Investigating Functional and Genetic Interactions Underlying Schizophrenia Risk in 22q11.2 Deletion Syndrome

A Thesis Submitted for the Degree of Doctor of Philosophy

School of Medicine, Cardiff University
September 2020

Áine Louise Moylett

Acknowledgements

First and foremost, I would like to thank my primary supervisor, Prof. Nigel Williams, for your continued guidance and unwavering encouragement throughout the PhD, I am so grateful to have had such open and enjoyable supervision. Our Monday meetings were always a pleasure and I will miss our conversations putting the world to rights. Secondly, thank you to my secondary supervisor Prof. Meng Li for your support and reassurance, even when things did not quite go to plan, I truly appreciate your consistent help over the years.

I would like to thank the MRC for funding my PhD studentship and giving me this opportunity.

I would like to thank Thomas Monfeuga and Leon Hubbard for all of your endless advice, kindness and patience with my bioinformatics work. Thank you to Nick Bray and Heath O'Brien for your collaboration and advice. A massive thank you to all of the members, past and present of the Li and Peall labs, you all made our lab such an enjoyable place to work. Thank you to Claudia, for your excellent teaching but also for inspiring me to undertake this PhD. Thank you to Maria for your continued encouragement during the golden gate days and my cell line generation, you were so supportive at every step. A special thank you to Dani, for letting me constantly ask you questions, helping with cell culture, viruses, R and this thesis, but also for being an incredible friend, I am so grateful for all the loud laughs we shared. Thank you also to everyone in the NMHRI, our daily lunches have been such a highlight of my PhD.

To my friends in Cardiff, thank you to Jasmine, Sylvia, Chiara and Holly, thank you for all the joy you have brought me over the years and for making Cardiff home. A special thank you to my pal, Emily, for being there for me day and night, sharing chocolate, gigs and walks when we needed it most and always making me smile. To Jack, thank you for your continued belief in me, for your all support and for that llama pen, I am so happy to have shared this experience with you.

Thank you to my friends at home, Hanna, Jennie, Max, Sarah, Lucy, Jamie, Chris, Tom, John and Xander. You have all been integral support systems over the past four years and a constant source of happiness and fun when I needed it most.

My biggest thank you is to my Mum, Dad and Robert, for all your love and support during my PhD. I am so grateful for you and all you have given me, including all the cake. Thank you for always believing in me - I am who I am because of you.

I dedicate this thesis to my Nanny, Grandad and Gran. I treasured my time with you all and I hope this makes you proud.

Thesis Summary

22q11.2 Deletion Syndrome (22q11.2DS) is a genetic disorder caused by a hemizygous deletion at chromosome 22q11.2. It is the most common chromosomal microdeletion and the strongest known molecular genetic risk factor associated with schizophrenia. However, the underlying mechanisms that lead to this neuropsychiatric risk remain largely unknown. The work in this thesis sought to investigate possible genetic and functional mechanisms that contribute to schizophrenia risk in 22q.11.2DS.

Potential schizophrenia candidate and disease modifier genes from within and outside of the 22q11.2 deletion region were explored. From within the deletion, *DGCR8* was initially selected as a gene of interest due to its key role in the microRNA biogenesis pathway and therefore gene expression regulation. Additional candidate genes were identified by assessing gene co-expression during fetal development in relation to *DGCR8* and predicated of loss of function and haploinsuffiency intolerance, leading to the selection of *HIRA* and *ZDHHC8*. Transcriptome wide association studies were performed in disease relevant tissues to identify schizophrenia modifier genes outside of the deletion by comparing 22q11.2DS patients with and without schizophrenia. However, this analysis identified no significant differences in gene expression.

CRISPR/Cas9 genome editing technology was utilised to knockout *DGCR8* in human embryonic stem cells. Mutant lines were generated and differentiated into cortical neuroprogenitor cells to investigate the role of DGCR8 in neurodevelopment. This work provided further evidence that *DGCR8* knockout lines derived from human embryonic stem cells may not be a viable method of modelling due to genomic instability, lack of protein reduction and so insufficient disease recapitulation. Finally, a lentiviral based CRISPR/Cas9 system in human neuroprogenitor cells (hNPCs) was established. Genetic manipulation of *DGCR8* in hNPCs further indicated a relationship between DGCR8 and *TBR1* in cortical development.

This thesis combines bioinformatic and cellular approaches to provide a basis for investigation of mechanisms underlying schizophrenia risk in 22q11.2DS.

Table of Contents

List of Figures

List of Tables

Abbreviations

22q11.2 IBBC	22q11.2 International Brain and Behaviour Consortium
22q11.2DS	22q11.2 Deletion Syndrome
ADHD	Attention Deficit Hyperactivity Disorder
AP	Anterior posterior
ASD	Autism Spectrum Disorder
CNS	Central nervous system
CNV	Copy number variant
COMT	Catechol-O-methyltransferase
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
CTIP2	Coup-TF interacting protein 2
DGCR8	DiGeorge Critical Region 8
DISC1	Disrupted in schizophrenia 1
DLPFC	Dorsolateral prefrontal cortex
dNTPs	Deoxyribonucleotides
DOX	Doxycycline
DSB	Double strand break
DV	Dorsoventral
EDTA	Ethylenediaminetetraacetic acid
eQTL	expression quantitative trait loci
FACS	Fluorescence activated cell sorting
Fezf2	Fez family zinc finger 2
FISH	Fluorescence in situ hybridization
FOXG1	Forkhead Box G1
GFP	Green fluorescent protein
GNB1L	G protein subunit Beta 1
gRNA	Guide RNA
GWAS	Genome wide association study
HEK	Human embryonic kidney
hESCs	Human embryonic stem cells

Het	Heterozygous
hPSCs	Human pluripotent stem cells
INDEL	Insertion/Deletion
iPSCs	Induced pluripotent stem cells
IQ	Intelligence quotient
KO	Knockout
LCR	Low copy repeat region
lncRNA	Long non-coding RNA
LD	Linkage disequilibrium
LTP	Long term potentiation
LVP	Lentiviral particles
mESCs	Mouse embryonic stem cells
min	Minute
miRNAs	MicroRNAs
MLPA	Multiplex PCR assay
MQPCR	Multiplex Quantitative PCR
NAHR	Non-allelic homologous recombination
NCAD	N-Cadherin
NEPs	Neuroepithelial progenitor cells
NHEJ	Non-homologous end joining
NPCs	Neuroprogenitor cells
NSCs	Neural stem cells
OCD	Obsessive compulsive disorder
P	p value
PAM	Protospacer adjacent motif
PAX6	Paired box protein Pax6
PBS	Phosphate-buffered saline
PBST	Phosphate-buffered saline with Tween
PCA	Principal component analysis
PCR	Polymerase chain reaction
PD	Parkinson’s disease
PPI	Pre-pulse inhibition
Pre-miRNA	Precursor MicroRNA

Pri-miRNA	Primary MicroRNA
PRODH	Proline dehydrogenase 1
PROVEAN	Protein Variation Effect Analyzer
PSCs	Pluripotent stem cells
QC	Quality control
R^2	Correlation Coefficient
RGCs	Radial Glial Cells
RISC	RNA-inducing silencing complex
S-COMT	Soluble-COMT
SATB2	SATB Homeobox 2
SCZ	Schizophrenia
SEPT5	Septin 5
sgRNA	Small guide RNA
SHH	Sonic hedgehog
siRNA	Small interfering RNA
SNPs	Single nucleotide polymorphism
SVZ	Subventricular zone
TBR1	T-box brain transcription factor 1
TBX1	T-box transcription factor 1
TET	Tetracycline
tracrRNA	Trans-activating crispr RNA
VCFS	Velocardiofacial syndrome
VZ	Ventricular zone
WT	Wildtype

1. General Introduction

1.1 The 22q11.2 Deletion Syndrome – overview

22q11.2 Deletion Syndrome (22q11.2DS) is a genetic disorder caused by the hemizygous deletion on the long arm (q) of chromosome 22 at region 11.2. Due to the wide ranging features associated with 22q11.2DS, over time it has been given other names, such as DiGeorge Syndrome, Velo-Cardio-Facial Syndrome (VCFS) and CATCH-22 Syndrome (Shprintzen, 2008). 22q11.2DS is the most common human microdeletion, occurring in approximately 1 in 4000 live births every year (Goodship *et al.*, 1998). The different syndromes associated with the 22q11.2 deletion and their date of discovery are summarised in Table 1.1. The 22q11.2 deletion is associated with a multitude of clinical features. Affected individuals can have mild to serious clinical features, routinely including congenital heart disease, immunodeficiency, autoimmune disease, facial and palatal abnormalities and cognitive and psychiatric disorders (McDonald-McGinn *et al.*, 2015). The diversity in phenotype leads to variable diagnosis depending on presentation, until the improvement in genetic testing which allowed for the amalgamation of the various syndromes. As a result, 22q11.2DS is designated to patients that harbour the copy number variant (CNV) mutation, regardless of their clinical presentation.

Table 1.1: Historical summary of syndromes associated with the 22q11.2 deletion. *Adapted from De Decker 2001 (De Decker and Lawrenson, 2001)*

Syndrome associated with 22q11.2 deletion	Date
Congenital thymic hypoplasia with hypocalcemia	1959
DiGeorge Syndrome described	1972
Takao Syndrome described (conotruncal anomaly face syndrome	1976
Velocardiofacial syndrome described (Shprintzen Syndrome)	1978
DiGeorge syndrome speculatively linked to chromosome 22	1981
Partial monosomy of chromosome 22 described	1982
"CATCH-22 syndrome" described	1989
Cayler syndrome associated with 22q11.2del	1994

1.1.1 Epidemiology

22q11.2DS is the most common chromosomal microdeletion syndrome. Birth defects tend to be the most frequent symptom leading to a diagnosis of 22q11.2DS. The most commonly cited incidence of 22q11.2DS is 1 in 4000 live births. This prevalence was based on screening for congenital disorders (mainly significant heart defects) in children born in the United Kingdom in 1993, this study established a prevalence of 25.7 per 100,000 live births (1:3891) (Goodship *et al.*, 1998). Both sexes and all racial and ethnic groups are affected by the deletion (McDonald-McGinn *et al.*, 1999). An equal distribution of the deletion has been found amongst males and females (Botto *et al.*, 2003). It has been reported that non-white patients may be diagnosed less frequently and less quickly, due to fewer craniofacial features in these populations (McDonald-McGinn *et al.*, 2005; Liu *et al.*, 2014). A systematic review of birth prevalence in Belgium, France, Singapore, Sweden, the United Kingdom and the United States of America reported 156 cases for 1,111,336 live births, with per- study rates ranging from 1:4525 to 1:9805 (Panamonta *et al.*, 2016). Approximately 90% of identified 22q11.2DS patients occur as *de novo* mutations, however parental transmission is estimated to occur in approximately 6-17% of cases (McDonald-McGinn *et al.*, 2001; Digillo *et al.*, 2003).

1.1.2 Molecular genetics of 22q11.2DS

Hemizygous deletions of varying sizes have been characterised at cytoband 22q11.2. The majority of patients, approximately 85-90%, harbour a 3Mb microdeletion, which is referred to as the typically deleted region (Carlson *et al.*, 1997; Shaikh, 2000). The remaining cases (5-10%) carry a smaller deletion, approximately 1.5Mb nested within the 3Mb deletion. Therefore, the majority of 22q11.2DS patients share the same 1.5Mb deletion. Additionally, there have been rare reported atypical deletions seen in 2-4% of patients (Lindsay, Greenberg, *et al.*, 1995; Kurahashi *et al.*, 1997).

The 22q11.2 deletion region is a known CNV hotspot, where CNVs are four times more likely to occur compared to the rest of the genome (Shaikh, 2000). The genetic architecture of the region underlies this genetic attribute. The presence of Low Copy Repeat regions (LCRs) adds genomic instability to the region, making it susceptible to meiotic error. The 22q11.2 region contains eight LCR regions (LCR A-H, as shown in

Figure 1.1A). LCR sequences share a high degree of sequence similarity, subsequently contributing to the region being a hotspot for non-allelic homologous recombination (NAHR) and therefore CNV generation (as reviewed in (Chen *et al.*, 2014). This mechanism is a common source of aberrant inter-chromosomal exchanges and results in high susceptibility for rearrangement (Shaikh et al. 2000). The eight LCRs within the 22q11.2 region encompass approximately 11% of the region (Babcock *et al.*, 2003). During meiosis alignment of these highly similar sequences can occur, leading to recombination events causing deletion or duplication of the regions surrounded by LCRs. This results in the appearance of CNVs as described in Figure 1.1B. These exchanges involve 8 large, paralogous LCRs, or segmental duplications (A–H) that are distributed along a 5.6 Mb segment of chromosome 22q11.2.

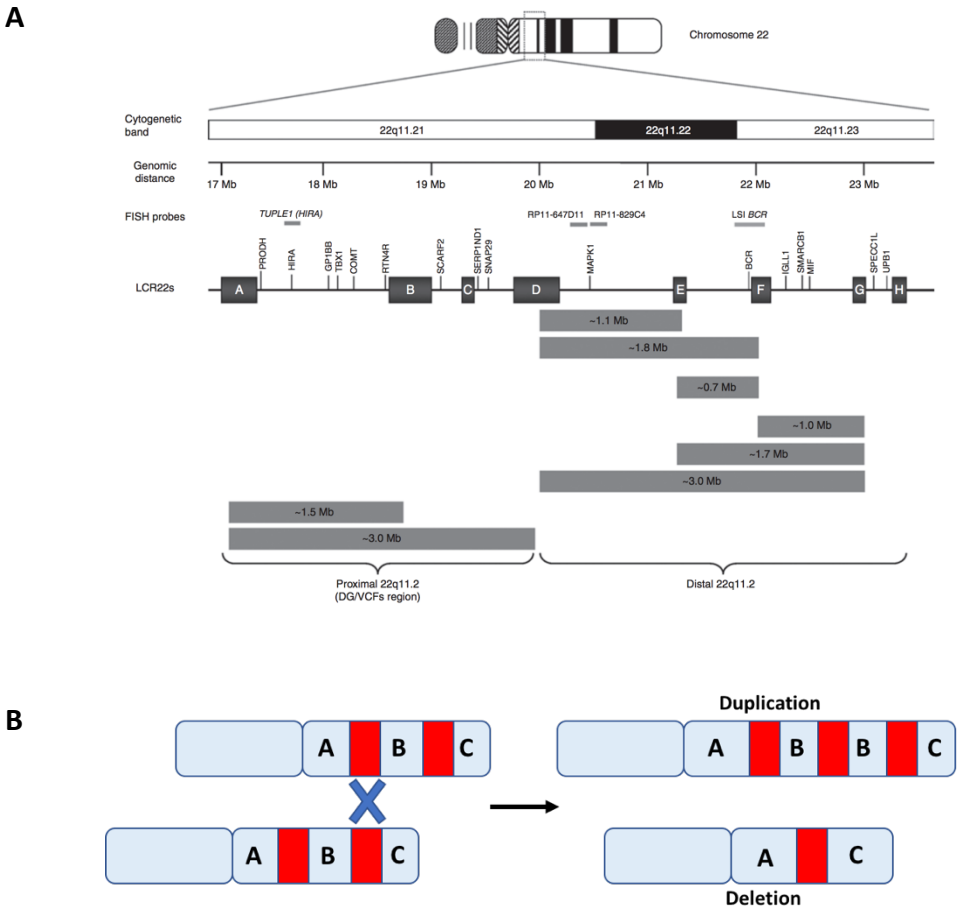


Figure 1.1 Genomic rearrangements in 22q11.2DS (A) 22q11.2 deletion region breakpoints and resulting deletions. *Proximal deletions occur between 22q11.2 LCR A to LCR D and distal deletions occur between LCR D to LCR H. LCR region coordinates defined from GRCh37/hg18. Figure obtained from Mikhail et al 2014*. **(B) Schematic representation of non-allelic homologous recombination (NAHR) resulting in inter-chromosomal deletion and duplication**. *Red boxes represent LCRs, with NAHR crossover events occurring directly between LCR regions on homologous chromosomes resulting in deletions and duplications.*

The typical 3Mb deletion, carried by the majority of patients, occurs between LCR A and LCR D, whereas the smaller 1.5Mb deletion occurs between LCR A and LCR B. Deletions have also been observed between LCR A to LCR C with patients harbouring a 2Mb deletion (Lindsay, Greenberg, *et al.*, 1995). Rarer deletions have been observed, central deletions have been observed which span LCR B-C and D (Rump *et al.*, 2014). Furthermore, 22q11.2 duplication syndrome is defined by the presence of a duplication starting from LCR A to LCR D and terminating between LCR D, E or G (Torres-Juan *et al.*, 2007).

1.1.3 Molecular diagnosis and testing for 22q11.2DS

The resolution of diagnostic techniques to detect 22q11.2 deletions have advanced over time. Fluorescence *in situ* hybridisation (FISH) was historically used to detect 22q11.2 deletions, typically using probe mapping to LCR A-B, however detection of nested deletions requiring custom probes (Driscoll *et al.*, 1993). A limitation of diagnosis using FISH is that for effective detection of the specific deletion, multiple probes are required and consequently the probes might miss atypical deletions. Whole region methodologies are becoming more clinically normal, such as multiplex ligation-dependent probe amplification (MLPA), single nucleotide polymorphisms (SNPs) and genome-wide arrays (Sørensen *et al.*, 2010; Racedo *et al.*, 2015). MLPA has been used to detect CNVs at 37 loci within the 22q11.2 region and other distal region with high sensitivity and specificity (Vorstman *et al.*, 2006). These methods are faster and allow for detailed mutation detection. Additionally, another effective and inexpensive method utilised for detection is quantitative polymerase chain reaction (qPCR) and multiplexed qPCR (MQPCR) (Weksberg *et al.*, 2005; Tomita-Mitchell *et al.*, 2010).

1.2 Clinical features of 22q11.2DS

Clinical manifestations of 22q11.2DS vary significantly between patients and can affect multiple organs across the body (Figure 1.2). More than 180 physical and behavioural features that are associated with 22q11.2DS have been compiled (Demily *et al.*, 2015). The typical features observed include congenital heart defects, immunodeficiency, facial abnormalities such as cleft palate, renal abnormalities, skeletal abnormalities, developmental delay and learning difficulties (Digilio *et al.*, 2005). The spectrum of anomalies associated with 22q11.2DS are incredibly broad and there is extensive

variability between patients (Ryan *et al.*, 1997). The main clinical features of 22q11.2DS are summarised in the section below.

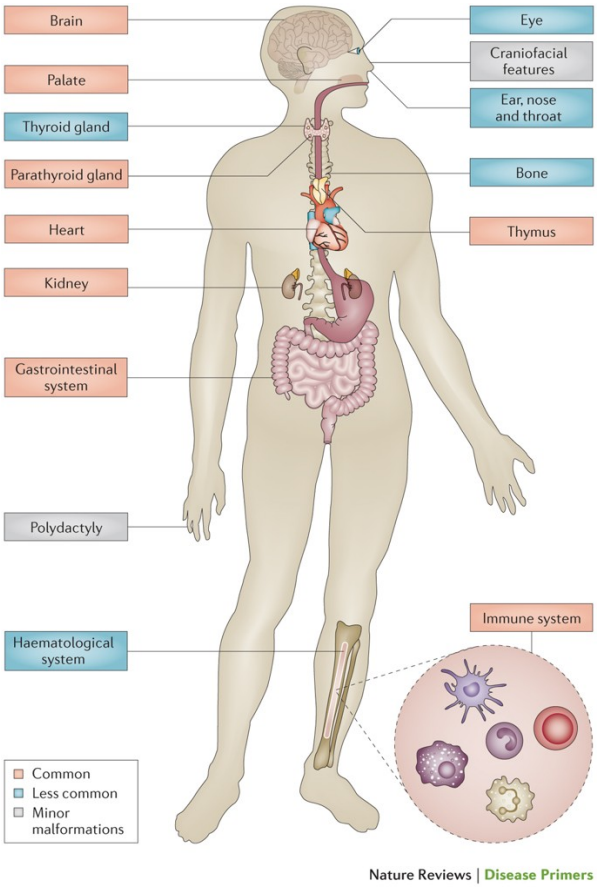


Figure 1.2 Organs and systems affected in 22q11.2 deletion syndrome. *Figure from McDonald-McGinn 2015 Nature reviews. Disease primers, (4271450447971) Copyright (2015).*

1.2.1 Physical Clinical Features of 22q11.2DS

The common physical abnormalities associated with 22q11.2DS include cardiac abnormality, dysmorphic facial features, cleft palate, thymic aplasia, endocrine abnormalities and hypocalcaemia/hypothyroidism, which will be described in more detail below (McDonald-McGinn *et al.*, 2015).

1.2.1.1 Congenital Heart Disorders

Cardiovascular abnormalities are one of the most common and well characterised features of 22q11.2DS. Cardiac anomalies become evident in the prenatal or neonatal period and therefore tend to lead to diagnosis (McDonald-Mcginn and Sullivan, 2011).

Approximately 80% of 22q11.2DS patients have congenital heart disorders (Momma, 2007, 2010). The most frequently observed cardiac malformations are conotruncal defects and include tetralogy of Fallot, pulmonary atresia with ventricular septal defect, truncus arteriosus and interrupted aortic arch as well as asymptomatic aortic arch malformations (Young, Shprintzen and Goldberg, 1980; Momma *et al.*, 1996). Age is correlated with type, severity, and prevalence of heart diseases in 22q11.2DS, with patients with full 22q11.2DS phenotypes presenting with more severe and recurrent congenital heart disorders (Momma, 2010). Congenital heart disorders are the main cause of mortality in 22q11.2DS, underlying approximately 90% of deaths (McDonald-McGinn *et al.*, 2001; Repetto *et al.*, 2014).

1.2.1.2 Craniofacial abnormalities

Facial abnormalities are another common manifestation of 22q11.2DS, presenting in nearly all children with a 22q11.2 deletion. Facial characteristics described in cases of the syndrome include long face, malar flatness, hypertelorism, hooded/swollen eyelids, wide nasal bridge, prominent nasal root, ear and earlobe abnormalities. The facial features are usually described as mild dysmorphic features, minor physical anomalies or typical facial appearance (Scambler *et al.*, 1992; Matsuoka *et al.*, 1998; McDonald-McGinn *et al.*, 1999). However, the characteristic facial appearance is often more subtle in infants with 22q11.2DS and become more prominent with age, therefore it is generally not used as a prompt for genetic testing (Digilio *et al.*, 2005). Additionally, approximately two thirds of 22q11.2DS patients have palatal abnormalities, but there is great variability in malformations (Shprintzen *et al.*, 1981; McDonald-McGinn *et al.*, 1999). Common palatal malformations include velopharyngeal insufficiencies observed in 27-32% of patients, furthermore cleft palate is seen in approximately 9-11% of cases (Ryan *et al.*, 1997; McDonald-McGinn *et al.*, 1999). Additional craniofacial features include asymmetric crying faces (found in 14% of patients) (McDonald-McGinn *et al.*, 1997).

1.2.1.3 Immunodeficiency in 22q11.2DS

Immunodeficiency in 22q11.2DS affects up to 75% of paediatric patients due to poor formation of the thymic tissues and subsequently deficient production of T-cells. Thymus hypoplasia is observed in approximately 80% of 22q11.2DS patients (Jawad *et*

al., 2001). The most frequent immunological abnormality in 22q11.2DS is low T-cell number. Additionally, functional T-cell deficiency is observed in a minority of patients (Sullivan *et al.*, 1998). The level of immunodeficiency is heterogeneous, with some patients having normal thymic development and normal T cell production compared to complete absence of T cell production (Sullivan, McDonald-McGinn and Zackai, 2002). Manifestations include chronic infections, with patients most at risk for repeated infections in the first years of life (Digilio *et al.*, 2005). 22q11.2DS is also associated with autoimmune diseases such as autoimmune thyroiditis or juvenile rheumatoid arthritis (Etzioni and Pollack, 1994; Jawad *et al.*, 2001; Gennery *et al.*, 2002).

1.2.1.4 Endocrine disorders

Hypoparathyroidism is reported in up to 30% of patients with 22q11.2DS (Lima *et al.*, 2011). Hypocalcaemia, a consequence of thyroid disfunction is observed in approximately 90% 22q11.2DS patients (Cheung *et al.*, 2014). Congenital hypocalcaemia is one of the strongest predictors of 22q11.2DS in new-borns (Ryan *et al.*, 1997). Tetany or seizures as a result of hypocalcaemia has been observed in undiagnosed 22q11.2DS adults (Van Den Bosch *et al.*, 2002). Growth disorders have also been reported, with patients presenting with short stature due to growth hormone deficiencies (Shprintzen *et al.*, 1981; McDonald-McGinn *et al.*, 1997; Weinzimer *et al.*, 1998).

1.2.2 Neurological and Neuropsychiatric features of 22q11.2DS

The 22q11.2 deletion is strongly associated with numerous neuropsychiatric conditions and cognitive impairments. Behavioural and developmental issues are one of the most commonly occurring features of the disorder (Swillen *et al.*, 1999). Up to 50% of patients have some degree of cognitive impairment, with typically low IQ scores compared to non-affected relatives (Niklasson *et al.*, 2009; Jolin, Weller and Weller, 2012). Furthermore, at any given age, at least 60% of individuals with 22q11DS meet diagnostic criteria for at least one psychiatric diagnosis, including psychotic disorders, attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorders (ASD), mood disorders, anxiety and others (Niklasson *et al.*, 2001; Jonas, Montojo and Bearden, 2014; Bertrán, Tagle and Irarrázaval, 2018). As with many of the 22q11.2DS features, the psychiatric, cognitive and behavioural phenotypes are highly variable between patients

and vary throughout a patient's lifetime. Importantly, no evidence has implicated either the length of the 22q11.2 deletion or the co-occurrence of other physical features, to contribute significantly to the heterogeneity of the cognitive or behavioural phenotype (M Karayiorgou *et al.*, 1995; Carlson *et al.*, 1997). There have been numerous studies investigating neuropsychiatric and cognitive phenotypes in 22q11.2DS patients, with the most recent efforts undertaken by the International 22q11.2 Brain Behaviour Consortium (22q11 IBBC). This is the largest scale international study into 22q11.2DS genotypes and phenotypes and the neurological and neuropsychiatric findings are discussed in further detail below.

1.2.2.1 Developmental and Speech Delay

Developmental delay and educational concerns are frequently reported in patients with 22q11.2DS. Significant delay of language onset is often seen in children with 22q11.2DS, with one study reporting approximately 70% of children did not speak or used few words at 24 months or older, indicating impairment of the onset of language (Scherer, D'Antonio and Kalbfleisch, 1999; Solot *et al.*, 2000). Furthermore, while the ability to understand speech is usually strong at the pre-school age, more general language delays and limitations in communicative abilities appear by school age (Gerdes *et al.*, 1999; Scherer, D'Antonio and Kalbfleisch, 1999). Additionally, children with 22q11.2DS have been shown to have gross and fine motor skill difficulties (Swillen *et al.*, 2005).

1.2.2.2 Cognitive impairment and IQ

Intelligence in children and adolescents with 22q11.2DS follows a normal distribution comparable to the general population (Swillen *et al.*, 1997), however patients have on average a lower full-scale IQ. Patients with 22q11.2DS have a mean IQ of approximately 70, within the borderline intellectual function range, compared to a mean IQ of 100 in the typically developing population (Swillen *et al.*, 1997; Moss *et al.*, 1999; Niklasson *et al.*, 2007). In the 22q11.2 IBBC cohort the average full-scale IQ was 71 points, with 46% of patients presenting with intellectual disability scores (< 70) (Schneider, Debbané, Anne S. Bassett, *et al.*, 2014). Furthermore, learning difficulties are very common in preschool and primary school, particularly within mathematical domains (De Smedt *et*

al., 2009). 33.3% of the patients have a mild intellectual disability, however more severe levels are uncommon in 22q11.2DS (Swillen *et al.*, 1999; De Smedt *et al.*, 2007).

1.2.2.3 Neuropsychiatric Disorders

Neuropsychiatric phenotypes have been widely observed in 22q11.2DS. Affected individuals have a substantially greater risk for various psychiatric disorders compared to the general population. Early studies which focused on adults consistently showed a greater than average prevalence of schizophrenia (SCZ) spectrum disorders (23%- 43%) (Green *et al.*, 2009; Fung *et al.*, 2010). In 2014, the 22q11 IBBC published investigations into psychiatric morbidity in the largest 22q11.2DS cohort to date (1402 patients) (Schneider, Debbané, Anne S Bassett, *et al.*, 2014). Participants were divided into 5 age groups: children (6-12 years), adolescents (13-17 years); emerging adults (18-25 years), young adults (26-35 years) and mature adults (≥36 years). The participant ages ranged from 6-68 years old. The incidence of each psychiatric disorder analysed in the study is summarised in Table 1.2. ADHD (37%), anxiety disorders (36%), disruptive disorders (14%) and ASD (13%) were the most common psychiatric phenotypes observed in children, while psychosis and mood disorders are more common in adolescence and young adulthood. Anxiety disorders were more prevalent than mood disorders at all ages, especially in paediatric age groups. In the adult cohort, there was high incidence of SCZ spectrum disorders (23-42%), which was mostly accounted for by SCZ (13-30%). 22q11.2DS patients showed an increase in prevalence of SCZ spectrum disorders with age, however there was a reduction in many forms of anxiety disorders across lifespan (As summarised in Table 1.2). Gender is implicated in affecting the incidence of psychotic disorders, but certain disorders were more common within each gender. Disruptive disorders and ADHD had a higher frequency in adult males, whereas mood and anxiety disorders were more prevalent in adult females.

Diagnosis	Children and adolescents		Adults			
	Children (6-12 years)	Adolescents (13-17 years)	Emerging Adults (18-25 years)	Young Adults (26-35 Years)	Mature Adults (≥36 Years)	Significant evolution
Any schizophrenia spectrum disorder	**2.0**	**10.8**	**23.5**	**41.5**	**41.7**	**Increase**
Schizophrenia	0.2	3.8	12.4	28.0	30.2	
Schizoaffective disorder	0.0	0.9	1.7	7.6	3.8	
Schizophreniform disorder	0.0	0.3	1.1	0.0	0.0	
Brief psychotic disorder	0.0	1.4	0.4	0.8	0.0	
Psychotic disorder not otherwise specified	1.8	4.1	9.0	9.3	13.5	
Delusion disorder	0.0	0.0	0.6	0.0	0.0	
Any anxiety disorder	**35.6**	**33.9**	**24.1**	**24.8**	**27.6**	**Decrease**
Separation anxiety disorder	6.3	1.5	1.8	0.0	0.0	Decrease
Specific phobia	21.9	17.0	7.2	3.8	2.8	Decrease
Social phobia	10.3	9.8	4.8	2.7	0.8	Decrease
Panic disorder	1.2	0.9	6.3	8.8	14.4	Increase
Posttraumatic stress disorder	0.4	1.4	0.8	0.0	2.7	
Obsessive-compulsive disorder	5.5	5.9	5.1	5.4	6.3	
Generalized anxiety disorder	8.3	10.5	9.8	12.2	11.0	
Anxiety disorder not otherwise specified	0.2	0.3	0.7	0.7	0.0	
Any mood disorder	**3.3**	**11.9**	**18.3**	**14.7**	**20.5**	
Major depressive disorder	2.2	9.0	10.8	12.0	15.8	Increase
Dysthymia	1.1	2.3	5.0	1.4	0.9	Across groups
Bipolar disorder or (hypo)manic episode in children	0.0	0.3	1.9	2.0	3.9	
Mood disorder not otherwise specified	0.0	1.2	2.2	0.0	1.6	
Substance-related disorder (substance abuse and dependence	0.0	0.5	2.5	6.3	4.6	

Diagnosis	Children (6-12 years)	Adolescents (13-17 years)	Emerging Adults (18-25 years)	Mature Adults (≥36 Years)	Significant evolution
ADHD	**37.1**	**23.9**	**15.6**	**Between children/ado and ado/adults**	
Autism spectrum disorders	**12.8**	**26.5**	**16.1**	**Across groups**	
Any disruptive disorder	**14.3**	**10.9**	**7.1**	**Across groups (non-significant trend)**	
Oppositional defiant disorder	14.3	14.8	6.1		
Conduct disorder	0.0	0.0	1.5		

Table 1.2. Incidence of psychiatric disorders (DSM-IV-TR) in five age groups of subjects with 22q11.2 deletion syndrome in 22q11.2 IBBC investigation (prevalence rates expressed in %). Significant evolution: significant increase/decrease over time, across or between age groups with $P \leq 0.05$ (ado = adolescents). Adapted from Schneider 2014.

1.2.2.4 Parkinson's Disease

It has been reported that 0.5% of early-onset Parkinson's Disease (PD) cases have occurred in patients with 22q11.2DS (Mok *et al.*, 2016). Furthermore, 22q11.2DS is estimated to increase risk for development of PD by 20- to 70-fold compared to the general population (Butcher *et al.*, 2013; Mok *et al.*, 2016). PD in 22q11.2DS patients is mostly indistinguishable from idiopathic PD in relation to the major clinical features, although the age of onset is earlier in 22q11.2DS (Boot *et al.*, 2018).

1.2.2.5 Neuroanatomical deficits

Brain malformations have been reported in 22q11.2DS patients. Overall brain volume has been shown to be diminished in 22q11.2 deletion carriers compared to typically developing controls (Simon *et al.*, 2005; Hopkins *et al.*, 2018). Furthermore, 22q11.2DS patients with SCZ have been showed to have smaller grey matter volume (Chow *et al.*, 2002). Conversely, cortical thickness as well as the volume of the corpus callosum has been shown to be increased in 22q11.2DS patients (Lin *et al.*, 2017; Sun *et al.*, 2018). A recent large scale meta-analysis found an overall reduction of mean volume in the frontal lobe, temporal lobe, parietal lobe, cerebellum and hippocampus, supporting the notion that structural abnormalities in 22q11.2DS and SCZ are convergent (Rogdaki *et al.*, 2020). While significant cortical malformations are relatively uncommon in 22q11.2DS, their frequency is much higher than that of the general population (Hopkins *et al.*, 2018).

1.3 The 22q11.2 deletion region: Genes and modelling the deletion

1.3.1 Genes within the deletion

Within the typical 3Mb 22q11.2 deletion spanning LCR A-D, there are 90 genes, 46 of these genes are protein coding (Figure 1.3). Within the smaller 1.5Mb nested deletion, there are 56 genes, 32 of which are protein coding genes as well as 6 microRNAs (miRNAs) and 9 long non-coding RNAs (lncRNAs) and some uncharacterised genes. A number of these genes are expressed in the mouse and human brain (Maynard *et al.*, 2003; Meechan *et al.*, 2009). The roles of many of these genes, in which haploinsufficiency could potentially contribute to the neurological/neuropsychiatric phenotypes of 22q11.2DS are well-described and will be discussed below.

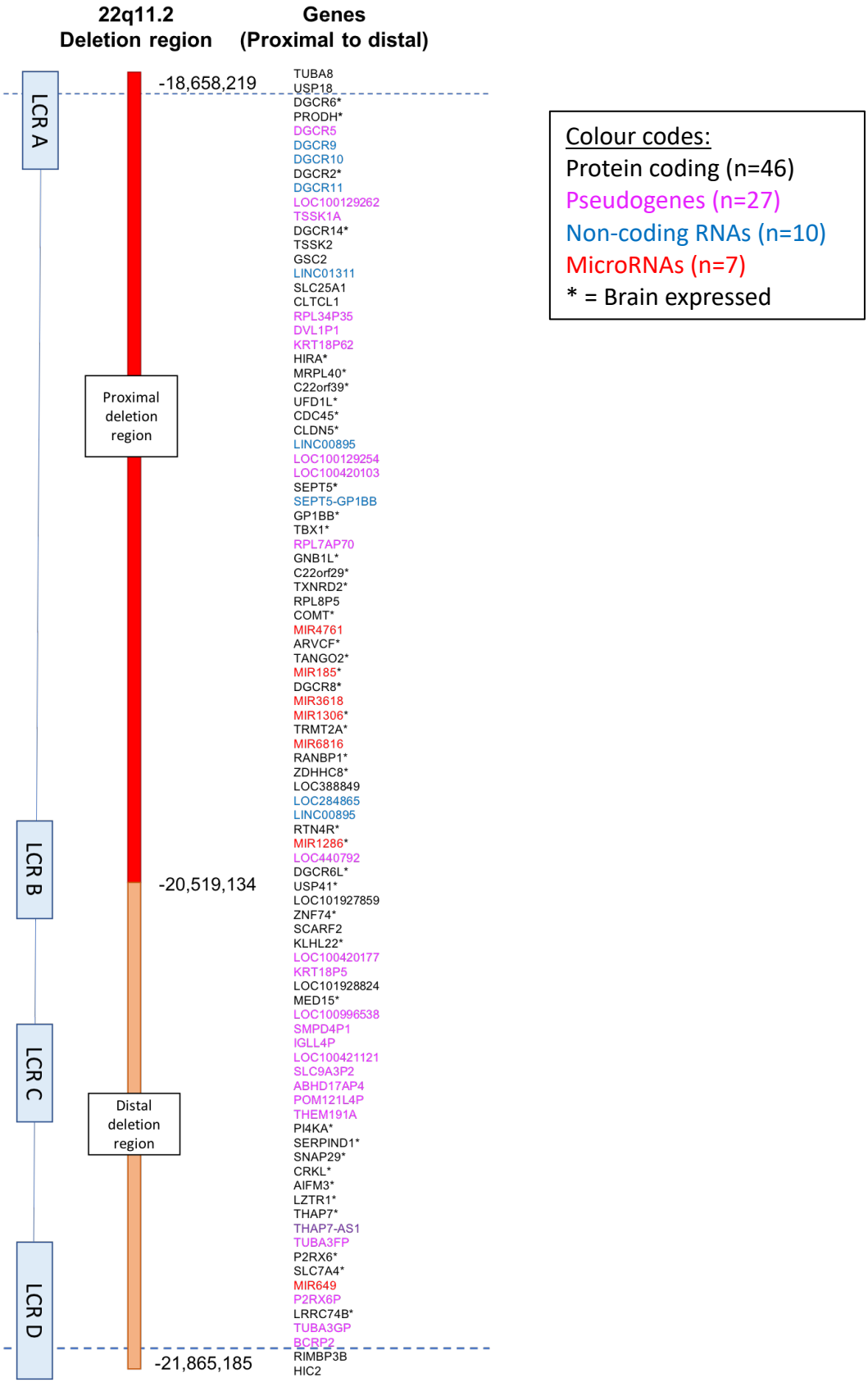


Figure 1.3: Gene content of the human 22q11.2 region: *The typical 3Mb 22q11.2 deletion spans approximately 90 genes, with protein coding, pseudogenes, non-coding RNAs and microRNAs highlighted. LCR A-D are indicated by blue boxes, deletion breakpoints are indicated by – (GCh38/hg19). Figure adapted from Guna et al 2015.*

1.3.2 Potential SCZ candidate genes within the deletion

The 22q11.2 deletion is the strongest known molecular genetic risk factor for SCZ, with 0.3% of all SCZ patients carrying the deletion (Maria Karayiorgou *et al.*, 1995; Karayiorgou, Simon and Gogos, 2010; Marshall *et al.*, 2017). There is little neuropsychiatric variability between patients harbouring the deletion between LCR A-B compared to LCR A-D deletions. This suggests haploinsuffiency of the genes within the 1.5Mb deletion are likely to be relevant to the psychiatric features of 22q11.2DS (Carlson *et al.*, 1997; Michaelovsky *et al.*, 2012). However, it is noteworthy that while numerous genes from the LCR A-D region have been individually associated with ASD, ADHD and other neuropsychiatric disorders, none have been unambiguously implicated (Paylor *et al.*, 2006; Bartsch *et al.*, 2011). Attempts have been made to associate common variants within genes in the 22q11.2 deletion with neuropsychiatric disorders; however, the results have been inconsistent, with positive and negative associations existing for numerous genes including *PRODH* and *COMT* (Li and He, 2006; Allen *et al.*, 2008; Kempf *et al.*, 2008; Ayalew *et al.*, 2012). Some genes within the 22q11.2 deletion region have key roles in brain development, neurotransmission and myelination, therefore have potentially stronger associations to SCZ risk (Motahari *et al.*, 2019). These genes have been further investigated for neuropsychiatric association and the following sections describes the most relevant genes within the 22q11.2 LCR A-B deletion.

1.3.2.1 PRODH

The gene *PRODH* encodes for the dual named proline oxidase or proline dehydrogenase 1. PRODH is an enzyme that catalyses the first step of proline degradation. Homozygous mutations in *PRODH* cause the neurological disorder hyperprolinemia type 1, which is associated with high levels of proline (Jacquet *et al.*, 2003). Prodh has been shown to have a suggested role in potentiation of pyramidal neuron excitatory transmission between different regions of the hippocampus (Cohen and Nadler, 1997). Homozygous *Prodh* mouse models display pre-pulse inhibition (PPI) and sensorimotor gating deficits (Gogos *et al.*, 1999; Heldt, Green and Ressler, 2004). It has been demonstrated that 22q11.2DS patients have elevated proline serum levels, which has been implicated in learning impairment as well as epilepsy and schizoaffective disorders (Raux *et al.*, 2007). Association studies between PRODH and SCZ have had mixed results and are reviewed in Willis 2008 (Willis *et al.*, 2008).

1.3.2.2 COMT

The *COMT* gene encodes for the enzyme Catechol-O-methyl transferase (COMT). COMT is an enzyme involved in the degradation of catecholamines, including the neurotransmitter dopamine in the synapse. The gene codes for two isoforms of the protein, with varied structure, one soluble isoform (S-COMT) and a membrane bound isoform (MB-COMT) (Tenhunen *et al.*, 1994). The MB-COMT is highly expressed in brain tissue and is responsible for dopamine inactivation in synapses (J. Chen *et al.*, 2004). COMT has a functional polymorphism, a valine is substituted by methionine at position 108 in the S-COMT isoform and position 158 in MB-COMT (Lachman *et al.*, 1996). The methionine polymorphism decreases enzymatic activity by approximately 40% in prefrontal cortex tissue (J. Chen *et al.*, 2004). Furthermore, this variant in *COMT* has been shown to impact prefrontal activation during working memory performance (Egan *et al.*, 2001). Therefore, 22q11.2DS carriers that also carry this variant have a suggested increased risk for impairment in prefrontal cognition, indicating that COMT could be a strong candidate gene for SCZ risk (Bearden *et al.*, 2004; Egan *et al.*, 2004).

1.3.2.3 TBX1

The T-box transcription factor (*TBX1*) gene is a member of the "T-box" family of transcription factors which all share the DNA binding domain, T-box. *TBX1* has been shown to be expressed during embryogenesis primarily in cardiovascular development and formation of pharyngeal arches (Vitelli *et al.*, 2002). Furthermore, mesodermal *Tbx1* expression has been shown to be essential for normal cortical development (Flore *et al.*, 2017). *Tbx1* heterozygous mutant mice have mild non-lethal phenotypes, whereas homozygous mutants are lethal and demonstrate many of the physical 22q11.2DS symptoms such as craniofacial defects, cleft palate and cardiac defects (Jerome and Papaioannou, 2001; Merscher *et al.*, 2001). *TBX1* is therefore a suggested genetic driver for many of the phenotypes seen in 22q11.2DS carriers. With regard to the neuropsychiatric phenotypes of 22q11.2DS, heterozygous *Tbx1* mutation mice have been shown to have reduced PPI (Paylor *et al.*, 2006).

1.3.2.4 GNB1L

The *GNB1L* gene (guanine nucleotide-binding protein, beta-1-like) encodes for a protein that contains six WD40 repeats, but the function of GNB1L remains relatively unknown (Funke, Pandita and Morrow, 2001). *GNB1L* has been shown to be expressed at high levels in embryonic human and mouse brain, but low levels in the adult brain (Gong *et al.*, 2000; Funke, Pandita and Morrow, 2001). *GNB1L* has been linked to SCZ in two independent case-control studies in different populations (Williams *et al.*, 2008; Li *et al.*, 2011). However, this result has not been replicated in larger, more recent association studies (Pardiñas *et al.*, 2018).

1.3.2.5 SEPT5

The *SEPT5* gene is a septin family member and is abundantly expressed in developing and adult human and mouse brains (Caltagarone *et al.*, 1998; Kinoshita, Noda and Kinoshita, 2000). Sept5 is expressed in presynaptic terminals and acts as a negative regulator of neurotransmitter release (Kinoshita, Noda and Kinoshita, 2000; Yang *et al.*, 2010). This mechanism has been implicated in abnormalities of neural connectivity in neuropsychiatric disorders (Honer, 2002). Furthermore, *SEPT5* has been recently associated as a potential driver of the synaptic pathology SCZ and ASD in 22q11.2DS based on a functional genomics approach (Forsyth *et al.*, 2020).

1.3.2.6 ZDHHC8

Zinc Finger DHHC-Type Containing 8 *(ZDHHC8)* encodes for a transmembrane palmitoyl-transferase. *Zdhhc8* expression is enriched in many brain regions in the developing and adult mouse brain, with protein expression restricted to the synapse (Maynard *et al.*, 2003, 2008; Meechan *et al.*, 2006). Palmitoylation is the covalent attachment of a saturated 16-carbon palmitic acid to cysteine residues of target proteins. Multiple lines of evidence suggest that palmitoylation plays a key role in neuronal development and synaptic plasticity (El-Husseini *et al.*, 2002; Hayashi, Thomas and Huganir, 2009; Yokoi, Fukata and Fukata, 2012). Furthermore, fatty acids such as long-chain ω-3 (omega-3) polyunsaturated fatty acids have been shown to reduce the risk of transition to SCZ in high-risk individuals (Amminger *et al.*, 2015). Initial studies associated SNPs within *ZDHHC8* with increased risk for SCZ (Mukai *et al.*, 2004; W. Y. Chen *et al.*, 2004).

However, this result failed to replicate in later studies with larger sample sizes (Glaser *et al.*, 2005; Pardiñas *et al.*, 2018). However, *Zdhhc8* mutant mice display abnormal axon growth, arborization and spine density in the hippocampus and cortex, as well as behavioural deficits (Mukai *et al.*, 2004, 2015). The role of *ZDHHC8* in SCZ risk in 22q11.2DS is further discussed in Chapter five of this thesis.

1.3.3 DGCR8

1.3.3.1 DGCR8, microRNA processing and other roles

DGCR8 (DiGeorge Critical Region Gene 8) encodes for an RNA binding protein which is a component of the microprocessor complex. DGCR8 interacts with DROSHA via the DGCR8 C-terminus to form the microprocessor complex (Yeom *et al.*, 2006). The microprocessor complex acts in primary miRNA (pri-miRNA) processing in the canonical miRNA biogenesis pathway (as depicted in Figure 1.4). MicroRNAs are small non-coding RNAs, approximately 22 nucleotides in length and regulate gene expression, therefore controlling a diverse range of cellular functions. MicroRNA biogenesis starts with synthesis of primary-miRNAs (pri-miRNA), these pri-miRNA transcripts are recognised by DGCR8 (Faller *et al.*, 2010; Quick-Cleveland *et al.*, 2014). DGCR8 then directs the RNase III-containing catalytic subunit enzyme Drosha to cleave the pri-miRNAs, which is the initial processing step (Lee *et al.*, 2003; Gregory *et al.*, 2004). The resulting precursor miRNA (pre-miRNA) is then exported to the cytoplasm for further processing by DICER to generate the final ~22nt mature miRNA (Grishok *et al.*, 2001; Ketting *et al.*, 2001). Mature miRNAs enter the RISC (RNA-induced silencing complex), in this complex miRNAs can anneal to their target and induce degradation and translational repression (Schwarz *et al.*, 2003). In the cell, the DGCR8 protein is localised mainly to the nucleus, specifically to the nucleolus, but there is also evidence showing it has been found in the cytoplasm (Shiohama *et al.*, 2007; Dai *et al.*, 2016). DGCR8 contains two double-strand RNA binding domains, which are essential for the microprocessor complex to interact with pri-miRNAs (Yeom *et al.*, 2006).

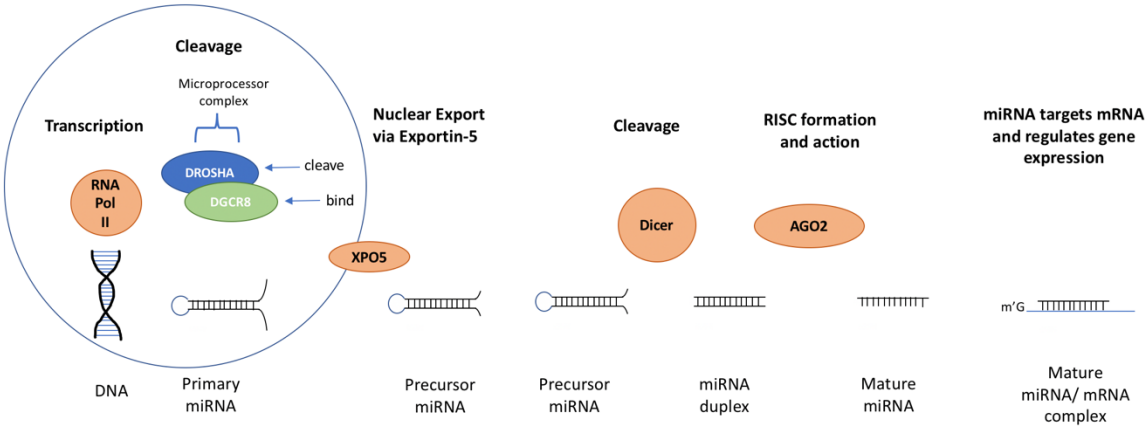


Figure 1.4: Canonical miRNA biogenesis pathway: *Figure illustrating stages of miRNA biogenesis, with filled circles representing the various proteins involved and the large blue circle representing the cell nucleus.*

The primary work investigating *DGCR8* shows ubiquitous expression of *DGCR8* in human fetal and adult tissue. Furthermore, *in situ* hybridization in the developing mouse embryo demonstrated expression in the neuroepithelium of the fetal brain, thymus and around the palate – indicating contribution to the clinical phenotypes observed in 22q11.2DS (Shiohama *et al.*, 2003).

DGCR8 is mostly associated with its role in miRNA processing, however it has miRNA independent functions. DGCR8 has been shown to directly interact with heterochromatin, maintaining heterochromatin organisation to protect against cellular senescence (Deng *et al.*, 2019). Furthermore, DGCR8 has been shown to be critical for corticogenesis. Independent of miRNA processing, DGCR8 has been suggested to directly regulate target mRNAs including *Tbr1,* further linking DGCR8 to the psychiatric phenotypes of 22q11.2DS (Marinaro *et al.*, 2017). The evidence for the relationship and regulation between Dgcr8 and *Tbr1* is further described in section 1.3.3.3.

1.3.3.2 DGCR8 and its Relevance to Schizophrenia

Multiple lines of evidence have implicated *DGCR8* in the aetiology of SCZ. Several animal models have been generated to elucidate the effects of *Dgcr8* deficiency. *Dgrc8* homozygous knockouts (KO) are lethal before the end of the pupal stage in *Drosophila melanogaster* models (reviewed in Guna, Butcher and Bassett, 2015). Additionally, embryonic lethality is observed in homozygous knockout mouse models, with death

observed at early post-implantation stage around E6.5 (Wang *et al.*, 2007). Further conditional knockout/knockdown and heterozygous experiments have helped to elucidate the role of *Dgcr8* in neurodevelopment. In *Drosophila melanogaster*, abnormal neuron morphology and neurophysiology was observed (Luhur *et al.*, 2014).

Dgcr8$^{+/-}$ mice show reduced adult hippocampal neurogenesis with decreased proliferation, which is rescued by restoration of IGF2 (Ouchi *et al.*, 2013). Dgcr8 deficiency has also been shown to affect neuronal circuitry. *Dgcr8* heterozygous mice and conditional *Dgcr8* deletion in pyramidal neurons of the mouse cortex have demonstrated numerous phenotypes. These experiment have shown deficits in layer V neurons; with impaired development of excitatory synaptic transmission, affecting their intrinsic electrical properties and a specific reduction in numbers of parvalbumin interneurons (Schofield *et al.*, 2011; Hsu *et al.*, 2012). Additionally, *Dgcr8* heterozygous mice demonstrate greater short-term synaptic depression in layer V prefrontal cortex electrophysiological recordings, indicating impairment in short-term synaptic plasticity (Fénelon *et al.*, 2011).

Behavioural deficits have also been observed in *Dgcr8* mutant mouse models. *Dgcr8*$^{+/-}$ mice display deficits in pre-pulse inhibition (PPI), cognitive performance and spatial working memory (Stark *et al.*, 2008; Ouchi *et al.*, 2013). There is further evidence linking *Dgcr8* and miRNA disruption to specific behavioural deficits observed in the D*f*(16)$^{A+/-}$ 22q11.2DS mouse model. Both *Dgcr8*$^{+/-}$ mice and 22q11.2DS mice models have an age-dependent increase in hippocampal long-term potentiation (LTP), which underlies learning and memory. Depletion of miR-25 and miR-185 in this model lead to dysregulation of SERCA2 (Sarcoendoplasmic reticulum Ca^{2+} ATPase) and increased LTP. Restoration of these miRNAs rescue the LTP in Dgcr8$^{+/-}$ mice (Earls *et al.*, 2012). Chun and colleagues identified *Dgcr8* happloinsuffiency as the cause for aberrant elevation of Drd2 (D2 dopamine receptor) in the 22q11DS mouse model. This reduction of *Dgcr8* leads to disruption of synaptic transmission of thalamic inputs into the auditory cortex (as described in Figure 1.5). This phenotype results in hypersensitivity to antipsychotics and therefore links *Dgcr8* to SCZ pathophysiology (Chun *et al.*, 2014). The *Dgcr8*-Drd2 pathway impairment in 22q11.2DS mice has been shown to affect fear memory, which can be restored by decreasing levels of Drd2 (Eom *et al.*, 2017). Additionally, progressive

ventricular enlargement, a common trait of neurological disorders, is observed in 22q11.2DS mice and is because of loss of *Dgcr8* meditated reduction of specific miRNAs causing elevated dopamine receptors (Eom *et al.*, 2020).

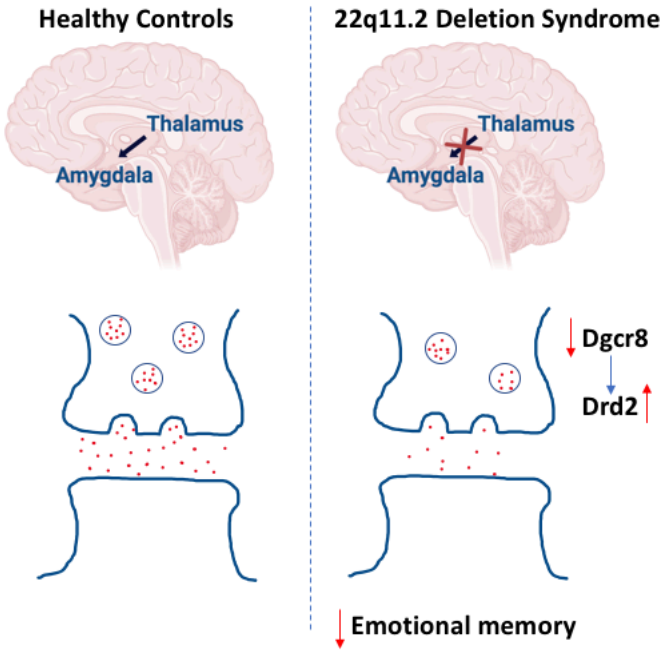


Figure 1.5: Mechanism underlying emotional memory deficits in 22q11.2DS mouse models due to relationship between *Dgcr8* and the Drd2 receptor. *Figure illustrating the relationship between Dgcr8 and Drd2 in 22q11.2DS mouse model. Haploinsufficiency of Dgcr8 causes increased Drd2 in the thalamus and reduced probability of glutamate release, leading to fear memory deficits (Eom et al., 2017, 2020). Figure generated using BioRender.com.*

1.3.3.3 Dgcr8 regulation of *Tbr1*

The pri-miRNA processing role is a primary focus of research surrounding DGCR8 function, however DGCR8 can also anneal to other stem-loop containing RNAs, including mRNAs which will then be cleaved by DROSHA (Maclas *et al.*, 2012; Seong *et al.*, 2014). Marinaro and colleagues investigate the loss of *Dgcr8* by conditional inactivation in apical progenitor cells of the mouse cortex (Marinaro *et al.*, 2017). Following conditional knockout, which caused complete loss of the Dgcr8 protein, there was increased generation of Tbr1 positive neurons in the developing cortex. They report this increase in Tbr1 positive neurons is due to miRNA independent functions of Dgcr8. The authors hypothesise that Dgcr8 regulates *Tbr1* at the post-transcriptional level. This occurs through an interaction between Dgcr8 and an evolutionarily conserved stem-loop hairpin structure formed by the *Tbr1* mRNA, similar to that of miRNAs, which the Microprocessor complex can exert a repressive function on. No further work has

investigated the relationship or interaction between DGCR8 and *TBR1.* TBR1 is a transcription factor essential in the development of layer VI neurons and therefore disruption of *TBR1* could have significant implications in brain development. This interaction could potentially indicate a role for DGCR8 in SCZ.

1.3.3.4 Models of *DGCR8* loss and effects on mRNA and protein expression

There are numerous *in vitro* and *in vivo* approaches described to model and investigate manipulation of *DGCR8.* As previously stated, *Dgcr8* null mice are embryonic lethal and therefore cannot be utilised for modelling, so other approaches have been explored to understand the function of DGCR8.

Many methods have been utilised to reduce expression at both the mRNA and protein level. It is noteworthy that across different models, genetically heterozygous and homozygous manipulation has varying and inconsistent effects on mRNA, protein and consequently miRNA expression, which is summarised below in Table 1.3. Homozygous Dgcr8 loss has been achieved in mouse embryonic stem cells (mESCs), which show complete protein loss, global loss of miRNA and pluripotency defects (Wang *et al.*, 2007; Cirera-Salinas *et al.*, 2017). Heterozygous mESCs models however do not display equivalent loss of protein and miRNA expression and are comparable to wildtype. Numerous mouse models have been generated, using either conditional knockouts in specific cell types or using *Dgcr8*$^{+/-}$ mouse models. Varying levels of reduction of *Dgcr8* expression is observed, with mRNA levels mostly being reported to reflect heterozygous loss, but miRNA reduction and protein expression reduction show mixed results (see Table 1.3).

In the literature, there is only one reported human embryonic stem cell (hESC) line modelling *DGCR8* loss, when full protein ablation occurred, there were either self-renew defects or karyotyping abnormalities (Deng *et al.*, 2019). DGCR8 expression has been investigated in cells derived from 22q11.2DS patients, with peripheral leukocytes showing both ~50% reduction at the mRNA and protein level. This indicates that heterozygous deletion does lead to heterozygous reduction in 22q11.2DS patients. Conversely, induced pluripotent stem cells (iPSCs) from 22q11.2DS patients with SCZ, show reduced *DGCR8* mRNA expression only when differentiated into neurospheres,

however reduction was not observed in fibroblasts or in undifferentiated iPSCs. Therefore, to recapitulate the 22q11.2DS phenotype, it is important to replicate DGCR8 expression in the model of interest, as the models potentially indicate that protein reduction is relevant to understanding phenotypes produced by loss of *DGCR8*.

Table 1.3: Reported *DGCR8* models in the literature and subsequent effects on mRNA, protein and miRNA expression.

Model	Type of DGCR8 loss	Observed levels of mRNA or protein reduction	Observed miRNA loss	Paper
Mouse ES cells	Complete knockout, remove exon 3, premature stop codon.	Neither reported.	Global loss of miRNAs.	(Wang *et al.*, 2007)
Mouse ES cells	Heterozygous knockout, remove exon 3, premature stop codon.	Neither reported.	No miRNA loss.	(Wang *et al.*, 2007)
22q11.2 del mouse model *Df(16)A*$^{+/-}$	1.3Mb chromosomal deficiency from *Dgcr2* to *Hira* (27 gene deletion).	Protein not reported, but 57% reduction at mRNA level in prefrontal cortex and 64% in hippocampus.	In most cases, miRNA expression was reduced ~20-70% but some were upregulated.	(Stark *et al.*, 2008)
Mouse model but only looked at PFC.	Dgcr8$^{+/-}$ mouse model.	Protein not reported, but in PFC at P5 no significant reduction but at P25 significant ~50% reduction at mRNA level.	Not reported.	(Schofield *et al.*, 2011)
Mouse model generated from human line XH157 then injected into blastocysts	Dgcr8$^{+/-}$ mouse model – first exon disrupted.	Protein not reported. mRNA extracted from hippocampus, showed ~20% reduction.	Not reported.	(Earls *et al.*, 2012)
Mouse model specifically pyramidal neurons in neocortex and hippocampus	Conditional Dgcr8$^{+/-}$ deletion.	Protein not reported PFC ~40% mRNA reduction.	Significant reduction of the 5 miRNAs they analysed.	(Hsu *et al.*, 2012)

Mouse brain: neural progenitor cells	Knockdown via shRNA expression.	Not reported.	No miRNA loss.	(Knuckles *et al.*, 2012)
Mouse model specifically looking at hippocampus	$Dgcr8^{+/-}$ mouse model. Interrupted intron 8.	Reduced wild-type protein and production of chimeric protein instead.	11 miRNAs show a decrease, but others remained comparable to WT.	(Stark *et al.*, 2008; Ouchi *et al.*, 2013)
Peripheral leukocytes from 22q11.2 del patients	22q11.2 del patients with one copy of *DGCR8.*	50% reduction observed at protein level (and mRNA).	Mixture of up and down regulated miRNAs.	(Sellier *et al.*, 2014)
iPSCs from 22q11.2 del with SCZ	iPSCs differentiated into neurons.	~50% reduction of *DGCR8* mRNA.	Mixture of up and down regulated miRNAs.	(Zhao *et al.*, 2015)
iPSCs from 22q11.2 deletion patients with SCZ	Fibroblasts before IPSC conversion.	Slightly increased mRNA expression compared to controls without deletion.	Not reported.	(Toyoshima *et al.*, 2016)
iPSCs from 22q11.2 deletion patients with SCZ	IPSCs derived from 22q11.2 del patients undifferentiated.	Slight reduction in mRNA expression but not significant.	Not reported.	(Toyoshima *et al.*, 2016)
iPSCs from 22q11.2 deletion patients with SCZ	IPSCs derived from 22q11.2 del patients undifferentiated into neurospheres.	Significant mRNA reduction compared to controls without deletion.	Downregulation.	(Toyoshima *et al.*, 2016)
Mouse telencephalon	Conditional deletion – heterozygous.	Reduced (~50%) protein levels for heterozygous.	~15% decrease.	(Marinaro *et al.*, 2017)
Mouse ESC line	Homozygous knockout	Complete loss of protein.	Not reported.	(Cirera-Salinas *et al.*, 2017)
Mouse telencephalon	Conditional deletion – knockout.	Complete loss of protein.	~35% decrease.	(Marinaro *et al.*, 2017)
Human ESC line	Homozygous *DGCR8* exon 3 deletion as reported in Wang et *al*, 2007.	Truncated non-functional protein produced but exhibited abnormal karyotype.	Not reported.	(Deng *et al.*, 2019)
Human ESC line	Homozygous deletion of *DGCR8* exon 2.	Complete protein loss but poor self-renewal capacity.	Not reported.	(Deng *et al.*, 2019)

Human ESC line	N-terminal truncated version of DGCR8 (retains miRNA processing).	Production of truncated protein with only loss of N terminal.	No global effects on miRNA processing.	(Deng *et al.*, 2019)

These lines of evidence indicate that haploinsuffiency of *DGCR8* could contribute to the SCZ risk in 22q11.2DS. The potential role of *DGCR8* and its contribution to SCZ risk in 22q11.2DS is further discussed in Chapters 3,4 and 5.

1.3.4 Multiple gene model

The relative importance of individual genes within the 22q11.2 region that contribute to the cognitive and behavioural phenotypes of 22q11.2DS have been extensively investigated. However, no one singular gene has been shown to be responsible for the neuropsychiatric risk caused by the deletion. There is accumulating evidence supporting the hypothesis that heterozygous loss of multiple genes spanned by the deletion increase the risk for the psychiatric symptoms associated with 22q11.2DS. Large GWAS and exome sequencing studies for idiopathic SCZ have not implicated a single risk locus within the 22q11.2 deletion, which could indicate that neuropsychiatric risk is due to the compound effect of haploinsuffiency of multiple genes (Pardiñas *et al.*, 2018). The 22q11.2 genes *Prodh* and *Comt* have been shown to have a transcriptional and behavioural interaction. This interaction is sensitive to pharmacological manipulation with psychoactive drugs by modulation of dopaminergic signalling in the frontal cortex (Paterlini *et al.*, 2005). Furthermore, bioinformatic analysis has revealed the existence of conserved transcription factor binding sites in 22q11.2 genes that are co-regulated and active during specific windows during development (Amati *et al.*, 2007). The effects of diminished dosage of multiple genes that act on similar pathways or have common cellular targets has been eluded to as a potential mechanism underlying 22q11.2DS. Meechan *et al* describe a hypothesis of functional groups of 22q11.2 genes, in which there is coordinated regulation to orchestrate specific cellular, physiological or morphogenetic processes. Subsets of specific genes involved in transcription regulation, cell cycle and mitochondrial function were determined bioinformatically and through cell biology methods. They determined that diminished dosage of these "groups" leads to specific phenotypic consequences relating to 22q11.2DS (Meechan *et al.*, 2007).

1.3.5 22q11.2 Mouse Models

There is a high level of genetic conservation between humans and mice. Genetic linkage analysis identified a section of the mouse chromosome 16 that is orthologous to the 1.5Mb deletion at the 22q11.2 locus in humans (Puech *et al.*, 1997). Similar to the human 22q11.2 locus, the mouse region is also flanked by LCRs making it susceptible to meiotic error (Puech *et al.*, 1997). Multiple mice strains have been engineered to model the 22q11.2 deletion, as shown in Figure 1.6. The first 22q11.2DS mouse model was generated by deleting an ~1.2Mb region (*Df1*) of mouse chromosome 16 and was subsequently named *Df1+/-* mice. The deleted region contains 22 functional genes orthologous to the human 22q11.2 region (Lindsay *et al.*, 1999). The mice exhibited congenital heart defects and some behavioural abnormalities (Lindsay *et al.*, 1999). Murine models with longer deletions have since been generated, carrying a 1.3Mb hemizygous deletion spanning 27 genes from *Dgcr2* to *Hira* such as *Lgdel*$^{+/-}$ and *Df(16)A+/-* (Merscher *et al.*, 2001; Stark *et al.*, 2008). Although there is high conservation of genes, the genes are arranged in a slightly different order and mice only have one functional *Dgcr6,* whereas humans have two as a result of a duplication (*DGCR6* and *DGCR6L*).

Behavioural deficits have been observed in many of the 22q11.2DS mouse models. PPI, which is disrupted in 22q11.2DS patients (Sobin, Kiley-Brabeck and Karayiorgou, 2005) has also been shown to be disrupted in mouse models (Paylor, 2001; Stark *et al.*, 2009). Furthermore, working memory deficits have been observed in the *Df(16)A*$^{+/-}$ model, that were in part due to the deficiency of *Dgcr8* (Stark *et al.*, 2008). Fear conditioning has also been used to assess cognition in 22q11.2 mouse models. Impaired conditioned fear memory and deficits in cued and contextual fear memory have been found in two different mouse models (Paylor, 2001; Stark *et al.*, 2008). Furthermore, aberrant spatial memory in adult 22q11.2DS mice has also been observed, this deficit has been linked to enhanced hippocampal long-term potentiation (LTP) (Earls *et al.*, 2010, 2012).

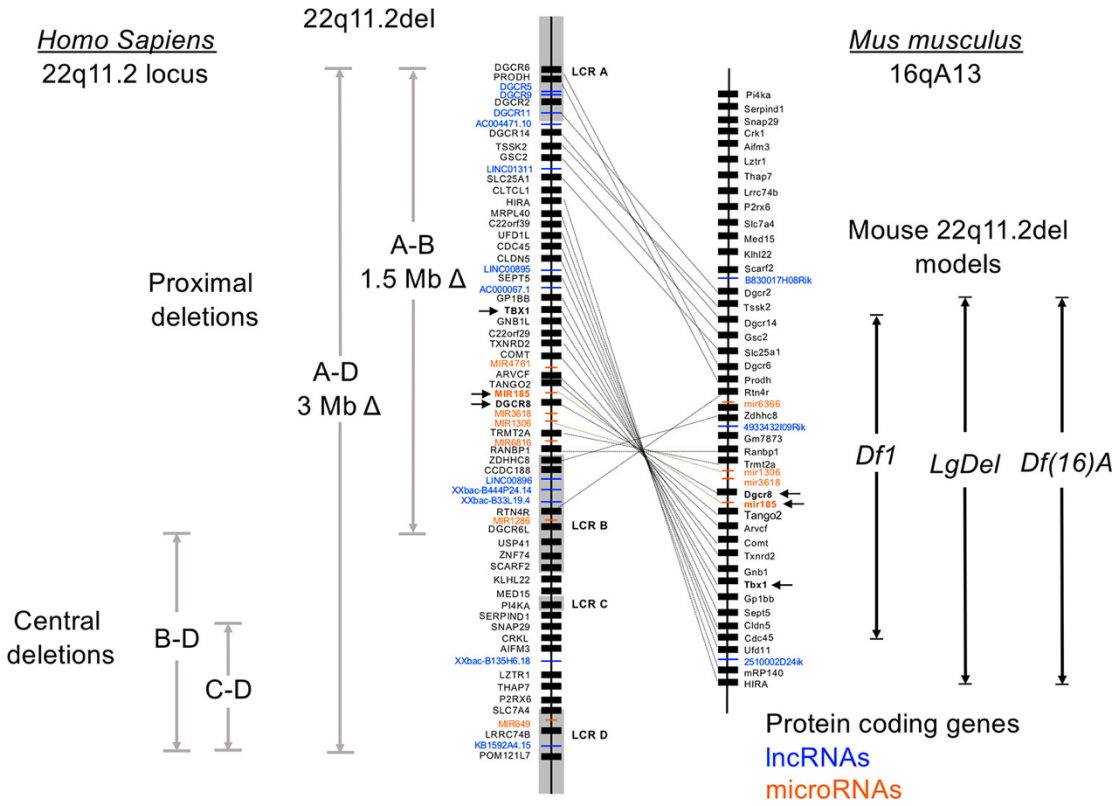


Figure 1.6: Human chromosome 22q11.2 region, syntenic mouse 16qA13 region and corresponding mouse models. *Black squares represent one gene. The corresponding genes on the murine locus are connected with lines to the human locus, with many conserved genes, but a variation in gene order. Three of the commonly used 22q11.2DS mouse models shown on the right. Figure adapted from Du et al 2020.*

1.4 Molecular mechanisms underlying the neuropsychiatric risk and phenotype variability in 22q11.2DS

All 22q11.2DS patients carry a hemizygous deletion at region 22q11.2. Although the deletion is common to all patients, the hallmark of 22q11DS genotype/phenotype association studies is the lack of consistent phenotype. There is significant phenotypic variability. Some patients exhibit the full syndromic phenotype including cardiovascular, craniofacial morphologies and psychiatric anomalies, whereas other patients can be relatively asymptomatic (Mclean *et al.*, 1993; Digillo *et al.*, 2003). There are several suggested hypothesises to explain the significant phenotypic variability, which are described in the next section.

1.4.1 The Deletion Size

The clinical heterogeneity observed in 22q11.2DS is not accounted for by the size of the deletion. There has been no robust or consistent correlation to distinguish between the 1.5Mb or 3Mb deletion and clinical and neuropsychiatric phenotypes (Lindsay, Goldberg, *et al.*, 1995; Carlson *et al.*, 1997). A recent meta-analysis found a lack of association between deletion size and the congenital heart disorders and palate abnormalities observed 22q11.2DS, suggesting the deletion size does not explain the incomplete penetrance of these symptoms (Rozas *et al.*, 2019). However, there is evidence showing that an atypical distal deletion, which does not overlap with the LCR A-B region might exhibit differing phenotypes (Rauch *et al.*, 2005).

1.4.2 Recessive mutations

22q11.2DS is caused by a hemizygous deletion, therefore affecting one copy of the genes spanning the region. Allelic variations on the non-deleted homologous region is a speculated contributor to phenotypic variability in 22q11.2DS. Mutations on the non-deleted allele have been found to occur in some 22q11.2DS patients that display atypical phenotypes. Children with biallelic mutations in the *PRODH* gene (including the presence of a 22q11.2 deletion) have Hyperprolinemia type I and display a homogeneous severe neurological phenotype (Bender *et al.*, 2005; Afenjar *et al.*, 2007). Furthermore, hemizygous mutations in *GP1BB* and the presence of the 22q11.2 deletion have been shown to cause the recessive disorder Bernard-Soulier syndrome, but the infants with mutation lacked the common 22q11.2DS phenotypes (Kunishima *et al.*, 2013). Conversely, mutation screening of the non-deleted *TBX1* in 22q11.2DS patients for cardiac phenotypes showed no difference between with or without congenital heart defects (Rauch *et al.*, 2004).

1.4.3 Genetic variability and modifiers outside the deletion

Another suggested driver for the phenotypic variability in 22q11.2DS is the presence of genetic modifiers outside of the deletion region. There are cases of 22q11.2DS individuals passing on the deletion to offspring, with children displaying mild to severe phenotypes varying from their parent. This provides evidence that other genetic factors outside of the 22q11.2 region contribute to phenotype (Leana-Cox *et al.*, 1996).

Furthermore, increased risk for SCZ has been associated with multiple classes of mutations (Rees, O'Donovan and Owen, 2015).

The presence of other CNVs in the genome has been explored for contribution to SCZ risk in 22q11.2DS. CNVs may contain dosage sensitive genes or genes that act in similar genetic or cellular pathways as the 22q11.2 genes, therefore having an additive effect on phenotype. It has been reported that SCZ patients with a known pathogenic CNV (such as a 22q11.2 deletion) carry an excess burden of common risk alleles compared to the general population (Tansey *et al.*, 2016). The presence of CNVs in 48 22q11.2DS patients with and without SCZ revealed that 22q11.2DS patients with psychosis on average had significantly larger CNVs than those without (Williams *et al.*, 2013). Furthermore, a recent study on a larger cohort of 22q11.2DS patients with and without SCZ (158 with any SCZ spectrum disorder and 171 non-psychotic patients) revealed differences in the CNVs between groups, with gene set enrichment showing enrichment of pathways related to SCZ in the CNVs present in 22q11.2DS-SCZ group (Bassett *et al.*, 2017). Although this study was unable to detect specific CNV differences between patients, it adds to the evidence for genetic variation outside of the 22q11.2 region contributing to SCZ risk.

A small, underpowered study with whole-genome sequencing data from 22q11.2DS patients with or without SCZ, identified an increased number of rare deleterious variants that were enriched in neurofunctional protein-coding genes in the 22q11.2DS-SCZ cohort. However this result did not survive multiple-testing correction (Merico *et al.*, 2015). Nevertheless, a multiple-hit pathway hypothesis postulates that a first hit (22q11.2 deletion) lowers the threshold for expression of additional genetic hits elsewhere in the genome (Zinkstok *et al.*, 2019). Another recent study with whole-exome sequence data and comparative-genomic hybridization array of a unique family with multiple members affected by 22q11.2DS revealed *de novo* and inherited rare and damaging variants outside of the 22q11.2 region. One such variant was a paternal 3p26.3-loss that has previously been associated with psychosis (Williams *et al.*, 2013; Michaelovsky *et al.*, 2019). Furthermore, polygenic risk for SCZ within 22q11.2DS was significantly greater patients with SCZ than those without (Cleynen *et al.*, 2020). Genetic

variability and genetic modifiers in 22q11.2DS are further discussed in detail in Chapter 3.

1.5 Genetic approaches to understand SCZ pathology

Family, twin and adoption studies have provided substantial evidence elucidating the roles of genetic and environmental factors contributing to SCZ. Heritability is a parameter used to estimate how much of the variability of a phenotypic trait is due to the genetic variability in a given population. For SCZ, the heritability estimate is above 80% (Sullivan, Kendler and Neale, 2003). However, it is noteworthy that environmental factors also play a critical role in the aetiology of SCZ, with a diverse array of factors causing increased risk, such as maternal malnutrition and urban environment (Susser *et al.*, 1996; Vassos *et al.*, 2012).

1.5.1 Linkage analysis and candidate gene studies

Early work investigating the genetic architecture of SCZ was based upon the idea that single genes or common risk variants were underlying the disorder. Linkage and candidate gene studies in families were commonly used to investigate SCZ genetics. Linkage analysis is based on the observation that genetic markers which are physically close together are transmitted/inherited together as a unit during meiosis (Lander and Kruglyak, 1995). The positive findings of these studies were generally not replicated in following studies, but results from subsequent meta-analyses suggest that many chromosomal regions may contain SCZ susceptibility loci, however it is the variants these loci harbour that confer the risk (Badner and Gershon, 2002; Ng *et al.*, 2009).

The candidate gene approach using case-control study design aimed to test if the frequency of susceptibility genes is higher in affected individuals than in unaffected. In contrast to linkage analysis, the candidate gene approach can detect genes with small effect alleles provided that the sample size is adequate. Numerous candidate genes have been tested for and well known identified genes include Disrupted in Schizophrenia 1 (DISC1) (Blackwood *et al.*, 2001), Neuregulin 1 (NRG1) (Stefansson *et al.*, 2002) and Dysbindin1 (DTNBP1) (Straub *et al.*, 2002). However, there has also been the problem of non-replication in these studies. A comprehensive study of some the most cited

candidate genes (including the above mentioned), were tested by genotyping a sample of 1870 cases and 2002 controls and no associations were found (Sanders *et al.*, 2008).

The rapid evolution and reduction in price of high throughput next generation sequencing and large-scale consortia efforts have made significant advances in identifying genetic variants contributing to SCZ risk.

1.5.2 Genome-wide association studies (GWAS)

GWAS are based on analysis of markers from whole-genome sequenced data, in order to identify differences in genetic variants between disease affected individuals compared to unaffected controls. This approach is hypothesis-free and has led to the identification of common variants such as single nucleotide polymorphisms (SNPs) or rare chromosomal CNVs (McCarroll, 2008; Bush and Moore, 2012). SNPs are a common occurrence in the genome (~1 in every 300 bases) (Kruglyak and Nickerson, 2001; Reich, Gabriel and Altshuler, 2003). GWAS have been successful in revealing SNPs that contribute to risk for numerous neuropsychiatric disorders. GWAS evaluate hundreds of thousands or millions of SNP markers, therefore it requires a large sample sizes to achieve adequate statistical power, as expected power increases with sample size and effect size (Klein, 2007). Testing such a large number of SNP markers, subsequently requires lots of multiple comparisons and therefore increases false positive rates. To avoid false positives, correction methods such as Bonferroni and FDR are widely used. A significance threshold is applied that controls for testing millions of DNA variants (based on 1 million independent tests in a comprehensive GWAS, the standard threshold adopted for 'genome-wide significance' is $P < 5 \times 10^{-8}$) (Hommel, 1988; Benjamini and Hochberg, 1995). Therefore, estimating a sufficient sample size to achieve adequate statistical power is crucial in the design stage of genetic association studies. When GWAS studies first began, it was directed by the common-disease common variant hypothesis (Reich and Lander, 2001; Pritchard and Cox, 2002), however this has since been refuted as it is observed that identified risk loci of moderate effect and intermediate frequency do not explain the percent of disease risk in a population. Subsequently, the genetic component of a disease has been attributed to 1) a large number of small effect common variants, 2) a small number of large-effect rare variants or 3) a combination of genetic, environmental and epigenetic interactions (Gibson, 2012). This genome-wide

data-driven approach has allowed for identification of common variants, however adequate sample size has been integral for success for a sufficiently-power GWAS in SCZ (Sullivan *et al.*, 2018). The most up-to-date SCZ GWAS included 11,260 cases and 24,542 controls, identifying 145 significant loci (Pardiñas *et al.*, 2018). Different types of variants have been associated with SCZ, including CNVs, de novo variants, common variants (GWAS) and rare variants (Avramopoulos, 2018).

Despite the success of these studies, biological interpretation of GWAS results remains challenging. Neighbouring variants are often correlated with one another and tend to be inherited together due to linkage disequilibrium (LD) (Gabriel *et al.*, 2002). LD results in multiple variants in a locus being present in an individual solely based on this correlation, therefore disentangling the associated causative variant is difficult. Furthermore, approximately >90% of associated variants lie within non-coding regions complicating functional understanding (Schaub *et al.*, 2012). There is evidence indicating these variants lie within DNA regulatory elements, including promotor and enhancer elements and enrichment within expression quantitative trait loci (Cookson *et al.*, 2009; Nicolae *et al.*, 2010; Maurano *et al.*, 2012). In order to identify causative variants and genes, integrating molecular signatures such as gene expression, DNA methylation or transcription factor binding with GWAS results could improve functional interpretation and ultimately genetics-driven understanding of pathology. Such methods and approaches for identifying disease causative genes are further discussed in Chapter 3.

1.6 Cortical development and its Relevance to Neuropsychiatric Disorders

The development of the cerebral cortex requires precise orchestration of complex sequential processes. Our comprehension of this multifaceted process has led to the understanding that many neuropsychiatric disorders are linked to aberrant events during neurodevelopment

1.6.1 Early Brain Development

At the gastrulation stage of early vertebrate embryonic development, neural fate is induced in the ectoderm (Wilson and Hemmati-Brivanlou, 1995). Neural induction is initiated by signals derived from the dorsal mesoderm to induce the formation of the

neural plate. The inductive signals, including noggin and chordin act by inhibiting bone morphogenic protein signalling to promote neural development (Piccolo *et al.*, 1996; Zimmerman, De Jesús-Escobar and Harland, 1996). Patterning of the neural plate is vital for development of the cellular and morphological intricacy of the central nervous system. This is instigated by the establishment of polarity along the anterior-posterior (AP) and dorsal-ventral (DV) axes of the neural plate. Subsequently, the neural plate lengthens along the AP axis and narrows and folds forming the neural tube (Copp, Greene and Murdoch, 2003). The neural tube is composed of neuroepithelial cells (NEPs), which give rise to multitude of different cell types that compose the central nervous system (Miyata, 2008). The NEPs are located in the ventricular zone (VZ), the most rostral region of the neural tube will give rise to the cerebral cortex.

1.6.2 Corticogenesis

The cerebral cortex is the largest and regarded as the most complex structure of the brain. The cortex regulates the highest cognitive functions involved in consciousness and perception (Frith and Dolan, 1996). Consequently, understanding corticogenesis and cortical function are of particular interest in neuropsychiatric disorders.

The cerebral cortex originates from the most rostral region of the neural tube, which later divides into the telencephalon and diencephalon. The dorsal region of the telencephalon gives rise to the cerebral cortex, whereas the ventral telencephalon differentiates into the basal ganglia. These differences are established by a gradient of diffusible morphogens, which are released from specific areas in a coordinated temporal fashion. Such morphogens include Wnt, which is released caudally, dorsal release of bone morphogenic proteins and sonic hedgehog which is released ventrally (Ciani and Salinas, 2005; Liu and Niswander, 2005; Lupo, Harris and Lewis, 2006). This patterning of the telencephalon occurs along the DV axis, which is the region that cortical excitatory neural precursors are generated (Sur and Rubenstein, 2005). The neocortex consists of various neuronal cell types, the most abundant cell type is excitatory pyramidal neurons (~85%) and GABAergic inhibitory interneurons represent the remaining cell population (Marín and Müller, 2014). Corticogenesis is the term used to describe the generation of pyramidal neurons from the NEPs located in the VZ of the neural tube (Tiberi, Vanderhaeghen and van den Ameele, 2012).

NEPs proliferate symmetrically in the VZ to generate the initial pool of progenitor cells, later the NEPs switch to asymmetric division to give rise to radial glia cells (RGCs) (Rakic, 1995; Sahara and O'Leary, 2009). RGCs reside within the ventricular zone (VZ) of the neural tube and oscillate radially along the apical-basal process through the zone during cell division (Taverna and Huttner, 2010). Asymmetric division of the RGCs generate basal progenitor cells (or intermediate progenitors), another class of cortical progenitor cells which are a transient progenitor population and eventually divide and differentiate into excitatory neurons (Malatesta, Hartfuss and Götz, 2000; Götz and Huttner, 2005). Asymmetric division continues the pool of progenitor cells by giving rise to one progenitor capable of self-renewal and the other cell committed to terminal differentiation. The apical domain of RGCs is key in regulating the balance between proliferation and differentiation, central signalling pathways are active in this, in order to support the proliferation of the progenitor pool including Wnt/ β-Catenin, Notch, Fgf and Shh (Tiberi, Vanderhaeghen and van den Ameele, 2012). Intermediate progenitors are localised in the subventricular zone, it is these cells that provide the expansion of cortical volume in primates and humans (Miyata *et al.*, 2004; Noctor, Martínez-Cerdeño and Kriegstein, 2008). Atypical proliferation of the neural progenitor population will directly impact the number of postmitotic neurons generated and have implications for brain size (Homem, Repic and Knoblich, 2015). Proliferation defects in opposite direction are associated with megalencephaly and microencephaly in which many patients have developmental delay and intellectual disability (Parrini *et al.*, 2016). Furthermore, many studies have identified reduced total volume and neuronal number in SCZ patients (Pakkenberg, 1990; Haijma *et al.*, 2013). Conversely, cortical surface area has been shown to be increased a cohort of ASD patients (Ohta *et al.*, 2016). Additionally, genes associated with neurogenesis have been implicated in for conferring genetic risk for development of neuropsychiatric disorders (Lee *et al.*, 2019)

1.6.3 Cortical layer formation

The waves of neuronal differentiation generate the layered structure of the cortex. The cortex consists of six stratified layers that are formed in an inside-out fashion, with early-born neurons forming the deep layers and subsequent late-born neurons migrate past them forming the proceeding layers as demonstrated in Figure 1.7 (Greig *et al.*, 2013; Molnár *et al.*, 2019). The cortical neurons of layers I to V are formed by asymmetric

division of cells directly generated from RGCs in the VZ or indirectly from intermediate progenitors in the subventricular zone (Noctor *et al.*, 2004). The expression of T-box brain protein 1 (Tbr1) distinguishes layer VI cortico-thalamic neurons (Hevner *et al.*, 2001). Tbr1 is a transcription factor essential in the development of layer VI neurons, it represses FEZ Family Zinc Finger 2 (Fezf2) and consequently is vital for arrangement of layer V sub-cerebral projection neurons, typically characterised by the expression of COUP-TF-interaction protein 2 (Ctip2) (Bedogni *et al.*, 2010; McKenna *et al.*, 2011). Ctip2 has downstream action to Fezf2 and regulates axonal projections to subcortical targets in order to determine fate (Chen *et al.*, 2008). Layer II and III is composed of neurons that project to other cortical regions such callosal projection neurons (Alcamo *et al.*, 2008). The repression of Ctip2 by special AT-rich sequence binding protein 2 (Satb2) is the critical step for identifying this class of neurons and ensuring appropriate development of their axonal projections (Britanova *et al.*, 2008). Such transcription factors have been extensively studies and are often used as markers of specific cortical neuron populations and layers.

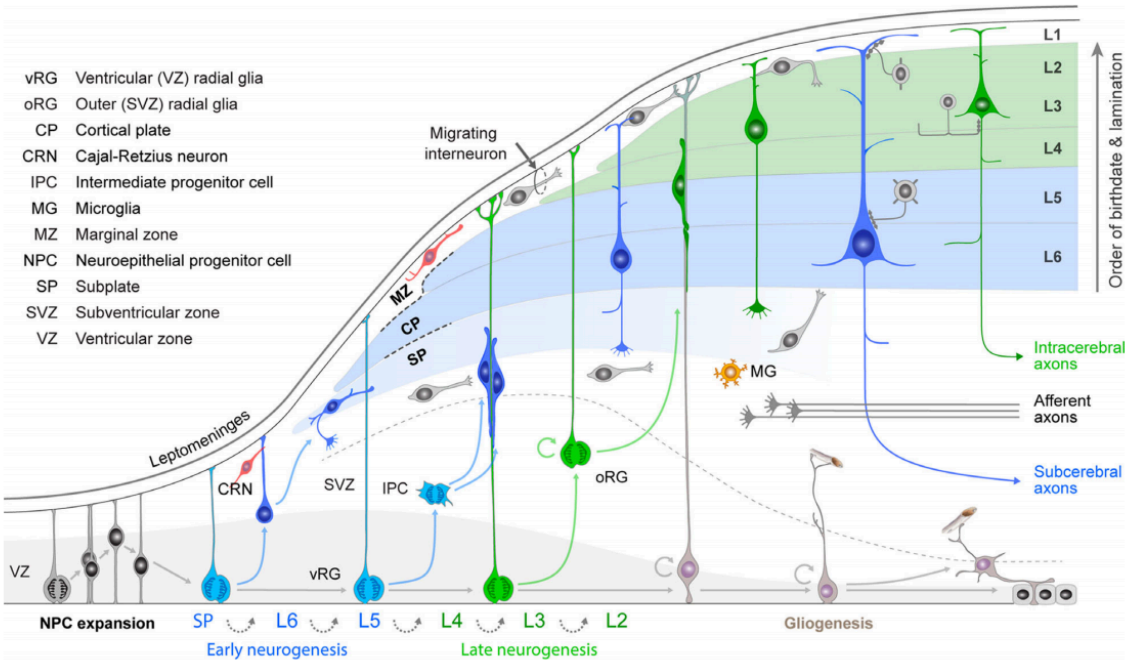


Figure 1.7 Schematic representation of development and formation of mammalian cortex.
As the cortical telencephalon develops, neuroepithelial cells give rise to radial glia cells (RGCs). Asymmetric division of RGCs generate intermediate progenitors, those committed to neuronal fate migrates radially from the ventricular zone (VZ) along the basal process into the cortical plate. The earliest born neurons migrate to form the preplate. Later migrating neurons split the preplate into the marginal zone (MZ) and subplate (SP). As neurogenesis continues, various subtypes of neurons are generated. Early-born nascent projection neurons settle in the deep layers (Layers 5 and 6; shown in blue), and later-born projection neurons migrate towards mid-neurogenesis stage. Neocortical projection neurons mature into cortical projection neurons, which show layer- and subtype-specific morphology and axonal projection patterns. Over time, interneurons arrive from ganglionic eminences and establish themselves in the cortical plate. At the end of corticogenesis, the remaining progenitors switch to the generation of astrocytes and oligodendrocytes. Figure from Molnar et al 2019.

Although much of our understanding of cortical development has stemmed from animal models, these models cannot recapitulate all cellular and molecular features of the human neocortex. Furthermore, pathologies of neurological disorders are often specific to particular brain regions or cell types. A number of neuronal cell types have been implicated in neuropsychiatric disorders including cortical projection neurons (Heckers, 2000) and inhibitory interneurons (Donegan and Lodge, 2017). Therefore, the ability to mimic development and maturation of these cell types can provide key insights for disease modelling, which has been made possible using human pluripotent stem cells (PSCs). In depth knowledge of the *in vivo* process required to direct cortical development and the characteristics of normal cortical development has enabled generation of efficient differentiation protocols of pluripotent stem cells (PSCs). Furthermore, human

PSC differentiation largely recapitulates human development, further validating the model. Consequently, differentiation of human PSCs toward cortical fate is often used for *in vitro* modelling of neuropsychiatric disorders, which is further discussed below (Soliman *et al.*, 2017).

1.7 SCZ disease modelling with human pluripotent stem cells

Historically, SCZ was studied using post-mortem tissue and animal models. Post-mortem tissue is uninformative for studying earlier disease progression as it can only represent the final stages of disease and can also be influenced by medication. Furthermore, it is now commonly accepted that SCZ is a neurodevelopmental disorder (Weinberger, 1986; Murray, Lewis and Lecturer, 1987), with evidence suggesting that deficits start *in utero* continuing through to adolescence, therefore post-mortem evaluation is not beneficial for understanding initial disease progression. Animal models, primarily rodents, are widely used for modelling SCZ and have been beneficial for reflecting the positive symptoms of SCZ, however there is little replication of the cognitive symptoms as reviewed by *Jones et al* (Jones, Watson and Fone, 2011).

Use of human embryonic stem cells (hESCs) have a become a fundamental tool for modelling human development and disease. hESCs are derived from the inner cell mass of human blastocysts and have the ability to continuously self-renew (Thomson, 1998). Furthermore, hESCs have extensive differentiation properties and so *in vitro* neural differentiation is a valid method to model neuropsychiatric disorders (Mertens *et al.*, 2016). It is possible to genetically edit hESCs to model disease variants using gene editing technologies.

An alternative to hESCs are induced pluripotent stem cells (iPSCs), which have become a fundamental tool for modelling human development and disease. Generation of iPSCs is based on the ability to reprogram differentiated cells, commonly skin fibroblasts cells, into iPSCs by expressing key pluripotency genes such as *OCT4*, *SOX2*, *c-MYC* and *KLF4* in somatic cells (Takahashi and Yamanaka, 2006; Takahashi *et al.*, 2007). Reprogrammed iPSCs will be genetically identical to the patient they are derived from, therefore can carry genetic variants associated with disease risk. Furthermore, use of adult somatic cells for iPSC generation removes the ethical issues associated with hESC use. iPSCs are believed to have the same self-renewal and differentiation properties as hESCs (Abu-

Dawud *et al.*, 2018). iPSCs have become increasingly important in psychiatric research due to the ability to investigate disease mutations in phenotypically relevant cell types.

Although hESCs and iPSCs share similar intrinsic properties, there are some notable differences. In terms of differentiation potential, there is more variability observed when differentiating iPSCs, particularly notable in neuronal differentiation (Hu *et al.*, 2010). Furthermore, extensive differences in gene expression signature have been observed between hESCs and iPSCs with transcriptomic comparison (Chin *et al.*, 2010). These differences may in part, be due to the "epigenetic memory" of the somatic cell used to derive the iPSCs, which has been show to affect gene expression (Kim *et al.*, 2010; Polo *et al.*, 2010).

Both hESCs and iPSCs are PSCs and therefore have extensive differentiation properties, consequently *in vitro* neural differentiation has become an invaluable tool used to investigate alterations in cellular and molecular processes. This approach allows for study of potential defects that arise during formation and maturation of neurons and therefore has become a key approach for modelling neuropsychiatric disorders. PSCs can be differentiated into neural progenitor cells (NPCs) and specific neuronal subtypes such as cortical glutamatergic pyramidal neurons, this process largely mimics vertebrate development (Boissart *et al.*, 2013; Espuny-Camacho *et al.*, 2013). Neural differentiation is achieved by replicating the specific signalling environment in development by exposing progenitors to 'inductive' molecules in order to generate the desired neural progenitor phenotype. PSCs can be differentiated into neural epithelial cells, then to neural progenitor cells, which are the precursor cells to neurons, early deficits in NPCs *in vivo* likely lead to deficits in later development, therefore they are considered an appropriate model to study neurodevelopmental defects. Monolayer differentiation of PSCs is a common approach for generating neural stems using a differentiation protocol of dual SMAD inhibition (Chambers *et al.*, 2009). With high replication of the sequential formation of cell types from the cortical layers observed in *in vivo* corticogenesis (Gaspard *et al.*, 2008; Handel *et al.*, 2016). Due the known sequential appearance of various cortical markers *in vivo,* such markers (as described in section 1.6.3) can be used to examine cortical development through differentiation of PSCs. This approach has been used for many SCZ disease modelling studies in PSCs.

iPSCs derived from patients with SCZ have been well studied and characterised. Brennand and colleagues were the first to report iPSCs from schizophrenia patients have alterations in neuronal connectivity and morphology (Brennand *et al.*, 2011). Additionally, synaptic deficits and transcriptomic alterations have been identified in neural differentiation of iPSCs derived from SCZ patients mutations in the known SCZ associated risk gene, *DISC1* (Wen *et al.*, 2014). Furthermore, iPSC studies from 22q11.2DS patients with SCZ have been used to elucidate underlying mechanisms in SCZ risk. Aberrant miRNA expression has been identified in neurons generated from 22q11.2DS patients with SCZ compared to healthy controls without the deletion (Zhao *et al.*, 2015; Toyoshima *et al.*, 2016). Use of iPSCs from 22q11.2DS patients have implicated certain 22q11.2 deletion genes in contributing to phenotypes observed in neural differentiation. Reduction in mitochondrial ATP in iPSCs derived excitatory neurons from 22q11.2DS patients with SCZ have also been reported, which was reportedly due to *MRPL40* loss (Li *et al.*, 2019). Transcriptomic profiling of iPSC derived neurons from 22q11.2DS patients with SCZ and schizoaffective disorder implicated *CDC45*-based regulation of affecting the cell cycle (Lin *et al.*, 2016).

1.7.1 CRISPR/Cas9 genome editing in human PSCs

Genome editing in hESCs has allowed for targeted investigation of specific genetic mutations, as all cells will share the same genetic background, avoiding outside variability. This has become an invaluable approach for modelling SCZ-related genetic mutations. Combined with the protocols for neuronal differentiation, this has enabled the study of potential deficits during the development and maturation of neurons due to SCZ-related mutations.

The CRISPR (Clusters of Regularly Interspaced Short Palindromic Repeats)/Cas9 technology has dramatically enhanced efficiency and speed of targeted gene editing and improved disease modelling in PSCs. CRISPR/Cas9 is a type II CRISPR system which confers immunity against foreign DNA in many bacteria and archaea (Labrie, Samson and Moineau, 2010). The system is composed of the Cas9 nuclease and a small guide RNA (sgRNA), a chimeric RNA molecule combining a CRISPR RNA (crRNA) and a transactivating RNA (tracrRNA), Cas9 nuclease relies on RNA guidance for target specificity (Jinek *et al.*, 2012). A protospacer adjacent motif (PAM) upstream of the

target binding region is also required. Complementary pairing of the spacer portion of the sgRNA to the targeted DNA next to the PAM site results in generation of a blunt DNA double-strand break (DSB) by the Cas9 nuclease (as reviewed in (Pickar-Oliver and Gersbach, 2019). Therefore, site-specific cleavage at any location containing a PAM site can be achieved by designing sgRNAs containing the appropriate complementary sequence. The generation of the DSBs triggers DNA repair mechanisms. Induced DNA repair mechanisms include non-homologous end joining (NHEJ) or homology directed repair (HDR). NHEJ is the most active in the cell, but it is susceptible to frequent mutation errors and so likely to introduce insertions and deletions (INDELs) (Heyer, Ehmsen and Liu, 2010; Rodgers and Mcvey, 2016). Whereas HDR is the dominant mechanism for precise DSB repair a it requires a template to guide repair and so a donor construct is required for use in CRISPR/Cas9 editing.

CRISPR/Cas9 based gene editing has been widely applied to human cells. Gene-based KO studies using CRISPR/Cas9 technologies are frequently used to rapidly induce gene mutations in PSCs by transfecting vectors expressing the Cas9 nuclease and the designed sgRNAs (Cho *et al.*, 2013; Cong *et al.*, 2013). Other delivery methods of the CRISPR/Cas9 components include lentiviruses or adenovirus, which have shown to have high transduction and editing efficacy (Kabadi *et al.*, 2014; Maggio *et al.*, 2014; Shalem *et al.*, 2014). The use of lentiviral-based CRISPR/Cas9 delivery is further discussed in Chapters 4 and 5. Although genome-editing based gene knockout enables studying of loss of specific genes, other gene-editing approaches using CRISPR/Cas9 have since been developed. Cas9 nuclease variants such as the “Dead” Cas9 (dCas9) used in conjunction with targeting sgRNAs provide a platform to control gene expression (Qi *et al.*, 2013). Furthermore, inducible CRISPR/Cas9 systems have been developed for human stem cell research, which further improve flexibility of genome editing. The generation of the iCRISPR platform, in which hESCs lines harbour a doxycycline-inducible Cas9 nuclease, meaning only sgRNA delivery is required, improving efficiency for gene targeting (González *et al.*, 2014).

1.8 Aims of the thesis

Hemizygous deletion at the 22q11.2 region is the strongest known molecular genetic risk factor schizophrenia, however the mechanisms underlying this risk remain largely unknown. The overall aim of this thesis is to identify SCZ candidate genes within and outside the 22q11.2 deletion and explore genetic manipulation of these candidate genes in human cell models through cortical differentiation. This work was carried out with a particular focus on DGCR8 from within the deletion region.

To identify other potential SCZ candidate genes within the 22q11.2 deletion region, alongside *DGCR8,* genes were selected based upon co-expression to *DGCR8* using RNA sequencing data from the fetal brain and literature review on pathogenic effects based predicted mutation intolerance. To investigate the possibility of genetic modifiers outside of the 22q11.2 deletion region, genome and transcriptome wide association studies were undertaken comparing whole genome sequenced data from 22q11.2DS patients with and without SCZ.

To investigate how altered levels of *DGCR8* could affect cortical development, I aim to derive and characterise a hESC line deficient of DGCR8 using CRISPR/Cas9 and investigate the effects of loss of *DGCR8* through cortical differentiation *in vitro*. Additionally, I aim to investigate genetic manipulation of prior selected SCZ candidate genes in human neuroprogenitor cells during cortical development using a lentiviral-based CRISPR/Cas9 system.

2: Material and Methods

2.1 Bioinformatic analysis

2.1.1 22q11.2 IBBC Cohort

The 22q11.2 IBBC (International 22q11.2 Deletion Syndrome Brain Behaviour Consortium) Cohort has been previously detailed in Gur *et al* 2017 (Gur *et al.*, 2017) and Bassett *et al* 2017 (Bassett *et al.*, 2017). Subjects with a 22q11.2 deletion were recruited from 22 international sites from Canada, USA and Europe and provided the appropriate consent. Studies were approved by the local institutional research ethics boards. Individuals were eligible for inclusion in this study if they met DSM diagnostic criteria for a major psychotic disorder (mostly schizophrenia) at any age or had no history of any psychotic illness when assessed at age ≥ 25 years. This data was collected by the IBBC cohort.

2.1.2 Whole Genome Sequencing and processing of DNA samples (IBBC)

Raw Illumina WGS data were mapped with PEMapper (Johnston *et al.*, 2017) to genome build GRCh38/hg38 with a median sequencing depth of 39×. The 22q11.2 deletion region was called based on sequencing depth, and variants were called independently with PECaller (Johnston *et al.*, 2017) for the genome-wide diploid and 22q11.2 haploid regions of the genomes. Using principal component analysis anchored with HapMap reference samples to infer ancestry, a European subsample of 435 subjects (214 schizophrenia, 221 no psychotic illness) were further defined based on subjects whose principal components were within 1 standard deviation of the average PC value for HapMap3 CEU subjects. This data was collected by the IBBC cohort.

2.1.3 Genome Wide Association Study (GWAS)

The details of the data processing here were performed with PLINK 1.9 (Chang *et al.*, 2015) (www.cog-genomics.org/plink/1.9/) and in R (Macintosh R Open 3.4.0 (Macintosh)), based on R-3.4.0 (R Statistics). I carried out GWAS using QC-ed whole genome sequencing data provided by the IBBC and described in Cleynen *et al* 2020

(Cleynen *et al.*, 2020). I performed GWAS using logistic regression adjusted for 10 principal-component analysis (PCA) covariates. To avoid overburdening the GWAS power by adding too many covariates to the regression model, the first twenty principal components were considered and tested for inclusion. I ran a logistical regression of the principal components, when running PCA analysis, the first five principal components were chosen and then covariates were chosen as those nominally significant ($p < 0.05$) in a logistic regression for association with the phenotype. The final set of covariates included the first five PCs (as recommended for most GWAS approaches (Morgan *et al.*, 2014)) and PCs 10, 13, 14, 15 and 19. Regular LD clumping was performed (r^2=0.1, $P<1x10^{-4}$; window size < 3Mb) to obtain independent index SNPs and an R2 threshold of 0.8 was applied (Pardiñas *et al.*, 2018).

2.1.4 Transcriptome Wide Association Study (TWAS)

Using the GWAS summary statistics (as described above), using multiple methods with various gene-expression reference panels I carried out different TWAS analysis. Panels used include the gene-expression reference panel from RNA sequencing from the dorsolateral prefrontal cortex (DLPFC) of individuals collected by the CommonMind Consortium which was obtained from the FUSION website (http://gusevlab.org/projects/fusion/). The CommonMind Consortium DLPFC consists of sequenced RNA from DLPFC from schizophrenia cases, bipolar cases and controls (n = 452) (Fromer *et al.*, 2016). The second gene-expression reference panel used to perform TWAS was using RNA sequencing from the cortex of 136 individuals collect by the GTEX consortium (Aguet *et al.*, 2017), the third gene-expression panel was from whole blood generated from 1264 individuals (Raitakari *et al.*, 2008; Nuotio *et al.*, 2014), both are obtained from the FUSION website (http://gusevlab.org/projects/fusion/). TWAS was also carried out using the gene-expression panel generated from RNA-sequencing from fetal brain (O'Brien *et al.*, 2018) (which can be obtained from https://doi.org/10.6084/m9.figshare.6881825).TWAS was performed in two ways, using FUSION (Gusev *et al.*, 2016) and PrediXcan (Gamazon *et al.*, 2015) as a comparison.

Using the mentioned reference panels and summary level data from the International 22q11.2 Brain Behaviour Consortium GWAS of 435 individuals with 22q11.2DS with

and without schizophrenia (Cleynen *et al.*, 2020) to identify genes associated with schizophrenia. TWAS p values were corrected using Bonferroni correction to account for multiple testing within each expression panel used, in which the threshold of significance (0.05) was divided by the number of experiments. This procedure is consistent with the correction applied in previous TWAS results of multiple expression reference (Gusev *et al.*, 2018).

2.1.5 Fetal brain RNA sequencing and co-expression analysis (Nick Bray and Heath O'Brien)

For co-expression analysis of genes within the 22q11.2 deletion region, RNA sequencing derived from fetal brain tissue was analysed, the samples and RNA-sequencing is in described in (O'Brien *et al.*, 2018). Heath O'Brien kindly analysed and calculated expression calculated the Pearson's correlation values for the 37 expressed genes identified. Using the provided significance and Pearson's correlation values, I then performed downstream analysis on R using the corrplot package to visualise and identify clusters of correlated genes based upon Pearson correlation (Wei and Simko, 2017).

2.2 Cell culture

2.2.1 hESC culture

I used the iCas9 cell line throughout this work, for derivation of the mutant DGCR8 cell line and as the parental control (González *et al.*, 2014). Human ESCs were maintained in TesR-E8 media (STEMCELL technologies) under standard culture conditions (37°C, 5% CO_2) in 6 well-plates coated with Matrigel® (Corning, VWR).

Cells were passaged every 3-4 days, when 70-90% confluent. hESCs were washed once with DPBS and then incubated in 0.02% EDTA (Sigma) for about 3-4minutes at 37°C. The EDTA was aspirated, then cells were manually dissociated into small clumps in fresh medium and seeded onto a new plate containing RevitaCell supplement at 1x concentration (Sigma). When iCas9 cells reached 70-90% confluency, approximately 1-2x10^6 cells, they were split at ratios between 1:3-1:5.

For freezing, approximately 1-2x10^6 cells were dissociated with EDTA (as described above), cells were collected in fresh media and centrifuged at 200g for 5 minutes, then

resuspended in 1ml of cold hESC medium with 10% DMSO (Sigma). Cryovials containing the cell suspension were transferred into a freezing container and put at -80°C, cooling at 1°C per minute. Once frozen, cryovials were transferred to liquid nitrogen tanks.

For thawing, cryo-vials were placed in a water bath at 37°C and gently swirled. The thawed cell suspension was transferred into pre-warmed 10ml of TesR-E8 media (STEMCELL Technologies) and centrifuged at 200g for 5 minutes. The cell pellet was then resuspended in TesR-E8 media and plated.

2.2.2 Monolayer differentiation into cortical glutamatergic neurons

Neural differentiation was induced using a modified version of the dual SMAD inhibition protocols developed by (Chambers *et al.*, 2009; Cambray *et al.*, 2012; Arber *et al.*, 2015). The timeline of the protocol for each cell line is outlined in Figure 2.1. The differentiation medium used, N2B27, was composed of 2:1 DMEM-F12 and Neurobasal, 1x N2 supplement, 1x B27 supplement (without vitamin A), 20 mM L-Glutamine, 20μm β-mercaptoethanol (all ThermoFisher Scientific) and 1x MycoZap Plus-CL antibiotics (Lonza).

Stem cells were initially plated on 12-well plates coated with Growth Factor Reduced Matrigel (Corning, VWR), in hESC medium until 80-90% confluency. At this confluency, the cells were switched to neural induction media supplemented with 10 μM SB-431542 (SB, TGF-B inhibitor) (StemCell technologies) and 100 nM LDN-193189 (LDN, BMP4 inhibitor) (Sigma-Aldrich), neural differentiation media was changed every other day. The cells were kept in neural induction medium until day 8 for the iCas9 line, when the LDN and SB were removed.

The iCas9 cells were passaged at a 2:3 ratio on fibronectin-coated 12-well plates. Fibronectin (Millipore) was coated onto 12 well plates (15μg/ml) and incubated at 37°C for at least an hour. An hour before passaging, the cells were treated with 100μM ROCK inhibitor (Y-27632, StemCell Technologies) to prevent cell death. The cells were then incubated in EDTA for 3 minutes and manually dissociated in N2B27 medium using a 2ml

serological pipette, keeping large cell clumps. These were resuspended in the appropriate volume of neural induction medium and seeded onto the new plate.

Around day 16, the neural progenitors that form the differentiating cultures were passaged again, at a ratio between 1:3-1:4 onto Poly-D-Lysin/Laminin-coated plates. 24-well plates were coated by incubating a solution of 10µg/ml of Poly-D-Lysin (Sigma) for 1h at room temperature. This was followed by 3 washes with DPBS and an overnight incubation at 37°C with 5µg/ml laminin solution (Sigma). A few days after the second passage, when cells displayed clear neuronal morphology, the media is switched to N2B27 with B27 supplement containing vitamin A, which promotes maturation. Again, neural differentiation medium was changed every other day.

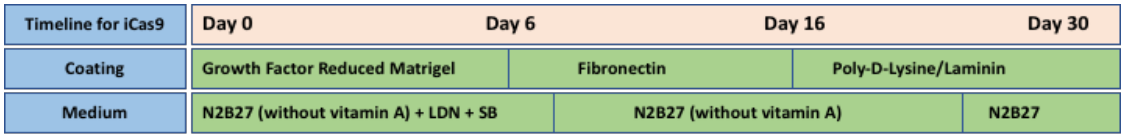


Figure 2.1: Monolayer differentiation of hESCs into cortical glutamatergic neurons.
Timing of iCas9 parental line shown in orange, with substrate and media composition used at each stage shown in green.

2.2.3 HEK 293T Culture

Adherent human embryonic kidney (HEK) 293T cells were seeded onto gelatine-coated 10cm^2 dishes or T75 flasks (Falcon) as a monolayer in HEK medium, composed of 450ml DMEM-F12 (Thermo Fisher), supplemented with 10% heat-inactivated fetal bovine serum (Biosera), 10mM L-Glutamine (Thermo Fisher), 1x MycoZap Plus-CL antibiotics (Lonza), 2mM non-essential amino acids (Thermo Fisher). Cells were passaged every 3-4 days using a 1:10 split ratio. To passage cells, the media was aspirated, and cells were washed with PBS. Cells were enzymatically dissociated using 1% trypsin (Gibco) for 5 minutes at 37°C. Trypsinisation was stopped by adding 2 volumes of HEK media. The cell suspension was collected into a universal tube (Gibco) and centrifuged for 200 x g for 5 minutes. The cell pellet was resuspended in 10ml of media and seeded at the appropriate ratio. Cultures were incubated at 37°C in the presence of 5% CO_2.

2.3 CRISPR/Cas9 targeting: gRNA design and synthesis

2.3.1 Guide RNA Design

Guide-RNAs (gRNAs) targeting the second exon of DGCR8, the fourth and sixth exons of HIRA and exon three and six of ZDHHC8 were designed using the online CRISPR Design tool from Massachusetts Institute of Technology (www.cripr.mit.edu). Two gRNAs were chosen per target gene based on the off-target scores generated by the online tool. A list of gRNAs generated are listed in the Table 2.1 below. The guides were designed with the appropriate overhang corresponding to the BbsI restriction enzyme for cloning in donor vector FgH1tUTG (addgene #70183) (Aubrey *et al.*, 2015).

Table 2.1 List of sgRNAs and their gene targets including overhangs.

Gene	Oligo name	Overhang	Protospacer	Overhang
DGCR8	sgRNA 1- Sense	5'-TCCCA	CTCATAGACCCGAACTGTAG	
DGCR8	sgRNA 1-Antisense	3'-T	GAGTATCTGGGCTTGACATC	CAAA-5'
DGCR8	sgRNA 2- Sense	5'-TCCCA	GGGGAGAACTTCCGGACCGC	
DGCR8	sgRNA 2-Antisense	3'-T	CCCCTCTTGAAGGCCTGGCG	CAAA-5'
HIRA	sgRNA 1- Sense	5'-TCCCA	TGTGTGCGGTGGTCAAACAG	
HIRA	sgRNA 1-Antisense	3'-T	CTGTTTGACCACCGCACACA	CAAA-5'
HIRA	sgRNA 2- Sense	5'-TCCCA	CTGGCTAGCCTCATGCAGCG	
HIRA	sgRNA 2-Antisense	3'-T	CGCTGCATGAGGCTAGCCAG	CAAA-5'
ZDHHC8	sgRNA 1- Sense	5'-TCCCA	CGTTCTTGTACAGCGGAGCC	
ZDHHC8	sgRNA 1-Antisense	3'-T	GCAAGAACATGTCGCCTCGG	CAAA-5'
ZDHHC8	sgRNA 2- Sense	5'-TCCCA	TGAAAGGGTTCACACCCCCG	
ZDHHC8	sgRNA 2-Antisense	3'-T	CGGGGGTGTGAACCCTTTCA	CAAA-5'

2.3.2 sgRNA assembly and cloning into lentiviral expression donor plasmid

In order to generate the construct, each of the single-stranded sgRNAs oligonucleotides were annealed to the complementary stand by mixing 8µl sense oligo + 8µl antisense oligo (10µM) with 2µl of 10X T4 Ligation buffer (NEB) in a 0.2ml tube, followed by melting and reannealing in a thermal cycler with the program: 96°C for 300secs, 85°C for 20secs, 75°C for 20secs, 65°C for 20secs, 55°C for 20secs, 45°C for 20secs, 35°C for 20secs, 25°C for 20 secs. Following annealing, phosphorylation of the overhangs was performed, sgRNA oligonucleotides from the previous step were mixed with 25µM ATP and 1µl of T4 PNK (NEB) and tubes were added to a thermocycler and incubated at 37°C for 60mins and then for inactivation at 65°C for 20mins.

Annealed and phosphorylated sgRNAs were ligated into the FgH1tUTG lentiviral construct. Briefly, 2μg of circular FgH1tUTG expression plasmid was digested with BsmBI (NEB) in 3.1 buffer for 5 hours at 55°C. 2μl of annealed and phosphorylated oligonucleotides were mixed with 2μl of DNA T4 Ligase (NEB), 100ng digested FgH1tUTG plasmid and incubated at 4°C overnight.

2.3.3 Bacterial Transformation

5μl of each ligation reaction was transformed into 25μl of NEB 10-beta competent *E.coli* (High efficiency) cells C3019I, the vial was flicked gently and left on ice for 30mins. Bacteria was heat shocked at 42°C for 30secs and placed on ice for 5mins. 950μl of SOC medium provided by the manufacturer was added to each vial. The vials were then incubated for 1.5hours at 37°C under constant agitation at 200rpm. Transformants were plated onto LB agar plates containing 100μg/ml Ampicillin (Sigma) and incubated overnight at 37°C. Bacterial colonies were manually picked and transferred to a bacterial tube with 2ml of LB media and 100μg/ml ampicillin (Sigma) and incubated overnight at 37°C and 200rpm.

2.3.4 Plasmidic DNA extraction: Miniprep

1.5ml of the above bacterial solution was transferred into a 2ml Eppendorf tube and centrifuged at 11000rcf for 1min. The supernatant was removed, and the pellet was resuspended in 100μl of solution A (25mM Tris HCl, Ph8, 10mM EDTA) and 3μl RNAseA were added to the sample. Samples were incubated with 100μl of lysis buffer (200mM NaOH, 1% SDS) for no more than 5 mins. 250μl of neutralisation buffer (5M potassium acetate, Ph5.5) was added and incubated for 10mins at RT. The solution was centrifuged at 12000g for 5mins. The supernatant was collected into a 1.5ml Eppendorf tube and a 1:1 volume of isopropanol was added and incubated for 15mins. The DNA was pelleted at 18000g for 15mins. The supernatant was removed, the pellet was washed with 500μl of 70% ethanol. Tubes were centrifuged at 14000g for 1min and the supernatant was removed. Tubes were air dried for 10 minutes and then DNA was resuspended in 30μl of ddH_2O.

2.3.5 Plasmidic DNA extraction: Maxiprep

250ml of bacterial solution was incubated overnight at 37°C at 200rpm. DNA extraction was performed using the HiSpeed Plasmid kit (Qiagen, 12662) following instructions provided by the manufacturer.

2.3.6 Sanger sequencing of plasmids and PCR products

Plasmids identified with the correct insert shown by enzymatic digestion were sent for sequencing to confirm the insert.

For sequencing of PCR products, before sequencing, primers were removed by enzymatic digestion. Per 5µl of PCR product, a combination of 0.5 µl of Exonuclease I (NEB) and 1µl of shrimp alkaline phosphatase (NEB) to remove any primers or dNTPs in the reaction mix. The mixture was incubated for 15 minutes at 37°C followed by an inactivation step of 80°C for 15 minutes.

Sanger sequencing was performed by LGC Genomics. Briefly 10ul at a concentration 100ng/µl of plasmidic DNA or 20ng/µl of PCR product was mixed with 4µl of 5uM of primer. The results were analysed with the BioEdit software (www.mbio.ncsu.edu/BioEdit). For identifying positive edited clones for DGCR8 CRISPR/Cas9 targeting. Each sequence derived from the targeted clone was aligned with the WT sequence and translated into the predicted protein in order to characterise the mutations generated from the targeting.

2.3.7 SNP Array

SNP array genotyping experiments were kindly performed by Alexandra Evans in Cardiff University Medical Research Centre (MRC), Hadyn Ellis Building. Genomic DNA was extracted as described in DNA extraction (2.6.1), 200 ng (50 ng/ µL) was required for genotyping. Samples were genotyped on the Infinium PsychArray-24 Kit (Illumina) or the Infinium Global Screening Array-24 (Illumina) and scanned using the iScan System (Illumina). Data were exported from Genome Studio and analysed using PennCNV (Fang and Wang, 2018). Sample level quality control was applied based on the standard

deviation of Log R ratio set at 0.3, minimum SNP number of 10 and minimum region size of 100,000bp

2.4 Lentivirus production

2.4.1 Lentiviral plasmids

All viral vectors in this thesis are third generation lentiviral vectors pCMV-VSV-G (Addgene #8454) , pRSV-Rev (Addgene #12253) and pMDLg-pRRE (Addgene #12251) obtained from Addgene (Aubrey *et al.*, 2015). Bacterial cells were grown on LB agar plates and supplemented with 100μg/ml ampicillin.

The bacteria contained in agar stabs were initially streaked onto LB agar plates with the antibiotic selection and grown overnight at 37°C. The following day, isolated colonies were picked and grown in conical tubes (Falcon) containing 3ml of LB supplemented with the selection antibiotic. Bacterial suspensions were grown overnight at 37°C with shaking at 200rpm. The following day the previously mentioned miniprep protocol was performed (2.2.6).

To confirm plasmid integrity, the extracted DNA was digested using a specific combination of restriction enzymes. Enzymatic digested was carried out at 37°C for 1 hour using the appropriate digestion buffer as required by the manufacturer. Once the plasmid identity was confirmed, the previously stated maxiprep protocol was used to gain larger quantities of DNA.

2.4.2 Production of lentivirus

Lentiviral particles were produced by transient transfection of 293T cells grown in 10cm Petri dishes using Lipofectamine^{TM} 3000 (Invitrogen, Carlbad, CA). Cells were plated at a density of 7 x 10^6 three days before transfection on 10cm Petri dishes. The morning of transfection, cell media was changed to lentiviral production media. Lentiviral medium consists of 500ml DMEM, 25ml Fetal bovine serum and 1 ml Sodium Pyruvate. Briefly, solution A, consisting of 1.5ml of OptiMEM and 41μl of Lipofectamine 3000 reagent were mixed. Solution B of 10μg of FgH1tUTG expression vector containing gRNA with viral packaging constructs pMDL (5μg), pRSV-rev (2.5μg) and pCMV-VSV-G (3μg) in 1.5ml of OptiMEM with 35μl of P3000 reagent was generated. Solution A was mixed into

solution B and incubated at RT for 10-20mins. The mixed solutions were added dropwise to the cells and incubated at 37°C, 5 CO_2 for 6 hours. The media was then changed to 12ml of preheated lentiviral packaging media.

Virus containing supernatant was collected at 24- and 48-hours post transfection, then centrifuged at 2000rpm for 10mins and passed through a 0.45-µm filter. Viral supernatants were stored at 4°C until ultracentrifugation.

2.4.3 Lentivirus concentration by ultracentrifugation

Viral supernatant was pipetted into 31.ml Thinwall Beckman tubes. Approximately 24ml of supernatant was transferred per tube. Tubes were ultra-centrifuged at 90,000g for 2.5 hours. The supernatant was then discarded. The pellet was resuspended in approximately 20µl of PBS (1000x concentrated). This was then collected in a microcentrifuge tube and briefly centrifuged for 1minute at maximum speed and then the supernatant was divided into 5µl aliquots and stored at -80°C.

2.4.4 Titration of lentivirus using flow cytometry

A lentivirus protocol was adapted using a previous method described (Barde, Salmon and Trono, 2010). Briefly, approximately 1-2 x10^5 cells per well on a 12 well plate. One day later, serial dilutions of the virus were added in the presence of 500µl of DMEM. 24 hours later the media was changed to fresh HEK media (stated previously). Cells were analysed for the percentage of GFP positive cells using flow cytometry 4 days after transduction. Before analysis, cells were dissociated in Accutase (Thermo Fisher) for 10 mins at 37°C. Cells were resuspended in 1ml DMEM and put through a FACs tube (Falcon), cells were centrifuged at 500g for 5mins at 4°C. Cells were then fixed in 1% PFA for 5mins at RT, then centrifuged 500g for 5mins at 4°C. Cells were counter stained with DAPI, centrifuged for a final time and then resuspended in cold DPBS and analysed on a BD LSR Fortessa (BD Biosciences). Lasers of the appropriate wavelength were used for exciting the samples and gates were set using the unstained samples as negative controls. The instrument was set up with the help of Mark Bishop, lab manager for the European Cancer Stem Cells Research Institute of Cardiff University.

Titre was calculated as followed:

$$Titre\left(\frac{transducing\ units}{ml}\right) = \frac{Number\ of\ target\ cells\ (count\ at\ day\ 1)\ x\ \left[\frac{\%\ of\ GFP\ positive\ cells}{100}\right]}{Volume\ of\ vector\ in\ ml}$$

Only dilutions yielding between 1% to 20% of GFP positive cells were used for titre calculations.

2.5 Lentiviral Transduction of hESCs and isolation of DGCR8 KO clonal cell lines

Lentivirus transduction was carried to deliver gRNA into iCas9 hESCs for the derivation of the DGCR8 KO line (see also chapter 4). Approximately 1x10^6 hESCs were seeded per well on a 12 well Matrigel coated plate. 24 hours after the cells had been passaged, E8 media was supplemented with 2μg/ml doxycycline (dox) which was maintained till single cell dissociation. Cells were transduced with 750000 lentiviral particles (LVP) or 1.5million LVP lentivirus with either one of two sgRNAs targeting DGCR8 in 500μl of E8 plus dox. 24 hours after transduction cells were washed with PBS.

Transduced cells were dissociated into single cells for clonal isolation. An hour before dissociation, transduced cells were treated with fresh E8 media with dox and 100μM ROCK inhibitor (Y-27632, STEMCELL Technologies) and incubated at 37C. At time of dissociation, the media was removed and kept. Cells were dissociated with Gentle Cell Dissociation Reagent (STEMCELL Technologies) for 10mins at 37C. Cells were resuspended in collected media into single sells. Cells were collected in a 15ml falcon tube and centrifuged at 980rcf for 4mins. The cells were resuspended in an appropriate volume of media and manually counted using trypan blue staining. Dissociated cells were then plated onto 6cm dishes at a density of either 20,000, 25,000 or 30,000 cells and were left to grow until small colonies emerged. Approximately 50 colonies for each gRNA were manually isolated expanded as clones.

Manually picked colonies were plated onto 48 well plates with E8 media and RevitaCell supplement at 1x concentration (Sigma). Clones were then visually screened for GFP fluorescence. Genomic DNA was extracted from GFP positive cell colonies at passage 2.

2.6 Genotyping

2.6.1 Genomic DNA extraction

Cultured cells were washed once with PBS and incubated at 37C overnight in lysis buffer (10 mM Tris-pH8.0-, 50 mM EDTA, 100 mM NaCl, 0.5% SDS) supplemented with 0.5mg/ml of Proteinase K (all components from Sigma). The next day, the lysis buffer was collected and mixed with an equal volume of isopropanol and DNA was precipitated at 15,000g for 20mins. The resulting pellet was washed with 70% ethanol, air-dried and resuspended in an appropriate volume of ddH_2O. DNA concentration was measuring using a Biospectrometre (Eppendorf).

2.6.2 PCR and DNA electrophoresis for screening

PCR was used for screening DGCR8-targeted clones. Each PCR reaction included approximately 100ng of template genomic DNA, 5µl of 5X Q5 Reaction Buffer, 0.125µl Q5 polymerase (NEB), 0.5µl of forward and reverse primers from a 10mM stock (Sigma), 0.5µl of 10mM dNTPs and ddH_2O up to 25µl. 35 amplification cycles were performed in a T100 Thermal Cycler (BioRad). Each cycle included 30 seconds of denaturation at 98°C, 30 seconds of denaturation at 68°C and 30 seconds of extension at 72°C. Finally, 15µl of the final PCR product was run on agarose gel (4% agarose).

Table 2.2 PCR primers amplifying DGCR8 targeted region

Gene	Fw sequence (5'-3')	Rev sequence (5'-3')	Product length (bp)
DGCR8	AGGGCTTGTAAAACTCTGGTCTTG	GCTCTCGGTAAAGCTCACGC	430

2.6.3 PCR Cloning

Candidate DGCR8 mutant hESC lines which were identified by PCR were selected for sequencing to verify the presence of indels and out-of-frame mutations. The targeted locus was amplified using the previously stated primers producing a 430bp product followed by cloning into the pGEM-T Easy vector (Promega) following the manufacturer's protocol. Firstly, the generated PCR product had to go through A-tailing

procedure, in which 3µl of the PCR product, 1µl of 10x Taq buffer, 2µl of 1mM dATP, 1µl of Taq polymerase and 2µl of ddH_2O were set up in a reaction tube and incubated at 70°C for 30 minutes. A ligation containing pGEM-T Easy vector (1µl), 2x rapid ligation buffer (5µl), T4 DNA ligase (1µl), A-tailing product (2µl) and 1µl of ddH_2O was set up and incubated at 4°C overnight. A volume of 2µl of the ligation reaction was used to transform competent cells (NEB® Stable Competent *E. coli* High Efficiency) with a heat-shock transformation. The competent cells were plated on LB agar plates with ampicillin 100µg/ml (Sigma) and incubated overnight at 37°C. For each transformation, the plasmid DNA was extracted from 10 minipreps as described above (2.2.6) and NotI-HF digestion was used to verify the presence of the insert. The plasmids that had the PCR product incorporated were sent for Sanger sequencing (See 2.2.8).

2.7 Transduction of hNPCs

2.7.1 Transduction

The lentivirus generated using the FgH1tUTG plasmid, without a gRNA cloned into it (referred to as Empty plasmid from now on) was used to determine transduction efficiency. As previously described, iCas9 hNPCs were passaged around day 14 onto poly-D-lysin/laminin 24 well plates in N2B27 (without retinoic acid). The next day, the medium was removed, cells were washed twice with DPBS, 250µl of prewarmed N2B27 was added to the cells, lentivirus was added to cells in three concentrations 0.5million, 1million and 1.5million lentiviral particles (LVP). The following day, the media containing lentiviral particles was removed and cells were washed with DPBS three times. 1ml of N2B27 was added cells, cells were maintained following the previously stated protocol of monolayer cortical differentiation. Cells were fixed or flow cytometry was performed at days 20, 25 and 30 to assess percentage of GFP positive cells and transduction efficiency. DNA was extracted at days 20 and 25.

For transduction with DGCR8 targeting, iCas9 cells were transduced at 24 hours after passaging onto poly-D-lysin/laminin 24 well plates. 1 million LVP was used to infect cells as this was determined to achieve sufficient transduction. 2µg/ml dox was added to the media at the point of transduction and was maintained throughout the differentiation.

DNA was extracted at days 20 and 25 from FACs sorted cells for TIDE analysis and cells were fixed at days 20, 25 and 30 for immunocytochemistry.

2.7.2 Flow cytometry to determine transduction efficiency

Flow cytometry was used to evaluate efficiency of lentiviral transduction of iCas9 hNPCs (described in Chapter 4). Before the analysis, cells were washed with DPBS and then dissociated in Accutase (Thermo Fisher) at 37°C for 10 minutes. The Accutase was then removed and cells were resuspended in 1ml of N2B27 and cells were put through a FACs tube (Falcon). Cells were centrifuged at 4°C at 500g for 5 minutes, cells were then resuspended in DPSB containing DAPI and incubated in this for 5 minutes. Cells were centrifuged again at 4°C at 500g for 5 minutes and finally resuspended in 1ml of cold DPBS and analysed on a BD LSR Fortessa cytometer (BD Biosciences). Lasers of the appropriate wavelength were selected for excitation of the samples and gates were set using the unstained and negative controls.

2.7.3 Fluorescence Activated Cell sorting (FACs) of transduced hNPCs

Transduced hNPCs cells were FACs sorted for DNA extraction from the transduced population. An hour before FACs analysis, cultures were treated with 100μM ROCK inhibitor (Y-27632, STEMCELL Technologies). The media was then removed, and cultures were washed with DPBS. Cells were then treated with 250μl of Accutase (Thermo Fisher) for 10 minutes at 37°C. To stop the Accutase reaction, 250μl of N2B27 was added and cells were gently dissociated by pipetting. Cells were collected in a 15ml Eppendorf (Eppendorf) and centrifuged at 200g for 5 minutes. The media was removed and a washing step with 0.5ml of PBS was carried out, followed by centrifugation at 200g for 5 minutes, cells were then resuspended in 1ml of DPBS and put through a FACs tube (Falcon) and immediately placed on ice. Cells were sorted on the FACS ARIA Fusion (BD Biosciences) by Mark Bishop, FACS operator of the European Cancer Stem Cell Research Institute of Cardiff University. Cells were sorted into 1.5ml Eppendorf tubes (Eppendorf) containing N2B27 and Rock Inhibitor. Between 300,000-500,000 cells were collected. After collection, cells were centrifuged at 200g for 5 minutes at 4°C and resuspended in DNA lysis buffer. DNA extraction was followed as previously stated.

2.7.4 PCR and gel electrophoresis for TIDE analysis

PCR was used to amplify the region targeted by sgRNAs in order to determine sgRNA cutting efficiency by TIDE analysis. Each PCR reaction for DGCR8 targeting, the reaction mix included approximately 100ng of genomic DNA extracted from a pool of transduced FACs sorted cells, 1x Q5 buffer (NEB), 200µM dNTPs (Sigma), 0.2µM of each forward and reverse primer (Sigma), 1.25 units of Q5 High fidelity polymerase (NEB) and ddH_2O to make the volume up to 25µl. Each PCR reaction for HIRA and ZDHHC8 targeting, the reaction mix included approximately 100ng of genomic DNA extracted from a pool of transduced FACs sorted cells, 1x Standard Buffer (NEB), 200µM dNTPs (Sigma), 0.2µM of each forward and reverse primer (Sigma), 5% DMSO and 1.25 units of Taq polymerase (NEB) ddH_2O to make the volume up to 25µl. 35 amplification cycles were performed in a Mastercycler X50s (Eppendorf). The appropriate annealing temperature was optimised for each primer pair (Table 2.3). 5µl of PCR product was run on an 1.5% agarose gel.

Table 2.3 List of primers and PCR conditions for targeted schizophrenia genes for assessment of sgRNA efficiency

Target	Forward	Reverse	Cycling Conditions	Amplicon size (bp)
DGCR8 sgRNA 1 target site	AGCGGACTTGTGCATGTTAG	TTGCTTTAGCCGTGAACCCG	98˚C - 2 minutes 98˚C – 30 seconds | 67˚C – 30 seconds | 34 x 72˚C – 30 seconds | 72˚C – 3 minutes	756
DGCR8 sgRNA 2 target site	AGCGGACTTGTGCATGTTAG	TTGCTTTAGCCGTGAACCCG	98˚C - 2 minutes 98˚C – 30 seconds | 67˚C – 30 seconds | 34 x 72˚C – 30 seconds | 72˚C – 3 minutes	756
HIRA sgRNA 1 target site	CCCTCCAGACTGTTCTGGTAAT	GTTTTCCAGGTGATGGTTAGCTC	95˚C – 30 seconds 95˚C – 30 seconds | 53˚C – 30 seconds | 34 x 68˚C – 1 minute | 68˚C – 5 minutes	766
HIRA sgRNA 2 target site	ACTGGGGTCGGTTAGATGCT	CTGGGTTTGTGTTAGCCCGTA	95˚C – 30 seconds 95˚C – 30 seconds | 53˚C – 30 seconds | 34 x 68˚C – 1 minute | 68˚C – 5 minutes	656
ZDHHC8 sgRNA 1 target site	CTGGCCGTTACTCGAAGGAG	TAGCGATAGTTTCGACGCCC	98˚C – 1 minute 98˚C – 30 seconds | 54˚C – 30 seconds | 34 x 68˚C – 1 minute | 68˚C – 5 minutes	766
ZDHHC8 sgRNA 2 target site	GAGACGAGGCTCCAACTCTG	CCTAAGGAAAGGCGGCTTCA	98˚C – 1 minute 98˚C – 30 seconds | 56˚C – 30 seconds | 34 x 68˚C – 1 minute | 68˚C – 5 minutes	601

2.7.5 TIDE analysis

To assess efficiency of each sgRNA targeting DGCR8, HIRA and ZDHHC8, tide analysis was performed. The sequence containing the targeted region was amplified by PCR and sent for Sanger sequencing (See 2.2.8). The sequence was uploaded to the online tool (https://tide.deskgen.com) to quantify sgRNA efficiency. TIDE analysis uses a decomposition algorithm to calculate the presence of insertions and deletions (INDELS) in a pool of edited sequences extracted from transduced samples compared to non-transduced samples in order to quantify the number of sequences edited by a sgRNA.

2.8 Western Blotting

Cultured cells were scratched off the culture plates in cold PBS with 2ml stereological pipette and pelleted by centrifuging at 1000g x 5mins at 4°C. For long term storage, the pellet was kept at -80°C. For protein extraction, pellets were lysed on ice using RIPA buffer (Abcam) with protease and phosphatase inhibitors (Sigma). The cell lysate was incubated on ice for 30 minutes and vortexed every 5 minutes during the incubation. Cell lysates were centrifuged for 15minutes at 12000g, 120µl of the resulting supernatant was combined with 1X Bolt® LDS Sample Buffer (ThermoFisher) and 1M DTT (Sigma), this product was boiled at 70°C for 10 minutes and then stored at -80°C. The remaining volume of the lysate was used for protein concentration quantification using the DC™ protein assay reagent (Bio-Rad) and compared against a protein standard (Bio-Rad), using the instructions supplied by the manufacturer. Absorbance at 705nm was used to determine protein quantification using a CLARIOStar Plus (BMG).

For western blotting, equal amounts of each protein sample were separated using 4-12% Bolt® Bis-Tris Plus gels (ThermoFisher) and then transferred to a PVDF membrane (0.45µm pore size, Amersham Hybond, GE Healthcare) via electroblotting. The PDVF membrane was blocked using 5% Bovine serum albumin (BSA, Sigma) in Tris Buffered Saline containing 0.1% Tween (TBS-T) for 2h at room temperature. Overnight incubation with primary antibodies at 4°C in fresh blocking solution. The membrane was washed three times in TBS-T for 10 minutes each and then incubated for 1 hour at room temperature with the HRP conjugated secondary antibodies (Abcam) in blocking solution. Before imaging, the membrane was washed 3x with TBS-T and the incubated

for 2-5 minutes with Crescendo HRP substrate (Millipore). Chemiluminescence was detected using iBright 2000 (ThermoFisher). Quantification was carried out in Fiji (Image j). All samples were normalised to GAPDH.

Table 2.4: Antibodies used for western blotting.

Target	Species	Dilution	Cat. Number	Supplier
DGCR8	Rabbit	1:1000	AB191875	Abcam
GAPDH	Mouse	1:5000	AB8245	Abcam

2.9 Immunocytochemistry

2.9.1 Immunofluorescence staining

Cultured cells were washed with DPBS and fixed with cold 3.7% PFA for 10-15 mins. After fixing, cells were washed with DPS three times for 5 mins at RT before staining. For staining, cells were first permeabilised with PBS-T (0.3% Triton-X-100 in PBS) for 10 minutes at room temperature. Blocking was then carried out in PBS-T with 2% BSA and 5% Donkey serum (Gentaur) for 20 minutes at room temperature. Cells were then incubated with primary antibodies overnight at 4°C (Table 2.5). They were then washed 3x in PBS-T for 20 minutes. Cells were then incubated with AlexaFluor® secondary antibodies (ThermoFisher) diluted in PBS-T and incubated for 2 hours at RT in the dark. Nuclei were stained with DAPI (Sigma), which was diluted 1:3000 in PBS. Cells were washed 3 times with PBS for 10 minutes, then were mounted with DAKO fluorescent mounting medium (Aligent) and then stored at 4°C.

Table 2.5: Primary antibodies used for immunocytochemistry

Target	Species	Dilution	Cat. Number	Supplier
OCT4	Goat	1:500	SC-8629	Santa Cruz
SOX2	Rabbit	1:500	Pa1-094	ThermoFisher
TRA-181	Mouse	1:200	Mab4381	Millipore
GFP	Goat	1:500	AF4240	R+D
PAX6	Mouse	1:1000	PAX6	DSHB
Nestin	Mouse	1:300	BD611659	BD
FOXG1	Rabbit	1:250	AB18259	Abcam
Ki67	Mouse	1:1000	ACK02	Leica Biosystems
N-CAD	Mouse	1:1000	18-0224	ThermoFisher
CTIP2	Rat	1:500	Ab18465	Abcam
TBR1	Rabbit	1:500	Ab31940	Abcam

2.9.2 Imaging and picture analysis

Stained cells were imaged using a Leica DM6000B inverted microscope. An average of 10 random fields were acquired per each well at a 20x magnification for quantification. Cell counting was either performed automatically, using Cell Profiler (cellprolifer.org), for DAPI and other nuclear markers. For overlap markers with GFP, cellular markers were manually counted using Fiji (Image J) software. Data analysis and representation was carried out using Excel and SPSS software (SPSS Inc. Released 2009. PASW Statistics for Windows, Version 18.0. Chicago: SPSS Inc). Unless otherwise stated, all the quantifications were collected from at least two independent experiments, with at least three biological replicates for each marker counted.

3. Identifying schizophrenia disease modifiers within and outside the 22q11.2 deletion region

3.1 Introduction

22q11.2 deletion syndrome is the strongest known molecular genetic risk factor for developing SCZ (M Karayiorgou *et al.*, 1995) with a prevalence of approximately 30% of 22q11.2DS adult patients developing SCZ (Murphy, Jones and Owen, 1999; Schneider, Debbané, Anne S Bassett, *et al.*, 2014). This is 6-7 times higher than the lifetime risk of SCZ in the general population, which is 4% (Saha *et al.*, 2005). The mechanisms underlying the genetic risk for SCZ remains largely unknown. Efforts to identify genes that confer such risk could help understand the underlying neuronal circuitry of psychiatric disorders seen in 22q11.2DS and potentially idiopathic SCZ.

Among the deleted genes in the 22q11.2 deletion, *DGCR8* is a strong SCZ candidate gene of interest. *DGCR8* encodes for a protein which is part of the microprocessor complex, which is an important component of the microRNA biogenesis pathway (Yeom *et al.*, 2006). MicroRNAs (miRNAs) are small non-coding RNAs (~22 nucleotides) involved in the regulation of gene expression (as described in Section 1.3.3.1). Each miRNA can potentially regulate the expression of numerous target genes, therefore disruptions to this system can have wide-spread effects, making DGCR8 an interesting candidate for epistatic interactions. Numerous miRNAs have been associated with SCZ, with miRNA expression dysregulation observed in multiple brain regions in post-mortem SCZ tissue (as reviewed in (Beveridge and Cairns, 2012). Furthermore, miR-185, which is within the 22q11.2 locus and also regulated by DGCR8 has been shown to cause neuro-morphological defect such as reduced dendritic complexity and contributes to brain volume (Xu *et al.*, 2013; Sellier *et al.*, 2014). miRNAs can act as a safeguarding response to reduce “excessive” gene expression responding to fluctuations in stress. They also have the potential to genomic mutations to ensure they are not expressed. Therefore, loss of DGCR8 could allow for such mutations to be expressed. The consequence of the hemizygous deletion of *DGCR8* in 22q11.2DS could lead to global downstream gene

dysregulation and therefore is an interesting candidate for gene-gene interaction effects.

Further to its role in miRNA preprocessing, *Dgcr8* conditional knockout or knockdown studies in mouse models have demonstrated that it is essential for embryonic neurogenesis. This role is independent of miRNA processing functions, further indicating a potential involvement in the SCZ risk of 22q11.2DS. *Dgcr8*$^{+/-}$ *mice* exhibit reduced neural progenitor cell proliferation and reduced adult hippocampal neurogenesis (Ouchi *et al.*, 2013). Furthermore, conditional *Dgcr8* deletion in mouse cortical progenitors in a mouse model lead to impairment of corticogenesis, due to premature differentiation of NPCs causing severe morphological defects of the cortex (Marinaro *et al.*, 2017). Equally, *Dgcr8* overexpression in mouse telencephalon, promotes expansion of NPCs, repressing neurogenesis, exhibiting a mirror phenotype (Hoffmann *et al.*, 2018). These studies indicate a key role for DGCR8 in cortical development and possibly neuropsychiatric risk.

The combinatory roles of DGCR8 in gene regulation and cortical development make it a strong candidate gene in conferring the psychiatric risk in 22q11.2DS. It is likely however, that multi-gene loss in 22q11.2DS of genes with shared expression or convergence on similar pathways would lead to a larger "hit" and collectively contribute to neuropsychiatric phenotype. Due to the size of the deletion, it is possible that haploinsuffiency of multiple genes, which converge on similar pathways or neuronal circuits contribute together to SCZ risk. There is accumulating evidence indicating that genes with similar expression patterns tend to be clustered together in the genome (Caron *et al.*, 2001; Lee and Sonnhammer, 2003; Singer *et al.*, 2005). Furthermore, protein-protein interactions demonstrate significant clustering of functionally related genes in the genome (Yi, Sze and Thon, 2007; Al-Shahrour *et al.*, 2010). Transcriptional data has been used to identify network-based approaches to human disease and inform functional disease mechanisms in order to prioritize candidate disease genes (Yue *et al.*, 2016; Gerring, Gamazon and Derks, 2019; Ma *et al.*, 2019). Therefore, mutations that affect multiple genes close together in the genome may lead to compounding deleterious effects, such as the case of copy number variants (CNVs, deletions of duplications > 1KB). Andrews and colleagues have demonstrated that pathogenic CNVs affect functional clusters of genes to a greater extent than benign CNVs and clusters of

functionally related genes in the human genome contribute to CNV-mediated developmental disorders (Andrews *et al.*, 2015). Therefore, analysis of gene expression could be used to potentially understand shared mechanisms underlying the disease pathology of 22q11.2DS.

As neuropsychiatric disorders have a complex genetic etiology, a multi-gene approach is required to elucidate the underlying pathology behind the SCZ risk in 22q11.2DS. Gene expression analysis in orthologous mouse genes from within the 1.5Mb deletion in the developing and adult brain, demonstrate that most 22q11.2 genes are brain expressed from early development through to maturity of the neuronal circuitry (Maynard *et al.*, 2003). This suggests that increased SCZ risk could be due to the distinct expression patterns of multiple genes. Although, it is noteworthy that approximately 84% of all genes are expressed in at least one brain structure, indicating the complexity of the brain (Negi and Guda, 2017). To date, the expression patterns of genes spanned by deletions at the 22q11.2 region in the developing human brain have not yet been explored. An aim of this thesis is to use gene brain expression data to help identify common expression pathways shared to identify candidate genes to the increased SCZ risk seen in 22q11.2DS.

As previously stated, SCZ is observed in approximately 30% of adults with 22q11.2DS, therefore there is significant phenotypic variability in patients. This variability could be due to the presence of other genetic risk factors located outside of the 22q11.2 locus, which potentially act as modifiers to disease in the presence of the deletion. This genetic variation could influence phenotypic variability, such as increasing the risk for SCZ. The presence of disease modifiers have been investigated in many Mendelian disorders (Cutting, 2010; Holmans, Massey and Jones, 2017). Modifiers for congenital heart symptoms observed in 22q11.2DS have been investigated, identifying a duplication CNV spanning the gene *SLC2A3* (Mlynarski *et al.*, 2015). This study lead to further identification of more CNVs associated with the congenital heart defects observed in 22q11.2DS (León *et al.*, 2017). This approach has not yet been used to identify specific modifiers of SCZ in 22q11.2DS, however the SCZ phenotype in 22q11.2DS has been reported to be associated with the presence of additional rare CNVs outside of the 22q11.2 deletion region (Bassett *et al.*, 2017).

Large scale sequencing approaches enable investigation of genetic differences in case-control studies, such as genome wide association studies (GWAS) and transcriptome wide association studies (TWAS) (Gusev *et al.*, 2018). These methods have been used to successfully identify common genetic variants in SCZ and other psychiatric disorders (Bush and Moore, 2012; Pardiñas *et al.*, 2018), but these methods have not yet been applied to investigating variants within 22q11.2DS. Previous work identified that 22q11.2 deletion carriers with SCZ have on average greater polygenic risk score for SCZ than 22q11.2 deletion carriers without psychosis, indicating additional genetic variance outside of the 22q11.2 region is relevant to the increased risk to SCZ (Cleynen *et al.*, 2020).

To this end, the aims of this chapter are to firstly select additional SCZ candidate genes in addition to *DGCR8* from within the 22q11.2 deletion region. This will be based on analysis of gene co-expression of the 22q11.2 genes during the developing fetal brain (O'Brien *et al.*, 2018). This approach can help identify potential modules of co-expressed genes from within the 22q11.2 locus and possibly genes that could be compromised during brain development. Following co-expression analysis, informative genetic studies on effects of loss of function mutations and predicted effects of haploinsuffiency will be utilised to further inform candidate gene selection. Candidate genes can be selected objectively based on predicted deleterious effects of hemizygous deletion, reflecting the genetic architecture underlying 22q11.2DS. This approach objectively prioritizes the 22q11.2 deletion genes that are most likely to contribute to disease pathology, with would then be taken forward to investigate in cell models later in this thesis.

The second aim is to identify potential modifiers of disease outside of the 22q11.2 deletion region. To achieve this, GWAS study was performed using whole-genome sequence data from 22q11.2 deletion carriers with SCZ compared to carriers without psychosis to identify genetic differences, which was collected by and previously analysed by the IBBC (Cleynen *et al.*, 2020). This was followed by transcriptome wide association study using the above generated GWAS summary statistics and expression weights from the prefrontal dorsolateral cortex, blood and fetal brain expression, in

order to identify cis-heritable changes in gene expression based in 22q11.2DS carriers with and without SCZ.

3.2 Results

3.2.1 Gene Co-expression of 22q11.2 Deletion Genes in Fetal Brain RNA Sequencing (in collaboration with Heath O'Brien)

To identify additional SCZ candidate genes within the 22q11.2 deletion region that are potentially functionally related with *DGCR8*, RNA sequencing data of fetal brain tissue (12-19PCW) provided by the Human Developmental biology Resource (O'Brien *et al.*, 2018) was used to investigate gene expression within the 22q11.2 deletion. Pearson's correlation values were calculated for all the genes within the 22q11.2 deletion region with measured gene expression in the fetal brain RNA sequencing data (all values were calculated by Heath O'Brien). Using the correlation and significance values generated by Heath, I took these values forward to construct plots of co-expression of the genes within the 22q11.2 region (Figure 3.1). The corrplot package hierarchically clusters genes based on the Pearson's correlation value, genes with higher Pearson's correlation values group together, above a threshold of 0.4. This analysis revealed ten distinct gene modules of clustered genes, with varying observational trends.

The genes within each module are summarised in Table 3.1, with the second and third module listed as 2a and 2b, as there appears to be a level of co-expression between both groups. Modules 2a and 2b contain the largest number of genes, the majority of these genes (16/21) located between LCR A and LCR B (low copy repeat region) and varying biological functions. Out of the seven genes in module 4, six of the genes are within low copy repeat region C and D within the 22q11.2 deletion, with *SCARF2* being the exception. In module 5, 6 out of 9 genes are located between LCR A-B, but *SNAP29* and *CRKL* are also in this module, which genetically reside next to each other in between LCR C-D. Interestingly, there were also incidences of significant negatively co-expressed genes, such as such as *SNAP29* and *CDC45* (module 5) and *DGCR5* and *LRRC74B* (module 6). This could indicate that there are patterns of genes that are not expressed in similar trajectories that could act in opposing periods.

As *DGCR8* is the candidate gene of interest, the genes most highly correlated with *DGCR8* are *LZTR1* (r value = 0.749, p = 8.1×10^{-23}) and *ZDHHC8* (r value = 0.752, p = 4.68×10^{-23}). *LZTR1* and *ZDHHC8* were also found to be highly co-expressed with each other (r value = 0.832, p = 4.98×10^{-23}). *DGCR8* is located in module 2.a, therefore the genes in modules 2.a and 2.b were to be considered as potential candidate genes.

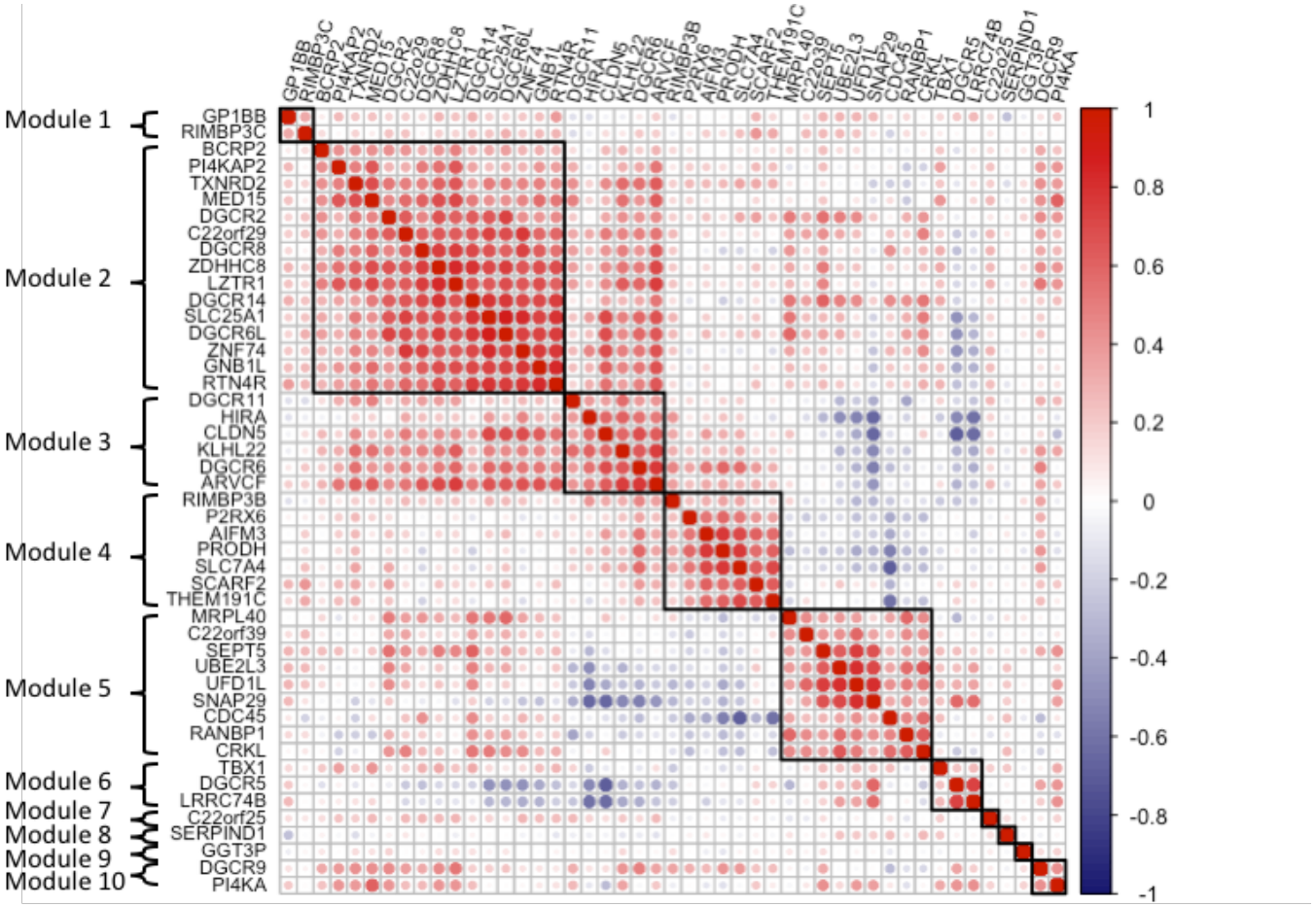


Figure 3.1 Correlation of gene Expression of 22q11.2 Deletion Region Genes in Fetal Brain RNA-Sequencing Data:
Correlogram heat map of gene expression data from RNA sequencing of fetal brain tissue (12-19PCW) provided by the Human Developmental Biology Resource and RNA sequencing carried out by the Bray lab of genes found within the 22q11.2 deletion and plotted using corrplot (Wei and Simko 2017). Pearson's correlation was calculated to show the co-expression pattern of the genes in the heat map. Red and blue represent positive and negative correlation respectively and the colour intensity represents the magnitude of the correlation. Red clusters identify groups of co-expressed genes.

Table 3.1 22q11.2 Gene modules generated from fetal brain expression

List of genes in modules generated from gene correlation analysis of RNA Sequencing data from fetal brain tissue of genes within the 22q11.2 deletion region.

Module 1	Module 2a	Module 2b	Module 3	Module 4	Module 5	Module 6	Module 7	Module 8	Module 9
GP1BB	BCRP2	DGCR11	RIMBP3B	MRPL40	TBX1	C22orf25	SERPIND1	GGT3P	DGCR9
RIMBP3C	PI4KAP2	HIRA	P2RX6	C22orf39	DGCR5				PI4KA
	TXNRD2	CLDN5	AIFM3	SEPT5	LRRC74B				
	MED15	KLH22	PRODH	UBE2L3					
	DGCR2	DGCR6	SLC7A4	UFD1L					
	C22orf29	ARVCF	SCARF2	SNAP29					
	DGCR8		THEM191C	CDC45					
	ZDHHC8			RANBP1					
	LZTR1			CRKL					
	DGCR14								
	SLC25A1								
	DGCR6L								
	ZNF74								
	GNB1L								
	RTN4R								

3.2.2 Prioritising Schizophrenia Candidate Genes Through Literature Analysis

Following identification of 22q11.2 genes with shared co-expression using fetal brain RNA sequencing data, the next step was to narrow down candidate genes alongside *DGCR8.* The following approach used to aid candidate gene selection was based upon literature review of available data generated from large scale exome sequencing projects. In these projects, large exome sequencing was carried out on healthy individuals (> 60K samples) to examine the presence of mutations, in order to predict genes that are intolerant to variation (Lek *et al.*, 2016; Cassa *et al.*, 2017). In these datasets, genes have been ranked according to their expected influence of mutation in two ways. One approach determined intolerance to homozygous loss of function and the other estimated a score of likelihood to be deleterious affected by haploinsufficiency (Shet score) Loss of function intolerance genes have the highest pathogenic potential and therefore act as a good proxy for disease susceptibility. Analysing the predicted effects of loss of function intolerance and effects of haploinsufficiency will identify which genes are most likely to be functionally affected by the deletion. This allowed me to objectively assess the potential effect of 22q11.2 deletion on a gene-by-gene basis. The candidate gene selection procedure is described in Figure 3.2.

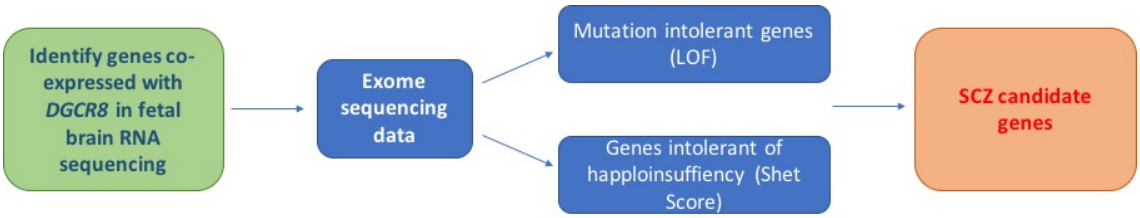


Figure 3.2: Flow diagram demonstrating schizophrenia candidate gene selection process. *Modules of co-expressed genes from the 22q11.2 region were identified using fetal brain RNA sequencing data, narrowing down candidates to those co-expressed to DGCR8. This is then supplemented with identification of genes that are loss of function intolerant or intolerant to happloinsuffiency from exome sequencing projects.*

Cassa and colleagues estimated the overall distribution of gene-based fitness effects and individual gene fitness cost in heterozygotes. It was calculated given gene-specific estimates of the *de novo* mutation rate, the observed number of protein truncating variant (PTV) alleles throughout each gene, and the number of chromosomes samples, they estimated the genome-wide distribution of selective effects for heterozygous PTVs, which they called a Shet score. Genes with a Shet score above 0.04 are indicated as potentially disease causing as recommended by the authors (Cassa *et al.*, 2017). These genes are more likely to be affected by haploinsufficiency and have a pathogenic effect. In their data, I found 9 genes from within the 22q11.2 region that have scores above this threshold and are listed in Table 3.2, with *DGCR8* having the highest score. *DGCR8* has a score within the top 200 genes from the genome. Of the genes with score above the happloinsuffiency threshold, 6 of them were also found to be loss-of-function (LOF) intolerant (Table 3.2). Lek and colleagues, used the ExAC dataset and were able to infer gene-level constraint against protein truncating variation (PTV). They developed an algorithm using the observed and expected PTV counts within each gene to separate genes into three categories: null (observed = expected), recessive (observed $\leq$50% of expected) and haploinsufficient (observed <10% of expected). Using this metric, they calculated the probability of being LOF intolerant, called pLI, with LoF intolerance having a metric $\geq$ 0.9. The Shet scores and LOF information is available for the public to acquire about all listed genes, making it an appropriate resource for prediction of disease relevant genes.

Combining this data with that of the results of Section 3.2.1, this highlights the genes *ZDHHC8* and *HIRA* as having the highest Shet and LoF intolerance scores and

potentially sharing a developmental gene expression pattern with *DGCR8* (as they also occur in modules 2a and 2b from the fetal brain RNA sequencing data). Together, this analysis has allowed for objective selection for potential SCZ candidate genes within the 22q11.2 deletion region. These candidate genes will be taken forward for investigation in human embryonic stem and neuroprogenitor cells which is later discussed in this thesis.

Table 3.2 Shet Scores and LOF status of 22q11.2 Deletion Genes:
Genes within the 22q11.2 deletion with 'Shet' Score > 0.04 to demonstrate effect of haploinsufficiency determined in Cassa et al, 2017 and loss of function intolerance status calculated in Lek et al, 2016.

Gene Symbol	Shet score	LOF intolerant
DGCR8	0.3913	LOF
HIRA	0.2728	LOF
UFD1L	0.2452	LOF
TBX1	0.2311	LOF
ZDHHC8	0.1133	LOF
CRKL	0.0724	
SEPT5	0.0646	LOF
SLC25A1	0.0575	
SNAP29	0.0510	

3.2.3 Genome Wide Association Study in 22q11.2 Deletion Carriers with and Without Schizophrenia using the IBBC (International 22q11.2 Deletion Syndrome Brain Behaviour Consortium) Sequencing Data

To investigate potential SCZ disease modifiers outside of the 22q11.2 deletion region, the 22q11.2 IBBC have generated whole genome sequenced data from 22q11.2DS patients with and without SCZ. GWAS has been performed previously by the IBBC comparing 22q11.2 deletion carriers with and without psychosis to investigate polygenic risk score for SCZ (Cleynen *et al.*, 2020). I used this dataset, with a slight difference in number of individuals and performed GWAS again. Among the cohort I used, there are 214 cases and 221 controls and this dataset had been through the QC steps by the IBBC, so I went on to perform GWAS analysis with the clean dataset (Cleynen *et al.*, 2020).

The GWAS was repeated and the results from the consortium were replicated, showing that no SNPs survive genome-wide correction, as shown by the Manhattan plot (Figure 3.3). The quantile-quantile plot (QQ plot) is shown in Figure 3.4. Although, it was likely that the GWAS approach would not identify any significant hits, as was found in the Cleynen *et al* study, it was important to demonstrate that this study is still underpowered for GWAS analysis, as demonstrated by the QQ plot. However, we know there are genetic differences between 22q11.2 DS carriers with and without SCZ from the increased polygenic risk score, so there is signal from this dataset. Therefore, other forms of investigation such as TWAS, which involve less multiple testing could potentially help elucidate the underlying genetic differences.

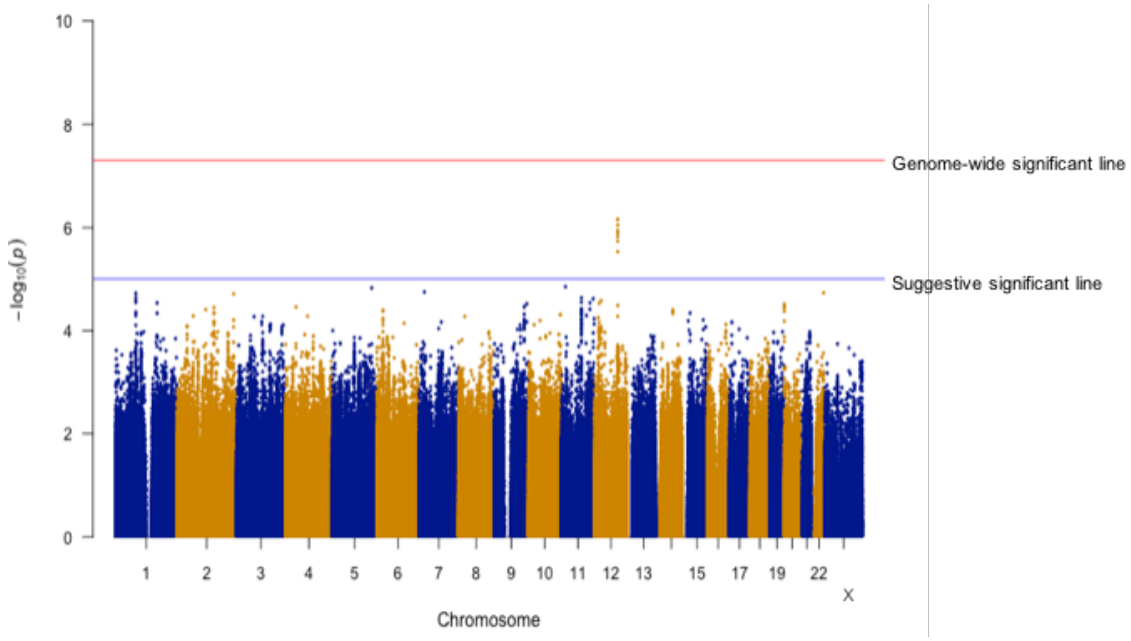


Figure 3.3: Manhattan Plot of 22q11.2DS Patients With and Without Schizophrenia GWAS Associations

Associations shown from the GWAS analysis of IBBC cohort of 22q11.2 deletion carriers with and without schizophrenia (214 cases, 221 controls). No genome-wide significant loci were identified.

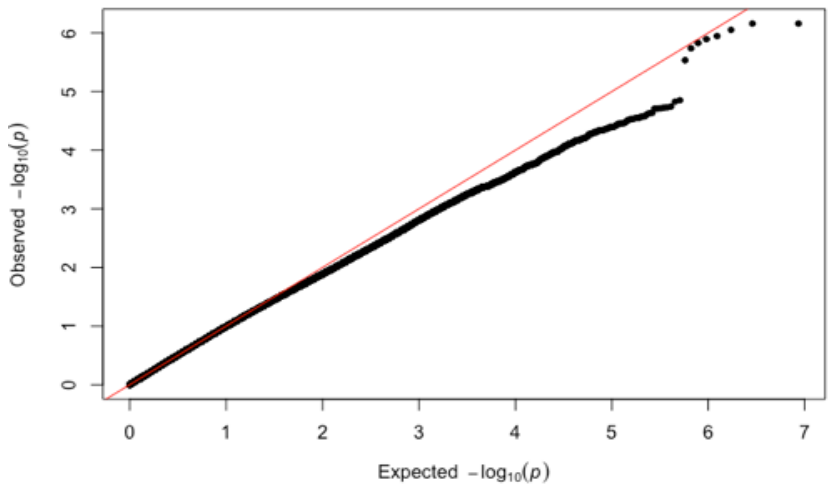


Figure 3.4: Quantile-Quantile Plot of Associations From GWAS Analysis of 22q11.2 Deletion Carriers With and Without Schizophrenia
Quantile-Quantile plot showing relationship between observed against expected p values from GWAS analysis, which indicates that the study is underpowered.

3.2.4 Transcriptome Wide Association Study in 22q11.2 Deletion Carriers with and Without Schizophrenia with Dorsolateral Prefrontal Cortex Expression Weights

TWAS is an approach to identify genes where their expression is significantly associated with complex traits. The 22q11.2DS with and without SCZ cohort is a small dataset and TWAS is an approach that could potentially increase the power to identify associations. Two factors can increase TWAS power over GWAS, firstly TWAS carries a reduced testing burden compared to the GWAS, due to testing substantially fewer genes compared with number of SNPs tested in a GWAS. Secondly, it could be expected to observe increased association signal when the expression of a risk gene is regulated by multiple local SNPs (Gusev *et al.*, 2016). It is an approach to potentially identify modifier genes contributing to the SCZ risk observed in 22q11.2DS. This method could identify a potential association due to changes in gene expression with genes that may interact with the 22q11.2 region which further compound SCZ risk.

I performed TWAS was performed using the programme FUSION following the protocols established by the Gusev Lab (Gusev *et al.*, 2016). The first TWAS was performed using the above stated GWAS summary statistics (Section 3.2.3) and eQTL expression weights from the dorsolateral prefrontal cortex (DLPFC) generated by the

Common Mind Consortium (Fromer *et al.*, 2016). Of the 5420 *cis*-heritable genes measured in the DLPFC, a TWAS test statistic was generated for 5376 genes from the 22q11.2DS cohort association summary statistics. The test statistics were corrected for multiple testing using Bonferroni correction, no genes survived the correction, with the blue line representing suggestive genome wide significance in Figure 3.5A. The most significant gene is *STEAP2,* but it is nominally significant (p = 0.175 after correction). The 10 most significant results are listed in Table 3.3.

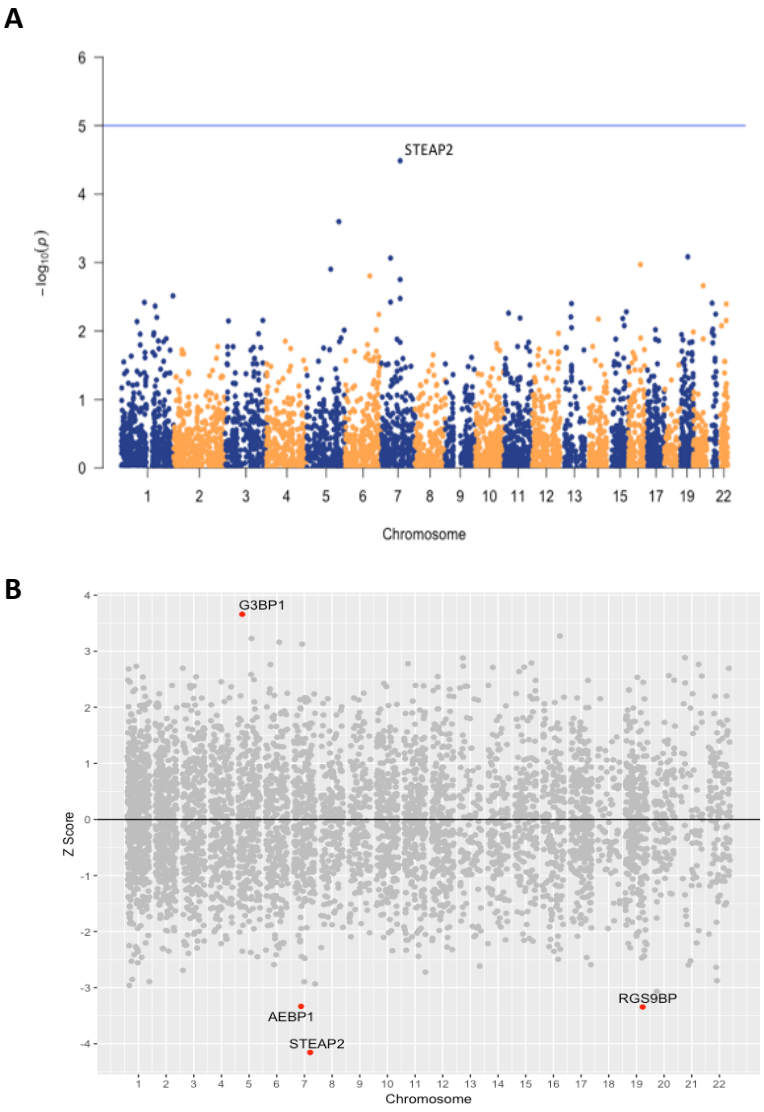


Figure 3.5: Results of TWAS Analysis of 22q11.2 Deletion carriers with and without SCZ using CMC expression data from the Prefrontal Dorsolateral Cortex using FUSION (Gusev et al, 2016). (A) Manhattan Plot of TWAS associations 22q11.2DS Patients With and Without Schizophrenia *Each point represents the expression of a gene measured in the Dorsolateral Prefrontal Cortex (DLPFC) on the X-axis and -log(10) of the p value on the y-axis. The most significant gene, STEAP2 is highlighted. Genome wide significance line is in blue but there are no genome wide significant associations*. **(B) Mirror Manhattan plot of transcriptome-wide association results.** *Each point represents a gene with the physical genomic position on the x-axis and the Z-score on the y-axis. The highlighted red points are the top 4 most significant genes.*

Table 3.3 Table of association values from the 10 most significant genes from the TWAS analysis. *Showing gene name, Z-Score which is an estimate of genetic covariance between 22q11.2DS carriers with SCZ and gene expression. P values are shown before correction and after Bonferroni correction.*

Gene ID	CHR	Z Score	P Value	Bonferroni
STEAP2	7	-4.15479	3.26E-05	0.1752
G3BP1	5	3.65903	0.000253	1
RGS9BP	19	-3.34491	0.000823	1
AEBP1	7	-3.33325	0.000858	1
NUP93	16	3.2712	0.00107	1
SRP19	5	3.22681	0.001252	1
FAM162B	6	3.16122	0.00157	1
STEAP1	7	3.12707	0.00177	1
DDX27	20	-3.0652	0.00218	1
EFCAB2	1	-2.96015	0.00307	1

Although no significant genes were identified in the TWAS with expression from the DLPFC, to determine the validity of the results and the method used I repeated the TWAS using two independent approaches. The GTEX generated expression weights from the cortex was used as an independent gene expression panel (Aguet *et al.*, 2017) and an alternative TWAS programme PrediXcan was used (Gamazon *et al.*, 2015). No genes were found to survive correction using either PrediXcan or FUSION with the GTEX generated expression weights, with the most significant results shown before and after correction in Figure 3.6.

There are 493 measured genes in common across all three methods and the Z-scores of any genes. Figure 3.7 shows the Z-score plot comparisons between each method of overlapping genes across all three methods. PrediXcan and FUSION with DLPFC weights show the highest correlation of Z-scores. The Z-scores of any genes found to be in the 15 most significant genes across each method were compared to see if gene expression change was in the same direction. There were four overlapping genes found between FUSION and PrediXcan with the Common Mind DLPFC expression weights (Figure 3.7). *STEAP2* was found to be the most significant TWAS association result in both these methods. The Z-scores were in same direction demonstrating concordance of analysis shown in Figure 3.7. There was only one overlapping gene using the independent GTEX cortex gene expression panel, *CKAP2*, the Z-Score was in

the same direction of change, but this is a limitation to this approach as there same genes might not be available across different expression panels making it difficult to compare the two methods.

Although there is a level of correlation between each other methods of TWAS and brain expression weights used, there is notable variation, which is more pronounced in the comparison between FUSION and PrediXcan. Most of the variation is likely accounted for by the lack of overlapping genes tested in both methods. Another reason for such variability is due to the differences in criteria used to determine how the gene prediction models were built. Different statistical models are used to train the gene expression prediction models for FUSION and PrediXcan. Furthermore, how SNPs that are present in the expression prediction models but are absent from the GWAS data are processed differently. In PrediXcan, these SNPs would not be included for expression prediction but in FUSION, the GWAS summary statistics are imputed using information from nearby SNPs present in the GWAS dataset. These imputed GWAS summary statistics can then be used to generate gene expression prediction (Fryett *et al.*, 2018).

A

Gene ID	Z Score	p value	Bonferroni
STEAP2	-3.79159	0.0001497	1
NA	-3.73535	0.0001875	1
PIRT	3.61906	0.0002957	1
SLC22A5	3.55503	0.0003779	1
STEAP1	3.45017	0.0005602	1
NA	-3.44030	0.0005811	1
SHANK2	3.39724	0.0006807	1
SP140	3.28613	0.0010157	1
RGS9BP	-3.26571	0.0010919	1
NPC2	3.22349	0.0012664	1
OXSM	-3.21432	0.0013075	1
RAD50	-3.15969	0.0015794	1
CKAP2	3.13082	0.0017432	1
PTCHD4	3.12275	0.0017917	1

B

Gene ID	Z Score	p value	Bonferroni
UCHL3	-3.1522	0.00162	1
AK8	-2.8336	0.0046	1
RP11-17E13.2	2.8152	0.00487	1
RP11-334C17.5	-2.7256	0.00642	1
FGFR4	2.6965	0.00701	1
CDC42BPA	-2.6921	0.0071	1
EFCAB2	-2.6900	0.00714	1
CENPJ	2.5943	0.00948	1
TM2D3	2.5763	0.00999	1
PLCH2	-2.5686	0.01021	1
CCDC66	-2.5453	0.0109	1
AC005363.9	-2.5365	0.0112	1
CKAP2	2.5152	0.0119	1
CCDC171	-2.5131	0.012	1

C

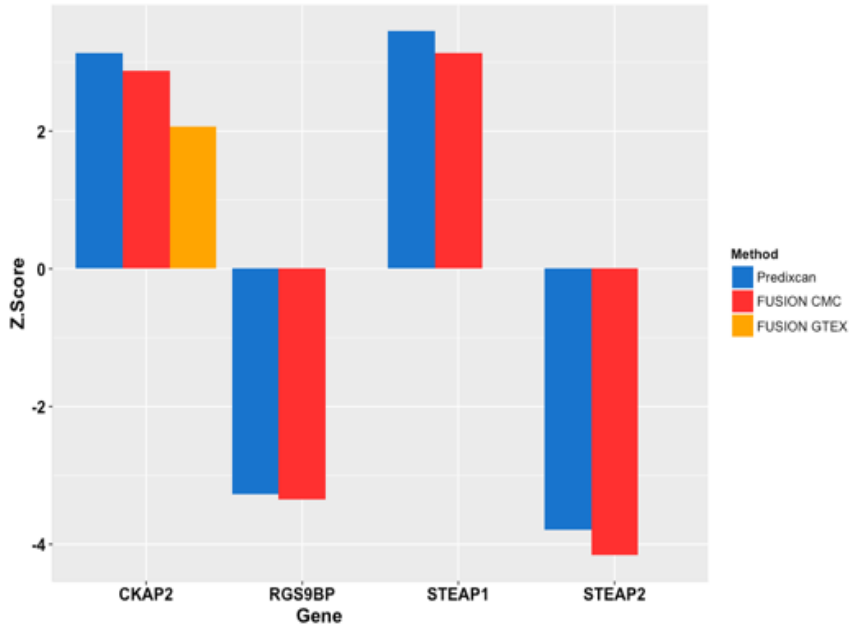


Figure 3.6: Comparison of Z Scores of transcriptome wide association results of the 15 most significant genes generated from FUSION with Common Mind Consortium (CMC) DLPFC expression weights, FUSION with GTEX generated cortex expression weights and PrediXcan with Common Mind DLPFC expression weights. (A) Table showing 15 most significant results generated from PrediXcan and Common Mind DLPFC expression weights. *Gene ID, Z-score and p value before and after Bonferroni correction are shown, no genes survive correction.* **(B) Table showing 15 most significant results generated from FUSION and Cortex expression weights generated by Gtex**. *Gene ID, Z-score and p value before and after Bonferroni correction are shown, no genes survive correction.* **(C) Plot to show direction of Z scores of transcriptome wide association results of overlapping genes from three independent methods.** *Z-scores between gene expression in the different methods of TWAS and 22q11.2DS patients with schizophrenia are plotted on the y-axis. The Z-scores of genes which were found in the top 15 most significant genes across all methods were compared (y-axis). There were four overlapping genes between FUSION with CMC (showing in red) and PrediXcan with CMC (shown in blue) expression weights. There was only one overlapping genes between associations generated from FUSION with GTEX expression weights (shown in yellow) and FUSION with CMC expression weights. All overlapping Z-scores were in concordant direction of expression.*

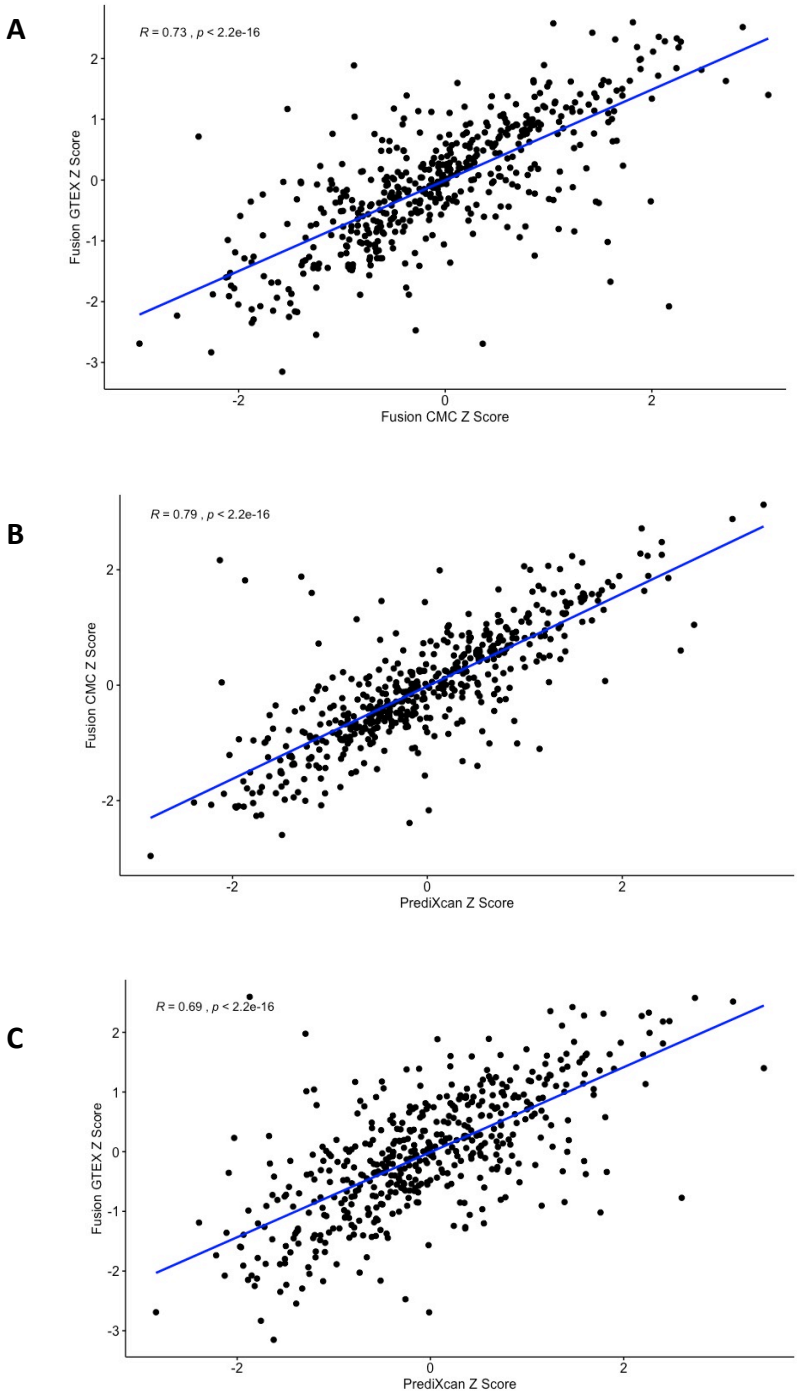


Figure 3.7: Correlation matrix plot of Z-scores from overlapping genes generated from multiple transcriptome wide association studies

(A) *Z-scores from FUSION with Common Mind Consortium (CMC) DLPFC expression weights plotted against FUSION with GTEX generated cortex expression weights.****(B)*** *Z-scores from FUSION with (CMC) DLPFC expression weights plotted against PrediXcan with CMC DLPFC expression weights.****(C)*** *Z-scores from PrediXcan with (CMC) DLPFC expression weights plotted against FUSION with GTEX generated cortex expression weights.*

3.2.5 Transcriptome Wide Association Study in 22q11.2 Deletion Carriers with and Without Schizophrenia using expression weights from Whole Blood and Fetal Whole Brain

To further investigate potential genetic differences in the 22q11.2DS with and without SCZ cohort, I carried out TWAS following the FUSION protocol using a potentially more powerful gene expression panel from the whole blood, generated from 1264 samples (Raitakari *et al.*, 2008; Nuotio *et al.*, 2014). This expression panel is publicly available. TWAS statistics were generated for 4613 out of 4701 genes measured in the whole blood expression weight panel. After Bonferroni correction for multiple testing, no genes survived correction (Figure 3.8). The most significant gene was *ARAP3* (p value = 0.000515 before correction, after correction p value = 1). TWAS was then carried out using FUSION with expression weights generated from the fetal brain expression (O'Brien *et al.*, 2018) generated from 120 samples from fetal brains from the second trimester. TWAS statistics were generated for 1319 out of 1329 genes with measurable expression in the fetal whole brain tissue, but again no genes survived correction (Figure 3.8), with most significant gene being *PTPN21* (p value before correction = 0.00112, after Bonferroni correction = 0.9688)

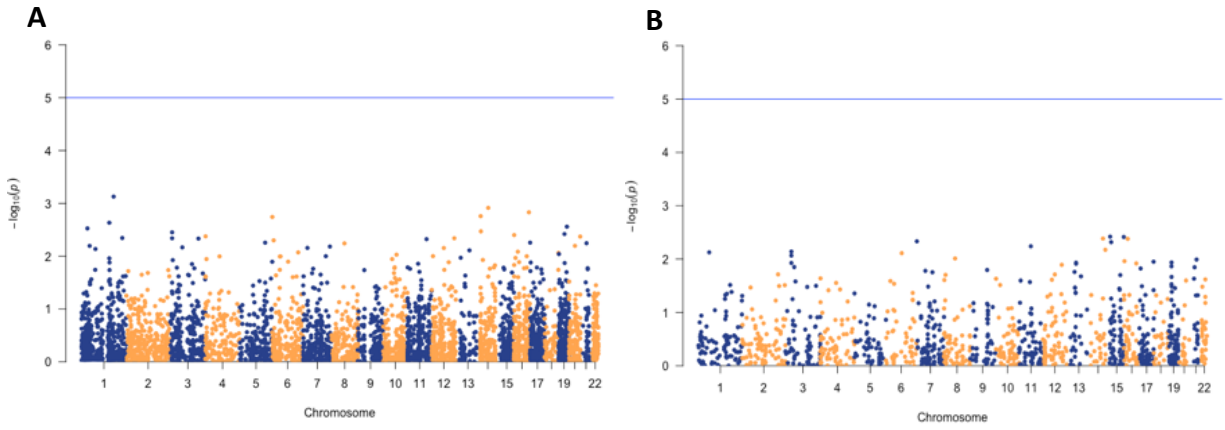


Figure 3.8: Manhattan Plots of transcriptome wide association results of 22q11.2DS with and without SCZ using FUSION with (A) whole blood expression weights or (B) fetal whole brain expression weights.

3.3 Discussion

This chapter describes two approaches to genetically identify potential genes that contribute to the known schizophrenia risk in 22q11.2 deletion syndrome. The first approach was to select other potential SCZ candidate genes, in addition to *DGCR8,* from within the 22q11.2 deletion region. The selected genes would then be taken forward to be studied *in vitro* in human embryonic stem cells and through cortical differentiation to understand their involvement in human brain development, which is later described in this thesis. The second approach was to investigate genetic modifiers of disease outside of the 22q11.2 deletion, through genetic comparisons of deletion carriers with and without SCZ using Genome Wide Association Study (GWAS) and Transcriptome Wide Association Study (TWAS).

DGCR8 was selected as the initial SCZ candidate gene from the beginning of this work, due to its roles in gene expression regulation and brain development. There is accumulating evidence that multiple gene loss of genes from 22q11.2 region that act on similar pathways or shared trajectories could be underlying SCZ risk. Subsequently, I wanted to select other SCZ candidate genes alongside the principal candidate gene DGCR8 based upon shared gene expression patterns. The expression profile of the 22q11.2 deletion region genes were first examined in fetal whole brain RNA sequencing data (O'Brien *et al.*, 2018). There is much evidence to indicate SCZ has an early neurodevelopmental component (Weinberger, 1987), with enrichment of SCZ risk loci found in methylation QTLs in the developing brain (Hannon *et al.*, 2015), hence looking at the expression profile of the 22q11.2 deletion genes during development. This process helped identify distinct modules of genes with correlated expression, indicating converging expression between multiple 22q11.2 genes in the developing brain.

Modules 2a and 2b make up the largest module and contains the majority of the genes between LCR A and B. Of the total 33 genes in LCR A-B, 25 of these genes have detected expression in the fetal brain and 16 of these are within this module. This section is the minimum critical deletion region to display the full spectrum of 22q11.2 phenotypes, suggesting phenotypes could be due to haploinsufficiency of the genes within the region. Motahari and colleagues categorized genes within the 1.5Mb and 3Mb typical deletion based upon all biological functions. The biological functions most represented

by the genes in modules 2a and 2b are transcription factors, RNA/ miRNA regulation and mitochondrial/metabolism (Motahari *et al.*, 2019). Therefore, this module contains key regulatory genes across varying biological processes and many of the genes in the critical 22q11.2 deletion region. The additional convergence of gene expression may indicate an important shared molecular signature of these genes. The presence of a large shared gene expression network, spanning multiple biological mechanisms at this important developmental window could indicate this as a key module of genes, which the diminished dosage could be contributing to the 22q11.2 phenotype.

There are further observed patterns of genomic location and biological function in the other modules. There is clear correlation of genes between LCR C-D and shown in module 4, with 6 out of the 7 genes within this region, further demonstrating that genes with similar expression levels tend to be clustered (Michalak, 2008). Module 5 contains an enrichment of the 22q11.2 genes involved in protein trafficking, *SEPT5, SNAP29* and *RANBP1*. Furthermore, independently *SEPT5* and *SNAP29* networks have been enriched in synaptic biological processes, indicating potentially why they also have correlated gene expression (Forsyth *et al.*, 2019).

With the gene expression modules identified in the 22q11.2 region, the next step for selecting SCZ candidate genes was to use literature review to investigate predicted loss of function status and effects of haploinsuffiency. Loss of function status is particularly relevant as mutation-intolerant genes have been found to be enriched in common SCZ alleles (Pardiñas *et al.*, 2018). Further to this, haploinsufficient genes exhibit higher levels of expression during early development and are usually found to be implicated in human dominant disease (Huang *et al.*, 2010). This makes this approach effective for selecting candidate genes. Nine genes were found to have a predicted Shet score higher than 0.04 as indicated in the paper as more likely to be affected by haploinsuffiency and of those, six of these genes were found to be loss of function intolerant. Of those genes, only two were found to have correlated expression to *DGCR8* leading to the independent selection of *HIRA* and *ZDHHC8* as potential SCZ candidate genes with shared expression with *DGCR8. HIRA* is a histone chaperone and has been identified as a epigenetic regulator (Nashun *et al.*, 2015; Ricketts and Marmorstein, 2017). Reduction of HIRA has been found to reduce neuroprogenitor cell proliferation, fundamentally

causing premature neuronal differentiation (Li and Jiao, 2017). Further to this, functional genomic approaches have identified *HIRA* and *DGCR8* as candidate drivers of disease-relevant alterations in gene regulation in SCZ (Forsyth *et al.*, 2019). *ZDHHC8* is a palmitoyl acyltransferase protein, the process of palmitoylation has been suggested to have a key role in pre- and post- synaptic proteins and proteins in neuronal development (El-Husseini and Bredt, 2002). ZDHHC8 has been shown to contribute to alterations in dendritic spine density in 22q11.2 deletion mouse models (Mukai *et al.*, 2008, 2015; Moutin *et al.*, 2017). The chosen candidate genes were objectively selected based on their shared gene expression profile and mutation intolerance.

Candidate gene selection approach was based solely upon gene expression correlation to *DGCR8* and mutation intolerance, however there are some limitations to this approach. Other factors could have been included to aid the selection process, such as looking at co-expression in adult brain. Information from animal models could also have been considered when factoring in shared biological roles between the genes in the 22q11.2 deletion. However, due to time constraints and interest in pursuing later described cell work, the candidate gene selection process seemed appropriate to move forward with.

Further work has since been performed to improve identification of loss-of-function intolerant genes using larger dataset of 125,748 whole exomes and 15,708 whole genome sequences (Karczewski *et al.*, 2019). This dataset became recently available and the work presented in this thesis was conducted using the available data at the time. A metric has been used to calculate probability of loss-of-function intolerance (pLI) with a threshold set of >0.9. Using this updated dataset, *ZDHHC8* (pLI =0.85) and *TBX1* (pLI = 0.84) are no longer above the predicted threshold to be loss-of-function intolerant but the remaining reported genes including *DGCR8, HIRA, UFD1L* and *SEPT5* remain LoF intolerant.

Genome wide association study and transcriptome wide association study approaches were used to identify SCZ gene modifiers from outside the 3MB deletion region in the largest cohort of 22q11.2DS patients to date in this study. The GWAS results replicate those found by IBBC Consortium, with no SNPs surviving correction. Sample size with

sufficient statistical power is a critical aspect to detecting significant genetic association in these studies (Visscher *et al.*, 2017). GWAS studies with hundreds of thousands of individuals, even have limited statistical power, missing associated variants due to polygenic effects and small effect sizes, therefore the small sample size in this study is a major hurdle in identifying genetic variants in 22q11.2 deletion carriers with schizophrenia. Hence why transcriptome wide association was chosen as the next form of analysis as instead of correcting for hundreds of thousands of SNPs, correction for the number of genes expressed are accounted for, reducing the burden of multiple testing. Further to this, TWAS can facilitate biological interpretation of the genetics.

TWAS performed using FUSION with gene expression measured in the DLPFC (Fromer *et al.*, 2016) did not reveal any TWAS associations that survived correction. Change in expression of *STEAP2* was found to the most significant result ($p = 0.175$ after Bonferroni correction) in the DLFPC. STEAP2 is a metalloreductase, facilitating reduction of iron and copper, with expression highest in the prostate, pancreas brain and fetal liver (Ohgami *et al.*, 2006). Fe(III) is important in DGCR8 function, it acts as a cofactor to the heme-binding domain, which is needed as an activator for the processing of pri-miRNA (Quick-Cleveland *et al.*, 2014). Further to this, *STEAP2* has previously been found to have reduced expression in the hippocampus a *DGCR8*-deficient mouse model (Ouchi *et al.*, 2013). This surrounding literature may indicate a potential link between *STEAP2* and *DGCR8* and schizophrenia in 22q11.2DS, but our study remains underpowered and a definitive conclusion cannot be drawn from the result.

The TWAS with the DLPFC expression generated the most significant result found across all the tissue expression panels investigated. Expression changes in the whole blood were less significant, even though this was an increased powered expression panel generated from more patient samples (1264 samples compared to 452 from the DLPFC). This could indicate that the most notable changes in expression are most observed in the disease relevant tissue type, which has been observed for schizophrenia (Bhalala *et al.*, 2018; Ma *et al.*, 2018; Scarr, Udawela and Dean, 2018). Another conclusion from this study is that as there were no significant results observed across all methods of TWAS and gene expression panel investigated, there might not be detectable changes in gene

expression between 22q11.2 deletion carriers with schizophrenia and deletion carriers without psychosis.

TWAS is understood to be a more powerful approach than GWAS, with reduced multiple testing burden due to probing genes compared to SNPs. Furthermore, using predicted gene expression weights removes environmental noise by focusing on the genetically regulated aspect, potentially increasing statistical power (Mancuso *et al.*, 2017). Gusev and colleagues demonstrated that TWAS increased power in the case of multiple causal variants with 92% power for TWAS compared to 18% power for GWAS (Gusev *et al.*, 2016). Ultimately this study is hindered by the small sample size, which is a difficulty when investigating genetic variants in rare diseases, but the sample size of this study has been calculated to provide 80% power to detect a nominally significant difference in polygenic risk score (Cleynen *et al.*, 2020). However, it should also be noted when compared with GWAS, TWAS is underpowered if the genetic risk is not mediated through expression or when expression heritability is too low, which could be relevant to this study (Ding *et al.*, 2020). Venturi and colleagues used a simulation-based method compare power between different TWAS methodologies and GWAS. The key component affecting the power is low trait heritability, which has been shown when it is less and 0.001%, results in low to zero power across all methodologies. However, importantly in relation to my TWAS study, TWAS based upon multiple SNP prediction, such as FUSION and PrediXcan always resulted in the highest power, except when expression heritability and training sample sizes were at their lowest, indicating sample size is an important factor for association detection (Veturi and Ritchie, 2018). It is difficult to calculate the exact power of this TWAS study as it would require extensive simulations accounting for the GWAS summary statistics as well as the selected expression weights. This analysis is not widely reported in the literature and would be challenging to perform.

Increased sample sizes and further investigation is required to characterize differences and the aetiology behind SCZ in 22q11.2DS outside of 22q11.2 deletion region.
Although no disease modifiers were identified in this study, elucidating how genes within deletion region itself contributes to the schizophrenia risk was a primary focus of my work. The next goal was to investigate the role of the selected schizophrenia

candidate genes, DGCR8, HIRA and ZDHHC8 in human embryonic stem cells models through cortical differentiation which is later discussed in this thesis.

4. CRISPR/Cas9 genome editing for derivation of hESCs with altered DGCR8 dosage

4.1 Introduction

A potential schizophrenia (SCZ) candidate gene of interest from within the 22q11.2 deletion locus is *DGCR8*. *DGCR8* encodes for a crucial component of the microprocessor complex, which is a key element in the canonical microRNA (miRNA) biogenesis pathway (Han *et al.*, 2004). MiRNAs play an important role in gene expression regulation and therefore DGCR8 greatly contributes to global transcription regulation. Consequently, the haploinsuffiency of *DGCR8* in 22q11.2DS is likely to lead to wide-spread gene expression dysregulation and potentially contribute to the phenotype of DiGeorge Syndrome.

DGCR8 is involved in the canonical miRNA biogenesis pathway (Han *et al.*, 2004). MiRNAs are small non-coding RNA molecules, approximately 22 nucleotides and have extensive roles in gene expression regulation, commonly gene expression repression (Schwarz *et al.*, 2003). Dysregulation of miRNAs has been observed in neuropsychiatric disorders. MiRNA alteration has been observed in post-mortem brain samples from both patients with SCZ and 22q11.2DS with SCZ (Fénelon *et al.*, 2011). Additionally, there is data to suggest there is an enrichment of rare CNVs overlapping miRNAs in SCZ compared to a control population (Warnica *et al.*, 2015). Furthermore, cortical miRNA biogenesis has been linked to SCZ pathology, with altered levels of miRNAs in the cerebral cortex in SCZ patients (Beveridge *et al.*, 2010). DGCR8 is ubiquitously expressed across many tissues including fetal and adult human and mouse brains, indicating an central role during development and throughout adult life (Shiohama *et al.*, 2003). Homozygous loss of *Dgcr8* is embryonic lethal. Consequent investigations using heterozygous *Dgcr8* mouse models have been essential in elucidating the role of Dgcr8, particularly in relation to SCZ risk. Such studies have revealed many neurological deficits, including impaired short-term synaptic plasticity, reduced excitatory synaptic transmission in pyramidal neurons and working memory deficits (Stark *et al.*, 2008; Schofield *et al.*, 2011).

Multiple lines of evidence indicate that loss of DGCR8 and the microprocessor complex potentially contribute to the neuropsychiatric risk in 22q11.2DS. The microprocessor complex has been found to be essential for neurogenesis in mice. Furthermore, loss of the microprocessor complex in forebrain neural progenitors lead to precocious differentiation shown by loss of Pax6 and exit of cells from the ventricular zone (Knuckles *et al.*, 2012). This phenotype was due to the regulation of the mRNA of pro-neuronal transcription factors such as *Neurog2* by the microprocessor complex. Loss of this regulation lead to aberrant neural stem cell self-renewal. *Dgcr8* knockout and knockdown studies in mouse models have demonstrated key roles for Dgcr8 in neurogenesis and revealed phenotypes of neuronal deficits. *Dgcr8* deletion in the mouse telencephalon lead to impairment of corticogenesis, with premature differentiation of neural progenitor cells (NPCs), potentially due to direct regulation of *Tbr1* transcripts by Dgcr8 (Marinaro *et al.*, 2017). Further to this, overexpression of *Dgcr8* in the mouse telencephalon demonstrated expansion of NPCs and repression of neurogenesis (Hoffmann *et al.*, 2018).

The discovery of the CRISPR/Cas9 system has significantly improved the ability for targeted gene editing and enabled disease modelling in hESCs (Musunuru, 2013). The process requires the expression of sgRNAs and a Cas9 nuclease in target cells, usually through plasmid constructs or viral vectors. These components will induce precise targeted double-strand DNA breaks in the genome. The iCRISPR platform comprises of hESCs with a doxycycline-inducible Cas9 stably integrated into their genome (González *et al.*, 2014). Use of iCas9 cell line eliminates delivery steps required for Cas9 expression in target cells and only necessitates delivery of sgRNAs. Lentiviral-mediated delivery is a commonly used vehicle in CRISPR/Cas9 editing. Lentiviral delivery can allow for high, stable and importantly permanent expression of sgRNAs with low toxicity in hESCs (Zufferey *et al.*, 1998; Buchschacher and Wong-Staal, 2000; Kabadi *et al.*, 2014).

Lentiviruses have proved to be an efficient gene delivery tool *in vitro* (Quinonez and Sutton, 2002). For lentiviral generation, a transfer plasmid encoding your insert of interest and packaging and envelope components are required, these elements are divided across multiple plasmids. Lentiviruses have a single stranded RNA genome that includes three major structural genes, *gag*, *pol* and *env* and several regulatory proteins

(Escors and Breckpot, 2010). The Gag encodes the three viral core proteins, matrix proteins, capsid proteins and nucleocapsid proteins. The *pol* gene encodes for the enzymes essential for viral replication. The *env* gene encodes for the envelope which is the viral surface glycoprotein and is crucial for viral entry into the host cell. The most updated lentiviral system is the development of 3rd generation lentiviral vectors, which have been developed for safe use in a clinical setting as it only requires three HIV-1 genes (Dull *et al.*, 1998). Third generation lentiviruses require four plasmids for production, the transfer plasmid, the envelope plasmid and the packaging system. The packaging system is split across two plasmids, one encoding the *Gag* and *Pol* genes and a separate plasmid encoding the regulatory element *Rev*, differing the system from 2nd generation lentiviruses.

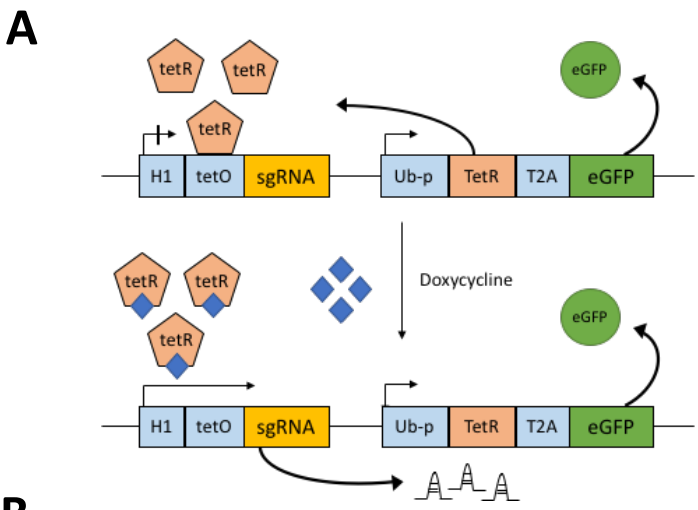
The use of CRISPR/Cas9 gene editing technology using the iCas9 iCRISPR platform with lentiviral transduction of sgRNAs was selected for derivation of *DGCR8* mutant cell lines in this thesis (González *et al.*, 2014). This approach was selected as prior attempts in our lab to generate a *DGCR8* knockout line had been performed in the H7 hESC cell line by transfection. This process generated genetically homozygous and heterozygous *DGCR8* cell lines, but all lines harboured a 390,000bp deletion at the 17p13.1 locus. This locus contains the *p53* tumour suppressor gene, a key cell cycle regulator (Prives and Hall, 1999). p53-mediated DNA damage response has been shown to be induced by CRISPR-Cas9 editing (Haapaniemi *et al.*, 2018) and therefore acquired mutations in this locus may confer a selection advantage for survival in edited cells. Furthermore, Liu and colleagues demonstrated mouse embryonic stem cells deficient for *Dgcr8* showed restricted neural differentiation due to the direct regulation of p53 by miR-302 (Liu *et al.*, 2017). Indicating that with loss of *Dgcr8*, there would be selective advantage for down-regulation of p53 in order to compensate for Dgcr8 deficiency. With this knowledge, it was selected to generate the *DGCR8* knockout hESC lines using a different parental cell line and different sgRNA delivery approach for gene editing, in order to determine whether the 17p13.1 deletion would be observed using an alternative approach. Furthermore, generation of a *DGCR8* mutant line without the CNV would still enable investigation into the effects of *DGCR8* deficiency to potentially understand its role in the SCZ risk in 22q11.2DS.

In this chapter, I describe the generation and validation of *DGCR8* mutant hESC lines using CRISPR/Cas9 technology. This chapter aims to investigate any phenotypes associated with *DGCR8* deficiency in hESCs and through cortical differentiation. This investigation will help to elucidate the role of *DGCR8* in neurodevelopment and the SCZ risk conferred by the 22q11.2 deletion.

4.2 Results

4.2.1 Construction of *DGCR8*-targeting guide RNA expression plasmid and Lentiviral Packaging Plasmids

For generation of *DGCR8* mutant lines, CRISPR/Cas9 technology was employed using a lentiviral-based delivery method in order to disrupt *DGCR8* in iCas9 hESCs. This method involved the cloning of a small guide RNA (sgRNA) into a third-generation lentiviral-based expression system, this system applies the use of the FH1tUTG plasmid (Aubrey *et al.*, 2015). This plasmid contains a doxycycline (dox)-inducible sgRNA expression cassette and ubiquitin promoter driven tetracycline (Tet) repressor linked via the T2A peptide to a green fluorescent protein (GFP). Two sgRNAs were designed targeting exon 2 of *DGCR8,* which is the first coding exon (Figure 4.1). SgRNAs were selected based on high quality score and low off targets score (Figure 4.1C).



C

Guide	Guide Sequence	Target Region	Score	PAM
sgRNA 1	CTCATAGACCCGAACTGTAG	chr22:+200073741	90	TGG
sgRNA 2	GGGGAGAACTTCCGGACCGC	chr22:-200073792	91	AGG

Figure 4.1 Genome editing strategy using CRISPR/Cas9 technology targeting *DGCR8.*
(A) *Schematic of the structure of the doxycycline inducible sgRNA expression plasmid, FH1tUTG. Addition of doxycycline treatment induces sgRNA expression, GFP expression is constitutive.* ***(B)*** *DGCR8 gene and protein structure including CRISPR/Cas9 targeted region in exon 2.* ***(C)*** *Table of sgRNA sequence, genomic target region, score and PAM site cut by Cas9 nuclease.*

Guide RNAs were cloned into the expression plasmid, successful cloning was first validated by restriction enzyme digestion of the region containing the insert. Loss of the BsmBI restriction site indicates insertion of sgRNA (Figure 4.2A). Figure 4.2A also demonstrates that plasmid integrity was maintained after the cloning process as digestion with restriction enzymes NotI and XbaI display the same digestion pattern as the empty plasmid. Positive clones were subsequently sent for sequencing to confirm insertion of the each sgRNA (Figure 4.2B). The lentiviral packaging plasmids, p-CMV-VSV-G, pMDL-pRRE and pRSV-Rev were extracted and purified using the miniprep methods, once positive clones were identified, sufficient quantities of the plasmids were generated and extracted using Qiagen Maxiprep kit. Due to the repetitive nature of

lentiviral plasmids, recombination is possible and so plasmid integrity was confirmed by enzymatic digestion (Figure 4.2C).

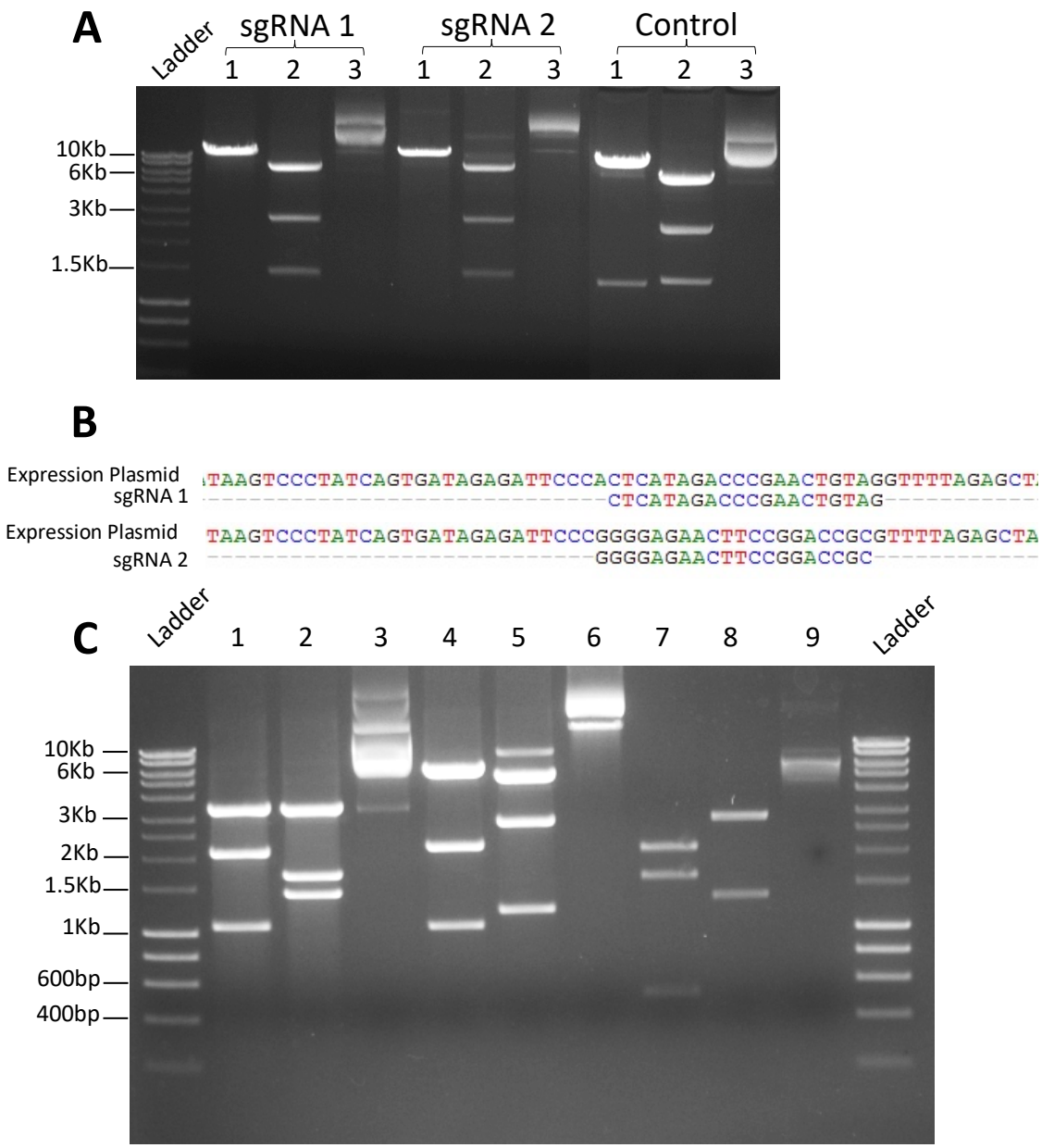


Figure 4.2 Generation of lentiviral construct targeting *DGCR8* and lentiviral packaging plasmids

(A) *DNA gel electrophoresis showing 1) BsmBI and BamHI, 2) NotI and XbaI 3) undigested enzymatic digestion patterns of FH1tUTG plasmid after cloning of sgRNAs compared to empty FH1tUTG plasmid.* ***(B)*** *Sanger sequencing confirming insertion of sgRNA 1 and sgRNA 2 into FH1tUTG expression plasmid.* ***(C)*** *DNA gel electrophoresis confirming plasmid integrity of lentiviral packaging plasmids using the following restriction enzymes: Lane 1: p-CMV-VSV-G digested with BglII and PvuI, expected band sizes 3377, 2085 and 1045. Lane 2: p-CMV-VSV-G digested with BamHI and NotI, expected band sizes 3392, 1692 and 1423 Lane 1: p-CMV-VSV-G undigested. Lane 4: pMDL-pRRE digested with PvuII, expected band sizes 5692, 2154, 1044. Lane 5: pMDL-pRRE digested with NotI and BsmbI, expected band sizes 4996, 2708 and 1186. Lane 6: pMDL-pRRE undigested. Lane 7: p-RSV-Rev digested with BamHI and BglI, expected band sizes 2059, 1594 and 527. Lane 8: p-RSV-Rev digested with PvuII, expected band sizes 1340 and 2840 Lane 9: p-RSV-Rev undigested.*

4.2.2 Generation of *DGCR8* mutant human embryonic stem cells

DGCR8 mutant hESCs were derived from iCas9 cells, a genetically modified pluripotent cell line which harbours dox-inducible Cas9 nuclease expression (González *et al.*, 2014). Gene editing using this cell line requires induction of *Cas9* expression followed by delivery of sgRNA molecules into the cells, in this case, lentiviral-based delivery. Doxycycline induces both the expression of the Cas9 nuclease and sgRNA expression in the FH1tUTG plasmid enabling tight regulation of this system (Aubrey *et al.*, 2015).

Lentiviral transduction was carried out in the presence of polybrene. iCas9 cells were transduced with a lentivirus containing one of the two sgRNAs at concentration of 750,000 lentiviral particles (LVP) or 1million LVP, in the presence of polybrene. The media was then changed 24 hours later, and transduced cells were maintained in doxycycline for 7 days. 48 hours after transduction GFP positive cells were observed in cultures transduced with both sgRNAs (Figure 4.3A). The cells were then dissociated and plated at low density and left to grow until small colonies emerged. The presence of GFP colonies was observed for cells targeted with sgRNA 1 (Figure 4.3B). For cells targeted with sgRNA 2, GFP was initially observed after transduction however no GFP positive colonies survived single cell dissociation and development into clonal colonies. Approximately 100 colonies were manually isolated and expanded as single clones in 48 well plates. The clones were then manually screened for GFP, from this process 6 potential clonal lines were identified. DNA was extracted from these GFP positive colonies and used for PCR screening of the *DGCR8* locus. This process led to identification of 3 positive clones indicating potential gene editing in the region of interest. This is demonstrated in the agarose gel used to analyse the PCR of the targeted *DGCR8* region shown in Figure 4.3C. The gel electrophoresis in Figure 4.3C displays the PCR of the amplified *DGCR8* target region, demonstrates that clone #1, #2 and #3 have an edited allele due to presence of two different bands compared to the singular band observed in the wildtype.

Sanger sequencing was then used to verify the PCR screening results to detect potential insertion/deletion mutations (INDELs) generated from the editing process. The sequencing results for the lines used subsequently in this thesis are presented in Figure

4.4. Sequencing revealed that on one allele, clones #1 and #2 harbour a 426bp inversion that begins before the start codon of exon 2 and would therefore disrupt the transcriptional start site and result in a heterozygous deletion *DGCR8*. Clone #3 also contains this same inversion and an in frame 3bp deletion on the second allele. Using the programme PROVEAN (protein variation effect analyser), this predicts that this 3bp deletion would have a neutral effect on the protein, with a score of -1.989 (Choi and Chan, 2015). This indicates that although this line is genetically homozygous, it will likely be functionally heterozygous. No other homozygous clonal lines were derived from this process, but the heterozygous lines were used for further characterisation and the lines used will be referred to as *DGCR8* mutant hereafter.

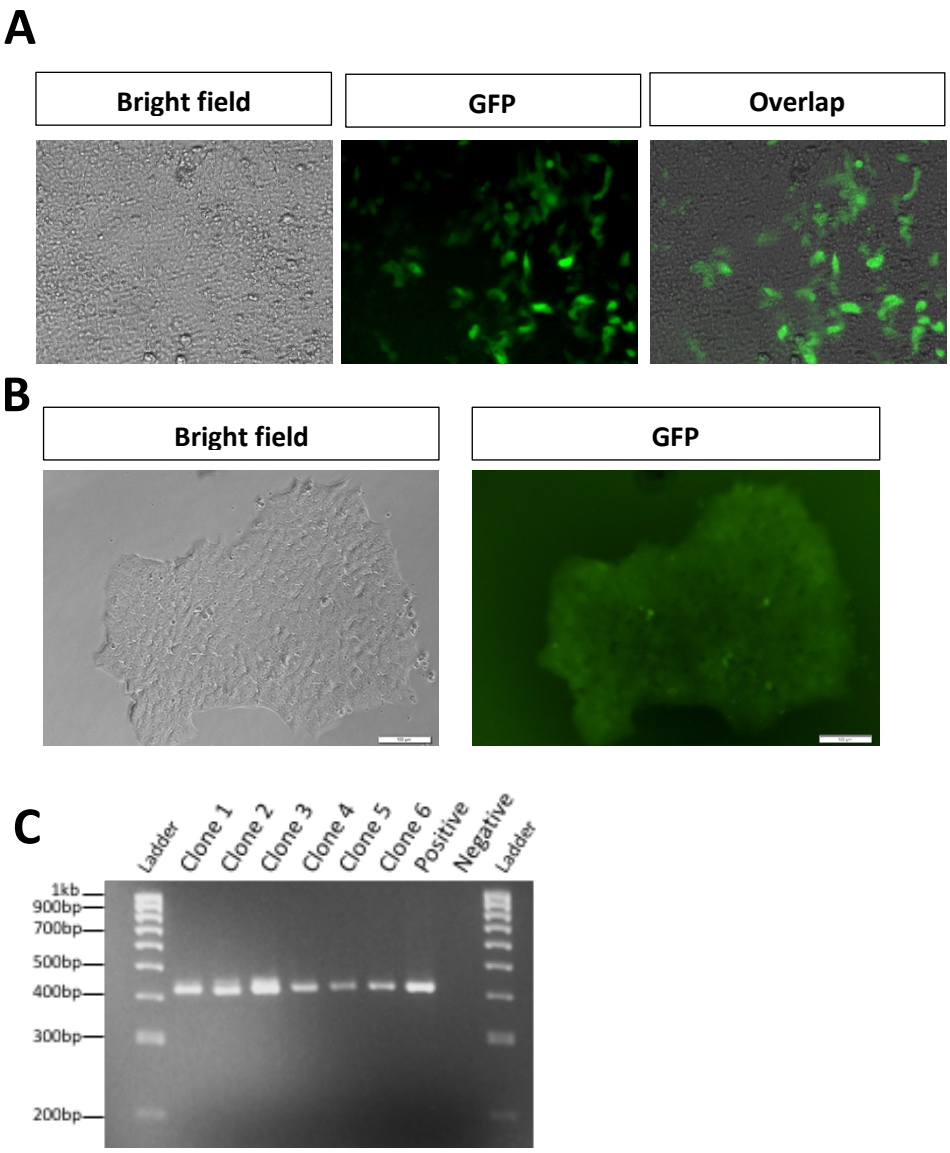


Figure 4.3 Lentiviral transduction and screening of targeted iCas9 hESCs
(A) *Bright field images taken 48hours after transduction iCas9 hESCs with lentivirus targeting DGCR8.* ***(B)*** *Example image of GFP positive hESC colony. Scale bars = 100µM*
(C) *Example of DNA gel electrophoresis showing PCR products of the targeted locus from 6 edited clones, with lanes 1-3 showing a different PCR product compared to non-transduced iCas9 hESCs in the positive lane, indicating gene editing has taken place.*

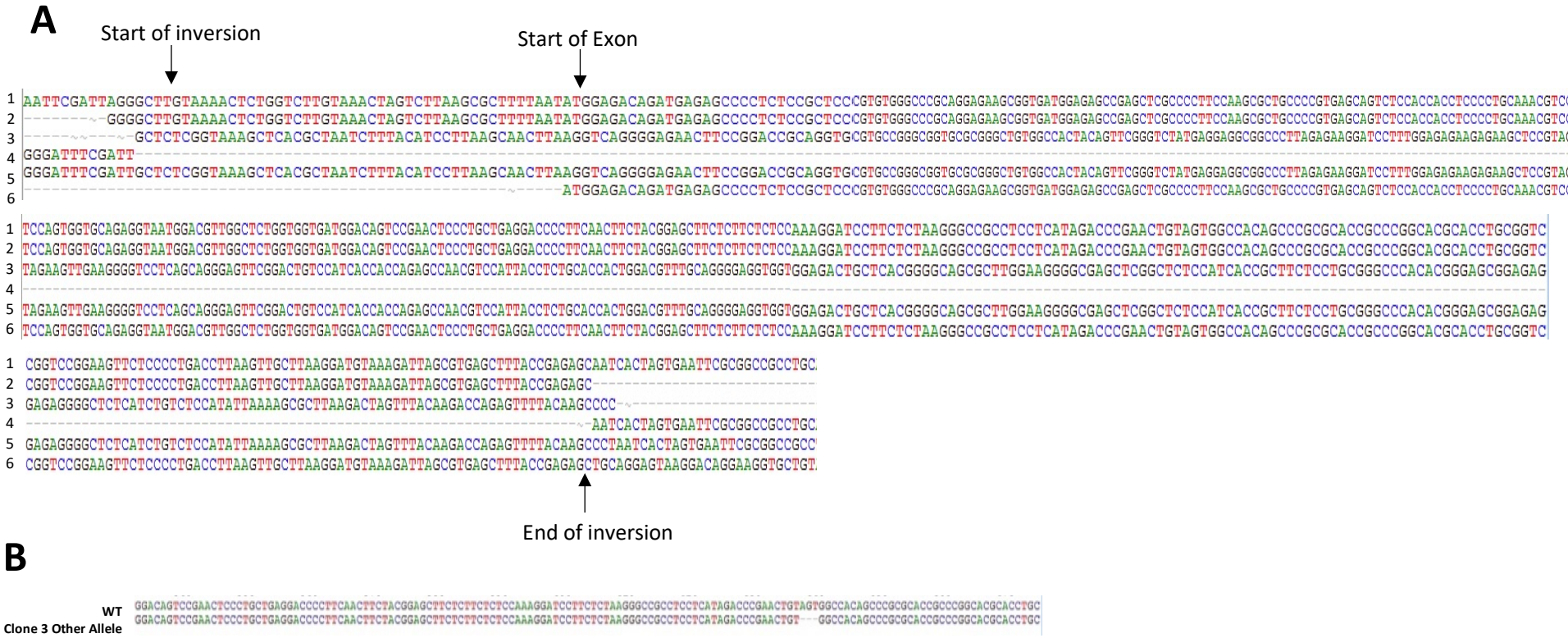


Figure 4.4: Sequencing results of *DGCR8*-mutant hESCs

(A) *Partial DNA sequence of DGCR8 exon 2. Row 1 showing the WT sequence from parental iCas9 line. Row 2 showing the reverse complement of the inverted sequence found in three clones. Row 3 showing non-inverted sequence found in three clones. Row 4 showing sequence found in all three clones without the 436bp inversion. Row 5 showing the full sequence found in three clones. Row 6 showing the sequence of DGCR8 exon 2.* ***(B)*** *Partial DNA sequence of DGCR8 exon 2 wildtype sequence from iCas9 parental line and 3bp deletion found in the other allele of clone 3.*

4.2.3 Validation and characterisation of *DGCR8* mutant hESC line

CNV analysis was performed on the *DGCR8* mutant lines in order to check for chromosomal duplications, deletions of rearrangements above 100,000bp. A duplication at chr8p23.2 is observed in the parental line, which is also observed in the edited lines. The CNV analysis screened the 17p13.1 region, in which a large deletion had been observed in prior generated *DGCR8* mutant hESC lines by a previous lab member. In my derived cell lines, importantly no CNVs were detected at the 17p13.1 region (Figure 4.5). Additionally, no other CNVs were observed as a by-product of the editing process, which demonstrates that any potential phenotypes of the cell line will be due the targeted edited process (Figure 4.5). CNV generation in hESCs is a common side effect of the editing process and can undermine any downstream analysis if off-target genetic mutations are not detected (Laurent *et al.*, 2011).

The pluripotency status of the genetically homozygous *DGCR8* mutant was established, as pluripotency has been shown to be affected when *DGCR8* is dysregulated in mouse ESCs (Wang *et al.*, 2007). The *DGCR8* mutant line was stained for the pluripotency markers OCT4, SOX2 and TRA-1-81 and these markers were expressed in comparable proportion to the wildtype line (Figure 4.6A). Furthermore, no spontaneous differentiation or deficits in cell viability and proliferation were observed in the *DGCR8* mutant line. This data confirmed normal CNV and pluripotency status of the CRISPR/Cas9 generated *DGCR8* mutant lines, indicating they are suitable for disease modelling.

Protein status of DGCR8 was investigated using western blot in the three derived clonal lines from protein samples extracted at the embryonic stem cell stage. No significant difference in DGCR8 protein levels was observed between the *DGCR8* mutant lines compared to the wildtype. Clone #1 was found to have approximately 12% reduction, clone #2 was found to have a 5% reduction and clone #3 had a 20%, reduction respectively (Figure 4.6B) Although clone #1 and #2 are genetically identical, there is a difference in the level of protein reduction, indicating there are other factors influencing DGCR8 protein production. Clone #3 is genetically homozygous and although the 3bp allele on the second allele was not predicted to affect the protein, this clone had the

largest level of protein reduction, hence clone #3 was used for most of the experiments presented in this thesis due to this decrease in protein expression. *DGCR8* mRNA expression was not quantified due to surrounding literature indicating that protein reduction is a relevant to affecting miRNA production and downstream phenotype and furthermore is observed in 22q11.2DS patients, therefore determining protein levels was prioritised as a line of investigation (Wang *et al.*, 2007; Sellier *et al.*, 2014).

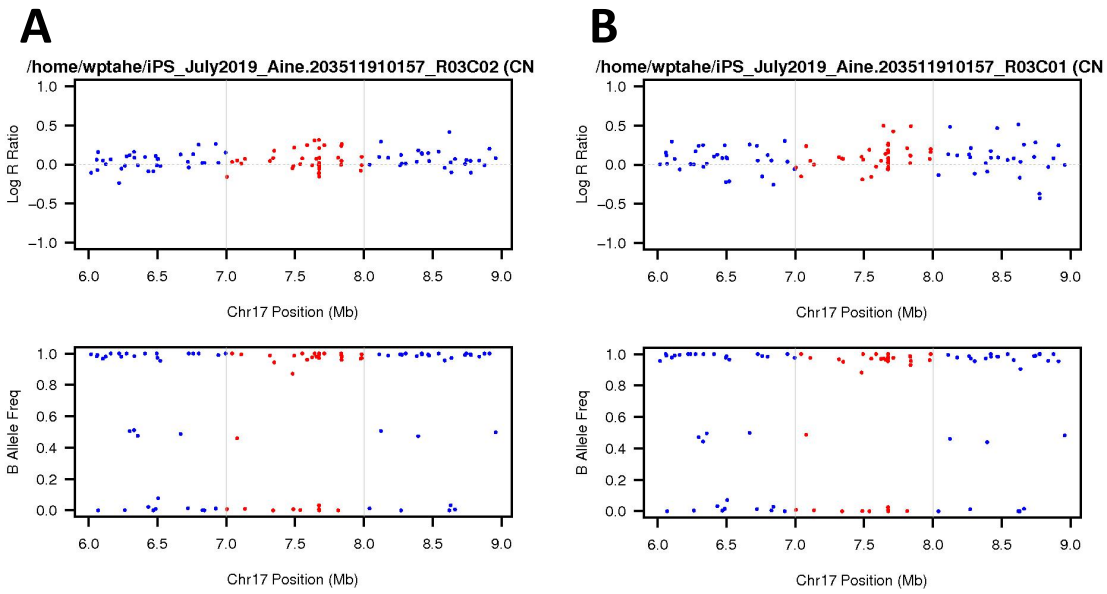


Figure 4.5 CNV status: Example SNP array plots of iCas9 and *DGCR8* mutant lines
(A) *iCas9 plot showing normal CNV status at locus chr 17p13.1, characterised by the Log R ratio at approximately zero.* ***(B)*** *Example plot from DGCR8 clone showing normal CNV status at locus chr 17p13.1, characterised by the Log R ratio at approximately zero.*

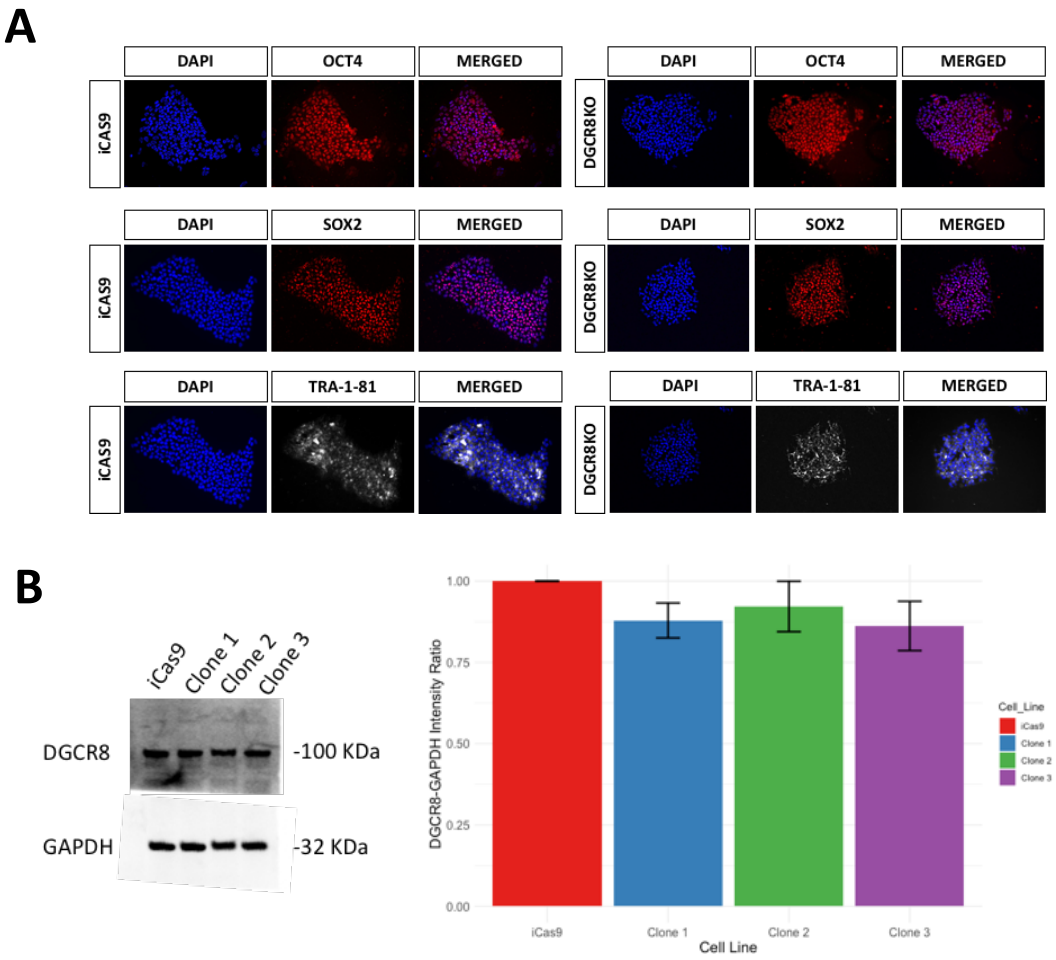


Figure 4.6 Validation of *DGCR8*-mutant hESC lines *(A)* *Undifferentiated iCas9 and DGCR8 mutant hESCs stained for pluripotency markers OCT4, SOX2 and TRA-1-81.* ***(B)*** *Western blot for DGCR8 in iCas9 parental line and edited clones (n =1 per clone, with 3 technical replicates values for independent replicates shown as mean ± SEM) and no significant difference observed between clones.*

4.2.4 Phenotypic observations during cortical neuronal differentiation

To verify neuronal induction efficiency and ability to differentiate into neurons, the *DGCR8* mutant line and the iCas9 parental line were differentiated into cortical pyramidal neurons using an adapted protocol from published studies (Chambers *et al.*, 2009; Cambray *et al.*, 2012; Arber *et al.*, 2015). Using this protocol, the iCas9 and *DGCR8* mutant stem cells differentiate into neuroprogenitor cells (NPCs) at approximately day 12 of the differentiation. At this timepoint, cells were fixed and stained for markers of cortical radial glia cells to determine NPC generation. Figure 4.7A shows that the at the NPC stage, the *DGCR8* mutant line demonstration expression of the typical neuron progenitor marker NESTIN, as expected. NESTIN is an intermediate filament detected

throughout the proliferating central nervous system during development (Dahlstrand, Lardelli and Lendahl, 1995). The proportion of the transcription factors PAX6 and FOXG1 were analysed in the mutant and wildtype lines. PAX6 and FOXG1 are expressed are telencephalic transcription factors, PAX6 is expressed in dorsal forebrain progenitors and FOXG1 in all forebrain cells (Greig *et al.*, 2013). 74% of the *DGCR8* mutant line was PAX6 positive cells, which is analogous to the wildtype with 70% PAX6 positive cells. FOXG1 was displayed in 73% of the *DGCR8* mutant line, which is comparable to the wildtype with 70% FOXG1 positive cells, therefore there were no significant differences between the *DGCR8* mutant and control cell line in these markers examined at day 12 (Figure 4.7B-E).

Furthermore, at the neuroprogenitor cell stage, no difference was observed in the proportion of cells positive for the proliferative marker Ki67 between the mutant and wildtype (Figure 4.7F-G). Cultures were stained for NCAD (N-Cadherin or Cadherin 2), a protein that *in vitro* marks the centre of neural rosettes, which are 2-dimensional structures that are equivalent to neural tube formation in development. The *DGCR8* mutant line showed normal neural rosette formation compared to the wildtype (Figure 4.7H). This data demonstrates that the *DGCR8* mutant line does not show deficits in efficiency of neural induction or proliferation of cortical progenitor cells at this stage of differentiation.

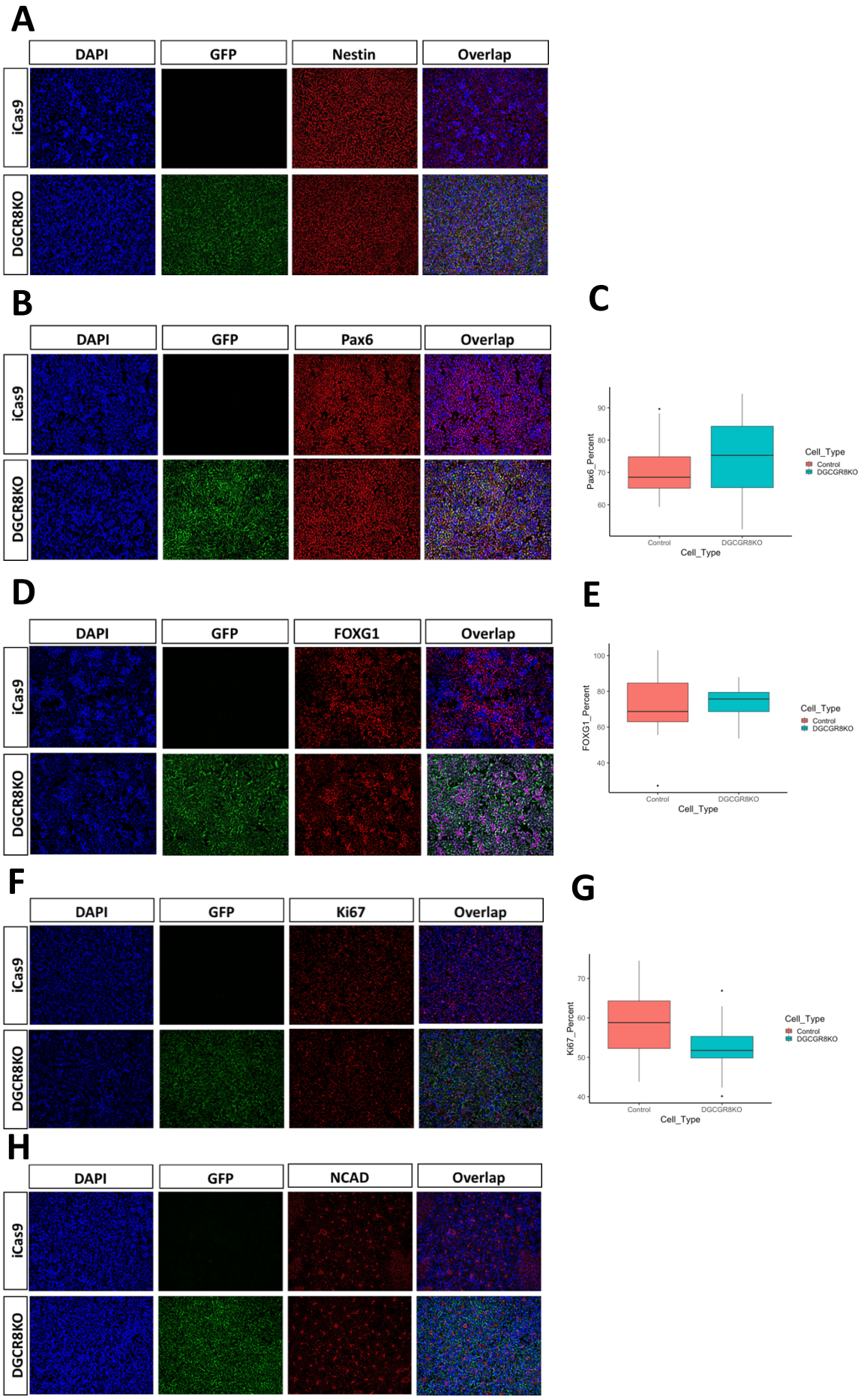


Figure 4.7 Expression of cortical progenitor markers in *DGCR8* mutant and WT cultures.
(Legend on the next page)

Figure 4.7 Expression of cortical progenitor markers in *DGCR8* mutant and WT cultures.
Immunofluorescence for Nestin (A), PAX6 (B), FOXG1 (D), Ki67 (F) and NCAD (H) in DGCR8 mutant and iCas9 cultures at day 15. All nuclei were counterstained with Dapi (blue). All cells were stained for GFP (green). Quantification of cells expressing PAX6 (C), FOXG1 (E) and Ki67 (G) in DGCR8 mutant and iCas9 NPCs at day 15 of differentiation. All quantification data were compared by t-test and no significant difference were found.

To investigate capacity for neuron generation at a later stage in differentiation, cultures were fixed and stained at day 25. During mammalian cortical development, the cortex is organised into six distinguishable layers, which appear in a strict order (Molyneaux *et al.*, 2007). The transcription factors TBR1 and CTIP2 appear sequentially in corticogenesis in deep layer formation and this is reflected in *in vitro* differentiation (Sadegh and Macklis, 2014). TBR1 is a marker specifically expressed by new-born layer VI cortico-thalamic projection neurons, furthermore it is indicated to be regulated by Dgcr8 (Marinaro *et al.*, 2017). There was no significant difference between the wildtype and *DGCR8* mutant lines for TBR1, with 19.8% and 20.2% of cells being TBR1 positive in the mutant and wildtype line respectively (Figure 4.8B and 4.8C). CTIP2 is a transcription factor strongly expressed by projecting neurons in layer V involved in axonal projection (Chen *et al.*, 2008). The *DGCR8* mutant line show no significant difference in the percentage of CTIP2 positive compared to the wildtype line (14.1% and 18.4% in the mutant and wildtype) shown in Figure 4.8A and 4.8C. Overall this data indicates the *DGCR8* mutant line showed no deficits in neural induction and cortical neuron differentiation compared to the wildtype cell line.

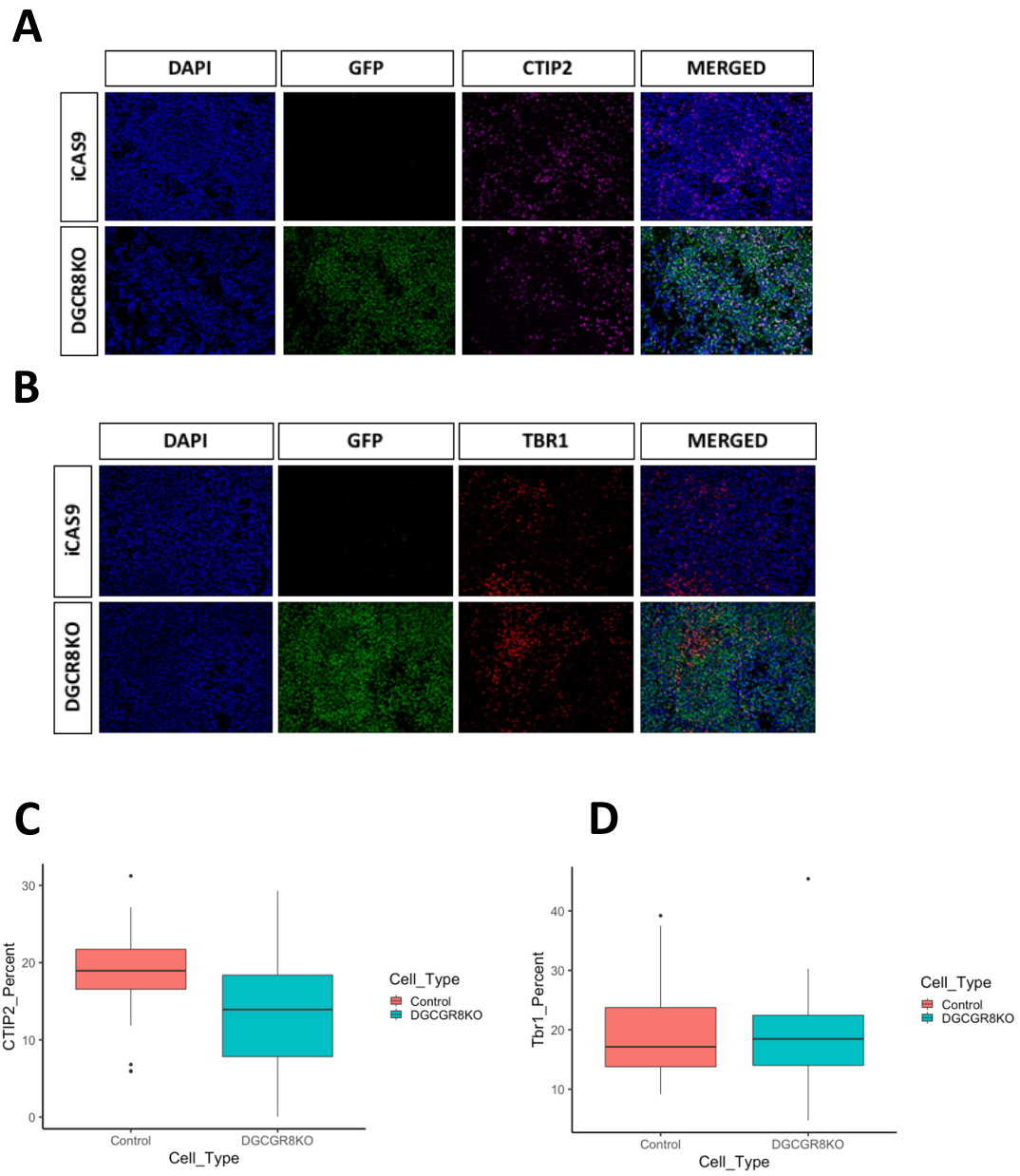


Figure 4.8 Expression of CTIP2 and TBR1 in DGCR8 mutant and WT cultures.
Immunofluorescence for CTIP2 (A) and TBR1 (B) in DGCR8 mutant and iCas9 cultures at day 25. All nuclei were counterstained with Dapi (blue). All cells were stained for GFP (green). Quantification of cells expressing CTIP2 (C) and TBR1 (D) in DGCR8 mutant and iCas9 NPCs at day 25 of differentiation. All quantification data were compared by t-test and no significant difference were found.

4.3 Discussion

This chapter describes the derivation of a *DGCR8* mutant hESC line, with the aim to investigate the function of DGCR8 in human brain development and its contribution for increased psychiatric risk in 22q11.2DS.

The use of the CRISPR/Cas9 gene editing technology allowed for generation of *DGCR8* mutant cell lines. These lines were derived using lentiviral-based delivery of sgRNAs, targeting the first translated exon of *DGCR8* in the iCas9 line (González *et al.*, 2014). All the cell lines were derived from transduction with the sgRNA 1 expressing lentivirus. All three lines contained a 426bp inversion that begins before the start codon and therefore disrupt one copy of *DGCR8*. Clone #3 also harboured a 3bp in frame deletion on the other allele, but this is an in-frame deletion, therefore it does not disrupt translation and was also predicted to have a neutral effect on protein structure. As all generated lines have the same inversion, it is possible they could be daughter cells, as after transduction the hESCs had a period where replication was possible before single cell dissociation. This could also be a result of genomic instability in this region and therefore a susceptibility to form this specific mutation. No true functional homozygous mutant was obtained from the process, which could be due to multiple factors. Firstly, this could be due to an inadequate number of clones screened, as there was low survival of clonal colonies. Secondly, the lack of surviving colonies could also indicate that complete loss of the DGCR8 protein may compromise self-renewal, *Dgcr8* deficient mESC lines demonstrate a slower proliferation rate (Cirera-Salinas *et al.*, 2017) and therefore if this phenotype were to occur in hESCs it would limit the ability to generate complete KO cell lines.

Before use of the derived cell lines for investigation of DGCR8 in psychiatric risk in neurodevelopment, the CNV and pluripotency status of the lines were assessed. It was important to assess CNV status, firstly because of the use of lentiviral-based delivery of sgRNAs, which relies on random genome integration and so undesired mutations could be introduced in the process. Secondly, to determine if the previously seen CNV at the 17p13.1 locus is observed in a *DGCR8* mutant cell line derived using a completely different protocol. However, the CNV status of the generated clonal lines were shown

to be normal. Although, there has been an established mechanistic link between *DGCR8* deficiency and regulation of p53, meta-analysis of genetic abnormalities in human induced pluripotent stem cells has since revealed that this is a commonly occurring mutation (Assou *et al.*, 2020). The lines generated in this thesis did not harbour any additional CNVs, indicating this CNV may have occurred due to the editing process, rather than a DGCR8 related mechanism. *DGCR8* mutant lines demonstrated a normal pluripotency status that was comparable to the parental line, indicating they are suitable for disease modelling.

DGCR8 protein levels were not significantly reduced in the *DGCR8* mutant cell lines, although the cell lines have disruption of one *DGCR8* allele. This could indicate that a potential compensatory mechanism within the cells to ensure sufficient DGCR8 protein levels are maintained for survival. Furthermore, although most autosomal gene expression is biallelic, some genes can respond to one allele inactivation by increasing the expression of the remaining allele, therefore heterozygous gene expression may not reduce protein expression proportionally (Cook, Gerber and Tapscott, 1998). Han and colleagues have also reported that *Dgcr8* heterozygous mESCs did not demonstrate a significant protein reduction (Han *et al.*, 2009). *Dgcr8* KO mESCs demonstrated an abnormal differentiation phenotype and cannot exit from the pluripotency state as pluripotency markers are not fully downregulated (Wang *et al.*, 2007; Cirera-Salinas *et al.*, 2017). Whereas the generated *DGCR8* mutant cell lines were successfully able to differentiate into the cortical lineage. Previous literature in mouse models with altered dosage of *Dgcr8* demonstrated aberrant phenotype in cortical development (Marinaro *et al.*, 2017; Hoffmann *et al.*, 2018). Notably, in these models, there was an observed reduction of Dgcr8 protein levels, therefore this could elucidate the lack of phenotype in the derived *DGCR8* mutant line. Additionally, work in mESCs demonstrated that heterozygous *Dgcr8* mESCs showed normal expression of miRNA, only in homozygous lines was there observed reduction of miRNA (Wang *et al.*, 2007), therefore further reduction of the DGCR8 protein could be required to induce a phenotype in hESCs. However, in mouse development by postnatal D5 in $Dgcr8^{+/-}$ mice show no changes in mRNA expression compared to wildtype, the reduction was observed at p25 and therefore protein levels should be assessed through cortical differentiation in the mutant cell lines (Schofield *et al.*, 2011). Furthermore, it would be beneficial to assess

mRNA levels of *DGCR8*, as it there is not adequate reduction of mRNA, it would underlie the lack of protein reduction. This was not performed due to focusing on later described work. The absence decreased protein levels and no obvious aberrant phenotype, along with the work in mouse cell lines could indicant significant mRNA and protein reduction is required to induce a phenotype. To generate attempt homozygous knock cell line generation, using a system with multiple guide RNAs, targeting numerous sites throughout *DGCR8* may have aided the generation of a homozygous line, whereas the use of one sgRNA may not have been sufficient.

22q11.2DS syndrome patient peripheral leukocytes have been shown to have approximately 50% reduced DGCR8 expression, demonstrating that heterozygous loss of *DGCR8* is sufficient to cause the same level of reduction at the protein level at the patient level and is relevant to contributing to the 22q11.2DS phenotype (Sellier *et al.*, 2014). However, this is the only study that has measured DGCR8 protein levels in 22q11.2DS patients. A different study in which iPSCs were generated from 22q11.2DS patients investigated *DGCR8* mRNA levels, which varied significantly between cell types, with no differences observed in fibroblasts and pluripotent iPSCs compared to control patients. A significant reduction of *DGCR8* mRNA was only observed when iPSCs were differentiated into neurospheres (Toyoshima *et al.*, 2016). Further investigation into other molecular and physiological characterisation is required to understand a potentially more subtle phenotype in the DGCR8 mutant, such as transcriptomic analysis. There is also potential that the phenotype could be observed through examining the properties of the neurons, such as through electrophysiological investigation, which would not be detected through morphological analysis. This would allow for a better understanding at the potential changes resulting from hemizygous loss of *DGCR8* and its role in development.

The generation of these *DGCR8* mutant cells line could provide further evidence that that DGCR8 is too critical for cell survival and so knockout is non-viable in human embryonic stem cells. This process could only generate heterozygous knockout cell lines, without reduction of DGCR8 protein levels. Previous attempts by our Lab using the H7 cell line and a transfection method also could not generate functionally homozygous mutants. Furthermore, all the lines harboured a deleterious 390,000bp CNV, indicating

this CNV was required for survival. Since starting this work Deng and colleagues reported attempts at generating a full *DGCR8* KO human embryonic stem cell line, their attempts that yielded complete loss of DGCR8 protein resulted in an abnormal karyotype or poor maintenance of self-renewal and significant spontaneous differentiation (Deng *et al.*, 2019). These combined investigations in generating *DGCR8* null hESC lines indicate that DGCR8 is essential for hESC maintenance and this may not be a viable approach to study the role of *DGCR8* in disease modelling. It appears technically challenging to develop a homozygous *DGCR8* hESC line, but equally heterozygous deletions of *DGCR8* do not seem to induce a significant level of protein expression reduction to cause a phenotype in hESCs and through cortical differentiation.

Consequently, hESCs might not be a suitable modelling system for *DGCR8* loss for elucidating its role in neurodevelopment. As homozygous and heterozygous mESCs have been previously developed and characterised, they could offer a better opportunity for understanding the effects of loss of *Dgcr8* (Wang *et al.*, 2007). Further work is required to characterise this hypothesis, potentially with use of conditional *DGCR8* knockout in cortical differentiation (which is discussed in Chapter 5), in order to understand the mechanisms underlying the neuropsychiatric phenotype in 22q11.2DS.

5. Gene editing of 22q11.2 deletion genes in human neuroprogenitors using a lentiviral-based delivery method

5.1 Introduction

The hemizygous deletion at the 22q11.2 locus is associated with a more than 20-fold increased risk for developing schizophrenia, however the mechanisms underlying this risk remain largely unknown. The "typical" deletion spans a 3Mb region, encompassing approximately 90 genes, including coding RNAs, noncoding RNAs and pseudogenes. Haploinsuffiency of a singular gene from the region has not been implicated in causing the increased SCZ risk. There are several lines of evidence indicating loss of multiple genes, which share molecular functions or multigenic interactions contribute to the 22q11.2DS phenotype (Williams, 2011; Jonas, Montojo and Bearden, 2014; Motahari *et al.*, 2019). With this hypothesis in mind, prior work discussed in Chapter 3 led to the selection of *DGCR8*, *HIRA* and *ZDHHC8* as potential SCZ candidate genes to investigate in a human neuroprogenitor cell (NPC) model. This selection was based upon shared gene expression in human fetal brain and predicted pathogenic effects based on loss-of-function and haploinsufficiency.

As discussed in the previous chapters, *DGCR8* has been implicated as a SCZ candidate gene in 22q11.2DS. DGCR8 has been shown to have a role in neurogenesis and corticogenesis. *Dgcr8* depletion and overexpression have been shown to lead to aberrant neurogenesis with a mirror phenotype (Marinaro *et al.*, 2017; Hoffmann *et al.*, 2018). Furthermore, deficits in short-term plasticity have been observed in *Dgcr8*$^{+/-}$ mutant mice (Fénelon *et al.*, 2011). Dgcr8 deficiency has been implicated as a contributor to synaptic abnormalities in 22q11.2DS mutant mice (Fénelon *et al.*, 2011; Schofield *et al.*, 2011). Furthermore, the work of Marinaro and colleagues have demonstrated a potential relationship between the transcript of the cortical neuron marker *Tbr1* and Dgcr8 (as described in Chapter 1, section 1.3.3.3). It was established that conditional loss of Dgcr8 in the mouse cortex lead to an increase in Tbr1 positive neurons, potentially due to regulation of *Tbr1* mRNA by Dgcr8 (Marinaro *et al.*, 2017),

however the relationship between DGCR8 and TBR1 has not yet been examined a human cell model.

HIRA is a histone chaperone with a critical role in the deposition of the histone variant H3.3 onto DNA in the DNA-synthesis-independent deposition pathway in nucleosome assembly (Ray-Gallet *et al.*, 2002; Tagami *et al.*, 2004). Eukaryotic cells have two modes of nucleosome assembly: 1) DNA replication-coupled chromatin assembly, which occurs in proliferating cells during S phase and 2) replication-independent chromatin assembly, which occurs in S phase and throughout the cell cycle. The HIRA chaperone complex consists of HIRA/UBN1/CABIN1 and transiently ASF1a, to facilitate H3.3 deposition. In mammalian ESCs and NPCs, HIRA is required for genome-wide H3.3 enrichment at active and repressed genes and been observed at many silent gene promoters (Guenther *et al.*, 2007; Goldberg *et al.*, 2010). HIRA has also been implicated in facilitating chromatin recovery after DNA damage (Adam, Polo and Almouzni, 2013).

HIRA is required for normal embryonic development. Homozygous *HIRA* deficient embryos are embryonic lethal by day 11 (Roberts *et al.*, 2002). HIRA has widespread involvement in mammalian development and has been implicated across numerous processes that contribute to many of the 22q11.2DS phenotypes. Conditional KO of *Hira* in embryonic mouse hearts leads to cardiac ventricular and atrial septal defects. This is due to histone chaperone activity affecting gene expression vital for heart development and cardiomyocyte homeostasis (Dilg *et al.*, 2016; Valenzuela *et al.*, 2016). The role of HIRA in neurogenesis and brain development remains largely unknown. Histone H3.3 levels rise to approximately 90% of total H3 content in mature cortical neurons due to being expressed independently of the cell cycle (Piña and Suau, 1987). In a *Drosophila* model, HIRA-mediated H3.3 deposition was found to be regulated by intellectual disability gene BRWD3, with increased H3.3 levels disrupting gene expression, dendritic morphogenesis and sensory organ differentiation (Chen *et al.*, 2015). Li and colleagues have demonstrated that neuroprogenitor cells are enriched in Hira and it peaks at E15 in mouse cerebral tissues. Hira knockdown lead to inhibition of NPC proliferation, with increased cell cycle exit and premature differentiation through regulation of β-catenin expression and its transcriptional targets (Li and Jiao, 2017). The wide-spread transcriptional implications of happloinsuffiency of *HIRA* make it an interesting

candidate to explore during neuronal development and as a potential SCZ risk gene in 22q11.2DS.

ZDHHC8 encodes a palmitoyltransferase which is highly expressed in the brain (Mukai *et al.*, 2004). Palmitoylation is the covalent attachment of a saturated 16-carbon palmitic acid to cysteine residues of target proteins. This post-translational modification is, in the majority of cases, reversible and therefore allows cells to respond to extracellular signals and can maintain homeostasis. Palmitoylation can control a number of cellular processes such as subcellular trafficking, protein localization and stability as well as enzyme activity (Greaves and Chamberlain, 2007; Aicart-Ramos, Valero and Rodriguez-Crespo, 2011). Multiple lines of evidence suggest that palmitoylation plays a key role in neuronal development and synaptic plasticity (El-Husseini *et al.*, 2002; Hayashi, Thomas and Huganir, 2009; Yokoi, Fukata and Fukata, 2012). Approximately 40% of synaptic proteins are palmitoylated (Sanders *et al.*, 2015) and palmitoylation is suggested to be essential for the proper function of the majority of pre- and post-synaptic proteins (as reviewed in (El-Husseini and Bredt, 2002)).

Initial genetic studies associated SNPs within *ZDHHC8* with increased risk for SCZ (Mukai *et al.*, 2004; W. Y. Chen *et al.*, 2004). However, this result failed to replicate in later studies with larger sample sizes (Glaser *et al.*, 2005; Pardiñas *et al.*, 2018). Nevertheless, depletion of ZDHHC8 has been implicated in abnormal neuronal morphology and behaviour. Primary cultures from Zdhhc8-deficient mice show deficits in axonal growth and terminal arborization, concordant with a 22q11.2DS murine model, which was attributed to decreased palmitoylation of cdc42 (Mukai *et al.*, 2015). Moreover, ZDHHC8 has been found to palmitoylate the post synaptic density protein PSD-95, an adaptor protein localised at the postsynaptic neurons, involved in formation and maintenance of synapse and excitatory maturation of neurons (Mukai *et al.*, 2008; Chen *et al.*, 2011). Behavioural defects observed *Zdhhc8* deficient mice include impaired spatial working memory and decreased pre-pulse inhibition (Mukai *et al.*, 2004, 2015). With such an evident role in neuronal formation and maturation, *ZDHHC8* is an interesting candidate for contribution to SCZ risk in 22q11.2DS.

CRISPR/Cas9 technology has provided the ability to introduce precise disease associated mutations within the genomes of human cells. This technology enables the modelling and assessment of molecular and cellular phenotypes resulting from induced genetic manipulations. The most common use of this system generates “knock-out” cell lines, with permanent mutations in the genome. CRISPR/Cas9 cell line generation requires editing when the cells are pluripotent, to allow for self-renewal and generation of clones from single cells, however this approach is not always applicable to every gene. Certain genes are essential for cell survival, proliferation or differentiation (Gao *et al.*, 2015). Moreover, expression can be spatially and/or temporally controlled, therefore it could be more beneficial to investigate gene loss during specific developmental windows (Nicholson *et al.*, 2008). Furthermore, differentiation can increase the chance of silencing, as transgenes can be located in regions where heterochromatin form following cell fate changes (Herbst *et al.*, 2012). Therefore, in some cases gene editing when commitment to cell fate has occurred is advantageous.

Gene editing in human neuroprogenitor cells (NPCs) and neurons using an inducible system is a viable option for modelling neurological disorders and has been explored for modelling neurodegenerative diseases (Heman-Ackah, Bassett and Wood, 2016). Lentiviral-based gene delivery offers advantages over other delivery methods, with stable integration of transgenes into the host genome and importantly the ability to infect dividing and non-dividing cells with low toxicity (Johnston *et al.*, 1999; Federici *et al.*, 2009). Lentiviral gene delivery has been shown to be effective in usually “hard to transfect” NPCs (Blits *et al.*, 2005; Rubio *et al.*, 2016) and has demonstrated long-term stable expression compared to other viral vector systems (Vroemen, Weidner and Blesch, 2005).

With this in mind, the aim of this chapter was to explore the feasibility and effects of CRISPR/Cas9-based gene manipulation of the selected SCZ gene candidates *DGCR8*, *HIRA* and *ZDHHC8* in NPCs derived from the iCas9 cell line via lentiviral mediated gene delivery in cortical neural development. Previous work in Chapter 4 had attempted to generate a *DGCR8* knockout hESC line, however reduction in at the protein level was not observed. There is evidence to indicate the role of DGCR8 is essential in stem cell maintenance and therefore knockout is detrimental for cell survival, therefore the

applying gene editing once commitment to neuronal fate provides a viable option for investigating the role of DGCR8 during cortical differentiation. To achieve this aim, efficient sgRNAs were designed and tested for each target gene. Before gene manipulation, the effects of a non-targeting lentivirus on cortical development were characterised. The non-targeting lentivirus showed an increased level of transduction in NPCs expressing deep layer cortical markers compared to non-transduced cells, with more transduced cells being TBR1 and CTIP2 positive compared to non-transduced cells. This is potentially due to increased lentiviral efficiency for infecting more mature intermediate progenitor cells. I then went on to specifically examine the phenotypic effects of *DGCR8* manipulation in cortical differentiation. *DGCR8* gene editing in the NPCs led to an increased proportion of TBR1 positive cells, further adding to the evidence for a relationship between DGCR8 and *TBR1* as discussed in Chapter 1, section 1.3.3.3).

Although, it was planned to look at genetic manipulation of all selected SCZ candidate genes in NPCs and investigate potential shared phenotypes between them, due to the COVID-19 pandemic this could not be achieved. The work below is described with this overall aim in mind, until investigations had to be halted.

5.2 Results

5.2.1 Generation and validation of sgRNA expression vectors targeting *DGCR8*, *HIRA* and *ZDHHC8*

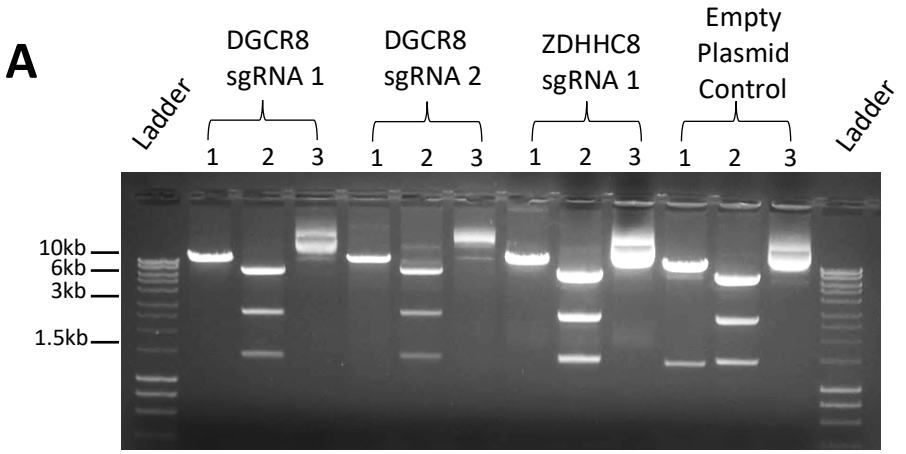
As utilised in Chapter four, the lentiviral vector FH1tUTG was used to express sgRNAs for targeted genetic manipulation of *DGCR8*, *HIRA* and *ZDHHC8* in iCas9-derived NPCs (Aubrey *et al.*, 2015). Guide RNA expression is dependent on the addition of doxycycline (Chapter 4 Figure 4.1A), as is the expression of the Cas9 nuclease in the iCas9 cell line (González *et al.*, 2014). The two sgRNAs designed targeting exon 2 of *DGCR8* that were described in Chapter 4 were also used for targeting in NPCs. Two sgRNAs were designed to target exon 4 and exon 6 of *HIRA* and two sgRNAs were designed to target exon 3 and exon 6 of *ZDHHC8.* SgRNAs were selected based on high quality score and low off-target number (Table 5.1).

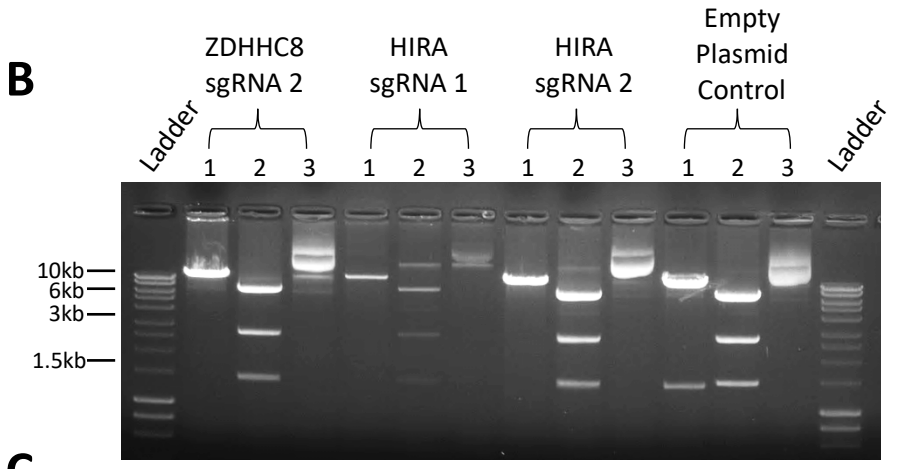
Table 5.1: Table of sgRNA sequences targeting *DGCR8, HIRA* and *ZDHHC8*, genomic target region, score and PAM site cut by Cas9 nuclease.

Guide	Guide Sequence	Target Region	Score	PAM
DGCR8 sgRNA 1	CTCATAGACCCGAACTGTAG	Chr22:+200073741	90	TGG
DGCR8 sgRNA 2	GGGGAGAACTTCCGGACCGC	Chr22:-200073792	91	AGG
HIRA sgRNA 1	TGTGTGCGGTGGTCAAACAG	Chr22:+19407246	93	TGG
HIRA sgRNA 2	CTGGCTAGCCTCATGCAGCG	Chr22:+19398036	93	TGG
ZDHHC8 sgRNA 1	CGTTCTTGTACAGCGGAGCC	Chr22:+20139514	91	CGG
ZDHHC8 sgRNA 2	TGAAAGGGTTCACACCCCCG	Chr22:+20140363	96	CGG

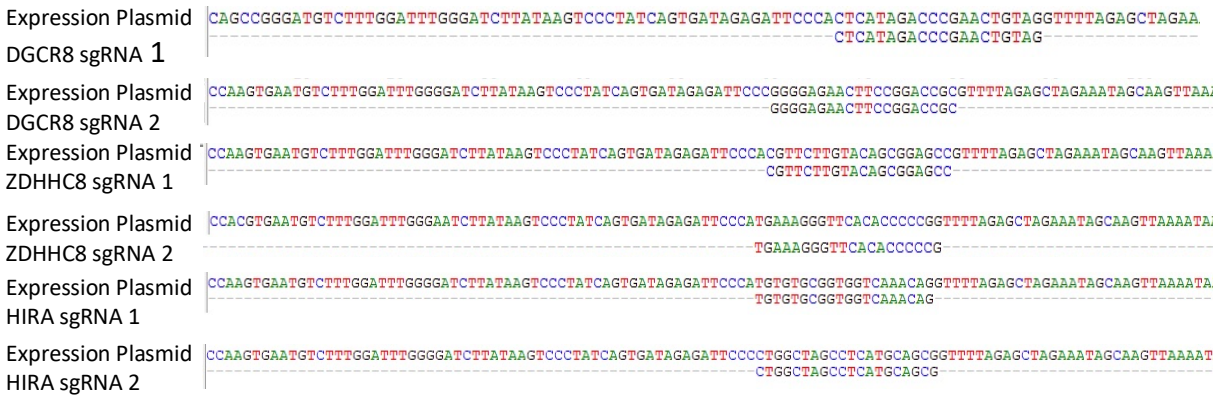
Insertion of sgRNAs into the FH1tUTG expression plasmid were confirmed by enzymatic digest. Loss of the BsmBI restriction site indicates insertion of the sgRNA (Figure 5.1A-B). The same Figure demonstrates that plasmid integrity was maintained after the cloning process, as digestion with NotI and XbaI shows the same digestion pattern as the empty plasmid. Positive clones identified by digestion were then sent for sequencing to confirm insertion of each sgRNA (Figure 5.1C). The individual expression plasmids were then used in conjunction with the 3rd generation lentiviral plasmids for lentivirus generation as used in Chapter four to target each gene in NPCs (Chapter Four, section 4.2.1).

Once each sgRNA-expressing lentivirus had been generated, the targeting efficiency of each sgRNA was assessed in order to use the most efficient sgRNA moving forward. Guide RNA targeting efficiency was determined using the online tool TIDE (https://tide.deskgen.com). NPCs were infected with each targeting lentivirus 24 hours after the second split of the differentiation (day 16/17). At day 20, NPCs were FAC sorted based on GFP and genomic DNA was extracted from the GFP positive population. For each gene, the targeted region was PCR-amplified and sanger-sequenced. The resulting sequencing data was inputted into the TIDE software. TIDE uses a decomposition algorithm to quantify the presence of InDels in a pool of edited cells compared to non-edited cells (Brinkman *et al.*, 2014). The analysis demonstrated which sgRNAs are more effective at cutting and targeting in a pool of infected cells. The more efficient sgRNAs were then to be used for subsequent targeting (Figure 5.1D).





C



D

Guide	Efficiency
DGCR8 sgRNA 1	12.4%
DGCR8 sgRNA 2	22.4%
HIRA sgRNA 1	48.6%
HIRA sgRNA 2	37.95%
ZDHCC8 sgRNA 1	17.6%
ZDHCC8 sgRNA 2	24.8%

Figure 5.1: Generation of lentiviral constructs targeting *DGCR8, HIRA* and *ZDHHC8* and their targeting efficiencies. Figure legend on next page.

Figure 5.1: Generation of lentiviral constructs targeting *DGCR8, HIRA* and *ZDHHC8* and their targeting efficiencies.
(A-B) *DNA gel electrophoresis showing 1) BsmBI and BamHI, 2) NotI and XbaI, 3) undigested enzymatic digestion patterns of FH1tUTG plasmid after cloning of sgRNAs targeting DGCR8, HIRA and ZDHHC8 compared to the empty FH1tUTG plasmid. Loss of BsmBI cut site confirming insertion.* ***(C)*** *Sanger sequencing confirming insertion of sgRNA 1 and sgRNA 2 for each target gene into FH1tUTG expression plasmid.* ***(D)*** *Table summarising the sgRNA targeting efficiency generated using TIDE analysis (https://tide.deskgen.com) for each targeted gene.*

5.2.2 Assessment of Lentiviral Transduction Efficiency

Before genetic manipulation using sgRNA-expressing lentiviruses by transducing NPCs, experimental lentiviral conditions had to be established. Firstly, it was important to determine the appropriate transduction titre. Secondly, it had to be verified if GFP expression was maintained throughout differentiation. Additionally, it had to be established if the lentivirus itself had any phenotypic effects on NPCs without sgRNA expression.

To assess transduction efficiency of the lentivirus, iCas9-derived NPCs were transduced with differing quantities of the lentivirus harbouring the FH1tUTG expression plasmid without a guide RNA (referred to as the empty lentivirus). Transduced cells were infected with either 0.75 or 1.5 million lentiviral particles (LVP) 24 hours after the second split of differentiation (day 16/17). Flow cytometry was carried out on day 20, 25 and 32 to determine percentage of GFP^+ cells (Figure 5.2). There was a marked increase in percentage of GFP^+ cells when a higher number of viral particles was used, with transduction using 1.5M LVP showing a much higher percentage of GFP^+ cells at each time point compared to 0.75M LVP (Figure 5.3). At day 20, 0.75M LVP had 8.6% GFP^+ cells, compared to 40.2% with 1.5M LVP. At day 25 there was 7.7% and 35% GFP^+ cells respectfully and at day 32 4.7% and 20% GFP^+ cells. For both concentrations, there was a reduction in GFP observed as differentiation progressed (Figure 5.3).

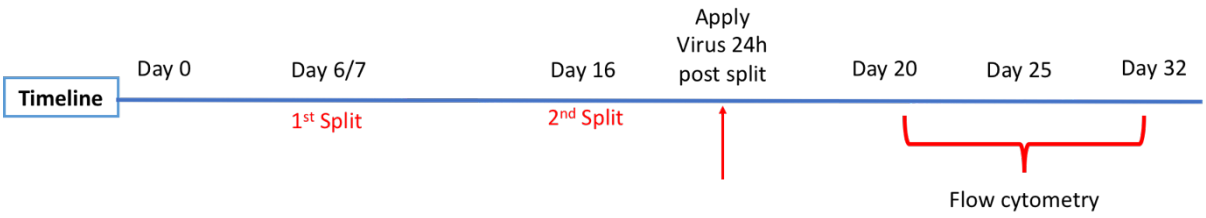


Figure 5.2: Schematic illustrating timeline for lentiviral transduction and flow cytometry
iCas9 cells are differentiated to cortical fate following established protocol, NPCs are transduced 24 hours after the second split of the differentiation and flow cytometry was performed on culture days 20, 25 and 32 to assess transduction efficiency by establishing percentage of GFP positive cells.

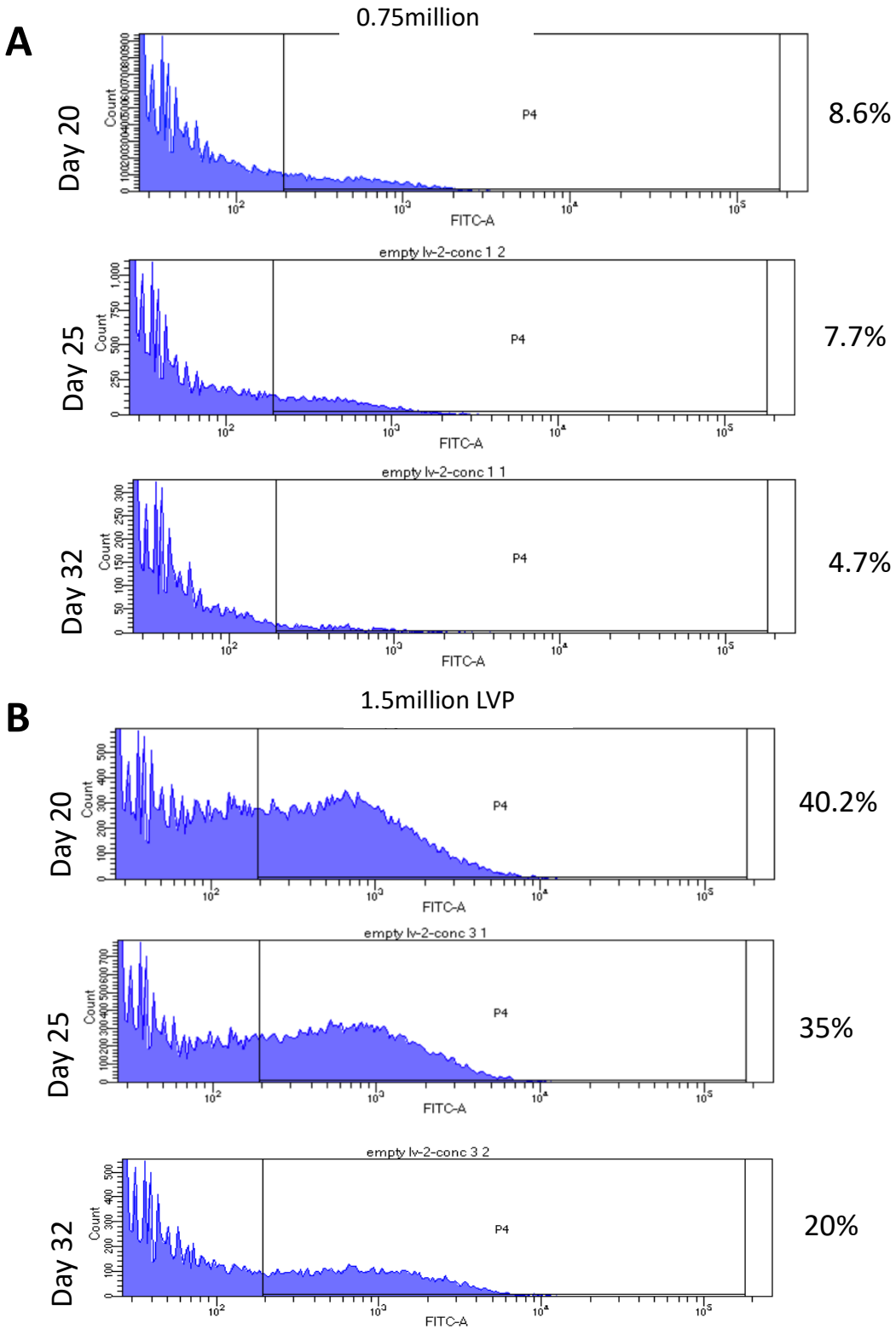


Figure 5.3: Flow cytometry analysis quantifying GFP+ cells in transduced neuroprogenitors with empty lentivirus

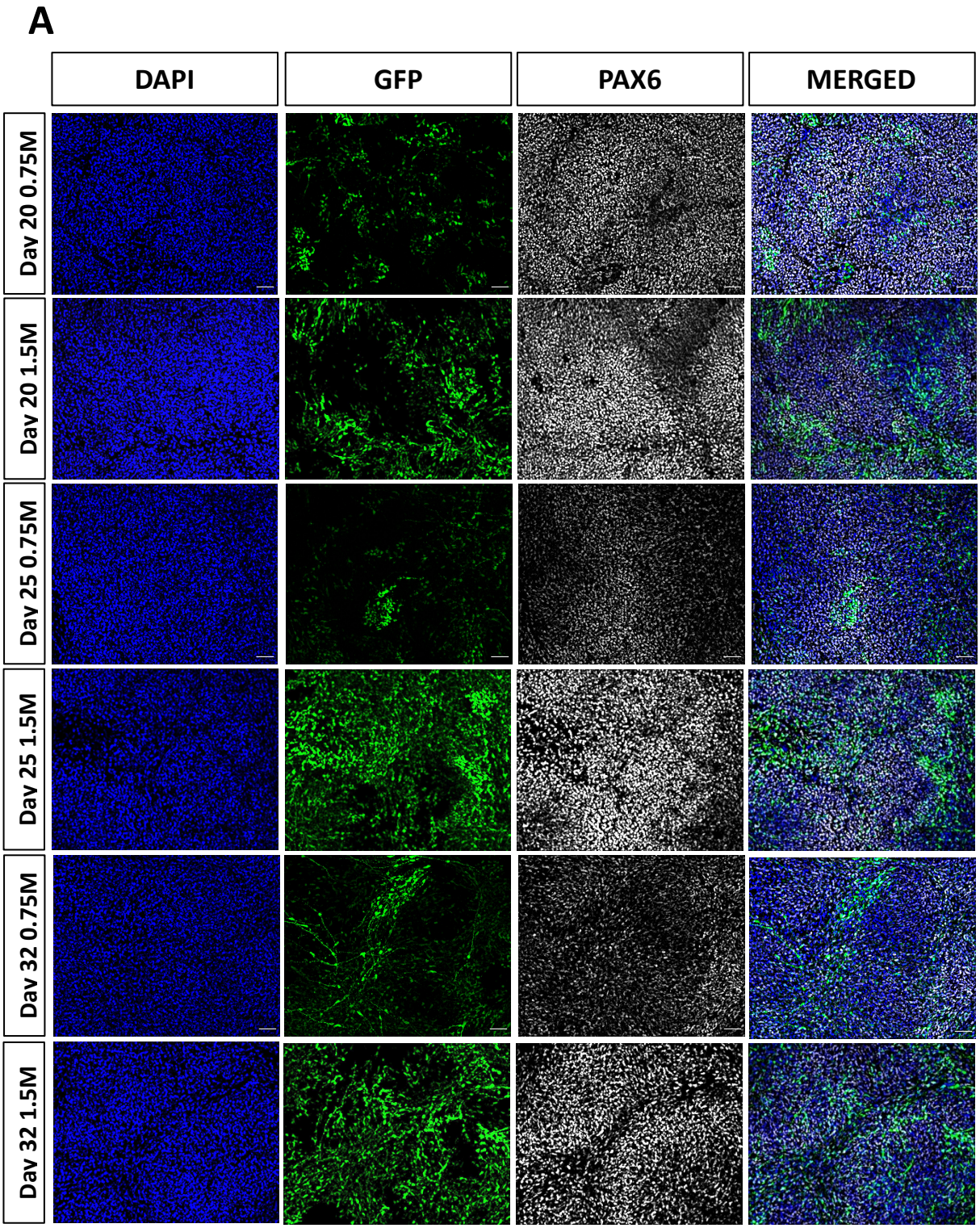
(A) *Flow cytometry histograms from iCas9-derived NPCs with transduced 0.75million viral particles of empty lentivirus quantifying percentage of GFP+ cells at day 20, 25 and 32.* ***(B)*** *Flow cytometry histograms from iCas9 NPCs with transduced 1.5million viral particles of empty lentivirus quantifying percentage of GFP+ cells at day 20, 25 and 32.*

5.2.3 Effects of empty lentivirus transduction on cortical progenitors

In order to investigate the effect of the lentivirus (without gene editing) on cortical progenitors, iCas9 cells were differentiated into NPCs and transduced with either 0.75M LVP or 1.5M LVP of the empty lentivirus 24 hours after the second split of the differentiation. The cultures were then fixed at day 20, 25 and 32 and various cortical markers were examined using immunofluorescent staining. This was used to determine whether there are any phenotypic differences between transduced and non-transduced cells in cortical differentiation in order to establish an appropriate control for edited cells as was considered for hESC genetic manipulation by CNV analysis (Chapter 2, section 2.3.7).

PAX6 is a transcription factor strongly expressed in dorsal forebrain progenitors, PAX6 expression is required for radial glial cell development and neural migration (Warren *et al.*, 1999). The percentage of PAX6 positive cells within the transduced (GFP^+) cells was compared to the percentage of PAX6 positive cells within the non-transduced (GFP^-) cells within the same well (referred to as $PAX6^+/GPF^+$ and $PAX6^+/GFP^-$ respectively). This was to assess that transduced cells followed the same expected cortical progression as non-transduced cells. At days 20 and 25, no significant difference was observed in percentage of PAX6 positive cells between the GFP^+ and GFP^- cells for NPCs transduced with either 0.75M or 1.5M LVP (Figure 5.4A-B). At day 32, cultures transduced with 0.75M LVP had a significant difference between the proportion of $PAX6^+/GFP^+$ and $PAX6^+/GFP^-$ cells. The GFP^+ NPCs show a higher percentage of PAX6 positive cells compared to non-transduced cells within the culture ($p = 0.014$). The same trend was observed for NPCs transduced with 1.5M LVP, but it was not significantly different. This could indicate a potential delay in maturation of transduced cells, as there is retained PAX6 expression in a higher proportion of transduced cells compared to non-transduced.

Notably across the three time points, there was no significant difference observed in percentage of $PAX6^+$ cells between GFP^+ cells transduced with 0.75M or 1.5M LVP, indicating increasing the number of viral particles does not enhance the effects (Figure 5.4C).


A
DAPI
GFP
PAX6
MERGED
Dav 20 0.75M
Dav 20 1.5M
Dav 25 0.75M
Dav 25 1.5M
Dav 32 0.75M
Dav 32 1.5M

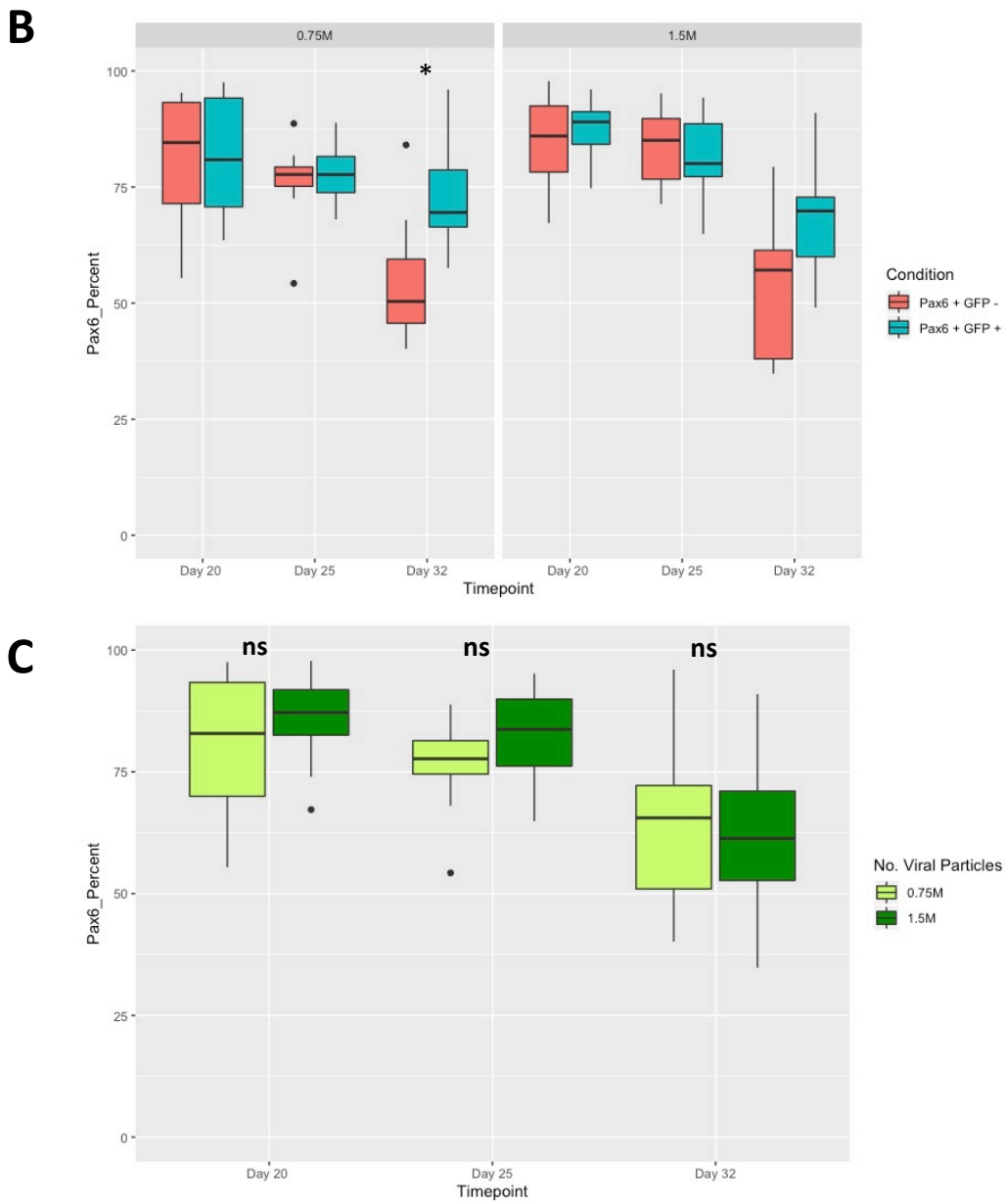


Figure 5.4: Proportion of PAX6 positive neuroprogenitors infected with 0.75M or 1.5M viral particles.

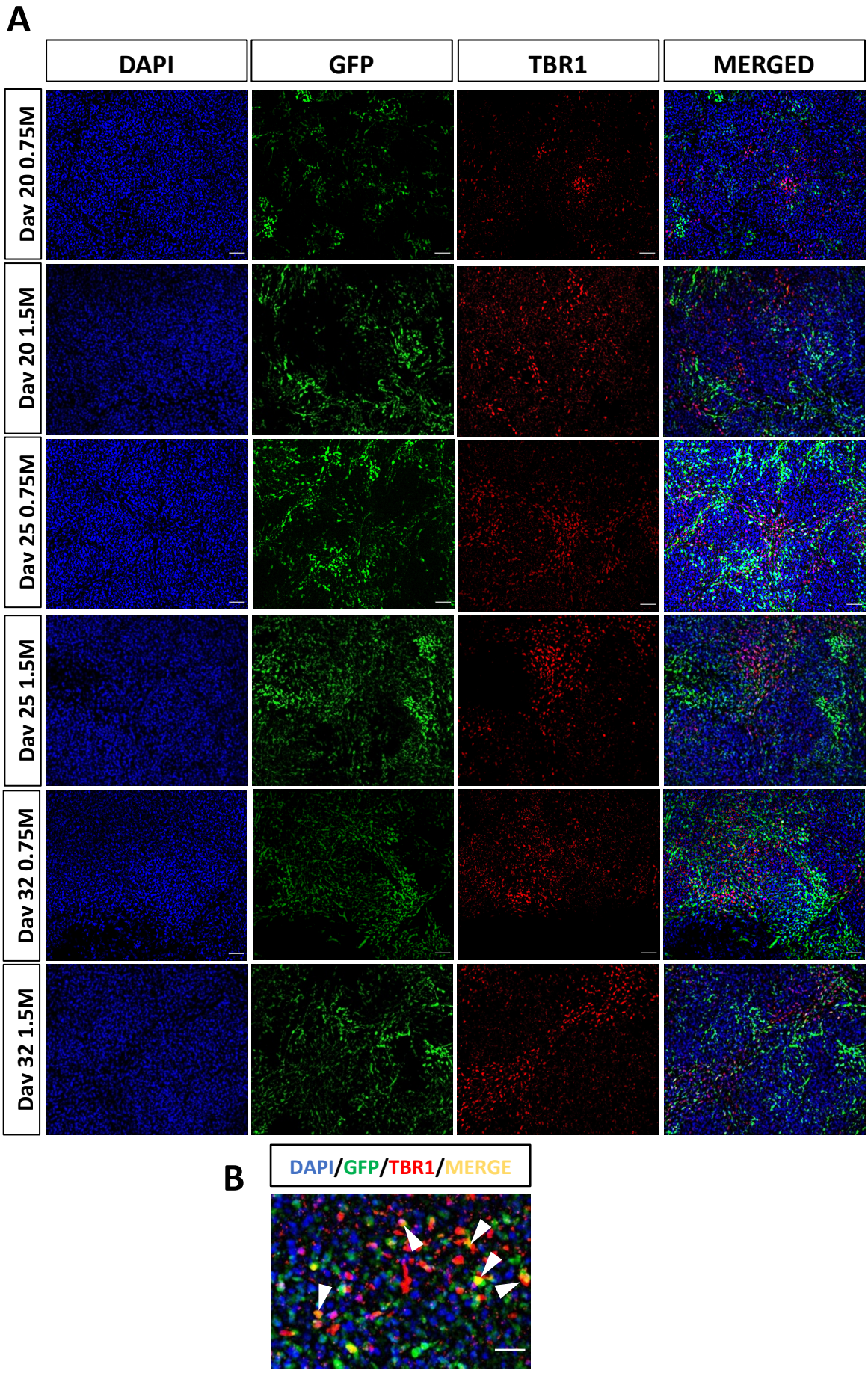
(A) *Immunofluorescent staining of GFP and PAX6 in iCas9 neuroprogenitors transduced with either 0.75M and 1.5M empty lentiviral particles at days 20, 25 and 32.* ***(B)*** *Quantification of PAX6+/GFP+ and PAX6+/GFP- in neuroprogenitors transduced with 0.75M or 1.5M viral particles at days 20,25 and 32. The percentage of PAX6+/GFP+ and PAX6+/GFP- were compared using one-way ANOVA (*p<0.05, **p<0.01,***p<0.001).* ***(C)*** *Comparison of PAX6+/GFP+ between NPCs transduced with 0.75M or 1.5M viral particles, compared using compared using one-way ANOVA. Data taken from 1 differentiation with three technical replicates.*

Considering that PAX6 was retained in a higher number of transduced cells, deep layer cortical markers were examined to see if the expected sequential appearance in cortical development was disrupted by lentiviral transduction. In order to assess this, the number of TBR1^{+} cells were examined. TBR1 is a transcription factor expressed by layer

VI cortico-thalamic project neurons and is important in regulating neuronal migration (Bedogni *et al.*, 2010; McKenna *et al.*, 2011).

As with PAX6, the percentage of TBR1 positive cells within the transduced (GFP^+) or non-transduced (GFP^-) cells within the same culture were examined ($TBR1^+/GFP^+$ and $TBR1^+/GFP^-$ respectfully). At day 20, there were no significance difference between TBR1+/GFP+ and $TBR1^+/GFP^-$ cells for either titre. Surprisingly, a significant increase was observed between in the percentage of $TBR1^+$ cells in the transduced GFP^+ cells compared to the GFP^- cells in cultures transduced with 0.75M LVP at day 25 and day 32 ($p= 0.001$ and 0.00048 respectively). A significant increase in the number of $TBR1^+/GFP^+$ compared to the $TBR1^+/GFP^-$ cells was also observed at day 25 for NPCs transduced with 1.5M LVP and the same trend was observed at day 32, but it was not significant (Figure 5.5C). This elevated TBR1 in transduced cells compared to non-transduced cells at day 20 and 25 is contradictory to higher retained PAX6 at day 32 in transduced cells. This could indicate that transduction may have different effects within the same culture, as this is a heterogenous population of cells, with variation in stage of differentiation and cell cycle. Lentiviruses are capable of infecting both dividing and non-dividing cells (Jandial *et al.*, 2008). Therefore, the higher proportion of TBR1 transduced cells compared to TBR1 in non-transduced cells could indicate that the lentivirus is more efficient at transducing more mature neuroprogenitors. The percentage of $TBR1^+/GFP^+$ cells did not differ at day 20 and day 25 between the cultures transduced with 0.75M LVP or 1.5M LVP, indicating the number of viral particles is not having an effect at these time points. However, there was a significant difference between the two conditions at day 32, with cultures infected with 0.75M LVP showing increased percent of $TBR1^+/GFP^+$ compared to the 1.5M LVP transduced cultures ($p= 0.02$) (Figure 5.5D).

The overall percentage of TBR1 was quantified in the entire culture, not distinguishing between GFP positive or negative cells, no significant difference was observed between cultures transduced with the two viral titres. Furthermore, no significant difference was observed between any of the conditions when compared to TBR1 expression in a non-transduced culture. This indicates that the presence of the lentivirus is not directly increasing the number of $TBR1^+$ cells, but potentially transducing TBR1 positive cells (and therefore intermediate progenitor cells) more effectively than radial cells (Figure 5.6).


A
DAPI
GFP
TBR1
MERGED
Dav 20 0.75M
Dav 20 1.5M
Dav 25 0.75M
Dav 25 1.5M
Dav 32 0.75M
Dav 32 1.5M
B
DAPI/GFP/TBR1/MERGE

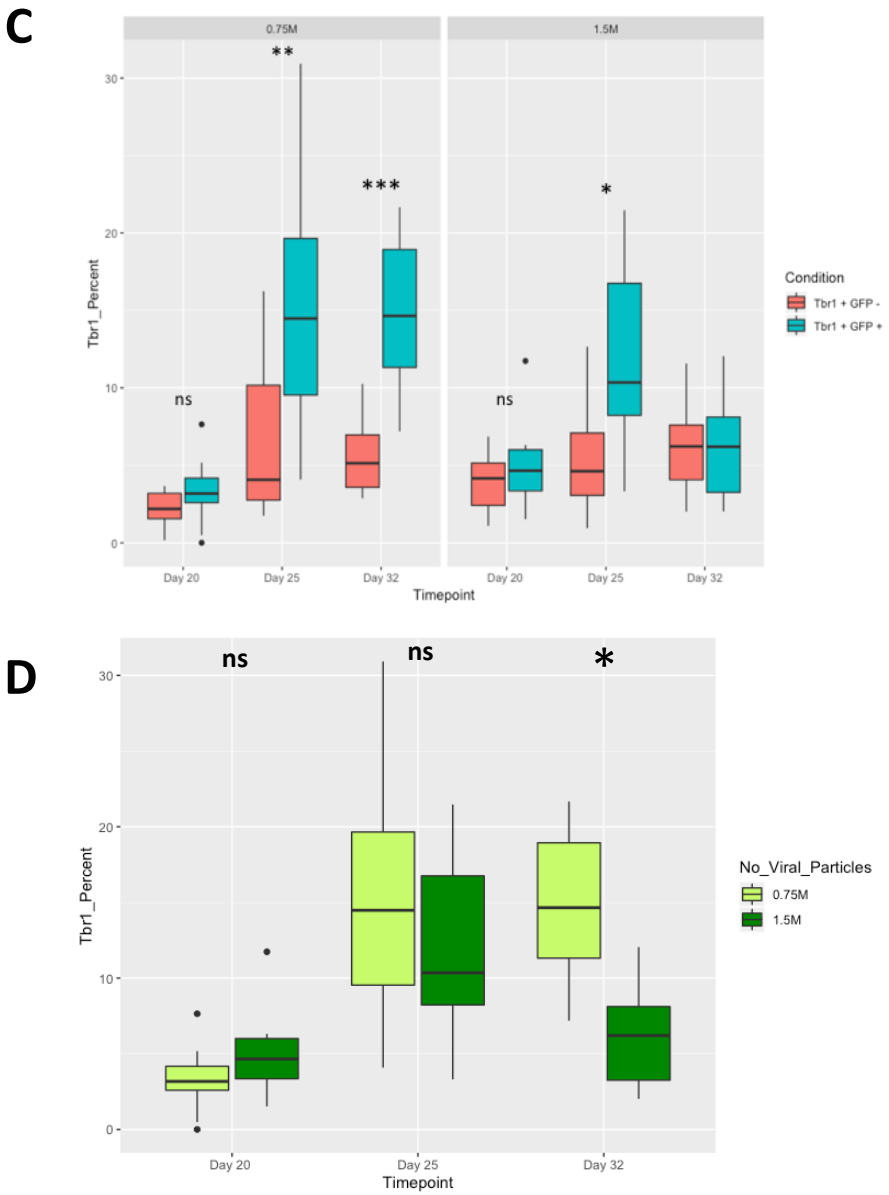


Figure 5.5: TBR1 positive neuroprogenitors infected with 0.75M or 1.5M viral particles.
(A) *Immunofluorescent staining of GFP and TBR1 in iCas9 neuroprogenitors transduced with empty lentivirus at days 20, 25 and 32.* ***(B)*** *Immunostaining showing example of TBR1+/GFP+ cells, shown by arrows.* ***(C)*** *Quantification of TBR1+/GFP+ and TBR1+/GFP- in neuroprogenitors transduced with 0.75M or 1.5M viral particles at days 20,25 and 32. The percentage of TBR1+/GFP+ and TBR1+/GFP- were compared using one-way ANOVA (*p<0.05, **p<0.01,***p<0.001).* ***(D)*** *Comparison of TBR1+/GFP+ between NPCs transduced with 0.75M or 1.5M viral particles, compared using compared using one-way ANOVA. Data taken from 1 differentiation with three technical replicates.*

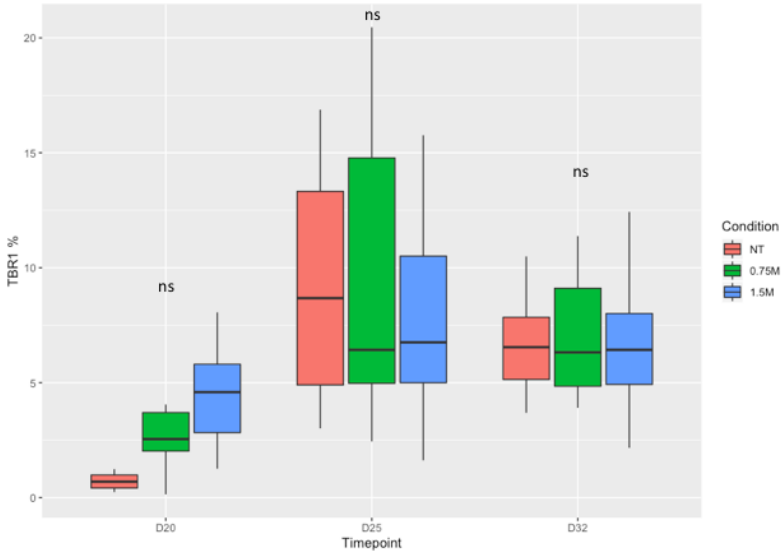
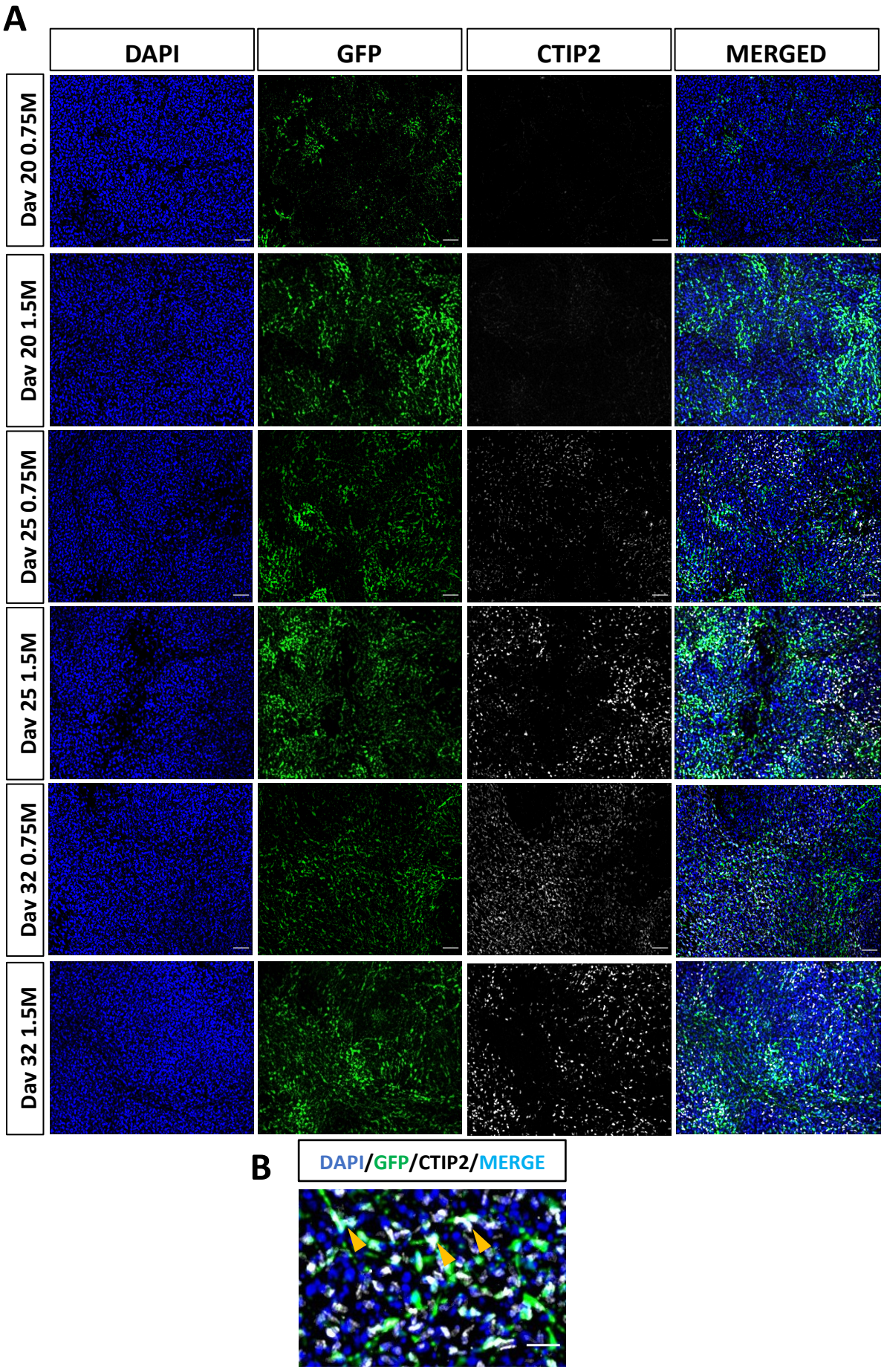


Figure 5.6: TBR1 positive cells in transduced and non-transduced neuroprogenitors
Quantification of overall TBR1 staining in the entire cultures transduced with 0.75M, 1.5M lentiviral particles or non-transduced (NT). Showing no significant difference between each condition. TBR1 percentage was compared using one-way ANOVA. Data taken from 1 differentiation with three technical replicates.

CTIP2 is a typical marker of layer V neurons and is important in axonal targeting of cortical projection neurons (Chen *et al.*, 2008). CTIP2 positive cells appear later than TBR1 and so monitoring their prevalence can indicate the progression of the differentiation (McKenna *et al.*, 2011). The proportion CTIP2$^+$ cells were examined within transduced GFP$^+$ cells and compared to the proportion of CTIP2$^+$ non-transduced GFP$^-$ cells within the same culture. The analysis revealed a consistent increase in the percentage of CTIP2$^+$ cells in the GFP$^+$ cells compared to the GFP$^-$ NPCs (Figure 5.7). For NPCs infected 0.75M LVP, at day 25 and day 32 there was a significantly increased percentage of CTIP2 in the transduced cells compared to non-transduced cells ($p = 3.77^{-9}$ and 9.91^{-13} respectively). The same significant trend was observed for the cultures transduced with 1.5M LVP, with significant differences observed at day 25 and 32 ($p = 1.53^{-7}$ and 4.8^{-7} respectively). No differences were observed at day 20 for either condition. This increase in number of CTIP2$^+$ transduced cells compared to non-transduced cells is concurrent with the previous finding of increased TBR1$^+$ cells. There is a consistent finding that a higher proportion of transduced cells express deep layer markers compared to non-transduced cells, which could indicate that the lentivirus is more effective at infecting the intermediate progenitors within the cell population.

There were no significant differences observed between any time point in the percentage of $CTIP2^+/GFP^+$ cells for either transduction condition, indicating different virus transduction titre do not have differing effects on CTIP2 expression (Figure 5.7). Furthermore, there was no difference in overall CTIP2 between the transduced conditions compared to non-transduced. Therefore, the lentivirus is not increasing the amount of $CTIP2^+$ cells in the culture, but more transduced cells are $CTIP2^+$ (Figure 5.7).


A
DAPI
GFP
CTIP2
MERGED
Dav 20 0.75M
Dav 20 1.5M
Dav 25 0.75M
Dav 25 1.5M
Dav 32 0.75M
Dav 32 1.5M
B
DAPI/GFP/CTIP2/MERGE

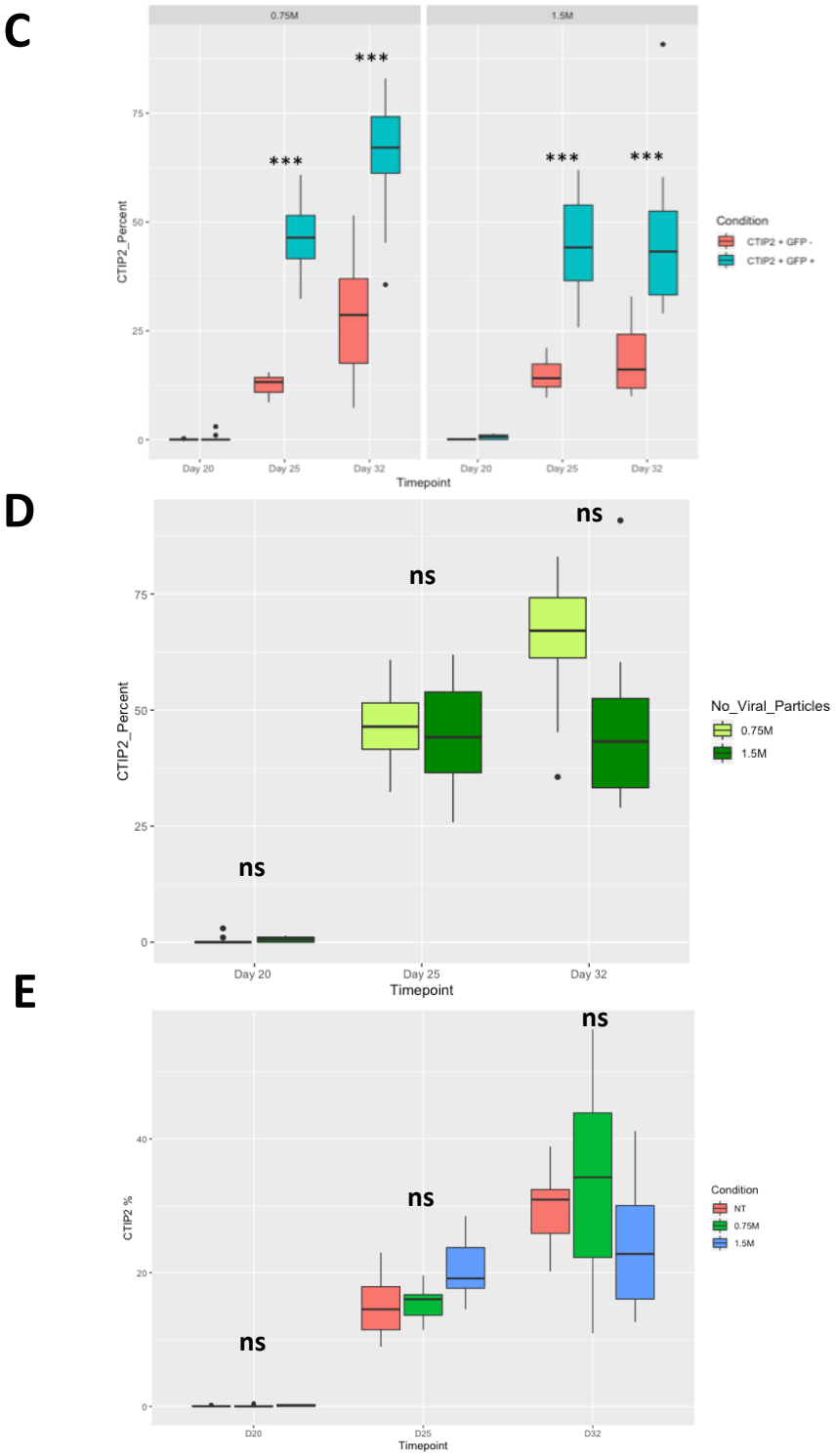
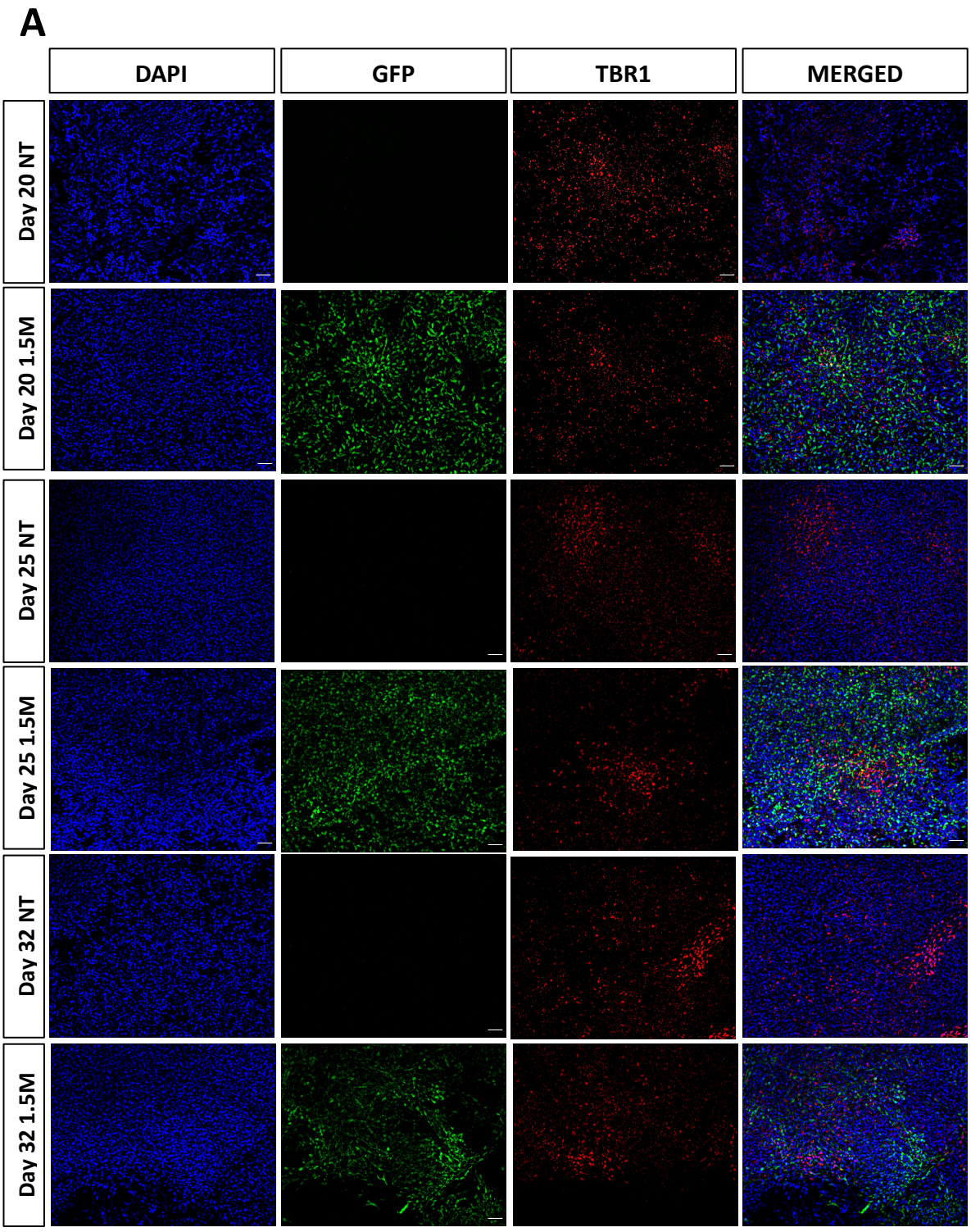


Figure 5.7: CTIP2 positive cells in GFP+ and GFP- neuroprogenitors infected with 0.75M or 1.5M viral particles. ***(A)*** *Example of immunofluorescent staining of GFP and CTIP2 in iCas9 neuroprogenitors transduced with empty lentivirus at days 20, 25 and 32.* ***(B)*** *Immunostaining of CTIP2+/GFP+ overlap, indicated by arrows.* ***(C)*** *Quantification of CTIP2+/GFP+ and CTIP2+/GFP- in neuroprogenitors transduced with 0.75M or 1.5M viral particles at days 20,25 and 32. The percentage of CTIP2+/GFP+ and CTIP2+/GFP- were compared using one-way ANOVA (*p<0.05, **p<0.01,***p<0.001).* ***(D)*** *Quantification of CTIP2 from immunostaining of CTIP2+/GFP+ in iCas9 neuroprogenitors transduced with 0.75M or 1.5M empty lentivirus particles at days 20, 25 and 32. The percentage of positive cells in both conditions were compared using one-way ANOVA.* ***(E)*** *Quantification of overall CTIP2 from immunostainings of non-transduced, transduced with 0.75M and 1.5M lentiviral particles. The percentage of positive cells in both conditions were compared using one-way ANOVA.*

Use of 1.5M lentiviral particles for transduction was chosen as the appropriate transduction concentration for further experiments, due to the increased transduction efficiency and higher observed levels of GFP, which was preferential for increased gene editing. Furthermore, the increased viral titre did not exaggerate the proportion of TBR1 and CTIP2 positive cells compared to the 0.75M viral particles. As shown with the comparisons to 0.75M viral particles, replications of the experiments consistently showed increased levels of $TBR1^{+}/GFP^{+}$ compared to $TBR1^{+}/GFP^{-}$, with significant differences observed at day 20, 25 and 32 ($p= 0.001$, 1.54^{-6} and 0.01 respectively) but again no difference observed in overall TRB1 levels compared to non-transduced cultures (Figure 5.8C).

Significant differences were again observed between $CTIP2^{+}/GFP^{+}$ and $CTIP2^{+}/GFP^{-}$, with increased numbers of $CTIP2^{+}$ cells in GFP^{+} cells compared to the GFP^{-} cells at day 25 and day 32 ($p=6.148^{-8}$ and $p=6.1^{-13}$) shown in Figure 5.9B. Importantly, there was no significant difference in the overall CTIP2 levels in the transduced cultures compared to non-transduced at any time point. Crucially, there is also no difference at day 20, indicating that the virus does not lead to premature CTIP2 expression, as CTIP2 is not usually expressed at this earlier stage of differentiation (Chen *et al.*, 2008). Additionally, TBR1 and CTIP2 expression in all conditions appeared in the appropriate order as observed during *in vivo* corticogenesis and this pattern is maintained in *in vitro* differentiation (Shi *et al.*, 2012; Espuny-Camacho *et al.*, 2013).

Due to the consistent differences of transduced cells expressing higher levels of TBR1 and CTIP2 compared to non-transduced cells within the same culture, for further experiments with editing lentiviruses, edited cells would be compared to cells transduced with the empty lentivirus rather than non-transduced cells within the same well.


A
DAPI
GFP
TBR1
MERGED
Day 20 NT
Day 20 1.5M
Day 25 NT
Day 25 1.5M
Day 32 NT
Day 32 1.5M

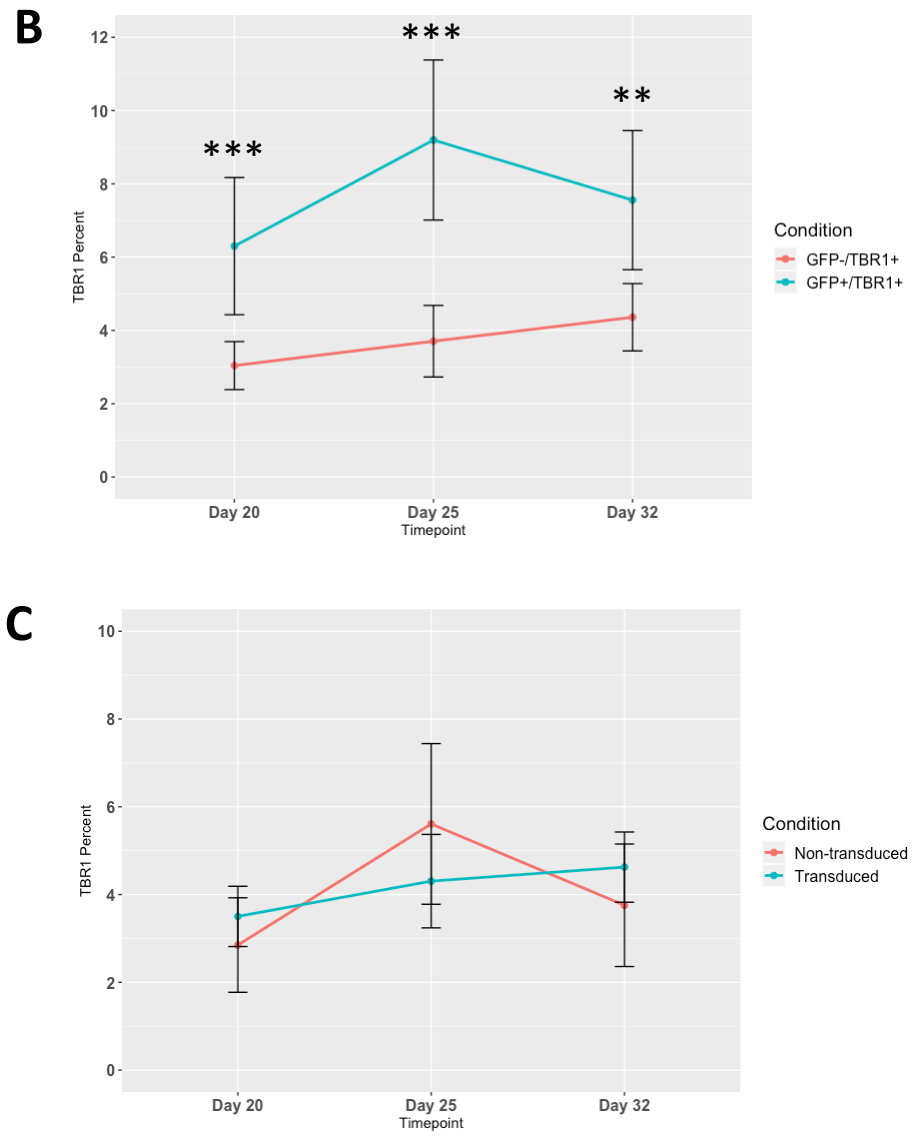


Figure 5.8: TBR1 positive neuroprogenitors transduced with 1.5M lentiviral particles or non-transduced ***(A)*** *Immunofluorescent staining of GFP and TBR1 in iCas9 neuroprogenitors transduced with empty lentivirus at days 20, 25 and 32 compared to non-transduced.* ***(B)*** *Comparison of TBR1+/GFP+ to TBR1+/GFP- percentage within the same cultures at days 20, 25 and 32. compared using one-way ANOVA followed by Tukey post-hoc test (*p<0.05, **p<0.01,***p<0.001). Values are expressed as mean ± SEM.* ***(C)*** *Comparison of overall TBR1 percentage in non-transduced cultures compared to transduced. Values are expressed as mean ± SEM. Data taken from 3 independent differentiations with two technical replicates.*

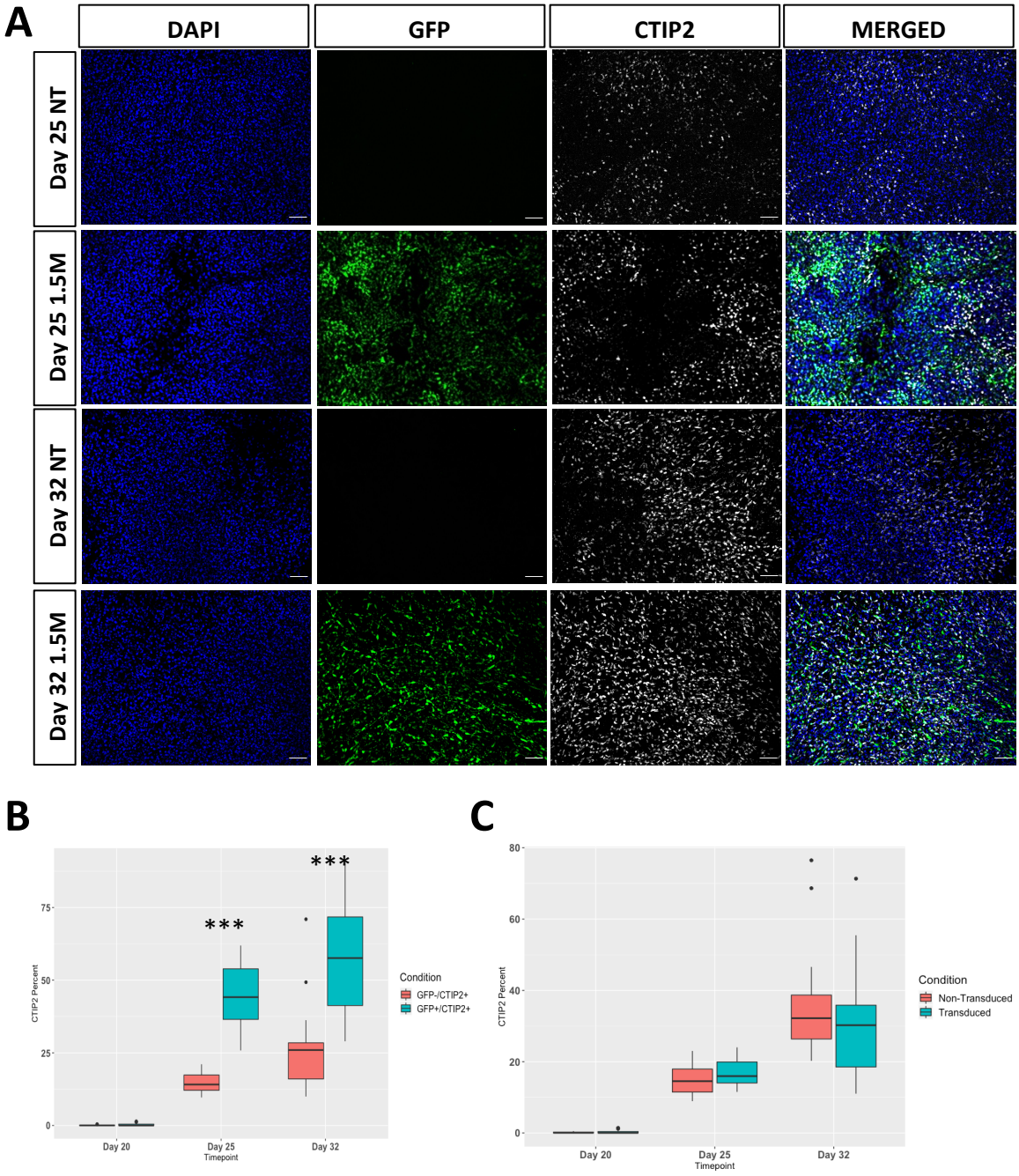


Figure 5.9: Percentages of CTIP2 positive neuroprogenitors transduced with 1.5M lentiviral particles or non-transduced ***(A)*** *Immunofluorescent staining of GFP and CTIP2 in iCas9 neuroprogenitors transduced with empty lentivirus at days 20, 25 and 32 compared to non-transduced.* ***(B)*** *Comparison of CTIP2+/GFP+ to CTIP2+/GFP- percentage within the same cultures at days 25 and 32, compared using one-way ANOVA followed by Tukey post-hoc test (*p<0.05, **p<0.01,***p<0.001).* ***(C)*** *Comparison of overall TBR1 percentage in non-transduced cultures compared to transduced. Data taken from 2 independent differentiations with two technical replicates.*

5.2.4 *DGCR8* gene editing in human neuroprogenitor cells

To investigate *DGCR8* loss in cortical neurogenesis, iCas9-derived neuroprogenitors were transduced using the lentiviral based CRISPR/Cas9 system expressing *DGCR8*-targeting sgRNAs in the presence of doxycycline.

Dgcr8 deletion in the developing mouse cortex has been shown to result in increased Tbr1$^+$ neurons, the suggested mechanism is due to Dgcr8 directly regulating the *Tbr1* mRNA transcript (Marinaro *et al.*, 2017), therefore it was selected as a marker to investigate in this model. The percentage of TBR1 positive cells was assessed in NPCs transduced with the control empty lentivirus or the lentivirus harbouring each of the sgRNA targeting *DGCR8* at day 25. The percentage of TBR1$^+$ cells within the GFP$^+$ population was significantly increased in *DGCR8*-manipulated NPCs compared to cells transduced with the empty virus (Figure 5.10A-B). TIDE analysis (as described in section 5.2.1) demonstrated that sgRNA 2 is the more efficient gRNA for editing. Additionally, sgRNA 2 transduced NPCs show a higher percentage of TBR1$^+$ cells compared to cells infected with sgRNA 1 or the empty lentivirus (Figure 5.10B). As sgRNA 2 had been previously shown to be the more efficient guide RNA, the effects of *DGCR8* editing on TBR1 was further explored using this system. The number of TBR1$^+$ cells were examined in the transduced cultures at day 20, 25 and 32 by immunofluorescent staining. At day 20, there was a trend of increased number of TBR1$^+$ cells in *DGCR8* targeted cells, but this was not significantly different compared to cells transduced with the empty lentivirus. A significantly increased difference was observed at day 25 between *DGCR8*-targeted cells and the empty-lentivirus transduced cells ($p=0.02$) but no difference was observed at day 32 (Figure 5.11). This significantly increased proportion of TBR1$^+$ cells in the genetically manipulated population at day 25 further reinforces the evidence for a relationship between DGCR8 and *TBR1* as described in Chapter 1 section 1.3.3.3 (Marinaro *et al.*, 2017). This has not been observed in a human cell model before and further indicates a role for DGCR8 in neurodevelopment. However, this difference is not observed at day 32, which could indicate a key temporal window in which DGCR8 acts during cortical development, which is at this timepoint in cell culture conditions. Equally, the percentage of GFP positive cells does decrease by day 32 (shown in flow cytometry Figure 5.3) and this could be due to the presence of fewer transduced cells, hence why the difference in TBR1$^+$ cells is lost by day 32. It would have been beneficial to examine

DGCR8 protein levels in transduced populations to determine if protein levels are reduced using this method and therefore if this is important in the mechanism for inducing *DGCR8* related phenotypes.

Overall this data indicates this lentiviral-based *DGCR8* manipulation is an effective method for gene editing in this model and can be used to evaluate phenotypic differences. Importantly, this work further reinforces a role for DGCR8 in regulation of $TBR1^{+}$ neurons during a specific window in corticogenesis.

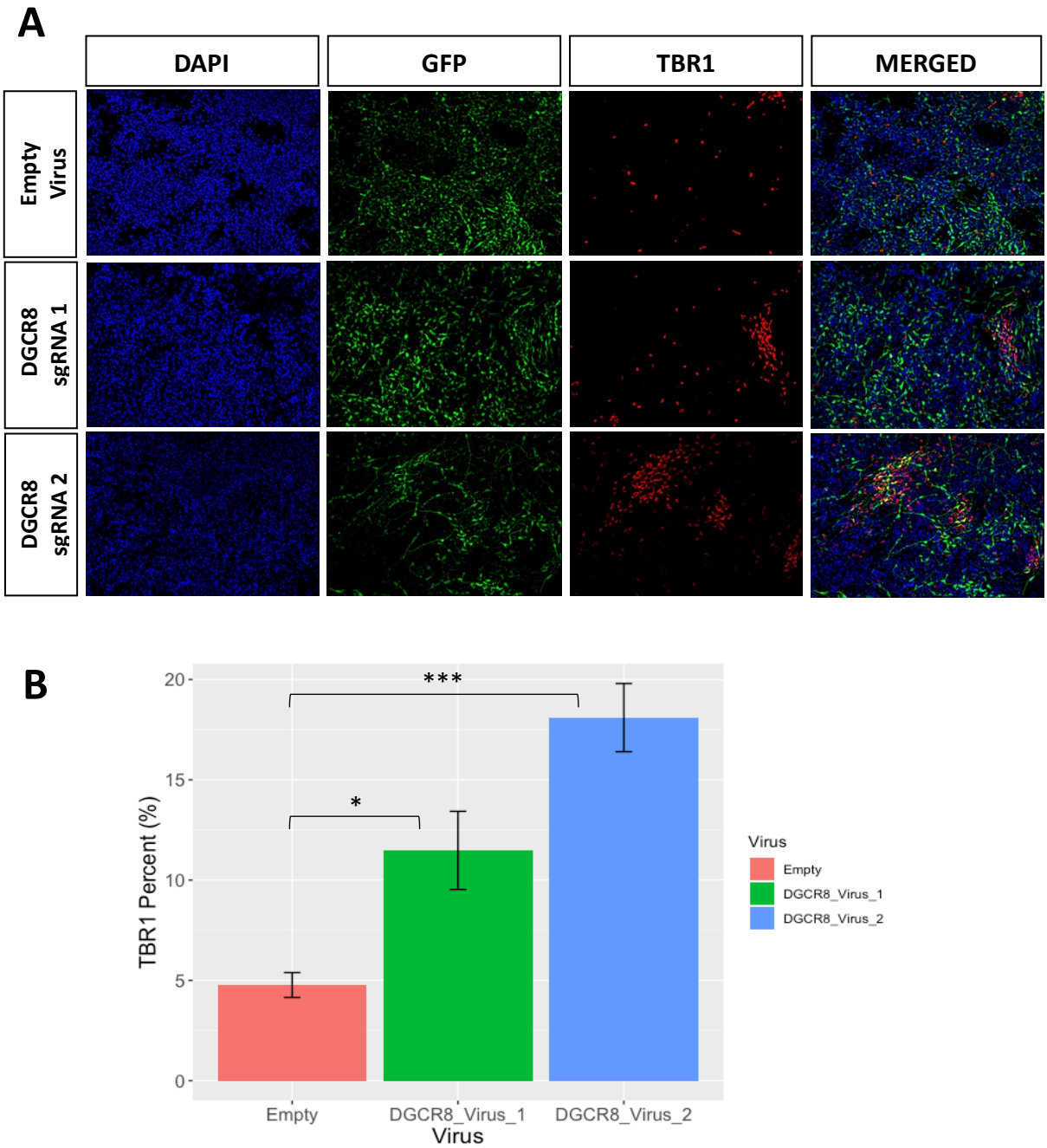
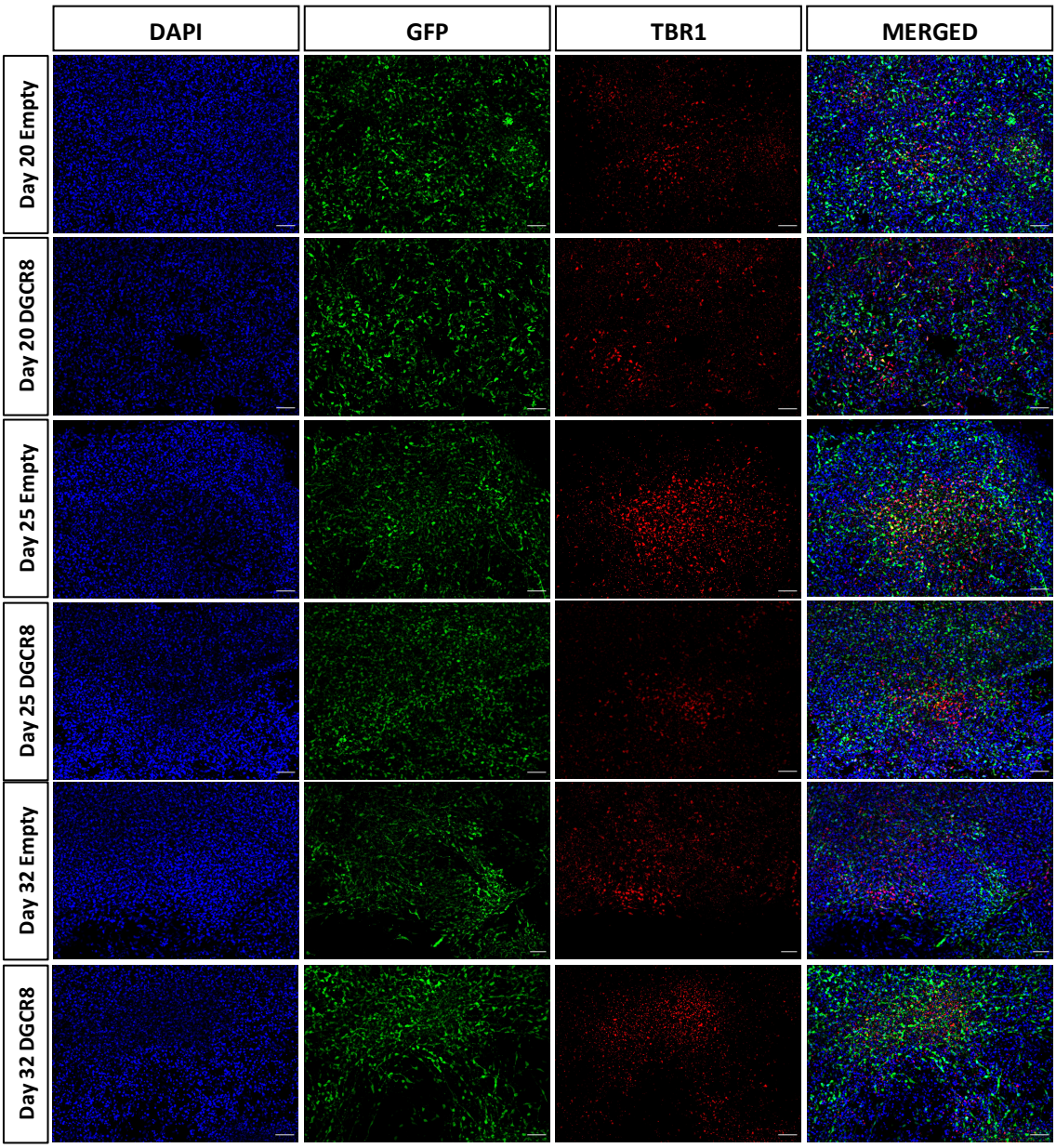


Figure 5.10: TBR1 expression in neuroprogenitors transduced with *DGCR8* editing lentiviruses

(A) *Immunofluorescence staining for TBR1 (red) and GFP (green) at day 25 iCas9 cells transduced with the empty lentivirus or lentivirus expression DGCR8 sgRNA1 or sgRNA 2. Nuclei are stained with Dapi (blue).* ***(B)*** *Quantification of TBR1+ cells in NPCs transduced with empty, sgRNA1-expressing or sgRNA2-expressing lentivirus. Data taken from 1 differentiation with 2 technical replicates. Data shown as mean ± SEM. Data compared using one-way ANOVA followed by post-hoc Tukey test (*p<0.05, **p<0.01, ***p<0.001).*

A


DAPI
GFP
TBR1
MERGED
Day 20 Empty
Day 20 DGCR8
Day 25 Empty
Day 25 DGCR8
Day 32 Empty
Day 32 DGCR8

B

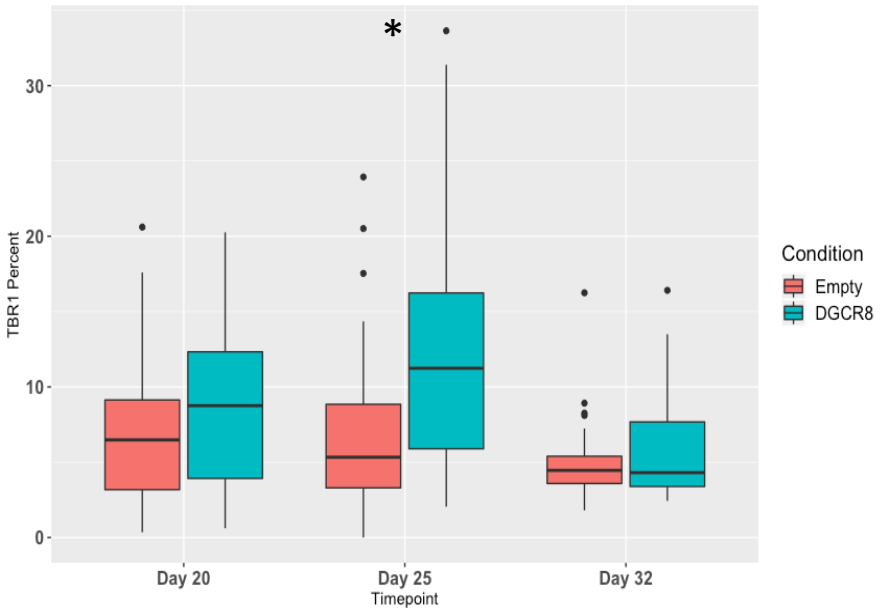


Figure 5.11: TBR1 expression in NPCs transduced with sgRNA-2 *DGCR8*-targeting lentivirus compared to empty lentivirus.
(A) *Immunofluorescence staining for TBR1 (red) and GFP (green) in iCas9 cells transduced with the empty lentivirus or lentivirus expressing DGCR8-targeting sgRNA 2 at days 20, 25 and 32. Nuclei are stained with DAPI (blue). Scale bars = 200μm.* ***(B)*** *Quantification of percentage of GFP+/TBR1+ cells in transduced with empty or sgRNA 2 DGCR8-targeting lentivirus at day 20, 25 and 32. Time points were compared by t-test (*p<0.05). Data taken from 2 independent differentiation with 2 technical replicates.*

5.3 Discussion

This chapter describes the use of a lentiviral-based CRISPR/Cas9 system for manipulation of 22q11.2DS SCZ candidate genes in iCas9-derived NPCs. The effects of the lentivirus in cortical differentiation were widely explored before investigating gene editing potential, followed by specific investigation into *DGCR8* manipulation during cortical differentiation.

DGCR8, *HIRA* and *ZDHHC8* had been selected as SCZ candidate genes to explore based on the work described in Chapter 3. Generation of mutant hESC lines can be a long and laborious process, with the need to for clonal isolation and expansion before extensive genetic screening to confirm targeting. Furthermore, some genes are essential for cell replication and survival and therefore may render the cells non-viable. It can be advantageous to investigate gene loss temporally or once cells have committed to a cell fate. The use of a lentiviral CRISPR/Cas9 methods allows for genetic manipulation during

a specific developmental window, without affecting the developmental process prior to this stage. Furthermore, lentiviral-based delivery has been shown to be efficient in terminally differentiated neurons, therefore allowing investigation of developmental windows relevant to SCZ development (Naldini *et al.*, 1996).

Lentiviral-based delivery of sgRNAs allows for stable expression due to the integration of the lentivirus into the host genome. Utilising this approach in the iCas9 cell line meant that there was tight regulation of the expression of sgRNA and the Cas9 nuclease, as both are controlled by doxycycline. Aubrey and colleagues, demonstrated that transduction of both an inducible Cas9 expression vector and sgRNA expression vector did not result in as efficient editing, therefore it was favourable to use the iCas9 cell line, rather transducing the Cas9 nuclease and the sgRNA separately (Aubrey *et al.*, 2015). The TIDE analysis demonstrated that the sgRNAs have varying targeting efficacies across the different candidate genes, but this could be due to efficiencies of the sgRNAs and also the genetic regions targeted. Using a strategy integrating multiple sgRNAs targeting individual genes may result in better targeting (Kabadi *et al.*, 2014).

Before investigating the effects of gene editing in the NPCs, it was important to determine the appropriate control for the transduced manipulated cells. The transduced cells would have an integrated lentivirus in their genome and also constitutively express GFP, therefore it had to be determined if this would have a phenotypic impact during differentiation, before any editing takes place. The markers PAX6, TBR1 and CTIP2 were analysed as these have known specific sequential and temporal expression patterns in cortical differentiation and therefore disruption of this could impact differentiation (Sadegh and Macklis, 2014).

It was consistently observed that with varying titres, transduced cells had a higher proportion of $TBR1^+$ and $CTIP2^+$ cells compared to non-transduced cells within the same culture. The higher proportion of $TBR1^+$ cells in the transduced population is observed from earliest timepoint, day 20 and is maintained to day 32. TBR1 is expressed by all postmitotic pyramidal neurons generated in the developing pallium (Englund *et al.*, 2005). Lentiviruses have the ability to infect both dividing and non-dividing cells. My data could indicate lentiviruses have a higher efficiency for infecting NPCs at different

stages of cortical development, particularly different efficiencies between radial glial cells and intermediate progenitors and therefore dividing and non-dividing cells. The increased proportion of $CTIP2^+$ cells in transduced cells is observed at day 25 and 32, but importantly no difference was observed at day 20, a timepoint where CTIP2 expression is not expected. Furthermore, as observed with TBR1, the overall levels of CTIP2 were not higher in transduced cultures compared to non-transduced. This combined with the no observed difference at day 20 would indicate that the lentivirus is not leading to maturation of transduced cells, but further points to enhanced efficiency for infecting more mature cells, but this has not been reported before in the literature.

Conversely, although PAX6 levels in transduced cells were not different at days 20 and 25, at day 32 a higher proportion of transduced cells were $PAX6^+$ compared to GFP^- cells in the same culture. PAX6 is highly expressed in the cortical proliferative zone (Estivill-Torrus *et al.*, 2002). This difference could indicate different effects of the lentivirus depending on the dividing state of the cell it infects, which could be examined by flow-cytometry examining levels of DAPI. Although, the lentiviral infected cells show increased proportions of TBR1 and CTIP2 positive cells, these were observed at the expected time points for their expression during differentiation. However, the retained PAX6 expression at day 32 could indicate the lentivirus is affecting cell proliferation, if at the point of transduction, the cells are dividing, and this has a longer-term effect. Transduced neurospheres with a recombinant adenovirus expressing GFP was demonstrated to affect proliferation human neuroprogenitor cell proliferation (Wu, Ye and Svendsen, 2002). However, *in vitro* transduction of NSCs with adenoviral vectors may lead to differentiation, which has not been reported with other viral vectors (Hughes *et al.*, 2002). Furthermore, it could be relevant to investigate if it is the GFP expression affecting neuronal differentiation, by using different delivery methods such a transfection and assessing the cortical markers. Puttonen and colleagues found that hNPCs transduced with a lentivirus encoding GFP reduced the neuronal characteristics of the cells, with reduction in PAX6, DCX and MAP-2 expression in transduced cells. They found this was due to the GFP, not the lentivirus itself, this effect was based on the size of the hNPCs, with smaller NPCs being more sensitive to GFP (Puttonen *et al.*, 2013). At the point of transduction, there could be a heterogenous population of dividing and non-dividing cells, therefore the lentivirus could have different effects depending on the

cell-cycle state, which is something to consider for future work with lentiviruses, it may be advantageous to transduce cell cultures with a cell-cycle synchronised population.

Once the appropriate control of comparison to transduced cells with an empty lentivirus was decided, investigation of *DGCR8* manipulation in NPCs was performed. The generated lentiviral particles proved efficient at infecting and editing NPCs. This is reflected in the number of $TBR1^+$ cells in cells transduced with the empty, sgRNA 1 or sgRNA 2, with higher numbers of $TBR1^+$ cells observed in *DGCR8* manipulated NPCs. The TIDE analysis demonstrated that sgRNA 2 was the more efficient sgRNA and interestingly cells transduced with sgRNA 2 demonstrated an amplified response in terms of increased TBR1 levels. This work further reinforces the suggested regulatory mechanism between DGCR8 and *TBR1* (Marinaro *et al.*, 2017). The increased proportion of $TBR1^+$ cells was observed at day 25, but this had returned to levels comparable to wildtype by day 32. This data could indicate a specific temporal window in this model, in which DGCR8 regulates *TBR1* during cortical differentiation, therefore it would be interesting to investigate the *TBR1* mRNA levels at this time point to see if they are dysregulated. However, this could also be due to reduction in GFP levels, which was observed at day 32 when transduction efficiency was analysed. Further work is required to investigate the underlying mechanism. It would be beneficial to investigate levels of the sequential transcriptional cascade that characterises cortical neurogenesis, surrounding TBR1, such as PAX6 and TBR2 to further elucidate the role DGCR8 has in cortical differentiation. Further work of RNA sequencing was planned to investigate effects of gene editing of *DGCR8, HIRA* and *ZDHHC8* in NPCs, however the COVID-19 pandemic halted this work. The further planned work included CRISPR/Cas9 editing of NPCs individually targeting *DGCR8, HIRA* and *ZDHHC8* and extracting RNA from the GFP positive population by FAC sorting at days 21, 25 and 30 of cortical differentiation. RNA sequencing was then to be carried out to investigate if there are dysregulated transcriptomic pathways shared when each gene is manipulated through cortical differentiation.

In summary, this work indicates that lentiviral mediated gene manipulation is viable method of gene editing in NPCs. These results highlight that careful consideration is required for selecting the appropriate controls when using lentiviruses, as there were observed phenotypic differences between transduced cells and non-transduced cells.

These differences were due to lentiviral transduction prior to any gene editing taking place throughout cortical differentiation of NPCs. Finally, genetic manipulation of *DGCR8* in NPCs demonstrate an independent confirmation of a relationship between DGCR8 and TBR1, indicating an important role of DGCR8 in corticogenesis, however further work is required to confirm the underlying mechanism.

6. General Discussion

6.1 Summary and Implications of Work

This PhD thesis aimed to investigate potential genetic and phenotypic mechanisms underlying schizophrenia risk in 22q11.2DS. To achieve this, the initial objective was to identify schizophrenia candidate and disease modifier genes from within and outside of the 22q11.2 deletion region respectively. Identifying candidate genes from within the 22q11.2 region, alongside the primary selected gene *DGCR8,* was based upon fetal brain gene co-expression and literature review. To identify potential modifier genes, analysis of *cis*-heritable gene expression changes was performed by carrying out transcriptome wide association studies. The subsequent chapters aimed to investigate selected schizophrenia candidate genes from within the 22q11.2 deletion region, with a particular focus on *DGCR8* in hESCs and hNPCs in cortical differentiation. Multiple techniques and approaches were utilised across the project to fulfil these aims.

In the third chapter, the objective was to identify schizophrenia candidate genes from within and outside of the 22q11.2 deletion region. In order to select potential schizophrenia candidate genes from within the 22q11.2 deletion region alongside *DGCR8*, firstly gene co-expression of genes within the 22q11.2 deletion in human fetal brain RNA sequencing data was analysed. The hypothesis behind this is that genes with a shared expression trajectory can be functional related and therefore haploinsuffiency of these genes could contribute together to the downstream phenotype. Furthermore, looking at expression during fetal development is particularly relevant as it is a key window for the development of neuropsychiatric disorders (Weinberger, 1987; van Dam *et al.*, 2017). This analysis revealed distinct clusters of 22q11.2 deletion genes with strong gene expression correlation, with many of the clusters containing genes physically close together in the genome and within the same deletion breakpoint regions. Physically close genes in the genome have been previously reported to have similar expression patterns (Woo, Walker and Churchill, 2010). Furthermore, the majority of genes from the typical deletion had highly correlated gene expression, supporting the hypothesis that multi-gene loss contributes to the phenotype of 22q11.2DS. This was supplemented with literature review into predicted loss of function

and haploinsufficiency intolerance genes in the 22q11.2 region, as mutation intolerance has been shown to be enriched in schizophrenia candidate genes (Pardiñas *et al.*, 2018). This resulted in the selection of DGCR8, HIRA and ZDHHC8 as potential coordinated candidate genes for schizophrenia risk.

In the second section of this chapter, transcriptome wide association studies were carried out to investigate heritable gene expression changes between 22q11.2DS with and without schizophrenia, for identification schizophrenia disease modifier genes. The rationale behind this was based upon the wide symptomatic heterogeneity between 22q11.2DS patients. There is evidence to indicate that this heterogeneity is underlined by the potential existence of genetic modifiers that have an additive effect, along with the deletion through epistatic interactions, based on the multiple-hit model (Cirillo *et al.*, 2014; Merico *et al.*, 2015). It has already been suggested that copy-number variants outside of the 22q11.2 region can potentially act as a second "hit" for schizophrenia in 22q11.2DS (Williams *et al.*, 2013). Gene expression changes were analysed in DLPFC, cortex, fetal brain and whole blood, however no significant associations were identified across all gene expression panels. The genes with the highest statistical significance, were identified in the DLPFC, indicating the importance of looking in disease relevant tissues. Although we know common variants associated with schizophrenia do play a role in aetiology in 22q11.2DS patients with schizophrenia (Cleynen *et al.*, 2020), my study could not further elucidate the specific biological mechanisms underlying this. This could be due to small sample size or lack of common changes in gene expression between 22q11.2 deletion carriers with schizophrenia and deletion carriers without psychosis.

In the fourth chapter, I generated a *DGCR8* mutant knock-out human embryonic stem cell (hESC) line using CRISPR/Cas9 genome editing technology. *DGCR8* is a particular gene of interest from the 22q11.2 region due to its important role in miRNA biogenesis, therefore happloinsuffiency would potentially lead to global gene expression dysregulation. Investigation of loss of DGCR8 could therefore have consequences for epistatic interactions, which are of interest in 22q11.2DS. Importantly, the cell line exhibited no abnormal proliferation phenotype, which has been observed in the literature and showed normal phenotype when stem cell markers were characterised.

Furthermore, the cell line showed no self-renew or differentiation restriction issues, which has also been previously noted (Wang *et al.*, 2007; Cirera-Salinas *et al.*, 2017; Liu *et al.*, 2017). The generated cell line displayed no difference in cortical markers compared to the parental line. However, although the derived cell lines were genetically heterozygous and homozygous, protein levels of DGCR8 were not significantly reduced in edited lines and this may underlie the lack of phenotype observed. If a homozygous *DGCR8* knockout cell line was obtained, protein levels and the CNV status of the cell line would have to be examined. Further to this, differentiation potential would have been examined, as well as transcriptomic analysis, examining both mRNA and miRNA expression to see if these are altered in a hESCs model as has been reported in different models in the literature (See Chapter 1, Table 1.3). These investigations would be used to understand potential pathways and systems affected by loss of DGCR8.

In the third results chapter, a lentiviral based CRISPR/Cas9 editing system was utilised for genetic manipulation of the selected schizophrenia candidate genes *DGCR8*, *HIRA* and *ZDHHC8* in NPCs during cortical differentiation. The aim of this chapter was to uncover potential shared disrupted mechanisms caused by the haploinsuffiency of these genes. The use of the lentivirus, without editing, was first characterised to determine that there were no abnormal effects caused by the lentivirus during cortical differentiation. This revealed significant differences between transduced and non-transduced NPCs within the same culture. A higher proportion of TBR1 and CTIP2 positive cells were observed in the transduced cells compared to the non-transduced cells. However, the lentivirus was not increasing the total number of TBR1 and CTIP2 positive cells present in the culture, suggesting it has a higher efficiency for infecting intermediate neuroprogenitors over early progenitors. With this in mind, when the lentiviral CRISPR/Cas9 construct was used for targeting *DGCR8.* Genetic manipulation of *DGCR8* in NPCs resulted in an increase in the proportion of $TBR1^+$ cells at day 25, further indicating an interaction between DGCR8 and *TBR1* as previously described (Marinaro *et al.*, 2017), but this has not been observed in a human cell model before. The Covid-19 pandemic halted the work at this stage and so further investigations of the other selected genes could not take place. The final experiment planned was RNA sequencing of FAC sorted GFP neuroprogenitor cells that were transduced with lentiviruses individually targeting *DGCR8, HIRA* and *ZDHHC8.* The aim of this experiment was to

uncover any shared dysregulated transcriptomic pathways between manipulated genes. At the point of the pandemic, all lentiviruses had been prepared and optimised and the process of optimising RNA extraction after FAC sorting was being established, however due to laboratory closure, this work could not be completed.

6.2 Limitations and possible solutions

There are multiple lines of evidence indicating genetic background contributes to differences in the clinical manifestations of 22q11.2 patients (Guo *et al.*, 2015, 2017; Morrow *et al.*, 2018; Michaelovsky *et al.*, 2019; Cleynen *et al.*, 2020). Large-scale genetic studies have been widely successful for identifying common risk loci contributing to disease risk. Although, I had access to the largest cohort of 22q11.2DS subjects with and without schizophrenia, which has previously been reported as sufficiently powered for comparison (Cleynen *et al.*, 2020), TWAS investigation across disease-relevant tissue gene expression panels did not undercover significant differences in gene expression between the groups. This may be due to limited cohort size, but it equally could be hindered by the possibility that patients in the non-schizophrenia cohort may go on to develop schizophrenia. To overcome this, 22q11.2DS patients with schizophrenia could be compared to idiopathic schizophrenia patients using both GWAS and TWAS. Genetic data from large idiopathic schizophrenia patient cohorts are available, therefore increasing power of the study and could potentially elucidate the presence of modifier genes in the 22q11.2DS.

It has been reported that 22q11.2DS patients with schizophrenia have a higher burden of additional rare mutations in protein-coding genes in general, but specifically neurofunctional genes compared to those without schizophrenia (Merico *et al.*, 2015; Bassett *et al.*, 2017). Another possibility is lack of common variants shared within the cohort, genetic variation may be patient specific. Pedigree studies in families with 22q11.2DS and psychosis have provided specific insights into the transmission of risk variants. Whole-exome sequencing and comparative-genomic hybridization array were carried out with a multiplex family with 22q11.2DS that had high incidence of psychosis. “Damaging” variants within 22q11.2 genes as well as a rare CNV deletion at 3p26.3 were both identified (Michaelovsky *et al.*, 2019). Therefore, while sufficiently powered cohort

studies might not be available yet, investigations in multiplex families may elucidate specific genetic modifier genes, although these are very rare and unique situations.

In order to study loss of DGCR8, attempts were made to generate a knockout human embryonic stem cell line to study and characterise through cortical differentiation. Although genetically heterozygous and homozygous cell lines were generated, reduction at the protein level was not observed. Interestingly, at the time of single cell isolation and clonal picking, there were no surviving colonies in cells targeting with sgRNA 2, although GFP was initially observed. TIDE analysis in the subsequent chapter revealed that sgRNA 2 was the more efficient gRNA, potentially indicating that with more efficient gene editing, cells do not survive the editing process.

The derivation of the cell lines in this thesis were using the iCas9 cell line and the less common method of lentiviral-based gene manipulation. This approach was selected based upon prior work performed in the lab also attempting to generate a *DGCR8* knockout line. The previous attempt at derivation was using a transfection method with the H7 cell line, however transcriptomic analysis revealed a CNV deletion spanning the region 17p13.1. This region contains the tumour suppressor gene p53, which has been shown to be upregulated in *Dgcr8*$^{-/-}$ mESCs and subsequently restricts neural differentiation, therefore loss of this region could be advantageous for cell survival (Liu *et al.*, 2017). Furthermore, this same CNV was also identified in induced pluripotent stem cells from 22q11.2DS patients, indicating a potential mechanism induced by the presence of the deletion and hypothetically due to the specific loss of *DGCR8*. However, this CNV has been identified as a common abnormality in human pluripotent cells, therefore it was important to establish if it occurs as a result of targeting *DGCR8* (Garitaonandia *et al.*, 2015; Assou *et al.*, 2020). The derived lines in this thesis had a normal pluripotency status and genome editing did not induce any off target CNVs, however there was little reduction of DGCR8 at the protein level, potentially indicating an underlying compensatory mechanism or the editing process was not sufficient enough to induce a pathogenic mutation. It has been shown in *Dcgr8*$^{+/-}$ mouse ESCs that there is not reduction of Dgcr8 protein due to a feedback control by the microprocessor, therefore this may also be the case for hESCs (Han *et al.*, 2009; Triboulet *et al.*, 2009). Therefore it is interesting that 22q11.2DS patients do have reduction of DGCR8 and

miRNAs levels with loss of one allele, indicating for disease relevant modelling purposes reduction of protein levels are required (Sellier *et al.*, 2014).

Another study which aimed to generate DGCR8 KO hESCs found that homozygous mutants had an abnormal karyotype with an inter-chromosomal translocation between chromosomes 19 and 22, or the cell line exhibited poor maintenance of self-renewal capacity and high levels of spontaneous differentiation. In hESCs, an N-terminal-truncated version of DGCR8 was generated, meaning the miRNA processing end of the protein was still fully functional and cells displayed normal karyotype and self-renewal capacity (Deng *et al.*, 2019). Therefore, this further indicates that full DGCR8 loss may not be feasible in a hESC clonal cell line model. The combination of these efforts indicates that a hESC KO cell line might not be a feasible approach for modelling *DGCR8* loss. Consequently, alternative methods that are transient may be more suitable for the study of *DGCR8* in human cell models. Alternatively, hESC line approaches might not be suitable for modelling *DGCR8* loss and it might be more appropriate to use the already generated and reported mESCs lines and further characterise these, as they appear to have less genetic and morphological issues associated with them (Wang *et al.*, 2007). Heme binding is required for the processing of at least a subset of pri-miRNA, therefore inhibition of this could be an approach for investigating the loss of miRNA biogenesis role of DGCR8, however other approaches are needed for looking at the miRNA-independent roles of DGCR8 (Partin *et al.*, 2018). Techniques such as siRNA based knockdown or lentiviral based approach discussed in chapter five could provide a feasible method for gene manipulation of *DGCR8* in hESCs, but also other cell types of interest.

6.3 Future work and directions

The work presented in this thesis aimed to uncover the complex mechanisms underlying schizophrenia risk in 22q11.2DS. After the establishment of the lentiviral based CRISR/Cas9 gene editing approach in NPCs, investigation into *DGCR8* manipulation demonstrated increase in TBR1 positive cells when compared to NPCs transduced with a non-targeting virus, which further suggested a mechanism between DGCR8 and TBR1. However, it was at this stage that the work was halted due to the Covid-19 pandemic. It would be interesting to establish if the mechanism of regulation of the *TBR1* transcript

by DGCR8 that has been observed in mouse studies is replicated in this human NPC model and underlies the observed increase in TBR1 positive cells (Marinaro *et al.*, 2017). Furthermore, phenotypically assessing the effects of *HIRA* and *ZDHHC8* gene manipulation in cortical development was another subsequent aim. *HIRA* has been found to be enriched in neuroprogenitors. Moreover, a mouse knockdown model of *Hira* revealed that Hira promotes cell cycle exit and premature neuron differentiation, due to β-Catenin regulation (Li and Jiao, 2017). Therefore, assessment of cortical markers such as PAX6 and TUJ1 in *HIRA* targeted NPCs would be interesting to confirm if this finding is replicated in human neuroprogenitor cells.

Secondly, the overall aim of using this approach was to assess the existence of shared mechanisms between the selected schizophrenia candidate genes. This was to be achieved by FACs sorting NPCs transduced with lentiviruses targeting each gene or the empty virus, as a comparison and performing bulk RNA sequencing on these populations. This could potentially elucidate transcriptomic changes within each targeted population, but also identity common dysregulated pathways between the 22q11.2 genes. Furthermore, there is available RNA sequencing data from 22q11.2DS iPSCs and so comparisons to this data could identify which genes contribute to specific transcriptomic changes. Subsequently, this analysis could then inform follow-up phenotypic experiments.

There are multiple lines of evidence to indicate that multi-gene loss is likely to contribute to the manifestation of 22q11.2DS (Motahari *et al.*, 2019), consequently, a multigene approach should be used to understand mechanisms underlying 22q11.2DS. Such approaches could incorporate use of a CRISPR/Cas9 based screening of 22q11.2 deletion genes. Generation of a library harbouring siRNAs or sgRNAs targeting 22q11.2 genes categorised by biological function or based on association with neuropsychiatric disorders. This approach could be applied to ESCs or NPCs, followed by single-cell RNA sequencing. The large scale approach of CRISPR libraries has been beneficial in drug screening studies (Kurata *et al.*, 2018; Wei *et al.*, 2019). Targeting multiple genes within the 22q11.2 region is more likely to mimic the overall cellular and transcriptome dysregulation of the syndrome and ultimately elucidate common mechanisms behind the disorder.

With the increased availability and reduced cost of whole genome sequencing technologies, these technologies in case-control study designs have been beneficial for elucidating genetic variants across many diseases. Larger cohorts of 22q11.2DS with schizophrenia could become available over time, therefore increasing the opportunity for variant identification. Sequencing of the 22q11.2 deletion has proved successful for identifying variants associated with conotruncal heart defects, identifying dysregulation of *CRKL* expression contributes to penetrance of conotruncal heart defects (Zhao *et al.*, 2020). However, this approach has not yet identified deleterious variants in schizophrenia in 22q11.2 deletion syndrome, alternatively it might be advantageous to investigate the potential existence of protective mechanisms. It has been reported that the polygenic risk scores of 22q11.2DS carriers with psychosis can be stratified into high and low risk groups *(Davies and* Fiksinski *et al., in press)*. Therefore, individuals with 22q11.2DS that have high polygenic risk for schizophrenia but do not display psychosis, offer a unique opportunity to screen for potentially protective variants. Furthermore, sequencing of patients with atypical deletions that still display particular symptoms could be beneficial for identifying 22q11.2 genes or outside genetic modifiers that contribute to 22q11.2DS phenotypes (Molck *et al.*, 2013). Importantly, identification of protective mechanisms could allow for interpretation of alleviating pathways to potentially target.

6.4 Conclusions

The findings presented in this thesis provide the basis for future work into molecular and cellular mechanisms underlying schizophrenia risk in 22q11.2 deletion syndrome. Use of gene expression studies and known pathogenic/ loss of function mutations has offered a non-bias approach to selecting schizophrenia relevant candidate genes. Although transcriptomic differences were not identified between 22q11.2DS patients with and without schizophrenia, future work with larger cohorts over time may provide insights into underlying genetic mechanisms leading to schizophrenia risk. Investigations in this thesis into loss of DGCR8 using a human embryonic stem cell model indicate that clonal human cell lines may not be an effective method for modelling DGCR8 loss, but transient methods of gene manipulation provide a valid alternative. A lentiviral-based gene editing mechanism in human neuroprogenitor cells has been established, providing a suitable option for investigation of selected schizophrenia candidate genes

in cortical differentiation. This model could be used to study these promising candidate genes and explore resulting transcriptomic signatures due to happloinsuffiency, in order to gain a deeper understanding of the underlying genetic, molecular and cellular mechanisms underlying schizophrenia risk in 22q11.2 deletion syndrome.

References

Abu-Dawud, R. *et al.* (2018) 'Pluripotent stem cells: Induction and self-renewal', *Philosophical Transactions of the Royal Society B: Biological Sciences*. doi: 10.1098/rstb.2017.0213.

Adam, S., Polo, S. E. and Almouzni, G. (2013) 'XTranscription recovery after DNA damage requires chromatin priming by the H3.3 histone chaperone HIRA', *Cell*, 155(1), p. 94. doi: 10.1016/j.cell.2013.08.029.

Afenjar, A. *et al.* (2007) 'Early neurological phenotype in 4 children with biallelic PRODH mutations', *Brain and Development*. doi: 10.1016/j.braindev.2007.01.008.
Aguet, F. *et al.* (2017) 'Genetic effects on gene expression across human tissues', *Nature*. doi: 10.1038/nature24277.

Aicart-Ramos, C., Valero, R. A. and Rodriguez-Crespo, I. (2011) 'Protein palmitoylation and subcellular trafficking', *Biochimica et Biophysica Acta - Biomembranes*. doi: 10.1016/j.bbamem.2011.07.009.

Al-Shahrour, F. *et al.* (2010) 'Selection upon genome architecture: Conservation of functional neighborhoods with changing genes', *PLoS Computational Biology*. doi: 10.1371/journal.pcbi.1000953.

Alcamo, E. A. *et al.* (2008) 'Satb2 Regulates Callosal Projection Neuron Identity in the Developing Cerebral Cortex', *Neuron*. doi: 10.1016/j.neuron.2007.12.012.

Allen, N. C. *et al.* (2008) 'Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: The SzGene database', *Nature Genetics*. doi: 10.1038/ng.171.

Amati, F. *et al.* (2007) 'Dynamic changes in gene expression profiles of 22q11 and related orthologous genes during mouse development', *Gene*. doi: 10.1016/j.gene.2006.12.026.

Amminger, G. P. *et al.* (2015) 'Longer-term outcome in the prevention of psychotic disorders by the Vienna omega-3 study', *Nature Communications*. doi: 10.1038/ncomms8934.

Andrews, T. *et al.* (2015) 'The clustering of functionally related genes contributes to CNV-mediated disease', *Genome Research*. doi: 10.1101/gr.184325.114.

Arber, C. *et al.* (2015) 'Activin a directs striatal projection neuron differentiation of human pluripotent stem cells', *Development (Cambridge)*. doi: 10.1242/dev.117093.

Assou, S. *et al.* (2020) 'Recurrent Genetic Abnormalities in Human Pluripotent Stem Cells: Definition and Routine Detection in Culture Supernatant by Targeted Droplet Digital PCR', *Stem Cell Reports*. doi: 10.1016/j.stemcr.2019.12.004.

Aubrey, B. J. *et al.* (2015) 'An Inducible Lentiviral Guide RNA Platform Enables the Identification of Tumor-Essential Genes and Tumor-Promoting Mutations InVivo', *Cell Reports*. The Authors, 10(8), pp. 1422–1432. doi: 10.1016/j.celrep.2015.02.002.

Avramopoulos, D. (2018) 'Recent Advances in the Genetics of Schizophrenia', *Molecular Neuropsychiatry*, 4(1), pp. 35–51. doi: 10.1159/000488679.

Ayalew, M. *et al.* (2012) 'Convergent functional genomics of schizophrenia: From comprehensive understanding to genetic risk prediction', *Molecular Psychiatry*. doi: 10.1038/mp.2012.37.

Babcock, M. *et al.* (2003) 'Shuffling of genes within low-copy repeats on 22q11 (LCR22) by Alu-mediated recombination events during evolution', *Genome Research*. doi: 10.1101/gr.1549503.

Badner, J. A. and Gershon, E. S. (2002) 'Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia', *Molecular Psychiatry*. doi: 10.1038/sj.mp.4001012.

Barde, I., Salmon, P. and Trono, D. (2010) 'Production and titration of lentiviral vectors', *Current Protocols in Neuroscience*, pp. 1–23. doi: 10.1002/0471142301.ns0100s37.

Bartsch, I. *et al.* (2011) 'Deletion of human GP1BB and SEPT5 is associated with Bernard-Soulier syndrome, platelet secretion defect, polymicrogyria, and develop-mental delay', *Thrombosis and Haemostasis*. doi: 10.1160/TH11-05-0305.

Bassett, A. S. *et al.* (2017) 'Rare genome-wide copy number variation and expression of schizophrenia in 22q11.2 deletion syndrome', *American Journal of Psychiatry*. doi: 10.1176/appi.ajp.2017.16121417.

Bearden, C. E. *et al.* (2004) 'Effects of a functional COMT polymorphism on prefrontal cognitive function in patients with 22q11.2 deletion syndrome', *American Journal of Psychiatry*. doi: 10.1176/appi.ajp.161.9.1700.

Bedogni, F. *et al.* (2010) 'Tbr1 regulates regional and laminar identity of postmitotic neurons in developing neocortex', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1002285107.

Bender, H. U. *et al.* (2005) 'Functional consequences of PRODH missense mutations', *American Journal of Human Genetics*. doi: 10.1086/428142.

Benjamini, Y. and Hochberg, Y. (1995) 'Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing', *Journal of the Royal Statistical Society: Series B (Methodological)*. doi: 10.1111/j.2517-6161.1995.tb02031.x.

Bertrán, M., Tagle, F. P. and Irarrázaval, M. (2018) 'Psychiatric manifestations of 22q11.2 deletion syndrome: a literature review', *Neurologia*. doi: 10.1016/j.nrl.2015.07.007.

Beveridge, N. J. *et al.* (2010) 'Schizophrenia is associated with an increase in cortical microRNA biogenesis', *Molecular Psychiatry*. Nature Publishing Group, 15(12), pp. 1176–1189. doi: 10.1038/mp.2009.84.

Beveridge, N. J. and Cairns, M. J. (2012) 'MicroRNA dysregulation in schizophrenia', *Neurobiology of Disease*. doi: 10.1016/j.nbd.2011.12.029.

Bhalala, O. G. *et al.* (2018) 'Identification of expression quantitative trait loci associated with schizophrenia and affective disorders in normal brain tissue', *PLoS Genetics*. doi: 10.1371/journal.pgen.1007607.

Blackwood, D. H. R. *et al.* (2001) 'Schizophrenia and affective disorders - Cosegregation with a translocation at chromosome 1q42 that directly disrupts brain-expressed genes: Clinical and P300 findings in a family', *American Journal of Human Genetics*. doi: 10.1086/321969.

Blits, B. *et al.* (2005) 'Lentiviral vector-mediated transduction of neural progenitor cells before implantation into injured spinal cord and brain to detect their migration, deliver neurotrophic factors and repair tissue', *Restorative Neurology and Neuroscience*.

Boissart, C. *et al.* (2013) 'Differentiation from human pluripotent stem cells of cortical neurons of the superficial layers amenable to psychiatric disease modeling and high-throughput drug screening', *Translational Psychiatry*. doi: 10.1038/tp.2013.71.

Boot, E. *et al.* (2018) 'Typical features of Parkinson disease and diagnostic challenges with microdeletion 22q11.2', *Neurology*. doi: 10.1212/WNL.0000000000005660.

Van Den Bosch, M. A. A. J. *et al.* (2002) 'Hypocalcemic tetany as an early sign of DiGeorge syndrome in an adult woman [4]', *American Journal of Medicine*. doi: 10.1016/S0002-9343(01)00955-X.

Botto, L. D. *et al.* (2003) 'A population-based study of the 22q11.2 Deletion: Phenotype, incidence, and contribution to major birth defects in the population', *Pediatrics*. doi: 10.1542/peds.112.1.101.

Brennand, K. J. *et al.* (2011) 'Modelling schizophrenia using human induced pluripotent stem cells', *Nature*. doi: 10.1038/nature09915.

Brinkman, E. K. *et al.* (2014) 'Easy quantitative assessment of genome editing by sequence trace decomposition', *Nucleic Acids Research*. doi: 10.1093/nar/gku936.

Britanova, O. *et al.* (2008) 'Satb2 Is a Postmitotic Determinant for Upper-Layer Neuron Specification in the Neocortex', *Neuron*. doi: 10.1016/j.neuron.2007.12.028.

Buchschacher, G. L. and Wong-Staal, F. (2000) 'Development of lentiviral vectors for gene therapy for human diseases', *Blood*. doi: 10.1182/blood.v95.8.2499.008k35_2499_2504.

Bush, W. S. and Moore, J. H. (2012) 'Chapter 11: Genome-Wide Association Studies', *PLoS Computational Biology*. doi: 10.1371/journal.pcbi.1002822.

Butcher, N. J. *et al.* (2013) 'Association between early-onset Parkinson disease and 22q11.2 deletion syndrome: Identification of a novel genetic form of Parkinson disease and its clinical implications', *JAMA Neurology*. doi: 10.1001/jamaneurol.2013.3646.

Caltagarone, J. *et al.* (1998) 'Localization of a novel septin protein, hCDCrel-1, in neurons of human brain', *NeuroReport*. doi: 10.1097/00001756-199808240-00042.

Cambray, S. *et al.* (2012) 'Activin induces cortical interneuron identity and differentiation in embryonic stem cell-derived telencephalic neural precursors', *Nature Communications*. doi: 10.1038/ncomms1817.

Carlson, C. *et al.* (1997) 'Molecular definition of 22q11 deletions in 151 velo-cardio-facial syndrome patients', *American Journal of Human Genetics*. doi: 10.1086/515508.

Caron, H. *et al.* (2001) 'The human transcriptome map: Clustering of highly expressed genes in chromosomal domains', *Science*. doi: 10.1126/science.1056794.

Cassa, C. A. *et al.* (2017) 'Estimating the selective effects of heterozygous protein-truncating variants from human exome data', *Nature Genetics*. Nature Publishing Group, 49(5), pp. 806–810. doi: 10.1038/ng.3831.

Chambers, S. M. *et al.* (2009) 'Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling', *Nature Biotechnology*. doi: 10.1038/nbt.1529.

Chang, C. C. *et al.* (2015) 'Second-generation PLINK: Rising to the challenge of larger and richer datasets', *GigaScience*. doi: 10.1186/s13742-015-0047-8.

Chen, B. *et al.* (2008) 'The Fezf2-Ctip2 genetic pathway regulates the fate choice of subcortical projection neurons in the developing cerebral cortex', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.0804918105.

Chen, J. *et al.* (2004) 'Functional analysis of genetic variation in catechol-O-methyltransferase (COMT): Effects on mrna, protein, and enzyme activity in postmortem human brain', *American Journal of Human Genetics*. doi: 10.1086/425589.

Chen, L. *et al.* (2014) 'Genome Architecture and Its Roles in Human Copy Number Variation', *Genomics & Informatics*. doi: 10.5808/gi.2014.12.4.136.

Chen, W. *et al.* (2015) ' Intellectual disability-associated dBRWD 3 regulates gene expression through inhibition of HIRA / YEM -mediated chromatin deposition of histone H3.3 ', *EMBO reports*. doi: 10.15252/embr.201439092.

Chen, W. Y. *et al.* (2004) 'Case-control study and transmission disequilibrium test provide consistent evidence for association between schizophrenia and genetic variation in the 22q11 gene ZDHHC8', *Human Molecular Genetics*. doi: 10.1093/hmg/ddh322.

Chen, X. *et al.* (2011) 'PSD-95 is required to sustain the molecular organization of the postsynaptic density', *Journal of Neuroscience*. doi: 10.1523/JNEUROSCI.5968-10.2011.

Cheung, E. N. M. *et al.* (2014) 'Prevalence of hypocalcaemia and its associated features in 22q11·2 deletion syndrome', *Clinical Endocrinology*. doi: 10.1111/cen.12466.

Chin, M. H. *et al.* (2010) 'Molecular analyses of human induced pluripotent stem cells and embryonic stem cells', *Cell Stem Cell*. doi: 10.1016/j.stem.2010.06.019.

Cho, S. W. *et al.* (2013) 'Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease', *Nature Biotechnology*. doi: 10.1038/nbt.2507.

Choi, Y. and Chan, A. P. (2015) 'PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels', *Bioinformatics*. doi: 10.1093/bioinformatics/btv195.

Chow, E. W. C. *et al.* (2002) 'Structural brain abnormalities in patients with schizophrenia and 22q11 Deletion Syndrome', *Biological Psychiatry*. doi: 10.1016/S0006-3223(01)01246-X.

Chun, S. *et al.* (2014) 'Specific disruption of thalamic inputs to the auditory cortex in schizophrenia models', *Science*. doi: 10.1126/science.1253895.

Ciani, L. and Salinas, P. C. (2005) 'WNTs in the vertebrate nervous system: From patterning to neuronal connectivity', *Nature Reviews Neuroscience*. doi: 10.1038/nrn1665.

Cirera-Salinas, D. *et al.* (2017) 'Noncanonical function of DGCR8 controls mESC exit from pluripotency', *Journal of Cell Biology*. doi: 10.1083/jcb.201606073.

Cirillo, E. *et al.* (2014) 'Intergenerational and intrafamilial phenotypic variability in 22q11.2 Deletion syndrome subjects', *BMC Medical Genetics*. doi: 10.1186/1471-2350-15-1.

Cleynen, I. *et al.* (2020) 'Genetic contributors to risk of schizophrenia in the presence of a 22q11.2 deletion', *Molecular Psychiatry*. doi: 10.1038/s41380-020-0654-3.

Cohen, S. M. and Nadler, J. V. (1997) 'Proline-induced potentiation of glutamate transmission', *Brain Research*. doi: 10.1016/S0006-8993(97)00352-1.

Cong, L. *et al.* (2013) 'Multiplex genome engineering using CRISPR/Cas systems', *Science*. doi: 10.1126/science.1231143.

Cook, D. L., Gerber, A. N. and Tapscott, S. J. (1998) 'Modeling stochastic gene expression: Implications for haploinsufficiency', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.95.26.15641.

Cookson, W. *et al.* (2009) 'Mapping complex disease traits with global gene expression', *Nature Reviews Genetics*. doi: 10.1038/nrg2537.

Copp, A. J., Greene, N. D. E. and Murdoch, J. N. (2003) 'The genetic basis of mammalian neurulation', *Nature Reviews Genetics*. doi: 10.1038/nrg1181.

Cutting, G. R. (2010) 'Modifier genes in Mendelian disorders: The example of cystic fibrosis', *Annals of the New York Academy of Sciences*. doi: 10.1111/j.1749-6632.2010.05879.x.

Dahlstrand, J., Lardelli, M. and Lendahl, U. (1995) 'Nestin mRNA expression correlates with the central nervous system progenitor cell state in many, but not all, regions of developing central nervous system', *Developmental Brain Research*. doi: 10.1016/0165-3806(94)00162-S.

Dai, L. *et al.* (2016) 'Cytoplasmic Drosha activity generated by alternative splicing', *Nucleic Acids Research*. doi: 10.1093/nar/gkw668.

van Dam, S. *et al.* (2017) 'Gene co-expression analysis for functional classification and gene–disease predictions', *Briefings in Bioinformatics*, 19(4), pp. 575–592. doi: 10.1093/bib/bbw139.

De Decker, H. P. and Lawrenson, J. B. (2001) 'The 22q11.2 deletion: From diversity to a single gene theory', in *Genetics in Medicine*. doi: 10.1097/00125817-200101000-00002.

Demily, C. *et al.* (2015) 'Neurocognitive and psychiatric management of the 22q11.2 deletion syndrome', *Encephale*. doi: 10.1016/j.encep.2014.10.005.

Deng, L. *et al.* (2019) 'Stabilizing heterochromatin by DGCR8 alleviates senescence and osteoarthritis', *Nature Communications*. doi: 10.1038/s41467-019-10831-8.

Digilio, M. *et al.* (2005) 'Clinical manifestations of Deletion 22q11.2 syndrome (DiGeorge/Velo-Cardio-Facial syndrome).', *Images in paediatric cardiology*.

Digillo, M. C. *et al.* (2003) 'Spectrum of clinical variability in familial deletion 22q11.2: From full manifestation to extremely mild clinical anomalies', *Clinical Genetics*. doi: 10.1034/j.1399-0004.2003.00049.x.

Dilg, D. *et al.* (2016) 'HIRA is required for heart development and directly regulates Tnni2 and Tnnt3', *PLoS ONE*. doi: 10.1371/journal.pone.0161096.

Ding, B. *et al.* (2020) 'Power analysis of transcriptome-wide association study', *bioRxiv*, p. 2020.07.19.211151. doi: 10.1101/2020.07.19.211151.

Donegan, J. J. and Lodge, D. J. (2017) 'Cell-based therapies for the treatment of schizophrenia', *Brain Research*. doi: 10.1016/j.brainres.2016.08.010.

Driscoll, D. A. *et al.* (1993) 'Prevalence of 22q11 microdeletions in DiGeorge and velocardiofacial syndromes: implications for genetic counselling and prenatal diagnosis.', *Journal of Medical Genetics*, 30(10), pp. 813 LP – 817. doi: 10.1136/jmg.30.10.813.

Dull, T. *et al.* (1998) 'A Third-Generation Lentivirus Vector with a Conditional Packaging System', *Journal of Virology*. doi: 10.1128/jvi.72.11.8463-8471.1998.

Earls, L. R. *et al.* (2010) 'Dysregulation of presynaptic calcium and synaptic plasticity in a mouse model of 22q11 deletion syndrome', *Journal of Neuroscience*. doi: 10.1523/JNEUROSCI.1425-10.2010.

Earls, L. R. *et al.* (2012) 'Age-dependent microRNA control of synaptic plasticity in 22q11 deletion syndrome and schizophrenia', *Journal of Neuroscience*. doi: 10.1523/JNEUROSCI.1312-12.2012.

Egan, M. F. *et al.* (2001) 'Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.111134598.

Egan, M. F. *et al.* (2004) 'Variation in GRM3 affects cognition, prefrontal glutamate, and risk for schizophrenia', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.0405077101.

El-Husseini, A. E. D. *et al.* (2002) 'Synaptic strength regulated by palmitate cycling on PSD-95', *Cell*. doi: 10.1016/S0092-8674(02)00683-9.

El-Husseini, A. E. D. and Bredt, D. S. (2002) 'Protein palmitoylation: A regulator of neuronal development and function', *Nature Reviews Neuroscience*. doi: 10.1038/nrn940.

Englund, C. *et al.* (2005) 'Pax6, Tbr2, and Tbr1 are expressed sequentially by radial glia, intermediate progenitor cells, and postmitotic neurons in developing neocortex', *Journal of Neuroscience*. doi: 10.1523/JNEUROSCI.2899-04.2005.

Eom, T. Y. *et al.* (2017) 'Schizophrenia-Related Microdeletion Impairs Emotional Memory through MicroRNA-Dependent Disruption of Thalamic Inputs to the Amygdala', *Cell Reports*. ElsevierCompany., 19(8), pp. 1532–1544. doi: 10.1016/j.celrep.2017.05.002.

Eom, T. Y. *et al.* (2020) 'Schizophrenia-related microdeletion causes defective ciliary motility and brain ventricle enlargement via microRNA-dependent mechanisms in mice', *Nature Communications*. doi: 10.1038/s41467-020-14628-y.

Escors, D. and Breckpot, K. (2010) 'Lentiviral vectors in gene therapy: Their current status and future potential', *Archivum Immunologiae et Therapiae Experimentalis*. doi: 10.1007/s00005-010-0063-4.

Espuny-Camacho, I. *et al.* (2013) 'Pyramidal Neurons Derived from Human Pluripotent Stem Cells Integrate Efficiently into Mouse Brain Circuits In Vivo', *Neuron*. doi: 10.1016/j.neuron.2012.12.011.

Estivill-Torrus, G. *et al.* (2002) 'Pax6 is required to regulate the cell cycle and the rate of progression from symmetrical to asymmetrical division in mammalian cortical progenitors', *Development*.

Etzioni, A. and Pollack, S. (1994) 'Autoimmune phenomena in DiGeorge syndrome.', *Israel Journal of Medical Sciences*.

Faller, M. *et al.* (2010) 'DGCR8 recognizes primary transcripts of microRNAs through highly cooperative binding and formation of higher-order structures', *RNA*. doi: 10.1261/rna.2111310.

Fang, L. and Wang, K. (2018) 'Identification of copy number variants from SNP arrays using PennCNV', in *Methods in Molecular Biology*. doi: 10.1007/978-1-4939-8666-8_1.

Federici, T. *et al.* (2009) 'Comparative analysis of HIV-1-based lentiviral vectors bearing lyssavirus glycoproteins for neuronal gene transfer', *Genetic Vaccines and Therapy*. doi: 10.1186/1479-0556-7-1.

Fénelon, K. *et al.* (2011) 'Deficiency of Dgcr8, a gene disrupted by the 22q11.2 microdeletion, results in altered short-term plasticity in the prefrontal cortex', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1101219108.

Flore, G. *et al.* (2017) 'Cortical Development Requires Mesodermal Expression of Tbx1, a Gene Haploinsufficient in 22q11.2 Deletion Syndrome', *Cerebral cortex (New York, N.Y. : 1991)*. doi: 10.1093/cercor/bhw076.

Forsyth, J. K. *et al.* (2019) 'Synaptic and Gene Regulatory Mechanisms in Schizophrenia, Autism, and 22q11.2 CNV Mediated Risk for Neuropsychiatric Disorders', *bioRxiv*. doi: 10.1101/555490.

Forsyth, J. K. *et al.* (2020) 'Synaptic and Gene Regulatory Mechanisms in Schizophrenia, Autism, and 22q11.2 Copy Number Variant–Mediated Risk for Neuropsychiatric Disorders', *Biological Psychiatry*. doi: 10.1016/j.biopsych.2019.06.029.

Frith, C. and Dolan, R. (1996) 'The role of the prefrontal cortex in higher cognitive functions', in *Cognitive Brain Research*. doi: 10.1016/S0926-6410(96)00054-7.

Fromer, M. *et al.* (2016) 'Gene expression elucidates functional impact of polygenic risk for schizophrenia', *Nature Neuroscience*. doi: 10.1038/nn.4399.

Fryett, J. J. *et al.* (2018) 'Comparison of methods for transcriptome imputation through application to two common complex diseases', *European Journal of Human Genetics*. doi: 10.1038/s41431-018-0176-5.

Fung, W. L. A. *et al.* (2010) 'Elevated prevalence of generalized anxiety disorder in adults with 22q11.2 deletion syndrome', *American Journal of Psychiatry*. doi: 10.1176/appi.ajp.2010.09101463.

Funke, B., Pandita, R. K. and Morrow, B. E. (2001) 'Isolation and characterization of a novel gene containing WD40 repeats from the region deleted in velo-cardio-facial/DiGeorge syndrome on chromosome 22q11', *Genomics*. doi: 10.1006/geno.2000.6506.

Gabriel, S. B. *et al.* (2002) 'The structure of haplotype blocks in the human genome', *Science*. doi: 10.1126/science.1069424.

Gamazon, E. R. *et al.* (2015) 'A gene-based association method for mapping traits using reference transcriptome data', *Nature Genetics*, 47(9), pp. 1091–1098. doi: 10.1038/ng.3367.

Gao, F. *et al.* (2015) 'Gene essentiality analysis based on DEG 10, an updated database of essential genes', *Methods in Molecular Biology*. doi: 10.1007/978-1-4939-2398-4_14.

Garitaonandia, I. *et al.* (2015) 'Increased risk of genetic and epigenetic instability in human embryonic stem cells associated with specific culture conditions', *PLoS ONE*. doi: 10.1371/journal.pone.0118307.

Gaspard, N. *et al.* (2008) 'An intrinsic mechanism of corticogenesis from embryonic stem cells', *Nature*. doi: 10.1038/nature07287.

Gennery, A. R. *et al.* (2002) 'Antibody deficiency and autoimmunity in 22q11.2 deletion syndrome', *Archives of Disease in Childhood*. doi: 10.1136/adc.86.6.422.

Gerdes, M. *et al.* (1999) 'Cognitive and behavior profile of preschool children with chromosome 22q11.2 deletion', *American Journal of Medical Genetics*. doi: 10.1002/(SICI)1096-8628(19990716)85:2<127::AID-AJMG6>3.0.CO;2-F.

Gerring, Z. F., Gamazon, E. R. and Derks, E. M. (2019) 'A gene co-expression network-based analysis of multiple brain tissues reveals novel genes and molecular pathways underlying major depression', *PLoS Genetics*. doi: 10.1371/journal.pgen.1008245.

Gibson, G. (2012) 'Rare and common variants: Twenty arguments', *Nature Reviews Genetics*. doi: 10.1038/nrg3118.

Glaser, B. *et al.* (2005) 'No association between the putative functional ZDHHC8 single nucleotide polymorphism rs175174 and schizophrenia in large European samples', *Biological Psychiatry*. doi: 10.1016/j.biopsych.2005.03.017.

Gogos, J. A. *et al.* (1999) 'The gene encoding proline dehydrogenase modulates sensorimotor gating in mice', *Nature Genetics*. doi: 10.1038/7777.

Goldberg, A. D. *et al.* (2010) 'Distinct Factors Control Histone Variant H3.3 Localization at Specific Genomic Regions', *Cell*. Elsevier Ltd, 140(5), pp. 678–691. doi: 10.1016/j.cell.2010.01.003.

Gong, L. *et al.* (2000) 'GNB1L, a gene deleted in the critical region for DiGeorge syndrome on 22q11, encodes a G-protein β-subunit-like polypeptide', *Biochimica et Biophysica Acta - Gene Structure and Expression*. doi: 10.1016/S0167-4781(00)00189-5.

González, F. *et al.* (2014) 'An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells', *Cell Stem Cell*, 15(2), pp. 215–226. doi: 10.1016/j.stem.2014.05.018.

Goodship, J. *et al.* (1998) 'A population study of chromosome 22q11 deletions in infancy', *Archives of Disease in Childhood*, 79(4), pp. 348–351. doi: 10.1136/adc.79.4.348.

Götz, M. and Huttner, W. B. (2005) 'The cell biology of neurogenesis', *Nature Reviews Molecular Cell Biology*. doi: 10.1038/nrm1739.

Greaves, J. and Chamberlain, L. H. (2007) 'Palmitoylation-dependent protein sorting', *Journal of Cell Biology*. doi: 10.1083/jcb.200610151.

Green, T. *et al.* (2009) 'Psychiatric Disorders and Intellectual Functioning Throughout Development in Velocardiofacial (22q11.2 Deletion) Syndrome', *Journal of the American Academy of Child and Adolescent Psychiatry*. doi: 10.1097/CHI.0b013e3181b76683.

Gregory, R. I. *et al.* (2004) 'The Microprocessor complex mediates the genesis of microRNAs', *Nature*. doi: 10.1038/nature03120.

Greig, L. C. *et al.* (2013) 'Molecular logic of neocortical projection neuron specification, development and diversity', *Nature Reviews Neuroscience*. doi: 10.1038/nrn3586.

Grishok, A. *et al.* (2001) 'Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing', *Cell*. doi: 10.1016/S0092-8674(01)00431-7.

Guenther, M. G. *et al.* (2007) 'A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells', *Cell*. doi: 10.1016/j.cell.2007.05.042.

Guna, A., Butcher, N. J. and Bassett, A. S. (2015) 'Comparative mapping of the 22q11.2 deletion region and the potential of simple model organisms', *Journal of Neurodevelopmental Disorders*. Journal of Neurodevelopmental Disorders, 7(1). doi: 10.1186/s11689-015-9113-x.

Guo, T. *et al.* (2015) 'Histone Modifier Genes Alter Conotruncal Heart Phenotypes in 22q11.2 Deletion Syndrome', *American Journal of Human Genetics*. doi: 10.1016/j.ajhg.2015.10.013.

Guo, T. *et al.* (2017) 'Genome-Wide Association Study to Find Modifiers for Tetralogy of Fallot in the 22q11.2 Deletion Syndrome Identifies Variants in the GPR98 Locus on 5q14.3', *Circulation: Cardiovascular Genetics*. doi: 10.1161/CIRCGENETICS.116.001690.

Gur, R. E. *et al.* (2017) 'A neurogenetic model for the study of schizophrenia spectrum disorders: The International 22q11.2 Deletion Syndrome Brain Behavior Consortium', *Molecular Psychiatry*. doi: 10.1038/mp.2017.161.

Gusev, A. *et al.* (2016) 'Integrative approaches for large-scale transcriptome-wide association studies', *Nature Genetics*, 48(3), pp. 245–252. doi: 10.1038/ng.3506.

Gusev, A. *et al.* (2018) 'Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights', *Nature Genetics*, 50(4), pp. 538–548. doi: 10.1038/s41588-018-0092-1.

Haapaniemi, E. *et al.* (2018) 'CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response', *Nature Medicine*. doi: 10.1038/s41591-018-0049-z.

Haijma, S. V. *et al.* (2013) 'Brain volumes in schizophrenia: A meta-analysis in over 18 000 subjects', *Schizophrenia Bulletin*. doi: 10.1093/schbul/sbs118.

Han, J. *et al.* (2004) 'The Drosha-DGCR8 complex in primary microRNA processing', *Genes and Development*. doi: 10.1101/gad.1262504.

Han, J. *et al.* (2009) 'Posttranscriptional Crossregulation between Drosha and DGCR8', *Cell*. doi: 10.1016/j.cell.2008.10.053.

Handel, A. E. *et al.* (2016) 'Assessing similarity to primary tissue and cortical layer identity in induced pluripotent stem cell-derived cortical neurons through single-cell transcriptomics', *Human Molecular Genetics*. doi: 10.1093/hmg/ddv637.

Hannon, E. *et al.* (2015) 'Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci', *Nature Neuroscience*. doi: 10.1038/nn.4182.

Hayashi, T., Thomas, G. M. and Huganir, R. L. (2009) 'Dual Palmitoylation of NR2 Subunits Regulates NMDA Receptor Trafficking', *Neuron*. doi: 10.1016/j.neuron.2009.08.017.

Heckers, S. (2000) 'Neural models of schizophrenia.', *Dialogues in clinical neuroscience*.

Heldt, S. A., Green, A. and Ressler, K. J. (2004) 'Prepulse inhibition deficits in GAD65 knockout mice and the effect of antipsychotic treatment', *Neuropsychopharmacology*. doi: 10.1038/sj.npp.1300468.

Heman-Ackah, S. M., Bassett, A. R. and Wood, M. J. A. (2016) 'Precision Modulation of Neurodegenerative Disease-Related Gene Expression in Human iPSC-Derived Neurons', *Scientific Reports*. doi: 10.1038/srep28420.

Herbst, F. *et al.* (2012) 'Extensive methylation of promoter sequences silences lentiviral transgene expression during stem cell differentiation in vivo', *Molecular Therapy*. doi: 10.1038/mt.2012.46.

Hevner, R. F. *et al.* (2001) 'Tbr1 regulates differentiation of the preplate and layer 6', *Neuron*. doi: 10.1016/S0896-6273(01)00211-2.

Heyer, W. D., Ehmsen, K. T. and Liu, J. (2010) 'Regulation of homologous recombination in eukaryotes', *Annual Review of Genetics*. doi: 10.1146/annurev-genet-051710-150955.

Hoffmann, N. *et al.* (2018) 'DGCR8 promotes neural progenitor expansion and represses neurogenesis in the mouse embryonic neocortex', *Frontiers in Neuroscience*, 12(APR), pp. 1–16. doi: 10.3389/fnins.2018.00281.

Holmans, P. A., Massey, T. H. and Jones, L. (2017) 'Genetic modifiers of Mendelian disease: Huntington's disease and the trinucleotide repeat disorders', *Human Molecular Genetics*. doi: 10.1093/hmg/ddx261.

Homem, C. C. F., Repic, M. and Knoblich, J. A. (2015) 'Proliferation control in neural stem and progenitor cells', *Nature Reviews Neuroscience*. doi: 10.1038/nrn4021.

Hommel, G. (1988) 'A stagewise rejective multiple test procedure based on a modified bonferroni test', *Biometrika*. doi: 10.1093/biomet/75.2.383.

Honer, W. G. (2002) 'Abnormalities of SNARE Mechanism Proteins in Anterior Frontal Cortex in Severe Mental Illness', *Cerebral Cortex*. doi: 10.1093/cercor/12.4.349.

Hopkins, S. E. *et al.* (2018) 'Neurologic challenges in 22q11.2 deletion syndrome', *American Journal of Medical Genetics, Part A*. doi: 10.1002/ajmg.a.38614.

Hsu, R. *et al.* (2012) 'Loss of microRNAs in pyramidal neurons leads to specific changes in inhibitory synaptic transmission in the prefrontal cortex', *Molecular and Cellular Neuroscience*, 50(3–4), pp. 283–292. doi: 10.1016/j.mcn.2012.06.002.

Hu, B. Y. *et al.* (2010) 'Neural differentiation of human induced pluripotent stem cells follows developmental principles but with variable potency', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.0910012107.

Huang, N. *et al.* (2010) 'Characterising and predicting haploinsufficiency in the human genome', *PLoS Genetics*. doi: 10.1371/journal.pgen.1001154.

Hughes, S. M. *et al.* (2002) 'Viral-mediated gene transfer to mouse primary neural progenitor cells', *Molecular Therapy*. doi: 10.1006/mthe.2001.0512.

Jacquet, H. *et al.* (2003) 'The severe form of type I hyperprolinaemia results from homozygous inactivation of the PRODH gene.', *Journal of medical genetics*. doi: 10.1136/jmg.40.1.e7.

Jandial, R. *et al.* (2008) 'Genetic modification of neural stem cells', *Molecular Therapy*. Nature Publishing Group, pp. 450–457. doi: 10.1038/sj.mt.6300402.

Jawad, A. F. *et al.* (2001) 'Immunologic features of chromosome 22q11.2 deletion syndrome (DiGeorge syndrome/velocardiofacial syndrome)', *Journal of Pediatrics*. doi: 10.1067/mpd.2001.118534.

Jerome, L. A. and Papaioannou, V. E. (2001) 'DiGeorge syndrome phenotype in mice mutant for the T-box gene, Tbx1', *Nature Genetics*. doi: 10.1038/85845.

Jinek, M. *et al.* (2012) 'A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity', *Science*. doi: 10.1126/science.1225829.

Johnston, H. R. *et al.* (2017) 'PEMapper and PECaller provide a simplified approach to whole-genome sequencing', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1618065114.

Johnston, J. C. *et al.* (1999) 'Minimum Requirements for Efficient Transduction of Dividing and Nondividing Cells by Feline Immunodeficiency Virus Vectors', *Journal of Virology*. doi: 10.1128/jvi.73.6.4991-5000.1999.

Jolin, E. M., Weller, R. A. and Weller, E. B. (2012) 'Occurrence of affective disorders compared to other psychiatric disorders in children and adolescents with 22q11.2 deletion syndrome', *Journal of Affective Disorders*. doi: 10.1016/j.jad.2010.11.025.

Jonas, R. K., Montojo, C. A. and Bearden, C. E. (2014) 'The 22q11.2 deletion syndrome as a window into complex neuropsychiatric disorders over the lifespan', *Biological Psychiatry*. doi: 10.1016/j.biopsych.2013.07.019.

Jones, C., Watson, D. and Fone, K. (2011) 'Animal models of schizophrenia', *British Journal of Pharmacology*. doi: 10.1111/j.1476-5381.2011.01386.x.

Kabadi, A. M. *et al.* (2014) 'Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector', *Nucleic Acids Research*, 42(19), pp. 1–11. doi: 10.1093/nar/gku749.

Karayiorgou, M *et al.* (1995) 'Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 92(17), pp. 7612–6. doi: 10.1073/PNAS.92.17.7612.

Karayiorgou, Maria *et al.* (1995) 'Schizophrenia susceptibility associated with interstitial deletions of chromosome 22qll', 92(August), pp. 7612–7616.

Karayiorgou, M., Simon, T. J. and Gogos, J. A. (2010) 'dysfunction and schizophrenia', 11(6), pp. 402–416. doi: 10.1038/nrn2841.2q11.2.

Karczewski, K. J. *et al.* (2019) 'Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes', *bioRxiv*. doi: 10.1101/531210.

Kempf, L. *et al.* (2008) 'Functional polymorphisms in PRODH are associated with risk and protection for schizophrenia and fronto-striatal structure and function', *PLoS Genetics*, 4(11). doi: 10.1371/journal.pgen.1000252.

Ketting, R. F. *et al.* (2001) 'Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans', *Genes and Development*. doi: 10.1101/gad.927801.

Kim, K. *et al.* (2010) 'Epigenetic memory in induced pluripotent stem cells', *Nature*. doi: 10.1038/nature09342.

Kinoshita, A., Noda, M. and Kinoshita, M. (2000) 'Differential localization of septins in the mouse brain', *Journal of Comparative Neurology*. doi: 10.1002/1096-9861(20001211)428:2<223::AID-CNE3>3.0.CO;2-M.

Klein, R. J. (2007) 'Power analysis for genome-wide association studies', *BMC Genetics*. doi: 10.1186/1471-2156-8-58.

Knuckles, P. *et al.* (2012) 'Drosha regulates neurogenesis by controlling Neurogenin 2 expression independent of microRNAs', *Nature Neuroscience*, 15(7), pp. 962–969. doi: 10.1038/nn.3139.

Kruglyak, L. and Nickerson, D. A. (2001) 'Variation is the spice of life', *Nature Genetics*. doi: 10.1038/85776.

Kunishima, S. *et al.* (2013) 'Bernard-Soulier syndrome caused by a hemizygous GPIbβ mutation and 22q11.2 deletion', *Pediatrics International*. doi: 10.1111/ped.12105.

Kurahashi, H. *et al.* (1997) 'Another critical region for deletion of 22q11: A study of 100 patients', *American Journal of Medical Genetics*. doi: 10.1002/(SICI)1096-8628(19971017)72:2<180::AID-AJMG10>3.0.CO;2-J.

Kurata, M. *et al.* (2018) 'CRISPR/Cas9 library screening for drug target discovery', *Journal of Human Genetics*. doi: 10.1038/s10038-017-0376-9.

Labrie, S. J., Samson, J. E. and Moineau, S. (2010) 'Bacteriophage resistance mechanisms', *Nature Reviews Microbiology*. doi: 10.1038/nrmicro2315.

Lachman, H. M. *et al.* (1996) 'Human catechol-O-methyltransferase pharmacogenetics: Description of a functional polymorphism and its potential application to neuropsychiatric disorders', *Pharmacogenetics*. doi: 10.1097/00008571-199606000-00007.

Lander, E. and Kruglyak, L. (1995) 'Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results', *Nature Genetics*. doi: 10.1038/ng1195-241.

Laurent, L. C. *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*. doi: 10.1016/j.stem.2010.12.003.

Leana-Cox, J. *et al.* (1996) 'Familial DiGeorge/velocardiofacial syndrome with deletions of chromosome area22q11.2: Report of five families with a review of the literature', *American Journal of Medical Genetics*. doi: 10.1002/(SICI)1096-8628(19961111)65:4<309::AID-AJMG12>3.0.CO;2-Y.

Lee, J. M. and Sonnhammer, E. L. L. (2003) 'Genomic gene clustering analysis of pathways in eukaryotes', *Genome Research*. doi: 10.1101/gr.737703.

Lee, P. *et al.* (2019) 'Genome wide meta-analysis identifies genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders', *bioRxiv*. doi: 10.1101/528117.

Lee, Y. *et al.* (2003) 'The nuclear RNase III Drosha initiates microRNA processing', *Nature*. doi: 10.1038/nature01957.

Lek, M. *et al.* (2016) 'Analysis of protein-coding genetic variation in 60,706 humans', *Nature*. Nature Publishing Group, 536(7616), pp. 285–291. doi: 10.1038/nature19057.

León, L. E. *et al.* (2017) 'Partial microduplication in the histone acetyltransferase complex member KANSL1 is associated with congenital heart defects in 22q11.2 microdeletion syndrome patients', *Scientific Reports*. doi: 10.1038/s41598-017-01896-w.

Li, D. and He, L. (2006) 'Association study of the G-protein signaling 4 (RGS4) and proline dehydrogenase (PRODH) genes with schizophrenia: A meta-analysis', *European Journal of Human Genetics*. doi: 10.1038/sj.ejhg.5201680.

Li, J. *et al.* (2019) 'Mitochondrial deficits in human iPSC-derived neurons from patients with 22q11.2 deletion syndrome and schizophrenia', *Translational Psychiatry*. doi: 10.1038/s41398-019-0643-y.

Li, Y. *et al.* (2011) 'Association study between GNB1L and three major mental disorders in Chinese Han populations', *Psychiatry Research*. doi: 10.1016/j.psychres.2010.04.019.

Li, Y. and Jiao, J. (2017) 'Histone chaperone HIRA regulates neural progenitor cell proliferation and neurogenesis via β-catenin.', *The Journal of cell biology*, 216(7), pp. 1975–1992. doi: 10.1083/jcb.201610014.

Lima, K. *et al.* (2011) 'Hypoparathyroidism and autoimmunity in the 22q11.2 deletion syndrome', *European Journal of Endocrinology*. doi: 10.1530/EJE-10-1206.

Lin, A. *et al.* (2017) 'Mapping 22q11.2 Gene Dosage Effects on Brain Morphometry', *The Journal of Neuroscience*, 37(26), pp. 6183–6199. doi: 10.1523/JNEUROSCI.3759-16.2017.

Lin, M. *et al.* (2016) 'Integrative transcriptome network analysis of iPSC-derived neurons from schizophrenia and schizoaffective disorder patients with 22q11.2 deletion', *BMC Systems Biology*. doi: 10.1186/s12918-016-0366-0.

Lindsay, E. A., Greenberg, F., *et al.* (1995) 'Submicroscopic deletions at 22q11.2: Variability of the clinical picture and delineation of a commonly deleted region', *American Journal of Medical Genetics*. doi: 10.1002/ajmg.1320560216.

Lindsay, E. A., Goldberg, R., *et al.* (1995) 'Velo-cardio-facial syndrome: frequency and extent of 22q11 deletions.', *American journal of medical genetics*. doi: 10.1002/ajmg.1320570339.

Lindsay, E. A. *et al.* (1999) 'Congenital heart disease in mice deficient for the DiGeorge syndrome region', *Nature*. doi: 10.1038/43900.

Liu, A. and Niswander, L. A. (2005) 'Bone morphogenetic protein signalling and vertebrate nervous system development', *Nature Reviews Neuroscience*. doi: 10.1038/nrn1805.

Liu, A. P. Y. *et al.* (2014) 'Under-recognition of 22q11.2 deletion in adult Chinese patients with conotruncal anomalies: Implications in transitional care', *European Journal of Medical Genetics*. doi: 10.1016/j.ejmg.2014.03.014.

Liu, Z. *et al.* (2017) 'Elevated p53 Activities Restrict Differentiation Potential of MicroRNA-Deficient Pluripotent Stem Cells', *Stem Cell Reports*, 9(5), pp. 1604–1617. doi: 10.1016/j.stemcr.2017.10.006.

Luhur, A. *et al.* (2014) 'Drosha-independent DGCR8/Pasha pathway regulates neuronal morphogenesis', *Proceedings of the National Academy of Sciences*. National Academy of Sciences, 111(4), pp. 1421–1426. doi: 10.1073/PNAS.1318445111.

Lupo, G., Harris, W. A. and Lewis, K. E. (2006) 'Mechanisms of ventral patterning in the vertebrate nervous system', *Nature Reviews Neuroscience*. doi: 10.1038/nrn1843.

Ma, C. *et al.* (2018) 'The integrated landscape of causal genes and pathways in schizophrenia', *Translational Psychiatry*. doi: 10.1038/s41398-018-0114-x.

Ma, S. *et al.* (2019) 'Gene modules associated with human diseases revealed by network analysis', *bioRxiv*. doi: 10.1101/598151.

Maclas, S. *et al.* (2012) 'DGCR8 HITS-CLIP reveals novel functions for the Microprocessor', *Nature Structural and Molecular Biology*. doi: 10.1038/nsmb.2344.

Maggio, I. *et al.* (2014) 'Adenoviral vector delivery of RNA-guided CRISPR/Cas9 nuclease complexes induces targeted mutagenesis in a diverse array of human cells', *Scientific Reports*. doi: 10.1038/srep05105.

Malatesta, P., Hartfuss, E. and Götz, M. (2000) 'Isolation of radial glial cells by fluorescent-activated cell sorting reveals a neural lineage', *Development*.

Mancuso, N. *et al.* (2017) 'Integrating Gene Expression with Summary Association Statistics to Identify Genes Associated with 30 Complex Traits', *American Journal of Human Genetics*. doi: 10.1016/j.ajhg.2017.01.031.

Marín, O. and Müller, U. (2014) 'Lineage origins of GABAergic versus glutamatergic neurons in the neocortex', *Current Opinion in Neurobiology*. doi: 10.1016/j.conb.2014.01.015.

Marinaro, F. *et al.* (2017) 'MicroRNA-independent functions of DGCR8 are essential for neocortical development and TBR1 expression', *EMBO reports*, 18(4), pp. 603–618. doi: 10.15252/embr.201642800.

Marshall, C. R. *et al.* (2017) 'Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects', *Nature Genetics*. doi: 10.1038/ng.3725.

Matsuoka, R. *et al.* (1998) 'Molecular and clinical study of 183 patients with conotruncal anomaly face syndrome', *Human Genetics*. doi: 10.1007/s004390050786.

Maurano, M. T. *et al.* (2012) 'Systematic localization of common disease-associated variation in regulatory DNA', *Science*. doi: 10.1126/science.1222794.

Maynard, T. M. *et al.* (2003) 'A comprehensive analysis of 22q11 gene expression in the developing and adult brain.', *Proceedings of the National Academy of Sciences of the United States of America*, 100(24), pp. 14433–8. doi: 10.1073/pnas.2235651100.

Maynard, T. M. *et al.* (2008) 'Mitochondrial localization and function of a subset of 22q11 deletion syndrome candidate genes', *Molecular and Cellular Neuroscience*. doi: 10.1016/j.mcn.2008.07.027.

McCarroll, S. A. (2008) 'Extending genome-wide association studies to copy-number variation', *Human Molecular Genetics*. doi: 10.1093/hmg/ddn282.

McDonald-McGinn, D. M. *et al.* (1997) 'The 22q11.2 deletion: Screening, diagnostic workup, and outcome of results; report on 181 patients', *Genetic Testing*. doi: 10.1089/gte.1997.1.99.

McDonald-McGinn, D. M. *et al.* (1999) 'The Philadelphia story: The 22q11.2 deletion: Report on 250 patients', in *Genetic Counseling*.

McDonald-McGinn, D. M. *et al.* (2001) 'Phenotype of the 22q11.2 deletion in individuals identified through an affected relative: Cast a wide FISHing net!', in *Genetics in Medicine*. doi: 10.1097/00125817-200101000-00006.

McDonald-McGinn, D. M. *et al.* (2005) 'The 22q11.2 deletion in African-American patients: an underdiagnosed population?', *American journal of medical genetics. Part A*, 134(3), pp. 242–246. doi: 10.1002/ajmg.a.30069.

McDonald-McGinn, D. M. *et al.* (2015) '22q11.2 deletion syndrome', *Nature Reviews Disease Primers*. doi: 10.1038/nrdp.2015.71.

McDonald-Mcginn, D. M. and Sullivan, K. E. (2011) 'Chromosome 22q11.2 deletion syndrome (DiGeorge syndrome/velocardiofacial syndrome)', *Medicine*. doi: 10.1097/MD.0b013e3182060469.

McKenna, W. L. *et al.* (2011) 'Tbr1 and Fezf2 regulate alternate corticofugal neuronal identities during neocortical development', *Journal of Neuroscience*. doi: 10.1523/JNEUROSCI.4131-10.2011.

Mclean, M. S. D. *et al.* (1993) 'Velo-cardio-facial Syndrome: Intrafamilial Variability of the Phenotype', *American Journal of Diseases of Children*. doi: 10.1001/archpedi.1993.02160350086013.

Meechan, D. W. *et al.* (2006) 'Gene dosage in the developing and adult brain in a mouse model of 22q11 deletion syndrome', *Molecular and Cellular Neuroscience*, 33(4), pp. 412–428. doi: 10.1016/j.mcn.2006.09.001.

Meechan, D. W. *et al.* (2007) 'When half is not enough: Gene expression and dosage in the 22q11 Deletion syndrome', *Gene Expression*. doi: 10.3727/000000006781510697.

Meechan, D. W. *et al.* (2009) 'Diminished dosage of 22q11 genes disrupts neurogenesis and cortical development in a mouse model of 22q11 deletion/DiGeorge syndrome', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.0905696106.

Merico, D. *et al.* (2015) 'Whole-genome sequencing suggests schizophrenia risk mechanisms in humans with 22q11.2 deletion syndrome', *G3: Genes, Genomes, Genetics*. doi: 10.1534/g3.115.021345.

Merscher, S. *et al.* (2001) 'TBX1 is responsible for cardiovascular defects in velo-cardio-facial/DiGeorge syndrome', *Cell*. doi: 10.1016/S0092-8674(01)00247-1.

Mertens, J. *et al.* (2016) 'Evaluating cell reprogramming, differentiation and conversion technologies in neuroscience', *Nature Reviews Neuroscience*. doi: 10.1038/nrn.2016.46.

Michaelovsky, E. *et al.* (2012) 'Genotype-phenotype correlation in 22q11.2 deletion syndrome', *BMC Medical Genetics*, 13(1), p. 122. doi: 10.1186/1471-2350-13-122.

Michaelovsky, E. *et al.* (2019) 'Risk gene-set and pathways in 22q11.2 deletion-related schizophrenia: a genealogical molecular approach', *Translational Psychiatry*. doi: 10.1038/s41398-018-0354-9.

Michalak, P. (2008) 'Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes', *Genomics*. doi: 10.1016/j.ygeno.2007.11.002.

Miyata, T. *et al.* (2004) 'Asymmetric production of surface-dividing and non-surface-dividing cortical progenitor cells', *Development*. doi: 10.1242/dev.01173.

Miyata, T. (2008) 'Development of three-dimensional architecture of the neuroepithelium: Role of pseudostratification and cellular "community"', *Development Growth and Differentiation*. doi: 10.1111/j.1440-169X.2007.00980.x.

Mlynarski, E. E. *et al.* (2015) 'Copy-number variation of the glucose transporter gene SLC2A3 and congenital heart defects in the 22q11.2 deletion syndrome', *American Journal of Human Genetics*. doi: 10.1016/j.ajhg.2015.03.007.

Mok, K. Y. *et al.* (2016) 'Deletions at 22q11.2 in idiopathic Parkinson's disease: A combined analysis of genome-wide association data', *The Lancet Neurology*. doi: 10.1016/S1474-4422(16)00071-5.

Molck, M. C. *et al.* (2013) 'Atypical copy number abnormalities in 22q11.2 region: Report of three cases', *European Journal of Medical Genetics*. doi: 10.1016/j.ejmg.2013.07.002.

Molnár, Z. *et al.* (2019) 'New insights into the development of the human cerebral cortex', *Journal of Anatomy*. doi: 10.1111/joa.13055.

Molyneaux, B. J. *et al.* (2007) 'Neuronal subtype specification in the cerebral cortex', *Nature Reviews Neuroscience*. doi: 10.1038/nrn2151.

Momma, K. *et al.* (1996) 'Cardiac anomalies associated with a chromosome 22q11 deletion in patients with conotruncal anomaly face syndrome', *American Journal of Cardiology*. doi: 10.1016/S0002-9149(96)00374-8.

Momma, K. (2007) 'Cardiovascular anomalies associated with chromosome 22q11.2 deletion', *International Journal of Cardiology*. doi: 10.1016/j.ijcard.2006.02.002.

Momma, K. (2010) 'Cardiovascular Anomalies Associated With Chromosome 22q11.2 Deletion Syndrome', *American Journal of Cardiology*. doi: 10.1016/j.amjcard.2010.01.333.

Morgan, C. *et al.* (2014) 'Reappraising the long-term course and outcome of psychotic disorders: The AESOP-10 study', *Psychological Medicine*. doi: 10.1017/S0033291714000282.

Morrow, B. E. *et al.* (2018) 'Molecular genetics of 22q11.2 deletion syndrome', *American Journal of Medical Genetics, Part A*, 176(10), pp. 2070–2081. doi: 10.1002/ajmg.a.40504.

Moss, E. M. *et al.* (1999) 'Psychoeducational profile of the 22q11.2 microdeletion: A complex pattern', *Journal of Pediatrics*. doi: 10.1016/S0022-3476(99)70415-4.

Motahari, Z. *et al.* (2019) 'In the line-up: Deleted genes associated with DiGeorge/22q11.2 deletion syndrome: Are they all suspects?', *Journal of Neurodevelopmental Disorders*. doi: 10.1186/s11689-019-9267-z.

Moutin, E. *et al.* (2017) 'Palmitoylation of cdc42 Promotes Spine Stabilization and Rescues Spine Density Deficit in a Mouse Model of 22q11.2 Deletion Syndrome', *Cerebral Cortex*. doi: 10.1093/cercor/bhw183.

Mukai, J. *et al.* (2004) 'Evidence that the gene encoding ZDHHC8 contributes to the risk of schizophrenia', *Nature Genetics*. doi: 10.1038/ng1375.

Mukai, J. *et al.* (2008) 'Palmitoylation-dependent neurodevelopmental deficits in a mouse model of 22q11 microdeletion', *Nature Neuroscience*. doi: 10.1038/nn.2204.

Mukai, J. *et al.* (2015) 'Molecular Substrates of Altered Axonal Growth and Brain Connectivity in a Mouse Model of Schizophrenia', *Neuron*. doi: 10.1016/j.neuron.2015.04.003.

Murphy, K. C., Jones, L. A. and Owen, M. J. (1999) 'High rates of schizophrenia in adults with velo-cardio-facial syndrome', *Archives of General Psychiatry*. doi: 10.1001/archpsyc.56.10.940.

Murray, R. M., Lewis, S. W. and Lecturer, L. (1987) 'Is schizophrenia a neurodevelopmental disorder?', *British Medical Journal (Clinical research ed.)*. doi: 10.1136/bmj.295.6600.681.

Musunuru, K. (2013) 'Genome editing of human pluripotent stem cells to generate human cellular disease models', *DMM Disease Models and Mechanisms*. doi: 10.1242/dmm.012054.

Naldini, L. *et al.* (1996) 'In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector', *Science*. doi: 10.1126/science.272.5259.263.

Nashun, B. *et al.* (2015) 'Continuous Histone Replacement by Hira Is Essential for Normal Transcriptional Regulation and De Novo DNA Methylation during Mouse Oogenesis', *Molecular Cell*. Elsevier, 60(4), pp. 611–625. doi: 10.1016/j.molcel.2015.10.010.

Negi, S. K. and Guda, C. (2017) 'Global gene expression profiling of healthy human brain and its application in studying neurological disorders', *Scientific Reports*. doi: 10.1038/s41598-017-00952-9.

Ng, M. Y. M. *et al.* (2009) 'Meta-analysis of 32 genome-wide linkage studies of schizophrenia', *Molecular Psychiatry*. doi: 10.1038/mp.2008.135.

Nicholson, L. *et al.* (2008) 'Spatial and temporal control of gene expression in drosophila using the inducible geneSwitch GAL4 system. I. Screen for larval nervous system drivers', *Genetics*. doi: 10.1534/genetics.107.081968.

Nicolae, D. L. *et al.* (2010) 'Trait-associated SNPs are more likely to be eQTLs: Annotation to enhance discovery from GWAS', *PLoS Genetics*, 6(4). doi: 10.1371/journal.pgen.1000888.

Niklasson, L. *et al.* (2001) 'Neuropsychiatric disorders in the 22q11 deletion syndrome', in *Genetics in Medicine*. doi: 10.1097/00125817-200101000-00017.

Niklasson, L. *et al.* (2007) 'Chromosome 22q11 deletion syndrome (CATCH 22): neuropsychiatric and neuropsychological aspects', *Developmental Medicine & Child Neurology*. doi: 10.1111/j.1469-8749.2002.tb00258.x.

Niklasson, L. *et al.* (2009) 'Autism, ADHD, mental retardation and behavior problems in 100 individuals with 22q11 deletion syndrome', *Research in Developmental Disabilities*. doi: 10.1016/j.ridd.2008.10.007.

Noctor, S. C. *et al.* (2004) 'Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases', *Nature Neuroscience*. doi: 10.1038/nn1172.

Noctor, S. C., Martínez-Cerdeño, V. and Kriegstein, A. R. (2008) 'Distinct behaviors of neural stem and progenitor cells underlie cortical neurogenesis', *Journal of Comparative Neurology*. doi: 10.1002/cne.21669.

Nuotio, J. *et al.* (2014) 'Cardiovascular risk factors in 2011 and secular trends since 2007: The Cardiovascular Risk in Young Finns Study', *Scandinavian Journal of Public Health*. doi: 10.1177/1403494814541597.

O'Brien, H. E. *et al.* (2018) 'Expression quantitative trait loci in the developing human brain and their enrichment in neuropsychiatric disorders', *Genome biology*. doi: 10.1186/s13059-018-1567-1.

Ohgami, R. S. *et al.* (2006) 'The Steap proteins are metalloreductases', *Blood*. doi: 10.1182/blood-2006-02-003681.

Ohta, H. *et al.* (2016) 'Increased Surface Area, but not Cortical Thickness, in a Subset of Young Boys With Autism Spectrum Disorder', *Autism Research*. doi: 10.1002/aur.1520.

Ouchi, Y. *et al.* (2013) 'Reduced Adult Hippocampal Neurogenesis and Working Memory Deficits in the Dgcr8-Deficient Mouse Model of 22q11.2 Deletion-Associated Schizophrenia Can Be Rescued by IGF2', *Journal of Neuroscience*. doi: 10.1523/jneurosci.2700-12.2013.

Pakkenberg, B. (1990) 'Pronounced Reduction of Total Neuron Number in Mediodorsal Thalamic Nucleus and Nucleus Accumbens in Schizophrenics', *Archives of General Psychiatry*. doi: 10.1001/archpsyc.1990.01810230039007.

Panamonta, V. *et al.* (2016) 'Birth Prevalence of Chromosome 22q11.2 Deletion Syndrome: A Systematic Review of Population-Based Studies', *Journal of the Medical Association of Thailand = Chotmaihet thangphaet*.

Pardiñas, A. F. *et al.* (2018) 'Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection', *Nature Genetics*, 50(3), pp. 381–389. doi: 10.1038/s41588-018-0059-2.

Parrini, E. *et al.* (2016) 'Genetic basis of brain malformations', *Molecular Syndromology*. doi: 10.1159/000448639.

Partin, A. C. *et al.* (2018) 'Publisher Correction: Heme enables proper positioning of Drosha and DGCR8 on primary microRNAs', *Nature Communications*. doi: 10.1038/s41467-018-06426-4.

Paterlini, M. *et al.* (2005) 'Transcriptional and behavioral interaction between 22q11.2 orthologs modulates schizophrenia-related phenotypes in mice', *Nature Neuroscience*. doi: 10.1038/nn1562.

Paylor, R. (2001) 'Mice deleted for the DiGeorge/velocardiofacial syndrome region show abnormal sensorimotor gating and learning and memory impairments', *Human Molecular Genetics*. doi: 10.1093/hmg/10.23.2645.

Paylor, R. *et al.* (2006) 'Tbx1 haploinsufficiency is linked to behavioral disorders in mice and humans: Implications for 22q11 deletion syndrome', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.0600206103.

Piccolo, S. *et al.* (1996) 'Dorsoventral patterning in Xenopus: Inhibition of ventral signals by direct binding of chordin to BMP-4', *Cell*. doi: 10.1016/S0092-8674(00)80132-4.

Pickar-Oliver, A. and Gersbach, C. A. (2019) 'The next generation of CRISPR–Cas technologies and applications', *Nature Reviews Molecular Cell Biology*. doi: 10.1038/s41580-019-0131-5.

Piña, B. and Suau, P. (1987) 'Changes in histones H2A and H3 variant composition in differentiating and mature rat brain cortical neurons', *Developmental Biology*. doi: 10.1016/0012-1606(87)90426-X.

Polo, J. M. *et al.* (2010) 'Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells', *Nature Biotechnology*. doi: 10.1038/nbt.1667.

Pritchard, J. K. and Cox, N. J. (2002) 'The allelic architecture of human disease genes: Common disease - Common variant... or not?', *Human Molecular Genetics*. doi: 10.1093/hmg/11.20.2417.

Prives, C. and Hall, P. A. (1999) 'The P53 pathway', *Journal of Pathology*. doi: 10.1002/(SICI)1096-9896(199901)187:1<112::AID-PATH250>3.0.CO;2-3.

Puech, A. *et al.* (1997) 'Comparative mapping of the human 22q11 chromosomal region and the orthologous region in mice reveals complex changes in gene organization', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.94.26.14608.

Puttonen, K. A. *et al.* (2013) 'Improved method of producing human neural progenitor cells of high purity and in large quantities from pluripotent stem cells for transplantation studies', *Cell Transplantation*. doi: 10.3727/096368912X658764.

Qi, L. S. *et al.* (2013) 'Repurposing CRISPR as an RNA-γuided platform for sequence-specific control of gene expression', *Cell*. doi: 10.1016/j.cell.2013.02.022.

Quick-Cleveland, J. *et al.* (2014) 'The DGCR8 RNA-Binding Heme Domain Recognizes Primary MicroRNAs by Clamping the Hairpin', *Cell Reports*. doi: 10.1016/j.celrep.2014.05.013.

Quinonez, R. and Sutton, R. E. (2002) 'Lentiviral vectors for gene delivery into cells', *DNA and Cell Biology*. doi: 10.1089/104454902762053873.

Racedo, S. E. *et al.* (2015) 'Mouse and human CRKL is dosage sensitive for cardiac outflow tract formation', *American Journal of Human Genetics*. doi: 10.1016/j.ajhg.2014.12.025.

Raitakari, O. T. *et al.* (2008) 'Cohort profile: The cardiovascular risk in young Finns study', *International Journal of Epidemiology*. doi: 10.1093/ije/dym225.

Rakic, P. (1995) 'A small step for the cell, a giant leap for mankind: a hypothesis of neocortical expansion during evolution', *Trends in Neurosciences*. doi: 10.1016/0166-2236(95)93934-P.

Rauch, A. *et al.* (2004) 'Assessment of association between variants and haplotypes of the remaining TBX1 gene and manifestations of congenital heart defects in 22q11.2 deletion patients.', *Journal of medical genetics*. doi: 10.1136/jmg.2003.010975.

Rauch, A. *et al.* (2005) 'Systematic assessment of atypical deletions reveals genotype-phenotype correlation in 22q11.2', *Journal of Medical Genetics*. doi: 10.1136/jmg.2004.030619.

Raux, G. *et al.* (2007) 'Involvement of hyperprolinemia in cognitive and psychiatric features of the 22q11 deletion syndrome', *Human Molecular Genetics*. doi: 10.1093/hmg/ddl443.

Ray-Gallet, D. *et al.* (2002) 'HIRA is critical for a nucleosome assembly pathway independent of DNA synthesis', *Molecular Cell*. doi: 10.1016/S1097-2765(02)00526-9.

Rees, E., O'Donovan, M. C. and Owen, M. J. (2015) 'Genetics of schizophrenia', *Current Opinion in Behavioral Sciences*. doi: 10.1016/j.cobeha.2014.07.001.

Reich, D. E., Gabriel, S. B. and Altshuler, D. (2003) 'Quality and completeness of SNP databases', *Nature Genetics*. doi: 10.1038/ng1133.

Reich, D. E. and Lander, E. S. (2001) 'On the allelic spectrum of human disease', *Trends in Genetics*. doi: 10.1016/S0168-9525(01)02410-6.

Repetto, G. M. *et al.* (2014) 'Case fatality rate and associated factors in patients with 22q11 microdeletion syndrome: A retrospective cohort study', *BMJ Open*. doi: 10.1136/bmjopen-2014-005041.

Ricketts, M. D. and Marmorstein, R. (2017) 'A Molecular Prospective for HIRA Complex Assembly and H3.3-Specific Histone Chaperone Function', *Journal of Molecular Biology*. Elsevier Ltd, 429(13), pp. 1924–1933. doi: 10.1016/j.jmb.2016.11.010.

Roberts, C. *et al.* (2002) 'Targeted Mutagenesis of the Hira Gene Results in Gastrulation Defects and Patterning Abnormalities of Mesoendodermal Derivatives Prior to Early Embryonic Lethality', *Molecular and Cellular Biology*. doi: 10.1128/mcb.22.7.2318-2328.2002.

Rodgers, K. and Mcvey, M. (2016) 'Error-Prone Repair of DNA Double-Strand Breaks', *Journal of Cellular Physiology*. doi: 10.1002/jcp.25053.

Rogdaki, M. *et al.* (2020) 'Magnitude and heterogeneity of brain structural abnormalities in 22q11.2 deletion syndrome: a meta-analysis', *Molecular Psychiatry*. doi: 10.1038/s41380-019-0638-3.

Rozas, M. F. *et al.* (2019) 'Association between phenotype and deletion size in 22q11.2 microdeletion syndrome: Systematic review and meta-analysis', *Orphanet Journal of Rare Diseases*. doi: 10.1186/s13023-019-1170-x.

Rubio, A. *et al.* (2016) 'Rapid and efficient CRISPR/Cas9 gene inactivation in human neurons during human pluripotent stem cell differentiation and direct reprogramming', *Scientific Reports*. The Author(s), 6, p. 37540. Available at: https://doi.org/10.1038/srep37540.

Rump, P. *et al.* (2014) 'Central 22q11.2 deletions', *American Journal of Medical Genetics, Part A*. doi: 10.1002/ajmg.a.36711.

Ryan, A. K. *et al.* (1997) 'Spectrum of clinical features associated with interstitial chromosome 22q11 deletions: A European collaborative study', *Journal of Medical Genetics*. doi: 10.1136/jmg.34.10.798.

Sadegh, C. and Macklis, J. D. (2014) 'Established monolayer differentiation of mouse embryonic stem cells generates heterogeneous neocortical-like neurons stalled at a stage equivalent to midcorticogenesis', *Journal of Comparative Neurology*. doi: 10.1002/cne.23576.

Saha, S. *et al.* (2005) 'A systematic review of the prevalence of schizophrenia', *PLoS Medicine*. doi: 10.1371/journal.pmed.0020141.

Sahara, S. and O'Leary, D. D. M. (2009) 'Fgf10 regulates transition period of cortical stem cell differentiation to radial glia controlling generation of neurons and basal progenitors.', *Neuron*. doi: 10.1016/j.neuron.2009.06.006.

Sanders, A. R. *et al.* (2008) 'No significant association of 14 candidate genes with schizophrenia in a large European ancestry sample: Implications for psychiatric genetics', *American Journal of Psychiatry*. doi: 10.1176/appi.ajp.2007.07101573.

Sanders, S. S. *et al.* (2015) 'Curation of the Mammalian Palmitoylome Indicates a Pivotal Role for Palmitoylation in Diseases and Disorders of the Nervous System and Cancers', *PLoS Computational Biology*. doi: 10.1371/journal.pcbi.1004405.

Scambler, P. J. *et al.* (1992) 'Velo-cardio-facial syndrome associated with chromosome 22 deletions encompassing the DiGeorge locus', *The Lancet*. doi: 10.1016/0140-6736(92)90734-K.

Scarr, E., Udawela, M. and Dean, B. (2018) 'Changed frontal pole gene expression suggest altered interplay between neurotransmitter, developmental, and inflammatory pathways in schizophrenia', *npj Schizophrenia*. doi: 10.1038/s41537-018-0044-x.

Schaub, M. A. *et al.* (2012) 'Linking disease associations with regulatory information in the human genome', *Genome Research*. doi: 10.1101/gr.136127.111.

Scherer, N. J., D'Antonio, L. L. and Kalbfleisch, J. H. (1999) 'Early speech and language development in children with velocardiofacial syndrome', *American Journal of Medical Genetics - Neuropsychiatric Genetics*. doi: 10.1002/(SICI)1096-8628(19991215)88:6<714::AID-AJMG24>3.0.CO;2-B.

Schneider, M., Debbané, M., Bassett, Anne S, *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.', *The American journal of psychiatry*, 171(6), pp. 627–639. doi: 10.1176/appi.ajp.2013.13070864.

Schneider, M., Debbané, M., Bassett, Anne S., *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', *American Journal of Psychiatry*. doi: 10.1176/appi.ajp.2013.13070864.

Schofield, C. M. *et al.* (2011) 'Monoallelic deletion of the microRNA biogenesis gene Dgcr8 produces deficits in the development of excitatory synaptic transmission in the prefrontal cortex', *Neural Development*. doi: 10.1186/1749-8104-6-11.

Schwarz, D. S. *et al.* (2003) 'Asymmetry in the assembly of the RNAi enzyme complex', *Cell*. doi: 10.1016/S0092-8674(03)00759-1.

Sellier, C. *et al.* (2014) 'Decreased DGCR8 expression and miRNA dysregulation in individuals with 22q11.2 deletion syndrome', *PLoS ONE*, 9(8). doi: 10.1371/journal.pone.0103884.

Seong, Y. *et al.* (2014) 'Global identification of target recognition and cleavage by the Microprocessor in human ES cells', *Nucleic acids research*. doi: 10.1093/nar/gku957.

Shaikh, T. H. (2000) 'Chromosome 22-specific low copy repeats and the 22q11.2 deletion syndrome: genomic organization and deletion endpoint analysis', *Human Molecular Genetics*. doi: 10.1093/hmg/9.4.489.

Shalem, O. *et al.* (2014) 'Genome-scale CRISPR-Cas9 knockout screening in human cells', *Science*. doi: 10.1126/science.1247005.

Shi, Y. *et al.* (2012) 'Human cerebral cortex development from pluripotent stem cells to functional excitatory synapses', *Nature Neuroscience*. doi: 10.1038/nn.3041.

Shiohama, A. *et al.* (2003) 'Molecular cloning and expression analysis of a novel gene DGCR8 located in the DiGeorge syndrome chromosomal region', *Biochemical and Biophysical Research Communications*. doi: 10.1016/S0006-291X(03)00554-0.

Shiohama, A. *et al.* (2007) 'Nucleolar localization of DGCR8 and identification of eleven DGCR8-associated proteins', *Experimental Cell Research*. doi: 10.1016/j.yexcr.2007.07.020.

Shprintzen, R. J. *et al.* (1981) 'The velo-cardio-facial syndrome: A clinical and genetic analysis', *Pediatrics*.

Shprintzen, R. J. (2008) 'Velo-cardio-facial syndrome: 30 Years of study', *Developmental Disabilities Research Reviews*. doi: 10.1002/ddrr.2.

Simon, T. J. *et al.* (2005) 'Volumetric, connective, and morphologic changes in the brains of children with chromosome 22q11.2 deletion syndrome: An integrative study', *NeuroImage*. doi: 10.1016/j.neuroimage.2004.11.018.

Singer, G. A. C. *et al.* (2005) 'Clusters of co-expressed genes in mammalian genomes are conserved by natural selection', *Molecular Biology and Evolution*. doi: 10.1093/molbev/msi062.

De Smedt, B. *et al.* (2007) 'Intellectual abilities in a large sample of children with Velo-Cardio-Facial Syndrome: An update', *Journal of Intellectual Disability Research*. doi: 10.1111/j.1365-2788.2007.00955.x.

De Smedt, B. *et al.* (2009) 'Mathematical learning disabilities in children with 22q11.2 deletion syndrome: A review', *Developmental Disabilities Research Reviews*. doi: 10.1002/ddrr.44.

Sobin, C., Kiley-Brabeck, K. and Karayiorgou, M. (2005) 'Lower prepulse inhibition in children with the 22q11 deletion syndrome', *American Journal of Psychiatry*. doi: 10.1176/appi.ajp.162.6.1090.

Soliman, M. A. *et al.* (2017) 'Pluripotent stem cells in neuropsychiatric disorders', *Molecular Psychiatry*. doi: 10.1038/mp.2017.40.

Solot, C. B. *et al.* (2000) 'Communication disorders in the 22Q11.2 microdeletion syndrome', *Journal of Communication Disorders*. doi: 10.1016/S0021-9924(00)00018-6.

Sørensen, K. M. *et al.* (2010) 'Detecting 22q11.2 deletions by use of multiplex ligation-dependent probe amplification on DNA from neonatal dried blood spot samples', *Journal of Molecular Diagnostics*. doi: 10.2353/jmoldx.2010.090099.

Stark, K. L. *et al.* (2008) 'Altered brain microRNA biogenesis contributes to phenotypic deficits in a 22q11-deletion mouse model', *Nature Genetics*. doi: 10.1038/ng.138.

Stark, K. L. *et al.* (2009) 'Analysis of prepulse inhibition in mouse lines overexpressing 22q11.2 orthologues', *International Journal of Neuropsychopharmacology*. doi: 10.1017/S1461145709000492.

Stefansson, H. *et al.* (2002) 'Neuregulin 1 and Susceptibility to Schizophrenia', *American Journal of Human Genetics*. doi: 10.1086/342734.

Straub, R. E. *et al.* (2002) 'Genetic variation in the 6p22.3 Gene DTNBP1, the human ortholog of the mouse dysbindin gene, is associated with schizophrenia', *American Journal of Human Genetics*. doi: 10.1086/341750.

Sullivan, K. E. *et al.* (1998) 'Lack of correlation between impaired T cell production, immunodeficiency, and other phenotypic features in chromosome 22q11.2 deletion syndromes (DiGeorge syndrome/velocardiofacial syndrome)', *Clinical Immunology and Immunopathology*. doi: 10.1006/clin.1997.4463.

Sullivan, K. E., McDonald-McGinn, D. and Zackai, E. H. (2002) 'CD4+ CD25+ T-cell production in healthy humans and in patients with thymic hypoplasia', *Clinical and Diagnostic Laboratory Immunology*. doi: 10.1128/CDLI.9.5.1129-1131.2002.

Sullivan, P. F. *et al.* (2018) 'Psychiatric genomics: An update and an Agenda', *American Journal of Psychiatry*. doi: 10.1176/appi.ajp.2017.17030283.

Sullivan, P. F., Kendler, K. S. and Neale, M. C. (2003) 'Schizophrenia as a Complex Trait: Evidence from a Meta-analysis of Twin Studies', *Archives of General Psychiatry*. doi: 10.1001/archpsyc.60.12.1187.

Sun, D. *et al.* (2018) 'Large-scale mapping of cortical alterations in 22q11.2 deletion syndrome: Convergence with idiopathic psychosis and effects of deletion size', *Molecular Psychiatry*. Springer US, pp. 1–13. doi: 10.1038/s41380-018-0078-5.

Sur, M. and Rubenstein, J. L. R. (2005) 'Patterning and plasticity of the cerebral cortex', *Science*. doi: 10.1126/science.1112070.

Susser, E. *et al.* (1996) 'Schizophrenia after prenatal famine further evidence', *Archives of General Psychiatry*. doi: 10.1001/archpsyc.1996.01830010027005.

Swillen, A. *et al.* (1997) 'Intelligence and psychosocial adjustment in velocardiofacial syndrome: A study of 37 children and adolescents with VCFS', *Journal of Medical Genetics*.

Swillen, A. *et al.* (1999) 'Neuropsychological, learning and psychosocial profile of primary school aged children with the velo-cardio-facial syndrome (22q11 deletion): Evidence for a nonverbal learning disability?', *Child Neuropsychology*. doi: 10.1076/0929-7049(199912)05:04;1-r;ft230.

Swillen, A. *et al.* (2005) 'Early motor development in young children with 22q.11 deletion syndrome and a conotruncal heart defect', *Developmental Medicine and Child Neurology*. doi: 10.1017/S0012162205001696.

Tagami, H. *et al.* (2004) 'Histone H3.1 and H3.3 Complexes Mediate Nucleosome Assembly Pathways Dependent or Independent of DNA Synthesis', *Cell*. doi: 10.1016/S0092-8674(03)01064-X.

Takahashi, K. *et al.* (2007) 'Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors', *Cell*. doi: 10.1016/j.cell.2007.11.019.

Takahashi, K. and Yamanaka, S. (2006) 'Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors', *Cell*. doi: 10.1016/j.cell.2006.07.024.

Tansey, K. E. *et al.* (2016) 'Common alleles contribute to schizophrenia in CNV carriers', *Molecular Psychiatry*. doi: 10.1038/mp.2015.143.

Taverna, E. and Huttner, W. B. (2010) 'Neural progenitor nuclei IN motion', *Neuron*. doi: 10.1016/j.neuron.2010.08.027.

Tenhunen, J. *et al.* (1994) 'Genomic organization of the human catechol O-methyltransferase gene and its expression from two distinct promoters', *European Journal of Biochemistry*. doi: 10.1111/j.1432-1033.1994.tb19083.x.

Thomson, J. A. (1998) 'Embryonic stem cell lines derived from human blastocysts', *Science*. doi: 10.1126/science.282.5391.1145.

Tiberi, L., Vanderhaeghen, P. and van den Ameele, J. (2012) 'Cortical neurogenesis and morphogens: Diversity of cues, sources and functions', *Current Opinion in Cell Biology*. doi: 10.1016/j.ceb.2012.01.010.

Tomita-Mitchell, A. *et al.* (2010) 'Multiplexed quantitative real-time PCR to detect 22q11.2 deletion in patients with congenital heart disease', *Physiological Genomics*. doi: 10.1152/physiolgenomics.00073.2010.

Torres-Juan, L. *et al.* (2007) 'Analysis of meiotic recombination in 22q11.2, a region that frequently undergoes deletions and duplications', *BMC Medical Genetics*. doi: 10.1186/1471-2350-8-14.

Toyoshima, M. *et al.* (2016) 'Analysis of induced pluripotent stem cells carrying 22q11.2 deletion', *Translational Psychiatry*. doi: 10.1038/tp.2016.206.

Triboulet, R. *et al.* (2009) 'Post-transcriptional control of DGCR8 expression by the Microprocessor', *RNA*. doi: 10.1261/rna.1591709.

Valenzuela, N. *et al.* (2016) 'Cardiomyocyte-specific conditional knockout of the histone chaperone HIRA in mice results in hypertrophy, sarcolemmal damage and focal replacement fibrosis', *DMM Disease Models and Mechanisms*. doi: 10.1242/dmm.022889.

Vassos, E. *et al.* (2012) 'Meta-analysis of the association of urbanicity with schizophrenia', *Schizophrenia Bulletin*. doi: 10.1093/schbul/sbs096.

Veturi, Y. and Ritchie, M. D. (2018) 'How powerful are summary-based methods for identifying expression-trait associations under different genetic architectures?', in *Pacific Symposium on Biocomputing*. doi: 10.1142/9789813235533_0021.

Visscher, P. M. *et al.* (2017) '10 Years of GWAS Discovery: Biology, Function, and Translation', *American Journal of Human Genetics*. doi: 10.1016/j.ajhg.2017.06.005.

Vitelli, F. *et al.* (2002) 'Tbx1 mutation causes multiple cardiovascular defects and disrupts neural crest and cranial nerve migratory pathways.', *Human molecular genetics*.

Vorstman, J. A. S. *et al.* (2006) 'MLPA: A rapid, reliable, and sensitive method for detection and analysis of abnormalities of 22q', *Human Mutation*. doi: 10.1002/humu.20330.

Vroemen, M., Weidner, N. and Blesch, A. (2005) 'Loss of gene expression in lentivirus- and retrovirus-transduced neural progenitor cells is correlated to migration and differentiation in the adult spinal cord', *Experimental Neurology*. doi: 10.1016/j.expneurol.2005.04.012.

Wang, Y. *et al.* (2007) 'DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal', *Nature Genetics*. doi: 10.1038/ng1969.

Warnica, W. *et al.* (2015) 'Copy number variable microrna in schizophrenia and their neurodevelopmental gene targets', *Biological Psychiatry*. doi: 10.1016/j.biopsych.2014.05.011.

Warren, N. *et al.* (1999) 'The transcription factor, Pax6, is required for cell proliferation and differentiation in the developing cerebral cortex', *Cerebral Cortex*. doi: 10.1093/cercor/9.6.627.

Wei, L. *et al.* (2019) 'Genome-wide CRISPR/Cas9 library screening identified PHGDH as a critical driver for Sorafenib resistance in HCC', *Nature Communications*. doi: 10.1038/s41467-019-12606-7.

Wei, T. and Simko, V. (2017) 'Package "corrplot: visualization of a correlation matrix" (v.0.84)', *URL https://CRAN.R-project.org/package=corrplot*.

Weinberger, D. R. (1986) 'The pathogenesis of schizophrenia: a neurodevelopmental theory', in *The Neurology of Schizophrenia*.

Weinberger, D. R. (1987) 'Implications of Normal Brain Development for the Pathogenesis of Schizophrenia', *Archives of General Psychiatry*. doi: 10.1001/archpsyc.1987.01800190080012.

Weinzimer, S. A. *et al.* (1998) 'Growth hormone deficiency in patients with a 22q11.2 deletion: Expanding the phenotype', *Pediatrics*. doi: 10.1542/peds.101.5.929.

Weksberg, R. *et al.* (2005) 'A method for accurate detection of genomic microdeletions using real-time quantitative PCR', *BMC Genomics*. doi: 10.1186/1471-2164-6-180.

Wen, Z. *et al.* (2014) 'Synaptic dysregulation in a human iPS cell model of mental disorders', *Nature*. doi: 10.1038/nature13716.

Williams, H. J. *et al.* (2013) 'Schizophrenia two-hit hypothesis in velo-cardio facial syndrome', *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*. doi: 10.1002/ajmg.b.32129.

Williams, N. M. *et al.* (2008) 'Strong evidence that GNB1L is associated with schizophrenia', *Human Molecular Genetics*. doi: 10.1093/hmg/ddm330.

Williams, N. M. (2011) 'Molecular mechanisms in 22q11 deletion syndrome', *Schizophrenia Bulletin*, 37(5), pp. 882–889. doi: 10.1093/schbul/sbr095.

Willis, A. *et al.* (2008) 'PRODH variants and risk for schizophrenia', *Amino Acids*. doi: 10.1007/s00726-008-0111-0.

Wilson, P. A. and Hemmati-Brivanlou, A. (1995) 'Induction of epidermis and inhibition of neural fate by Bmp-4', *Nature*. doi: 10.1038/376331a0.

Woo, Y. H., Walker, M. and Churchill, G. A. (2010) 'Coordinated expression domains in mammalian genomes.', *PloS one*, 5(8), p. e12158. doi: 10.1371/journal.pone.0012158.

Wu, P., Ye, Y. and Svendsen, C. N. (2002) 'Transduction of human neural progenitor cells using recombinant adeno-associated viral vectors', *Gene Therapy*. doi: 10.1038/sj.gt.3301646.

Xu, B. *et al.* (2013) 'Derepression of a neuronal inhibitor due to miRNA dysregulation in a schizophrenia-related microdeletion', *Cell*. doi: 10.1016/j.cell.2012.11.052.

Yang, Y. M. *et al.* (2010) 'Septins regulate developmental switching from microdomain to nanodomain coupling of Ca2+ influx to neurotransmitter release at a central synapse', *Neuron*. doi: 10.1016/j.neuron.2010.06.003.

Yeom, K. H. *et al.* (2006) 'Characterization of DGCR8/Pasha, the essential cofactor for Drosha in primary miRNA processing', *Nucleic Acids Research*. doi: 10.1093/nar/gkl458.

Yi, G., Sze, S. H. and Thon, M. R. (2007) 'Identifying clusters of functionally related genes in genomes', *Bioinformatics*. doi: 10.1093/bioinformatics/btl673.

Yokoi, N., Fukata, M. and Fukata, Y. (2012) 'Synaptic Plasticity Regulated by Protein-Protein Interactions and Posttranslational Modifications', in *International Review of Cell and Molecular Biology*. doi: 10.1016/B978-0-12-394308-8.00001-7.

Young, D., Shprintzen, R. J. and Goldberg, R. B. (1980) 'Cardiac malformations in the velocardiofacial syndrome', *American Journal of Cardiology*. doi: 10.1016/0002-9149(80)90515-9.

Yue, Z. *et al.* (2016) 'Identification of breast cancer candidate genes using gene co-expression and protein-protein interaction information', *Oncotarget*. doi: 10.18632/oncotarget.9132.

Zhao, D. *et al.* (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*. doi: 10.1371/journal.pone.0132387.

Zhao, Y. *et al.* (2020) 'Complete Sequence of the 22q11.2 Allele in 1,053 Subjects with 22q11.2 Deletion Syndrome Reveals Modifiers of Conotruncal Heart Defects', *American Journal of Human Genetics*. doi: 10.1016/j.ajhg.2019.11.010.

Zimmerman, L. B., De Jesús-Escobar, J. M. and Harland, R. M. (1996) 'The Spemann organizer signal noggin binds and inactivates bone morphogenetic protein 4', *Cell*. doi: 10.1016/S0092-8674(00)80133-6.

Zinkstok, J. R. *et al.* (2019) 'Neurobiological perspective of 22q11.2 deletion syndrome', *The Lancet Psychiatry*. doi: 10.1016/S2215-0366(19)30076-8.

Zufferey, R. *et al.* (1998) 'Self-Inactivating Lentivirus Vector for Safe and Efficient In Vivo Gene Delivery', *Journal of Virology*. doi: 10.1128/jvi.72.12.9873-9880.1998.