Figure 5.5), along with their previously noted broad PCA clustering (Figure 5.4) high sequence duplication (Supplementary File 2) and reduced 200-300bp peak during RNAseq library quantification (Figure 5.3) resulted in the removal of these samples from the combined analysis. This decision was taken to ensure only high-quality samples were used in analyses and increase confidence in the results obtained. The data presented here for days 15, 20 and 60 therefore corresponds to the two *DLG2*$^{-/-}$ lines combined, while day 30 results are for one *DLG2*$^{-/-}$ line only (KO1) in triplicate. However, data for KO2 at day 30 available as part of the separate line analyses presented in supplementary information (Supplementary Tables 3c, 4c) and shows similar results to KO1 at day 30 despite the issues with samples quality previously noted.

Differential gene expression analysis of the day 15, 20, 30 and 60 samples showed large numbers of protein coding genes were significantly differentially expressed (Bonferroni $p < 0.05$) in *DLG2*$^{-/-}$ cells compared to WT, peaking at day 30 (Figure 5.5). This included both up and downregulated differentially expressed genes (DEGs) but when considered as whole 5.74% of human protein coding genes were differentially expressed in *DLG2*$^{-/-}$ cells at D15, 36.9% at D20, 47.0% at D30 and 21.5% at D60. Therefore, these data clearly show a large-scale dysregulation of gene expression in cultures lacking the *DLG2* gene during the course of differentiation from hESC to cortical projection neurons; this includes all developmental stages investigated, not only the later time points

known to contain more mature and electrically active neurons but also earlier time points containing NPCs and immature neurons.

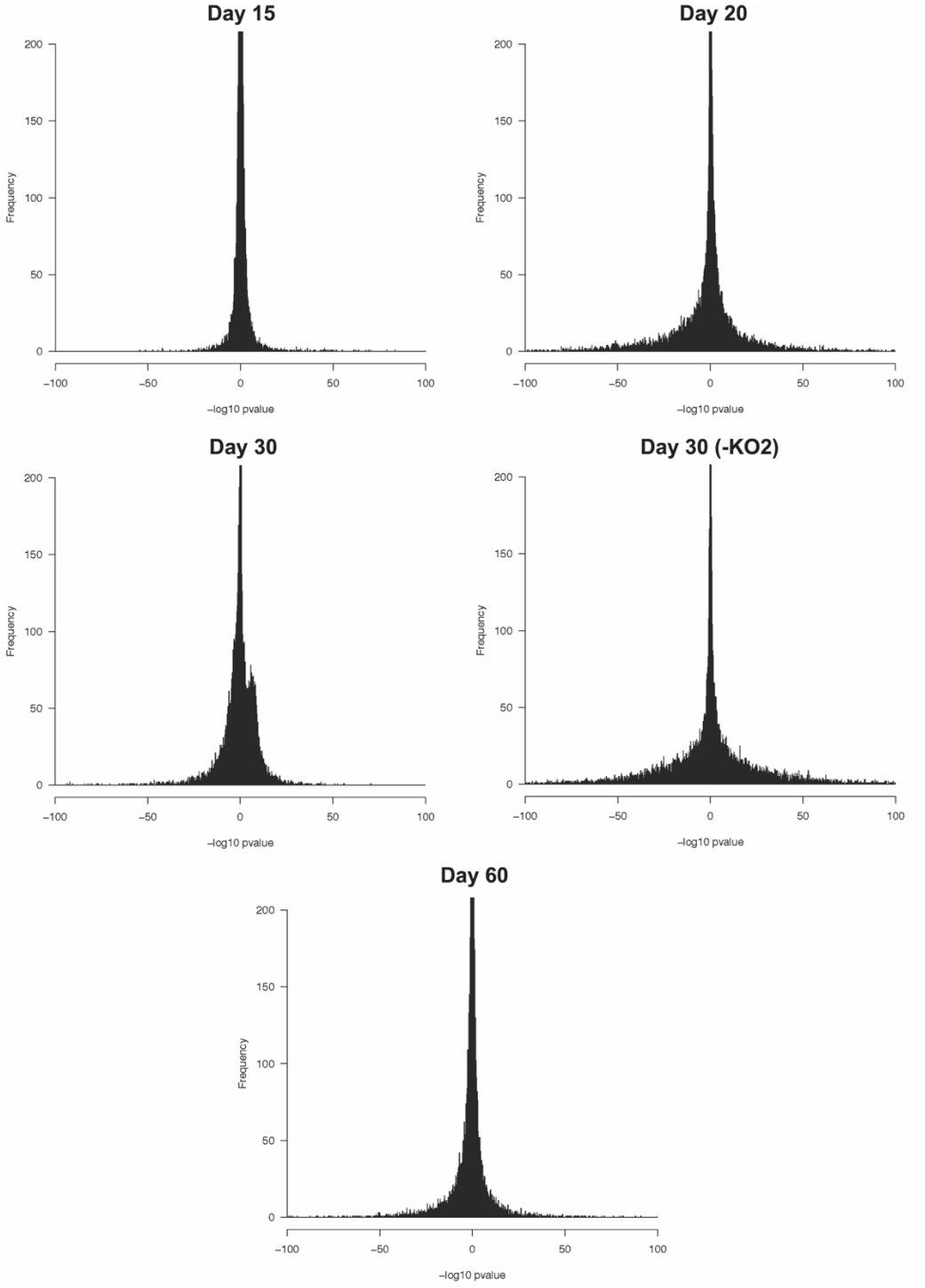


Figure 5.5. p-value distribution in genes differentially expressed between *DLG2*$^{-/-}$ and WT cultures
-log10 p-value distribution of protein coding genes differentially expressed between *DLG2*$^{-/-}$ and genetically WT cultures at four time points of cortical differentiation from hESCs. Positive -log10 p-values indicate up-regulated genes and a negative -log10 p-values down-regulated genes. Unless otherwise specified differential gene expression is for both *DLG2*$^{-/-}$ cell lines (KO1 & KO2) combined, compared to WT. For all histograms Frequency is limited to 200 and -log10 p-value to -/+ 100, bin size 0.1.

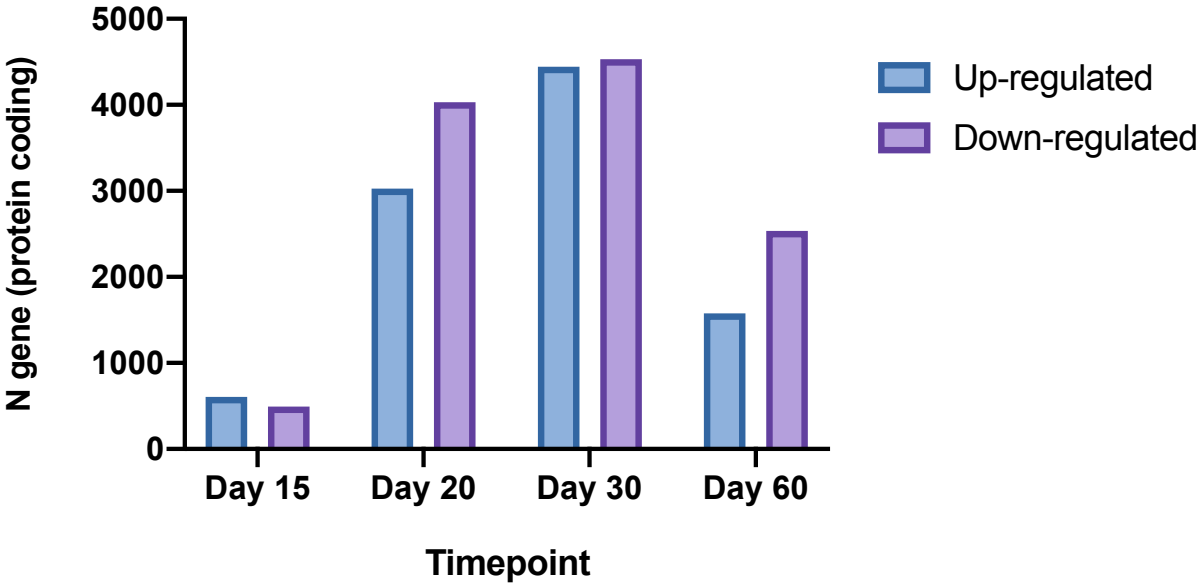


Figure 5.6. Differentially expressed protein coding genes between *DLG2*$^{-/-}$ and WT cultures
Up and down-regulated protein coding genes significantly differentially expressed (Bonferroni $p < 0.05$) between *DLG2*$^{-/-}$ (KO1 & KO2 combined) and genetically WT cultures at four time points of cortical differentiation from hESCs.

5.4. DEGs suggest neurodevelopmental deficits in *DLG2*$^{-/-}$ NSCs

Even at the earliest time point of investigated (day 15) large numbers of protein coding genes (~1100) were differentially expressed in *DLG2*$^{-/-}$ compared to WT (Figure 5.6) clearly indicating a role for DLG2 in the regulation of early corticoneurogenesis. A point further emphasised by the top 3 differentially expressed genes (by Bonferroni corrected P value) in *DLG2*$^{-/-}$ cells at day 15 belonging to the inhibitor of DNA binding (ID) family of transcriptional factors (Figure 5.7 A), these genes also being among the most differentially expressed at this timepoint when KO1 and KO2 are analysed separately (Supplementary Tables 2b, 2c). The *ID* genes encoding highly conserved proteins that act as key regulators of stem cell differentiation (Jung et al., 2010; Niola et al., 2012). However, to further interpret the gene expression data and clarify the role of *DLG2* in neurodevelopment, Fisher's exact tests were performed to analyse the sets of genes differentially expressed at each experimental time point for functional enrichment, using an R script written and kindly provided by Dr. Andrew Pocklington (MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University) (Supplementary Figure 3). These analyses enabled the assignment of properties to the differentially expressed gene sets by identifying functional classes of genes (those associated with a specific biological process or phenotype) that were overrepresented within the set. Two sources of gene classification were used for this functional enrichment analysis, a

strictly annotated (removing those with the IEA – inferred from Electronic Annotation – evidence code) set of GO terms (Ashburner et al., 2000; Carbon et al., 2019) and MP (mouse phenotype) data from the mouse genome informatics (MGI) resource (Smith and Eppig, 2009). The MP data was restricted to phenotypes reported in single-gene functional studies and where the mouse gene had a 1:1 mapping onto a human gene. For each of the 4 time points, significantly differentially expressed (Bonferroni $p< 0.05$) protein coding genes were separated into up-regulated and down-regulated gene sets; functional enrichment for both GO and MP terms was then performed on each of these sets. A complete set of protein coding genes expressed throughout all samples and time points was used as the background (comparator) gene set for these analyses rather than all human protein coding genes, thereby preventing inflated enrichment for neural related terms by virtue of the fact that these gene sets were derived cultures undergoing neuronal differentiation.

The results of these initial analyses showed significant functional enrichment (Bonferroni $p < 0.05$) for large numbers of GO and MP terms in both up-regulated and down-regulated gene sets at all time points (Supplementary Tables 3a & 4a). There can be considerable overlap in GO and MP terms due to gene pleiotropy in addition to the hierarchical nature of these ontologies, making it unclear how many relatively independent sets of gene are driving observed functional enrichment. However, these initial analyses can be refined to reduce the number of terms to a more manageable size and group terms whose enrichment is driven by a common set of genes. To this end sets of genes were recursively removed from the significant GO and MP terms, beginning with the genes present in the term with the largest odds ratio (OR), re-testing and only retaining terms remaining significant (Bonferroni $p<0.05$) before proceeding to remove the set of genes present in the term with the next highest OR, re-testing and so on. This recursive refinement by OR continued until concise lists of enriched GO and MP terms, with independent evidence for enrichment, were produced.

At the earliest time point of cortical differentiation studied, day 15, functional enrichment results indicate that key developmental processes are already dysregulated in *DLG2*$^{-/-}$ cells (Figure 5.7). For GO terms (Figure 5.7 B), down-regulated genes show enrichment for *nervous system development* while correspondingly the up-regulated genes show an enrichment for the negative regulation of key neurodevelopmental signalling pathways, namely *receptor serine threonine kinase signalling* and *BMP signalling* (Derynck and Zhang, 2003; Hegarty et al., 2013). Separate analysis of the *DLG2*$^{-/-}$ lines also showing a disruption to nervous system development and developmental signalling pathways at day 15 of differentiation is a feature of both KO1 and KO2 (Supplementary Tables 3b, 3c). Furthermore, the MP terms (Figure 5.6 C) are in broad agreement with the those for GO terms, down-regulated genes being enriched for phenotypes related to nervous system function such as *abnormal nervous system physiology*, *abnormal long-term potentiation* and

abnormal spatial learning. Taken together these data therefore provide further evidence for the role of *DLG2* as a component of signalling pathways regulating early corticoneurogenesis and indicate that biological processes regulated by DLG2 are not restricted to neurons but are also present in NSCs.

A

NCBI	Symbol	Log_2 fold change	SE	P	$P_{corrected}$
3397	ID1	3.49072091	0.1083495	1.01x10-227	$2.24x10^{-223}$
3398	ID2	2.29879346	0.0787895	3.87x10-187	$8.56x10^{-183}$
3399	ID3	2.70473374	0.09288251	2.01x10-186	$4.44x10^{-182}$
1277	COL1A1	1.48926057	0.05148364	5.48x10-184	$1.21x10^{-179}$
151449	GDF7	3.68039226	0.13684965	2.60x10-159	$5.76x10^{-155}$

B

GO Term (up-regulated genes)	$N^{overlap}$	OR	P	$P_{corrected}$
anatomical structure morphogenesis	97	2.92892854	$5.24x10^{-17}$	$2.24x10^{-13}$
endoplasmic reticulum lumen	37	5.00465204	$1.36x10^{-13}$	$5.80x10^{-10}$
negative regulation of transmembrane receptor protein serine/threonine kinase signalling pathway	17	7.69200095	$1.86x10^{-09}$	$7.95x10^{-06}$
cardiac septum morphogenesis	11	9.86954656	$1.65x10^{-07}$	0.000707416
biological adhesion	40	2.67071914	$2.36x10^{-07}$	0.001009087
regulation of mononuclear cell migration	8	17.2748025	$3.21x10^{-07}$	0.001373233
regulation of heart morphogenesis	8	17.2748025	$3.21x10^{-07}$	0.001373233
cell death	42	2.56344284	$3.32x10^{-07}$	0.001421622
collagen metabolic process	11	6.49942353	$5.13x10^{-06}$	0.021929904
negative regulation of BMP signalling pathway	8	10.361395	$6.07x10^{-06}$	0.025950901
lipoprotein particle receptor binding	7	12.9310817	$7.51x10^{-06}$	0.032126946
negative regulation of response to wounding	10	7.0176152	$7.69x10^{-06}$	0.032901999
skin development	14	4.56223898	$1.05x10^{-05}$	0.044943941

GO Term (down-regulated genes)	$N^{overlap}$	OR	P	$P_{corrected}$
cholesterol biosynthetic process	14	28.5050567	$4.81x10^{-14}$	$2.06x10^{-10}$
nervous system development	70	2.05005074	$3.62x10^{-07}$	0.001546666
negative regulation of gene expression	66	1.96561715	$2.52x10^{-06}$	0.010790069
nuclear chromosome part	32	2.54603025	$7.71x10^{-06}$	0.032958177

C

MP Term (up-regulated genes)	$N^{overlap}$	OR	P	$P_{corrected}$
abnormal craniofacial morphology	94	3.45960604	$1.02x10^{-20}$	$3.02x10^{-17}$
abnormal renal/urinary system morphology	71	3.34577197	$1.04x10^{-15}$	$3.06x10^{-12}$
no abnormal phenotype detected	106	2.09507624	$3.24x10^{-10}$	$9.56x10^{-07}$
abnormal heart septum morphology	32	3.33631539	$3.95x10^{-08}$	0.0001164
aneurysm	10	12.9907994	$7.95x10^{-08}$	0.00023459
abnormal tooth development	13	7.88775484	$1.12x10^{-07}$	0.00033105
abnormal maxilla morphology	18	4.96867617	$2.11x10^{-07}$	0.00062112
abnormal hypaxial muscle morphology	17	5.05941981	$3.60x10^{-07}$	0.00106114
abnormal metanephric mesenchyme morphology	8	12.9509258	$1.64x10^{-06}$	0.00484026
abnormal iris morphology	15	5.02271847	$1.84x10^{-06}$	0.00541738

MP Term (down-regulated genes)	$N^{overlap}$	OR	P	$P_{corrected}$
abnormal nervous system physiology	81	2.0573792	$5.36x10^{-08}$	0.00015815
abnormal long term potentiation	20	4.39429779	$2.35x10^{-07}$	0.00069224
abnormal spatial learning	21	4.17117369	$2.62x10^{-07}$	0.00077373

Figure 5.7. Day 15 DEGs, GO and MP functional enrichment
(A) Top 5 differentially expressed genes (DEGs) in *DLG2*$^{-/-}$ cultures at day 15 of cortical differentiation. **(B+C)** Fisher's exact test for GO term (B) and MP term (C) functional enrichment in sets of up-regulated and down-regulated protein coding genes significantly differentially expressed (Bonferroni $p < 0.05$) between *DLG2*$^{-/-}$ and WT cultures at day 15 of cortical differentiation from hESCs. Enrichment resulting from unique sets of differentially expressed genes was achieved following recursive refinement by odds ratio (OR). All protein coding genes expressed at all experimental timepoints were included in the analysis as a comparator gene set, $N^{overlap}$ = number of unique differentially expressed genes captured by term.

5.5. *DLG2*$^{-/-}$ NPCs display increased proliferation and adhesion to the ECM

Enriched GO and MP terms were surveyed to determine common themes and identify potential *DLG2*$^{-/-}$ phenotypes and DLG2 regulated biological processes that could be experimentally validated. Cellular adhesion to the ECM and proliferation were two such processes, identified as being dysregulated in *DLG2*$^{-/-}$ NPCs through functional enrichment analyses for GO terms at days 15 (Figure 5.7) and 20 (Figure 5.8). Several terms relating to both processes also being found in the sperate analysis of KO1 and KO2 at days 15 and 20, further indicating the similarity between the *DLG2*$^{-/-}$ cells lines (Supplementary Tables 3b, 3c). Up-regulated differentially expressed genes were enriched for GO terms at D15 that included *biological adhesion* and *collagen metabolic process* (Figure 5.7 B) similarly D20 up-regulated genes were enriched for *extracellular matrix organisation*, *cell adhesion* and others (Figure 5.8 A). Enrichment for processes related to ECM adhesion in proliferative cells at early time points of differentiation (in D15 NSCs and D20 NPCs) is perhaps unsurprising, given that 4 out of the 5 most significantly differentially expressed genes at day 15 are related to this process; *ID1*, *ID2* and *ID3* are all upregulated and have established roles in maintaining cellular adhesion to the ECM (Jung et al., 2010; Niola et al., 2012) while CO*L1A1* encodes the pro-α1 chain of collagen, a key ECM component in the developing cortex (Long et al., 2018). Based on these results we would therefore predict that *DLG2*$^{-/-}$ cells would show increased adhesion to the ECM compared to WT and an experimental assay was performed to test this. Cells at day 26 of cortical differentiation were selected for this assay, this time point chosen as being similar to but slightly advanced of the time point for the corresponding functional enrichment analyses, allowing time for any changes at the RNA level to clearly manifest in culture. For this assay equal numbers of WT and *DLG2*$^{-/-}$ cells were plated on to seven different ECM substrates, including three collagen types, in triplicate and the number of adhering cells quantified in colorimetric manner. *DLG2*$^{-/-}$ cells did show significantly increased adhesion to various ECM substrates compared to WT cells particularly to types I and II collagen and fibronectin, with others including type IV collagen and laminin showing an increased trend (Figure 5.9 A); interestingly both laminin and fibronectin, are known mediators of ID protein regulated adhesion (Niola et al., 2012). These assays therefore confirm the prediction of increased adhesion and validate the results of the functional enrichment analyses.

Cell proliferation was a second biological process identified as being dysregulated in *DLG2*$^{-/-}$ cells in the day 20 functional enrichment results (Figure 5.8). The refined functional enrichment results show that genes related to the mitotic cell cycle are enriched in the day 20 down-regulated gene set (Figure 5.8B). However, the enrichment results prior to refinement by odds ratio indicate that the down-regulated genes are related to the negative regulation of the cell cycle with enrichment for terms such as *negative regulation of cell cycle*, *negative regulation of cell cycle G2/M phase*

transition and others while the day 20 up-regulated gene set is in enriched for the *positive regulation of cell proliferation* (Figure 5.8 C, Supplementary Table 3a). The refined enrichment analysis for MP terms also supports a phenotype of increased proliferation in *DLG2*$^{-/-}$ cells with the day 20 up-regulated gene set being enriched for terms including abnormal *kidney cell proliferation* and *altered tumour pathology*. Again, perhaps unsurprising given that ID1, ID2, ID3 show significantly increased expression in the *DLG2*$^{-/-}$ during early neurodevelopment and are known to regulate the cell cycle with their expression increasing the proliferation of cortical NSCs (Jung et al., 2010; Niola et al., 2012). An assay was therefore performed to further investigate proliferation in WT and *DLG2*$^{-/-}$ cells, on day 26 of cortical differentiation, chosen for similar reasons to those described for the ECM adhesion assay. This assay involved plating equal numbers of WT and *DLG2*$^{-/-}$ cells in triplicate and monitoring their proliferation at 2-day intervals using a colorimetric approach to quantify the numbers of cells in culture at a given time. The results of this assay (Figure 5.9c) agree with both experimental observations and the functional enrichment analyses in that *DLG2*$^{-/-}$ cells, show significantly increased proliferation from day 26 of cortical differentiation compared to WT. As a negative control an additional assay was performed to investigate proliferation in WT and *DLG2*$^{-/-}$ cells maintained as hESCs, rather than during the course of cortical differentiation. There was no significant difference in proliferation (Figure 5.9b) between the three cell lines at the hESC stage, thereby indicating that an increase in proliferation for *DLG2*$^{-/-}$ cells is a cell type specific neurodevelopmental phenotype that mirrors the changes seen in gene expression.

A

GO Term (up-regulated genes)	$N^{overlap}$	OR	P	$P_{corrected}$
extracellular matrix organization	95	4.07997925	$6.94x10^{-21}$	$2.97x10^{-17}$
cell adhesion	148	2.29031317	$1.62x10^{-14}$	$6.91x10^{-11}$
regulation of intracellular signal transduction	349	1.66318243	$6.35x10^{-14}$	$2.71x10^{-10}$
regulation of protein phosphorylation	264	1.75785846	$4.33x10^{-13}$	$1.85x10^{-09}$
collagen trimer	17	14.7558819	$4.40x10^{-09}$	$1.88x10^{-05}$
Golgi apparatus part	174	1.72680854	$6.52x10^{-09}$	$2.79x10^{-05}$
apical part of cell	53	2.72312786	$5.37x10^{-08}$	0.00022954
collagen fibril organization	15	13.0191163	$7.45x10^{-08}$	0.00031848
basement membrane	21	6.51604095	$8.07x10^{-08}$	0.00034494
negative regulation of response to wounding	25	4.94009929	$1.28x10^{-07}$	0.00054542
negative regulation of cellular response to growth factor stimulus	37	3.22321773	$2.76x10^{-07}$	0.00117934
cardiac septum morphogenesis	22	5.30964046	$3.35x10^{-07}$	0.00143235
cell leading edge	72	2.12799923	$5.04x10^{-07}$	0.00215339
PDZ domain binding	28	3.92788057	$5.15x10^{-07}$	0.0022002
regulation of smooth muscle cell migration	15	8.13256447	$1.32x10^{-06}$	0.00566234
ERBB signalling pathway	30	3.18193882	$4.25x10^{-06}$	0.01817615
keratinocyte differentiation	25	3.62062103	$5.53x10^{-06}$	0.02364806
cardiac cell development	15	6.50505033	$6.02x10^{-06}$	0.02573851
occluding junction	27	3.35263539	$6.42x10^{-06}$	0.0274351
protein tyrosine phosphatase activity	25	3.50356949	$8.22x10^{-06}$	0.03512998

GO Term (down-regulated genes)	$N^{overlap}$	OR	P	$P_{corrected}$
translation	214	8.09577618	$4.63x10^{-66}$	$1.98x10^{-62}$
rRNA processing	153	6.5730146	$2.38x10^{-42}$	$1.02x10^{-38}$
SRP-dependent cotranslational protein targeting to membrane	80	30.5213064	$2.94x10^{-39}$	$1.26x10^{-35}$
mitochondrial matrix	206	3.65995146	$1.60x10^{-34}$	$6.84x10^{-31}$
inner mitochondrial membrane protein complex	90	11.458667	$1.20x10^{-33}$	$5.12x10^{-30}$
mitochondrial translation	83	10.5484948	$3.94x10^{-30}$	$1.68x10^{-26}$
nucleoside monophosphate metabolic process	113	5.0887114	$1.29x10^{-26}$	$5.53x10^{-23}$
organellar ribosome	65	14.1095097	$1.37x10^{-26}$	$5.86x10^{-23}$
RNA splicing, via transesterification reactions	143	3.43082734	$6.34x10^{-23}$	$2.71x10^{-19}$
mitotic cell cycle	256	2.16832902	$1.81x10^{-19}$	$7.76x10^{-16}$
mitochondrial respiratory chain complex I biogenesis	47	12.9292014	$4.72x10^{-19}$	$2.02x10^{-15}$
anaphasx10-promoting complex-dependent catabolic process	56	8.07930671	$1.80x10^{-18}$	$7.69x10^{-15}$
endopeptidase complex	42	15.8650812	$2.73x10^{-18}$	$1.17x10^{-14}$
mitochondrial electron transport, NADH to ubiquinone	32	19.308808	$5.68x10^{-15}$	$2.43x10^{-11}$
ribosomal large subunit biogenesis	45	7.15923011	$3.19x10^{-14}$	$1.36x10^{-10}$
DNA replication	85	3.35445938	$5.38x10^{-14}$	$2.30x10^{-10}$
spliceosomal snRNP complex	42	6.67662445	$7.13x10^{-13}$	$3.05x10^{-09}$
cristae formation	23	34.6165386	$2.44x10^{-12}$	$1.04x10^{-08}$
catalytic step 2 spliceosome	54	3.88767404	$6.52x10^{-11}$	$2.79x10^{-07}$
chromatin assembly or disassembly	65	3.17388037	$1.97x10^{-10}$	$8.43x10^{-07}$
spliceosomal tri-snRNP complex	23	13.8431108	$3.53x10^{-10}$	$1.51x10^{-06}$
tRNA processing	49	3.70011041	$1.34x10^{-09}$	$5.75x10^{-06}$
nuclear chromosome, telomeric region	60	3.02390144	$3.14x10^{-09}$	$1.34x10^{-05}$
mitochondrial small ribosomal subunit	20	15.0339796	$3.25x10^{-09}$	$1.39x10^{-05}$
sister chromatid cohesion	56	3.02186863	$1.04x10^{-08}$	$4.45x10^{-05}$
protein targeting to mitochondrion	31	5.18799481	$1.70x10^{-08}$	$7.26x10^{-05}$
cofactor biosynthetic process	67	2.59819885	$2.35x10^{-08}$	0.00010035
proteasome regulatory particle	17	17.0381074	$2.96x10^{-08}$	0.00012678
intrinsic component of mitochondrial membrane	34	4.45431926	$3.00x10^{-08}$	0.00012808
mitochondrial intermembrane space	33	4.51904174	$3.91x10^{-08}$	0.000167
termination of RNA polymerase II transcription	38	3.81856277	$5.48x10^{-08}$	0.00023444
deoxyribonucleotide metabolic process	20	8.5893158	$1.21x10^{-07}$	0.00051665
ncRNA 3'-end processing	18	9.01531389	$3.89x10^{-07}$	0.00166392
double-strand break repair	68	2.30956113	$4.00x10^{-07}$	0.00170827
secondary alcohol biosynthetic process	21	6.31431151	$6.66x10^{-07}$	0.00284688
telomere maintenance via semi-conservative replication	18	7.72948656	$1.06x10^{-06}$	0.00454503
DNA replication-independent nucleosome organization	29	3.63637663	$3.64x10^{-06}$	0.01555356
aerobic respiration	26	3.91052716	$5.18x10^{-06}$	0.02214193

B

MP Term (up-regulated genes)	$N^{overlap}$	OR	P	$P_{corrected}$
abnormal muscle fiber morphology	102	2.63102943	$3.36x10^{-13}$	$9.90x10^{-10}$
abnormal glomerular mesangium morphology	32	5.80896863	$1.91x10^{-10}$	$5.63x10^{-07}$
insulin resistance	49	3.05626311	$1.23x10^{-08}$	$3.64x10^{-05}$
ventricular septal defect	71	2.63694896	$9.80x10^{-10}$	$2.89x10^{-06}$
decreased skeletal muscle size	28	5.29720974	$8.86x10^{-09}$	$2.61x10^{-05}$
abnormal wound healing	52	3.29417662	$7.27x10^{-10}$	$2.14x10^{-06}$
postnatal lethality, incomplete penetrance	154	1.87686939	$5.74x10^{-10}$	$1.69x10^{-06}$
abnormal long bone morphology	132	2.00295798	$3.75x10^{-10}$	$1.11x10^{-06}$
abnormal kidney cell proliferation	14	6.74423103	$9.82x10^{-06}$	0.02895773
abnormal smooth muscle morphology	32	2.96112615	$6.08x10^{-06}$	0.01794939
abnormal liver lobule morphology	54	2.30988518	$1.95x10^{-06}$	0.00574573
limb grasping	59	2.2799958	$9.40x10^{-07}$	0.00277263
abnormal lens morphology	77	2.14745195	$1.60x10^{-07}$	0.00047098
abnormal fibroblast migration	22	5.97420251	$9.71x10^{-08}$	0.00028655
altered tumor pathology	48	2.9094555	$5.15x10^{-08}$	0.00015195
abnormal visceral yolk sac morphology	60	2.5716913	$3.48x10^{-08}$	0.00010267

MP Term (down-regulated genes)	$N^{overlap}$	OR	P	$P_{corrected}$
embryonic lethality prior to organogenesis	269	2.20542918	$4.95x10^{-21}$	$1.46x10^{-17}$
preweaning lethality, complete penetrance	268	1.76533653	$1.51x10^{-12}$	$4.47x10^{-09}$
embryonic lethality before implantation	78	3.1132374	$6.20x10^{-12}$	$1.83x10^{-08}$
embryonic lethality between implantation and somite formation, complete penetrance	94	1.99122192	$4.27x10^{-07}$	0.0012586
abnormal inner cell mass morphology	39	3.17690308	$7.40x10^{-07}$	0.00218369
abnormal mitochondrial physiology	53	2.42287454	$2.67x10^{-06}$	0.00788774

C

GO Term (up-regulated genes)	$N^{overlap}$	OR	P	$P_{corrected}$
regulation of cell proliferation	241	1.61673901	$1.69x10^{-09}$	$7.22x10^{-06}$
positive regulation of cell proliferation	128	1.6051496	$8.99x10^{-06}$	0.03844106

GO Term (down-regulated genes)	$N^{overlap}$	OR	P	$P_{corrected}$
negative regulation of cell cycle G2/M phase transition	64	6.69706941	$6.99x10^{-19}$	$2.99x10^{-15}$
negative regulation of cell cycle process	124	2.64671794	$1.85x10^{-14}$	$7.89x10^{-11}$
negative regulation of cell cycle phase transition	95	3.10760699	$3.81x10^{-14}$	$1.63x10^{-10}$
negative regulation of cell cycle	175	1.86961758	$3.35x10^{-10}$	$1.43x10^{-06}$

Figure 5.8. Day 20 GO and MP functional enrichment

(A+B) Fisher's exact test for GO term (A) and MP term (B) functional enrichment in sets of up-regulated and down-regulated protein coding genes significantly differentially expressed (Bonferroni $p < 0.05$) between *DLG2*$^{-/-}$ and WT cultures at day 20 of cortical differentiation from hESC. Enrichment resulting from unique sets of differentially expressed genes was achieved following recursive refinement by odds ratio (OR). **(C)** Selected results from GO term functional enrichment analyses (as in A) prior to refinement by odds ratio. For all analyses all protein coding genes expressed at all experimental timepoints were included in the analysis as a comparator gene set, $N^{overlap}$ = number of unique differentially expressed genes captured by term.

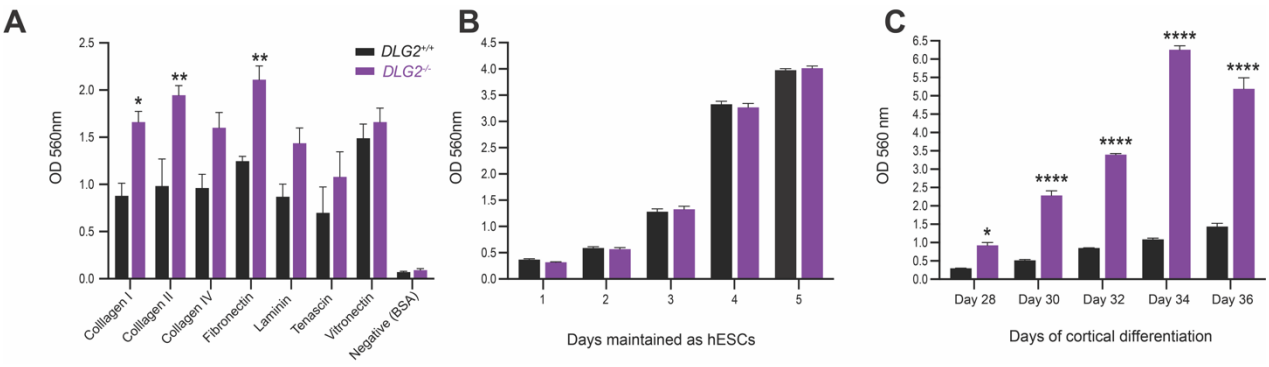


Figure 5.9. Adhesion and proliferation assays

(A) 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. **(B)** Proliferation of *DLG2*$^{-/-}$ and WT lines maintained as hESCs for 5 days. While time ($F_{4,35}$=1924, P<0.0001; n≥3) did have a significant effect on hESC proliferation, genotype ($F_{1,35}$=0.3780, P=0.8470; n≥3) did not. **(C)** 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). Data sets were analysed by two-way ANOVA. Stars above bars in each graph represents Bonferroni-corrected post hoc tests, *P<0.05; **P<0.01; ***P<0.001; ****P<0.0001 vs. WT controls. From (Sanders et al., 2020).

5.6. *DLG2*$^{-/-}$ neurons display altered morphology, migration and electrical properties

Functional enrichment results for genes differentially expressed at day 30 (Figure 5.10), a time point corresponding to the emergence of postmitotic neurons in culture, further characterised the *DLG2*$^{-/-}$ phenotype and highlighted several neuron specific biological processes potentially regulated by DLG2 dependent signalling that were further investigated through experimental assays including neuronal morphology, locomotion and electrical properties. Once again to verify that the results for the two *DLG2*$^{-/-}$ cells lines were broadly similar, functional enrichment analyses were also performed for both lines separately which showed that terms relating to these 3 processes were over-represented in both KO1 and KO2 at day 30 (Supplementary Tables 3b, 3c).

Genes down regulated at day 30 (Figure 5.10 A) were enriched for GO terms related to *nervous system development* generally but more specifically several terms related to neuronal morphology including *regulation of neuron projection development, axonogenesis*, *dendrite morphogenesis* and others. These results at the level of the transcriptome clearly indicated defects in the morphology of *DLG2*$^{-/-}$ neurons, in agreement with the observations from morphometric assays reported in a Chapter 4 (Section 4.7) that *DLG2*$^{-/-}$ neurons display a less mature structure than WT characterised by a similar number of primary neurites projecting from the soma but significantly reduced neurite branching (Figure 4.10).

Another key aspect of nervous system development dysregulated in *DLG2*$^{-/-}$ neurons was neuronal migration, as indicated by functional enrichment for GO terms in the D30 down-regulated gene set. Results from earlier time points had indicated a role for DLG2 in regulating cell migration with regulation of *mononuclear cell migration* enriched at day 15 (Figure 5.7 B) and *cell leading edge* (the area of a migrating cell nearest the direction of movement) enriched at day 20 (Figure 5.8 A). Although not present in the refined enrichment results at day 30 (Figure 5.10 A) the unrefined results show clear enrichment in down-regulated genes for *neuronal migration* and *cell leading edge* GO terms. To investigate this phenotype *DLG2*$^{-/-}$ and WT neurons were tracked for 70 hours from day 40 of differentiation with the location of individual neurons recorded at 2-hour intervals. These results validated the RNAseq analyses and showed that *DLG2*$^{-/-}$ neurons had a significantly slower speed of migration and displacement from their origin over the course of the assay (Figure 5.11).

A

GO Term (up-regulated genes)	$N^{overlap}$	OR	P	$P_{corrected}$
mitochondrion	610	3.18315405	$4.32x10^{-77}$	$1.85x10^{-73}$
structural constituent of ribosome	115	61.9063948	$6.78x10^{-58}$	$2.90x10^{-54}$
translational initiation	102	17.0959496	$6.19x10^{-41}$	$2.65x10^{-37}$
SRP-dependent cotranslational protein targeting to membrane	83	44.3526819	$1.83x10^{-40}$	$7.80x10^{-37}$
mitochondrial matrix	225	3.93213319	$1.08x10^{-38}$	$4.63x10^{-35}$
rRNA processing	150	5.39497005	$1.32x10^{-34}$	$5.66x10^{-31}$
mitochondrial translational elongation	77	22.8217784	$1.50x10^{-33}$	$6.41x10^{-30}$
mitochondrial respiratory chain complex assembly	74	16.4372974	$9.61x10^{-30}$	$4.11x10^{-26}$
ATP synthesis coupled electron transport	63	18.6144149	$2.19x10^{-26}$	$9.38x10^{-23}$
NADH dehydrogenase complex	41	54.2639352	$5.04x10^{-21}$	$2.15x10^{-17}$
endopeptidase complex	41	12.051786	$1.58x10^{-15}$	$6.77x10^{-12}$
cofactor metabolic process	157	2.23451887	$5.35x10^{-13}$	$2.29x10^{-09}$
ribosomal large subunit biogenesis	44	5.81891405	$8.30x10^{-12}$	$3.55x10^{-08}$
cristae formation	23	30.3176109	$2.15x10^{-11}$	$9.21x10^{-08}$
ficolin-1-rich granule lumen	63	3.33848169	$2.59x10^{-10}$	$1.11x10^{-06}$
aerobic respiration	33	6.70045704	$5.93x10^{-10}$	$2.53x10^{-06}$
transcription-coupled nucleotide-excision repair	43	4.20889856	$3.21x10^{-09}$	$1.37x10^{-05}$
mitochondrial intermembrane space	36	5.00289783	$5.23x10^{-09}$	$2.24x10^{-05}$
protein targeting to mitochondrion	33	5.4435219	$7.73x10^{-09}$	$3.31x10^{-05}$
chaperone-mediated protein folding	30	4.94533158	$1.13x10^{-07}$	0.00048134
azurophil granule lumen	42	3.36101196	$2.05x10^{-07}$	0.00087859
spliceosomal tri-snRNP complex	21	7.90177265	$2.47x10^{-07}$	0.00105716
endoplasmic reticulum	418	1.3546295	$1.08x10^{-06}$	0.0046374
cofactor biosynthetic process	67	2.27269214	$1.21x10^{-06}$	0.00516432
monosaccharide catabolic process	28	4.34176291	$1.24x10^{-06}$	0.00530246
peptidyl-prolyl cis-trans isomerase activity	21	6.14414725	$1.63x10^{-06}$	0.00697536
antioxidant activity	32	3.37486242	$5.30x10^{-06}$	0.02266803
organelle disassembly	32	3.37486242	$5.30x10^{-06}$	0.02266803
fatty acid oxidation	31	3.26878962	$1.10x10^{-05}$	0.04682408

GO Term (down-regulated genes)	$N^{overlap}$	OR	P	$P_{corrected}$
nervous system development	580	2.43401735	$3.10x10^{-48}$	$1.32x10^{-44}$
regulation of cell morphogenesis	144	2.95695095	$2.43x10^{-18}$	$1.04x10^{-14}$
positive regulation of neuron projection development	79	4.17123438	$2.05x10^{-15}$	$8.79x10^{-12}$
axonogenesis	111	2.9092436	$3.02x10^{-14}$	$1.29x10^{-10}$
protein serine/threonine kinase activity	168	2.15578619	$8.64x10^{-13}$	$3.69x10^{-09}$
DNA binding transcription factor activity	330	1.62956451	$2.36x10^{-11}$	$1.01x10^{-07}$
synapse assembly	40	5.71471442	$1.23x10^{-10}$	$5.24x10^{-07}$
regulation of axonogenesis	47	4.17038605	$8.84x10^{-10}$	$3.78x10^{-06}$
Rho guanyl-nucleotide exchange factor activity	36	5.1387089	$4.39x10^{-09}$	$1.88x10^{-05}$
covalent chromatin modification	113	2.13415206	$5.30x10^{-09}$	$2.27x10^{-05}$
nuclear body	225	1.66499178	$6.60x10^{-09}$	$2.82x10^{-05}$
peptidyl-serine phosphorylation	71	2.54174381	$3.88x10^{-08}$	0.00016581
ubiquitin-like protein ligase activity	86	2.28800894	$4.01x10^{-08}$	0.00017148
regulation of synaptic plasticity	42	3.59849873	$7.72x10^{-08}$	0.00033032
negative regulation of nervous system development	59	2.76163418	$8.30x10^{-08}$	0.00035503
glutamate receptor activity	19	9.7370702	$3.04x10^{-07}$	0.0013005
regulation of dendritic spine development	25	5.82865684	$3.14x10^{-07}$	0.00134224
transcriptional repressor complex	37	3.65486352	$3.60x10^{-07}$	0.00153734
axon part	33	4.03385107	$4.25x10^{-07}$	0.0018161
nuclear periphery	41	3.19167159	$7.47x10^{-07}$	0.00319228
modification-dependent protein binding	56	2.52712853	$1.11x10^{-06}$	0.00472936
beta-catenin binding	36	3.4230988	$1.29x10^{-06}$	0.005537
microtubule end	17	8.70825978	$2.42x10^{-06}$	0.01034053
voltage-gated sodium channel activity	17	8.70825978	$2.42x10^{-06}$	0.01034053
central nervous system neuron development	19	6.95357089	$2.54x10^{-06}$	0.01086685
Rab GTPase binding	69	2.16611434	$2.98x10^{-06}$	0.01275856
Rac GTPase binding	30	3.66484703	$4.36x10^{-06}$	0.01865421
dendrite morphogenesis	22	5.12585691	$4.69x10^{-06}$	0.02005671

B

MP Term (up-regulated genes)	$N^{overlap}$	OR	P	$P_{corrected}$
abnormal redox activity	51	2.93070651	$1.66x10^{-07}$	0.00048909
abnormal cellular respiration	48	2.75644305	$1.17x10^{-06}$	0.00343835

MP Term (down-regulated genes)	$N^{overlap}$	OR	P	$P_{corrected}$
lethality during fetal growth through weaning	673	2.00251967	$1.41x10^{-36}$	$4.16x10^{-33}$
abnormal CNS synaptic transmission	269	3.01733845	$2.35x10^{-33}$	$6.93x10^{-30}$
perinatal lethality	393	2.02605933	$1.46x10^{-23}$	$4.31x10^{-20}$
abnormal miniature excitatory postsynaptic currents	52	5.15349529	$1.62x10^{-12}$	$4.78x10^{-09}$
abnormal hippocampus pyramidal cell layer	41	7.0317964	$3.20x10^{-12}$	$9.44x10^{-09}$
limb grasping	90	2.83861342	$1.68x10^{-11}$	$4.95x10^{-08}$
abnormal contextual conditioning behavior	67	3.14155184	$4.31x10^{-10}$	$1.27x10^{-06}$
perinatal lethality, complete penetrance	100	2.24762023	$6.42x10^{-09}$	$1.89x10^{-05}$
embryonic growth retardation	171	1.77554192	$1.66x10^{-08}$	$4.91x10^{-05}$
abnormal sternum morphology	72	2.50801003	$4.45x10^{-08}$	0.00013121
small olfactory bulb	27	5.77243772	$1.16x10^{-07}$	0.00034216
abnormal channel response	20	8.54243133	$3.38x10^{-07}$	0.00099622
abnormal parietal lobe morphology	21	6.72716227	$9.46x10^{-07}$	0.00278984
abnormal embryonic/fetal subventricular zone morphology	21	5.97888905	$2.31x10^{-06}$	0.00682513
absence seizures	21	5.97888905	$2.31x10^{-06}$	0.00682513
abnormal discrimination learning	18	7.68466934	$2.54x10^{-06}$	0.00748425
decreased fetal size	74	2.02671061	$6.99x10^{-06}$	0.02062479
pericardial effusion	44	2.62821621	$7.78x10^{-06}$	0.02294458
abnormal synaptic vesicle number	21	4.89179774	$1.12x10^{-05}$	0.03318681

C

GO Term (down-regulated genes)	$N^{overlap}$	OR	P	$P_{corrected}$
membrane depolarization	31	3.97742635	$1.12x10^{-06}$	0.00479914
voltage-gated channel activity	62	2.38144269	$1.16x10^{-06}$	0.00496851
voltage-gated ion channel activity	62	2.38144269	$1.16x10^{-06}$	0.00496851
action potential	43	2.98584961	$1.23x10^{-06}$	0.00524208
neuron migration	26	4.44480016	$2.65x10^{-06}$	0.01132792
membrane depolarization during action potential	25	4.57856187	$3.10x10^{-06}$	0.01323508
cell leading edge	93	1.88791853	$4.29x10^{-06}$	0.01832325

Figure 5.10. Day 30 GO and MP functional enrichment
(A+B) Fisher's exact test for GO term (A) and MP term (B) functional enrichment in sets of up-regulated and down-regulated protein coding genes significantly differentially expressed (Bonferroni $p < 0.05$) between *DLG2*$^{-/-}$ and WT cultures at day 30 of cortical differentiation from hESCs. Enrichment resulting from unique sets of differentially expressed genes was achieved following recursive refinement by odds ratio (OR). **(C)** Selected results from GO term functional enrichment analyses (as in A) prior to refinement by odds ratio. For all analyses all protein coding genes expressed at all experimental timepoints were included in the analysis as a comparator gene set, $N^{overlap}$ = number of unique differentially expressed genes captured by term.

In addition to defects in morphology and migration functional enrichment results also indicated *DLG2*$^{-/-}$ neurons to have abnormal electrical properties. The day 30 down-regulated gene set was enriched for the GO term *voltage-gated sodium channel activity* following odds ratio refinement (Figure 5.10 A) while prior to refinement numerous GO terms relating to the electrical properties of neurons were significantly enriched including *membrane depolarization*, *voltage-gate ion channel activity* and *action potential* (Figure 5.10 C). To investigate this further, various electrophysiological studies were performed on *DLG2*$^{-/-}$ and WT neurons at days 50 and 60 of cortical differentiation by Dr Daniel Whitcombe and Tom Steward (Bristol Medical School, University of Bristol) as well as Dr Ying Zhu (Neuroscience and Mental Health Research Institute, Cardiff University), the results of which are included here to provide a more complete picture of the *DLG2*$^{-/-}$ phenotype. Full results are available in (Sanders et al., 2020) but in summary these

show day 50 *DLG2*$^{-/-}$ neurons to be less excitable than WT, they possess a significantly more depolarised resting membrane potential than WT (Figure 5.12 A) and a reduced ability to fire APs following current injection (Figure 5.12C). APs that were generated by *DLG2*$^{-/-}$ neurons showed characteristics of less mature neurons including a smaller amplitude, longer half-width and a slower maximum rate of depolarisation and repolarisation ($\delta V/\delta t$) (Figure 5.12 E-H). Other electrophysiological properties were unchanged between *DLG2*$^{-/-}$ and WT neurons which includes the percentage of neurons displaying spontaneous excitatory postsynaptic currents (sEPSCs) along with their frequency and amplitude (Figure 5.12 I-M). It was hypothesised that this lack of effect on synaptic transmission may be due to a compensatory mechanism by DLG4, which shows a trend towards increased expression in synaptosomes from *DLG2*$^{-/-}$ neurons (Figure 5.12 O). As *DLG2*$^{-/-}$ neurons have a reduced ability to fire APs and produce less mature APs, these experimental results therefore validate the functional enrichment analyses.

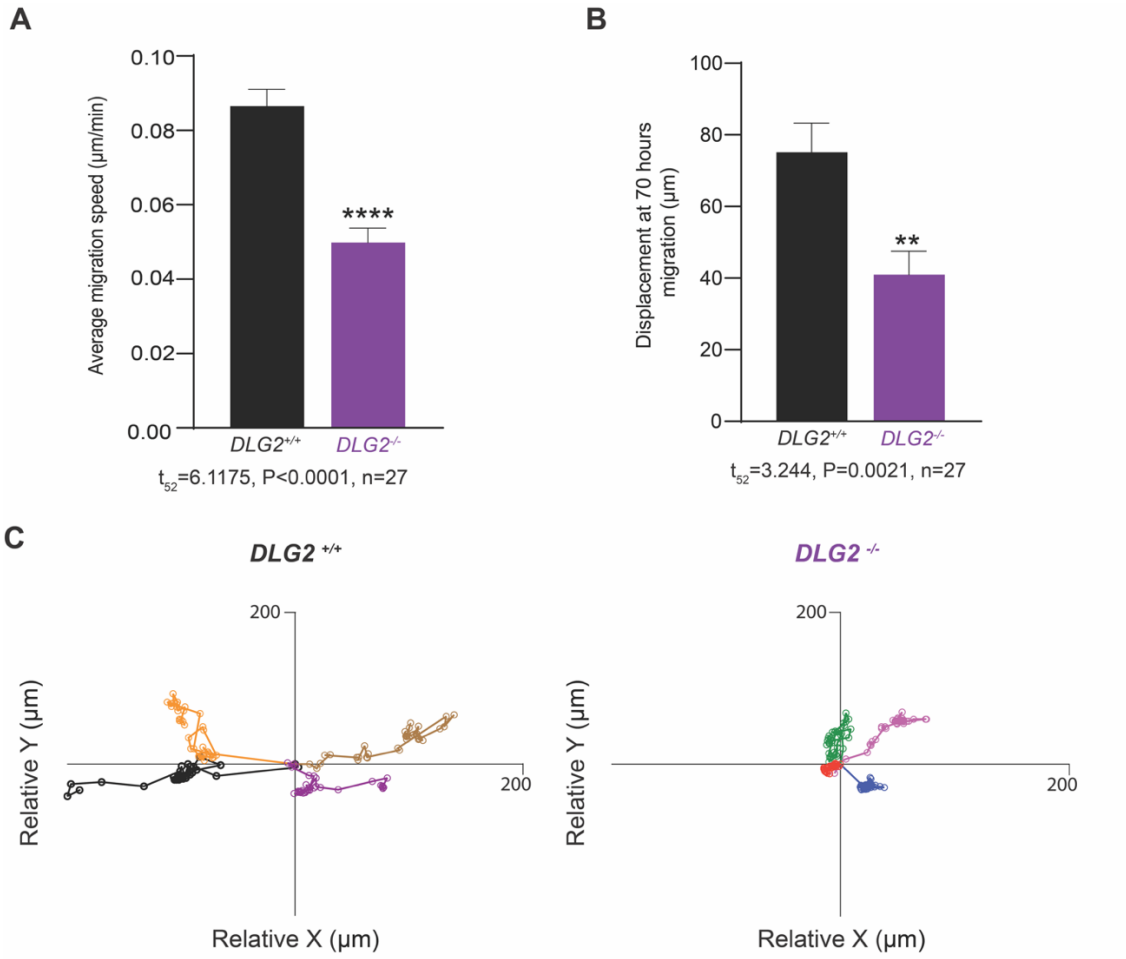


Figure 5.11. Migration assay

(A) 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). **(B)** 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). **(C)** Representative traces of neuronal migration from a given origin over 70 hours, from day 40 of cortical differentiation. Data analysed by unpaired two-tailed Student's t-test. Stars above bars in each graph represents Bonferroni-corrected post hoc tests, *P<0.05; **P<0.01; ***P<0.001; ****P<0.0001 vs. WT controls From (Sanders et al., 2020).

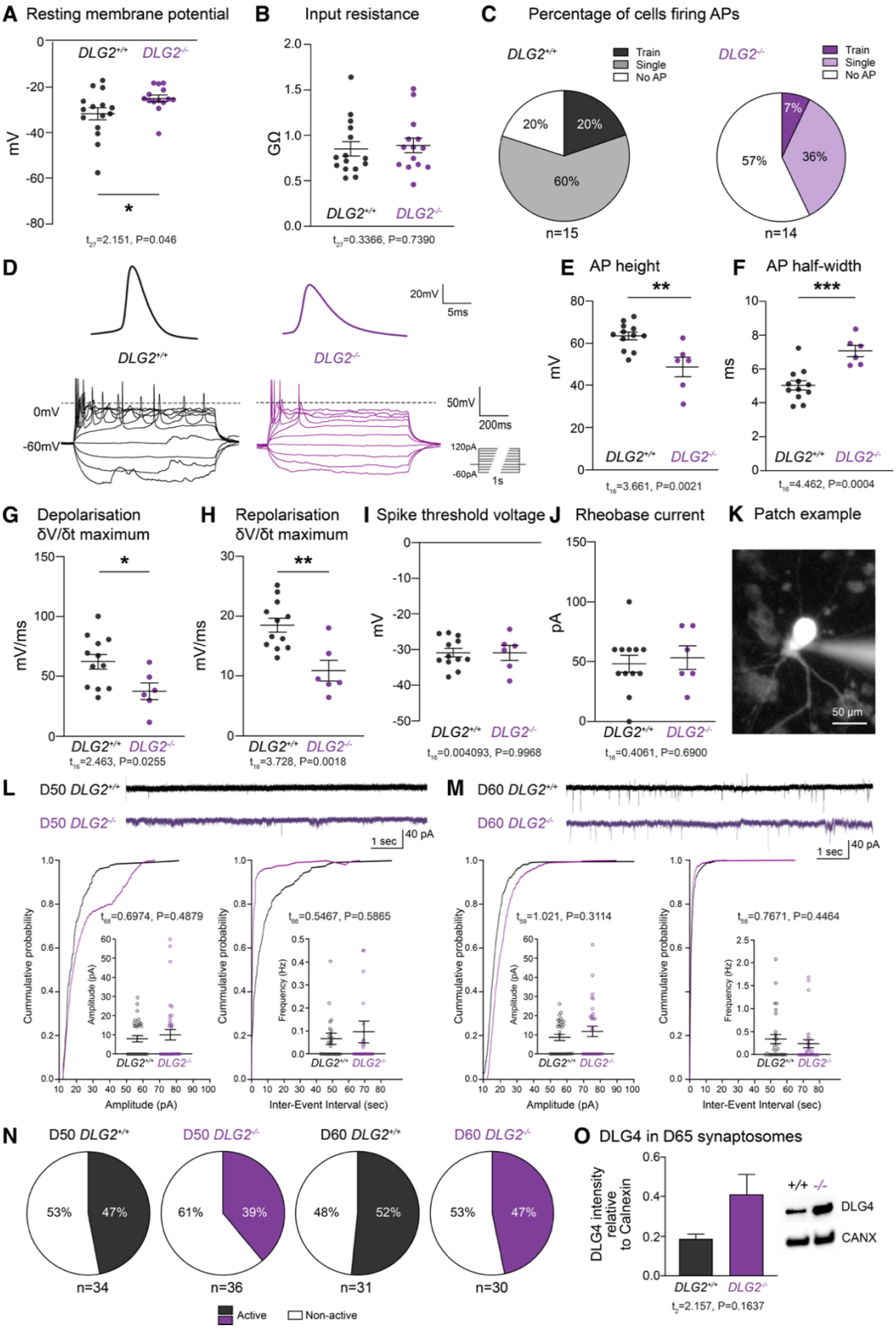


Figure 5.12. Electrophysiological properties of *DLG2*$^{-/-}$ neurons
(A) Resting membrane potential **(B)** input resistance (day 50) **(C)** %cells firing APs on current step injection. **(D)** Example traces of first overshooting AP and AP trains **(E-H)** AP height, half-width, maximum depolarising and repolarising speed **(I-J)** spike threshold voltage & rheobase current (day 50) **(K)** Example whole-cell patch clamp. Amplitude and frequency of spontaneous EPSCs, days **(L)** 50 and **(M)** 60. **(N)** %cells displaying sEPSCs **(O)** DLG4 western blot. From (Sanders et al., 2020).

5.7. Aspects of the *DLG2*$^{-/-}$ phenotype show recovery at later time points

Both differential gene expression analyses as well as the subsequent functional enrichment for GO and MP terms suggests the *DLG2*$^{-/-}$ phenotype show a degree of recovery over the course of cortical differentiation. When considering differential gene expression as a whole (Figure 5.6) this did not show a continual increase over the course of the differentiation or a plateau at later time points, rather it peaked at day 30 of cortical differentiation where 47% of protein coding genes were differentially expressing in *DLG2*$^{-/-}$ cells.; however, at day 60 differential gene expression was only 21.5%, below the levels found at both D30 and D20. This recovery in expression at later experimental time points includes genes such as *BCL11B* encoding the key developmental transcription factor and cortical layer V marker CTIP2 (Supplementary Table 2a). At the mRNA level *BCL11B* expression is significantly reduced in *DLG2*$^{-/-}$ cells at early experimental time points but recovers fully to WT by day 60. This recovery in *BCL11B* expression is in at least partial agreement with the CTIP2 expression data presented in Chapter 4; in this case CTIP2 expression is significantly reduced at all experimental time points but shows significant recovery towards WT levels as differentiation progresses. Presumably the delayed recovery of CTIP2 expression compared to *BCL11B* is due to the time taken for changes in the transcriptome to be reflected at the protein level within cells.

The results of the functional enrichment analyses also support a recovery for certain *DLG2*$^{-/-}$ phenotypes over time. Genes relating to various aspects of neuronal development such as morphology, migration and electrical properties, are downregulated in *DGL2*$^{-/-}$ cells at day 30 of differentiation (Figure 5.10 A) but by day 60 of this is no longer the case (Figure 5.13 A), suggesting recovery for these phenotypes at the mRNA level. Genes significantly down-regulated in *DLG2*$^{-/-}$ cells at day 20 are also enriched for various GO terms related to mitochondria, aerobic respiration and translation (Figure 5.8 A) while 10 days later similar terms are found to be enriched in the day 30 up-regulated gene set (Figure 5.10 A), suggesting a delay in the expression of genes regulating these processes in *DLG2*$^{-/-}$ cells; however, the recovery may not be complete in this case as by day 60 several mitochondria related GO terms are once again enriched in the down-regulated gene set (Figure 5.13 A). Finally, another phenotype that shows evidence of incomplete recovery is cell cycle abnormalities, which are present at early time points such as day 20 (Figure 5.8 A), appear to show recovery by day 30 (Figure 5.10A) but are once again present at day 60 (Figure 5.13 A).

A

GO Term (up-regulated genes)	$N^{overlap}$	OR	P	$P_{corrected}$
sterol transport	15	7.33638048	$2.07x10^{-07}$	0.000886568
Golgi apparatus	147	1.57279407	$2.08x10^{-06}$	0.008914141
canonical glycolysis	11	7.28539808	$9.26x10^{-06}$	0.039586378
oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors	14	5.19941248	$9.96x10^{-06}$	0.042582708

GO Term (down-regulated genes)	$N^{overlap}$	OR	P	$P_{corrected}$
DNA replication	91	7.06627202	$2.24x10^{-32}$	$9.60x10^{-29}$
RNA binding	364	2.06350587	$2.73x10^{-26}$	$1.17x10^{-22}$
DNA repair	143	3.31457266	$2.25x10^{-25}$	$9.64x10^{-22}$
DNA-dependent DNA replication	60	9.09881152	$3.37x10^{-25}$	$1.44x10^{-21}$
mitochondrion	289	1.86112375	$7.01x10^{-17}$	$3.00x10^{-13}$
G1/S transition of mitotic cell cycle	51	5.54187608	$4.12x10^{-16}$	$1.76x10^{-12}$
sister chromatid cohesion	50	4.37619591	$3.81x10^{-13}$	$1.63x10^{-09}$
RNA splicing, via transesterification reactions	88	2.62619857	$4.56x10^{-12}$	$1.95x10^{-08}$
DNA replication initiation	19	20.443704	$1.02x10^{-11}$	$4.36x10^{-08}$
DNA strand elongation	18	19.3585027	$5.17x10^{-11}$	$2.21x10^{-07}$
preribosome	34	4.96369704	$2.12x10^{-10}$	$9.05x10^{-07}$
negative regulation of cell cycle G2/M phase transition	40	4.08173233	$3.40x10^{-10}$	$1.45x10^{-06}$
catalytic activity, acting on a tRNA	42	3.85133867	$4.44x10^{-10}$	$1.90x10^{-06}$
nucleobase-containing small molecule metabolic process	121	1.98913167	$1.23x10^{-09}$	$5.28x10^{-06}$
cellular respiration	48	3.29029034	$1.28x10^{-09}$	$5.48x10^{-06}$
mitochondrial RNA metabolic process	19	7.8619509	$2.31x10^{-08}$	$9.87x10^{-05}$
DNA-templated transcription, termination	39	3.1914871	$7.68x10^{-08}$	0.00032826
tRNA export from nucleus	18	6.45092766	$3.22x10^{-07}$	0.00137592
snoRNA binding	16	7.1634933	$6.29x10^{-07}$	0.00268768
RNA polymerase activity	18	5.69128632	$1.02x10^{-06}$	0.00437679
site of double-strand break	17	6.09018386	$1.13x10^{-06}$	0.00482829
maturation of 5.8S rRNA	16	6.61203895	$1.20x10^{-06}$	0.00511732
strand displacement	15	7.32391548	$1.20x10^{-06}$	0.00513856
DNA replication-independent nucleosome organization	23	4.12533266	$1.57x10^{-06}$	0.00672838
intraciliary transport involved in cilium assembly	19	4.86389354	$2.43x10^{-06}$	0.01038767
cellular protein complex disassembly	40	2.60067819	$2.86x10^{-06}$	0.0122372
mitochondrial electron transport, NADH to ubiquinone	18	5.09159832	$2.90x10^{-06}$	0.01239912
ligase activity, forming carbon-oxygen bonds	18	4.83673967	$4.70x10^{-06}$	0.02009508
condensed chromosome kinetochore	14	6.26320911	$8.26x10^{-06}$	0.03530447
proteasome regulatory particle	12	8.04990406	$8.32x10^{-06}$	0.03557425

B

MP Term (up-regulated genes)	$N^{overlap}$	OR	P	$P_{corrected}$
abnormal muscle physiology	100	1.74515531	$1.74x10^{-06}$	0.00514645
abnormal myocardium layer morphology	57	2.12602963	$1.89x10^{-06}$	0.00557848
abnormal cutaneous collagen fibril morphology	11	9.27486164	$1.93x10^{-06}$	0.00570555

MP Term (down-regulated genes)	$N^{overlap}$	OR	P	$P_{corrected}$
embryonic lethality prior to organogenesis	190	2.3155312	$4.54x10^{-19}$	$1.34x10^{-15}$
chromosomal instability	38	4.46652042	$1.55x10^{-10}$	$4.56x10^{-07}$
abnormal cell cycle	70	2.55350396	$1.49x10^{-09}$	$4.40x10^{-06}$
preweaning lethality, complete penetrance	172	1.67562687	$2.33x10^{-08}$	$6.88x10^{-05}$
abnormal DNA repair	33	3.49022403	$1.64x10^{-07}$	0.00048295

Figure 5.13. Day 60 GO and MP functional enrichment

(A+B) Fisher's exact test for GO term (A) and MP term (B) functional enrichment in sets of up-regulated and down-regulated protein coding genes significantly differentially expressed (Bonferroni $p < 0.05$) between *DLG2*$^{-/-}$ and WT cultures at day 30 of cortical differentiation from hESCs. Enrichment resulting from unique sets of differentially expressed genes was achieved following recursive refinement by odds ratio (OR). All protein coding genes expressed at all experimental timepoints were included in the analysis as a comparator gene set, $N^{overlap}$ = number of unique differentially expressed genes captured by term.

5.8. Discussion

5.8.1. RNAseq supports a novel role for *DLG2 in cortical neurogenesis*

To further investigate the currently unreported role for DLG2 in early cortical neurogenesis identified from ICC and western blotting analyses (see Chapter 4), RNAseq was performed on *DLG2*$^{-/-}$ and WT cells at various time points of cortical differentiation. These sequencing data supported previous analyses and provided further evidence of a role for *DLG2* in early cortical neurogenesis as *DLG2*$^{-/-}$ cells displayed large-scale disruption to gene expression compared to WT at all time points investigated (Figure 5.5), including early developmental stages where NSCs, NPCs and immature neurons are the predominant cell types in culture (days 15, 20 & 30 respectively). Although transcriptional changes are observed at all time points investigated, of particular interest is that the peak of differential gene expression, 47% of protein coding genes, did not occur at day 60, where mature neurons are present in the culture and DLG2 is known to function within the PSD, but rather at day 30 when the cell types present in culture are a mix of NPCs and new-born immature neurons.

A developmental role for a DLG protein is itself not novel. The single orthologue of human DLG1-4, invertebrate Dlg, is known to have such a role and regulates diverse cellular processes relating to apical-basal polarity, proliferation and differentiation as a component of the Scribble signalling module (Elsum et al., 2012; Stephens et al., 2018). In humans this developmental role for Dlg is believed to be taken by DLG1 (also called human DLG or hDlg) interacting with SCRIB (Humbert et al., 2008; Massimi et al., 2008; Elsum et al., 2012; Stephens et al., 2018) with the other homologues assuming the role of scaffolding proteins. DLG1 and DLG2 are of course members of the same MAGUK subfamily and as such have a very similar domain structure, DLG1 possessing an L27 domain being the notable difference (Zheng et al., 2011; Zhu et al., 2016). This structural similarity combined with the data presented here suggests that both DLG1 and DLG2 have inherited aspects of the developmental regulation functions ascribed to invertebrate Dlg.

5.8.2. DLG2 mutations act at early stages of cortical differentiation in NSCs/NPCs

ICC and western blotting analyses identified significant differences in the protein expression and morphology of *DLG2*$^{-/-}$ neurons from day 30 of cortical differentiation although the phenotype of cells assayed prior to this at day 20 was less clear (see Chapter 4). RNAseq was selected as a suitable approach to determine if a *DLG2*$^{-/-}$ phenotype was restricted to postmitotic neurons or existed in earlier cell types. Initial RNAseq analysis did show large numbers of differentially expressed genes in *DLG2*$^{-/-}$ cells at earlier time points of cortical differentiation and prior to neurogenesis, while functional enrichment analysis for GO and MP terms showed numerous biological processes and

phenotypes were either under or over-represented the cells, indicating that there are indeed *DLG2*$^{-/-}$ phenotypes in NSCs and NPCs in addition to those established for postmitotic neurons.

The functional enrichment analyses for the day 15 and 20 gene sets, corresponding to NSC and NPC time points, indicated a dysregulation of several developmentally relevant biological processes (Figures 5.7 & 5.8). This includes two processes that were experimentally validated, adhesion to the ECM and cell cycle progression, with *DLG2*$^{-/-}$ cells showing both increased adhesion to various ECM substrates and proliferation compared to WT controls (Figure 5.9), in agreement with the increased trend in KI67 expression identified from western blots (Figure 4.7). Cell-ECM adhesion play a critical role in normal development (Rozario and DeSimone, 2010) including cortical differentiation, during which various transcriptional programmes inhibit cell adhesion to the ECM, triggering the delamination and subsequent migration of cells (Itoh et al., 2013). Collagen is the most abundant protein within the ECM where it forms a cross-linked scaffold with itself, other ECM components and cell surface receptors (Frantz et al., 2010). Interestingly the gene encoding the pro-α1 chain of collagen I (*COL1A1*) is one of the most significantly upregulated genes in *DLG2*$^{-/-}$ cells at day 15 of cortical differentiation (Figure 5.7 A), this may lead to greater collagen secretion into the ECM leading to a denser scaffold and increased adhesion. The regulation of cell adhesion has previously been associated with MAGUK proteins, although cell-cell rather than cell-ECM adhesion, where these scaffolding proteins are known to stabilise tight junctions (Funke et al., 2005). These proteins are also known to be targeted for degradation by proteases to weaken cell adhesion prior to detachment (Ivanova et al., 2011); it is therefore possible that cell-ECM contacts which form in the absence of the normal scaffolding proteins, such as DLG2, cannot be efficiently degraded by proteases inhibiting cell detachment from the ECM.

The increased proliferation observed in *DLG2*$^{-/-}$ cultures during early cortical differentiation may also relate to the delayed expression of cell-type identity observed in *DLG2*$^{-/-}$ neuronal cultures, including the significant reduction in expression of cortical layer V maker CTIP2 that shows recovery towards WT levels at later time points (see Chapter 4). Cell cycle regulation in NPCs plays a key role in determining neuronal subtype identity, with their internal state changing over the course of multiple cell divisions and the precise state immediately prior to cell cycle exit determining the final identity of generated neurons, (Telley et al., 2019). NPCs are intrinsically capable of producing postmitotic neurons, the initial identity of which appears to be independent of environmental signals; however transcriptional programmes activated in NPCs in response to environmental signals are believed to have roles in subsequent neuronal maturation and final cell type identity (Telley et al., 2019). Initially NPCs exist in an "introverted" state and are not responsive to such external signals with the switch to an environment sensing "extroverted" state being under cell cycle control, once in this receptive state it is the length of the cell cycle that

appears key for the ability of cells to sense and respond to its environment, with the G_1 phase in particular appearing critical for environment sensing and fate decision (Telley et al., 2019). This raises the possibility that the increased proliferation in *DLG2*$^{-/-}$ NPCs, with implied reduction in cell-cycle length, may reduce the period these cells are receptive to external signals and hence delay the expression of transcription programmes required for the development of mature neuronal identity; a process that could be exacerbated by any disruption of signalling between cell surface receptors and the transcriptional machinery within the nucleus.

The significant upregulation of the transcription regulators *ID1, ID2 and ID3* in *DLG2*$^{-/-}$ at day 15 of cortical differentiation (Figure 5.7 A) may also be a factor underlying both of the experimentally validated NPC phenotypes, being key regulators of early neurodevelopment including NPC proliferation and adhesion to the ECM (Jung et al., 2010; Niola et al., 2012). These act by inhibiting the binding of basic helix-loop-helix transcription factors to DNA, with knockout of ID1-3 increasing cell cycle length and decreasing adhesion to the ventricular surface of the embryonic cortex *in vivo* and to ECM substrates *in vitro* (Jung et al., 2010; Niola et al., 2012).

5.8.3. Multiple *DLG2*$^{-/-}$ phenotypes indicated a delayed expression of cell-type identity

A diverse collection of cortical projection neurons subtypes exists which were initially characterised by the connection they made within the brain (hodology), although more recently this has expanded to include their gene expression profiles, morphology and electrophysiological properties (Migliore and Shepherd, 2005; Molyneaux et al., 2007). The results of ICC and WB blotting analysis presented previously showed that two such characteristics were disrupted in *DLG2*$^{-/-}$ neurons compared to WT, the expression of the cortical layer V marker CTIP2 and neurite morphology, suggesting a delay in the development of mature cell-type identity (Chapter 4). The results of the functional enrichment analysis also indicate that *DLG2*$^{-/-}$ postmitotic neurons are less mature than their WT counterparts (Figure 5.10). This includes an over-representation of differentially expressed genes relating to various aspects of neuronal morphology such as *regulation of neuron projection development, axonogenesis*, *dendrite morphogenesis* and others. This agrees with the results of a morphometric analysis noted in Chapter 4 that *DLG2*$^{-/-}$ neurons have a more immature morphology than WT with significantly reduced neurite branching (Figure 4.10). A second phenotype present in the *DLG2*$^{-/-}$ cells relates to migration with both an over-representation of differentially expressed genes related to *neuronal migration* and *cell leading edge*. This was validated by experimental assay that showed *DLG2*$^{-/-}$ cells had a slower speed of migration and migration less distance from their origin in a given period (Figure 5.11).

A third phenotype in the down-regulated genes at day 30 relates to electrical properties with an over-representation for genes in GO terms relating to voltage gated *sodium channel activity*,

membrane depolarization and *action potentia*l. Electrophysiological assays confirmed these findings and showed that *DLG2*$^{-/-}$ neurons have a reduced ability to fire APs and produce less mature APs (Figure 5.12). These electrophysiology results, along with those of migration assay previously described, also suggest that the *DLG2*$^{-/-}$ phenotype is not restricted to a specific neuronal subtype, as neurons for these experiments were selected at random rather than by expression of a particular marker. ICC analyses at days 50 and 60 (the time points analysed by electrophysiology) shows for both WT and *DLG2*$^{-/-}$ cultures a more cells are expressing the marker TBR1, which showed no difference between the cell lines, than are expressing CTIP2, which was significantly reduced in *DLG2*$^{-/-}$ cells (Figure 4.8). The likelihood that the majority of neurons analysed were TBR1 positive, yet still significant deficits were observed for *DLG2*$^{-/-}$ compared to WT, indicated that *DLG2*$^{-/-}$ phenotype is widespread among neurons and not restricted to those of a layer V identity. For future experiments selecting neurons of a specific subtype identity for electrophysiological assays would be of use to further refine the *DLG2*$^{-/-}$ phenotype, although such a selection of live neurons would be difficult with the currently generated hESC lines; however, this could be possible if the lines were transduced with suitable lentiviral reporters, such as a CTIP2 promoter-GFP vector.

There are several potential links between the phenotypes identified for *DLG2*$^{-/-}$ NPCs and neurons. Within neurons it is possible that a disruption to voltage gated sodium channel activity, as indicated by functional over-representation analyses, may not only contribute to the observed electrophysiological defects but also to those in migration and morphology; with voltage gated sodium channel activity shown to enhance the spontaneous calcium transients required for neuronal maturation in the developing neocortex, including migration and dendritic branching (Bando et al., 2016). Aberrant signalling resulting from disrupted voltage gated sodium channel activity in new-born neurons may also impact the transcriptional programmes of NPCs, if these NPCs are in an extroverted stated and receptive to external signals (Telley et al., 2019), potentially leading to abnormal neuronal identity upon cell cycle exit. There is also a potential link between the increased cell-ECM adhesion observed in NPCs and decreased migration of neurons, if the ECM is particularly dense or cell-ECM attachment cannot be efficiently severed this would impede the ability of neurons to migrate from their site of birth. One molecule that may link many of the observed phenotypes both in NPCs and neurons is CASK, known to regulate proliferation and ECM adhesion through an interaction with ID1 (Qi et al., 2005; Jung et al., 2010; Niola et al., 2012) and multiple aspects of neurogenesis through the transcription factor TBR1 (Hsueh et al., 2000; Wang et al., 2004; Huang and Hsueh, 2017). More specifically CASK regulates neuronal migration and branching through interaction with the transcription factor CTIP1 (Kuo et al., 2010), the gene encoding CTIP1 (*BCL11A)* is also significantly downregulated in *DLG2*$^{-/-}$ cells at day 30 of differentiation (Supplementary Table 2a). Interestingly CASK is also a MAGUK protein and thus

shares the same PSD tandem as the DLG subfamily members, although it does contain additional domains (Zheng et al., 2011; Zhu et al., 2016), it is also known to interact with DLG1 in mammalian epithelia (Porter et al., 2019), a similar DLG2-CASK interaction in NPCs and new-born cortical projection neurons could therefore explain many of the observed *DLG2*$^{-/-}$ phenotypes.

5.8.4. RNAseq indicates mitochondrial abnormalities in *DLG2*$^{-/-}$ NPCs and neurons

Additional evidence of a role for *DLG2* in the regulation of developmentally relevant processes can be found in the genes down-regulated in the *DLG2*$^{-/-}$ at day 20 which show a functional enrichment for numerous terms related to the mitochondria and cellular respiration (Figure 5.8). NSCs and NPCs are known to have relatively low energy requirements, to the extent that when quiescent they may be entirely glycolytic (Folmes and Terzic, 2016) and in other cases possessing immature spherical mitochondria with poorly developed cristae (Beckervordersandforth, 2017). However, the process of neuronal differentiation is associated with a large increase in cellular respiration along with mitochondrial mass, mitochondrial DNA copy number and electron transport chain capacity (Cheng et al., 2010; Wang et al., 2010). Indeed specific terms relating to *cristae formation,* the *inner mitochondrial membrane* and the *mitochondrial electron transport chain* as well as the more general term *aerobic respiration* are all over-represented in the day 20 down-regulated gene-set, suggesting *DLG2*$^{-/-}$ NPCs at this time point may have reduced energy requirements, potentially as a result of reduced *translation* as this term is also over-represented among the same gene set. Although corticoneurogenesis itself does not appear to be delayed in *DLG2*$^{-/-}$cells these processes being down-regulated in *DLG2*$^{-/-}$ NPCs relative to WT could therefore be seen as evidence of a delay in the production of cellular components required for subsequent neuronal maturation, in agreement with the immature phenotype of *DLG2*$^{-/-}$ neurons previously described. Such a delay is further supported by many of the related terms including *mitochondrion*, *aerobic respiration* and *translation* that are over-represented among down-regulated genes at day 20 being enriched among up-regulated genes at day 30 (Figure 5.8).

Although mitochondria related terms appear to show recovery between days 20 and 30 they are once again over-represented among the down-regulated gene set at day 60 (Figure 5.12). This later time point of cortical differentiation contains more mature and electrically active neurons so it is possible that reduced mitochondrial energy production is linked to the abnormal electrophysiological properties of *DLG2*$^{-/-}$ cells, such as the reduction in number and maturity of action potentials, although many of these show recovery by this time point (Figure 5.12) (Sanders et al., 2020). Alternatively, this may be a similar phenotype to that proposed at day 20 with decreased translation leading to lower energy requirements, terms relating to translation are also over-represented in the day 60 down-regulated gene set (Figure 4.12) including *tRNA* and

preribosome (Abetov et al., 2019). This decreased translation may be in NPCs, as these persist in culture to later time points, but also neurons as DLG2 does have a known role in the establishment of LTP (Carlisle et al., 2008), a process that requires both the remodelling of existing and the generation of new protein complexes (Bliim et al., 2016, 2019).

5.8.5. DLG2 mediated signalling may regulate multiple signalling pathways during early cortical differentiation

Based on the observed *DLG2*$^{-/-}$ phenotypes relating to proliferation, adhesion and various aspects of neuronal subtype identity it seems DLG2 shares many of the neurodevelopmental roles associated with its paralogue DLG1. Due to these similarities in function as well as a similar domain structure it is possible that DLG2 may interact with key developmental signalling complexes in a similar manner to DLG1, including SCRIB and CASK, as previously described. Functional enrichment analyses for GO terms have also shown that *DLG2* knockout impacts the expression of genes related to multiple signalling pathways including *regulation of transmembrane receptor protein serine/threonine kinase signalling pathway* and *regulation of BMP signalling pathway* (Figure 5.7). Receptor serine threonine kinase signalling covers a broad range of pathways with numerous developmental roles; DLG2 itself has also been shown to link cell surface receptors to components of another serine/threonine kinase pathway, the MAPK/ERK signalling cascade (Pocklington et al., 2006; Guo et al., 2012) know to regulate cell proliferation and synaptic plasticity (Pocklington et al., 2006; Cagnol and Chambard, 2010) and defects of which have been shown to result in abnormal cortical development (Li et al., 2014a; Xing et al., 2016). Another signalling pathway implicated by functional enrichment analyses is BMP signalling, itself a receptor serine/threonine kinase signalling pathway (Hegarty et al., 2013). The inhibition of BMP signalling during early development differentiates neuroectoderm from ectoderm, playing a critical role in the establishment of neural fate (Bond et al., 2012) and is a key feature of the *in vitro* cortical differentiation protocol that was adapted for use in this research (Chambers et al., 2009). In contrast once a neural fate is established BMP signalling is required for neurogenesis, with transient BMP signalling sufficient to generate postmitotic neurons while sustained BMP signalling is required for the successive generation of neuronal subtypes from NPCs (Tozer et al., 2013). This offers another potential explanation for the *DLG2*$^{-/-}$ phenotype in relation to neurogenesis and altered subtype identity, with impaired BMP signalling in *DLG2*$^{-/-}$ NPCs being sufficient to generate neurons but not of the normal cortical layer V identity. BMP signalling has also been implicated in two specific aspects of neurodevelopment that are disrupted in *DLG2*$^{-/-}$ cells, neuronal migration and neurite branching (Gámez et al., 2013).

Particularly considering the proliferation phenotype observed in *DLG2*$^{-/-}$ cells during cortical differentiation another serine/threonine kinase pathway of potential interest is the Hippo signalling pathway, which is involved in inhibiting cell proliferation and promoting apoptosis (Zhao et al., 2011). This pathway is often disrupted in cancer leading to uncontrolled cell proliferation (White et al., 2019). *DLG2* has a known role in cancer as a tumour suppressor (Shao et al., 2019; Keane et al., 2020), giving additional emphasis to the proliferation phenotype observed on *DLG2* knock-outs, with studies also showing an interaction between DLG2 protein and components of the Hippo signalling pathway such as Ras Association Domain Family Member 6 (RASSF6) (Kwan et al., 2016). In addition to regulating proliferation, Hippo signalling plays important roles in regulating cell's apical-basal polarity (Genevet and Tapon, 2011) and mitotic spindle orientation relative to the cell surface (VanHook, 2015). As studies have shown a role for the relative orientation of the mitotic spindle during cell division in NPC fate determination (Lancaster and Knoblich, 2012) this provides another potential link between *DLG2*, proliferation and neuronal identity.

5.8.6. Possible DLG2-scaffolded signalling complexes in early neurodevelopment

Data from mass spectrometry, western blotting, immunocytochemistry and RNAseq indicate both that DLG2 is expressed during early cortical neurogenesis and is potentially involved in several developmental signalling pathways, given the large-scale changes in gene expression and various phenotypes identified in *DLG2*$^{-/-}$ NPCs and neurons. As DLG2 is known to function as a scaffolding protein these data raise the possibility that it may play a similar role during early cortical differentiation, holding the components of one or more signalling pathways in the correct location and/or orientation for efficient signal transduction. If this hypothesis were correct DLG2 deficiency may not prevent the signalling pathways in which it normally acts as a scaffold from functioning, as although less optimally organised the key components are still present, but could make that signalling more stochastic and less efficient. These signalling deficits could delay rather than prevent the activation of key transcriptional programmes required for normal cortical development which would explain why hierarchical cluster analysis which shows *DLG2*$^{-/-}$ samples from later experimental time points clustering with WT samples from earlier time points (Figure 5.4 B). Such as delay may also explain why several features of the *DLG2*$^{-/-}$ phenotype show recovery at later experimental time points including overall differential gene expression, the expression of key neuronal markers, and certain electrophysiological properties. It is also of interest that the greatest expression variability between samples as determined PCA (Figure 5.4) was the day 30 *DLG2*$^{-/-}$ samples, this includes KO1 day 30 which showed good QC metrics in all other regards. As day 30 corresponds to the start of neurogenesis and is likely the stage at which cells are changing most rapidly in terms of their identity/external state (it also corresponds to the greatest differential gene

expression between WT and *DLG2*$^{-/-}$ cultures); therefore, it could be expected that the greatest effects of stochastic signalling would be observed at this time point, not only increasing differences between cell lines but also in samples from the same line.

5.8.7. Summary

The RNAseq and subsequent analyses described in this chapter identify *DLG2*$^{-/-}$ phenotypes at various developmental stages during the directed differentiation of hESCs to cortical projection neurons and similar to previous western blotting and ICC data these show a broad similarity between the two DLG2 KO lines. As differential gene expression between *DLG2*$^{-/-}$ and WT cell is shown to occur at the earliest time points investigated and prior to the formation of mature neurons this cannot be the result of errors in synaptic signalling, indicating a novel role (or roles) for *DLG2* in the regulation of early cortical neurogenesis in addition to its established function within the PSD. Functional enrichment analyses performed on sets of genes differentially expressed between *DLG2*$^{-/-}$ and WT cells identified various dysregulated developmental processes and further refined the phenotypes associated with *DLG2* knock-out. The phenotypes included increased proliferation and adhesion to the ECM in *DLG2*$^{-/-}$ NPCs as wells as abnormal morphology, migration and electrical activity in *DLG2*$^{-/-}$ neurons, all of which were validated by experimental assays. Potential mechanisms underlying the *DLG2*$^{-/-}$ phenotypes are also suggested by the analyses of RNAseq data which includes a delay in the activation of several key neurodevelopmental transcription programmes, possibly due to stochastic signalling resulting from the lack of a DLG2 scaffold in signalling complexes, although further experiments are required to better determine the precise mechanism of action for DLG2. Such delayed signalling due to a lack of a DLG2 scaffold could explain how many (but not all) of the identified *DLG2*$^{-/-}$ phenotypes show recovery at later time points in culture.

6. The *DLG2*$^{-/-}$ phenotype and neuropsychiatric disease

6.1. Introduction

RNAseq analysis of *DLG2*$^{-/-}$ and WT cells at four time points of cortical differentiation provided further support for a hypothesised role for DLG2 mediated signalling during early cortical neurogenesis, with large numbers of protein coding genes differentially expressed in *DLG2* deficient cultures at all experimental time points investigated including those predominantly composed of NPCs. Furthermore, this RNAseq analysis provided evidence for several DLG2 dependent biological process in both NPCs, such as proliferation and ECM adhesion, and new-born neurons, such as neuronal morphology, migration and electrophysiological properties, which were verified by experimental assays. Additional experiments are however required to better resolve the mechanism of action for DLG2 during neurodevelopment and more precisely determine how DLG2 mediated signalling regulates the large-scale changes in gene expression previously described.

The 'bulk' RNAseq performed for this research yielded robust and informative results yet one of its limitations is that it was performed on a mixed population of cells in culture. With an analysis of morphology, staining for key markers along with reference to published data one can determine what the predominant cell type is in a culture at a given time; however, there will inevitably be other cells types in the population. Although resulting in a shallower read depth that bulk sequencing single cell RNAseq (Mi et al., 2018) would overcome this heterogeneity limitation and identify the contribution of defined cell types such as NPCs and neurons, as well their sub-types, to each *DLG2*$^{-/-}$ phenotype. To compliment this identification of proteins that DLG2 interacts with during neurodevelopment, including the NPC stage, would also be of great benefit in understanding its mechanism of action prior to PSD formation. Such protein-protein interaction analysis would be greatly aided if a successful DLG2-FLAG knock-in could be generated (as described in Chapter 3) for use in pull-down experiment although peptide affinity purification using the NMDA receptor subunit 2 C-terminal peptide "SIESDV" (Husi and Grant, 2001) to pull-down DLG2 and other PDZ containing proteins would still provide valuable information. Such a pull down has currently been performed for 30 and 60 samples but requires both additional analysis of the mass spectrometry data as well as expansion to earlier time points of cortical differentiation. Various developmentally relevant signalling pathways indicated by RNAseq analysis including receptor serine/threonine kinase pathways such as BMP or HIPPO signalling could also be manipulated through the application of agonists and antagonists in an attempt to rescue the *DLG2*$^{-/-}$ phenotype and confirm whether DLG2 does act through these pathways when regulating early neurodevelopment.

These additional experiments would have to form the basis of future studies into the role of DLG2 in early cortical neurogenesis; however, various additional analyses of the currently generated 'bulk' RNAseq data to investigate the transcriptional programmes disrupted in *DLG2*$^{-/-}$ cultures

during cortical differentiation are possible, several of which are now in preprint (Sanders et al., 2020). Given the established link between *DLG2* and schizophrenia one question that can be answered from the existing RNAseq data is whether this newly identified role for DLG2 in cortical projection neuron development shows disease relevance. This can be address due to the availability of high quality schizophrenia GWAS data (Pardiñas et al., 2018) and statistically robust tools for the gene set analysis of this data, such MAGMA (de Leeuw et al., 2015), which enable sets of genes differentially expressed in *DLG2* KO cells to be tested for schizophrenia risk variant enrichment.
Although rare variant studies convincingly implicate components of mature synaptic signalling, such as DLG2-scaffolded NMDA receptor complexes, in schizophrenia aetiology (Kirov et al., 2012a; Fromer et al., 2014; Purcell et al., 2014; Szatkiewicz et al., 2014; Pocklington et al., 2015; Genovese et al., 2016) these show little evidence of GWAS association, with even more broadly classified synapse-related gene sets only accounting for a small proportion of the overall GWAS signal (Pardiñas et al., 2018). However, the common variant association signal is highly enriched for LoFi genes which capture roughly 50% of genic SNP-based heritability for schizophrenia. As LoFi genes are under strong selective pressure they are believed to play key roles in normal development (Pardiñas et al., 2018). The implied involvement of developmental processes, harbouring LoFi genes, in the schizophrenia GWAS signal and the role for *DLG2* in regulating cortical development, indicated in previous chapters, provided the rational for testing gene sets differentially expressed in *DLG2*$^{-/-}$ cells for schizophrenia common variant enrichment.
The aim of the research presented in this chapter was to determine if genes differentially expressed in *DLG2*$^{-/-}$ cells at any of the four time points studied were enriched for schizophrenia common risk variants, thereby determining if the novel role identified for *DLG2* in early cortical neurogenesis was relevant to neuropsychiatric disease. Additionally, sets of *DLG2* dysregulated genes identified by GO term analysis as being involved in *DLG2*$^{-/-}$ phenotypes were also tested to gain greater insight into the neurodevelopmental processes underpinning schizophrenia aetiology. To this end sets of genes differentially expressed in *DLG2*$^{-/-}$ cells were tested for schizophrenia common risk variant enrichment using competitive gene-set enrichment tests run in MAGMA.

6.2. Experimental procedure

Following RNAseq analysis of *DLG2*$^{-/-}$ and WT cells during cortical differentiation eight sets of differentially expressed protein coding genes were identified, being genes up and down-regulated at each of the four experimental time points (see Chapter 5). These eight gene sets were tested for schizophrenia and (as a control) Alzheimer's disease common risk variant enrichment using MAGMA (version 1.07) (de Leeuw et al., 2015). GWAS summary statistics for schizophrenia were taken from the PGC2+CLOZUK meta-analysis (Pardiñas et al., 2018) and Alzheimer's from the

IGAP meta-analysis (Lambert et al., 2013). Gene locations being taken from human genome build 37 (available at https://ctg.cncr.nl/software/magma) and MAGMA gene annotation was in a 35,10 window around genes (35kb upstream and 10kb downstream), meaning that SNPs lying within this region were assigned to a gene. Gene-wide analysis was performed using the MAGMA multi model and the 1,000 Genomes European panel (available at https://ctg.cncr.nl/software/magma) used as a reference data set to estimate the correlation between pairs of SNPs. When calculating gene-wide association statistics under the multi model, a fixed number of permutations (20,000) was used for each gene. A one-sided MAGMA gene-level analysis was performed for each of the eight sets of differentially expressed genes identified for *DLG2*$^{-/-}$ cells either without conditioning, conditioning on all expressed genes at a given experimental time point, or all expressed genes across all experimental time points, using the *condition-residualize* function. Finally, a two-sided MAGMA gene-set analysis was performed on several sub-sets of differentially expressed genes, identified in day 30 down-regulated genes through functional enrichment for GO or MP terms. Conditioning was on both all expressed genes (either at a time point or across all time points) as well as the original gene-set form which they were derived. A two-sided test being selected to detected both significant depletion or significant enrichment and thereby determine whether any over-represented functional genes sets appear to play a significantly greater, or lesser role, in schizophrenia than day 30 down-regulated genes as a whole. The gene sets used for the research presented in this chapter were taken from the combined *DLG2*$^{-/-}$ vs WT RNAseq data described previously (see Chapter 5); however, significant results were verified using data from KO1 and KO2 separately. Further details of these analyses can be found in Chapter 2, *Section 2.19 MAGMA gene set enrichment* and also follow within this chapter.

6.3. Enrichment for schizophrenia common risk variants

As previously reported, large numbers of protein coding genes are differentially expressed (both up and down-regulated) in *DLG2*$^{-/-}$ cells at four time points of cortical differentiation (see Chapter 5). In order to determine if any sets of differentially expressed genes were enriched for schizophrenia common risk variants a competitive gene-set enrichment test was performed using MAGMA (Figure 6.1, Supplementary Table 5). The results of the initial analysis (without conditioning) showed enrichment for common variant association in gene sets comprised of all expressed genes (WT and *DLG2*$^{-/-}$) at each time point as well as *DLG2*$^{-/-}$ differentially expressed genes at several time points (Figure 6.1 A). This is unsurprising given that schizophrenia is a neuropsychiatric disorder and these gene sets were from cells of a neural lineage. In order to determine if any genes sets were enriched above all expressed genes as a whole the competitive gene-set enrichment was re-run conditioning on all genes expressed in KO1, KO2 and WT cells using MAGMAs *condition-*

residualize function. Two approaches were used for this, one conditioning on all expressed genes at a given experimental time point (Figure 6.1 B) and the other conditioning on a single gene set consisting of all genes expressed at one or more time points (Figure 6.1 C). While the first conditioning approach (at a time point) provides a good internal control for each cell-type, the second (across all timepoints) allows for a better comparison across cell-types; however, both approaches produced a similar pattern of enrichment and as such subsequent enrichment analyses presented here used conditioning across all time points, analyses using at time point conditioning are presented in supplementary information (Supplementary Tables 5, 6 & 7).

Enrichment analysis revealed a strong association ($P_{corrected}$ = 8.37 x 10^{-8}) in the day 30 down-regulated gene set only, with enrichment for other gene sets no longer significant following conditioning (Figure 6.1 C). As the greatest differential gene expression between *DLG2*$^{-/-}$ and WT cells occurs at day 30 of cortical differentiation (Figure 5.5) this raises the possibility that it is simply the number of differentially expressed genes at this time point driving the strong enrichment for schizophrenia common risk variants. However, the day 30 down-regulated and day 30 up-regulated genes sets contain similar numbers of genes (4530 and 4442 respectively) yet day 30-upregulated genes show no enrichment (Figure 6.1 C). Also of note is that out of all gene-sets tested the day 30 down-regulated has the largest regression coefficient (β) indicating enrichment due to a greater effect size rather than numbers of genes (Figure 6.1 D). This was significantly greater than all up-regulated gene sets (Supplementary Table 5) and also the day 20 and day 60 down-regulated gene-sets (Figure 6.1 D). The comparison between day 30 and 20 being particularly compelling as both time points have large numbers of down-regulated genes (4030 at day 20, 4530 at day 30); however, only the day 30 downregulated genes show enrichment (Figure 6.1 C) and this is with a significantly greater effect size (p = 4.56 x 10^{-3}) than that found at day 20 (Figure 6.1 D). Finally, to verify that the observed significant enrichment for schizophrenia risk genes was a shared feature of both *DLG2*$^{-/-}$ lines the analysis was repeated using day 30 down-regulated genes from KO1 and KO2 separately (Figure 6.1 E). This showed a strong association in both lines of a relatively similar magnitude to that from the combined *DLG2*$^{-/-}$ vs WT analysis (KO1 $P_{corrected}$ = 2.46 x 10^{-7}, KO2 $P_{corrected}$ = 7.27 x 10^{-8}) providing both additional validation of the enrichment results and evidence for the similar behaviour of the knock-out lines.

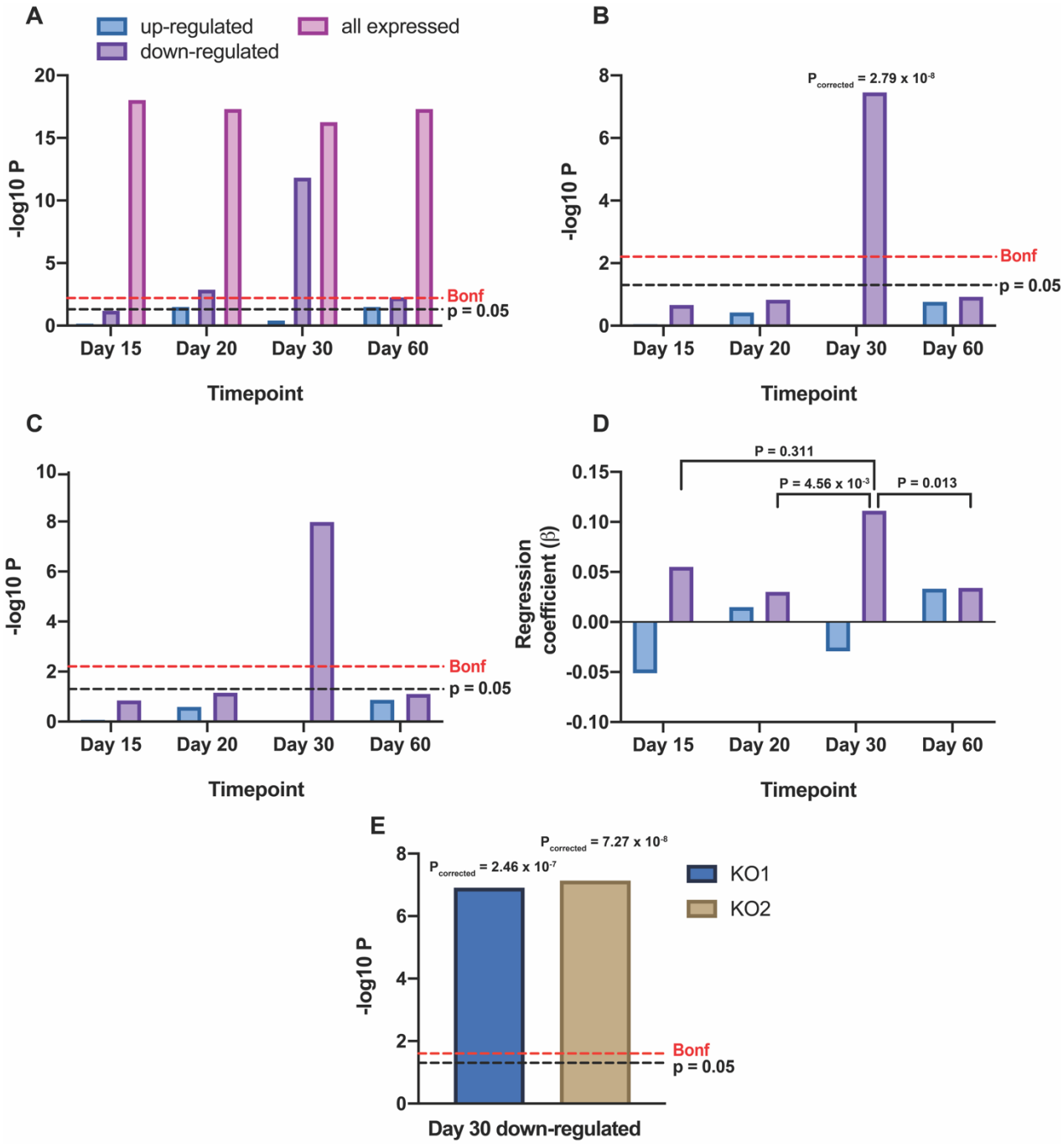


Figure 6.1 Enrichment for schizophrenia common risk variants
(A) Schizophrenia common risk variant enrichment for all expressed genes in WT and *DLG2*$^{-/-}$ cultures at four time points of cortical differentiation (all expressed) and in genes differentially expressed in *DLG2*$^{-/-}$ cultures at each time point (up or down-regulated), without conditioning. **(B)** Schizophrenia common risk variant enrichment in genes up & down-regulated at each timepoint of cortical differentiation (DLG2$^{-/-}$ relative to WT), conditioning on all expressed genes at a given time point. **(C)** & **(D)** Schizophrenia common risk variant enrichment in genes up & down-regulated at each timepoint of cortical differentiation (DLG2$^{-/-}$ relative to WT), conditioning on all expressed genes across all time points. **(E)** Schizophrenia common risk variant enrichment in genes down-regulated in KO1 and KO2 (separately) at day 30 of cortical differentiation, conditioning on all expressed genes across all time points. In all cases dotted red line indicates $P_{corrected}$ = 0.05 following Bonferroni correction for 12 tests (A), 8 tests (B-C) and 2 tests (E).

6.4. Enrichment for Alzheimer's common risk variants as a negative control

In order to confirm that the enrichment for schizophrenia common risk variants identified in genes down-regulated at day 30 of cortical differentiation is a genuine disease specific effect, rather than revealing general relevance to any brain disorders, an additional MAGMA gene-set enrichment was performed to act as a negative control. Summary statistic for this enrichment were taken from the largest available Alzheimer's GWAS, the IGAP meta-analysis (Lambert et al., 2013). Enrichment for Alzheimer's common risk variants was selected as a negative control as, unlike a neurodevelopmental disorder such as schizophrenia, Alzheimer's is a neurodegenerative disorder and therefore we would not expect to see enrichment at day 30 of cortical differentiation, a time point corresponding to active neurogenesis in culture. As described previously this enrichment was run conditioning on all genes expressed in KO1, KO2 and WT cells across all time points and showed no enrichment for the day 30 down-regulated gene set (Figure 6.2, Supplementary Table 6), therefore confirming the relevance of these genes to schizophrenia aetiology. Although not the focus of this research there was modest enrichment for Alzheimer's common risk variants in the day 20 up-regulated gene-set (Figure 6.2).

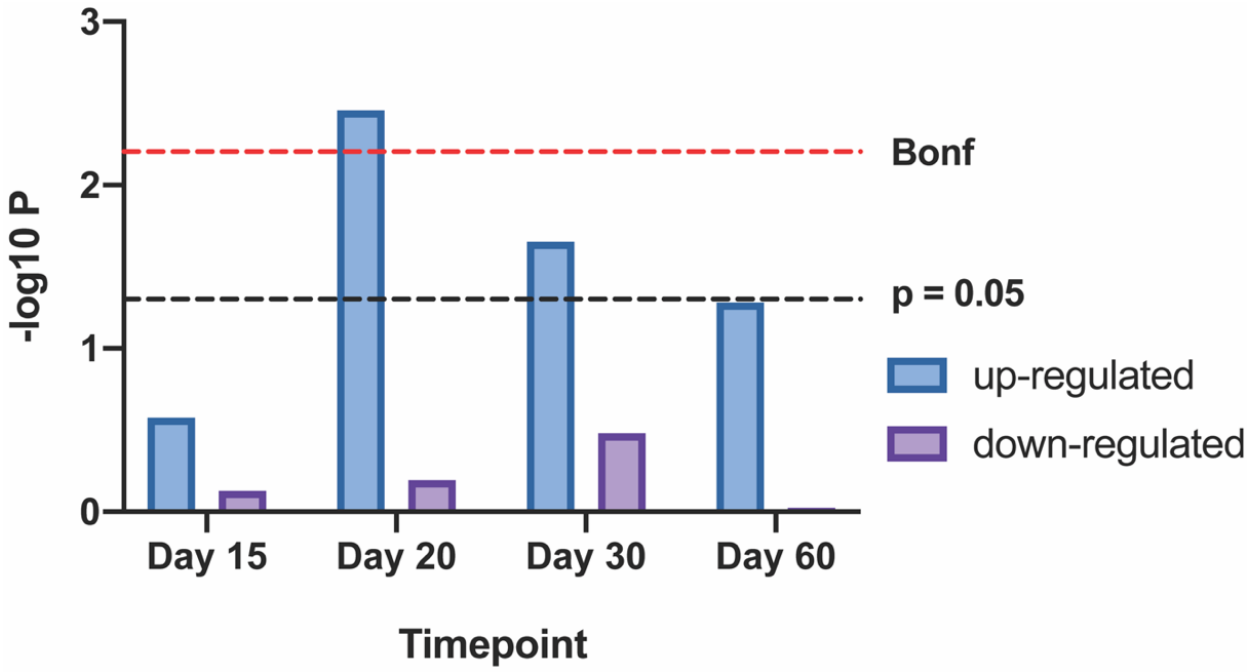


Figure 6.2 Enrichment for Alzheimer's disease common risk variants
Enrichment for Alzheimer's disease common risk variants in genes up & down-regulated at each timepoint of cortical differentiation ($DLG2^{-/-}$ relative to WT), conditioning on all expressed genes. Dotted red line indicates $P_{corrected} = 0.05$ following Bonferroni correction for 8 tests.

6.5. Relevance of specific processes and phenotypes to disease

After identifying that genes down-regulated in *DLG2*$^{-/-}$ cells at day 30 of cortical differentiation are enriched for common schizophrenia risk variants this process was refined to determine the relevance of specific biological processes and phenotypes to the disease. As previously described a Fisher's exact test was used to test enrichment in the day 30 down-regulated gene set for both GO and MP terms (see Chapter 5). After iteratively refining results by odds ratio, 28 GO terms and 19 mammalian phenotypes were identified with independent evidence for over-representation in day 30 down-regulated genes (Figure 5.10). The sets of day 30 down-regulated genes found in these over-represented GO terms and mammalian phenotypes were tested for schizophrenia common risk variant enrichment using MAGMA, to determine if any biological process or phenotype is of significantly more/less relevance to disease than the day 30 down-regulated gene set as a whole (Figure 6.3, Supplementary Table 7). For this purpose a two-sided enrichment test was performed, enabling both significant enrichment or depletion to be investigated. Tests were conditioned on both all genes expressed in *DLG2*$^{-/-}$ and WT cells across all time points and all day 30 down-regulated genes, in order to determine if any functional gene sets (GO or MP terms) have a greater of lesser role in schizophrenia than the day 30 down-regulated gene set as a whole. This analysis showed that no terms were significantly depleted, indicating that all processes and phenotypes over-represented in the day 30 down-regulated may potentially contribute to schizophrenia. However, following Bonferroni-correction two GO terms showed moderate enrichment for common schizophrenia risk variants: *voltage gated sodium channel activity* and *synapse assembly* (Figure 6.3 A) suggestive of a more central role to disease, in contrast no mammalian phenotypes were significantly enriched compared to the overall day 30 down-regulated gene set (Figure 6.3 B). Finally to validate these results for the separate *DLG2*$^{-/-}$ lines, genes within the GO terms *voltage gated sodium channel activity* and *synapse assembly* that were downregulated in KO1 and KO2 at day 30 of cortical differentiation were also tested for schizophrenia common variant enrichment, conditioning in the same manner as above (Supplementary Table 8). These results showed significant enrichment for both terms in both lines separately, again providing evidence of the broad similarity between the *DLG2*$^{-/-}$ lines.

A

GO Term	N genes	β	P	$P_{corrected}$
voltage-gated sodium channel activity	**17**	**0.975**	**8.90×10^{-4}**	**0.025**
synapse assembly	**39**	**0.697**	**1.51×10^{-3}**	**0.042**
DNA binding transcription factor activity	325	0.179	5.42×10^{-3}	0.152
nervous system development	571	0.138	7.24×10^{-3}	0.203
Rho guanyl-nucleotide exchange factor activity	35	-0.344	0.091	2.552
positive regulation of neuron projection development	78	0.226	0.099	2.763
transcriptional repressor complex	37	-0.259	0.153	4.290
beta-catenin binding	36	-0.308	0.156	4.367
regulation of dendritic spine development	24	0.320	0.163	4.568
peptidyl-serine phosphorylation	69	0.186	0.170	4.770
glutamate receptor activity	19	-0.377	0.206	5.779
axonogenesis	110	0.144	0.209	5.863
protein serine/threonine kinase activity	165	0.094	0.292	8.166
axon part	33	0.221	0.309	8.667
Rac GTPase binding	28	-0.214	0.319	8.935
regulation of synaptic plasticity	42	0.176	0.322	9.012
microtubule end	17	0.253	0.376	10.53
nuclear body	220	0.055	0.463	12.98
dendrite morphogenesis	22	0.157	0.546	15.29
regulation of axonogenesis	46	0.073	0.677	18.94
nuclear periphery	40	-0.069	0.689	19.29
negative regulation of nervous system development	56	0.051	0.734	20.56
Rab GTPase binding	68	-0.035	0.792	22.19
covalent chromatin modification	111	0.028	0.798	22.34
modification-dependent protein binding	56	0.020	0.887	24.83
central nervous system neuron development	19	0.022	0.935	26.19
regulation of cell morphogenesis	142	-2.69×10^{-3}	0.978	27.39
ubiquitin-like protein ligase activity	84	-1.42×10^{-5}	1.000	28.00

B

MP Term	N genes	β	P	$P_{corrected}$
abnormal CNS synaptic transmission	265	0.189	8.98×10^{-3}	0.171
lethality during fetal growth through weaning	665	0.094	0.039	0.736
small olfactory bulb	25	0.475	0.053	1.012
perinatal lethality	388	0.110	0.069	1.311
abnormal miniature excitatory postsynaptic currents	52	0.229	0.148	2.806
absence seizures	21	0.351	0.165	3.131
limb grasping	87	0.126	0.295	5.612
abnormal hippocampus pyramidal cell layer	40	0.166	0.305	5.787
pericardial effusion	44	-0.150	0.348	6.612
abnormal discrimination learning	18	-0.210	0.392	7.447
abnormal contextual conditioning behavior	67	0.119	0.398	7.566
abnormal sternum morphology	71	0.116	0.412	7.829
abnormal parietal lobe morphology	20	-0.245	0.423	8.033
abnormal embryonic/fetal subventricular zone morphology	20	0.186	0.477	9.066
embryonic growth retardation	168	0.048	0.580	11.03
abnormal synaptic vesicle number	21	0.140	0.591	11.24
decreased fetal size	72	0.053	0.676	12.84
perinatal lethality, complete penetrance	97	-0.035	0.766	14.55
abnormal channel response	19	-0.076	0.776	14.74

Figure 6.3 Enrichment for schizophrenia common risk variants in GO terms and mammalian phenotypes
(A) Enrichment for schizophrenia common risk variants in GO terms over-represented amongst genes down-regulated in *DLG2*$^{-/-}$ cultures at day 30 of cortical differentiation, conditioned on all expressed genes and on the day 30 down-regulated gene set. $P_{corrected}$ = P value following Bonferroni correction for 28 tests. **(B)** Enrichment for schizophrenia common risk variants in gene MP terms over-represented amongst genes down-regulated in *DLG2*$^{-/-}$ cultures at day 30 of cortical differentiation, conditioning on all expressed genes and on the day 30 down-regulated gene set. $P_{corrected}$ = P value following Bonferroni correction for 19 tests.

6.6. Discussion

6.6.1. Genes down-regulated in *DLG2*$^{-/-}$ cells at day 30 of cortical differentiation are enriched for schizophrenia common risk variants

It is perhaps unsurprising that sets of genes expressed in WT and *DLG2*$^{-/-}$ cells at various time points of differentiation from hESCs to cortical projection neurons are enriched for schizophrenia common risk variants, given the neural lineage of these cells and the nature of schizophrenia as a neuropsychiatric disease. However, one gene set remains strongly enriched for schizophrenia common risk variants even after conditioning the analysis on all genes expressed in *DLG2*$^{-/-}$ and WT cells across all time points of the experiments, linking several of the *DLG2*$^{-/-}$ phenotypes identified in previous chapters to neuropsychiatric disease. These results are also consistent with published genetic data indicating a role for *DLG2* in schizophrenia pathogenesis (Kirov et al., 2012a; Nithianantharajah et al., 2013; Marshall et al., 2017; Michaelovsky et al., 2019). Given the established role of DLG2 in the postsynaptic density of mature neurons it is interesting to note that this strong enrichment for schizophrenia genetic risk was not found in the genes from the day 60 experimental time point, containing the most mature neurons, but rather in the day 30 down-regulated gene set. This strong enrichment at time point corresponding to active neurogenesis (day 30), places emphasis on the novel role for *DLG2* in corticoneurogenesis reported in previous chapters as it pertains to schizophrenia genetic risk. This is of course in agreement with the long established neurodevelopmental component of schizophrenia aetiology (Murray and Lewis, 1987; Weinberger, 1987; Harrison, 1997, 1999), with recent work also supporting a role for various schizophrenia risk factors during corticoneurogenesis (Walker et al., 2019) as well as more generally highlighting the contribution of developmentally relevant LoFi genes to schizophrenia genetic risk (Pardiñas et al., 2018). The lack of GWAS enrichment in day 15 or 20 gene sets also indicates that the novel NSC and NPC phenotypes identified in previous chapters, such as increased proliferation and adhesion to the extracellular matrix, although of interest in a developmental context are not directly relevant to the schizophrenia common variant signal, although their enrichment for rare variants remains to be tested.

The evidence of schizophrenia common risk variant enrichment in the day 30 down-regulated gene set is robust, being highly significant after Bonferroni correction ($P_{corrected} = 8.37 \times 10^{-8}$). This enrichment for day 30 down-regulated genes appears not to be a result of the number of genes tested, as indicated by its large effect size (β) that showed a significant increase over the majority of gene sets tested; furthermore, no other gene set showed significant enrichment despite several containing a broadly similar numbers of genes to the day 30 down-regulated set. A lack of enrichment for Alzheimer's common risk variants in the day 30 down-regulated gene set provides evidence that results are a genuine effect relevant to schizophrenia pathophysiology, rather than more generally

relevant to brain disorders. Although neurodegenerative disorders such as Alzheimer's disease are not the focus of this research it should be noted that the day 20 up-regulated gene set was modestly enriched for Alzheimer's common risk variants. Although surprising that genes from an NPC time point (day 20) would be enriched for a neurodegenerative disease such as Alzheimer's, the day 20 up-regulated gene set was enriched for GO terms relating to protein phosphorylation (Figure 5.8), offering a possible explanation given the link between dysregulation of this process and Alzheimer's aetiology (Kumar et al., 2015; Oliveira et al., 2017; Zhang et al., 2019).

6.6.2. *DLG2* knockout reveals multiple biological processes disrupted in schizophrenia

Subsets of genes down-regulated in *DLG2*$^{-/-}$ cells at day 30 of cortical differentiation can be identified based on enrichment for GO terms and mammalian phenotypes as described in Chapter 5 (Figure 5.8). Following a two-sided test none of these identified functional gene sets were significantly depleted for schizophrenia common variant enrichment, indicating that the schizophrenia GWAS signal is distributed across these gene sets and thus implicating the disruption of multiple biological processes in schizophrenia (Figure 6.3). Such *DLG2* relevant biological processes include those identified both through RNAseq analysis and experimental assays, such as the regulation of neuronal morphology and electrical activity (see Chapters 4 and 5). Although no functional gene sets were depleted for enrichment, indicating a role for all in schizophrenia developmental pathophysiology*, voltage-gated sodium channel activity* and *synapse assembly* did show a modest but significantly increased enrichment suggesting these two terms best capture the most relevant pathophysiological processes. The *synapse assembly* set contains genes encoding various cell adhesion molecules (CAMs) with established roles in synapse development in addition to pre and postsynaptic specialisation (Südhof, 2018) that are differentially expressed in day 30 *DLG2*$^{-/-}$ cells. These CAMs include Neuroligin (*NLGN1*) and Neurexin (*NRXN1*), associated with schizophrenia (Kirov et al., 2009; Sun et al., 2011), and others such as Ephrin receptor B2 (*EPHB2*) which have additional roles in axon guidance and neuron branching (Xu and Henkemeyer, 2012), providing a potential link to other identified *DLG2*$^{-/-}$ phenotypes. This enrichment for synapse assembly genes is of interest as although synaptic dysfunction has been robustly implicated in schizophrenia pathophysiology the current genetic support for disease models focuses on postsynaptic signalling and synaptic plasticity (Kirov et al., 2012a; Fromer et al., 2014; Szatkiewicz et al., 2014; Pocklington et al., 2015; Genovese et al., 2016). Genes related to the regulation of synaptic plasticity were indeed over-represented in the day 30 down-regulated set confirming their relevance; however, these were not enriched for schizophrenia genetic risk more than the gene set as whole. Taken together this may suggest that a greater emphasis should be placed on synapse assembly during early cortioneurogenesis in the aetiology of schizophrenia although this obviously

requires further investigation. As disruption of synaptic NMDA and ARC receptor complexes is the schizophrenia relevant biological processes with the most genetic support (Kirov et al., 2012a; Fromer et al., 2014; Purcell et al., 2014; Szatkiewicz et al., 2014; Pocklington et al., 2015; Genovese et al., 2016) this would suggest testing to determine whether the rare variant signal in these complexes is solely a mature functional signal, or if it can be captured be genes expressed during early development.

The increased schizophrenia common variant enrichment for *voltage-gated sodium channel activity* genes (Figure 6.3) provides disease relevance to the aberrant electrophysiological properties of *DLG2*$^{-/-}$ neurons as previously described (see Chapter 5), given the well-established role of these channels in the resting membrane potential and action potential generation (Catterall, 2012). This is supported by genetic studies implicating voltage gated sodium channels in several disorders including schizophrenia and ASD (Wolff et al., 2017; Rees et al., 2019). Dysfunction in these channels may also be linked to additional *DLG2*$^{-/-}$ phenotypes, including deficits in neuronal morphology and migration, as voltage gated sodium channel activity has been shown to regulate spontaneous calcium transients critical for neuronal migration and dendritic branching within the developing neocortex (Bando et al., 2016). Finally, the development of voltage-dependent ion channels in immature neurons and subsequent spontaneous electrical activity has been shown to have a key role in the establishment and maturation of neuronal networks, with any disturbance of this process potentially leading to long-lasting neuronal deficits (Luhmann et al., 2016).

6.6.3. Summary

Using a competitive gene set enrichment test implemented in MAGMA, genes downregulated in *DLG2*$^{-/-}$ cells at day 30 of cortical differentiation, a time point corresponding to early neurogenesis, are shown to be strongly enriched ($P_{corrected} = 8.37 \times 10^{-8}$) for schizophrenia common risk variants, a result verified in both KO1 and KO2 seperately. This enrichment was conditioned on all genes expressed in both *DLG2*$^{-/-}$ and WT cells across all experimental time points and appears robust, as dysregulated gene sets of a similar size do not show enrichment. In addition the significantly enriched day 30 down-regulated set has a larger effect size than the majority of other gene sets tested, as determined by a significantly increased β. The day 30 down-regulated gene set is also not enriched for Alzheimer's disease indicating the observed schizophrenia enrichment is a disease specific effect and not related to brain disorders more generally. Various subsets of functional genes were identified within the day 30 downregulated gene set and upon testing none of these were depleted for schizophrenia association, indicating a contribution from multiple biological processed in the disease. Two specific subsets of genes displayed evidence for increased enrichment, being those related to voltage gated sodium channel activity and synapse assembly, suggesting a greater

contribution to schizophrenia pathophysiology from these processes. This work has therefore indicated that DLG2 mediated signalling regulates schizophrenia relevant gene expression during early corticoneurogenesis and provides may provide insight into processes contributing to the neurodevelopmental component of disease aetiology; however, the precise molecular mechanism of action for DLG2 remains unclear and requires further study.

7. General discussion

7.1. Research Summary

DLG2 in common with other MAGUKs is known to act as a scaffolding protein within the PSD of mature neurons, with roles in the assembly, localisation and function of neurotransmitter receptors complexes that are associated with the regulation of synaptic transmission and plasticity (Zheng et al., 2011; Frank et al., 2016; Zhu et al., 2016). In addition to these established roles in later neurodevelopment (subsequent to synaptogenesis) it was hypothesised that DLG2-mediated signalling may also regulate early cortical neurogenesis due to the association between *DLG2* and neuropsychiatric diseases such as schizophrenia, which possess a known neurodevelopmental component with risk factors acting both in cortical excitatory neurons and their progenitors (Hill and Bray, 2012; Willsey et al., 2013; O'Brien et al., 2018; Clifton et al., 2019; Polioudakis et al., 2019; Walker et al., 2019). This hypothesis was reinforced by *DLG2* mRNA expression both *in vitro* and *in vivo* during early neurodevelopment and prior to synaptogenesis (Kang et al., 2011; van de Leemput et al., 2014; Yoo et al., 2020) as well as the known involvement of the invertebrate orthologue of DLG2 in the regulation of developmental signalling pathways (Humbert et al., 2008; Stephens et al., 2018).

The research presented here supports this hypothesis and indicates a novel role for DLG2 in the regulation of early cortical neurogenesis, which was identified through the characterisation of WT and *DLG2*$^{-/-}$ cells during the course of differentiation from hESC to cortical projection neurons. This involved multiple complimentary techniques including western blotting mass spectrometry, immunocytochemistry, RNA sequencing and various phenotypic assays. Enrichment for schizophrenia common risk variants in genes differentially expressed in *DLG2*$^{-/-}$ cultures suggested that processes regulated by DLG-mediated signalling during excitatory cortical neurogenesis may be disease relevant and could potentially provide insight into the neurodevelopmental component of schizophrenia pathophysiology.

7.2. DLG2-mediated signalling in early cortical differentiation

An initial search of transcriptome databases revealed that *DLG2* is expressed earlier in the neurodevelopmental time course than previously appreciated and prior to synaptogenesis. *DLG2* mRNA being present from 8 weeks post-conception in humans (Kang et al., 2011) as well as *in vitro* throughout all stages of differentiation from hESCs to cortical projection neurons (van de Leemput et al., 2014). Efforts to verify DLG2 protein expression during early cortical differentiation using western blotting were hampered by issues with antibody specificity so mass spectrometry was employed as an alternative approach and this robustly identified DLG2

expression in WT cells at day 30 of cortical differentiation, a time point corresponding to the early emergence of postmitotic neurons in culture. Although resource constraints precluded the investigation of earlier time points using mass spectrometry, western blotting analysis of *DLG2*$^{-/-}$ cells at day 20 of cortical differentiation indicated increased expression of PAX6 (a key dorsal forebrain developmental transcription factor) and KI67 (a marker of cell proliferation) compared to WT, suggesting a role for DLG2-mediated signalling (and by implication, active DLG2 expression) in NPCs.

In addition to confirm *DLG2* mRNA expression in WT cells at all experimental time points RNAseq provided further evidence for DLG2 expression during early cortical differentiation, with *DLG2*$^{-/-}$ cells showing large scale gene expression changes compared to WT from as early as day 15. At these early time points of differentiation, corresponding to the NPC stage, genes differentially expressed in *DLG2*$^{-/-}$ cells were over-presented in GO terms relating to biological processes including proliferation and adhesion. Experimental assays validated these findings, with *DLG2*$^{-/-}$ NPCs showing increased proliferation and adhesion to several ECM substrates compared to controls. Such a role for DLG2-mediated signalling in the regulation of proliferation would be novel in the context of cortical neurogenesis but in agreement with the known role of DLG2 more widely, previously studies having identified *DLG2* as a tumour suppressor gene with abnormally low *DLG2* expression linked to ovarian cancer, osteosarcoma and recently neuroblastoma (Shao et al., 2019; Keane et al., 2020). Although the precise mechanism by which DLG2 regulates cell cycle progression is unclear, invertebrate Dlg is known to have a similar role and acts via interaction with the Scribble cell polarity module. The GK domain of Dlg interacts with the PDZ2 domain of Scrib via the adaptor protein GUK-holder as well as directly with another component of the Scribble module, Lethal giant larvae (Lgl) in a phosphorylation dependent manner (Mathew et al., 2002; Zhu et al., 2014). Loss of function of either *Scrib*, *Dlg* or *Lgl* results in excessive cell proliferation, tumour formation and overgrowth of larvae providing a clear link to the regulation of cell cycle progression (Humbert et al., 2008; Stephens et al., 2018). In humans this role of Dlg is believed to be taken by DLG1 (also called human DLG or hDlg) interacting with SCRIB (Massimi et al., 2008; Elsum et al., 2012; Stephens et al., 2018). However, DLG1 and DLG2 are both members of the same MAGUK subfamily and with the exception of the L27 domain of DLG1 have a very similar structure which includes a GK domain (Zheng et al., 2011; Zhu et al., 2016). Interactions between the DLG4 GK domain and LGL2 (a homologue of Lgl) *in vitro* have also been established (Zhu et al., 2014), raising the possibility that DLG2-LGL2 or DLG2-SCRIB complexes may regulate cell cycle progression in cortical NPCs.

An alternative or perhaps complementary mechanism by which DLG2 may regulate NPC proliferation is through the ID proteins, which inhibit the binding of basic helix-loop-helix

transcription factors to DNA and whose knockout has been shown to increase cell cycle length (Jung et al., 2010; Niola et al., 2012). During early corticoneurogenesis at day 15 of differentiation, the expression of *ID1*, *ID2* and *ID3* was upregulated in *DLG2*$^{-/-}$ cells compared to controls, with these being the most significantly differentially expressed genes at this time point. This is consistent with a decrease in cell cycle length, inferred from the increased proliferation observed in *DLG2*$^{-/-}$ cultures. In addition to DLG2 affecting *ID* gene expression either directly or indirectly to regulate NPC proliferation, there is also the possibility that the two proteins may interact. This would be similar to CASK, which is known to regulate proliferation through interaction with ID1 (Qi et al., 2005). DLG2 and CASK possess a comparable structure which contains the PSG tandem common to all MAGUKs, although CASK contains three additional domains, two L27 and one CaM (Zheng et al., 2011; Zhu et al., 2016). If DLG2 does not directly interact with the ID proteins, it may do so in complex with CASK; DLG1-CASK interactions have been shown in mammalian epithelia where they regulate mitotic spindle orientation (Porter et al., 2019). Interestingly, in Drosophila neuroepithelial cells mitotic spindle orientation and proliferation are under the control of a common signalling pathway (Franco and Carmena, 2019), raising the possibility that DLG1 or potentially DLG2 complexes may play similar roles in spindle orientation and proliferation in cortical NPCs. It should be noted that proteins that share similar domain structures do not necessarily share binding partners or act through similar mechanism. This can be seen with DLG2 and DLG4 each of which contain one GK, one SH3 and three PDZ domains yet have distinct roles in the regulation of synaptic plasticity (Carlisle et al., 2008), making further work to validate any potential DLG2 interactions essential.

Cell adhesion is another process that has been previously associated with MAGUK proteins, although this is in the context of cell-cell adhesion where these scaffolding proteins are known to stabilise tight junctions (Funke et al., 2005). In contrast, genes differentially expressed in *DLG2*$^{-/-}$ cultures during early cortical differentiation were shown to be enriched for biological processes related to cell-ECM adhesion, with experimental assays confirming increased adhesion between *DLG2*$^{-/-}$ NPCs and various ECM substrates including the most abundant ECM component Collagen I (Frantz et al., 2010). Interestingly *COL1A1* expression, the gene encoding Collagen I, is also significantly upregulated in *DLG2*$^{-/-}$, being the most significantly differentially expressed at day 15 of differentiation after the *ID* genes. It should be noted that although collagen production is usually more associated with glia, *in vitro COL1A1* expression in NPCs has previously been shown at comparable time points of cortical differentiation to this day 15 data (van de Leemput et al., 2014) and antibody staining shows the protein is expressed in neurons within the human cerebral cortex (Uhlén et al., 2015) (https://www.proteinatlas.org/ENSG00000108821-COL1A1/tissue/cerebral+cortex). This elevated *COL1A1* expression may therefore suggest that the

increased adhesion in *DLG2*$^{-/-}$ cells may in part be due to a greater concentration of proteins surrounding cells producing a denser and more cross-linked ECM, potentially resulting in an increased number and/or stronger cell-ECM interactions. MAGUK proteins are also known to be targeted by proteases for degradation which weakens cell contacts prior to detachment (Ivanova et al., 2011), raising the possibility that cell-ECM contacts which form in the absence of the normal scaffolding proteins, such as DLG2, cannot be efficiently targeted by proteases leading to increased adhesion.

In addition to their role in cell cycle regulation, the ID proteins have also been shown to regulate cell-ECM adhesion, with ID1-3 knockout decreasing adhesion to the ventricular surface of the embryonic cortex *in vivo* and to ECM substrates *in vitro* (Niola et al., 2012). This increased adhesion in *DLG2*$^{-/-}$ cells may indicate a defect in the normal process of cortical differentiation, in which transcriptional programmes inhibit NPC adhesion to the ECM, triggering the detachment and subsequent migration of cells (Itoh et al., 2013). Further evidence for an early disruption to cortical differentiation during the NPC stage can been seen by *DLG2*$^{-/-}$ differentially expressed genes being enriched for receptor serine/threonine kinase signalling pathways including BMP signalling, a key regulator of early neurodevelopment (Hegarty et al., 2013). Taken together these data from western blotting, mass spectrometry RNAseq and cellular assays clearly suggest that DLG2-mediated signalling is present during early cortical differentiation, although the full range of processes regulated by this signalling and the precise mechanisms underlying it require clarification. Give the evidence that neuronal sub-type identify is determined by the internal state of NPCs immediately prior to cell-cycle exit (Telley et al., 2019) it may be the perturbation of gene expression in *DLG2*$^{-/-}$ NPCs that underlies subsequently observed neuronal phenotypes.

7.3. DLG2-mediated signalling regulates schizophrenia relevant aspects of corticoneurogenesis

Despite the large scale differential gene expression and resulting phenotypes observed in *DLG2*$^{-/-}$ NPCs these cells are able to exit the cell cycle and differentiate to cortical projection neurons. At day 30 of cortical differentiation, a timepoint corresponding to early neurogenesis, *DLG2*$^{-/-}$ cultures contain a similar proportion of postmitotic neurons as controls, as determined by the expression of the markers NEUN, TBR1 and MAP2, indicating that cell cycle exit is a *DLG2* independent process (although *DLG2*$^{-/-}$ NPCs may have increased levels of PAX6 and KI67 as previously described). However, the neurons produced by *DLG2*$^{-/-}$ cultures appear to exhibit a more immature subtype identity as determine by their gene expression, morphology and electrophysiological properties. This includes reduced expression of the neuronal transcription factor and cortical layer V marker

CTIP2, a lack of arborisation with long but relatively unbranched neurites, reduced neuronal migration as well as a reduction in both the number and maturity of generated action potentials. Such a delay is also supported by hierarchical clustering analysis of the RNAseq data, which shows *DLG2*$^{-/-}$ samples from later time points of differentiation clustering with WT samples from earlier time points.

Although several of these neuronal phenotypes persist from day 30 cultures containing new-born neurons to day 60 cultures, in which more mature neurons are present, some do show a degree of recovery; *CTIP2* expression recovers fully at the mRNA level by day 60 and while protein expression does lags behind this it does increase relative to WT levels between days 30 and 60. RNAseq analyses also shows that more generally the peak of differential gene expression is at day 30 while day 60 - the latest time point of cortical differentiation investigated - had a substantially lower percentage of genes differentially expressed in *DLG2*$^{-/-}$ cells. Functional enrichment analyses for GO terms among genes differentially expressed in *DLG2*$^{-/-}$ cultures are also suggestive of recovery, with numerous biological processes initially downregulated in *DLG2*$^{-/-}$ either being upregulated or showing no change from WT at later time points.

Genes differentially expressed in *DLG2*$^{-/-}$ cultures at days 15 and 20 of cortical differentiation were not enriched for schizophrenia common variants, indicating that the phenotypes identified at these time points - although of interest in a developmental context and likely underlying the subsequently observed neuronal phenotypes - may not be directly disease relevant (although their rare variant association remains to be tested). In contrast, genes downregulated in *DLG2*$^{-/-}$ cultures at day 30 of cortical differentiation are highly enriched for common genetic variants conferring risk to schizophrenia, providing a link between the identified *DLG2*-dependent defects in neuronal subtype identity and neuropsychiatric disease. This same gene set is not enriched for Alzheimer's common risk variants, validating the relevance of genes downregulated at day 30 of cortical differentiation to schizophrenia, rather the brain disorders more generally. Interestingly, genes differentially expressed in cultures containing more mature neurons at day 60 show no enrichment for schizophrenia common risk variants potentially indicating that DLG2 knockout does not dysregulate any disease-relevant aspects of later gene expression although synaptic genes in adult neurons may not yet be expressed at day 60 in culture, hence its involvement cannot be definitively excluded. Also supporting a role for *DLG2* in the disruption of early (rather than late) neurogenesis in schizophrenia and other disorders further analysis of these data revealed a cascade of transcriptional programmes active during corticoneurogenesis that were enriched for both schizophrenia common and rare variants as well as genetic risk factors for a range of neuropsychiatric and neurodevelopmental disorders (Sanders et al., 2020).

Biological processes well known to contribute to mature neuronal function also have roles in early neurodevelopment, where their disruption may contribute to disease risk. One such example being voltage-gated sodium (Na_v) channel activity, which is vital for sodium influx into neurons during action potential generation (Catterall, 2012). Na_v channel genes are particularly enriched for schizophrenia common risk variants among the genes downregulated in *DLG2*$^{-/-}$ cells at day 30, supporting earlier genetic studies that have reported rare mutations in Na_v channel genes in disorders including schizophrenia and ASD (Wolff et al., 2017; Rees et al., 2019). Dysfunction in these channels may be linked to the migration and morphology defects observed in *DLG2*$^{-/-}$ neurons, with Na_v channel activity shown to enhance the spontaneous calcium transients that are critical for neuronal maturation in the developing neocortex, including neuronal migration and dendritic branching (Bando et al., 2016). Mutations in voltage-gated calcium (Ca_v) channels that generate these transients are also strongly associated with neuropsychiatric disorders: both *CACNA1C* and *CACNA1I* (encoding $Ca_v1.2$ and $Ca_v3.3$ respectively) are located at single-gene genome wide significance schizophrenia loci (Pardiñas et al., 2018), while rare mutations in *CACNA1C* are also found in Timothy syndrome and Brugada syndrome (Splawski et al., 2004; Watanabe and Minamino, 2016). The $Ca_v1.2$ channel in particular has been shown to play a key role in the generation of spontaneous calcium currents in both the axons and dendrites of developing cortical neurons, with mutations in *CACNA1C* showing phenotypes related to those presented in this research for *DLG2* deficient cells including reductions in neurite growth and impaired radial migration (Kamijo et al., 2018).

7.4. Stochastic signalling due to an absent DLG2 scaffold may explain the *DLG2*$^{-/-}$ phenotype

DLG2 is a well-studied scaffold protein of the PSD, where it connects cell-surface receptors to downstream signalling molecules and maintains them in the correct location and orientation for efficient signal transduction. Given its structure and known function (and that of invertebrate *Dlg*, discussed above), it seems likely that DLG2 plays a role in the formation of signal transduction complexes at earlier time points in neural development. If this were the case, then signalling would still be expected to occur in the absence of *DLG2* - as the various required components would still be present in the cell - but it would be more stochastic, as without a DLG2 scaffold these components would not be held in the optimal position for efficient signalling. The resulting 'noisy' signalling could delay, rather than prevent, the activation of various transcriptional programs required for normal neurodevelopment and may explain why several features of the *DLG2*$^{-/-}$ phenotype show recovery at later experimental time points, such as: overall differential gene expression (Figure 5.6); expression of key neuronal transcription factors such as *BCL11B* (CTIP2)

(Supplementary Table 2); and various biological processes identified through functional enrichment analyses for GO terms (Figures 5.7, 5.8, 5.10 & 5.13).

Where neurogenesis is considered, DLG2-scaffolded signalling complexes appear to regulate schizophrenia relevant neurogenic transcriptional programmes responsible for the expression of cell-specific properties that underlie neuronal subtype identity. However, as neurogenesis is linked to cell cycle exit, a process that is not impacted by DLG2 loss, this suggests that presence of at least two broad signalling pathways; DLG2-independent signalling being responsible for cell cycle exit and the subsequent generation of postmitotic neurons, while DLG2-dependent signalling is responsible for the expression of correct neuronal subtype identity (Figure 7.1). Given the diverse features of the *DLG2*$^{-/-}$ phenotype identified here and the multiple signalling pathways with which DLG2 may interact, this increase in signal noise leading to the delayed activation of transcriptional programs may be the most plausible underlying mechanism of action, although further work is clearly required to clarify this.

7.5. Future studies

Additional experiments are required to clarify the mechanism of action for *DLG2* and further investigate its role in cortical neurogenesis and neuropsychiatric disease. A key experiment would be to use mass spectrometry to identity DLG2-interacting proteins during early cortical differentiation, in NPCs and new-born neurons, in order to confirm the presence of DLG2 scaffolded complexes at these time points and gain insight into the signalling pathways they regulate. As highly specific anti-DLG2 antibodies are not currently available for use in pull-down experiments, this would ideally require the generation of a knock-in cell line. Previous efforts to knock-in a FLAG tag to *DLG2* using CRISPR/Cas9 genome editing and a HDR template were not successfully; however, an alternative HITI-based approach (Suzuki et al., 2016; Suzuki and Izpisua Belmonte, 2018) was recently used to knock-in a tag to *DLG4*, suggesting a similar approach may work for *DLG2* (Willems et al., 2019). Such a knock-in could also be used in conjunction with ICC and western blotting to validate the temporal pattern of DLG2 expression and further to investigate the spatial pattern of DLG2 expression during the course of cortical differentiation; to determine how the subcellular localisation of DLG2 varies over time as well as the proteins with which it colocalises, both of which would provide useful details regarding DLG2 function.

Given the evidence that genes relating to electrical activity are enriched for schizophrenia risk in *DLG2*$^{-/-}$ cells at day 30, additional electrophysiological recordings could be made to better investigate this time point. This should perhaps be particularly focused on spontaneous electrical activity, given both the immature nature of day 30 neurons and the role of this activity in their subsequent maturation (Luhmann et al., 2016). As voltage-gated sodium channels were also

implicated by enrichment analyses, agonists and antagonists to these channels (De Lera Ruiz and Kraus, 2015) could be applied to further investigate their role in the electrophysiological properties of developing *DLG2*$^{-/-}$ neurons and the effect of perturbing these properties on other observed phenotypes.

Given the evidence that DLG2 loss dysregulates the expression of pathways involved in schizophrenia, efforts should be made to confirm which *DLG2*$^{-/-}$ phenotypes show recovery over the course of cortical differentiation; which phenotypes persist; and what are their long-term consequences on brain function. Several phenotypes did show recovery at the mRNA level, others showed a trend towards recovery at the protein level while others - such as cell morphology changes - showed no evidence of recovery during the time period investigated. As changes at the protein and cellular level often lag behind those at the mRNA, the obvious approach would be to extend cortical differentiation over a longer period in order to characterise phenotypes at later time points using ICC, western blotting, mass-spec, RNAseq and other suitable techniques. In conjunction with this, rescue experiments could be performed such as co-culturing WT and *DLG2*$^{-/-}$ cells during the course of differentiation to determine if this accelerates the recovery of the *DLG2*$^{-/-}$ phenotype. If this were the case it would indicate a role for external signalling (in the form of secreted molecules or cell-cell connections) in the *DLG2*$^{-/-}$ phenotype; however, if co-culture did not accelerate recovery this would indicate the *DLG2*$^{-/-}$ phenotype was the result of solely cell intrinsic processes. As a second rescue experiment, *DLG2* could be reintroduced to the *DLG2*$^{-/-}$ cell lines through an inducible lentiviral construct; this would allow *DLG2* expression only during defined periods and could be used to determine if there was a critical time window during the course of cortical differentiation (such as the NPC stage) within which *DLG2* expression is required for normal cortical neurogenesis, or whether expression of *DLG2* is required throughout.

As neurons can only mature to a certain point *in vitro* and maintaining the cells in culture for long periods poses several challenges, transplantation of cultured cells into rodents would be an option for any future investigation of later stages of cortical differentiation. Even if transplantation was not required there is a clear place for rodents in any future studies of the *DLG2*$^{-/-}$ phenotype in the form of *DLG2*$^{-/-}$ rodent models. Previous studies have generated such models (Nithianantharajah et al., 2013; Winkler et al., 2018; Yoo et al., 2020) although these were examined in the context of adult behavioural and cognitive deficits rather than embryonic neurodevelopment. Given the clear *in vitro* *DLG2*$^{-/-}$ phenotypes identified in this research a next logical step would be to validate these *in vivo*, through a comparison of both embryonic and adult brains in WT and *DLG2*$^{-/-}$ mice. As these *in vitro* phenotypes are associated with early cortical projection neuron development the first step would be to examine the cortex of embryonic mouse brains to determine if these are replicated, if this was the case then brains from later embryonic days and if necessary, adults could then be studied to

reveal the long-term consequences of *DLG2* loss and the extent to which early phenotypes recover. Immunohistochemistry could be performed on sections of cortex staining for markers dysregulated in the *in vitro* studies such as CTIP2, while general neuronal markers could be used be used to quantify neuronal numbers and examine neuronal morphology. *In utero* electroporation of a fluorescent marker would facilitate studies of neuronal migration while either cortical slices or cultured primary neurons could be studied for electrophysiological properties. Given the focus on this research these *in vivo* studies should initially be focused on cortical projection neuron development but could readily be expanded to investigate the consequences of DLG2 deficiency on local connectivity through a study of the expression profile, morphology, migration and electrical activity of cortical interneurons. Similarly, the research presented here could be repeated (and future experiments also conducted) on human cortical interneurons generated *in vitro* through the directed differentiation of hESCs, for which there are established protocols (Maroof et al., 2013).

Additional analyses of the RNAseq data presented here have been performed to uncover distinct neurodevelopmental transcriptional programmes disrupted in schizophrenia and a wide spectrum of related disorders as a result of *DLG2* knockout (Sanders et al., 2020). This research also identifying potential key genetic regulators of these programmes. Although the bulk RNAseq used for these analyses allows for a greater depth of reads and/or coverage, allowing the expression of more genes to be quantified, one disadvantage is that it was taken from a somewhat heterogenous population of cells from a given time point of cortical differentiation, meaning that cell type specific phenotypes and transcriptional programmes could be unequivocally identified. A single-cell RNAseq analysis at similar time points of cortical differentiation could be performed to address this, potentially in both $DLG2^{-/-}$ cells as used here (to determine the role of *DLG2* in the absence of confounding factors) and patient derived $DLG2^{+/-}$ cells (to investigate a more directly disease-relevant model system). Taken together, single-cell RNAseq, interacting protein data and the various cellular assays described here may allow potentially disease relevant, *DLG2* dependent, signalling pathways acting in early cortical development to be refined. Following their clarification these DLG2-dependent signalling pathways could subsequently be manipulated in rescue experiments, in an attempt to recover $DLG2^{-/-}$ phenotypes.

7.6. Conclusions

The research presented here supports that hypothesis that DLG2-mediated signalling regulates aspects of early cortical neurogenesis in addition to its established functions as a structural scaffolding protein within the PSD. Biological processes that appear to be regulated by DLG2-dependent signalling include cell cycle regulation and adhesion in NPCs, which may underpin subsequent phenotypes, as well as aspects of neurogenesis that are responsible for the generation of

mature neuronal subtype identity and are highly enriched for schizophrenia common risk variants. The precise mechanism of action for DLG2 during early cortical differentiation remains unclear and will require further experimentation to clarify. However, it is hypothesised that DLG2-scaffolded signalling complexes act at earlier stages of development than previously appreciated, and the absence of *DLG2* stochastic signalling results in the delayed activation of transcriptional programmes and the subsequent expression of cell-specific properties required for normal development. As precise timing is required during normal brain development, it is possible that even a temporary disruption to corticoneurogenesis (defects to neuronal migration, branching and electrical properties as observed in *DLG2*$^{-/-}$ cells) may have long-lasting impacts on the formation and subsequent functioning of neuronal networks.

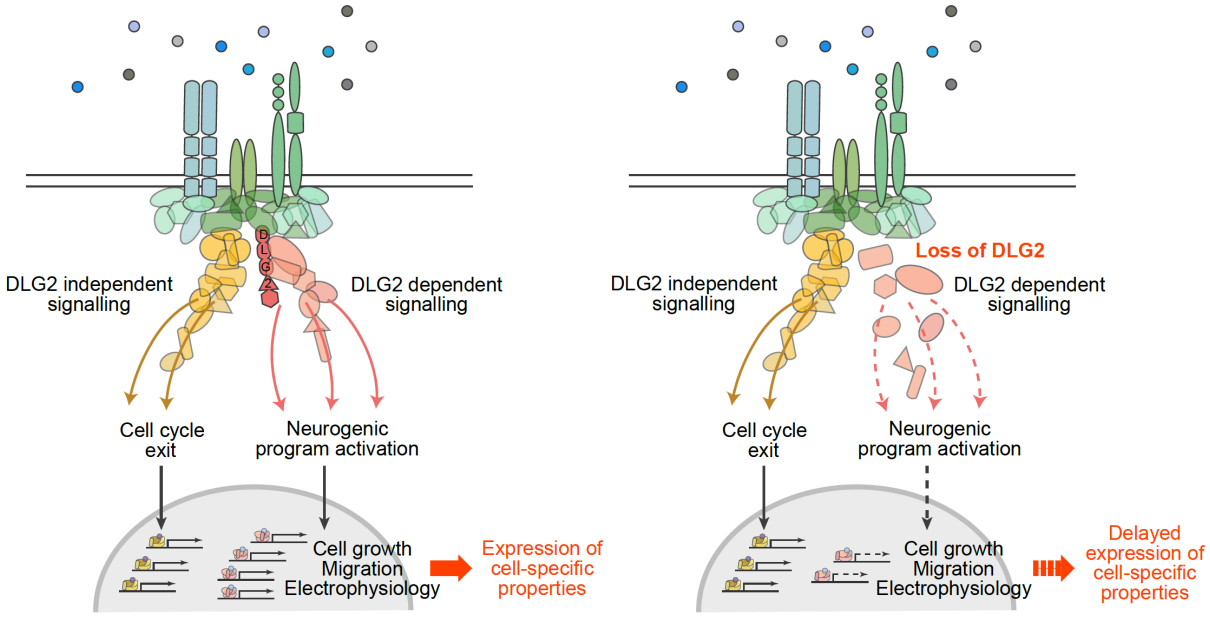


Figure 7.1 DLG2 in cortical neruogenesis
Proposed model of action for DLG2-mediated signalling in neurodevelopment: External cues transduced by DLG2-scaffolded complexes activate neurogenic programs underlying cell growth, migration and development of electrophysiological signalling properties; this 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 signalling that delays expression of cell specific properties. Adapted from (Sanders *et al.*, 2020).

8. References

Abetov, D.A., Kiyan, V.S., Zhylkibayev, A.A., Sarbassova, D.A., Alybayev, S.D., Spooner, E., Song, M.S., Bersimbaev, R.I., and Sarbassov, D.D. (2019). Formation of mammalian preribosomes proceeds from intermediate to composed state during ribosome maturation. J. Biol. Chem. *294*.

Addington, J., and Heinssen, R. (2012). Prediction and Prevention of Psychosis in Youth at Clinical High Risk. Annu. Rev. Clin. Psychol. *8*, 269–289.

Akbarian, S., Viñuela, A., Kim, J.J., Potkin, S.G., Bunney, W.E., and Jones, E.G. (1993). Distorted Distribution of Nicotinamide-Adenine Dinucleotide Phosphate—Diaphorase Neurons in Temporal Lobe of Schizophrenics Implies Anomalous Cortical Development. Arch. Gen. Psychiatry *50*, 178–187.

Altman, D.G., and Bland, J.M. (2003). Interaction revisited: The difference between two estimates. BMJ *326*.

American Psychiatric Association (2013). Diagnostic and statistical manual of mental disorders—fifth edition: DSM-5. (Arlington, VA).

Andrews, S. (2010). FastQC: A Quality Control tool for High Throughput Sequence Data. Available Online at: Https://Www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc/.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. *25*, 25–29.

Bai, F., and Witzmann, F.A. (2007). Synaptosome Proteomics. Subcell. Biochem. *43*, 77–98.

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.

Barbato, A. (1998). Psychiatry in transition: Outcomes of mental health policy shift in Italy. Aust. N. Z. J. Psychiatry *32*, 673–679.

Barch, D.M., and Ceaser, A. (2012). Cognition in schizophrenia: Core psychological and neural mechanisms. Trends Cogn. Sci. *16*, 27–34.

Barnett, D.W., Garrison, E.K., Quinlan, A.R., Stŕmberg, M.P., and Marth, G.T. (2011). Bamtools: A C++ API and toolkit for analyzing and managing BAM files. Bioinformatics *27*.

Bayatti, N., Moss, J.A., Sun, L., Ambrose, P., Ward, J.F.H., Lindsay, S., and Clowry, G.J. (2008). A molecular neuroanatomical study of the developing human neocortex from 8 to 17 postconceptional weeks revealing the early differentiation of the subplate and subventricular zone. Cereb. Cortex *18*, 1536–1548.

Beckervordersandforth, R. (2017). Mitochondrial Metabolism-Mediated Regulation of Adult Neurogenesis. Brain Plast. (Amsterdam, Netherlands) *3*, 73–87.

Benros, M.E., Eaton, W.W., and Mortensen, P.B. (2014). The epidemiologic evidence linking autoimmune diseases and psychosis. Biol. Psychiatry *75*.

Benson, D.L., and Huntley, G.W. (2012). Synapse adhesion: A dynamic equilibrium conferring stability and flexibility. Curr. Opin. Neurobiol. *22*.

Betizeau, M., Cortay, V., Patti, D., Pfister, S., Gautier, E., Bellemin-Ménard, A., Afanassieff, M., Huissoud, C., Douglas, R.J., Kennedy, H., et al. (2013). Precursor Diversity and Complexity of Lineage Relationships in the Outer Subventricular Zone of the Primate. Neuron *80*, 442–457.

Bliim, N., Leshchyns'ka, I., Sytnyk, V., and Janitz, M. (2016). Transcriptional regulation of long-term potentiation. Neurogenetics *17*.

Bliim, N., Leshchyns'ka, I., Keable, R., Chen, B.J., Curry-Hyde, A., Gray, L., Sytnyk, V., and Janitz, M. (2019). Early transcriptome changes in response to chemical long-term potentiation induced via activation of synaptic NMDA receptors in mouse hippocampal neurons. Genomics *111*.

Boeckers, T.M. (2006). The postsynaptic density. Cell Tissue Res. *326*.

Boettcher, M., and McManus, M.T. (2015). Choosing the Right Tool for the Job: RNAi, TALEN, or CRISPR. Mol. Cell *58*, 575–585.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Bond, A.M., Bhalala, O.G., and Kessler, J.A. (2012). The dynamic role of bone morphogenetic proteins in neural stem cell fate and maturation. Dev. Neurobiol. *72*, 1068–1084.

Bosch, M., Castro, J., Saneyoshi, T., Matsuno, H., Sur, M., and Hayashi, Y. (2014). Structural and molecular remodeling of dendritic spine substructures during long-term potentiation. Neuron *82*.

Bossinger, O., Klebes, A., Segbert, C., Theres, C., and Knust, E. (2001). Zonula adherens formation

in Caenorhabditis elegans requires dlg-1, the homologue of the Drosophila gene discs large. Dev. Biol. *230*, 29–42.

Bray, N.J., and O'Donovan, M.C. (2018). The genetics of neuropsychiatric disorders. Brain Neurosci. Adv. *2*, 239821281879927.

Brenman, J.E., Christopherson, K.S., Craven, S.E., McGee, A.W., and Bredt, D.S. (1996). Cloning and characterization of postsynaptic density 93, a nitric oxide synthase interacting protein. J. Neurosci. *16*, 7407–7415.

Brenman, J.E., Rick Topinka, J., Cooper, E.C., McGee, A.W., Rosen, J., Milroy, T., Ralston, H.J., and Bredt, D.S. (1998). Localization of postsynaptic density-93 to dendritic microtubules and interaction with microtubule-associated protein 1A. J. Neurosci. *18*, 8805–8813.

Brennand, K.J., Simone, A., Jou, J., Gelboin-Burkhart, C., Tran, N., Sangar, S., Li, Y., Mu, Y., Chen, G., Yu, D., et al. (2011). Modelling schizophrenia using human induced pluripotent stem cells. Nature *473*, 221–225.

Bressan, R.A., Erlandsson, K., Stone, J.M., Mulligan, R.S., Krystal, J.H., Ell, P.J., and Pilowsky, L.S. (2005). Impact of schizophrenia and chronic antipsychotic treatment on [123I]CNS-1261 binding to N-methyl-D-aspartate receptors in vivo. Biol. Psychiatry *58*, 41–46.

Breunig, J.J., Haydar, T.F., and Rakic, P. (2011). Neural Stem Cells: Historical Perspective and Future Prospects. Neuron *70*, 614–625.

Britanova, O., de Juan Romero, C., Cheung, A., Kwan, K.Y., Schwark, M., Gyorgy, A., Vogel, T., Akopov, S., Mitkovski, M., Agoston, D., et al. (2008). Satb2 Is a Postmitotic Determinant for Upper-Layer Neuron Specification in the Neocortex. Neuron *57*, 378–392.

Broad Institue (2019). Picard Tools. Available Online at: Https://Broadinstitute.Github.Io/Picard/.

Brown, A.S. (2011). The environment and susceptibility to schizophrenia. Prog. Neurobiol. *93*, 23–58.

Brown, A.S. (2012). Epidemiologic studies of exposure to prenatal infection and risk of schizophrenia and autism. Dev. Neurobiol. *72*, 1272–1276.

Bundy, H., Stahl, D., and MacCabe, J.H. (2011). A systematic review and meta-analysis of the fertility of patients with schizophrenia and their unaffected relatives. Acta Psychiatr. Scand. *123*, 98–106.

Byrne, S.M., Mali, P., and Church, G.M. (2014). Genome editing in human stem cells. In Methods in Enzymology, (Academic Press Inc.), pp. 119–138.

Bystron, I., Blakemore, C., and Rakic, P. (2008). Development of the human cerebral cortex: Boulder Committee revisited. Nat. Rev. Neurosci. *9*, 110–122.

Cagnol, S., and Chambard, J.C. (2010). ERK and cell death: Mechanisms of ERK-induced cell death - Apoptosis, autophagy and senescence. FEBS J. *277*.

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.

Cannon, M., Jones, P.B., and Murray, R.M. (2002). Obstetric complications and schizophrenia: Historical and meta-analytic review. Am. J. Psychiatry *159*, 1080–1092.

Cantor-Graae, E., and Pedersen, C.B. (2013). Full spectrum of psychiatric disorders related to foreign migration: A danish population-based cohort study. JAMA Psychiatry *70*.

Cantor-Graae, E., and Selten, J.P. (2005). Schizophrenia and migration: A meta-analysis and review. Am. J. Psychiatry *162*.

Carbon, S., Douglass, E., Dunn, N., Good, B., Harris, N.L., Lewis, S.E., Mungall, C.J., Basu, S., Chisholm, R.L., Dodson, R.J., et al. (2019). The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. *47*, D330–D338.

Cardno, A.G., and Gottesman, I.I. (2000). Twin studies of schizophrenia: From bow-and-arrow concordances to star wars Mx and functional genomics. Am. J. Med. Genet. - Semin. Med. Genet. *97*.

Carlisle, H.J., Fink, A.E., Grant, S.G.N., and O'dell, T.J. (2008). Opposing effects of PSD-93 and PSD-95 on long-term potentiation and spike timing-dependent plasticity. J. Physiol. *586*, 5885–5900.

Caroni, P., Donato, F., and Muller, D. (2012). Structural plasticity upon learning: Regulation and functions. Nat. Rev. Neurosci. *13*.

Catterall, W.A. (2012). Voltage-gated sodium channels at 60: Structure, function and pathophysiology. J. Physiol. *590*, 2577–2589.

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.

Chambers, S.M., Qi, Y., Mica, Y., Lee, G., Zhang, X.J., Niu, L., Bilsland, J., Cao, L., Stevens, E., Whiting, P., et al. (2012). Combined small-molecule inhibition accelerates developmental timing and converts human pluripotent stem cells into nociceptors. Nat. Biotechnol. *30*.

Chen, F., Pruett-Miller, S.M., Huang, Y., Gjoka, M., Duda, K., Taunton, J., Collingwood, T.N., Frodin, M., and Davis, G.D. (2011). High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. Nat. Methods *8*, 753–757.

Chen, X., Levy, J.M., Hou, A., Winters, C., Azzam, R., Sousa, A.A., Leapman, R.D., Nicoll, R.A., and Reese, T.S. (2015). PSD-95 family MAGUKs are essential for anchoring AMPA and NMDA receptor complexes at the postsynaptic density. Proc. Natl. Acad. Sci. U. S. A. *112*.

Cheng, A., Hou, Y., and Mattson, M.P. (2010). Mitochondria and neuroplasticity. ASN Neuro *2*, e00045.

Chesney, E., Goodwin, G.M., and Fazel, S. (2014). Risks of all-cause and suicide mortality in mental disorders: A meta-review. World Psychiatry *13*, 153–160.

Chiang, T.W.W., Le Sage, C., Larrieu, D., Demir, M., and Jackson, S.P. (2016). CRISPR-Cas9D10A nickase-based genotypic and phenotypic screening to enhance genome editing. Sci. Rep. *6*, 1–17.

Choquet, D., and Triller, A. (2013). The dynamic synapse. Neuron *80*.

Cingolani, L.A., and Goda, Y. (2008). Actin in action: The interplay between the actin cytoskeleton and synaptic efficacy. Nat. Rev. Neurosci. *9*.

Clancy, M.J., Clarke, M.C., Connor, D.J., Cannon, M., and Cotter, D.R. (2014). The prevalence of psychosis in epilepsy; a systematic review and meta-analysis. BMC Psychiatry *14*.

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.

Crawford, T.Q., and Roelink, H. (2007). The Notch response inhibitor DAPT enhances neuronal differentiation in embryonic stem cell-derived embryoid bodies independently of Sonic Hedgehog signaling. Dev. Dyn. *236*, 886–892.

Dahoun, T., Trossbach, S. V., Brandon, N.J., Korth, C., and Howes, O.D. (2017). The impact of

Disrupted-in-Schizophrenia 1 (DISC1) on the dopaminergic system: A systematic review. Transl. Psychiatry *7*, e1015.

Dakoji, S., Tomita, S., Karimzadegan, S., Nicoll, R.A., and Bredt, D.S. (2003). Interaction of transmembrane AMPA receptor regulatory proteins with multiple membrane associated guanylate kinases. Neuropharmacology *45*.

Danielson, E., Zhang, N., Metallo, J., Kaleka, K., Shin, S.M., Gerges, N., and Lee, S.H. (2012). S-SCAM/MAGI-2 is an essential synaptic scaffolding molecule for the GluA2-containing maintenance pool of AMPA receptors. J. Neurosci. *32*.

Das, D.K., Tapias, V., D''Aiuto, L., Chowdari, K. V., Francis, L., Zhi, Y., Ghosh, A., Surti, U., Tischfield, J., Sheldon, M., et al. (2015). Genetic and Morphological Features of Human iPSC-Derived Neurons with Chromosome 15q11.2 (BP1-BP2) Deletions. Mol. Neuropsychiatry *1*, 116–123.

Davies, G., Welham, J., Chant, D., Torrey, E.F., and McGrath, J. (2003). A Systematic Review and Meta-analysis of Northern Hemisphere Season of Birth Studies in Schizophrenia. Schizophr. Bull. *29*.

Dehay, C., Kennedy, H., and Kosik, K.S. (2015). The Outer Subventricular Zone and Primate-Specific Cortical Complexification. Neuron *85*, 683–694.

Delint-Ramirez, I., Fernandez, E., Bayes, A., Kicsi, E., Komiyama, N.H., and Grant, S.G.N. (2010). In Vivo Composition of NMDA Receptor Signaling Complexes Differs between Membrane Subdomains and Is Modulated by PSD-95 And PSD-93. J. Neurosci. *30*, 8162–8170.

Derynck, R., and Zhang, Y.E. (2003). Smad-dependent and Smad-independent pathways in TGF-β family signalling. Nature *425*, 577–584.

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.

Duan, W., Zhang, Y.-P., Hou, Z., Huang, C., Zhu, H., Zhang, C.-Q., and Yin, Q. (2016). Novel Insights into NeuN: from Neuronal Marker to Splicing Regulator. Mol. Neurobiol. *53*, 1637–1647.

Duflocq, A., Chareyre, F., Giovannini, M., Couraud, F., and Davenne, M. (2011). Characterization of the axon initial segment (AIS) of motor neurons and identification of a para-AIS and a juxtapara-AIS, organized by protein 4.1B. BMC Biol. *9*.

Eberhardt, M., Salmon, P., von Mach, M.A., Hengstler, J.G., Brulport, M., Linscheid, P., Seboek, D., Oberholzer, J., Barbero, A., Martin, I., et al. (2006). Multipotential nestin and Isl-1 positive mesenchymal stem cells isolated from human pancreatic islets. Biochem. Biophys. Res. Commun. *345*, 1167–1176.

Einhauer, A., and Jungbauer, A. (2001). The FLAG™ peptide, a versatile fusion tag for the purification of recombinant proteins. J. Biochem. Biophys. Methods *49*, 455–465.

Elias, G.M., Funke, L., Stein, V., Grant, S.G., Bredt, D.S., and Nicoll, R.A. (2006). Synapse-Specific and Developmentally Regulated Targeting of AMPA Receptors by a Family of MAGUK Scaffolding Proteins. Neuron *52*, 307–320.

Elsum, I., Yates, L., Humbert, P.O., and Richardson, H.E. (2012). The Scribble–Dlg–Lgl polarity module in development and cancer: from flies to man. Essays Biochem. *53*, 141–168.

Emrich, S.J., Barbazuk, W.B., Li, L., and Schnable, P.S. (2007). Gene discovery and annotation using LCM-454 transcriptome sequencing. Genome Res. *17*, 69–73.

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.

Falk, A., Heine, V.M., Harwood, A.J., Sullivan, P.F., Peitz, M., Brüstle, O., Shen, S., Sun, Y.M., Glover, J.C., Posthuma, D., et al. (2016). Modeling psychiatric disorders: From genomic findings to cellular phenotypes. Mol. Psychiatry *21*, 1167–1179.

Farkas, L.M., and Huttner, W.B. (2008). The cell biology of neural stem and progenitor cells and its significance for their proliferation versus differentiation during mammalian brain development. Curr. Opin. Cell Biol. *20*, 707–715.

Farrell, M., Lichtenstein, M., Harner, M.K., Crowley, J.J., Filmyer, D.M., Lázaro-Muñoz, G., Dietterich, T.E., Bruno, L.M., Shaughnessy, R.A., Biondi, T.F., et al. (2020). Treatment-resistant psychotic symptoms and the 15q11.2 BP1–BP2 (Burnside-Butler) deletion syndrome: case report and review of the literature. Transl. Psychiatry *10*, 1–8.

Favaro, P.D., Huang, X., Hosang, L., Stodieck, S., Cui, L., Liu, Y.Z., Engelhardt, K.A., Schmitz, F., Dong, Y., Löwel, S., et al. (2018). An opposing function of paralogs in balancing developmental synapse maturation. PLoS Biol. *16*.

Feng, W., and Zhang, M. (2009). Organization and dynamics of PDZ-domain-related supramodules in the postsynaptic density. Nat. Rev. Neurosci. *10*, 87–99.

Flaherty, E., Deranieh, R.M., Artimovich, E., Lee, I.S., Siegel, A.J., Levy, D.L., Nestor, M.W., and Brennand, K.J. (2017). Patient-derived hiPSC neurons with heterozygous CNTNAP2 deletions display altered neuronal gene expression and network activity. Npj Schizophr. *3*, 1–4.

Florio, M., and Huttner, W.B. (2014). Neural progenitors, neurogenesis and the evolution of the neocortex. Dev. *141*, 2182–2194.

Folmes, C.D.L., and Terzic, A. (2016). Energy metabolism in the acquisition and maintenance of stemness. Semin. Cell Dev. Biol. *52*, 68–75.

Foss-Feig, J.H., Adkinson, B.D., Ji, J.L., Yang, G., Srihari, V.H., McPartland, J.C., Krystal, J.H., Murray, J.D., and Anticevic, A. (2017). Searching for Cross-Diagnostic Convergence: Neural Mechanisms Governing Excitation and Inhibition Balance in Schizophrenia and Autism Spectrum Disorders. Biol. Psychiatry *81*, 848–861.

Franco, M., and Carmena, A. (2019). Eph signaling controls mitotic spindle orientation and cell proliferation in neuroepithelial cells. J. Cell Biol. *218*.

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.

Frantz, C., Stewart, K.M., and Weaver, V.M. (2010). The extracellular matrix at a glance. J. Cell Sci. *123*.

Friedrichs, F., Henckaerts, L., Vermeire, S., Kucharzik, T., Seehafer, T., Möller-Krull, M., Bornberg-Bauer, E., Stoll, M., and Weiner, J. (2008). The Crohn's disease susceptibility gene DLG5 as a member of the CARD interaction network. J. Mol. Med. *86*.

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.

Funke, L., Dakoji, S., and Bredt, D.S. (2005). Membrane-Associated Guanylate Kinases Regulate Adhesion and Plasticity At Cell Junctions. Annu. Rev. Biochem. *74*, 219–245.

Gage, S.H., Jones, H.J., Burgess, S., Bowden, J., Davey Smith, G., Zammit, S., and Munafò, M.R.

(2017). Assessing causality in associations between cannabis use and schizophrenia risk: A two-sample Mendelian randomization study. Psychol. Med. *47*, 971–980.

Gámez, B., Rodriguez-Carballo, E., and Ventura, F. (2013). BMP signaling in telencephalic neural cell specification and maturation. Front. Cell. Neurosci.

Garey, L.J., Ong, W.Y., Patel, T.S., Kanani, M., Davis, A., Mortimer, A.M., Barnes, T.R.E., and Hirsch, S.R. (1998). Reduced dendritic spine density on cerebral cortical pyramidal neurons in schizophrenia. J. Neurol. Neurosurg. Psychiatry *65*, 446–453.

Geller, J.L. (1992). A historical perspective on the role of state hospitals viewed from the era of the "revolving door." Am. J. Psychiatry *149*, 1526–1533.

Gelman, D.M., and Marín, O. (2010). Generation of interneuron diversity in the mouse cerebral cortex. Eur. J. Neurosci. *31*, 2136–2141.

Genevet, A., and Tapon, N. (2011). The Hippo pathway and apico–basal cell polarity. Biochem. J. *436*, 213–224.

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.

Gilmore, E.G., and Herrup, K. (1997). Cortical development: Layers of complexity. Curr. Biol. *7*.

Glantz, L.A., and Lewis, D.A. (2000). Decreased dendritic spine density on prefrontal cortical pyramidal neurons in schizophrenia. Arch. Gen. Psychiatry *57*, 65–73.

Glausier, J.R., and Lewis, D.A. (2013). Dendritic spine pathology in schizophrenia. Neuroscience *251*, 90–107.

González, F., Zhu, Z., Shi, Z.D., Lelli, K., Verma, N., Li, Q. V., and Huangfu, D. (2014). An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. Cell Stem Cell *15*, 215–226.

Grande, I., Berk, M., Birmaher, B., and Vieta, E. (2016). Bipolar disorder. Lancet *387*, 1561–1572.

Grant, S.G.N. (2012). Synaptopathies: Diseases of the synaptome. Curr. Opin. Neurobiol. *22*.

Guillin, O., Abi-Dargham, A., and Laruelle, M. (2007). Neurobiology of Dopamine in Schizophrenia. Int. Rev. Neurobiol. *78*, 1–39.

Guo, M.L., Xue, B., Jin, D.Z., Mao, L.M., and Wang, J.Q. (2012). Interactions and phosphorylation of postsynaptic density 93 (PSD-93) by extracellular signal-regulated kinase (ERK). Brain Res. *1465*, 18–25.

Hafner, A.S., Penn, A.C., Grillo-Bosch, D., Retailleau, N., Poujol, C., Philippat, A., Coussen, F., Sainlos, M., Opazo, P., and Choquet, D. (2015). Lengthening of the stargazin cytoplasmic tail increases synaptic transmission by promoting interaction to deeper domains of PSD-95. Neuron *86*.

Haijma, S. V., Van Haren, N., Cahn, W., Koolschijn, P.C.M.P., Hulshoff Pol, H.E., and Kahn, R.S. (2013). Brain volumes in schizophrenia: A meta-analysis in over 18 000 subjects. Schizophr. Bull. *39*.

Hall, J., Trent, S., Thomas, K.L., O'Donovan, M.C., and Owen, M.J. (2015). Genetic risk for schizophrenia: Convergence on synaptic pathways involved in plasticity. Biol. Psychiatry *77*.

Hamshere, M.L., Stergiakouli, E., Langley, K., Martin, J., Holmans, P., Kent, L., Owen, M.J., Gill, M., Thapar, A., O'Donovan, M., et al. (2013). Shared polygenic contribution between childhood attention-deficit hyperactivity disorder and adult schizophrenia. Br. J. Psychiatry *203*.

Hansen, D. V., Lui, J.H., Parker, P.R.L., and Kriegstein, A.R. (2010). Neurogenic radial glia in the outer subventricular zone of human neocortex. Nature *464*, 554–561.

Hansen, D. V., Lui, J.H., Flandin, P., Yoshikawa, K., Rubenstein, J.L., Alvarez-Buylla, A., and Kriegstein, A.R. (2013). Non-epithelial stem cells and cortical interneuron production in the human ganglionic eminences. Nat. Neurosci. *16*, 1576–1587.

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.

Hartberg, C.B., Sundet, K., Rimol, L.M., Haukvik, U.K., Lange, E.H., Nesvåg, R., Melle, I., Andreassen, O.A., and Agartz, I. (2011). Subcortical brain volumes relate to neurocognition in schizophrenia and bipolar disorder and healthy controls. Prog. Neuro-Psychopharmacology Biol. Psychiatry *35*, 1122–1130.

Haubensak, W., Attardo, A., Denk, W., and Huttner, W.B. (2004). Neurons arise in the basal

neuroepithelium of the early mammalian telencephalon: A major site of neurogenesis. Proc. Natl. Acad. Sci. U. S. A. *101*, 3196–3201.

Hegarty, S. V., O'Keeffe, G.W., and Sullivan, A.M. (2013). BMP-Smad 1/5/8 signalling in the development of the nervous system. Prog. Neurobiol. *109*, 28–41.

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.

Hilker, R., Helenius, D., Fagerlund, B., Skytthe, A., Christensen, K., Werge, T.M., Nordentoft, M., and Glenthøj, B. (2018). Heritability of Schizophrenia and Schizophrenia Spectrum Based on the Nationwide Danish Twin Register. Biol. Psychiatry *83*, 492–498.

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.

Hodge, R.D., Bakken, T.E., Miller, J.A., Smith, K.A., Barkan, E.R., Graybuck, L.T., Close, J.L., Long, B., Johansen, N., Penn, O., et al. (2019). Conserved cell types with divergent features in human versus mouse cortex. Nature *573*, 61–68.

Hoffman, G.E., Schrode, N., Flaherty, E., and Brennand, K.J. (2019). New considerations for hiPSC-based models of neuropsychiatric disorders. Mol. Psychiatry *24*, 49–66.

Hoffmann, A., Ziller, M., and Spengler, D. (2019). Progress in iPSC-Based Modeling of Psychiatric Disorders. Int. J. Mol. Sci. *20*, 4896.

Horlbeck, M.A., Witkowsky, L.B., Guglielmi, B., Replogle, J.M., Gilbert, L.A., Villalta, J.E., Torigoe, S.E., Tjian, R., and Weissman, J.S. (2016). Nucleosomes impede cas9 access to DNA in vivo and in vitro. Elife *5*.

Hornung, V., and Latz, E. (2010). Intracellular DNA recognition. Nat. Rev. Immunol. *10*, 123–130.

Horresh, I., Poliak, S., Grant, S., Bredt, D., Rasband, M.N., and Peles, E. (2008). Multiple molecular interactions determine the clustering of Caspr2 and Kv1 channels in myelinated axons. J. Neurosci. *28*.

Hotulainen, P., and Hoogenraad, C.C. (2010). Actin in dendritic spines: Connecting dynamics to function. J. Cell Biol. *189*.

Howes, O.D., and Murray, R.M. (2014). Schizophrenia: An integrated sociodevelopmental-

cognitive model. Lancet *383*, 1677–1687.

Hsueh, Y.P., Wang, T.F., Yang, F.C., and Sheng, M. (2000). Nuclear translocation and transcription regulation by the membrane- associated guanylate kinase CASK/LIN-2. Nature *404*.

Huang, T.-N., and Hsueh, Y.-P. (2017). Calcium/calmodulin-dependent serine protein kinase (CASK), a protein implicated in mental retardation and autism-spectrum disorders, interacts with T-Brain-1 (TBR1) to control extinction of associative memory in male mice. J. Psychiatry Neurosci. *42*, 37–47.

Huganir, R.L., and Nicoll, R.A. (2013). AMPARs and synaptic plasticity: The last 25 years. Neuron *80*.

Humbert, P.O., Grzeschik, N.A., Brumby, A.M., Galea, R., Elsum, I., and Richardson, H.E. (2008). Control of tumourigenesis by the Scribble/Dlg/Lgl polarity module. Oncogene *27*.

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.

Hutton, S.R., and Pevny, L.H. (2011). SOX2 expression levels distinguish between neural progenitor populations of the developing dorsal telencephalon. Dev. Biol. *352*, 40–47.

Iida, J., Hirabayashi, S., Sato, Y., and Hata, Y. (2004). Synaptic scaffolding molecule is involved in the synaptic clustering of neuroligin. Mol. Cell. Neurosci. *27*.

Ingason, A., Giegling, I., Hartmann, A.M., Genius, J., Konte, B., Friedl, M., Schizophrenia Working Group of the Psychiatric Genomics Consortium (PGC), S.W.G. of the P.G.C., Ripke, S., Sullivan, P.F., St Clair, D., et al. (2015). Expression analysis in a rat psychosis model identifies novel candidate genes validated in a large case-control sample of schizophrenia. Transl. Psychiatry *5*, e656.

Irie, M., Hata, Y., Takeuchi, M., Ichtchenko, K., Toyoda, A., Hirao, K., Takai, Y., Rosahl, T.W., and Südhof, T.C. (1997). Binding of neuroligins to PSD-95. Science (80-.). *277*.

Itoh, Y., Tyssowski, K., and Gotoh, Y. (2013). Transcriptional coupling of neuronal fate commitment and the onset of migration. Curr. Opin. Neurobiol. *23*, 957–964.

Ivanova, S., Gregorc, U., Vidergar, N., Javier, R., Bredt, D.S., Vandenabeele, P., Pardo, J., Simon, M.N., Turk, V., Banks, L., et al. (2011). MAGUKs, scaffolding proteins at cell junctions, are substrates of different proteases during apoptosis. Cell Death Dis. *2*, e116–e116.

Jaaro-Peled, H., Ayhan, Y., Pletnikov, M. V., and Sawa, A. (2010). Review of pathological hallmarks of schizophrenia: Comparison of genetic models with patients and nongenetic models. Schizophr. Bull. *36*, 301–313.

Javitt, D.C., and Zukin, S.R. (1991). Recent advances in the phencyclidine model of schizophrenia. Am. J. Psychiatry *148*, 1301–1308.

Johnstone, M., Thomson, P.A., Hall, J., McIntosh, A.M., Lawrie, S.M., and Porteous, D.J. (2011). DISC1 in schizophrenia: Genetic mouse models and human genomic imaging. Schizophr. Bull. *37*, 14–20.

Joyce, E.M., and Roiser, J.P. (2007). Cognitive heterogeneity in schizophrenia. Curr. Opin. Psychiatry *20*, 268–272.

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.

Kamitaki, N., Sekar, A., Handsaker, R.E., de Rivera, H., Tooley, K., Morris, D.L., Taylor, K.E., Whelan, C.W., Tombleson, P., Loohuis, L.M.O., et al. (2020). Complement genes contribute sex-biased vulnerability in diverse disorders. Nature *582*, 577–581.

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.

Kantrowitz, J.T., and Javitt, D.C. (2010). N-methyl-d-aspartate (NMDA) receptor dysfunction or dysregulation: The final common pathway on the road to schizophrenia? Brain Res. Bull. *83*, 108–121.

Kashyap, V., Rezende, N.C., Scotland, K.B., Shaffer, S.M., Persson, J.L., Gudas, L.J., and Mongan, N.P. (2009). Regulation of Stem cell pluripotency and differentiation involves a mutual regulatory circuit of the Nanog, OCT4, and SOX2 pluripotency transcription factors with polycomb Repressive Complexes and Stem Cell microRNAs. Stem Cells Dev. *18*, 1093–1108.

Keane, S., Améen, S., Lindlöf, A., and Ejeskär, K. (2020). Low DLG2 gene expression, a link between 11q-deleted and MYCN-amplified neuroblastoma, causes forced cell cycle progression, and predicts poor patient survival. Cell Commun. Signal. *18*.

Kemp, P.J., Rushton, D.J., Yarova, P.L., Schnell, C., Geater, C., Hancock, J.M., Wieland, A., Hughes, A., Badder, L., Cope, E., et al. (2016). Improving and accelerating the differentiation and functional maturation of human stem cell-derived neurons: role of extracellular calcium and GABA. J. Physiol. *594*, 6583–6594.

Kerchner, G.A., and Nicoll, R.A. (2008). Silent synapses and the emergence of a postsynaptic mechanism for LTP. Nat. Rev. Neurosci. *9*.

Khandaker, G.M., Zimbron, J., Dalman, C., Lewis, G., and Jones, P.B. (2012). Childhood infection and adult schizophrenia: A meta-analysis of population-based studies. Schizophr. Res. *139*.

Khandaker, G.M., Zimbron, J., Lewis, G., and Jones, P.B. (2013). Prenatal maternal infection, neurodevelopment and adult schizophrenia: A systematic review of population-based studies. Psychol. Med. *43*, 239–257.

Khashan, A.S., Abel, K.M., McNamee, R., Pedersen, M.G., Webb, R.T., Baker, P.N., Kenny, L.C., and Mortensen, P.B. (2008). Higher risk of offspring schizophrenia following antenatal maternal exposure to severe adverse life events. Arch. Gen. Psychiatry *65*, 146–152.

Kim, E., Cho, K.O., Rothschild, A., and Sheng, M. (1996). Heteromultimerization and NMDA receptor-clustering activity of Chapsyn- 110, a member of the PSD-95 family of proteins. Neuron *17*, 103–113.

Kim, E., DeMarco, S.J., Marfatia, S.M., Chishti, A.H., Sheng, M., and Strehler, E.E. (1998a). Plasma membrane Ca2+ ATPase isoform 4B binds to membrane-associated guanylate kinase (MAGUK) proteins via their PDZ (PSD-95/Dlg/ZO-1) domains. J. Biol. Chem. *273*.

Kim, J.H., Liao, D., Lau, L.F., and Huganir, R.L. (1998b). SynGAP: A synaptic RasGAP that associates with the PSD-95/SAP90 protein family. Neuron *20*.

Kim, S., Kim, D., Cho, S.W., Kim, J., and Kim, J.S. (2014). Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome Res. *24*, 1012–1019.

Kirov, G., Rujescu, D., Ingason, A., Collier, D.A., O'Donovan, M.C., and Owen, M.J. (2009).

Neurexin 1 (NRXN1) deletions in schizophrenia. Schizophr. Bull. *35*.

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

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

Kirov, G., Rees, E., Walters, J.T.R., Escott-Price, V., Georgieva, L., Richards, A.L., Chambert, K.D., Davies, G., Legge, S.E., Moran, J.L., et al. (2014). The penetrance of copy number variations for schizophrenia and developmental delay. Biol. Psychiatry *75*.

Kooyman, I., Dean, K., Harvey, S., and Walsh, E. (2007). Outcomes of public concern in schizophrenia. Br. J. Psychiatry *191*.

Kornau, H.C., Schenker, L.T., Kennedy, M.B., and Seeburg, P.H. (1995). Domain interaction between NMDA receptor subunits and the postsynaptic density protein PSD-95. Science (80-.). *269*.

Kosmicki, J.A., Samocha, K.E., Howrigan, D.P., Sanders, S.J., Slowikowski, K., Lek, M., Karczewski, K.J., Cutler, D.J., Devlin, B., Roeder, K., et al. (2017). Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. Nat. Genet. *49*.

Kowalczyk, T., Pontious, A., Englund, C., Daza, R.A.M., Bedogni, F., Hodge, R., Attardo, A., Bell, C., Huttner, W.B., and Hevner, R.F. (2009). Intermediate neuronal progenitors (basal progenitors) produce pyramidal-projection neurons for all layers of cerebral cortex. Cereb. Cortex *19*.

Kristiansen, L. V., Beneyto, M., Haroutunian, V., and Meador-Woodruff, J.H. (2006). Changes in NMDA receptor subunits and interacting PSD proteins in dorsolateral prefrontal and anterior cingulate cortex indicate abnormal regional expression in schizophrenia. Mol. Psychiatry *11*.

Kruger, J.M., Favaro, P.D., Liu, M., Kitlinska, A., Huang, X., Raabe, M., Akad, D.S., Liu, Y., Urlaub, H., Dong, Y., et al. (2013). Differential Roles of Postsynaptic Density-93 Isoforms in Regulating Synaptic Transmission. J. Neurosci. *33*, 15504–15517.

Krystal, J.H., Karper, L.P., Seibyl, J.P., Freeman, G.K., Delaney, R., Bremner, J.D., Heninger, G.R., Bowers, M.B., and Charney, D.S. (1994). Subanesthetic Effects of the Noncompetitive NMDA Antagonist, Ketamine, in Humans: Psychotomimetic, Perceptual, Cognitive, and Neuroendocrine Responses. Arch. Gen. Psychiatry *51*, 199–214.

Krystal, J.H., Bennett, A., Abi-Saab, D., Belger, A., Karper, L.P., D'Souza, D.C., Lipschitz, D., Abi-Dargham, A., and Charney, D.S. (2000). Dissociation of ketamine effects on rule acquisition and rule implementation: Possible relevance to NMDA receptor contributions to executive cognitive functions. Biol. Psychiatry *47*, 137–143.

Kumamoto, T., and Hanashima, C. (2014). Neuronal subtype specification in establishing mammalian neocortical circuits. Neurosci. Res. *86*, 37–49.

Kumar, P., Jha, N.K., Jha, S.K., Ramani, K., and Ambasta, R.K. (2015). Tau phosphorylation, molecular chaperones, and ubiquitin E3 Ligase: Clinical relevance in Alzheimer's disease. J. Alzheimer's Dis. *43*.

Kuo, T.Y., Hong, C.J., Chien, H.L., and Hsueh, Y.P. (2010). X-linked mental retardation gene CASK interacts with Bcl11A/CTIP1 and regulates axon branching and outgrowth. J. Neurosci. Res. *88*.

Kwan, J., Sczaniecka, A., Arash, E.H., Nguyen, L., Chen, C.C., Ratkovic, S., Klezovitch, O., Attisano, L., McNeill, H., Emili, A., et al. (2016). DLG5 connects cell polarity and Hippo signaling protein networks by linking PAR-1 with MST1/2. Genes Dev. *30*, 2696–2709.

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.

Lancaster, M.A., and Knoblich, J.A. (2012). Spindle orientation in mammalian cerebral cortical development. Curr. Opin. Neurobiol. *22*, 737–746.

Laurent, L.C., Ulitsky, I., Slavin, I., Tran, H., Schork, A., Morey, R., Lynch, C., Harness, J. V., Lee, S., Barrero, M.J., et al. (2011). Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture. Cell Stem Cell *8*, 106–118.

Laursen, T.M., and Munk-Olsen, T. (2010). Reproductive patterns in psychotic patients. Schizophr. Res. *121*, 234–240.

Lee, I.S., Carvalho, C.M.B., Douvaras, P., Ho, S.M., Hartley, B.J., Zuccherato, L.W., Ladran, I.G., Siegel, A.J., McCarthy, S., Malhotra, D., et al. (2015). Characterization of molecular and cellular phenotypes associated with a heterozygous CNTNAP2 deletion using patient-derived hiPSC neural cells. Npj Schizophr. *1*, 1–5.

Lee, S.H., Ripke, S., Neale, B.M., Faraone, S. V., Purcell, S.M., Perlis, R.H., Mowry, B.J., Thapar, A., Goddard, M.E., Witte, J.S., et al. (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat. Genet. *45*.

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.

van de Leemput, J., Hess, J.L., Glatt, S.J., and Tsuang, M.T. (2016). Genetics of Schizophrenia: Historical Insights and Prevailing Evidence. Adv. Genet. *96*, 99–141.

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.

De Lera Ruiz, M., and Kraus, R.L. (2015). Voltage-Gated Sodium Channels: Structure, Function, Pharmacology, and Clinical Indications. J. Med. Chem. *58*, 7093–7118.

Leucht, S., Tardy, M., Komossa, K., Heres, S., Kissling, W., Salanti, G., and Davis, J.M. (2012). Antipsychotic drugs versus placebo for relapse prevention in schizophrenia: A systematic review and meta-analysis. Lancet *379*, 2063–2071.

Levers, T.E., Edgar, J.M., and Price, D.J. (2001). The fates of cells generated at the end of neurogenesis in developing mouse cortex. J. Neurobiol. *48*, 265–277.

Levy, J.M., Chen, X., Reese, T.S., and Nicoll, R.A. (2015). Synaptic Consolidation Normalizes AMPAR Quantal Size following MAGUK Loss. Neuron *87*, 534–548.

Lewis, D.A. (2012). Cortical circuit dysfunction and cognitive deficits in schizophrenia - implications for preemptive interventions. Eur. J. Neurosci. *35*, 1871–1878.

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.

Li, S., Mattar, P., Dixit, R., Lawn, S.O., Wilkinson, G., Kinch, C., Eisenstat, D., Kurrasch, D.M., Chan, J.A., and Schuurmans, C. (2014a). RAS/ERK signaling controls proneural genetic programs in cortical development and gliomagenesis. J. Neurosci. *34*, 2169–2190.

Li, Y., Wei, Z., Yan, Y., Wan, Q., Du, Q., and Zhang, M. (2014b). Structure of Crumbs tail in complex with the PALS1 PDZ-SH3-GK tandem reveals a highly specific assembly mechanism for the apical Crumbs complex. Proc. Natl. Acad. Sci. U. S. A. *111*.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

Liddle, P.F. (1987). The symptoms of chronic schizophrenia. A re-examination of the positive-negative dichotomy. Br. J. Psychiatry *151*, 145–151.

Lieberman, J.A., Perkins, D., Belger, A., Chakos, M., Jarskog, F., Boteva, K., and Gilmore, J. (2001). The early stages of schizophrenia: Speculations on pathogenesis, pathophysiology, and therapeutic approaches. In Biological Psychiatry, (Biol Psychiatry), pp. 884–897.

Lin, M., Pedrosa, E., Hrabovsky, A., Chen, J., Puliafito, B.R., Gilbert, S.R., Zheng, D., and Lachman, H.M. (2016). Integrative transcriptome network analysis of iPSC-derived neurons from schizophrenia and schizoaffective disorder patients with 22q11.2 deletion. BMC Syst. Biol. *10*.

Linden, D.E.J. (2012). The Challenges and Promise of Neuroimaging in Psychiatry. Neuron *73*, 8–22.

Liu, P., Kaplan, A., Yuan, B., Hanna, J.H., Lupski, J.R., and Reiner, O. (2014). Passage number is a major contributor to genomic structural variations in mouse iPSCs. Stem Cells *32*, 2657–2667.

Liu, Y., Beyer, A., and Aebersold, R. (2016). On the Dependency of Cellular Protein Levels on mRNA Abundance. Cell *165*, 535–550.

Lodato, S., and Arlotta, P. (2015). Generating Neuronal Diversity in the Mammalian Cerebral Cortex. Annu. Rev. Cell Dev. Biol. *31*.

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.

Lord, C., Elsabbagh, M., Baird, G., and Veenstra-Vanderweele, J. (2018). Autism spectrum disorder. Lancet *392*, 508–520.

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.

Luhmann, H.J., Sinning, A., Yang, J.W., Reyes-Puerta, V., Stüttgen, M.C., Kirischuk, S., and Kilb, W. (2016). Spontaneous neuronal activity in developing neocortical networks: From single cells to large-scale interactions. Front. Neural Circuits *10*.

MacLaren, E.J., Charlesworth, P., Coba, M.P., and Grant, S.G.N. (2011). Knockdown of mental disorder susceptibility genes disrupts neuronal network physiology in vitro. Mol. Cell. Neurosci. *47*, 93–99.

Maden, M. (2007). Retinoic acid in the development, regeneration and maintenance of the nervous system. Nat. Rev. Neurosci. *8*.

Malhotra, D., and Sebat, J. (2012). CNVs: Harbingers of a rare variant revolution in psychiatric genetics. Cell *148*.

Malinow, R., and Malenka, R.C. (2002). AMPA Receptor Trafficking and Synaptic Plasticity. Annu. Rev. Neurosci. *25*.

Mao, Y., Ge, X., Frank, C.L., Madison, J.M., Koehler, A.N., Doud, M.K., Tassa, C., Berry, E.M., Soda, T., Singh, K.K., et al. (2009). Disrupted in Schizophrenia 1 Regulates Neuronal Progenitor Proliferation via Modulation of GSK3β/β-Catenin Signaling. Cell *136*, 1017–1031.

Maroof, A.M., Keros, S., Tyson, J.A., Ying, S.W., Ganat, Y.M., Merkle, F.T., Liu, B., Goulburn, A., Stanley, E.G., Elefanty, A.G., et al. (2013). Directed differentiation and functional maturation of cortical interneurons from human embryonic stem cells. Cell Stem Cell *12*, 559–572.

Marshall, C.R., Howrigan, D.P., Merico, D., Thiruvahindrapuram, B., Wu, W., Greer, D.S., Antaki, D., Shetty, A., Holmans, P.A., Pinto, D., et al. (2017). Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. Nat. Genet. *49*, 27–35.

Marwaha, S., and Johnson, S. (2004). Schizophrenia and employment: A review. Soc. Psychiatry Psychiatr. Epidemiol. *39*, 337–349.

Mason, J.O., and Price, D.J. (2016). Building brains in a dish: Prospects for growing cerebral organoids from stem cells. Neuroscience *334*, 105–118.

Massimi, P., Narayan, N., Thomas, M., Gammoh, N., Strand, S., Strand, D., and Banks, L. (2008). Regulation of the hDlg/hScrib/Hugl-1 tumour suppressor complex. Exp. Cell Res. *314*.

Mathew, D., Gramates, L.S., Packard, M., Thomas, U., Bilder, D., Perrimon, N., Gorczyca, M., and Budnik, V. (2002). Recruitment of Scribble to the synaptic scaffolding complex requires GUK-holder, a novel DLG binding protein. Curr. Biol. *12*.

McDonald, C., Grech, A., Toulopoulou, T., Schulze, K., Chapple, B., Sham, P., Walshe, M., Sharma, T., Sigmundsson, T., Chitnis, X., et al. (2002). Brain volumes in familial and non-familial schizophrenic probands and their unaffected relatives. Am. J. Med. Genet. - Neuropsychiatr. Genet. *114*, 616–625.

McGee, A.W., Dakoji, S.R., Olsen, O., Bredt, D.S., Lim, W.A., and Prehoda, K.E. (2001). Structure of the SH3-guanylate kinase module from PSD-95 suggests a mechanism for regulated assembly of MAGUK scaffolding proteins. Mol. Cell *8*.

McGrath, J., Saha, S., Chant, D., and Welham, J. (2008). Schizophrenia: A concise overview of incidence, prevalence, and mortality. Epidemiol. Rev. *30*.

McGrath, J.J., Eyles, D.W., Pedersen, C.B., Anderson, C., Ko, P., Burne, T.H., Norgaard-Pedersen, B., Hougaard, D.M., and Mortensen, P.B. (2010). Neonatal vitamin D status and risk of schizophrenia: A population-based case-control study. Arch. Gen. Psychiatry *67*, 889–894.

McGrath, J.J., Petersen, L., Agerbo, E., Mors, O., Mortensen, P.B., and Pedersen, C.B. (2014). A comprehensive assessment of parental age and psychiatric disorders. JAMA Psychiatry *71*.

Mertens, J., Marchetto, M.C., Bardy, C., and Gage, F.H. (2016). Evaluating cell reprogramming, differentiation and conversion technologies in neuroscience. Nat. Rev. Neurosci. *17*, 424–437.

Mi, D., Li, Z., Lim, L., Li, M., Moissidis, M., Yang, Y., Gao, T., Hu, T.X., Pratt, T., Price, D.J., et al. (2018). Early emergence of cortical interneuron diversity in the mouse embryo. Science *360*, 81–85.

Michaelovsky, E., Carmel, M., Frisch, A., Salmon-Divon, M., Pasmanik-Chor, M., Weizman, A., and Gothelf, D. (2019). Risk gene-set and pathways in 22q11.2 deletion-related schizophrenia: a genealogical molecular approach. Transl. Psychiatry *9*, 1–9.

Migliore, M., and Shepherd, G.M. (2005). Opinion: An integrated approach to classifying neuronal phenotypes. Nat. Rev. Neurosci. *6*.

Millan, M.J., Andrieux, A., Bartzokis, G., Cadenhead, K., Dazzan, P., Fusar-Poli, P., Gallinat, J., Giedd, J., Grayson, D.R., Heinrichs, M., et al. (2016). Altering the course of schizophrenia: Progress and perspectives. Nat. Rev. Drug Discov. *15*, 485–515.

Miller, B., Messias, E., Miettunen, J., Alaräisänen, A., Järvelin, M.R., Koponen, H., Räsänen, P., Isohanni, M., and Kirkpatrick, B. (2011). Meta-analysis of paternal age and schizophrenia risk in male versus female offspring. Schizophr. Bull. *37*.

Miura, H., Gurumurthy, C.B., Sato, T., Sato, M., and Ohtsuka, M. (2015). CRISPR/Cas9-based generation of knockdown mice by intronic insertion of artificial microRNA using longer single-stranded DNA. Sci. Rep. *5*.

Miyata, T., Kawaguchi, A., Saito, K., Kawano, M., Muto, T., and Ogawa, M. (2004). Asymmetric production of surface-dividing and non-surface-dividing cortical progenitor cells. Development *131*, 3133–3145.

Moghaddam, B., and Javitt, D. (2012). From revolution to evolution: The glutamate hypothesis of schizophrenia and its implication for treatment. Neuropsychopharmacology *37*, 4–15.

Molyneaux, B.J., Arlotta, P., Menezes, J.R.L., and Macklis, J.D. (2007). Neuronal subtype specification in the cerebral cortex. Nat. Rev. Neurosci. *8*, 427–437.

Moslem, M., Olive, J., and Falk, A. (2019). Stem cell models of schizophrenia, what have we learned and what is the potential? Schizophr. Res. *210*, 3–12.

Mueser, K.T., and McGurk, S.R. (2004). Schizophrenia. In Lancet, (Lancet), pp. 2063–2072.

Mukherjee, K., Sharma, M., Urlaub, H., Bourenkov, G.P., Jahn, R., Südhof, T.C., and Wahl, M.C. (2008). CASK Functions as a Mg2+-Independent Neurexin Kinase. Cell *133*.

Mullen, R.J., Buck, C.R., and Smith, A.M. (1992). NeuN, a neuronal specific nuclear protein in vertebrates. Development *116*, 201–211.

Murai, K., Sun, G., Ye, P., Tian, E., Yang, S., Cui, Q., Sun, G., Trinh, D., Sun, O., Hong, T., et al. (2016). The TLX-miR-219 cascade regulates neural stem cell proliferation in neurodevelopment and schizophrenia iPSC model. Nat. Commun. *7*.

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

Murray, C.J.L., Lopez, A.D., World Health Organization, World Bank, and Harvard School of Public

Health (1996). The Global burden of disease : a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020.

Nebel, R.A., Zhao, D., Pedrosa, E., Kirschen, J., Lachman, H.M., Zheng, D., and Abrahams, B.S. (2016). Reduced CYFIP1 in human neural progenitors results in dysregulation of schizophrenia and epilepsy gene networks. PLoS One *11*.

Nehme, R., Zuccaro, E., Ghosh, S.D., Li, C., Sherwood, J.L., Pietilainen, O., Barrett, L.E., Limone, F., Worringer, K.A., Kommineni, S., et al. (2018). Combining NGN2 Programming with Developmental Patterning Generates Human Excitatory Neurons with NMDAR-Mediated Synaptic Transmission. Cell Rep. *23*, 2509–2523.

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.

Nithianantharajah, J., Komiyama, N.H., McKechanie, A., Johnstone, M., Blackwood, D.H., Clair, D.S., Emes, R.D., Van De Lagemaat, L.N., Saksida, L.M., Bussey, T.J., et al. (2013). Synaptic scaffold evolution generated components of vertebrate cognitive complexity. Nat. Neurosci. *16*, 16–24.

Nowakowski, T.J., Pollen, A.A., Sandoval-Espinosa, C., and Kriegstein, A.R. (2016). Transformation of the Radial Glia Scaffold Demarcates Two Stages of Human Cerebral Cortex Development. Neuron *91*.

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'Brien, H.E., Hannon, E., Hill, M.J., Toste, C.C., Robertson, M.J., Morgan, J.E., McLaughlin, G., Lewis, C.M., Schalkwyk, L.C., Hall, L.S., et al. (2018). Expression quantitative trait loci in the developing human brain and their enrichment in neuropsychiatric disorders. Genome Biol. *19*, 194.

Ogawa, Y., Horresh, I., Trimmer, J.S., Bredt, D.S., Peles, E., and Rasband, M.N. (2008). Postsynaptic Density-93 Clusters Kv1 Channels at Axon Initial Segments Independently of Caspr2. J. Neurosci. *28*, 5731–5739.

Ogawa, Y., Oses-Prieto, J., Kim, M.Y., Horresh, I., Peles, E., Burlingame, A.L., Trimmer, J.S., Meijer, D., and Rasband, M.N. (2010). ADAM22, a Kv1 channel-interacting protein, recruits membrane-associated guanylate kinases to juxtaparanodes of myelinated axons. J. Neurosci. *30*, 1038–1048.

Okada, N., Fukunaga, M., Yamashita, F., Koshiyama, D., Yamamori, H., Ohi, K., Yasuda, Y., Fujimoto, M., Watanabe, Y., Yahata, N., et al. (2016). Abnormal asymmetries in subcortical brain volume in schizophrenia. Mol. Psychiatry *21*, 1460–1466.

Okamoto, S., Amaishi, Y., Maki, I., Enoki, T., and Mineno, J. (2019). Highly efficient genome editing for single-base substitutions using optimized ssODNs with Cas9-RNPs. Sci. Rep. *9*, 1–11.

Olabi, B., Ellison-Wright, I., McIntosh, A.M., Wood, S.J., Bullmore, E., and Lawrie, S.M. (2011). Are there progressive brain changes in schizophrenia? a meta-analysis of structural magnetic resonance imaging studies. Biol. Psychiatry *70*.

Oliva, C., Escobedo, P., Astorga, C., Molina, C., and Sierralta, J. (2012). Role of the maguk protein family in synapse formation and function. Dev. Neurobiol. *72*, 57–72.

Oliveira, J., Costa, M., De Almeida, M.S.C., Da Cruz E Silva, O.A.B., and Henriques, A.G. (2017). Protein Phosphorylation is a Key Mechanism in Alzheimer's Disease. J. Alzheimer's Dis. *58*.

Olson, E.C. (2014). Analysis of preplate splitting and early cortical development illuminates the biology of neurological disease. Front. Pediatr. *2*.

Orlovska, S., Pedersen, M.S., Benros, M.E., Mortensen Dr., P.B., Agerbo Dr., E., and Nordentoft Dr., M. (2014). Head injury as risk factor for psychiatric disorders: A nationwide register-based follow-up study of 113,906 persons with head injury. Am. J. Psychiatry *171*.

van Os, J., and Kapur, S. (2009). Schizophrenia. Lancet *374*, 635–645.

Owen, M.J. (2012). Implications of genetic findings for understanding schizophrenia. Schizophr. Bull. *38*.

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., Sawa, A., and Mortensen, P.B. (2016). Schizophrenia. Lancet *388*, 86–97.

Paatero, A.O., Turakainen, H., Happonen, L.J., Olsson, C., Palomäki, T., Pajunen, M.I., Meng, X., Otonkoski, T., Tuuri, T., Berry, C., et al. (2008). Bacteriophage Mu integration in yeast and mammalian genomes. Nucleic Acids Res. *36*.

Pak, C.H., Danko, T., Zhang, Y., Aoto, J., Anderson, G., Maxeiner, S., Yi, F., Wernig, M., and Südhof, T.C. (2015). Human Neuropsychiatric Disease Modeling using Conditional Deletion Reveals Synaptic Transmission Defects Caused by Heterozygous Mutations in NRXN1. Cell Stem Cell *17*, 316–328.

PALAY, S.L. (1956). Synapses in the central nervous system. J. Biophys. Biochem. Cytol. *2*.

Paquet, D., Kwart, D., Chen, A., Sproul, A., Jacob, S., Teo, S., Olsen, K.M., Gregg, A., Noggle, S., and Tessier-Lavigne, M. (2016). Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. Nature *533*, 125–129.

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.

Parker, C.M., and Clifton, N.E. (2018). GENEsis Gene Conversion Program.

Parker, M.J., Zhao, S., Bredt, D.S., Sanes, J.R., and Feng, G. (2004). PSD93 Regulates Synaptic Stability at Neuronal Cholinergic Synapses. J. Neurosci. *24*.

Picchioni, M.M., and Murray, R.M. (2007). Schizophrenia. Br. Med. J. *335*, 91–95.

Pluta, K., Luce, M.J., Bao, L., Agha-Mohammadi, S., and Reiser, J. (2005). Tight control of transgene expression by lentivirus vectors containing second-generation tetracycline-responsive promoters. J. Gene Med. *7*, 803–817.

Pocklington, A.J., Cumiskey, M., Armstrong, J.D., and Grant, S.G.N. (2006). The proteomes of neurotransmitter receptor complexes form modular networks with distributed functionality underlying plasticity and behaviour. Mol. Syst. Biol. *2*.

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.

Porter, A.P., White, G.R.M., Mack, N.A., and Malliri, A. (2019). The interaction between CASK and

the tumour suppressor Dlg1 regulates mitotic spindle orientation in mammalian epithelia. J. Cell Sci. *132*.

Pouget, J.G., Gonçalves, V.F., Spain, S.L., Finucane, H.K., Raychaudhuri, S., Kennedy, J.L., and Knight, J. (2016). Genome-wide association studies suggest limited immune gene enrichment in schizophrenia compared to 5 autoimmune diseases. Schizophr. Bull. *42*.

Prelich, G. (2012). Gene overexpression: Uses, mechanisms, and interpretation. Genetics *190*, 841–854.

Proudfoot, N.J., Furger, A., and Dye, M.J. (2002). Integrating mRNA processing with transcription. Cell *108*, 501–512.

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.

Qi, J., Su, Y., Sun, R., Zhang, F., Luo, X., Yang, Z., and Luo, X. (2005). CASK inhibits ECV304 cell growth and interacts with Id1. Biochem. Biophys. Res. Commun. *328*.

Qi, Y., Zhang, X.J., Renier, N., Wu, Z., Atkin, T., Sun, Z., Ozair, M.Z., Tchieu, J., Zimmer, B., Fattahi, F., et al. (2017). Combined small-molecule inhibition accelerates the derivation of functional cortical neurons from human pluripotent stem cells. Nat. Biotechnol. *35*, 154–163.

Quadrato, G., Brown, J., and Arlotta, P. (2016). The promises and challenges of human brain organoids as models of neuropsychiatric disease. Nat. Med. *22*, 1220–1228.

Radhakrishnan, R., Wilkinson, S.T., and D'Souza, D.C. (2014). Gone to pot-a review of the association between cannabis and psychosis. Front. Psychiatry *5*.

Radonjić, N. V., Ayoub, A.E., Memi, F., Yu, X., Maroof, A., Jakovcevski, I., Anderson, S.A., Rakic, P., and Zecevic, N. (2014). Diversity of Cortical Interneurons in Primates: The Role of the Dorsal Proliferative Niche. Cell Rep. *9*, 2139–2151.

Rakic, P. (1972). Mode of cell migration to the superficial layers of fetal monkey neocortex. J. Comp. Neurol. *145*.

Rakic, P. (1988). Specification of cerebral cortical areas. Science (80-.). *241*, 170–176.

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

Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., et al. (2013b). Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. Cell *154*, 1380–1389.

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. (2019). 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.S., Legge, S.E., Pardiñas, A.F., et al. (2020). De novo mutations identified by exome sequencing implicate rare missense variants in SLC6A1 in schizophrenia. Nat. Neurosci. *23*.

Reggiani, C., Coppens, S., Sekhara, T., Dimov, I., Pichon, B., Lufin, N., Addor, M.C., Belligni, E.F., Digilio, M.C., Faletra, F., et al. (2017). Novel promoters and coding first exons in DLG2 linked to developmental disorders and intellectual disability. Genome Med. *9*.

Reissner, C., and Missler, M. (2014). MAGUKs end a tale of promiscuity. Proc. Natl. Acad. Sci. U. S. A. *111*, 17350–17351.

Reissner, C., Runkel, F., and Missler, M. (2013). Neurexins. Genome Biol. *14*, 213.

del Rio, T., and Feller, M.B. (2006). Early Retinal Activity and Visual Circuit Development. Neuron *52*, 221–222.

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.

Roth, T.L., Puig-Saus, C., Yu, R., Shifrut, E., Carnevale, J., Li, P.J., Hiatt, J., Saco, J., Krystofinski, P., Li, H., et al. (2018). Reprogramming human T cell function and specificity with non-viral genome targeting. Nature *559*, 405–409.

Rowe, R.G., and Daley, G.Q. (2019). Induced pluripotent stem cells in disease modelling and drug discovery. Nat. Rev. Genet. *20*, 377–388.

Rozario, T., and DeSimone, D.W. (2010). The extracellular matrix in development and morphogenesis: a dynamic view. Dev. Biol. *341*, 126–140.

Rujescu, D., Bender, A., Keck, M., Hartmann, A.M., Ohl, F., Raeder, H., Giegling, I., Genius, J., McCarley, R.W., Möller, H.J., et al. (2006). A pharmacological model for psychosis based on N-methyl-D-aspartate receptor hypofunction: molecular, cellular, functional and behavioral abnormalities. Biol. Psychiatry *59*, 721–729.

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.

Sahoo, T., Theisen, A., Rosenfeld, J.A., Lamb, A.N., Ravnan, J.B., Schultz, R.A., Torchia, B.S., Neill, N., Casci, I., Bejjani, B.A., et al. (2011). Copy number variants of schizophrenia susceptibility loci are associated with a spectrum of speech and developmental delays and behavior problems. Genet. Med. *13*, 868–880.

Sakuma, T., Nishikawa, A., Kume, S., Chayama, K., and Yamamoto, T. (2014). Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. Sci. Rep. *4*, 4–9.

Sanders, B., Andrea, D.D., Collins, M.O., Rees, E., Steward, T.G.J., Zhu, Y., Chapman, G., Legge, S.E., Pardiñas, A.F., Harwood, A.J., et al. (2020). Synaptic protein DLG2 controls neurogenic transcriptional programs disrupted in schizophrenia and related disorders. BioRxiv 2020.01.10.898676.

Sansom, S.N., Griffiths, D.S., Faedo, A., Kleinjan, D.-J., Ruan, Y., Smith, J., Heyningen, V. van, Rubenstein, J.L., and Livesey, F.J. (2009). The Level of the Transcription Factor Pax6 Is Essential for Controlling the Balance between Neural Stem Cell Self-Renewal and Neurogenesis. PLoS Genet. *5*.

Sato, Y., Tao, Y.X., Su, Q., and Johns, R.A. (2008). Post-synaptic density-93 mediates tyrosine-phosphorylation of the N-methyl-d-aspartate receptors. Neuroscience *153*, 700–708.

Satomura, A., Nishioka, R., Mori, H., Sato, K., Kuroda, K., and Ueda, M. (2017). Precise genome-wide base editing by the CRISPR Nickase system in yeast. Sci. Rep. *7*, 1–10.

Scannevin, R.H., and Huganir, R.L. (2000). Postsynaptic organisation and regulation of excitatory synapses. Nat. Rev. Neurosci. *1*.

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.

Schizophrenia Commission (2012). The abandoned illness: a report from the Schizophrenia Commission.

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.

Schneider, M., Debbané, M., Bassett, A.S., Chow, E.W.C., Fung, W.L.A., Van Den Bree, M.B.M., Owen, M., Murphy, K.C., Niarchou, M., Kates, W.R., et al. (2014). Psychiatric disorders from childhood to adulthood in 22q11.2 deletion syndrome: Results from the international consortium on brain and behavior in 22q11.2 deletion syndrome. Am. J. Psychiatry *171*, 627–639.

Schnell, E., Sizemore, M., Karimzadegan, S., Chen, L., Bredt, D.S., and Nicoll, R.A. (2002). Direct interactions between PSD-95 and stargazin control synaptic AMPA receptor number. Proc. Natl. Acad. Sci. U. S. A. *99*.

Schubert, M.S., Cedrone, E., Neun, B., Behlke, M.A., and Dobrovolskaia, M.A. (2018). Chemical Modification of CRISPR gRNAs Eliminate type I Interferon Responses in Human Peripheral Blood Mononuclear Cells. J. Cytokine Biol. *03*.

Sekar, A., Bialas, A.R., De Rivera, H., Davis, A., Hammond, T.R., Kamitaki, N., Tooley, K., Presumey, J., Baum, M., Van Doren, V., et al. (2016). Schizophrenia risk from complex variation of complement component 4. Nature *530*, 177–183.

Selemon, L.D., and Goldman-Rakic, P.S. (1999). The reduced neuropil hypothesis: A circuit based model of schizophrenia. Biol. Psychiatry *45*, 17–25.

Semple, B.D., Blomgren, K., Gimlin, K., Ferriero, D.M., and Noble-Haeusslein, L.J. (2013). Brain development in rodents and humans: Identifying benchmarks of maturation and vulnerability to injury across species. Prog. Neurobiol. *106–107*, 1–16.

Shao, Y.W., Wood, G.A., Lu, J., Tang, Q.-L., Liu, J., Molyneux, S., Chen, Y., Fang, H., Adissu, H., McKee, T., et al. (2019). Cross-species genomics identifies DLG2 as a tumor suppressor in osteosarcoma. Oncogene *38*, 291–298.

Sharma, T., Lancaster, E., Sigmundsson, T., Lewis, S., Takei, N., Gurling, H., Barta, P., Pearlson, G., and Murray, R. (1999). Lack of normal pattern of cerebral asymmetry in familial schizophrenic patients and their relatives - The Maudsley Family Study. Schizophr. Res. *40*, 111–120.

Silva, C.G., Peyre, E., and Nguyen, L. (2019). Cell migration promotes dynamic cellular interactions to control cerebral cortex morphogenesis. Nat. Rev. Neurosci. 1.

Singh, T., Walters, J.T.R., Johnstone, M., Curtis, D., Suvisaari, J., Torniainen, M., Rees, E., Iyegbe, C., Blackwood, D., McIntosh, A.M., et al. (2017). The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. Nat. Genet. *49*.

Skene, N.G., Bryois, J., Bakken, T.E., Breen, G., Crowley, J.J., Gaspar, H.A., Giusti-Rodriguez, P., Hodge, R.D., Miller, J.A., Muñoz-Manchado, A.B., et al. (2018). Genetic identification of brain cell types underlying schizophrenia. Nat. Genet. *50*, 825–833.

Smeland, O.B., Wang, Y., Frei, O., Li, W., Hibar, D.P., Franke, B., Bettella, F., Witoelar, A., Djurovic, S., Chen, C.H., et al. (2018). Genetic Overlap between Schizophrenia and Volumes of Hippocampus, Putamen, and Intracranial Volume Indicates Shared Molecular Genetic Mechanisms. Schizophr. Bull. *44*, 854–864.

Smith, C.L., and Eppig, J.T. (2009). The mammalian phenotype ontology: enabling robust annotation and comparative analysis. Wiley Interdiscip. Rev. Syst. Biol. Med. *1*, 390–399.

Smyth, A.M., and Lawrie, S.M. (2013). The neuroimmunology of schizophrenia. Clin. Psychopharmacol. Neurosci. *11*.

Soltani, M.H., Pichardo, R., Song, Z., Sangha, N., Camacho, F., Satyamoorthy, K., Sangueza, O.P., and Setaluri, V. (2005). Microtubule-associated protein 2, a marker of neuronal differentiation, induces mitotic defects, inhibits growth of melanoma cells, and predicts metastatic potential of cutaneous melanoma. Am. J. Pathol. *166*, 1841–1850.

Song, A.J., and Palmiter, R.D. (2018). Detecting and Avoiding Problems When Using the Cre–lox System. Trends Genet. *34*, 333–340.

Speicher, M.R., and Carter, N.P. (2005). The new cytogenetics: Blurring the boundaries with molecular biology. Nat. Rev. Genet. *6*, 782–792.

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.

Srikanth, P., Han, K., Callahan, D.G., Makovkina, E., Muratore, C.R., Lalli, M.A., Zhou, H., Boyd, J.D., Kosik, K.S., Selkoe, D.J., et al. (2015). Genomic DISC1 Disruption in hiPSCs Alters Wnt Signaling and Neural Cell Fate. Cell Rep. *12*, 1414–1429.

Stark, R., Grzelak, M., and Hadfield, J. (2019). RNA sequencing: the teenage years. Nat. Rev. Genet. *20*, 631–656.

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.

Stiles, J., and Jernigan, T.L. (2010). The basics of brain development. Neuropsychol. Rev. *20*, 327–348.

Stone, J.M., Erlandsson, K., Arstad, E., Squassante, L., Teneggi, V., Bressan, R.A., Krystal, J.H., Ell, P.J., and Pilowsky, L.S. (2008). Relationship between ketamine-induced psychotic symptoms and NMDA receptor occupancy - A [123I]CNS-1261 SPET study. Psychopharmacology (Berl). *197*, 401–408.

Südhof, T.C. (2012). The presynaptic active zone. Neuron *75*.

Südhof, T.C. (2018). Towards an Understanding of Synapse Formation. Neuron *100*.

Sullivan, P.F., Kendler, K.S., and Neale, M.C. (2003). Schizophrenia as a Complex Trait: Evidence from a Meta-analysis of Twin Studies. Arch. Gen. Psychiatry *60*.

Sullivan, P.F., Daly, M.J., and O'Donovan, M. (2012). Genetic architectures of psychiatric disorders: The emerging picture and its implications. Nat. Rev. Genet. *13*.

Sun, Q., and Turrigiano, G.G. (2011). PSD-95 and PSD-93 play critical but distinct roles in synaptic scaling up and down. J. Neurosci. *31*, 6800–6808.

Sun, C., Cheng, M.C., Qin, R., Liao, D.L., Chen, T.T., Koong, F.J., Chen, G., and Chen, C.H. (2011). Identification and functional characterization of rare mutations of the neuroligin-2 gene (NLGN2) associated with schizophrenia. Hum. Mol. Genet. *20*.

Susser, E.S., and Lin, S.P. (1992). Schizophrenia After Prenatal Exposure to the Dutch Hunger Winter of 1944-1945. Arch. Gen. Psychiatry *49*, 983–988.

Suzuki, K., and Izpisua Belmonte, J.C. (2018). In vivo genome editing via the HITI method as a tool for gene therapy. J. Hum. Genet. *63*, 157–164.

Suzuki, K., Tsunekawa, Y., Hernandez-Benitez, R., Wu, J., Zhu, J., Kim, E.J., Hatanaka, F., Yamamoto, M., Araoka, T., Li, Z., et al. (2016). In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. Nature *540*, 144–149.

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.

Taber, K.H., Hurley, R.A., and Yudofsky, S.C. (2010). Diagnosis and Treatment of Neuropsychiatric Disorders. Annu. Rev. Med. *61*, 121–133.

Tan, X., and Shi, S.H. (2013). Neocortical neurogenesis and neuronal migration. Wiley Interdiscip. Rev. Dev. Biol. *2*, 443–459.

Tao, Y.-X., Rumbaugh, G., Wang, G.-D., Petralia, R.S., Zhao, C., Kauer, F.W., Tao, F., Zhuo, M., Wenthold, R.J., Raja, S.N., et al. (2003). Impaired NMDA receptor-mediated postsynaptic function and blunted NMDA receptor-dependent persistent pain in mice lacking postsynaptic density-93 protein. J. Neurosci. *23*, 6703–6712.

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.

Terkelsen, K.C., and Menikoff, A. (1995). Measuring the Costs of Schizophrenia: Implications for the Post-Institutional Era in the US. Pharmacoeconomics *8*, 199–222.

Thalhammer, A., and Cingolani, L.A. (2014). Cell adhesion and homeostatic synaptic plasticity. Neuropharmacology *78*.

Thapar, A., and Riglin, L. (2020). The importance of a developmental perspective in Psychiatry: what do recent genetic-epidemiological findings show? Mol. Psychiatry 1–9.

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

Thomas, H. V., Dalman, C., David, A.S., Gentz, J., Lewis, G., and Allebeck, P. (2001). Obstetric complications and risk of schizophrenia: Effect of gender, age at diagnosis and maternal history

of psychosis. Br. J. Psychiatry *179*, 409–414.

Tomoda, T., Sumitomo, A., Jaaro-Peled, H., and Sawa, A. (2016). Utility and validity of DISC1 mouse models in biological psychiatry. Neuroscience *321*, 99–107.

Toyoshima, M., Akamatsu, W., Okada, Y., Ohnishi, T., Balan, S., Hisano, Y., Iwayama, Y., Toyota, T., Matsumoto, T., Itasaka, N., et al. (2016). Analysis of induced pluripotent stem cells carrying 22q11.2 deletion. Transl. Psychiatry *6*.

Tozer, S., Le Dréau, G., Marti, E., and Briscoe, J. (2013). Temporal control of BMP signalling determines neuronal subtype identity in the dorsal neural tube. Dev. *140*.

Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V. V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P., et al. (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat. Biotechnol. *33*, 187–198.

Turrigiano, G.G. (2008). The Self-Tuning Neuron: Synaptic Scaling of Excitatory Synapses. Cell *135*.

Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Tissue-based map of the human proteome. Science (80-.). *347*.

Van, L., Boot, E., and Bassett, A.S. (2017). Update on the 22q11.2 deletion syndrome and its relevance to schizophrenia. Curr. Opin. Psychiatry *30*, 191–196.

VanHook, A.M. (2015). Hippo signaling for spindle orientation. Sci. Signal. *8*, ec331–ec331.

Varese, F., Smeets, F., Drukker, M., Lieverse, R., Lataster, T., Viechtbauer, W., Read, J., Van Os, J., and Bentall, R.P. (2012). Childhood adversities increase the risk of psychosis: A meta-analysis of patient-control, prospective-and cross-sectional cohort studies. Schizophr. Bull. *38*.

Varrault, A., Journot, L., and Bouschet, T. (2019). Cerebral Cortex Generated from Pluripotent Stem Cells to Model Corticogenesis and Rebuild Cortical Circuits: In Vitro Veritas? Stem Cells Dev. *28*, 361–369.

Vassos, E., Pedersen, C.B., Murray, R.M., Collier, D.A., and Lewis, C.M. (2012). Meta-analysis of the association of urbanicity with schizophrenia. Schizophr. Bull. *38*.

Vita, A., and Barlati, S. (2018). Recovery from schizophrenia: Is it possible? Curr. Opin. Psychiatry *31*, 246–255.

Vita, A., De Peri, L., Deste, G., and Sacchetti, E. (2012). Progressive loss of cortical gray matter in schizophrenia: A meta-analysis and meta-regression of longitudinal MRI studies. Transl. Psychiatry *2*.

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, 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.

Wang, T.F., Ding, C.N., Wang, G.S., Luo, S.C., Lin, Y.L., Ruan, Y., Hevner, R., Rubenstein, J.L.R., and Hsueh, Y.P. (2004). Identification of Tbr-1/CASK complex target genes in neurons. J. Neurochem. *91*.

Wang, W., Osenbroch, P., Skinnes, R., Esbensen, Y., Bjørås, M., and Eide, L. (2010). Mitochondrial DNA Integrity Is Essential For Mitochondrial Maturation During Differentiation of Neural Stem Cells. Stem Cells *28*, 2195–2204.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. Nat. Rev. Genet. *10*, 57–63.

Warland, A., Kendall, K.M., Rees, E., Kirov, G., and Caseras, X. (2020). Schizophrenia-associated genomic copy number variants and subcortical brain volumes in the UK Biobank. Mol. Psychiatry *25*.

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

Wei, Z., Zheng, S., Spangler, S.A., Yu, C., Hoogenraad, C.C., and Zhang, M. (2011). Liprin-Mediated Large Signaling Complex Organization Revealed by the Liprin-α/CASK and Liprin-α/Liprin-β Complex Structures. Mol. Cell *43*.

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

Wen, Z., Nguyen, H.N., Guo, Z., Lalli, M.A., Wang, X., Su, Y., Kim, N.S., Yoon, K.J., Shin, J., Zhang, C., et al. (2014). Synaptic dysregulation in a human iPS cell model of mental disorders. Nature

515, 414–418.

White, S.M., Murakami, S., and Yi, C. (2019). The complex entanglement of Hippo-Yap/Taz signaling in tumor immunity. Oncogene *38*, 2899–2909.

WHO (1992). The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. (Geneva).

Wiegreffe, C., Simon, R., Peschkes, K., Kling, C., Strehle, M., Cheng, J., Srivatsa, S., Liu, P., Jenkins, N.A., Copeland, N.G., et al. (2015). Bcl11a (Ctip1) Controls Migration of Cortical Projection Neurons through Regulation of Sema3c. Neuron *87*, 311–325.

Wienert, B., Shin, J., Zelin, E., Pestal, K., and Corn, J.E. (2018). In vitro–transcribed guide RNAs trigger an innate immune response via the RIG-I pathway. PLOS Biol. *16*, e2005840.

Willems, J., Jong, A.P.H. de, Scheefhals, N., and MacGillavry, H.D. (2019). ORANGE: A CRISPR/Cas9-based genome editing toolbox for epitope tagging of endogenous proteins in neurons. BioRxiv 700187.

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.

Winkler, D., Daher, F., Wüstefeld, L., Hammerschmidt, K., Poggi, G., Seelbach, A., Krueger-Burg, D., Vafadari, B., Ronnenberg, A., Liu, Y., et al. (2018). Hypersocial behavior and biological redundancy in mice with reduced expression of PSD95 or PSD93. Behav. Brain Res. *352*.

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*.

Won, S., Levy, J.M., Nicoll, R.A., and Roche, K.W. (2017). MAGUKs: multifaceted synaptic organizers. Curr. Opin. Neurobiol. *43*, 94–101.

Wood, A.J., Lo, T.W., Zeitler, B., Pickle, C.S., Ralston, E.J., Lee, A.H., Amora, R., Miller, J.C., Leung, E., Meng, X., et al. (2011). Targeted genome editing across species using ZFNs and TALENs. Science (80-.). *333*, 307.

Woods, D.F., Hough, C., Peel, D., Callaini, G., and Bryant, P.J. (1996). Dlg protein is required for junction structure, cell polarity, and proliferation control in Drosophila epithelia. J. Cell Biol. *134*,

1469–1482.

Xing, L., Larsen, R.S., Bjorklund, G.R., Li, X., Wu, Y., Philpot, B.D., Snider, W.D., and Newbern, J.M. (2016). Layer specific and general requirements for ERK/MAPK signaling in the developing neocortex. Elife *5*.

Xu, W. (2011). PSD-95-like membrane associated guanylate kinases (PSD-MAGUKs) and synaptic plasticity. Curr. Opin. Neurobiol. *21*.

Xu, N.J., and Henkemeyer, M. (2012). Ephrin reverse signaling in axon guidance and synaptogenesis. Semin. Cell Dev. Biol. *23*.

Xu, B., Roos, J.L., Levy, S., Van Rensburg, E.J., Gogos, J.A., and Karayiorgou, M. (2008). Strong association of de novo copy number mutations with sporadic schizophrenia. Nat. Genet. *40*.

Xu, X., Gao, D., Wang, P., Chen, J., Ruan, J., Xu, J., and Xia, X. (2018). Efficient homology-directed gene editing by CRISPR/Cas9 in human stem and primary cells using tube electroporation. Sci. Rep. *8*.

Yamagata, M., Sanes, J.R., and Weiner, J.A. (2003). Synaptic adhesion molecules. Curr. Opin. Cell Biol. *15*.

Yan, S., Schubert, M., Young, M., and Wang, B. (2017). Applications of Cas9 nickases for genome engineering.

Yarrington, R.M., Verma, S., Schwartz, S., Trautman, J.K., and Carroll, D. (2018). Nucleosomes inhibit target cleavage by CRISPR-Cas9 in vivo. Proc. Natl. Acad. Sci. U. S. A. *115*, 9351–9358.

Ye, F., Zeng, M., and Zhang, M. (2018). Mechanisms of MAGUK-mediated cellular junctional complex organization. Curr. Opin. Struct. Biol. *48*, 6–15.

Yin, H., Song, C.Q., Suresh, S., Wu, Q., Walsh, S., Rhym, L.H., Mintzer, E., Bolukbasi, M.F., Zhu, L.J., Kauffman, K., et al. (2017). structure-guided chemical modification of guide RNA enables potent non-viral in vivo genome editing. Nat. Biotechnol. *35*, 1179–1187.

Yoo, T., Kim, S.G., Yang, S.H., Kim, H., Kim, E., and Kim, S.Y. (2020). A DLG2 deficiency in mice leads to reduced sociability and increased repetitive behavior accompanied by aberrant synaptic transmission in the dorsal striatum. Mol. Autism *11*.

Yoon, K.J., Nguyen, H.N., Ursini, G., Zhang, F., Kim, N.S., Wen, Z., Makri, G., Nauen, D., Shin, J.H., Park, Y., et al. (2014). Modeling a genetic risk for schizophrenia in iPSCs and Mice reveals neural

stem cell deficits associated with adherens junctions and polarity. Cell Stem Cell *15*, 79–91.

Yoshimi, K., Kunihiro, Y., Kaneko, T., Nagahora, H., Voigt, B., and Mashimo, T. (2016). SsODN-mediated knock-in with CRISPR-Cas for large genomic regions in zygotes. Nat. Commun. *7*.

Yuan, S.H., Martin, J., Elia, J., Flippin, J., Paramban, R.I., Hefferan, M.P., Vidal, J.G., Mu, Y., Killian, R.L., Israel, M.A., et al. (2011). Cell-surface marker signatures for the Isolation of neural stem cells, glia and neurons derived from human pluripotent stem cells. PLoS One *6*.

Zecevic, N., Chen, Y., and Filipovic, R. (2005). Contributions of cortical subventricular zone to the development of the human cerebral cortex. J. Comp. Neurol. *491*, 109–122.

Zeng, M., Ye, F., Xu, J., and Zhang, M. (2018). PDZ Ligand Binding-Induced Conformational Coupling of the PDZ–SH3–GK Tandems in PSD-95 Family MAGUKs. J. Mol. Biol. *430*.

Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C.G., et al. (2018). Ensembl 2018. Nucleic Acids Res. *46*, D754–D761.

Zhang, T., Chen, D., and Lee, T.H. (2020). Phosphorylation signaling in APP processing in Alzheimer's disease. Int. J. Mol. Sci. *21*.

Zhang, X.H., Tee, L.Y., Wang, X.G., Huang, Q.S., and Yang, S.H. (2015). Off-target effects in CRISPR/Cas9-mediated genome engineering. Mol. Ther. - Nucleic Acids *4*, e264.

Zhang, Y., Pak, C.H., Han, Y., Ahlenius, H., Zhang, Z., Chanda, S., Marro, S., Patzke, C., Acuna, C., Covy, J., et al. (2013). Rapid single-step induction of functional neurons from human pluripotent stem cells. Neuron *78*.

Zhao, X., and Bhattacharyya, A. (2018). Human Models Are Needed for Studying Human Neurodevelopmental Disorders. Am. J. Hum. Genet. *103*, 829–857.

Zhao, B., Tumaneng, K., and Guan, K.-L. (2011). The Hippo pathway in organ size control, tissue regeneration and stem cell self-renewal. Nat. Cell Biol. *13*, 877–883.

Zhao, D., Lin, M., Chen, J., Pedrosa, E., Hrabovsky, A., Fourcade, H.M., Zheng, D., and Lachman, H.M. (2015). MicroRNA profiling of neurons generated using induced pluripotent stem cells derived from patients with schizophrenia and schizoaffective disorder, and 22q11.2 del. PLoS One *10*.

Zheng, C.Y., Seabold, G.K., Horak, M., and Petralia, R.S. (2011). MAGUKs, synaptic development, and synaptic plasticity. Neuroscientist *17*, 493–512.

Zhu, Z., and Huangfu, D. (2013). Human pluripotent stem cells: An emerging model in developmental biology. Dev. *140*, 705–717.

Zhu, J., Shang, Y., Xia, C., Wang, W., Wen, W., and Zhang, M. (2011). Guanylate kinase domains of the MAGUK family scaffold proteins as specific phospho-protein-binding modules. EMBO J. *30*.

Zhu, J., Shang, Y., Chen, J., and Zhang, M. (2012). Structure and function of the guanylate kinase-like domain of the MAGUK family scaffold proteins. Front. Biol. (Beijing). *7*, 379–396.

Zhu, J., Shang, Y., Wan, Q., Xia, Y., Chen, J., Du, Q., and Zhang, M. (2014). Phosphorylation-dependent interaction between tumor suppressors Dlg and Lgl. Cell Res. *24*.

Zhu, J., Shang, Y., and Zhang, M. (2016). Mechanistic basis of MAGUK-organized complexes in synaptic development and signalling. Nat. Rev. Neurosci. *17*, 209–223.

9. Supplementary Information

Supplementary Document 1. Calculating the penetrance of deletions at the *DLG2* locus. Excerpt from unpublished work by Dr Andrew Pocklington (MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University) calculating the penetrance of deletions in the *DLG2* locus in schizophrenia and other developmental disorders (relevant to Section 1.3).

Supplementary Figure 1. ICC and western blot characterisation of the KO1 and KO2 cell lines (separately) between days 20–60 of cortical differentiation (relevant to Section 4.4 & 4.5).

Supplementary Figure 2. RNA QC summary with RINs for all 36 samples used in RNAseq c-DNA library construction. (relevant to Section 5.3).

Supplementary Figure 3. R script for functional gene-set enrichment test: Fisher's exact test plus refinement by odds ratio. Kindly provided by Dr Andrew Pocklington (relevant to Sections 5.4, 5.5, 5.6 & 5.7).

Supplementary Table 1. Full predicted off target sites for gRNA1/gRNA2 used in conjunction with Cas9 nuclease and off-target validation of gRNA1+gRNA2 used in conjunction with Cas9 nickase to generate the KO1 and KO2 cell lines, including sanger sequencing primers (relevant to Section 4.3).

Supplementary Table 2a. *DESEQ2* output. Full list of differentially expressed protein coding genes (up and down-regulated) between *DLG2*$^{-/-}$ (KO1 & KO2 combined) and WT cells at four time points of cortical differentiation from hESCs (relevant to section 5.3).

Supplementary Table 2b. *DESEQ2* output. Full list of differentially expressed protein coding genes (up and down-regulated) between KO1 (only) and WT cells at four time points of cortical differentiation from hESCs (relevant to section 5.3).

Supplementary Table 2c. *DESEQ2* output. Full list of differentially expressed protein coding genes (up and down-regulated) between KO2 (only) and WT cells at four time points of cortical differentiation from hESCs (relevant to section 5.3).

Supplementary Table 3a. Fisher's exact test for GO term functional enrichments in sets of up-regulated and down-regulated protein coding genes significantly differentially expressed (Bonferroni $p < 0.05$) between *DLG2*$^{-/-}$ and WT cultures at days 15, 20, 30 and 60 of cortical differentiation from hESC, without odds ratio refinement (relevant to Sections 5.4, 5.5, 5.6 & 5.7).

Supplementary Table 3b. Fisher's exact test for GO term functional enrichments in sets of up-regulated and down-regulated protein coding genes significantly differentially expressed (Bonferroni $p < 0.05$) between KO1 (only) and WT cultures at days 15, 20, 30 and 60 of cortical differentiation from hESC, without odds ratio refinement (relevant to Sections 5.4, 5.5, 5.6 & 5.7).

Supplementary Table 3c. Fisher's exact test for GO term functional enrichments in sets of up-regulated and down-regulated protein coding genes significantly differentially expressed

(Bonferroni $p < 0.05$) between KO2 (only) and WT cultures at days 15, 20, 30 and 60 of cortical differentiation from hESC, without odds ratio refinement (relevant to Sections 5.4, 5.5, 5.6 & 5.7).

Supplementary Table 4a. Fisher's exact test for MP term functional enrichments in sets of up-regulated and down-regulated protein coding genes significantly differentially expressed (Bonferroni $p < 0.05$) between *DLG2*$^{-/-}$ and WT cultures at days 15, 20, 30 and 60 of cortical differentiation from hESC, without odds ratio refinement (relevant to Sections 5.4, 5.5, 5.6 & 5.7).

Supplementary Table 4b. Fisher's exact test for MP term functional enrichments in sets of up-regulated and down-regulated protein coding genes significantly differentially expressed (Bonferroni $p < 0.05$) between KO1 (only) and WT cultures at days 15, 20, 30 and 60 of cortical differentiation from hESC, without odds ratio refinement (relevant to Sections 5.4, 5.5, 5.6 & 5.7).

Supplementary Table 4c. Fisher's exact test for MP term functional enrichments in sets of up-regulated and down-regulated protein coding genes significantly differentially expressed (Bonferroni $p < 0.05$) between KO2 (only) and WT cultures at days 15, 20, 30 and 60 of cortical differentiation from hESC, without odds ratio refinement (relevant to Sections 5.4, 5.5, 5.6 & 5.7).

Supplementary Table 5. Enrichment for schizophrenia common risk variants in all genes expressed in WT and *DLG2*$^{-/-}$ cultures (without conditioning) and genes differentially expressed (up or down-regulated) in *DLG2*$^{-/-}$ cultures (without conditioning, conditioned on all expressed genes at a timepoint or conditioned on all expressed genes across all time points) at four time points of cortical differentiation from hESCs, including β comparisons with the day 30 down-regulated gene set (relevant to section 6.3).

Supplementary Table 6. Enrichment for Alzheimer's disease common risk variants in genes differentially expressed (up or down-regulated) in *DLG2*$^{-/-}$ cultures (conditioned on all expressed genes at a timepoint or conditioned on all expressed genes across all time points) at four time points of cortical differentiation from hESCs (relevant to section 6.4).

Supplementary Table 7. Enrichment for schizophrenia common risk variants in GO and MP terms over-represented amongst genes down-regulated in *DLG2*$^{-/-}$ cultures at day 30 of cortical differentiation, conditioned on all expressed genes (either at the day 30 timepoint or across all time points) and on the day 30 down-regulated gene set (relevant to section 6.5).

Supplementary Table 8. Enrichment for schizophrenia common risk variants for genes within the GO terms *voltage gated sodium channel activity* and *synapse assembly downregulated in KO1* and *KO2* (separately) at day 30 of cortical differentiation. Conditioning on all expressed genes across all time points and on day 30 down-regulated genes (relevant to section 6.5).

Supplementary File 1. *FastQC* output as QC for trimmed reads following RNAseq (relevant to Section 5.3).

Supplementary File 2. *BamTools* output including % sequence duplication of mapped trimmed reads following RNAseq (Relevant to Section 5.3).