CYFIP1-dependent Regulatory Networks during Cortical Differentiation of Human Pluripotent Stem Cells: Implications in Neuropsychiatric Diseases

Thesis submitted for the degree of Doctor of Philosophy

School of Medicine, Cardiff University

2019

Daniel Cabezas de la Fuente

DECLARATION

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed …… Date ……

STATEMENT 1

This thesis is being submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy, PhD

Signed …… Date ……

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated, and the thesis has not been edited by a third party beyond what is permitted by Cardiff University's Policy on the Use of Third-Party Editors by Research Degree Students. Other sources are acknowledged by explicit references. The views expressed are my own.

Signed …… Date ……

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed …… Date ……

STATEMENT 4: PREVIOUSLY APPROVED BAR ON ACCESS

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loans **after expiry of a bar on access previously approved by the Academic Standards & Quality Committee.**

Signed …… Date ……

Acknowledgements

First and foremost, I wish to thank my supervisor Prof. Meng Li, for giving me the chance to work in this project and for her support throughout the whole process. I would also like to thank Dr Andrew Pocklington and Dr Robert Andrews for being my mentors and guides in the uncharted world of bioinformatics.

A massive thank you to all the members in the Li and Peall labs, for helping me through all this journey. Thank you Lucia and Marija, for always listening to my problems, teaching me and helping me find the right solution. Sorry Lucia for always stealing your snacks. To Áine, for being my friend, for always coping with me and making me laugh when I needed it the most and helping me to correct this thesis. However, I owe the biggest thank you to Claudia, who has given me the biggest help with cell culture and to her pink post-it notes that reminded me to stay focused. Thank you all for being amazing colleagues and friends.

Thank you to all my friends in Cardiff. To Dan, Kevin and Paul for all the great moments we have lived together, and have made me feel at home. To Loukia and Cai, who have been next to me at all times wether they were good or bad and always tried their best to make me smile. Thank you for your encouraging and motivating me.

I also need to thank all my friends in Igualada and Lleida. Thank you Anna, Marta, Guille, Carlota, Oriol and Eva for always keeping in touch despite the distance, and for always loving me and accepting me for who I am.

Lastly, I would like to thank my family without whom I would have never got this far. Thank you to my mother Gloria, the most hardworking person I have ever met and whom I look up to. This thesis is dedicated to the memory of Justa Elvira and Joaquin de la Fuente, my loving grandparents, who raised me to value knowledge. I hope I made you proud.

Abstract

Copy number variants affecting the chromosomic locus ch15q11.2 are associated with an increased risk for the development of neuropsychiatric disorders. CYFIP1 is one of the four genes within the 15q11.2 region and is the most studied thus far. While multiple studies have demonstrated a role for CYFIP1 in the physiology and morphology of neuronal cells, how altered expression of CYFIP1 and the other 15q11.2 genes confer risk to neuropsychiatric disorders remain largely unknown.

Increased levels of CYFIP1 caused a delayed in neurogenesis while decreased levels caused the progenitor cells to prematurely differentiate into neural cells.This thesis investigates the transcriptional changes associated with altered levels of CYFIP1 in order to elucidate the mechanisms behind the observed cellular phenotypes. RNAseq analysis revealed altered overlapping pathways in these cells that could explain the observed phenotype. This also revealed new functions associated with CYFIP1-regulated genes involving mitochondrial and cholesterol metabolism. Common variant analysis indicated that amongst the CYFIP1-regulated genes, those that are targets of FMRP show a significant enrichment for genetic variants conferring increased risk for schizophrenia.

In-depth analysis indicated that *AKT3* was one of the top candidate genes involved in the development of the abnormal phenotype observed in cells with altered CYFIP1 expression. Genetic manipulation of *ATK3* using a lentiviral CRISPR/Cas9 system showed evidence that AKT3 is involved, in part, in the maintenance of the neural progenitor cells in CYFIP1tg cultures.

This thesis demonstrates that the defects caused by altered CYFIP1 expression can be explained partly by changes in AKT3 activity elucidating a potential mechanism behind the neuropsychiatric phenotype associated with 15q11.2 CNVs.

Contents

List of figures

List of tables

List of abbreviations

Abbreviation	Full name
ABI	Ableson interacting protein
ADHD	attention deficit and hyperactivity disorder
ADHD	Attention deficit hyperactivity disorder
ARCCA	Advanced Research computing @ Cardiff
AS	Angelman syndrome
ASD	Autism spectrum disorder
βCAT	β-Catenin
BP	Breakpoint
BSA	Bovine serum albumin
CNS	Central nervous system
CNV	Copy number variant
CYFIP1	cytoplasmic FMRP-interacting portion 1
CYFIP2	cytoplasmic FMRP-interacting portion 2
DGE	differential gene expression
DLL	Delta-like
eGFP/GFP	Enhanced green fluorescent protein
EHS5	hairy and enhancer of split homolog 5
eIF4E	Eukaryotic translation initiation factor 4E
eQTL	expression quantitative trait loci
FC	Fold change
FDR	False discovery rate
FMRP	Fragile X mental retardation protein
FOXG1	forkhead box protein G1
FPKM	fragments per million reads
FRET	Förster resonance energy transfer
FXS	Fragile X syndrome
gDNA	genomic DNA
GSEA	Gene set enrichment analysis
GSK3β	glycogen synthase 3 β
GWAS	Genome-wide association study
HES1	hairy and enhancer of split homolog1
hPSCs	Human pluripotent stem cells
HSPC300	hematopoietic stem progenitor cell 300

ID	Intellectual disability
InDels	Insertions/Deletions
IP	intermediate progenitors
iPSC	Induced pluripotent stem cells
IPTG	isopropyl β-D-1-thiogalactopyranoside
IZ	Intermediate zone
KD	Knockdown
KEGG	Kyoto Encyclopaedia for Genes and Genomes
KO	Knockout
LD	Linkage disequilibrium
MAGMA	Multi-marker analysis of genomic annotation
mTOR	mammalian target of rapamicin
NAP	NCK-associated proteins
NCAD	N-Cadherin
NDD	neurodevelopmental disorder
NES	Normalised enrichment score
NGS	next generation sequencing
NIPA1/2	non-imprinted in Prader-Willi and Angelman syndrome 1/2
NPCs	neural progenitor cells
PBS-T	Phosphate buffer saline - Triton X 100
PCA	Principal component analysis
PH	pleckstrin homology
PIP_3	phosphatidylinositol (3,4,5)-triphosphate
PRS	Polygenic risk score
PTEN	phosphatase and tensin homolog
PWS	Prader-Willy syndrome
RGC	radial glial cells
RIN	RNA integrity number
RNAseq	RNA sequencing/whole-transcriptome shotgun sequencing
SETBP1	Set binding protein 1
SNP	single nucleotide polymorphism
SOC	Super optimal broth with catabolite repression
SVZ	subventricular zone
SZ	Schizophrenia
TBS-T	Tris buffer saline - Tween 20
TUBB3	β-III-tubulin

TUBGCP5	tubulin gamma complex associated protein 5
VCA	Verprolin-homology, Central and Acidic regions
VSVG	Vesicular stomatitis virus G
VZ	ventricular zone
WAVE	WASP (Wiskott-Aldrich syndrome protein)-family verprolin homologous protein
WRC	wave regulatory complex
WT	wild type
X-Gal	5-Bromo-4-Chloro-3-Indolyl β-D-Galactopyranoside

Chapter 1

General introduction

1.1 Introduction

Neuropsychiatric disorders such as schizophrenia (SZ) and autism spectrum disorders (ASD) are debilitating psychiatric diseases affecting around 1% of the world population (Tandon, Nasrallah and Keshavan, 2009; McRae *et al.*, 2017). Individuals suffering from these disorders experience reduced life quality and expectancy and tend to require long-term care (Chesney, Edward, Goodwind, Guy M. and Fazel, Seena, 2014). The development of these disorders not only has a huge negative impact on the patients themselves but also represent a great economic burden to the society (Buescher *et al.*, 2014; Chong *et al.*, 2016)

SZ is characterised by a combination of positive and negative symptoms. The positive symptoms include delusions and hallucinations, while the negative symptoms consist of impaired speech, reduced motivation, social isolation, and cognitive impairment (Owen, Sawa and Mortensen, 2016). This disorder has an early onset manifesting during the late adolescence or early adulthood. However, the development of SZ tends to be preceded by a prodromal phase, in which no apparent symptoms are present (Makeev, 1978). ASD is commonly known by the manifestation of severe psychiatric symptoms which include social anxiety, opposing defiant disorder, intellectual disability (ID), attention deficit and hyperactivity disorder (ADHD) (Masi *et al.*, 2017). ASD is also accompanied by multiple medical conditions such as sleep disorders, epilepsy and mitochondrial disorders, amongst others (Oliveira *et al.*, 2005; Weissman *et al.*, 2008; Bauman, 2010). Both SZ and ASD have a heritable component, at 80% in SZ and 40-60% in ASD respectively (Cardno and Gottesman, 2000; Shih, Belmonte and Zandi, 2004; Gaugler *et al.*, 2014; Polderman *et al.*, 2015). Both SZ and ASD have heterogeneous mechanisms underlying the cause of these disorders, combining the effect of the individual genetic background with environmental factors (Brugha *et al.*, 2014; Owen, Sawa and Mortensen, 2016).

Hundreds of chromosomic loci have been associated with an increased risk of developing these conditions. However, how these risk loci contribute to the aetiology of the disorders and the cellular mechanisms behind them remain largely unknown. This could be explained in part by the large pleiotropic effect of the genes involved and the existing overlap of common risk variants between SZ, ASD and other neuropsychiatric disorders (Lee *et al.*, 2013).

The vague understanding of the pathological mechanisms underlying the disorders presents a major challenge in the development of new therapies. Gain- and loss-of-function animal models have been traditionally used by scientists to elucidate the developmental origin of multiple disorders and diseases. The study of neuropsychiatric disorders in

animals is often not feasible or unable to lead to conclusive results. Two main reasons are behind this: 1) chromosomal alterations harbouring multiple genes confer increased risk of development of psychiatric disorders, which make it difficult to generate an animal model to study the disease. 2) The complexity of the human neocortex compared to classic model organisms.

Human pluripotent stem cells (hPSC)-derived models allow the study in relevant human cells, albeit *in vitro*. For example, combining the use of DNA editing tools such as CRISPR/Cas9 to mimic known mutations and patient-derived induced pluripotent stem cells (iPSCs) carrying defined risk variants, with the generation of neuronal types recapitulating the human embryonic cortical development and whole-transcriptome shotgun sequencing, commonly known as RNA sequencing (RNAseq), allow researchers to investigate the effects of mutations in one or multiple risk genes on transcription and cellular behaviour at stages of interest during *in vitro* neural development. This approach represents an advantage in characterising the pathological effects of risk variants and in identifying altered pathways and ultimately provides a platform for identification of drug targets and screening studies.

1.2 Cortical development

The brain is considered the most complex organ in the human body, and the cortex is the largest and most complex structure. The cortex contains the higher cognitive functions involved in consciousness and perception, making it particularly interesting in the context of psychiatric disorders (Frith and Dolan, 1996). Consisting of six stratified cortical layers, the cortex is formed in an inside-out fashion, where newly born neurons settle on top of the layer that preceded them (Rakic, 1974) (**Figure 1.1**). Structural, cellular and molecular studies of neocortical development have been based on mouse and rat models. These animals largely recapitulate the development of the human neocortex in terms of layer organisation and regionalisation into areas containing specific functions. However, the rodent cortex is 1,000-fold smaller compared to that of the human and is non-folded (Rakic, 2009). Moreover; while the process of neurogenesis takes 7 days in the mouse, it takes 70 days in the human. These implications reveal the inherent limitations of extrapolating findings in animal models.

The neocortex is formed by a wide variety of cells, of which pyramidal excitatory neurons are the most abundant (approximately 85%). The remaining neuronal cells are GABAergic inhibitory interneurons. Both cell types are formed in the prosencephalon (or forebrain). Pyramidal neurons are born in the dorsal telencephalon, interneurons are generated in the ventral telencephalon in the medial and caudal ganglionic eminence.

Interneurons constitute approximately 10% of the cerebral neocortex mass in rodents and ~20% in primates (Marín & Müller, 2014). These cells use tangential migration to reach their final destination in the cortex.

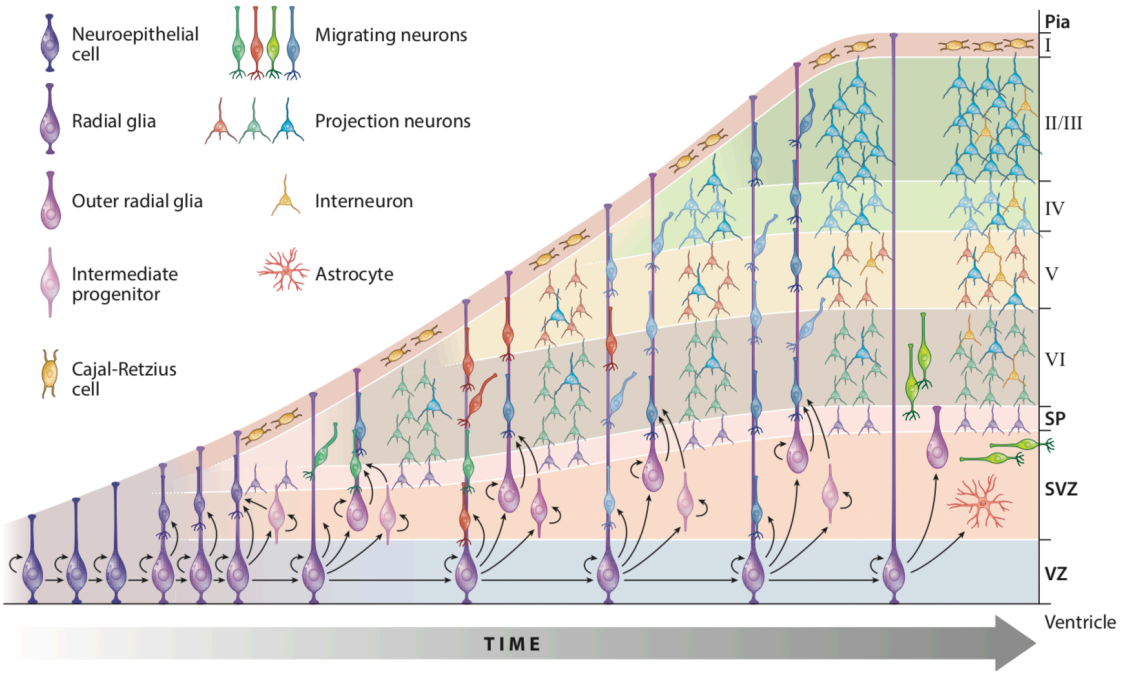


Figure 1.1 Development and structure of the mammalian cortex
As the cortical telencephalon develops, neuroepithelial cells will give raise to radial glia. These cells will divide and expand in numbers to finally generate glutamatergic neurons. Before terminally differentiating, radial glia will generate intermediate progenitors that will undergo several rounds of division before generating neurons. With time, interneurons arrive from ganglionic eminences through tangential migration and stablish themselves in the cortical plate. SP: subplate; SVZ: subventricular zone; VZ: ventricular zone. Adapted from (Greig *et al.*, 2013).

1.2.1 The early stages of cortical development

The central nervous system (CNS) originates from the patterning of the embryonic neural plate, which generates the neural tube, a hollow structure that develops into the brain and spinal cord. Four subdivisions are present in the neural tube that form the distinct regions of the CNS: the prosencephalon (forebrain), mesencephalon (midbrain), rhombencephalon (hindbrain) and spinal cord (Smith and Schoenwolf, 1997; Sadler, 2005). The neural tube is populated by neuroepithelial cells which give rise to radial glial cells (RGC), which reside in the ventricular zone (VZ) of the neural tube (Mukhtar and Taylor, 2018). As embryonic development progresses, the prosencephalon further develops into the telencephalon: the part of the brain containing the cortex as well as multiple subcortical nuclei.

In the telencephalon, RGCs are in contact with the ventricular (apical) surface where the cell body lies, and pial (basal) surface through the extension of a cellular process (Rakic, 1971, 1972; Hartfuss *et al.*, 2001; Noctor *et al.*, 2002). It has been shown in the mouse brain that apical progenitors generate an epithelial niche where they undergo a series of symmetric divisions expanding the pool of progenitor cells. Apical progenitors cells generate basal neural progenitor cells (or intermediate progenitors), a transient progenitor population that divide and differentiate to excitatory neurons (Rakic, 1995; Malatesta *et al.*, 2000; Miyata *et al.*, 2001; Noctor *et al.*, 2001). Basal neural progenitors migrate towards the subventricular zone (SVZ) where intermediate progenitors undergo the majority of neuron-producing cell divisions (Kowalczyk *et al.*, 2009). It is these cells that contribute to a major expansion of the cortical volume in primates and humans.

In rodents, intermediate progenitors only exist for a short period of time (hours) in the SVZ, dividing only once before generating two terminally differentiated neurons (Noctor *et al.*, 2004). Conversely, in primates, the SVZ is more complex and is subdivided into inner and outer SVZ, which is often split by a thin fibre layer (Smart *et al.*, 2002; Zecevic, Chen and Filipovic, 2005; Fish *et al.*, 2008). TBR2 (*EOMES*) expressing intermediate progenitors are maintained for a longer period of time (up to a week) in the outter SVZ, where they undergo multiple neurogenic divisions (Kriegstein, Noctor and Martínez-Cerdeño, 2006) (**Figure 1.2**). According to the radial unit hypothesis, new-born neurons will migrate along the process of the RGC to reach the basal surface and form the cortical plate (Rakic, 1988). The different waves of differentiation will give rise to the multi-layered structure of the cortex. The final number of neuronal cells in the cortical plate will be dependent on the balance between progenitor proliferation and terminal differentiation. These processes are controlled by multiple signalling molecules that are released locally, and through gradients that diffuse at different concentrations in the brain through cerebrospinal fluid. The following section will review some of the relevant cellular pathways underpinning the work carried out in this thesis.These pathways are involved in the governing the ratio between neural progenitor proliferation and neural differentiation.

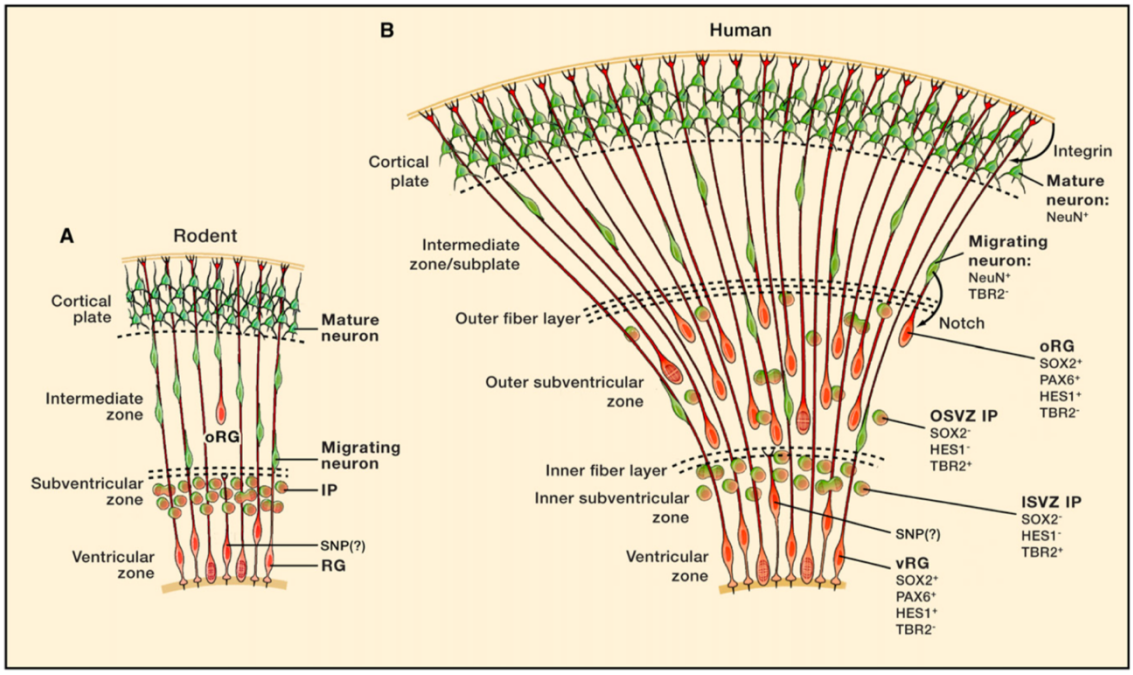


Figure 1.2 Comparison of rodent and human neocortical development
A comparative view of the rodent (A)and human (B) corticogenesis. The presence of an outer subventricular zone creates a niche for intermediate progenitors to undergo transient amplification before terminally differentiating into postmitotic neurons generating the obvious difference in size with the rodent cortex. IP: intermediate progenitor; ISVZ: inner subventricular zone; OSVZ: outer subventricular zone; vRG: ventricular radial glia. Figure taken from (Lui, Hansen and Kriegstein, 2011).

1.2.2 Pathways involved in neocortical development

(I) NOTCH signalling pathway

The Notch signalling pathway has a highly conserved role in embryonic morphogenesis and neurogenesis in both vertebrates and invertebrates (de la Pompa *et al.*, 1997; Artavanis-Tsakonas, Rand and Lake, 1999; Lai and Kimble, 2004). This signalling pathway involves a family of cell-surface receptors, NOTCH 1-4 in humans, activated by the different orthologs of the Delta-like (DLL) and Jagged ligands, and requires physical interaction between the signalling cells (Gaiano and Fishell, 2002). The interaction between Notch and its ligands promotes the proteolytic cleavage and release of the intracellular domain, which translocates to the nucleus. In the nucleus, the intracellular Notch domain interacts with centromere-binding protein 1; suppressor of Hairless and Lag-1, collectively known as CSL, switching its repressor function to co-activator mediating the transcription of Notch target genes (Lai and Kimble, 2004). The activity of Notch signalling is negatively regulated by the protein Numb, which can bind the Notch ligands DLL and Jagged, promote the degradation of the active Notch intracellular domain, or promoting the endocytosis of Notch from the cell membrane (Giebel and Wodarz, 2012).

In neural development, Notch signalling regulates the generation of neurons in waves. Notch promotes the expression of the basic helix-loop-helix transcription factors hairy and enhancer of split homolog-1 and hairy and enhancer of split homolog-5 (HES1 and HES5 respectively). HES1 and HES5 act as repressors of the proneural genes neurogenein-2 (NEUROG2) and achaete-scute homologue 1 (ASCL1), promoting the maintenance of the proliferative state in one of the basal progenitors. Notch drives this effect through the process of lateral inhibition, where a specified cell exercises an inhibitory pressure over the immediate surrounding cells, in order to stop them from becoming the same cellular type (Haddon *et al.*, 1998). DLL and Jagged are expressed at different levels by different cell populations. Cells with a higher expression of the Notch ligands induce high levels of HES1 and HES5, repressing terminal differentiation (**Figure 1.3 A**) (Heitzler and Simpson, 1991).

Giano and colleagues demonstrated the importance of this signalling for maintaining the balance between proliferation and differentiation. By overexpressing the activated notch intracellular domain, the authors observed increased numbers of RGC indicating that Notch is associated with this state (Gaiano, Nye and Fishell, 2000). On the other hand, loss-of-function of the downstream effectors, and in particular CBF1, promote premature differentiation and depletion of the RGC in favour of the postmitotic neuronal fate (Mizutani *et al.*, 2007; Imayoshi *et al.*, 2010).

(II) β-Catenin (βCAT)

One of the central regulators of neural development is βCAT (encoded by the *CTNNB1* gene). This protein exerts a number of roles in the cell. When it is first synthesized, βCAT binds to E-Cadherin participating in cell adherens junctions, excess of free βCAT in the cytoplasm can translocate to the cell nucleus where they act as transcriptional co-activators. Moreover, βCAT could have a function in regulating cell division and polarity by interacting with the centrosome (Chilov *et al.*, 2011; Mbom, Nelson and Barth, 2013; Marín and Müller, 2014). βCAT is a point of convergence of multiple cellular signalling pathways. Many studies have characterised the importance of this protein in the development of the cortex. βCAT is required for the establishment of the telencephalon fate, maintaining the structural integrity of the neuroepithelium and promoting the survival of NPCs (Gunhaga *et al.*, 2003; Machon *et al.*, 2003; Backman *et al.*, 2005; Junghans *et al.*, 2005)

A well-described mechanism is the connection between Wnt and βCAT. In normal conditions, free βCAT is rapidly phosphorylated by the glycogen synthase 3 β (GSK3β) promoting the degradation of the former. Binding of Wnt to its receptor Frizzled and the co-receptor low-density lipoprotein receptor-related protein 5 and 6 (LRP5 and LRP6) inhibit

the destruction activity promoted by GSK3β, which allows the accumulation of βCAT and translocation to the cell nucleus acting as a co-activator of TCF/LEF allowing the expression of Wnt target genes. However, βCAT can be activated by multiple stimuli in the cell, such as adhesion proteins. Zhang and colleagues demonstrated that reduction of N-Cadherin (NCAD) was directly related to the loss of activated βCAT and an increase in progenitor differentiation (Zhang *et al.*, 2010). A similar event was observed following direct removal of βCAT (Mutch *et al.*, 2010). The authors showed that this change was not caused by Wnt signalling pathways but by a combination of direct activation of βCAT through AKT and, perhaps, by a reduction in phosphorylation through GSK3β (**Figure 1.3 B**) (Zhang *et al.*, 2010, 2013).

Expression of a dominant-negative form of βCAT in neural progenitors promotes differentiation of the RGCs to a neuronal fate, indicating a role for βCAT in the maintenance of undifferentiated RGCs (Mutch *et al.*, 2010). In support of this finding, expression of stabilised βCAT in mouse promoted self-renewal of the RG, resulting in a massive expansion of the progenitors residing in the VZ (Chenn and Walsh, 2002). Further analysis of transgenic mice with elevated levels of stabilised βCAT revealed defects in neuronal migration and disruption of the cortical lamination (Chenn and Walsh, 2003). Loss of GSK3β results in the over-proliferation of SOX2 positive RGs with a concomitant activated βCAT (Kim *et al.*, 2009). The same effect was also seen by transgenic expression of a constitutively active form of βCAT in mice (Chenn and Walsh, 2002).

(III) PI3K-AKT signalling

The PI3K-AKT pathway can be activated by a variety of signals including, amongst others, growth factors and the extracellular matrix (Nicholson and Anderson, 2002). Upon reception of the right stimuli, PI3K converts phosphatidylinositol (PI) to PI 3,4,5-triphosphate (Fruman, Meyers and Cantley, 1998). These phosphorylated lipids remain anchored to the cell membrane and recruit proteins with a pleckstrin homology (PH) domain. This allows the interaction of phosphoinositide-dependent kinase 1 (PDK1) and AKT, leading to the phosphorylation of AKT in S308 (Cantley, 2002). Complete activation of AKT requires a second phosphorylation in S473, which can be given by other proteins with kinase activity, including the mechanistic target of rapamycin complex (mTOR), or even through a mechanism of autophosphorylation (Hemmings and Restuccia, 2012). However, AKT can also be activated independently from PI3K activity. This can be given by response to specific growth factors, DNA damage or increases in Ca^{2+} concentration (Derkach, Barria and Soderling, 1999; Zheng *et al.*, 2010; Saki, Toulany and Rodemann, 2013). Activation of AKT is negatively regulated by phosphatase and tensin homolog

(PTEN), protein phosphatase 2 (PP2A) and the PH-domain leucine-rich-repeat-containing protein phosphatases (PHLPP1/2) (Hemmings and Restuccia, 2012).

The PI3K-AKT pathway is required for radial migration in new-born neurons. Overexpression of Akt in mouse cortex at E13.5 increased the velocity in which cells migrated towards the cortical plate. However, expression of an inactive form of the protein reduced the neuronal migration, stalling the cells in the SVZ/VZ (Itoh *et al.*, 2016). Mutations affecting the PI3K-AKT signalling are a cause of malformations during cortical development, which can be implicated in the development of ASD, developmental delay and epilepsy. Research has shown that alterations in this signalling pathway generate cortical dyslamination (alterations in the cortex laminar structure) and enlarged soma. Treatment with rapamycin, a drug inhibiting mTOR during embryonic developmental can revert the abnormal migration caused by altered PI3K-AKT signalling (Baek *et al.*, 2015).

Despite being a well-known pathway in postmitotic neurons, current knowledge about a role for PI3K-AKT signalling in NPCs is rather limited. Multiple studies in mice have shown that increased activity of AKT pathway induces cell proliferation, whereas the inhibition reduces the proliferation of cortical progenitor cells (Cheng *et al.*, 2018). AKT phosphorylates over 100 proteins, two of which, forkhead box protein G1 (FOXG1) and βCAT, may be responsible for the above-mentioned effects. A FOXG1 gradient is required for the formation and dorsoventral patterning of the mammalian telencephalon (Martynoga *et al.*, 2005; Kumamoto and Hanashima, 2017). Mutations affecting FOXG1 function cause structural brain abnormalities, severe intellectual disability and autism (Ariani *et al.*, 2008). During cortical development, FOXG1 acts as a repressor of transcription. When AKT is active, it phosphorylates FOXG1, displacing it to the cytoplasm and allowing the transcriptional expression of its target genes (Regad *et al.*, 2007). Furthermore, AKT participates in the regulation of βCAT signalling, independently of WNT activation, through two mechanisms. Firstly, AKT can directly phosphorylate βCAT, leading to its activation and translocation to the nucleus. Secondly, AKT phosphorylates and inactivates GSK3β, the protein that represses the activity of βCAT (**Figure 1.3 B**). This further amplifies the signalling effect by βCAT and has a direct effect over the cell cycle and the proliferation/differentiation balance of cortical progenitor cells (Zhang *et al.*, 2010, 2013).

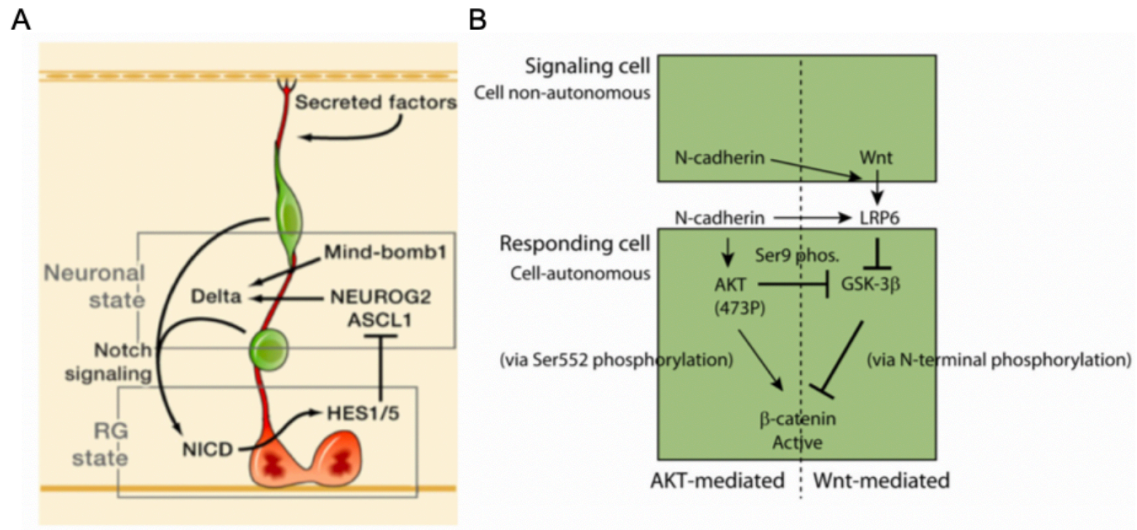


Figure 1.3 Pathways involved in radial glia maintenance and differentiation
(A) Schematic representation of Notch signalling in cortical development. Migrating new-born neurones secrete the Delta ligand that will bind to the RGC activating HES1/5 and inhibiting the expression of proneural genes. (B) Regulation of βCAT through ATK and WNT signalling by directly inactivating GSK3β. Figure adapted from (Lui, Hansen and Kriegstein, 2011; Zhang *et al.*, 2013).

1.2.3 Aberrant cortical development

The neural tube expands during cortical development in order to form the brain. Alterations in the process of neurogenesis can have devastating effects on normal brain structure and function. These alterations are associated with neurodevelopmental disorders (NDD) which affect normal cognition, learning abilities and self-control amongst others (Thapar, Cooper and Rutter, 2017) amongst these disorders we can find SZ and ASD.

Several studies have identified a reduction of the total volume and neuronal number in subcortical structures, such as the thalamus, in SZ patients (Pakkenberg, 1990; Young *et al.*, 2000; Csernansky *et al.*, 2004). Moreover, it is widely known that SZ patients show abnormalities in cortical grey matter volume (Gur *et al.*, 1999; Vita *et al.*, 2012; Xiao *et al.*, 2013; Yue *et al.*, 2016). It has been hypothesised that this thinning is caused by antipsychotic drugs. However, research in drug-naïve SZ patients show similar, if not more pronounced defects, indicating pre-existing structural abnormalities (Haijma *et al.*, 2013). Concomitant with these defects, the dorsolateral prefrontal cortex show increased neuronal density, suggesting that the defects originate from alterations in neural connectivity (Selemon, Rajkowska and Goldman-Rakic, 1995). Importantly, all these abnormalities can be recapitulated in non-human primates. By radiating gestating macaques in the first gestational trimester, Racik and colleagues were able to inhibit neurogenesis in cortical and subcortical areas. These monkeys presented all the above mentioned structural abnormalities as well as cognitive deficits reminiscent of SZ, suggesting a neurodevelopmental origin for the disorder (reviewed by (Selemon and Zecevic, 2015).

Contrary to the findings in SZ brains, the cortical surface area in a subset of ASD patients appears to be increased (Ohta *et al.*, 2016). This could be explained, at least in part, by an increased number of neurons in the prefrontal cortex (Courchesne *et al.*, 2011). This theory is supported by blood analysis of ASD patients as well as in post-mortem brains, where large numbers of cell-cycle genes have found to be dysregulated (Pramparo *et al.*, no date; Chow *et al.*, 2012). It is known that controlling cell-cycle is one of the critical elements for balancing between proliferation and differentiation of the NPCs and have a repercussion over the cerebral cortex area, thus being a likely mechanism for the aetiology of ASD (Lange, Huttner and Calegari, 2009; Pilaz *et al.*, 2009). In support of a neurodevelopmental origin of ASD hypothesis, genetic studies in ASD patients have identified many genes involved in neurogenesis as candidate genes conferring a strong risk for the development of the disorder (De Rubeis *et al.*, 2014; Iossifov *et al.*, 2014; Sanders *et al.*, 2015).

Indeed, the research in ASD and SZ suggest for a genetic neurodevelopmental origin for these disorders. Thanks to the development of high throughput sequencing, researchers have been able to identify rare mutations associated with the development of malformations and neurological disorders (Hu, Chahrour and Walsh, 2014). The following section will review the implication and advances of genetic studies in the research for the origin of psychiatric disorders.

1.3 Psychiatric genetics in SZ and ASD

1.3.1 Overview

Traditionally the study of the origin of NDDs has been based on the knowledge of Mendelian principles on inheritance, where the mutation of one gene was associated to a phenotype. Tools such as karyotyping and basic linkage analysis were the techniques used to identify the genetic basis of lissencephaly. This approach identified two genes: *LIS1* and doublecortin (*DCX*) (Reiner *et al.*, 1993; Lo Nigro *et al.*, 1997; des Portes *et al.*, 1998; Gleeson *et al.*, 1998).

The liability of ASD is driven mainly by common variants, and *de novo* mutations representing the individual liability (Yoo, 2015). Early ASD studies were based on candidate gene approaches to identify common variation of the genetic structure caused by single nucleotide polymorphisms (SNPs) (Yoo, 2015). SNPs are the most common variations in the genome. They account for differences between single nucleotides in the same position found in different individuals (Twyman, 2009). SNPs are single base substitutions that can occur in 1:300 nucleotides and vary within populations (Schwab and

Wildenauer, 2013). On the other hand, copy number variants (CNV) involve differing numbers of copies of sections of the genome (McCarroll and Altshuler, 2007). Linkage studies were able to identify single genes affected by these variations in the genome as the causal genes for syndromic ASD (**Table 1.0.1**). However, these studies failed to identify the origin of idiopathic ASD. Recent developments in the use of next generation sequencing and the introduction of whole exome sequencing has identified protein-disrupting mutations which could underlie idiopathic ASD (Geschwind and State, 2015).

As with ASD, initial research describing SZ incorporated the theory of monogenic inheritance due to its simplicity and the familial aggregation of the disorder (Bleuler and Jung, 1908). This theory was rapidly abandoned after the first systematic family study for SZ revealing that the data did not fit the Mendelian models of inheritance (Rüdin, 1916). Nonetheless, familial, twin and adoption studies establish a strong evidence for a genetic origin in SZ (reviewed in Henriksen, Nordgaard and Jansson, 2017). These studies revealed that the prevalence of SZ was higher in patients with an affected relative as compared to the general population. These studies show that monozygotic (genetically identical) twins have a higher rate of SZ than dizygotic twins (up to 65% and 28%, respectively). The same study estimated a heritability of 80% (Cardno and Gottesman, 2000).

Thanks to the development of molecular genetics, linkage studies were perfected. These types of studies are based on the premise that alleles that are close together are more likely to be transmitted as a unit during meiosis (Lander and Kruglyak, 1995). In SZ and ASD, linkage studies identified genomic regions that harboured variants increasing the risk for the disorder. Some of the most relevant regions identified using this technique included the 1q21-22 and 6p24-p22 in SZ (Moises *et al.*, 1995; Straub *et al.*, 1995; Brzustowicz *et al.*, 2000; Schwab *et al.*, 2000; Rosa *et al.*, 2002); and 3p25 and 7q35 in ASD (Lauritsen *et al.*, 1999; Ylisaukko-oja *et al.*, 2006). However, these results are difficult to reproduce mainly due to the small effect driven by these variants and the small sample sizes making it impossible to detect the linkage (Risch and Merikangas, 1996).

Causal gene	Syndrome	Locus	Clinical phenotype
NLGN3a	-	Xq13.1	Autism, Asperger syndrome, PDD-NOS
NLGN4a	-	Xp22.33	Autism, Asperger syndrome, X-linked mental retardation, PDD-NOS
SHANK3a	-	22q13.3	Autism with severe language and social deficits
NRXN1	Pitt Hopkins-Like syndrome-2	2p16.3	Autism with seizures, facial dysmorphism spoken language deficits
MeCP2	Rett's syndrome	Xq28	Autism, learning disability, Angelman syndrome phenotype, variant of Rett syndrome
HOXA1		7p15.3	ASD susceptibility
PTEN	Cowden syndrome	10q22.31	ASD with macrocephaly
FMR1	Fragile X syndrome (FXS)	Xq27.3	Elongated cranial features, low social interaction, developmental delay, hyperactivity
NSD1	Sotos syndrome	5q35.2-q35.3	Overgrowth in childhood, learning disabilities, ADHD
TSC1/ TSC2	Tuberous sclerosis	9q34.13/ 16p13.3	ADHD, impulsivity, social impairment, epilepsy
CACNA1C	Timothy syndrome	12p13.33	Severe verbal delay, developmental delay, social avoidance
NF1	Neurofibromatosis type 1	17q11.2	ADHD, social anxiety
TRSP1	Trichorhinophalangeal syndrome	8q24.12	Distinctive facial features, high-functioning autism

Table 1.0.1 Causal genes of syndromic and non-syndromic ASD.
Table summarising individual genes identified to be the origin of syndromic and non-syndromic cases of ASD. Table adapted from (Caglayan, 2010; Sztainberg and Zoghbi, 2016).

Even though the human exome harbours ~85% of disease-related variants, it is unclear whether whole exome sequencing is able to capture the genetic variants associated with complex disorders (Lacey, Chung and Lin, 2014), therefore the remaining non-coding sequences of the genome should also be investigated. The rapid development and reduction in price of next generation sequencing (NGS) represented a major breakthrough in the research for variants causing SZ and ASD, it has allowed the study of large cohorts. In genome-wide association studies (GWAS), the genome of individuals is surveyed in order to identify known common genetic variants. The following section will cover the fundamental aspects of GWAS and how they can be applied in the research for SZ and ASD.

1.3.2 GWAS for neuropsychiatric disorders

GWAS are observational studies of genetic variants of different individuals at the genome-wide level. Using a phenotype-first approach, the objective of GWAS is to identify genetic risk factors for complex diseases through a hypothesis-free method (Bush and Moore, 2012). Typically, a GWAS is focused on a case-control setup where individuals are grouped according to a specific trait or phenotype. In order to detect genetic variants in the individuals, DNA is genotyped using probes for at least 100,000 SNPs (Cirillo *et al.*, 2018). Differences in allele frequency can be investigated between each group reporting an odds ratio for each group to have a specific SNP, determining the association for that particular SNP towards the development of the phenotype.

GWAS have led to the identification of common and rare variants such as CNVs and SNPs that contribute to the development of a disease (McCarroll *et al.*, 2008). In 2005, the first GWAS for age-related macular degeneration marked a milestone with the identification of two SNPs with altered frequency between 96 patients and 50 controls (Klein *et al.*, 2005). Since then, the publication of GWAS has increased significantly with more than 3900 human GWAS made available online (https://www.ebi.ac.uk/gwas/). Over time, GWAS studying larger cohorts have identified CNVs that could explain, in part, the development of neuropsychiatric disorders including SZ and ASD Table 1.2 (Sebat, 2013).

CNVs confer a significant increased risk for SZ (odds ratio 2-60) (Marshall *et al.*, 2017). This association has been observed in a variety of studies including case control and family-based studies. Walsh *et al*. and the International SZ Consortium found up to 3-fold enrichment in CNVs in cases compared to controls (Walsh *et al.*, 2008; Ripke *et al.*, 2014). More than 30 GWAS for SZ have characterised SNPs that contribute to the development of psychiatric diseases. The most recent study identified 145 SZ associated loci through meta-analysis (Pardiñas *et al.*, 2018). This is an indication of the effect of

the sample size in the power of the GWAS. Sample sizes for ASD have been more limited compared to SZ making the analysis largely underpowered, limiting the ability to identify genome wide significant loci (**Table 1.2**).

Genetic studies comparing autistic probands with unaffected siblings and parents identified an increased occurrence of *de novo* CNVs up to 5-fold higher in ASD patients (Autism Genome Project Consortium *et al.*, 2007; Marshall *et al.*, 2008; Weiss *et al.*, 2008; Fischbach and Lord, 2010). Duplications in 15q11-q13, including Prader-Willi and Angelman syndromes (PWS and AS, respectively); and FXS have been consistently identified in patients with autism (Chung, Tao and Tso, 2014; Yoo, 2015). New CNVs are being reported in ASD as result of NGS studies. This demonstrates the relevance of NGS in identifying and confirming *de novo* mutations as a cause for ASD, since many causal CNVs haven't been reported yet for ASD (Yoo, 2015). To date, the largest published meta-analysis for ASD has only been able to identify a locus in 10q24.32 containing genes involved in proteasomal degradation pathway (Consortium, 2017). Altered intelligence is another psychiatric trait associated with chromosomal defects in 15q11.2 region. A large-scale meta-analysis is available for intelligence which identified over 200 loci associated with this trait (Savage *et al.*, 2018).

Table 1.2 CNVs conferring risk to the development of SZ and ASD.

Causal CNV	Condition	Reference
1q21.1 del/dup	SZ OR del = 8.3 dup = 3.45 ASD OR del = 1.6 dup = 8.0	(Autism Genome Project Consortium *et al.*, 2007; Brunetti-Pierri *et al.*, 2008; International Schizophrenia Consortium *et al.*, 2008; Mefford *et al.*, 2008; Stefansson *et al.*, 2008; Walsh *et al.*, 2008; Kirov *et al.*, 2009; Need *et al.*, 2009)
15q13.3 del/dup	SZ OR del = 10.7 dup = NA ASD OR del = 10.8 dup = NA	(International Schizophrenia Consortium *et al.*, 2008; Stefansson *et al.*, 2008; Ben-Shachar *et al.*, 2009; Miller *et al.*, 2009)

16p11.2 del/dup	SZ OR del = 0.9 dup = 9.4 ASD OR del = 9.5 dup = 11.8	(Sebat *et al.*, 2007; Christian *et al.*, 2008; International Schizophrenia Consortium *et al.*, 2008; Kumar *et al.*, 2008; Marshall *et al.*, 2008; Walsh *et al.*, 2008; Weiss *et al.*, 2008; Bijlsma *et al.*, 2009; Glessner *et al.*, 2009; McCarthy *et al.*, 2009)
16p13.11 del/dup	SZ OR del = 25.8 dup = 2 ASD OR del = 10.7 dup = 1.5	(International Schizophrenia Consortium *et al.*, 2008; Kirov *et al.*, 2009; Need *et al.*, 2009; Ingason *et al.*, 2011)
17p12 del/dup	SZ OR del = 9.5 dup = NA ASD OR del = 16.0 dup = NA	(Autism Genome Project Consortium *et al.*, 2007; International Schizophrenia Consortium *et al.*, 2008; Stefansson *et al.*, 2008; Weiss *et al.*, 2008; Kirov *et al.*, 2009)
22q11.2 del/dup	SZ OR del = NA dup = 0.4 ASD OR del = NA dup = 3.3	(Xu *et al.*, no date; Autism Genome Project Consortium *et al.*, 2007; International Schizophrenia Consortium *et al.*, 2008; Marshall *et al.*, 2008; Glessner *et al.*, 2009; Kirov *et al.*, 2009; Need *et al.*, 2009)
22q13.3 del/dup	SZ OR del = NA dup = NA ASD OR del = NA dup = NA	(Sebat *et al.*, 2007; International Schizophrenia Consortium *et al.*, 2008; Marshall *et al.*, 2008; Glessner *et al.*, 2009)

Table summarising CNVs identified in patients with SZ and ASD conferring risk to the development of the disorders. OR reported for SZ and ASD (Malhotra and Sebat, 2012; Rees *et al.*, 2014; Kirov, 2015)*. NA = not available.*

These studies show the complexity of psychiatric disorders, multiple genes can be affected, leading to alterations in multiple pathways (Kushima *et al.*, 2018). It is also known that the same mutation can have distinct effects in different patients. In turn, these altered pathways are likely to explain the phenotypical similarities between multiple neuropsychiatric disorders, indicating that the combination of all variants in the genetic background of the individual play a determining role in the development of the disorder (Craddock and Owen, 2010; Kirov, 2015; Consortium, 2017; Savage *et al.*, 2018; Grove *et al.*, 2019). Consistent with this hypothesis, alterations in the biology of FMRP have been associated to the development of ID, ASD and SZ (Consortium, 2017).

1.3.3 Heritability and genetic basis of SZ and ASD

Rather than a single-locus, SZ follows a polygenic heritability model (Gottesman and Shields, 1967; O'Rourke *et al.*, 1982). This indicates that the genetic risk is the combined association of alleles that can be present in a population in high to low frequencies, called common or rare variants (or alleles), respectively. Each allele presents a small contribution to the development of the schizophrenic phenotype (Craddock, O'Donovan and Owen, 2005). The accumulative effect of SNPs represent 23% of the liability in SZ, many of these being common variants identified through GWAS (Lee *et al.*, 2012).

The accumulative contribution of each variant can be calculated using a polygenic risk score (PRS) (Wray, Goddard and Visscher, 2007). The PRS can be calculated from GWAS summary statistics for each SNP tested. Therefore, a PRS can be computed for each tested individual, and the average PRS is expected to be higher in the case (manifesting a disorder) than in control population (**Figure 1.4 A, B**).

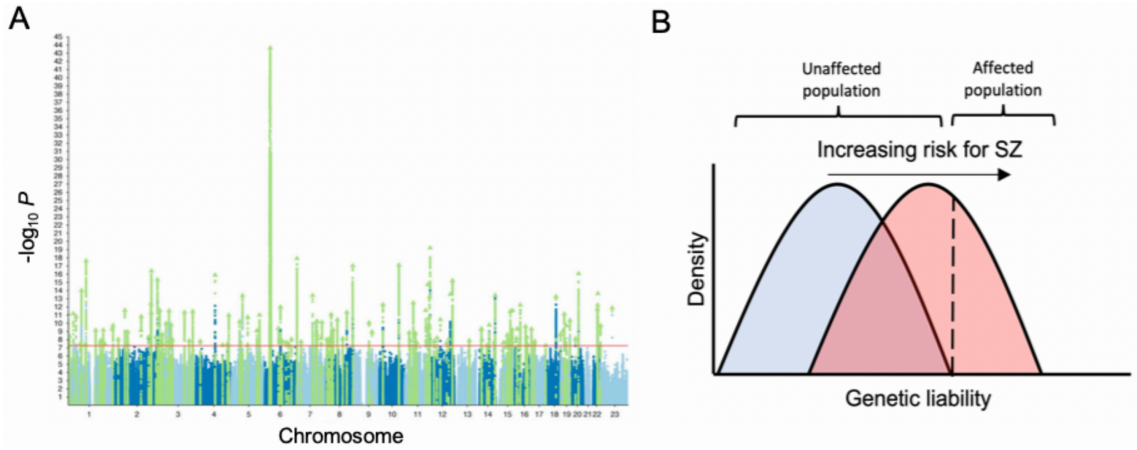


Figure 1.4 Polygenic signal of SZ.
Manhattan plot with the identified variants in the SZ GWAS from (Pardiñas et al., 2018). Red line indicates the genome-wide threshold of significance. (B) Graphical representation of the accumulative effect of polygenic risk score (PRS). Upon reaching a threshold, the accumulation of risk variants causes the appearance of the SZ phenotype.

Using the summary statistics of the Psychiatric Genetics Consortium for SZ and ASD GWAS, PRS for SZ has been used to accurately predict if individuals with first-episode psychosis will develop SZ or a different type of psychiatric disorders (Vassos *et al.*, 2017). In a recent study, PRS revealed that the heritable liability for SZ is mediated through pathways relevant to cognition including verbal memory, processing speed and visual memory (> 33%) (Toulopoulou *et al.*, 2019). In ASD, PRS has been used as a measurement of cognitive ability and intelligence quotient (Clarke *et al.*, 2016), or to identify specific alterations associated with the methylation profile in new-born children with the development of ASD (Hannon *et al.*, 2018). However, this approach is not applicable to every cohort. It has been shown that PRS for SZ is highly associated with ancestry. For any given population, there is a non-random association of alleles at different loci. This is known as linkage disequilibrium (LD). Therefore, corrections for LD should be applied when working with individuals of diverse ethnicity (Curtis, 2018).

1.3.4 Predicting dysregulated cellular processes using GWAS data

The association of alleles with a phenotype are only one part of the information obtained from GWAS. Both ASD and SZ are caused by the combinatorial effect of multiple variants. Because of the large number of SNPs analysed, very stringent correction tests need to be applied, hence reducing the number of significant variants. Mechanisms to reduce this stringency have been developed in order to reduce multiple testing and therefore aggregate the effects of individual SNPs (obtained from GWAS) into individual genes. PRS, gene and gene-set analysis are powerful alternatives to single-SNP analyses.

In gene analysis, the multiple SNPs identified through sequencing are mapped to genes. By doing this, one can test for the aggregated association of all SNPs with the phenotype. However, in some cases SNPs are located outside the coding sequence of a gene. 90% of the SNPs affecting expression quantitative trait loci are located within 15 kb from the 5' and 3' gene boundaries (Pickrell *et al.*, 2010). However, the effect of these SNPs can vary due to great differences in LD. The use of genotype data can help to aggregate SNPs in LD blocks before identifying the overlap with gene regions (White *et al.*, 2019).

In gene set analysis (also referred to as pathway analysis in genetics), individual genes are grouped in sets of genes with common characteristics, either biological or functional. Pathway analysis is an approach that has been applied to GWAS since 2007 (Wang, Li and Bucan, 2007). This method presents multiple advantages: firstly, aggregating genes

in groups reduces the number of tests required allowing the identification of weaker effects. Secondly, gene sets provide direct insight into specific biological pathways or cellular functions contributing to the aetiology of the phenotype (de Leeuw *et al.*, 2015).

The procedure of gene set analysis consists of three steps: 1) Gene sets are generated by the user with the genes to be investigated, 2) genetic variants (SNPs) are mapped onto genes, and 3) pathway enrichment is performed. Gene set p-values can be calculated using two different strategies: a one-step approach, where gene set p-values are calculated from genotype data, and a two-step approach, where single gene p-values are calculated first, from which the final pathway p-value is determined. Multiple software, such as VEGAS, PLINK, INRICH, are available to perform this kind of analysis. The statistical power of most of these programmes are strongly dependent on local LD, which in turn reduces the power to detect association dependent on multiple SNPs (Moskvina *et al.*, 2012). An approach to overcome this is by incorporating the information of the LD from different markers to detect multi-marker effects. This framework of analysis is used by the GWAS analysis tool Multi-marker Analysis of GenoMic Annotation (MAGMA) (de Leeuw *et al.*, 2015) (**Figure 1.0.5 A**).

Using this methodology, Jansen and colleagues were able to identify that genes targeted by the Fragile X mental retardation protein (FMRP) were enriched amongst candidate genes for autism, and that risk variants for ASD were more significant in FMRP target genes than in any other group of genes (**Figure 1.0.5 B**). In a different study carried out by Schijven and colleagues, the same analysis workflow was used to identify altered gene ontology (GO) terms that could capture the genetic association for ASD. Indeed, the analysis with MAGMA suggests that genes involved in synaptic plasticity and neuron differentiation could be driving the pathogenic effects causing ASD (**Figure 1.0.5 C**).

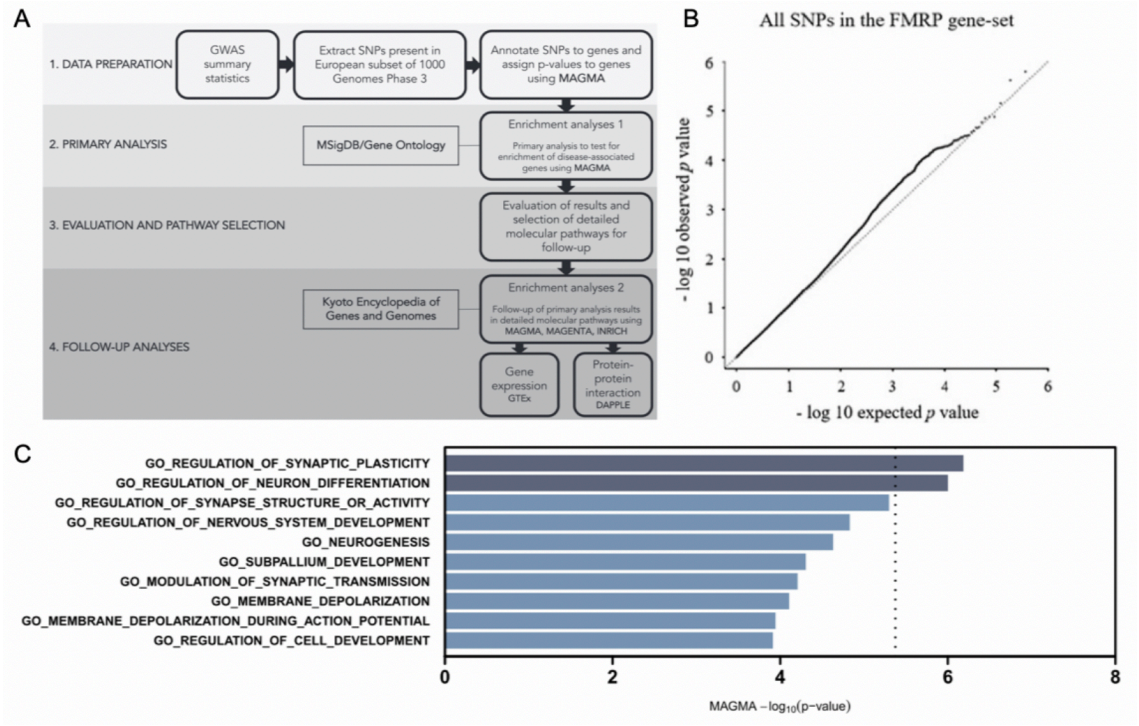


Figure 1.0.5 Genetic enrichment analysis with MAGMA
(A) Standard workflow for gene set enrichment analysis in MAGMA. (B) Quantile-quantile showing relationship between observed and expected pvalues of risk variants in the FMRP gene set. (C) Gene set enrichment analysis of GO with ASD associated genes tested in MAGMA. A and C from (Schijven et al., 2018) B from (Jansen et al., 2017).

1.3.5 Ch15q11.2 CNVs, a risk for psychiatric disorders

CNVs can range between 1,000 basepairs to several megabases (Mb) and can present as duplication (extra copy) or deletion (loss) of a segment of chromosomal DNA. They can be inherited or appear *de novo,* mainly through non-homologous recombination of low-copy repeat regions during meiosis (Gu, Zhang and Lupski, 2008). Both SNPs and CNVs have been associated to the development of human diseases and have been proposed as mechanism for the development of psychiatric disorders (Need *et al.*, 2009; Thapar and Cooper, 2013; Nishioka *et al.*, 2018).

CNVs affecting the long arm of chromosome 15 are estimated to be in the general population with a frequency of 0.3% (Grozeva *et al.*, 2012). Low copy DNA repeat clusters, also called duplicons, are present in this area. Misalignment of these sequences during meiosis are responsible for non-allelic homologous recombination causing the formation of chromosomal abnormalities (Locke *et al.*, 2004). Five breakpoints (BP) are present in the proximal region of the long arm of chromosome 15 (**Figure 1.6Figure 1.6**). Deletions and duplications in this region can involve combinations of these BPs (Butler, 2017).

Deletions of the paternal copy of the 15q11-q13 region are the cause for PWS, whereas deletions of the maternal chromosome are the cause for AS (Nicholls and Knepper, 2001; Butler, 2017). This difference in the manifestation of the phenotype is due to the fact that multiple loci in the 15q11-q13 region are subject to genomic imprinting. During imprinting, specific alleles inherited by one of the parents are methylated leading to silencing of the allele. Therefore, a deletion in the paternal copy of ch 15q11-q13 combined with the maternal imprinting of this region would lead to developing PWS. Conversely, the inheritance of a non-functional maternal copy of this same chromosomal region, combined with the imprinting of the paternal equivalent are the cause of the development of AS. These disorders are characterised by neurodevelopmental phenotypes of variable severity. The deletions can be classified as type I when deletion occurs between BP1 and BP3 (6.6 Mb), or type II when deletion occurs between BP2 and BP3 (5.3 Mb) (Butler *et al.*, 2008). Individuals carrying the type I deletion present more severe intellectual and behavioural deficits suggesting the presence of dosage-sensitive genes playing an important role in neural development (Butler *et al.*, 2004).

The BP1-BP2 region spans approximately 500 Mb and contains four genes: NIPA1 (non-imprinted in Prader-Willi and Angelman syndrome 1), NIPA2, CYFIP1 (cytoplasmic FMRP-interacting portion 1) and TUBGCP5 (tubulin gamma complex associated protein 5, (**Figure 1.6**)) (Chai *et al.*, 2003). Deletions affecting the 15q11.2 region have recently been classified as having a mild effect size (intellectual disabilities OR = 1.7, schizophrenia OR = 1.5 and epilepsy OR = 3.1), explaining only a small proportion of the phenotype presented by the carriers of this CNV (elise Jønch *et al.*, 2019).

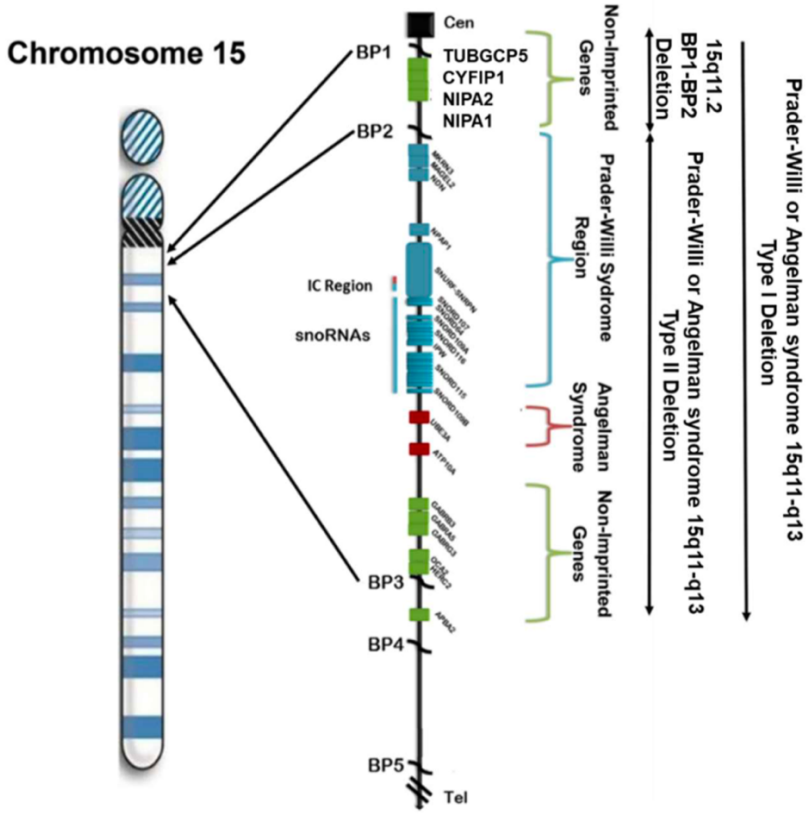


Figure 1.6 Ideogram of the 15q11.2 locus and the BPs affecting this region.
Schematic representation of genes between BP1 and BP5 in the 15q11.2 region. Non-imprinted genes are shown in green. In blue, genes expressed from the paternal allele (PWS critical region). In red, genes expressed only from the maternal allele (AS critical region). Figure was taken from (Butler, 2017).

NIPA1 and *NIPA2* encode for magnesium (Mg2+) transporters. NIPA1 is highly expressed in the developing and adult brain (Goytain *et al.*, 2007; van der Zwaag *et al.*, 2009) and mutations affecting this gene are the cause of hereditary spastic paraplegia (Fink, 2003). NIPA2 is expressed in several organs including the brain. Mutations in *NIPA2* have been implicated in the aetiology of childhood absence epilepsy (Goytain, Hines and Quamme, 2008; Jiang *et al.*, 2012; Xie *et al.*, 2014).

TUBGCP5 is part of the gamma-tubulin complex required for microtubule nucleation at the centrosome (Murphy *et al.*, 2001). This protein is highly expressed in subthalamic nuclei in the brain, a region related to the biology of ADHD and obsessive-compulsive disorder (Chai *et al.*, 2003; Doornbos *et al.*, 2009; Brem *et al.*, 2014).

CYFIP1 regulates mRNA translation via binding to FMRP, a protein with high expression in the CNS related to the development of fragile X syndrome (FXS). Moreover, CYFIP1 can also interact with the WAVE regulatory complex (WRC) to control the actin cytoskeleton dynamics (Napoli *et al.*, 2008; De Rubeis *et al.*, 2013). Therefore, CYFIP1 can be part of two multi-protein complexes (Chai *et al.*, 2003). Multiple studies have been reported to characterise the biological function of this protein, which will be reviewed later in this chapter.

CNVs affecting the region between BP1-BP2 of chromosome 15 (15q11.2) have been identified to contribute to the development of neurodevelopmental disorders. Murthy and colleagues reported the first study demonstrating the association between microdeletions affecting the 15q11.2 and neurodevelopmental disorders (Murthy *et al.*, 2007). The clinical phenotype presented by patients carrying a deletion at 15q11.2 include developmental delay and psychiatric disorders including ASD and SZ (Doornbos *et al.*, 2009; Burnside *et al.*, 2011; von der Lippe *et al.*, 2011; Abdelmoity *et al.*, 2012; Madrigal *et al.*, 2012; Cafferkey *et al.*, 2014; Jerkovich and Butler, 2015; Vanlerberghe *et al.*, 2015; Picinelli *et al.*, 2016). Microduplications affecting 15q11.2 are less abundant in the literature. However, these patients also present developmental delay and ASD as well as dysmorphic features, motor coordination issues and seizures (van der Zwaag *et al.*, 2009; Burnside *et al.*, 2011; Picinelli *et al.*, 2016).

GWAS have strengthened the association between CNVs in the 15q11.2 and psychiatric disorders. 15q11.2 microdeletions or microduplications are present at frequencies of 0.55-0.82% in SZ and above 0.8% in ASD cases (Stefansson *et al.*, 2008; Kirov *et al.*, 2009; Rees *et al.*, 2014). In the context of deletions, it has been estimated that on average CNVs are inherited from unaffected parents in 51% of the cases and 35% from affected parents. On the other hand, *de novo* and idiopathic mutations represent 5-22%

of the cases (Cafferkey *et al.*, 2014; Butler, 2017). This difference in the inheritance can be attributed to the low penetrance estimated for this CNV, where the presence of the mutation represents a small twofold increase over the overall population in the development of the disease (Rosenfeld *et al.*, 2013).

The data presented above strongly indicate that genes present in the 15q11.2 BP-BP2 have an important role in brain development and normal neuronal function. Dysregulated expression of these genes can significantly increase the risk of developing neuropsychiatric disorders such as SZ and ASD. In this context, CYFIP1 is considered the most promising causal risk gene in 15q11.2 CNVs, due to its known biological functions to be reviewed in the following section.

1.4 Biology of CYFIP1

1.4.1 Expression of CYFIP1 in the developing neocortex

In situ hybridisation analysis showed that *Cyfip1* is expressed in the mouse developing brain as early as E12 (van der Zwaag *et al.*, 2009). The signal appears to be enriched in the VZ of the forebrain. This data is in agreement with expression data available in the Eurexpress database (http://www.eurexpress.org) (**Figure 1.7 A**). *Cfip1* expression is maintained throughout embryonic development up to day E18.5. In the human brain, *CYFIP1* has been detected at week 8 post-conception (http://www.brainspan.org), although there is no data available prior to this stage.

Research carried out by Yoon and colleagues provided further support that Cyfip1 is expressed in the developing cortex (Yoon *et al.*, 2014). In this study, the expression of Cyfip1 appeared to be restricted to the VZ, specifically, on the ventricular surface of radial glial cells; and colocalised with NCAD and BCAT. The expression was lower in the intermediate zone. Cyfip1 protein expression was not found in the cortical plate, opposed to what has been found in the *in situ* hybridisation studies (Yoon *et al.*, 2014) (**Figure 1.7 B, C**).

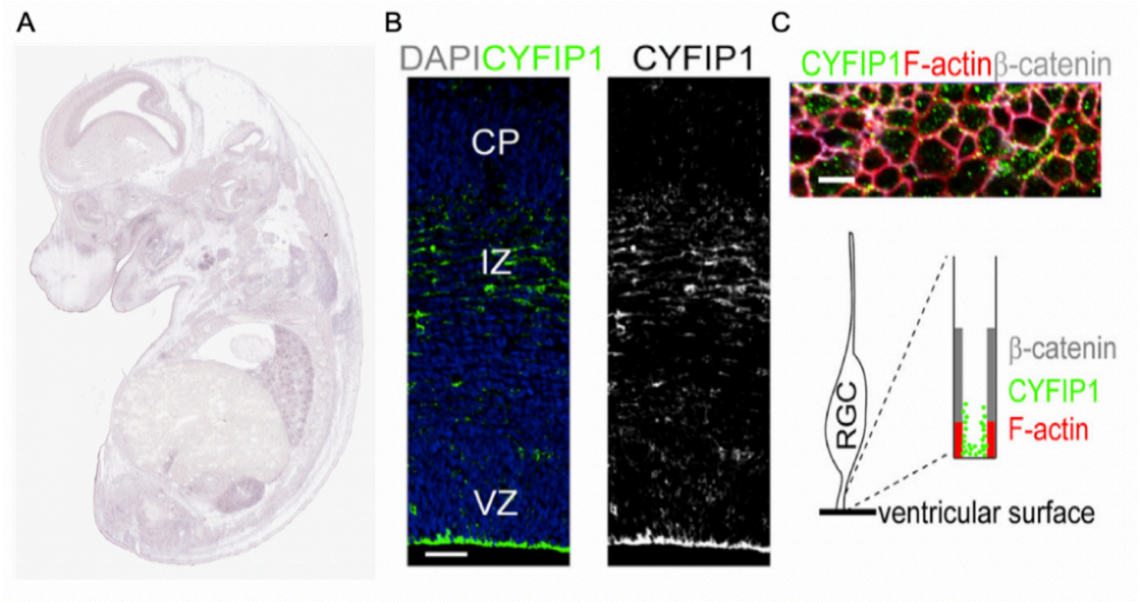


Figure 1.7 Expression pattern of Cyfip1 in the embryonic mouse brain.
(A) In situ hybridisation for Cyfip1 in mouse embryo E14.5 (from eurexpress.org). (B) Specific localisation of the expression of CYFIP1 in the VZ of mice brains at E15.5 and (C) in the RGC, where it accumulates close to the ventricular surface in regions rich in F-actin (adapted from Love et al 2014). CP: cortical plate, IZ: intermideate zone, VZ: ventricular zone.

1.4.2 Functions of CYFIP1 in the neural system

CYFIP1 was originally identified as a target of the small Rho GTPase Rac1 and was named as p140Sra-1 (Kobayashi *et al.*, 1998). Involved in the processes of actin remodelling and organisation, Rac1 plays an important role in the growth and maintenance of neuronal structures. The identification of CYFIP1 was achieved by screening a mouse embryonic library using a yeast two-hybrid system to identify proteins binding to the N-terminus of FMRP. In addition to CYFIP1, the assay also identified CYFIP2. Both proteins are largely evolutionary conserved and share nearly 88% of their amino acidic sequence. Despite the similarities, CYFIP2 can bind to both FXR1P and FXR2P, as well as FMRP, whereas CYFIP1 shows a specific association with FMRP.

However, Schenk and colleagues found that only a fraction of CYFIP1 could be immunoprecipitated with FMRP. The authors interpreted this finding as an independent activity of CYFIP1 from FMRP. Indeed, CYFIP1 and CYFIP2 had been previously identified as interactors of members of the WAVE (WASP (Wiskott-Aldrich syndrome protein)-family verprolin homologous protein) proteins forming the WRC, which controls the actin cytoskeleton dynamics (Eden *et al.*, 2002; Steffen *et al.*, 2004).

Two theories have been postulated on the role of CYFIP1 in the WRC. The first was proposed by Chen and colleagues, found that CYFIP1 played an inhibitory role in the WRC (Chen *et al.*, 2010). However, further research carried out by De Rubeis and

colleagues suggests that the presence of CYFIP1 in the WRC is required for the actiavation of the complex (De Rubeis *et al.*, 2013). To date, the later is the most accepted hypothesis on the activity of CYFIP1 in the WRC. Both hypothesis will be summarised in the following section.

(I) Role of Cyfip1 in the WRC

The WRC is formed by WAVE1/2/3, ABI (Abelson interacting protein), NAP (NCK-associated proteins), either CYFIP1 or CYFIP2 and HSPC300 (hematopoietic stem progenitor cell 300) (Chen *et al.*, 2010). These proteins form a hetero-pentameric complex controlling the actin polymerisation through Arp2/3. This process is fundamental for cell adhesion, migration, vesicle trafficking and neurite extension (Takenawa and Suetsugu, 2007). The presence of a VCA (Verprolin-homology, Central and Acidic regions) domain in WAVE proteins allows them to bind to Arp2/3 and G-actin monomers promoting actin polymerisation (Mendoza, 2013).

It was initially proposed that CYFIP1 and CYFIP2 could act as negative regulators of the WRC activity. In this model, the WRC would remain in a constitutively inactive form, with the VCA domain occupied by CYFIP1 and CYFIP2 preventing its interaction with Arp2/3. Specific cellular processes requiring the rearrangement of the actin cytoskeleton, such as the extension of lamellipodia, induce the activation of Rac1, which promotes the release of CYFIP1 and CYFIP2 from the VCA (Chen *et al.*, 2010). Negatively charged lipids can recruit the WRC to the cell membrane promoting the oligomerisation of the complex, which enhances its efficacy at generating cell protrusions such as lamellipodia (Oikawa *et al.*, 2004; Padrick *et al.*, 2008).

De Rubeis and colleagues investigated the mechanism CYFIP1 uses to alternate between the FMRP and the WRC complexes. Studying the crystallography structure of CYFIP1, the authors characterised mobile areas in CYFIP1 that could indicate changes in the 3D structure of the protein (**Figure 1.8 A**). To study this, the authors took advantage of Förster resonance energy transfer, where two fluorophores were linked to each end of the protein. If a conformational change occurs in the protein, a change in the fluorescence is observed. Using this methodology, the authors hypothesise that CYFIP1 has two conformations: a planar (open) and a globular (closed) conformation allowing the oscillation between the FMRP-eIF4E and the WAVE complex (**Figure 1.8 B**). The same study concluded that this conformational change was driven, in part, by neural stimulation through brain-derived neurotrophic factor. This neurotrophic factor promotes the activation of Rac1, which consequently phosphorylates CYFIP1, promoting a structural change from the globular to planar conformation. Activation by Rac1 releases CYFIP1 from the FMRP-eIF4E-CYFIP1 complex and promotes its recruitment to the

WRC allows protein synthesis and actin polymerisation (De Rubeis *et al.*, 2013) (**Figure 1.8 C**).

To validate this, the authors designed two mutant forms of CYFIP1 that were unable to interact with the WRC or the eIF4E. Each isoform was transfected separately into primary neurons, where CYFIP1 had been previously KD In this experiment, the authors observed that expression of the CYFIP1 form unable to interact with FMRP was able to restore normal actin levels. On the other hand, expression of the mutant form unable to bind to the WRC only rescued the translational regulation of some of the known FMRP targets (De Rubeis *et al.*, 2013). These results are conflicting with previous studies reporting a role for CYFIP1 as an inhibitory element in the actin polymerisation within the WRC. This difference could be explained by differences in the regulation of oscillation CYFIP1 between the two complexes.

(II) Cyfip1 regulation of FMRP-dependent mRNA translation

CYFIP1 and CYFIP2 interact with FMRP at the same domain that the latter uses for homo- and hetero-dimerisation with itself and other proteins, suggesting a potential role for CYFIP1 and CYFIP2 as modulators of FMRP activity (A. Schenck *et al.*, 2001). FMRP can form a complex with CYFIP1 and the initiation factor 4E (eIF4E), binding to mRNAs containing the short-sequences motifs TGGA, GAC and TAY (Anderson *et al.*, 2016). In this complex, CYFIP1 acts as a 4E-binding protein thus blocking the initiation of translation (Richter and Sonenberg, 2005; Napoli *et al.*, 2008).

FMRP is a ubiquitously expressed mRNA-binding protein, with high expression in the CNS. Mutations causing the loss-of-function of FMRP are the cause of fragile X syndrome (FXS), causing neurodevelopmental delay and autistic phenotype (Jacquemont *et al.*, 2007). In fact, FXS is a frequent cause of monogenic ASD and the most common form of inherited intellectual disabilities. FMRP acts as a translational repressor, and in neuronal cells, promotes the localisation of mRNAs to the dendrites (Fernández, Rajan and Bagni, 2013). Multiple studies have reported the identification of mRNAs targeted by FMRP (Brown *et al.*, 2001; Chen *et al.*, 2003; Darnell *et al.*, 2011), of which a significant proportion have been identified as ASD risk genes through GWA studies, while a small fraction overlaps with mood- and SZ-associated genes dendrites (Fernández, Rajan and Bagni, 2013).

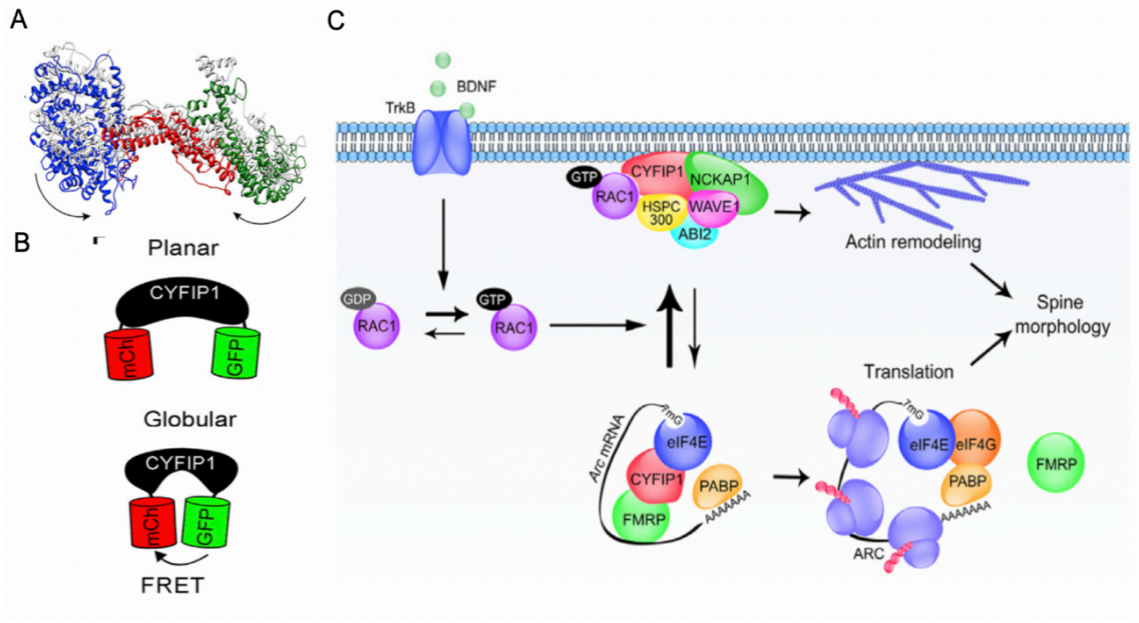


Figure 1.8 Structural conformation of CYFIP1 and oscillation between the FMRP and the WRC complexes.
(A) 3D model of the conformational structure of CYFIP1 predicting two conformations. Arrows indicate the main movements during structural conformational changes. (B) Sketch of the Förster resonance energy transfer model structure used to detect the changes in CYFIP1 conformation upon Rac1 activation. (C) Proposed model for the oscillation of CYFIP1 between complexes. Activation by Rac1 promotes the dissociation of CYFIP1 from the FMRP complex towards the WRC, allowing the protein synthesis of mRNAs targeted by FMRP and the actin polymerisation. Figure adapted from (De Rubeis *et al.*, 2013).

1.4.3 CYFIP1-dependent dysregulation in neural development

Animal and iPSC models in which *CYFIP1* dosage is manipulated have been used as a tool for understanding its role in neural development and neuropsychiatric disorders. The following section will focus on the current stage of research involving the role of CYFIP1 in neural development. Mouse embryos deficient in Cyfip1 (*Cyfip1* $^{-/-}$) die at embryonic day 8.5, demonstrating a pivotal role of this gene for embryonic development (Pathania *et al.*, 2014). The same study also showed that Cyfip1 appeared to be enriched in excitatory synapses in cultured hippocampal neurons. Reduction of Cyfip1 caused a decrease in dendritic branching, whereas high Cyfip1 expression increased the complexity in dendritic arborisation.

Increased dosage of CYFIP1 was also found to cause disruption of dendritic growth in human neuroblastoma cells and cortical cells of CYFIP1 transgenic mice (Oguro-Ando *et al.*, 2015). Neurites appeared to be shorter, accompanied by an increase in the number of branching and cell size in both models. Microarray analysis of cortical cells at day E13 of wild type (WT) and transgenic mice identified disruption in the mTOR pathway as the potential cause for the observed phenotypical changes. This hypothesis was confirmed by rescuing Cyfip1-caused morphological alterations with rapamycin, a specific inhibitor of mTOR (Oguro-Ando *et al.*, 2015).

Activity of CYFIP1 has a direct repercussion in dendritic spine morphology. *Cyfip1* KD in primary cortical neurons with shRNA caused a reduction of mature spine morphology (stubby and mushroom) whereas immature-looking spines were increased in number (De Rubeis *et al.*, 2013). Moreover, mean spine length was increased in *Cyfip1*-silenced primary cultured neurons. This correlates with the finding in FMRP KO mice models. The only consistent finding in FXS patients and reported FMRP KO mouse models is an abnormality in dendritic spines that are too long, thin and slightly more numerous (Irwin *et al.*, 2001; De Rubeis *et al.*, 2013).

CYFIP1 is one of the four genes present in the 15q11.2 BP1-BP2 CNV. Because large chromosomal aberrations are hard to replicate in animal models, CNV carrier derived iPSCs represent a useful tool to investigate the effect of CNVs involving multiple genes. The first report of 15q11.2 iPSCs came from Yoon *et al*. (Yoon *et al.*, 2014), who observed the abnormal organisation of the neural rosettes. This rosette defect was phenocopied by control iPSCs with shRNA mediated CYFIP1 KD, demonstrating CYFIP1 as the causal factor for this phenotype. The same authors also reported ectopic radial glial cell distribution outside the VZ in E13.5 mouse cortex following *in utero* electroporation of Cyfip1 shRNA, thus establishing a requirement for CYFIP1 in maintaining cortical progenitor polarity (Yoon *et al.*, 2014).

Another study by Das and colleagues showed morphological changes in dendritic spines in iPSC-derived neuronal cells carrying 15q11.2 deletion. However, no quantitative analysis was reported for this observation (Das *et al.*, 2015). The same paper also reported a reduction of *CYFIP1, NIPA1, NIPA2* and *TUBGCP5* at the mRNA level in undifferentiated iPSCs and the derived NPCs (Das *et al.*, 2015).

The effect of reduced CYFIP1 expression in iPSC-derived NPCs was further explored by Nebel and colleagues. In this study, the authors differentiated ‘control’ iPSCs into cortical NPCs and performed shRNA knock-down of CYFIP1 followed by RNAseq (Nebel *et al.*, 2016). Transcriptome analysis revealed dysregulation of genes involved in cytoskeletal regulation, a finding that is expected. Moreover, genes dysregulated in CYFIP1 KD NPCs appeared to be over-represented amongst SZ and epilepsy-associated risk genes (Nebel *et al.*, 2016).

1.4.4 Work leading to this thesis

In our laboratory, multiple hESCs lines have been engineered to either express excess levels of CYFIP1 through nucleofection of a constitutively active expression cassette, or loss of CYFIP1 via CRISPR/Cas9 induced mutagenesis. In this thesis, these cells are referred to as CYFIP1tg and CYFIP1ko cells, respectively, although sometimes they are also collectively mentioned as 'CYFIP1 mutant cells' for simplicity. These cells were respectively shown to have increased and decreased levels of CYFIP1 at the mRNA and protein level, making them suitable to study the alterations associated to abnormal expression of this protein during *in vitro* cortical glutamatergic differentiation. In addition to these, multiple clones of iPSC lines were derived from two patients harbouring a 15q11.2 BP1-BP2 deletion (named EA8 and EA62, respectively) and two non-affected control subjects. **Table 1.3** summarises all the phenotypical characteristics observed in cells with altered CYFIP1 expression.

Immediately after neural induction, the CYFIP1tg and CYFIP1ko cultures contained a similar proportion of cells expressing PAX6, FOXG1 and nestin as compared to their respective parental control lines. These findings indicated that increase or decrease of dosage of CYFIP1 do not affect the process of neural induction nor the regional identity or proliferation in the early stages of the differentiation.

However, structural abnormalities were found in the arrangement of NPCs when forming neural rosettes, an *in vitro* structure that recapitulates the embryonic neural tube (Abranches *et al.*, 2009). Rosettes act as a signalling hub where multiple signalling molecules are released by the neighbouring cells, affecting the way cells divide and differentiate. The apical zone of the neural rosettes can be labelled by cell adhesion molecules such as NCAD. In normal conditions, apical zones stained for NCAD take a ring-like shape. In CYFIP1tg and CYFIP1ko NPCs, the shape of the neural rosettes is affected and tends to take the shape of a collapsed or open ring, as well as being smaller in size. Moreover, CYFIP1tg cells presented a decreased number of neural rosettes. Alterations in NCAD distribution and neural rosettes structures were previously described by Yoon and colleagues (Yoon *et al.*, 2014), where the KD of CYFIP1 alone produced a scattered distribution of the NCAD in neural rosettes.

These findings were recapitulated in CYFIP1tg cells using a 3D model of differentiation. Following a previously reported protocol, PSCs were grown in suspension in order to generate embryoid bodies-like structures and later embedded in Matrigel® droplets to allow the expansion of the neuroepithelial tissue (Lancaster and Knoblich, 2014). The study of the thickness VZ-like areas generated by the cortical organoids showed that

CYFIP1tg organoids had a significantly smaller and thinner areas at the same analysed stage.

At later stages of differentiation, when mature neurons arise, strong phenotypical differences were identified between CYFIP1tg, CYFIP1ko and their respective control parental cells. CYFIP1tg maintained a significantly higher number of $PAX6^+$ NPCs while the number of $NEUN^+$ neurons was higher in the control population. This effect was found in the opposite direction in CYFIP1ko, which showed a lower number of $PAX6^+$ NPCs and a higher number of $NEUN^+$ neurons than its control parental line. This finding was further explored using EdU incorporation essays. The analysis revealed that CYFIP1tg retained a higher population of cells mitotically active compared to its control at later stages of the differentiation.

Further experiments revealed that these differences in maturation were due to changes in cell-cycle exit rate of the NPCs. Co-labelling of EdU with the post-mitotic mature marker NEUN. While CYFIP1tg cultures had a lower proportion of double-stained $EdU/NEUN^+$ neurons, CYFIP1ko showed the opposite effect. In addition to this, DNA content of the cells by flow cytometry showed that CYFIP1tg contained a significant number of cells in the S and M phases, whereas the control cells had exited the cell cycle. On the other hand, CYFIP1ko contained a larger number of cells in the G0/G1 phase and a reduced number in the M phase than their controls. Altogether, this indicated that CYFIP1tg and CYFIP1ko show opposite phenotypes in terms of cell-cycle progression.

While CYFIP1tg have a tendency for symmetrical, more proliferative, divisions than asymmetrical divisions generating post-mitotic neurons; CYFIP1ko show a premature exit of the cell cycle and generation mature neurons.

In addition to these findings, the work generated in this thesis indicated a potential dysregulation of the mitochondria. The analysis of this organelle in CYFIP1tg and CYFIP1ko showed opposing differences. While CYFIP1tg had a higher number of mitochondria than its control parental line, mitochondria in CYFIP1ko NPCs were less abundant and reduced in size compared to the control.

Finally, 15q11.2 deletion carrying cells were incorporated into the study. These iPSC derived cells were incorporated in the later stages of the research, therefore the phenotypical analysis of these was not carried out at the same depth as the CYFIP1tg and CYFIP1ko cells. The effects of the deletion were studied at the mRNA and protein level by RT-qPCR and Western blot analysis. Compared to the controls, CYFIP1 was found to be reduced at the mRNA and protein level in EA8 and EA62 NPCs and neurons. NIPA1 was not found to be reduced at the mRNA level. Furthermore, its expression at the protein level was equivalent to the control cells. On the other hand, NIPA2 was

strongly reduced in cells with a 15q11.2 deletion at the mRNA level, and the protein product was only detectable, and highly reduced compared to the control, at the NPC stage. TUBGCP5 showed variable results between the different studied cells at both mRNA and protein level, which were not found to be significant between mutant and control cells. During cortical differentiation, EA8 and EA62 cells showed a phenotype similar to the observed in CYFIP1ko. Differences in the ratio between cell proliferation and differentiation were observed in both patient-derived cells compared to controls, where the cells affected by the CNV differentiated prematurely into cortical glutamatergic neurons.

Table1.3 Summary of changes in cells with altered levels of CYFIP1 expression.

	CYFIP1tg	CYFIP1ko	15q11.2 del (EA8 and EA62)
Differentiation	NPC maintenance	Premature differentiation	Premature differentiation
Neural rosettes	Reduced number and abnormal structure	Abnormal structure	Abnormal structure
Cell-cycle	Higher proportion of cells in S and M phases	Higher number of cells in G1/G0 phases and reduced in phase M	-
Mitochondria	Increased number	Decreased number and size	-

1.5 Aims

This thesis aims to identify and manipulate the mechanisms behind the phenotypes observed in neural cells with altered expression of CYFIP1 during *in vitro* cortical differentiation. Furthermore, their association in the development of neuropsychiatric disorders will be analysed. The principal objectives are of this thesis are:

1. Analyse the transcriptional changes caused by altered CYFIP1 dosage in CYFIP1tg, CYFIP1ko and patient-derived cells carrying a 15q11.2 deletion, during *in vitro* cortical glutamatergic differentiation.

2. Characterise pathways and genes that could be contributing to the previously characterised phenotype in the studied neural cells.

3. Investigate the association between the CYFIP1-regulated genes and their enrichment for common variants conferring risk for neuropsychiatric disorders.

4. Manipulate through lentiviral CRISPR/Cas9 mediated mutagenesis, the expression of genes contributing to the development of the altered phenotype observed in the neural cells.

Chapter 2

Material and methods

2.1 Cell culture

2.1.1 hESCs and iPSCs

The CYFIP1tg and CYFIPko lines were derived from H7 (wicell.org) and iCas9 hESCs, respectively (González *et al.*, 2014, Tambrini et al, unpublished). Increased dosage of CYFIP1 was achieved by transgenic integration of a CYFIP1 constitutive expression vector. CYFIP1ko was generated through CRISPR/Cas9 directed genome editing. The iPSCs used in this study include two control lines derived from healthy donors, and two clones each of patient-derived iPSCs carrying a deletion in the 15q11.2 BP1-BP2 region. **Table 2.1** summarises all the cell lines used in this thesis. All of the cell lines were cultured in TesR-E8 media (Life Technologies) in standard culture conditions (37 ºC, 5% CO_2) in 6 well-plates coated with Matrigel® (Corning, VWR).

When cultures reached 70-90% confluence, the cells were passaged to freshly coated Matrigel plates. Briefly, the cells were washed with DPBS once and incubated in 0.02% EDTA (Sigma) for 3-4 minutes in the incubator. The EDTA was aspirated from the well and the cells dissociated into small clumps in TesR-E8 by gently scratching the well. The cells were seeded using a split ratio between 1:4 and 1:6. This allowed maintaining the cells in the same well for 3-4 days before doing a new split. Media for iCas9 and CYFIPko cells was supplemented with 1x RevitaCell (Gibco) when passaging in order to avoid extreme cell death.

To freeze cells, 70% confluent cultures were dissociated with EDTA as previously described, and the cell suspension was centrifuged at 200 g for 5 minutes at room temperature. The cell pellet was resuspended in 1 mL of cold TesR-E8 medium with 10% DMSO (Sigma) and transferred into cryo-vials (ThermoFisher). To achieve a cooling temperature of 1ºC/min, the vials were placed in Mr Frosty™ freezing container (ThermoFisher) at -80º C. For long term storage, the frozen cryo-vials were placed in liquid nitrogen tanks.

For thawing, cryo-vials placed in a 37º C water bath with gently shaking. Once thawed, the cells were transferred to 9 mL of pre-warmed TesR-E8 medium and centrifuged at 200 g for 5 minutes. Finally, the cell pellet was resuspended in 1 mL of TeseR-E8 medium and seeded in fresh Matrigel®-coated plates.

Table 2.1 Cell lines analysed in this thesis

Cell Line	Cell Type	Source	Clonal lines analysed
H7	hESC	wicell.org	1
iCas9	hESC	hue 9	1
CYFIP1tg	hESC	Genetically modified H7	1
CYFIP1ko	hESC	Genetically modified iCas9	1
EA8	iPSC	Patient-derived iPSC carrying 15q11.2 BP1-BP2 deletion	2
EA62	iPSC	Patient-derived iPSC carrying 15q11.2 BP1-BP2 deletion	2
900	iPSC	Control iPSC line	1
202	iPSC	Control iPSC line	1

2.1.2 Cortical neuron differentiation in monolayer

The protocol presented in this thesis used to generate pyramidal neurons is an adapted version of previously published ones (Chambers *et al.*, 2009; Cambray *et al.*, 2012; Arber *et al.*, 2015). The complete outline of the procedure and timeline for each cell line is shown in **Figure 2.1**.

Stem cells were initially plated in growth factor reduced Matrigel (Corning, VWR) and grown in TeseR-E8 medium until they reached 80-90% confluence. At this point, the cells were washed with DPBS, and the medium was replaced with neural induction medium composed by a mixture of N2B27 (without vitamin A) medium supplemented with the dual SMAD inhibitors LDN-193189 (TOCRIS) and SB-431542 (TOCRIS). The complete composition of the medium is described in **Table 2.2**. The switch to this medium defines the day 0 of differentiation. Due to important differences in the neural induction and maturation in H7 and iCas9 derived cells, the protocol was adapted accordingly for each cell line. iPSCs followed the same timeline of differentiation as the iCas9. The cells were kept in neural induction medium for 8 days in iCas9 cells and 12 days for H7. At this point, the LDN-193189 and SB-431542 were removed.

iCas9 and H7 cells were split at days 6 and 10 respectively in a 2:3 ratio onto fibronectin-coated plates. Fibronectin (Millipore) coating was performed by incubating a solution of 15 µg/mL in PBS at 37 ºC for at least 1h. Before splitting, the cells were treated for 1h with 100 µM ROCK inhibitor (Y-27632, STEMCELL Technologies). After this, cells were incubated in EDTA for 3 minutes and dissociated manually in N2B27 medium using a 2 mL serological pipette. The cells were carefully resuspended in the appropriate volume of neural induction media and seeded onto the fibronectin-coated plate.

A second split in a 1:4 ratio was done 8-10 days later. At this stage, the differentiating neural progenitors were plated on Poly-D-Lysin/Laminin coated plates. The coating was performed by first incubating a solution of 10 µg/mL of Poly-D-Lysin (Sigma) for 1h at room temperature. The Poly-D-Lysin solution was removed, and the wells were washed with DPBS three times. This was followed by an incubation overnight of 5 µg/mL Laminin solution (Sigma) at 37 ºC. Several days after the second passage, when cells display a clear neuronal morphology, the medium is switched to N2B27 supplemented with vitamin A in order to promote neuronal maturation. Cells were fed every other day throughout the whole protocol of differentiation.

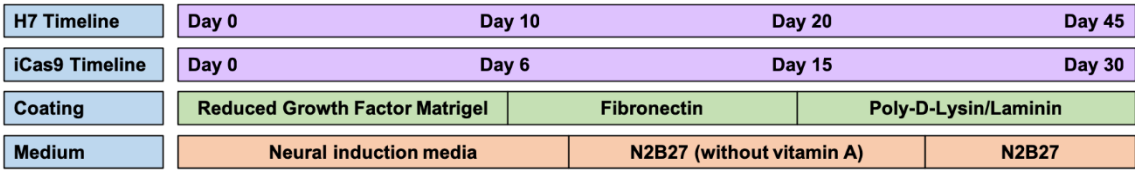


Figure 2.1 Timeline of cortical glutamatergic differentiation.
Timing of monolayer cortical differentiation for both parental lines is shown in purple, while substrates and media used at each stage are shown in green and orange respectively.

Table 2.2 Composition of media used for neural differentiation and HEK cell maintenance.

Medium	Composition
N2B27 (without vitamin A)	2:1 DMEM-F12 and Neurobasal, 1x N2, 1x B27 without vitamin A, 2mM L-Glutamine, 0.1 mM β-mercaptoethanol (All from ThermoFisher), 1x Mycozap (Lonza)
Neural induction medium	N2B27 (without vitamin A), 10 µM SB-431542 (TOCRIS), 100 nM LDN-193189 (TOCRIS)
N2B27 (with vitamin A)	2:1 DMEM-F12 and Neurobasal, 1x N2, 1x B27 with vitamin A, 2mM L-Glutamine, 0.1 mM β-mercaptoethanol (All from ThermoFisher), 1x Mycozap (Lonza)

2.1.3 HEK maintenance

Human embryonic kidney (HEK293T) cells were cultured in gelatine-coated P10 plates or T75 flasks (Falcon). The coating was done by incubating a solution of 0.2% gelatine (Sigma) for 10 minutes at 37 ºC. Cells were routinely passaged every three days using a 1:10 split ratio. To passage the cells, the media was removed, and cells were washed with PBS. Cells were enzymatically dissociated using 1% trypsin (Gibco) for three minutes at 37 ºC. Trypsinisation was stopped by adding 2 volumes of HEK medium composed of: 450 mL DMEM-F12, 0.1 mM β-mercaptoethanol, 2 mM non-essential amino acids (all from Thermo Fisher), 1x Mycozap (Lonza), 1x heat-inactivated foetal bovine serum (Biosera). The cell suspension was recovered in a universal tube (Gibco) and centrifuged at 200 x g for 5 minutes. The cell pellet was then resuspended in 10 mL of media and seeded in the required ratio.

2.2 Immunofluorescence

Cultured cells were washed with 1x DPBS and fixed with cold 3.7% PFA for 10 – 15 minutes. Cells were washed 3x post-fixation with DBPS. Cells were then permeabilised with three washes of PBS-T (0.3% Triton-X-100 in PBS) for 10 minutes each. Blocking was done with 2% BSA and 3% donkey serum in PBS-T for 20 minutes at room temperature. Cells were incubated with primary antibodies in blocking solution at 4º C overnight (**Table 2.3**). The following day, the cells were washed 3 times in PBS-T for 10 minutes at room temperature. Incubation with AlexaFluor® (Thermo Fisher) secondary antibodies was done in PBS-T for 2 hours at room temperature, in the dark. Nuclei were counterstained with DAPI in PBS for 5 minutes. After three consecutive washes with PBS, cells were mounted with DAKO (Agilent) fluorescent mounting medium.

Table 2.3 Primary antibodies used for immunoflourecense.

Target	Species	Dilution	Cat. Number	Supplier
DCX	Goat	1:1000	SC-8067	Santa Cruz
FOXG1	Rabbit	1:1000	Ab18259	Abcam
MAP2	Rabbit	1:1000	AB5622	Abcam
N-Cadh	Mouse	1:1000	18-0224	ThermoFisher
NES	Mouse	1:1000	BD611659	BD
PAX6	Rabbit	1:1000	901301	BioLegend
TBR1	Rabbit	1:500	Ab31940	Abcam
TBR2	Rabbit	1:500	Ab23345	Abcam
TUJ1	Mouse	1:1000	T8600	Sigma

2.3 Imaging and picture analysis

Stained cells were imaged with a Leica DM6000 B (Leica Microsystems) inverted microscope. For cell counting, an average of 10 random fields were acquired for each well at 20x magnification. DAPI and most nuclear markers were counted with Cell Profiler (cellprofiler.org). The remaining cellular markers were counted manually with Fiji (Image J) software. Statistical analysis and visual representation were carried out with R. Unless stated otherwise, all immunohistochemistry quantifications were collected from two independent experiments and, at least, two biological replicates for each marker counted.

2.4 Western blotting

Cells were scratched off in ice-cold PBS and pelleted by centrifuging at 1000 x g for 5 minutes at 4ºC. For long term storage, the cell pellet was kept at -80 ºC. To extract the proteins, cells were lysed in a cocktail of RIPA buffer (Abcam) with protease and phosphatase inhibitors (Sigma). The cell lysate was incubated for 30 minutes in ice, and vortexed every 5 minutes throughout the process. The supernatant was cleared by centrifuging the suspension at 14,000 rpm for 5 minutes at 4 ºC. 120 µL of the supernatant were transferred to a fresh 1,5 mL tube and mixed with 4x Bolt® LDS buffer

(ThermoFisher) and 1 M DTT (Sigma) reaching a final volume of 175 µL. The final product was boiled at 72 ºC for 10 minutes. The remaining volume of the lysate was used to quantify the protein concentration using the DC™ protein assay reagent (Bio-Rad) and compared against a protein standard (Bio-Rad), following the instructions supplied by the manufacturer. Absorbance at 705 nm to determine protein quantification was carried out in a CLARIOStar Plus (BMG Labtech).

For each sample, equal amounts of protein were loaded in 4-12% Bolt® Bis-Tris Plus gels (ThermoFisher). Proteins were transferred into a 0.45 µM pore size PDVF membrane (Amersham) through electro-blotting. 5% Bovine serum albumin (BSA, Sigma) (w/v) was diluted in Tris Buffered Saline (Formedium) containing 0.1% Tween® 20 (Sigma) (TBS-T) and used to block the PDVF membrane for 2 h at room temperature. Incubation of primary antibodies was performed overnight at 4 ºC in fresh blocking solution (**Table 2.4**). The following day, the membrane was washed 3x with TBS-T for 20 minutes each at room temperature. HRP conjugated secondary antibodies (Abcam) were incubated for 1h at room temperature in blocking solution. After washing the membrane 3x with TBS-T, it was incubated for 2 to 5 minutes with Crescendo HRP substrate (Millipore). Chemiluminescence was detected using iBrigth 2000 (ThermoFisher). Quantification was carried out in Fiji (Image J). All samples were normalised to GAPDH.

Table 2.4 List of primary antibodies for Western blotting.

Target	Species	Dilution	Cat. Number	Supplier
AKT3	Rabbit	1:1000	149825	CST
GAPDH	Mouse	1:5000	Ab8245	Abcam
GSK3B	Mouse	1:1000	98325	CST
LRP6	Rabbit	1:1000	3395S	SCT
pAKT (S473)	Rabbit	1:1000	40585	CST
pGSK3B (S9)	Rabbit	1:1000	55585	CST
pLRP6	Rabbit	1:1000	3668S	CST

2.5 Production and titration of lentiviral particles

2.5.1 Plasmid preparation

To generate the lentiviral CRISPR/Cas9 construct, plasmids from (Kabadi *et al.*, 2014) were obtained from Addgene (Addgene, #53186 - #53190) in the form of agar stabs. Two more plasmids, already present in our lab, were also required for the synthesis of the lentivirus: pMD2.G (Addgene #12259) containing the VSV-G lentiviral envelope and the psPax2 (Addgene #12260) containing the lentiviral packaging sequences. All plasmids are listed in **Table 2.5**. Bacterial cells were grown in LB agar plates or LB medium depending on the application required. Antibiotic selection was applied by adding either 100 µg/mL ampicillin (Sigma) or 50 µg/mL kanamycin (Sigma) as required.

The bacteria contained in agar stabs were first streaked onto LB agar plates with the required selection antibiotic and grown overnight at 37 ºC. The following day, single colonies were picked and grown in conical tubes (Falcon) containing 2 mL of LB supplemented with the required antibiotic. Bacterial suspensions were grown overnight at 37 ºC in the incubator with shacking at 200 rpm.

The following morning, bacterial cultures were pelleted in microfuge tubes at 11,000 x g for 1 minute. The supernatant was discarded, and bacterial pellet was thoroughly resuspended in of solution A (25 mM Tris-HCl, pH 8,0, 10 mM EDTA) with µg/mL RNAse A (Qiagen). An equal volume of lysis buffer (400 mM NaOH, 2% SDS) was added to the cells and mixed by gently inverting the tubes. After 5 minutes, lysis was stopped by adding neutralising buffer (5M KAc pH 5,5) mixed by inversion and incubated for 3 minutes. The cell lysate was cleared by centrifuging at 12,000 x g for 2 minutes. The aqueous fraction containing the plasmidic DNA was transferred to a fresh tube. DNA was concentrated and purified using alcohol method. Briefly, an equal volume of isopropanol (Sigma Aldrich) was added to the DNA suspension and mixed by shacking. The suspension was centrifuged at 18,000 x g for 1 minute. Isopropanol was removed, and the DNA pellet was subsequently washed 70% ethanol (Sigma) and centrifuged again. Ethanol was removed, and the DNA pellet was left to air-dry for 10 minutes and resuspended in an appropriate amount of ddH_2O.

To confirm plasmid the plasmid integrity, each extracted plasmidic DNA was digested using a specific combination of restriction enzymes. Enzymatic digestion was carried out at 37 ºC for 15 minutes using the appropriate digestion buffer as required by the manufacturer.

When large amounts of DNA were required, a starter culture of 250 mL of LB inoculated with the bacterial strain to be amplified was grown overnight at 37 ºC with shacking at 200 rpm. Bacterial cultures were pelleted by centrifuging at 5,000 x g for 10 minutes. Cell pellets were resuspended and lysed, as described above. The cell lysate was cleared by centrifuging at 14,000 x g for 20 minutes in a fixed angle rotor. DNA was purified using the PureYield™ (Promega) column system using a vacuum manifold following the manufacturer's instructions. When amplifying plasmids for cell transfection, an extra step of endotoxin removal wash was included when recovering the plasmidic DNA. The DNA was finally eluted with nuclease and endotoxin-free water (Promega).

Table 2.5 List of plasmids used for lentiviral construct

Plasmid	Use	ID[a]	Reference	Resistance
phH1-gRNA	Expression of *S. pyogenes* sgRNA controlled by a human H1 promoter.	#53186	(Kabadi *et al.*, 2014)	Kan
pmU6-gRNA	Expression of *S. pyogenes* sgRNA controlled by a mouse U6 promoter.	#53187	(Kabadi *et al.*, 2014)	Kan
phU6-gRNA	Expression of *S. pyogenes* sgRNA controlled by a human U6 promoter.	#53188	(Kabadi *et al.*, 2014)	Kan
ph7SK-gRNA	Expression of *S. pyogenes* sgRNA controlled by a human 7SK promoter.	#53189	(Kabadi *et al.*, 2014)	Kan
pLV hUbC-Cas9-T2A-GFP	3rd generation lentiviral transfer vector co-expressing Cas9 and GFP under a human ubiquitin promoter.	#53190	(Kabadi *et al.*, 2014)	Amp
pMD2.G	VSV-G envelope for lentiviral packaging.	#12259	[b]	Amp
psPAX2	2nd generation lentiviral packaging plasmid.	#12260	[b]	Amp

a) Addgene ID, b) no publication associated to this plasdmid, the construct was deposited to Adggene by Dider Trono. Amp: ampicillin; Kan: kanamycin.

2.5.2 Assembly of sgRNAs into lentiviral vector

Four individual sgRNAs targeting ATK3 were designed with the Zang lab tool (www.crispr.mit.edu). The top four guides with the lowest off-target scores were selected. In order to facilitate the correct insertion of the sgRNAs into the promoter plasmids, additional bases, complementary to the plasmids overhangs, were included on each end of the sequences **Table 2.6**. In order to generate the construct, each of the sgRNAs were annealed with their complementary strand. To perform the annealing, equal volumes of 10 µM of each complementary guide were mixed with of 10x ligase buffer (NEB) in a 10

µL reaction. The mix was incubated at 96 ºC for 5 minutes in the thermocycler, and the temperature was decreased by 0.3 ºC per second until reaching a final temperature of 25 ºC. Overhangs were phosphorylated by incubating the annealed guides with 25 mM ATP (Sigma) and of T4 Polynucleotide Kinase (NEB). The mix was incubated for 1h at 37 ºC and followed by a 20 min incubation at 65 ºC to inactivate the enzyme.

Table 2.6 List of promoter plasmids and sgRNAs inserted into each including overhangs

Promoter	Oligo name	Overhang	Protospacer	Overhang
mU6	AKT3 sgRNA-Sense-1	5'-TTGTTTG	TAAGGTAAATCCACATCTTG	
	AKT3 sgRNA-Antisense-1	3'-AAC	ATTCCATTTAGGTGTAGAAC	CAAA-5'
hU6	AKT3 sgRNA-Sense-2	5'-CACCG	TTGCCACTGAAAAGTTGTTG	
	AKT3 sgRNA-Antisense-2	3'-C	AACGGTGACTTTTCAACAAC	CAAA-5'
7SK	AKT3 sgRNA-Sense-3	5'-CCTCG	TCCCCTCAACAACTTTTCAG	
	AKT3 sgRNA-Antisense-3	3'-C	AGGGGAGTTGTTGAAAAGTC	CAAA-5'
H1	AKT3 sgRNA-Sense-4	5-'TCCCA	GAAAAGTTGTTGAGGGGATA	
	AKT3 sgRNA-Antisense-4	3'-T	CTTTTCAACAACTCCCCTAT	CAAA-5'

Each annealed and phosphorylated guide was ligated into the corresponding complementary plasmid containing a promoter. First, 1 µg of each vector was digested with BbsI (NEB) for 15 minutes at 37 ºC. 50 ng of the linearised product were mixed with 1 µL of the annealed guides and incubated overnight at 4 ºC with T4 DNA ligase (NEB) in a 10 µL reaction. 5 µL of the ligation product were used to transform XL1 blue chemically competent bacteria (Agilent). Briefly, competent bacteria were incubated in ice with the ligation product for 30 minutes. After this, the bacteria were heat-shocked in the water bath at exactly 42 ºC for 30 seconds. The bacteria were placed again in ice for 2 minutes, and 1 mL of super optimal broth with catabolite repression outgrowth medium (SOC, ThermoFisher) was added into each tube. The bacteria were allowed to recover for 2 h incubated at 37 ºC and shaking at 280 rpm. The bacterial suspension was plated on LB agar plates containing kanamycin, 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG, Promega) and 80 µg/mL 5-Bromo-4-Chloro-3-Indolyl (X-Gal, ThermoFisher). Plates were incubated overnight at 37 ºC. The following day, individual bacterial colonies were picked for plasmid preparation. Thanks to the lacZ complementation, white-blue discrimination was used to pick colonies with an insert. Insert presence was confirmed by restriction enzyme digestion and PCR amplification of the purified plasmids.

The previous step generated four individual constructs containing a single sgRNA and promoter. In order to assemble the final construct, 200 ng of each sgRNA construct and expression vector containing the Cas9-T2A-GFP were combined with 1 µL T4 DNA ligase, 1 µL BsmBI FastDigest (Fisher Scientific), 2 µL 10x T4 ligase buffer (NEB) and 20

mM DTT in a 20 µL reaction. The final construct was generated by golden gate reaction incubating the mixture at 37 ºC for 10 min, 16 ºC 15 min, 37 ºC for 30 min and 80 ºC for 5 min. 5 µL of the ligation product were transformed in Stable chemically competent bacteria (NEB) as described above and plated on LB agar plates with 100 µg/mL of ampicillin.

2.5.3 Packaging and titration of lentiviral particles

Lentiviral particles were packaged in HEK cells. The system combines three second-generation lentiviral plasmids: pMD2.G (Addgene #12259) containing the VSV-G lentiviral envelope, psPax2 (Addgene #12260) containing the lentiviral packaging and the final construct generated through golden gate assembly with the hUbC-Cas9-T2A-GFP and the four sgRNAs and their promoters. The three plasmids were co-transfected using lipofectamine 3000 (Thermo Fisher) using the forward transfection method. On the day before transfection, 7 * 10^6 HEK cells were plated on a P10 plate in lentiviral packaging media supplemented with 5% FBS and 100 mM of sodium pyruvate (Sigma). The following day, when the cells reached a 90-99% confluence, they were co-transfected with 20 µg of the Cas9 construct, 6 µg of the pMD2.G and 10 µg of the psPax2. 6h post-transfection the medium was replaced with fresh lentiviral packaging medium. Lentiviral supernatant was collected at 24 and 52h post-transfection and stored at 4 ºC until use.

The lentiviral supernatant collected at the two time points were pooled into a single tube and centrifuged at 2,000 x g for 5 minutes at 4 ºC to remove cell debris. The supernatant was cleared through a 45 µm pore PDVF membrane (Fisherbrand). The cleared supernatant was then centrifuged at 92,000 x g in an L 100 XP ultracentrifuge (Beckman Coulter) for 2.5h at 4 ºC in the vacuum. The supernatant was decanted into 20% Chemgene (Starlab) for disinfection. The tubes were airdried upside down for 3 minutes to remove the excess of fluid in them. The pellet, not always visible, was resuspended in an adequate of PBS volume to obtain a 1,000-fold concentrated virus. Finally, the viral suspension was aliquoted in single-use aliquots and stored in the -80 ºC if not used immediately. The complete process is described in **Figure 2.2**.

Lentiviral titer was calculated as described in (Barde, Salmon and Trono, 2010). Briefly, 150,000 HEK cells were plated on a 12 well plate the day prior to infection. The following day, one well was used for counting to determine the exact cell number. The media was removed, and the cells were carefully washed once with DPBS and replaced for 500 µL of pre-warmed HEK media. Serial dilutions of the lentiviral particles were done in DPBS of up to 10^{-4} µL of viral suspension and a non-infected control. The HEK cells were transduced with 10 µL of the diluted virus in duplicates. 24h post-transduction the media was discarded, cells were washed with DPBS, and 2 mL of pre-warmed HEK media was

added to each well. Three days post-transduction the cells were dissociated in single-cell suspension with Acutase (STEMCELL Technologies) incubating the cells for 10 minutes in the incubator. Three volumes of HEK media were added to the cells to stop the dissociation and transferred into a FACS tube (Falcon) passing through a 0.45 µm cell strainer. The cell suspension was centrifuged for 5 minutes at 500 x g. The supernatant was removed and cells were fixed in 1% PFA for 5 minutes at room temperature. The cell suspension was centrifuged for 5 minutes at 500 x g, and the cells pellets were resuspended in 1 mL of DPBS. Expression of GFP was assessed using a Fortessa flow cytometer (BD). Final viral titer was calculated with the following formula:

$$Titer = \frac{Number\ of\ target\ cells\ (day\ 1) * \left[\frac{\%\ GFP - positive\ cells}{100}\right]}{Volume\ of\ lentivirus\ (in\ mL)}$$

Each collection yielded an approximate titer of 10^9 lentiviral particles/mL.

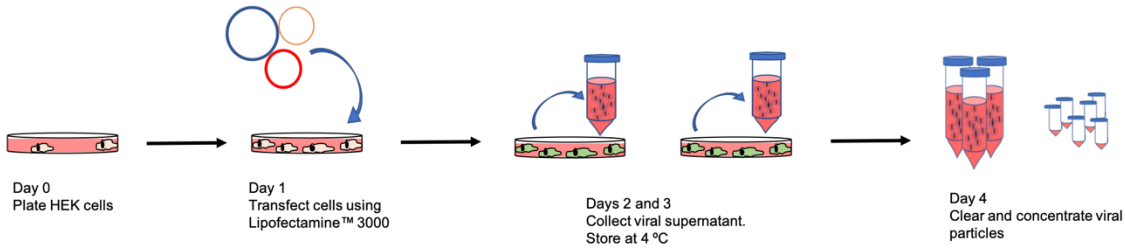


Figure 2.2 Workflow for lentiviral production.
Summarised steps for packaging and concentrating lentiviral particles. On day 0, HEK cells are plated. The following day, the cells are transfected with the necessary plasmids to generate the viral particles. On days 2 and 3, the supernatant containing the lentivirus are collected and pooled. Lastly, the lentiviral supernatant is cleared from cell debris, concentrated and aliquoted in single-use tubes.

2.5.4 Lentiviral transduction in NPCs

To determine the efficiency of the lentiviral infection, control NPCs (H7) were dissociated with Acutase at day 18 of differentiation as previously described. 200,000 NPCs were plated onto poly-D-lysin/laminin 24 well plates in N2B27 (without vitamin A) medium. The following day, the medium was removed, and cells were washed 3 times with DPBS, and 250 µL of fresh pre-warmed medium was added. Cells were transduced in triplicates where two wells were used for immunostaining and the remaining well for DNA extraction. Three different units (0; 200,000; 300,000 and 400,000) of lentiviral particles were tested to determine the lentiviral infectivity. The following day, the media containing the lentiviral particles was removed, and cells were washed three times with DPBS. 1 mL of fresh pre-warmed N2B27 was added. Cells were grown for 5 more days to allow

expression of GFP and Cas9 activity. After this, cells were fixed, or genomic DNA (gDNA) was extracted (see next section).

For the lentiviral *AKT3* targeting, CYFIP1tg and their isogenic control cultures were transduced at day 18 of differentiation using the conditions described above. Since a 1:1 (cell:virus) ratio was deemed sufficient to achieve efficient transduction, each well of a 12 well plate was infected with 6,000,000 viral particles. The number of cells was estimated from previous experiments carried out in our laboratory where NPCs were dissociated and counted at this stage. Following the previously stated protocol of monolayer cortical differentiation, cells were passaged onto poly-D-lysine/laminin-coated plates at day 20. Cells were grown in N2B27 (without vitamin A) for 10 more days before switching to complete N2B27. Cells were fixed and DNA or protein were extracted, as required, at days 35 and 45.

2.6 sgRNA efficiency test

2.6.1 Genomic DNA extraction

Cultured cells were washed once with DPBS. The cells were incubated at 37 ºC overnight in lysis buffer (10 mM Tris-pH8.0, 50 mM EDT, 100 mM NaCl, 0.5% SDS) supplemented with 0.5 mg/mL of Proteinase K (all from Sigma). The following day, the cell lysate was mixed with an equal volume of isopropanol, and the DNA was precipitated at 15,000 x g for 20 minutes at 4 ºC. The pellet was washed with 70% ethanol, air-dried and resuspended in an appropriate amount of ddH_2O. The DNA concentration was measured with Nanodrop 2000 (ThermoFisher).

2.6.2 PCR and DNA electrophoresis

PCR was used for screening for the insertion of the multiple sgRNAs in the lentiviral donor template and to assess the cutting efficiencies of such guides. Each PCR reaction included about 100 ng of template plasmid or genomic DNA, 1x Standard Buffer (NEB), 200 µM dNTPs (Sigma), 0.2 µM of each forward and reverse primer (Sigma) and 1.25 units of Taq polymerase (NEB) per reaction. Usually, 35 amplification cycles were performed in a Mastercycler X50s (Eppendorf). Each cycle included 30 seconds of denaturation at 95 ºC, 30 seconds of annealing at 51 – 57 ºC and 30 – 60 seconds of extension at 62 ºC. The annealing temperature and length of extension interval were optimised for each pair of primers (**Table 2.7**). Finally. 5 µL of the PCR product were run on an agarose gel (0,7 to 2.5 % agarose, according to the size of the PCR amplicon).

Table 2.7 List of primers and PCR conditions

Target	Forward	Reverse	Use	Cycling conditions	Amplicon size (bp)
sgRNA PCR	TCGGGTTTATTACAGGGACAGCAG	TCTAAGGCCGAGTCTTATGAGCAG	PCR across gRNA region	95°C 5 min 95°C 30 sec | 56°C 30 sec | 35 X 62°C 2 min | 68°C 5 min	1,800
AKT3	TGCGTATCGTTGAAAGCCCT	ACCCAAACTTTTTCACAGGTTTCT	CRISPR/Cas9 target region	95°C 5 min 95°C 30 sec | 51°C 30 sec | 35 X 62°C 1 min | 68°C 5 min	826
CCNG2	GGAGGTCACCAGCCTCTATC	CAAGTCAACGGGGGTAAGGA	Off-target screening	95°C 5 min 95°C 30 sec | 53°C 30 sec | 35 X 62°C 1 min | 68°C 5 min	671
SNRPF	CAACCTTTTGCAGTGGCCTT	ATTTCTGTTGTGCCCTGTCTGTA	Off-target screening	95°C 5 min 95°C 30 sec | 52°C 30 sec | 35 X 62°C 1 min | 68°C 5 min	953
COL7A1	GGACCTGAGGGCTCCTATTG	CCCGAGGACCAATCACACT	Off-target screening	95°C 5 min 95°C 30 sec | 53°C 30 sec | 35 X 62°C 1 min | 68°C 5 min	535
BORCS5	TGCCCAGCTCTTTTCCTTGA	GAGGGTGTGGAAAGGAAGGC	Off-target screening	95°C 5 min 95°C 30 sec | 53°C 30 sec | 35 X 62°C 1 min | 68°C 5 min	470
AKT1	GTCGCTGGCCCTAAGAAACA	CCAGTGCTTGTTGCTTGCC	Off-target screening	95°C 5 min 95°C 30 sec | 54°C 30 sec | 35 X 62°C 1 min | 68°C 5 min	416

Primers used for PCR and their application.

2.6.3 Sanger sequencing and TIDE analysis

Prior to sequencing, primers were removed from PCR products by enzymatic digestion. Every 5 µL of PCR product were treated with a combination of exonuclease I (Exo I, NEB) and shrimp alkaline phosphatase (rSAP, NEB) to digest and dephosphorylate the remaining primers and dNTP. The mixture was incubated for 15 minutes at 37 ºC, and enzymes were inactivated by incubation at 80 ºC for 15 minutes.

Sanger sequencing was performed by LGC Genomics. Briefly, between 40 and 400 ng of cleaned PCR product or plasmidic DNA were mixed with 5 µM of the necessary primer for sequencing. The sequence containing the targeted region was amplified by PCR and sequenced in order to quantify the efficiency of each sgRNA using the online tool TIDE (https://tide.deskgen.com). TIDE uses a decomposition algorithm to quantify the presence of insetions and deletions (InDels) in a pool of edited cells Sequences generated from treated samples were aligned against non-transduced samples in TIDE. This uses an algorithm of decomposition to identify changes in the pool of sequences generated and approximate the number of sequences edited by a sgRNA.

2.7 RNA sequencing and transcriptome analysis

2.7.1 RNA extraction, purification and RIN determination

RNA was collected for sequencing at three time-points of differentiation, for each cell type, in triplicate. Cells were lysed with TriReagent (Ambion) for 5 minutes at room temperature to allow the dissociation of nucleoproteins. Total RNA was isolated using the phenol/chloroform method. 200 µL of chloroform were added to each sample per each 1 mL of TriReagent used and mixed through vigorous shaking without vortexing. The mix was incubated for 5 minutes at room temperature and centrifuged at 12,000 x g for 15 at 4º C. The colourless upper phase containing the RNA was transferred to a new tube, and an equal volume of 100% ethanol was added. To remove any gDNA, the mixture was briefly vortexed and transferred to PureLinkTM RNA Mini Kit column (Ambion) and centrifuged at 12,000 x g for 1 min at room temperature. Columns were washed twice with Wash Buffer II and spun for 15 seconds at 12,000 x g. Columns were dried for 1 minute. RNA was finally eluted in a volume of 30 µL of nuclease-free water (Ambion).

RNA integrity number (RIN) and purity were determined using the Bioanalyzer 2100 (Agilent). A freshly prepared gel-dye mix was loaded into an RNA 6000 Nano chip (Agilent). Once the gel was ready, the RNA markers and ladders were loaded into their respective wells. RNA was heated to 70º C for 2 minutes to reduce the number of secondary structures, and 1 µL of each RNA sample was added into each well. The RNA chip was vortexed for 60 seconds at 2,400 rpm. Finally, the chip was analysed within 5 minutes after vortexing and processed with the Bioanalyzer.

2.7.2 Library preparation and sequencing

Library prep and sequencing was performed at the Oxford Genomics Centre (http://www.well.ox.ac.uk/ogc). Here, 3 µg of total RNA were used to construct the library for sequencing. Briefly, the mRNA fraction was isolated by poly-A selection before conversion to cDNA and incorporation of dUTP. cDNA was end-repaired, A-tailed and adapters were ligated to each end. Before PCR amplification, samples undergo uridine digestion. Final libraries were size selected and multiplexed.

Paired-end sequencing was performed by distributing the multiplexed samples over multiple lanes of a flow cell and sequenced was performed on an Illumina HiSeq 4000 (Illumina). For all samples, a minimum of 28 million reads were obtained, and an average library size of 33 million reads.

2.7.3 Sequencing QC and mapping

Figure 2.3 describes the workflow for QC and mapping of the sequencing using an adapted workflow previously designed by Dr Robert Andrews (Cardiff University). The pipeline was run in the Advanced Research computing @ Cardiff (ARCCA). First, FastQC (v 0.11.2) is used to determine the quality of the reads for each of the samples. Bases with a low score in the Phred scale were removed with Trimmomatic to avoid misalignments when mapping the reads to the reference genome. FastQC is used again to determine the accuracy of the trimmed sequences. The remaining sequences are aligned to the Ensembl human reference genome GRCh38.84 (hg38) using STAR (2.5.1b). To avoid the loss of reads mapping to multiple regions of the genome, the native behaviour of STAR was modified. Using this approach, conflicting sequences were given an alignment score and assigned to the region with the highest score. in the case where two, or more, regions had the same score, STAR would randomly assign the read to one of them. Final gene counts were obtained calling featureCounts from the HTSeq library.

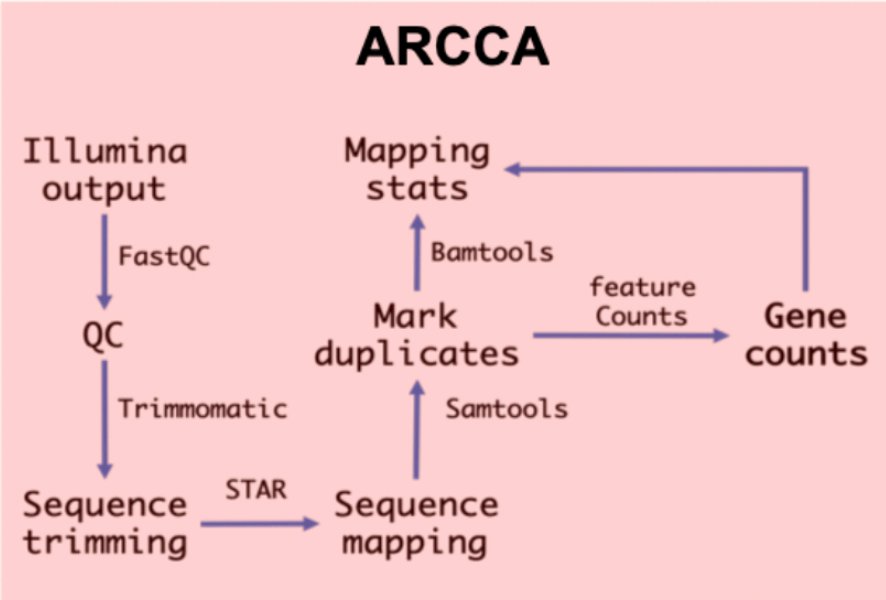


Figure 2.3 ARCCA workflow.
Summarised workflow followed in ARCCA in order to map the raw sequencing data to the human reference genome and obtain gene counts.

2.8 Differential gene expression, gene annotation and QC

2.8.1 Differential gene expression

Differential gene expression (DGE) between genotypes was done using R/Bioconductor package DESeq2 (Love, Huber and Anders, 2014). Due to the large number of datasets for processing, an automatic R script for DGE analysis was generated. Two files were required by this script, the first containing the gene counts and the second containing the metadata for each dataset. In this way, each 'mutant' versus control comparison was automatically performed in batch, reducing the amount of hands on work. To reduce the computational power required for the analysis, genes with less than 10 gene counts in at least one of the samples were removed from the analysis. The design formula applied in the generalised linear model regression to estimate the dispersions and the log2 fold changes of the model is ~ *phenotype* + *time*, where *phenotype* represents the genetic background of the samples (mutant or control) and the variable *time* represents the stage of differentiation of the samples. Using this formula, multiple pairwise comparisons of different samples and developmental stages can be analysed using the same script.

For this purpose, DGE was calculated using the DESeq method as descrived in (Love, Huber and Anders, 2014). The p values are calculated in a three-step process: 1) "Size factors" are estimated in order to normalise the gene counts for each sample (e.g. sequencing depth, library size, GC content…). 2) Estimation of the dispersion i.e. the difference between an observed gene count value and the mean value of all samples, calculating the log2 fold change between sample and controls. 3) Negative binomial generalised linear regression model testing the null hypothesis that the logarithmic fold change between control and samples equals zero. In order to avoid false positives caused by abnormal differences in gene counts between samples, outliers are filtered by calculating the Cook's distance. A gene with a Cook's distance above the calculated threshold by DESeq is flagged and the p value for this gene is not calculated, therefore discarded for downstream analysis.

2.8.2 Gene annotation and DGE QC

The outcome of the sequencing alignment returns only Ensembl IDs for each of the genes of the human genome. To produce a more user-friendly output, and to obtain the required data structure for downstream analysis, a biomaRt (Durinck *et al.*, 2005, 2009) query was introduced into the script. After the DGE analysis, the Ensembl identifiers were used to retrieve matching Entrez IDs as well as gene names, biotypes and description. When multiple IDs for the same gene were obtained from biomaRt, only the first was kept.

Graphical representation of the data is included at different stages of the DGE analysis. Boxplots of the unnormalized and normalised gene counts indicated an effective normalisation of the samples prior to and after DGE. Principal component analysis (PCA) plot and dendrograms of the samples were used to identify outliers. When samples were considered outliers, they were removed from the dataset and re-analysed.

2.9 BrainSpan correlation analysis

The BrainSpan human brain developmental transcriptome dataset was downloaded from http://www.brainspan.org. The dataset contains gene expression data from multiple developmental stages and brain regions. Each sample column was matched to their metadata file and samples missing metadata were removed from further analysis. To match the same gene expression values, gene counts obtained from RNA sequencing were transformed to fragments per million reads (FPKM). To do so, gene length for every transcript in the human genome were obtained from the GRCh38.84.gtf file using the R library rtracklayer (Lawrence, Gentleman and Carey, 2009). Finally, the unnormalized gene counts stored in a DESeqDataSet object were transformed to FPKM with the GenomicFeatures library (Lawrence *et al.*, 2013).

The comparative analysis was limited to embryonic stages; therefore post-natal samples were removed. In order to reduce the computational power required, genes with less than 1 FPKM in at least one sample were removed from the analysis. The R library Swamp (Lauss, 2018) was used to identify, and correct for batch effects in the BrainSpan and PSC generated datasets. To make the gene expression comparable between the two datasets, the values were transformed into Z-scores. Z-scores of different biological replicates for the same sample were combined using the formula:

$$Combined\ Zscore = \frac{\sum Zscores}{\sqrt{N}\ samples}.$$

Finally, the Spearman's correlation between *in vitro* and *in vivo* transcriptome datasets was calculated. Euclidean distances were calculated for each comparison between samples and plotted using the R package pheatmaps.

2.10 Pathway analysis

2.10.1 Pathway enrichment

All Entrez genes with a p adjusted < 0.1 (10% false discovery rate (FDR)) were used for pathway analysis. The genes were tested for enrichment against pathways in the Kyoto Encyclopaedia for Genes and Genomes (KEGG) (Kanehisa and Goto, 2000). Genes were ranked by decreasing fold change and tested using a Gene Set Enrichment Analysis (GSEA) test. The test determines the contribution of the ranked genes with an associated phenotype. Significant pathways (Storey q value < 0.1) were selected for further analysis.

Significant protein-coding Entrez genes (Bonferroni p adj < 0.05) were used for GO enrichment. GO terms were downloaded from NCBI (version 13th of September 2017). Significant genes were tested using a hypergeometric test. Non-significant protein-coding Entrez genes were used as background genes in the hypergeometric test for enrichment.

2.10.2 Signal clustering

Since many GO gene sets share a significant proportion of the genes, an iterative procedure was implemented in order to identify sets whose enrichment was explained by the same set of differentially expressed genes (DEGs). Briefly, A) genes in the smallest enriched gene set (Bonferroni p < 0.05) were removed from all other enriched terms, B) re-calculated DEG enrichment for the reduced terms, C) identified all the reduced terms, if any, that were no longer significant (uncorrected p > 0.05) and removed them from analysis, recording that the enrichment for this group was explained by the initial small gene set, and D) returning to step A using the next smallest significant gene set that had not been explained by another gene set before. This process was repeated until there were no enriched terms left (**Figure 2.4**).

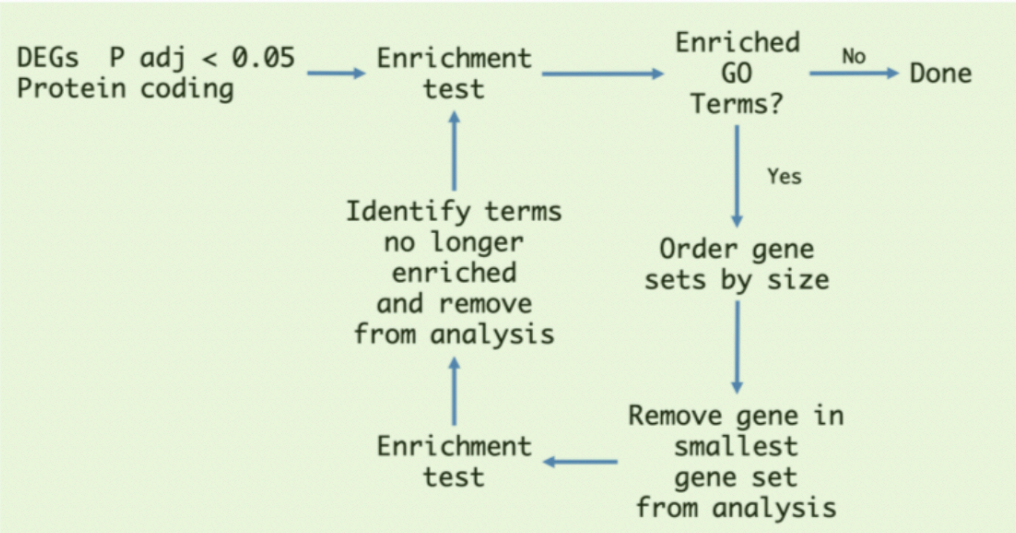


Figure 2.4 Signal clustering algorithm.
Iterative algorithm used to identify a common signal amongst groups of genes that share similar functions (functional gene sets) in order to reduce the number of gene ontologies resulting from previous steps.

2.11 Genetic association with neuropsychiatric disorders

GWAS summary statistics results were downloaded for SZ (Pardiñas *et al.*, 2018), autism spectrum disorder (https://www.med.unc.edu/pgc/), intelligence (Savage *et al.*, 2018) and Alzheimer's disease (Lambert *et al.*, 2013). SNPs with a minor allele frequency < 1% were removed. If available, only SNPs with an imputation > 0.8 were retained for analysis.

Using MAGMA, SNPs were mapped onto genes of the GRCh37 assembly. An annotation window of 5 and 10 Kb upstream and downstream respectively was defined around each gene for SNPs mapping outside the transcription region. Gene analysis was performed in order to calculate gene p-values for each trait. The European panel of the 1,000 Genomes phase 3 was used as raw genotype data in order to calculate the LD values between SNPs. If duplicated SNPs were present, only the first entry was kept. Synonymous SNP rs IDs was accounted for using the SNPdb file containing information of rs IDs that have been merged into a single ID. The final gene p-value was calculated using a combination of the principal component linear regression of the SNPs on a gene, mean SNP association and top SNP association p-values. An aggregated p-value for each gene was then generated from the three computed p-values.

To determine the association of the DEGs with each of the psychiatric traits, for each cell type DEGs in each analysed time-point, were divided into three groups: 1) all DEGs, 2) upregulated DEGs and 3) downregulated DEGs. To determine the association with neuropsychiatric traits, gene sets were tested for association with GWAS results. Functional gene sets were constructed using the previously described algorithm of those groups of genes with a significant association for a neuropsychiatric trait. The enrichment was recalculated for the new generated gene sets in MAGMA.

Chapter 3

CYFIP1 manipulation and 15q11.2 deletion lead to shared transcriptomic changes in stem cell-derived cortical neurons

3.1 Introduction

Psychiatric disorders can arise from a combination of multiple factors, including environmental and genetic components (Hyman, 2000; Uher and Zwicker, 2017). CNVs, including the ones affecting the 15q11.2 BP1-BP2 region, have been found responsible for increased liability for the development of neuropsychiatric disorders including SZ and ASD. Because CNVs span through large chromosomic regions affecting multiple genes, it can make the generation of animal models technically challenging. In this context, the use of iPSCs derived from patients carrying a particular mutation represent an advantage when studying the associated cellular phenotype. A relevant example is the work carried out by Yoon and colleagues, where the authors reported disrupted neural rosettes in NPCs derived from iPSCs carrying a 15q11.2 deletion (Yoon *et al.*, 2014).

However, due to the differences between individuals, it can be difficult to dissociate the effects of a mutation from the distinct genetic background of each individual. A solution to overcome this comes from the use of genetically modified hESCs and their isogenic parental line, which are subject to the same degree of genetic variability. Another advantage of using engineered hESCs models is investigating the implication of a specific gene within a CNV in the development of a phenotype. Therefore, the use of patient-derived iPSCs and genetically modified hESCs represent two complementary approaches in stem cell-based disease modelling. An ex-PhD student of our laboratory generated multiple hESCs lines with increased dosage and loss of CYFIP1, as well as patient-derived iPSCs harbouring a 15q11.2 BP1-BP2 deletion. As summarised in Chapter 1, the phenotype of these cells includes abnormal neural rosette structure and altered neural differentiation kinetics. Remarkably, we observed similar phenotype in cells with loss-of-CYFIP1 and those carrying the 15q.11.2 deletion, indicating that CYFIP1 is the major responsible gene for the observed phenotype. However, the cellular mechanism underlying these cellular pathologies remain unknown. Therefore, we performed RNAseq analysis to identify transcripts and gene pathways regulated by CYFIP1 and 15q11.2 deletion.

RNAseq uses deep-sequencing to analyse a population of RNAs (typically, but not restricted to, the mRNA fraction of the total RNA). The process of sequencing generates millions of short reads ranging from 25-400 bp with new technologies achieving lengths of several gigabases. This improvement in technology allows for the sequencing of whole mRNA molecules (Rank *et al.*, 2009; Sharon *et al.*, 2013; Leggett and Clark, 2017). The combination of read length and read type (single- or paired-end sequencing) enables the generation of a qualitative and quantitative measure of the gene expression, identify exon-exon junctions and novel coding regions, characterise isoform expression or

generate *de novo* assemblies of transcriptional maps (Wang, Gerstein and Snyder, 2009). Altogether, RNAseq allows unbiased quantification of the transcriptome, in order to identify changes in gene expression between different cells.

Because of the demonstrated effect of CYFIP1 in the structure of neural rosettes, a source of multiple signals required for NPC survival and neuronal differentiation (Banda 2015, Elkabetz 2008), it is likely that alterations in CYFIP1 expression have a repercussion in neural development. In support of this, *Cyfip1KD* in mouse embryonic development leads to abnormal localisation of the radial glia and neuronal cells (Yoon *et al.*, 2014). RNAseq of these cells can serve as a powerful unbiased approach to identify direct and indirect downstream targets of CYFIP1 whose changes at the transcriptional level lead to the defects observed during neuronal differentiation.

In this chapter, I use RNAseq to generate the transcriptome profile of multiple cells with altered levels of CYFIP1 at different stages of cortical glutamatergic differentiation. The gene expression of each cell type and timepoint analysed is compared to the BrainSpan dataset of the human developmental transcriptome in order to identify the real regional identity and age of the *in vitro* differentiation model. Finally, I address the similarities and differences at the gene expression level of the CYFIP1 mutant cells used in this thesis by comparing them against cells harbouring a deletion in the 15q11.2 region.

3.2 Experimental design and RNAseq QC

3.2.1 Cortical differentiation (with Dr Tamburini)

In order to elucidate the role of CYFIP1 during human cortical development, the CYFIP1tg, CYFIP1ko hESCs, as well as the patient-derived iPSCs were differentiated into cortical glutamatergic cells. Total RNA was extracted from three biological replicates for CYFIP1-manipulated cells and their respective controls at three stages of differentiation. The stages chosen for the study were: 1) neuroepithelium, around day 7-10, cultures at this stage contains around 80% of newly committed cortical progenitors as marked by the expression of forebrain marker FOXG1 and close to 100% of the pan neuroepithelial marker Nestin. 2) NPCs, day 15-20. This is the stage when neural rosettes were most prominent, as marked by NCAD staining, while intermediate progenitors expressing TBR2 cells were also present. 3) Neurons, days 30 and 45. At this timepoint, the cultures contain a significant number of $MAP2^+$ postmitotic neurons with $TBR1^+$ populations of >30% (**Figure 3.1**).

High RNA integrity and purity are a major requirement in transcriptomic experiments. After gDNA removal, the RIN of the samples was determined before sequencing. The RIN is an indicative value of the purity and integrity of the RNA. Samples with a lower RIN contain a more degraded RNA, which could cause the artificial selection of a small subset of transcripts (Imbeaud *et al.*, 2005). This method also allows to determine the presence of traces of gDNA in the sample. The assumption is that a low RIN could cause a poor outcome in the sequencing due to degraded RNA and the presence of contaminants, limiting the number of reads that will truly identify mRNA. Thus, the RIN is a key criterion to take into account when deciding which samples are used for sequencing. The analysis indicated a high quality of the RNA (average RIN 9.6 ± 0.3). Using the same analysis, the plotted electropherogram indicated that no gDNA was present in the samples. This was shown by a flat baseline consistent throughout all the samples analysed.

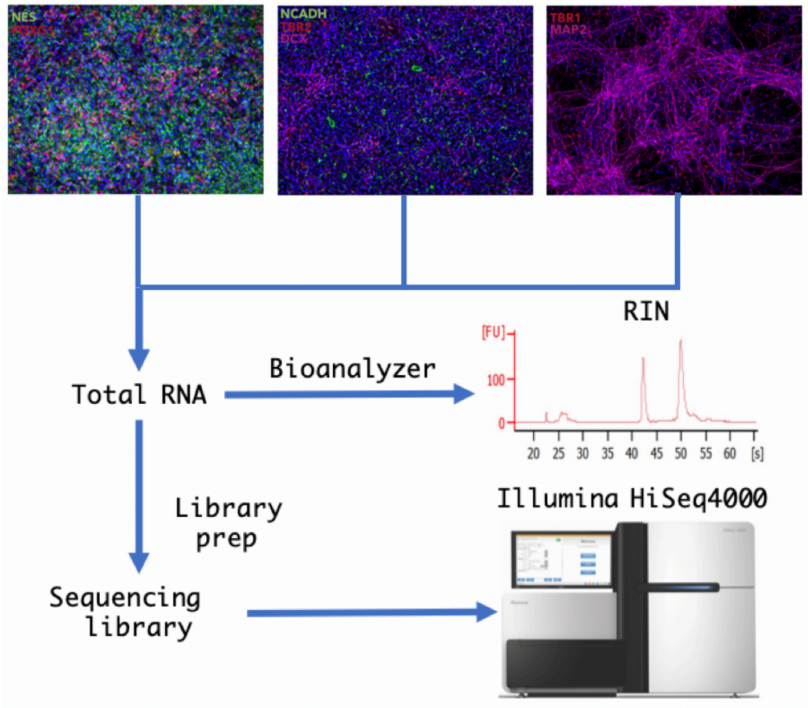


Figure 3.1. Workflow for cortical differentiation and sample preparation for RNAseq
hPSCs were differentiated to cortical glutamatergic neurons. During the process, total RNA was extracted at three well-stablished stages of differentiation. RNA was purified and the RIN was determined for each sample before proceeding to the library preparation. Paired-end sequencing was performed in an Illumina HiSeq 4000.

3.2.2 RNAseq analysis and differential gene expression

Sequencing of the samples generated a number of reads superior to the minimum required for differential gene expression (average library size 74 ± 11 million reads). FastQC was used to detect the presence sequences with poor base calling score, including repetitive sequences and those belonging to the library adapters. An abundant number of Illumina sequencing adapters were removed with trimmomatic. This process had a minimal impact on the number of reads for downstream analysis (70 ± 10 million

reads). Approximately 98% of the remaining sequences were successfully mapped to the human genome (98% ± 0.3% on target reads), with duplication rates below 25 % (average 18.8% ± 2.3% duplicated reads per sample).

Given the great number of comparisons to process, an R script based on a DESeq2 workflow for differential gene expression was generated (**Figure 3.2 A**). In order to obtain DEGs, the script pulled information from two files: 1) a file containing the raw gene counts, obtained from the mapping, for the samples to be analysed and 2) a metadata file with relevant information associated to each sample (e.g. stage of differentiation, phenotype, etc). The pipeline used to align sequencing reads to the human genome, obtaining only Ensembl IDs as gene identifiers. These are not informative, and manual search of each Ensembl gene identifier is not feasible in large-scale experiments where thousands of genes are interrogated. Therefore, a biomaRt query was included in the process in order to make the results of the DGE easier to interpret. Moreover, downstream analysis requires the use of other gene identifiers, i.e. Entrez IDs. For each studied gene, the biomaRt query pulled the associated gene names, Entrez IDs, gene biotypes and description.

To gain confidence in the DGE analysis, relevant visual QC analysis was introduced throughout the process. Appropriate normalisation needs to be performed to avoid expression biases due to library size. Given an adequate formula for dispersion, DESeq2 automatically approximates the size factors and corrects for these differences. Boxplots for unnormalized and normalised libraries are generated. Moreover, the gene dispersion estimates generated are shown for each comparison. These are important to visualise as they highlight any outliers that could alter the results of the DGE. A volcano plot is generated in order to highlight those genes with the largest changes in expression and are highly significant. Finally, using the top 500 DEGs, the samples are represented in a principal component analysis (PCA) plot to identify outliers between replicates easily. The output of the QC is simplified in **Figure 3.2 B-E**.

This streamlined process was proved to be an appropriate tool for the analysis of RNAseq data and obtain DEGs. The script works in a stand-alone way reducing the time of hands in work for the researcher and generates publication-ready figures of the analysis produced. This pipeline has been used by several members in the Li laboratory to analyse their RNAseq expression data.

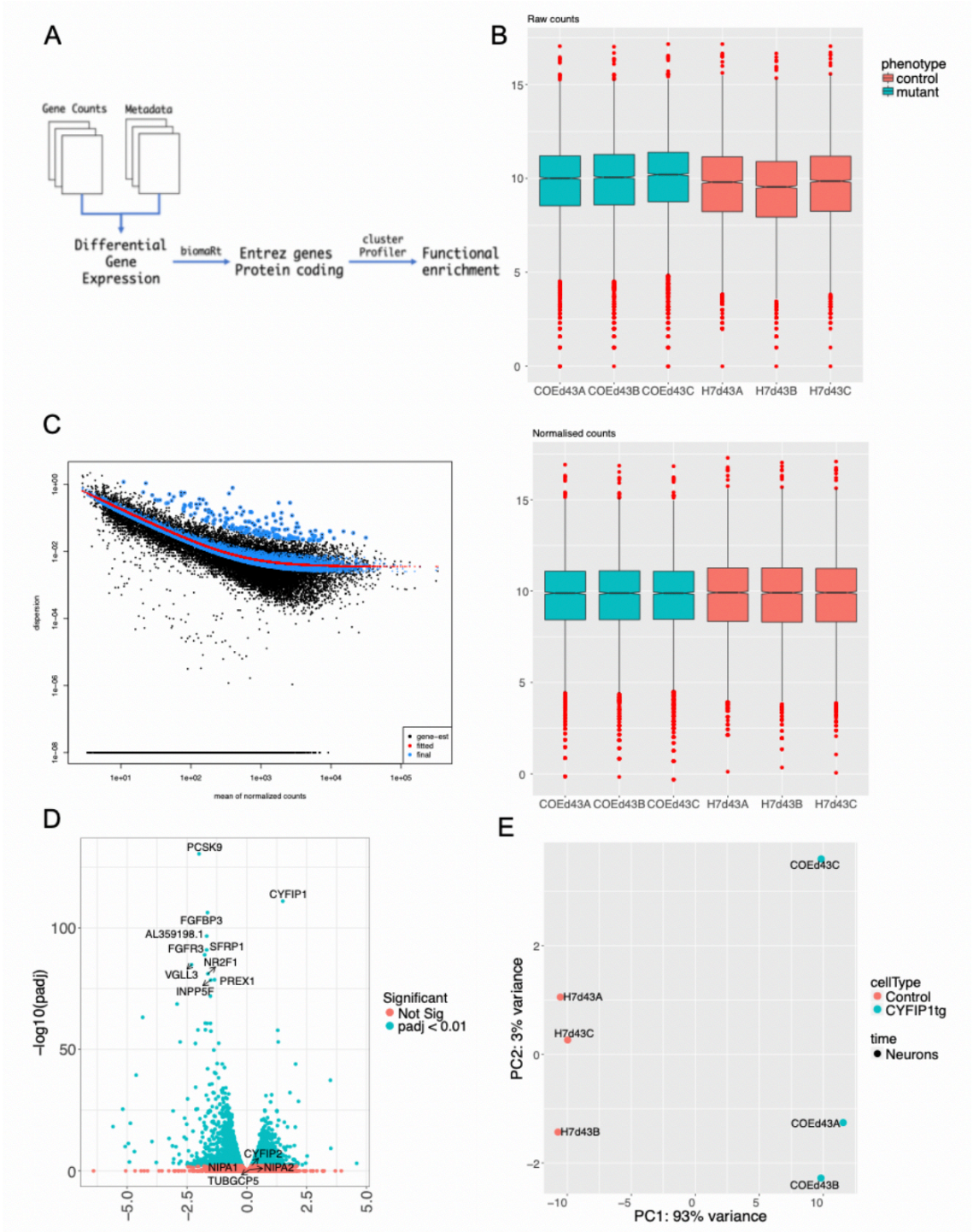


Figure 3.2. Established workflow for DGE and QC

(A) Workflow for DGE analysis in R using gene count data. (B) Boxplot of unnormalized and normalised library sizes in order to identify biases in the process of normalisation. (C) Gene expression estimates adjusted to fit the model. (D) Volcano plot highlighting genes with high differences in gene expression and significance. (E) PCA plot of the samples analysed used to identify outliers that could incorporate undesired biological variability.

3.3 Results

3.3.1 PSC-derived transcriptome mirrors foetal human cortical development

The transcriptomic data generated was produced from three different experiments. Due to its sensitivity, 'omics' datasets are subjected to batch effects that can introduce artefacts in the analysis (Yi *et al.*, 2018). These are non-biological variations introduced through different laboratory conditions, sources of samples or even small changes in the protocol (Leek *et al.*, 2010). This variation can make it difficult for comparing and interpreting datasets generated at different times. To evaluate the reproducibility of the differentiation model first, the R package Swamp was used to detect and correct for technical variability in the samples. After this, Spearman's correlation (Spearman's rho) was determined for all biological replicates used in this analysis (**Figure 3.3**). All the samples showed a high correlation for the same stage of differentiation (rho > 0.9) and two main clusters were identified: one containing the early, less differentiated neuroepithelial and NPCs samples, and a second formed by the neuronal samples. This indicates that our differentiation model creates similar expression profiles independently from the experimental origin, making it highly reproducible.

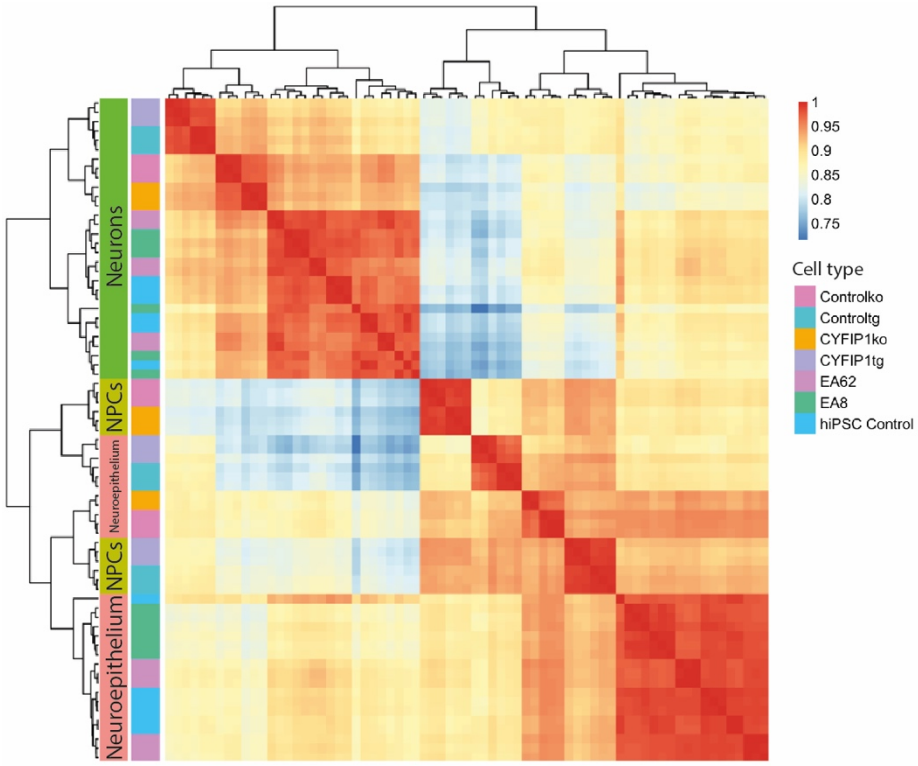


Figure 3.3. Correlation between the samples analysed.
For each biological replicate, Spearman's correlation was calculated using all DEGs in every timepoint analysed for CYFIP1 engineered and 15q11.2 deletion carrying cells. Unsupervised hierarchical clustering with Euclidean distances clustered the samples in two clearly defined groups: the first containing the neuronal stages and the second, with the neuroepithelial and NPC stages.

Reliability in cellular differentiation models will impact on how accurately they can recapitulate the tissue to be studied. The cell differentiation model used in our laboratory aims to generate dorsal telencephalic neurons. To determine how close the *in vitro* model used mimics the *in vivo* tissue, our PSC-derived gene expression data set was compared to that of the human embryonic brain obtained from the Allen Brain Atlas of the developing Transcriptome (BrainSpan, www.brainspan.org). This database contains the transcriptomic profile of multiple developing and adult human brain regions. The samples available range from 8 weeks, post-conception up to 60 years of age (Miller *et al.*, 2014). Rather than directly comparing the gene expression from both datasets, gene counts were transformed to Z-scores, and multiple replicates for the same structure and developmental stage were combined into a single one following the method described in Chapter 2.

Spearman correlation was calculated between the BrainSpan and PSC-derived dataset. The comparison revealed that all samples were highly correlated with embryonic stages of the BrainSpan dataset (rho > 0.57), with the highest correlation observed at weeks 8 and 9 post-conception (**Figure 3.4Figure 3.4 A**). The correlation decreased gradually when compared to the time-points closer to childbirth. In order to validate the regional identity of the *in vitro* generated neuronal cells, the transformed gene expression values were compared against the first two developmental stages in the BrainSpan, corresponding to the early mid-gestational period. The highest correlation was indeed observed with those structures in the telencephalon. In contrast, the correlation was lower with the thalamus and the upper rhombic lip, a structure that will develop into the cerebellum (**Figure 3.4 B**). For these comparisons, p value significance tests were not calculated.

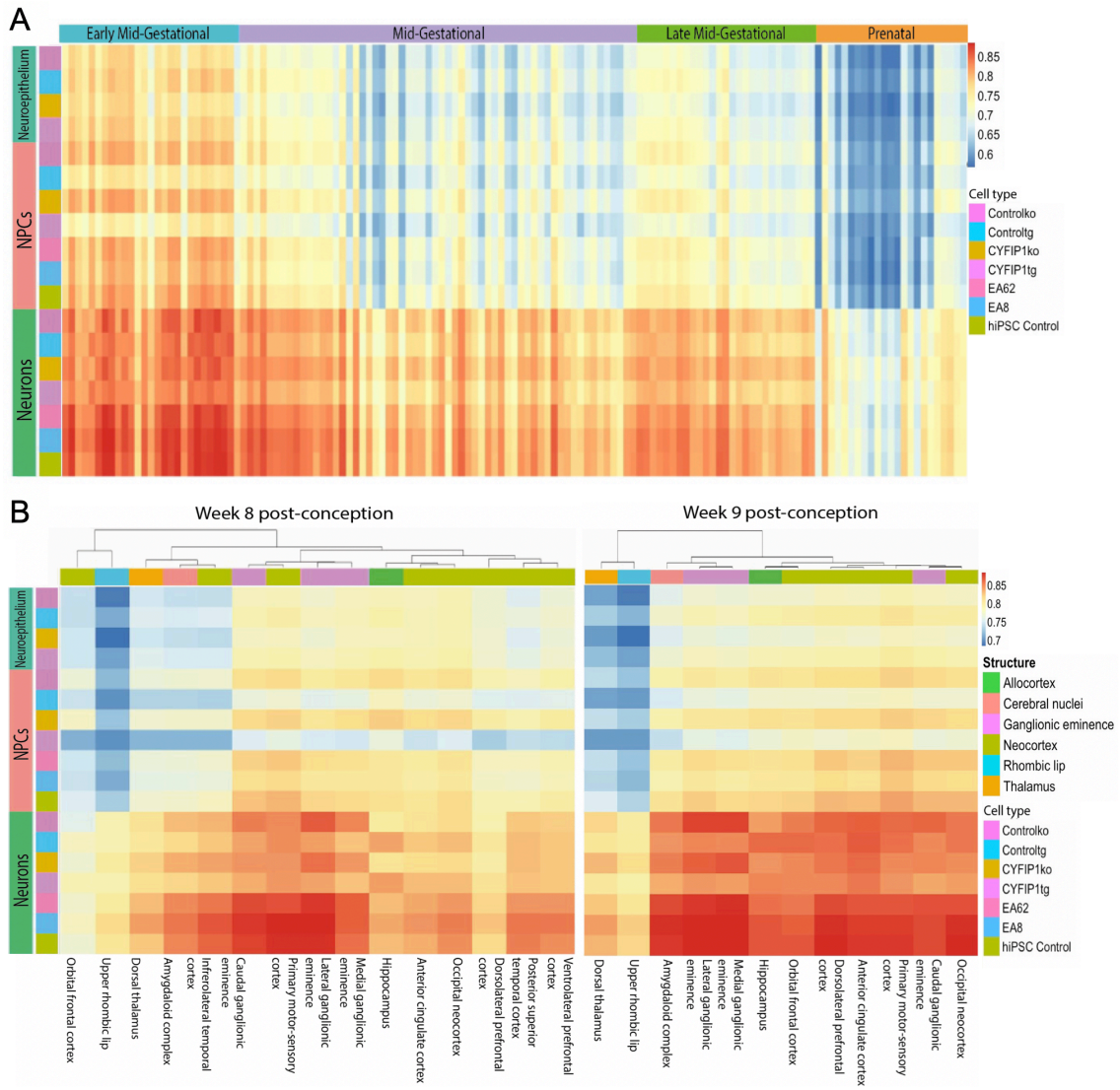


Figure 3.4. Recapitulation of the human embryonic developmental transcriptome in the in vitro model.

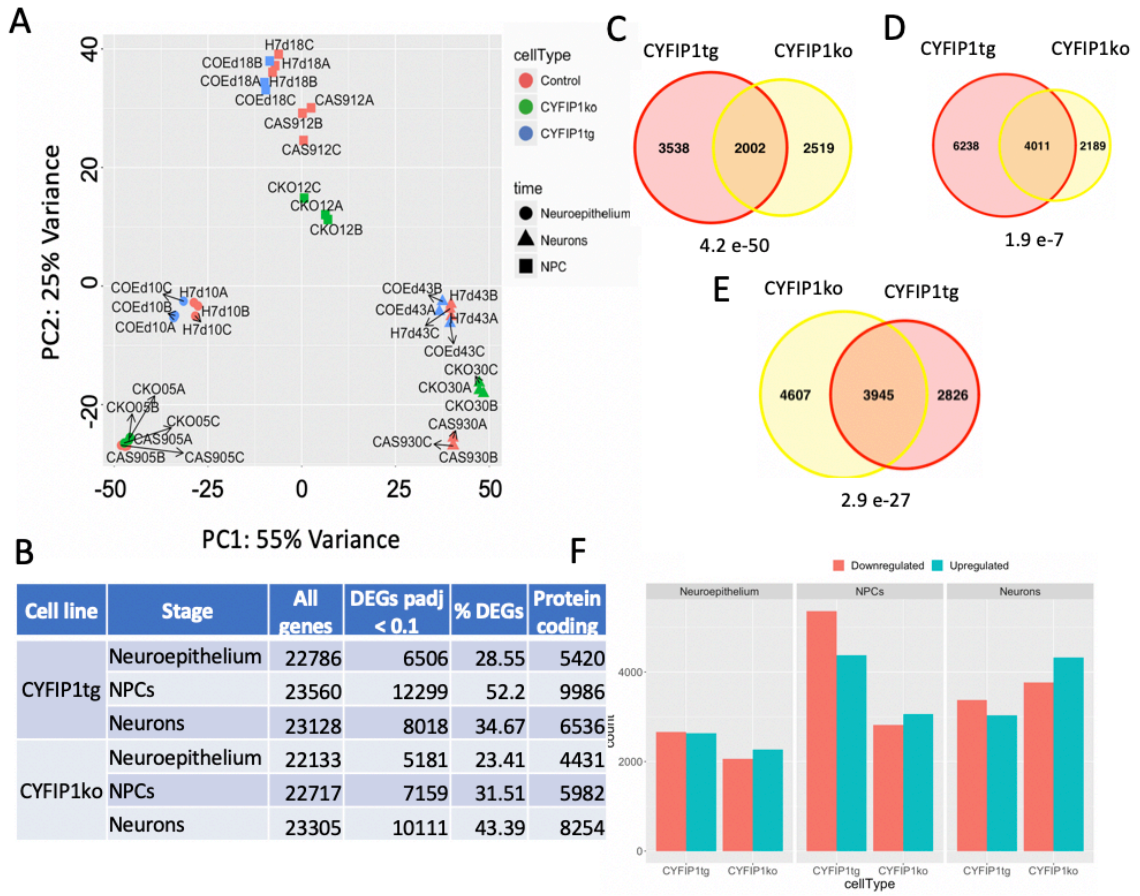
(A) Gene expression of PSC-derived datasets was compared against the BrainSpan embryonic stages. The highest correlation was observed in the first trimester of pregnancy corresponding to weeks 8 and 9 of embryonic development. (B) Spearman's correlation at weeks 8 (left) and 9 (right) post-conception. The greatest correlation corresponded to telencephalic samples embryonic samples belonging to the neocortex and ganglionic eminences, whereas the lowest was observed for the thalamus (midbrain) and upper rhombic lip (hindbrain).

3.3.2 Dysregulated genes related to CYFIP1 manipulation

(I) Gain and loss of CYFIP1 affects the expression of a large number of genes

PCA plot of the samples was generated using the top 500 dysregulated genes. Samples were found to be segregated by developmental stage and cell type (**Figure 3.5 A**). The distribution of the samples revealed three clearly defined clusters corresponding to the three developmental stages studied: neuroepithelial cells, NPCs and neurons. The first principal component (PC1) captured the largest variability in the samples, corresponding to the time-course neural differentiation. The second principal component (PC2) explained the differences between genotypes, segregating CYFIP1tg and CYFIP1ko, which clustered together with their parental control lines. With an FDR cut-off < 10% (Benjamini-Hochberg corrected p-value < 0.1), 4000 and 7700 DEGs were identified between CYFIP1tg and control and CYFIP1ko and control, respectively, which represents approximately 45% or less of the total number of DEGs in each dataset (**Figure 3.5 B**).

CYFIP1tg and CYFIP1ko shared a large number of DEGs in each time points (**Figure 3.5 C-E**). The most significant overlap was found at the neuroepithelial stage (p value calculated through hypergeometric test using all significant protein coding genes. The total population was defined as all protein coding genes detected in a particular developmental stage). Amongst the shared DEGs at each stage, the proportion of downregulated genes and upregulated appeared to be similar, although CYFIP1tg cells yielded a higher number of downregulated genes (**Figure 3.5 D**). These results indicate that alterations in CYFIP1 expression have a repercussion over a large number of genes. For this analysis, results using a 5% FDR threshold chut-off can be found in Apendix I (**Supplementary Figure 1**).



Cell line	Stage	All genes	DEGs padj < 0.1	% DEGs	Protein coding
CYFIP1tg	Neuroepithelium	22786	6506	28.55	5420
	NPCs	23560	12299	52.2	9986
	Neurons	23128	8018	34.67	6536
CYFIP1ko	Neuroepithelium	22133	5181	23.41	4431
	NPCs	22717	7159	31.51	5982
	Neurons	23305	10111	43.39	8254

Figure 3.5 Differential gene expression in CYFIP1tg and CYFIP1ko cells.
(A) PCA plot of the CYFIP1tg and CYFIP1ko and their isogenic control parental lines. Samples clustered by stage of differentiation and cell line. (B) Number of expressed genes and significant DEGs at 10% FDR after DESeq2 analysis comparing each CYFIP1tg and CYFIP1ko time-points with their respective isogenic control parental line. The percentage of DEGs indicates the proportion of differentially expressed genes over the total amount of transcripts sequenced for that sample. Overlapping DEGs between CYFIP1tg and CYFIP1ko for neuroepithelium (C), NPCs (D) and neurons (E). Neuroepithelial cells and NPCs shared a significant overlap of DEGs (hypergeometric p-value overall genes tested for that developmental stage). (F) The directionality of the fold change of the genes commonly expressed between CYFIP1tg and CYFIP1ko.

As part of the library preparation, the samples were enriched for mRNAs through poly-A tail affinity. This is important to deplete the ribosomal RNA (rRNA) in the samples. The analysis of the gene biotypes revealed that over 80% of the DEGs across the three time-points were indeed protein-coding genes (**Figure 3.6 A**). The remaining types of characterised RNAs belonged primarily to long non-coding RNAs (lncRNAs) and pseudogenes. lncRNAs is a group formed by RNAs of more than 200 nucleotides long which are not translated into proteins.

More than 10% of the DEGs were lncRNAs, which can be poly-A tailed. More than 10,000 lncRNAs have been characterised in humans, and they show tissue-specific expression pattern, with the brain being the organ where most of them are expressed (Derrien *et al.*, 2012). At different developmental stages, lncRNAs showed big changes

in FC with highly significant p-values (**Figure 3.6 B, C**), which indicates their relevance in the normal process of cortical development. Multiple lncRNAs with known roles in brain development were found to be altered in cells with altered CYFIP1 expression. The myocardial infarction associated transcript (MIAT) appeared to be downregulated in all CYFIP1ko time-points analysed. MIAT has been associated through GWAS studies with ocular movement disorders in SZ, and it is downregulated in schizophrenic brains (Takahashi *et al.*, 2003; Barry *et al.*, 2014). MIAT is also involved in the splicing of the known SZ-associated genes DISC1 and ERBB4 (Barry *et al.*, 2014). Other differentially expressed lncRNAs included HOTAIR, which is known to promote the recruitment of Polycomb repressive complex 2 (PRC2), causing transcriptional repression (Tsai *et al.*, 2010). NEAT1 and MALAT1 regulate synaptogenesis by modulating specific mRNA splicing events (Clemson *et al.*, 2009; Bernard *et al.*, 2010) and PAUPAR which re-presses the expression of multiple genes, including PAX6 (Vance *et al.*, 2014).

In summary, both increase or decrease of CYFIP1 expression have a big impact on gene expression of cortical neural cells, causing the dysregulation of thousands of transcripts. Interestingly, the changes were not limited to protein-coding genes, but also to lncRNAs which exercise a role in regulating the expression of the protein-coding transcripts.

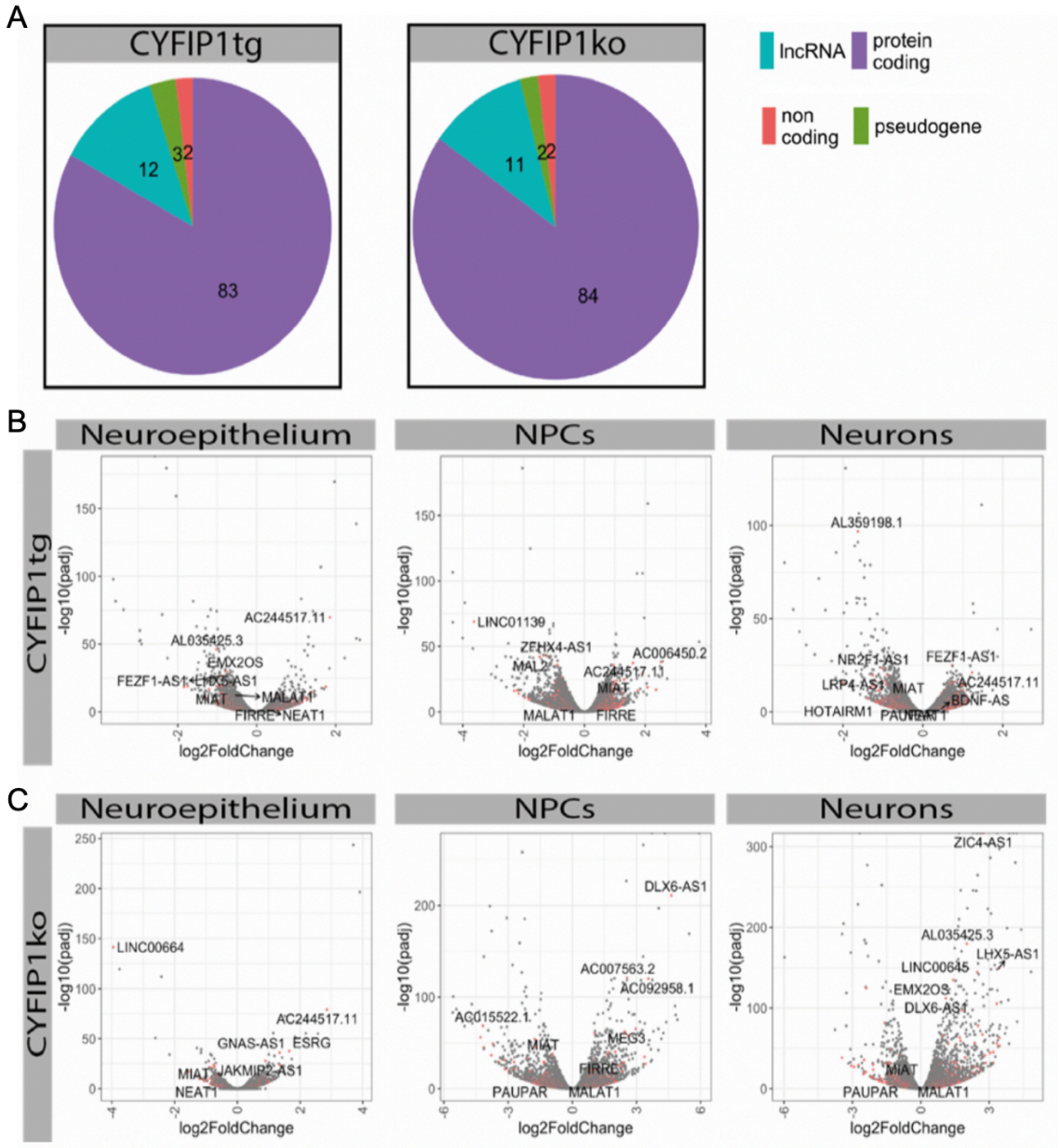


Figure 3.6. Characterisation of the DEGs in CYFIP1 cell lines.
(A) Average percentage for each cell line of the different gene biotypes identified over the total number of DEGs. Volcano plots for showing the dispersion of the adjusted p-values by FC for significant lncRNAs (in red) expressed at each developmental stage in CYFIP1tg (B) and CYFIP1ko cells (C). The plots show how lncRNAs are more affected in CYFIP1ko cells at all stages with greater p-values and FC.

(II) Altered CYFIP1 dosage affects the expression of NIPA1 and cortical regulators but not CYFIP2 or FMR1

I then inspected in detail the expression of *CYFIP1* itself and genes close to *CYFIP1* in the control and CYFIP1 mutant samples. As shown in **Figure 3.7**, *CYFIP1* levels were higher at neuroepithelial and NPC stages than in neurons, consistent with the newly identified role for CYFIP1 in modulating NPC maintenance and terminal differentiation. *CYFIP1* is one of the four genes in the 15q11.2 BP1-BP2 region, which also contains

TUBGCP5, *NIPA1* and *NIPA2*. Intriguingly, *NIPA1* was downregulated in both CYFIP1tg and CYFIP1ko NPCs. However, no significant changes were observed in *NIPA2* and *TUBGCP5* consequent of CYFIP1 manipulation.

CYFIP2 is another member of the CYFIP family. CYFIP2 is highly conserved with CYFIP1 in amino acidic sequence, structure and function, and like CYFIP1, is also expressed in the developing cortex. Both CYFIP1 and CYFIP2 interact with FMRP and form a transcriptional regulatory complex of specific mRNAs (a. Schenck *et al.*, 2001). Interestingly, CYFIP2 showed an opposite trend of temporal expression pattern to that of CYFIP1, with the highest expression level observed in the neuronal stage in cultures of all genotypes. Moreover, NPCs and neurons in CYFIP1tg cells had a significantly higher expression of CYFIP2 than the control. On the other hand, *FMR1*, the coding gene for FMRP, expression was maintained at a similar level across all stages of differentiation in both the control and CYFIP1 mutant cells.

The process of cortical development involves a tight regulation of multiple genes in a timely fashion. Defects in the regulation of these genes can lead to the development of functional and structural brain abnormalities. Graphical representation of some well-known stage-specific markers in cortical neural development were observed to be changed at the transcriptomic level in CYFIP1tg and CYFIP1ko cells (**Figure 3.7 B**) indicating fundamental changes in genes involved in the process of neural differentiation. Both CYFIP1 overexpression and CYFIP1-deficiency caused changes in cadherin 1 (*CDH1*), *HES1*, *Nestin* and *FOXG1*. These genes are required for the correct timing of neurogenesis and patterning of the brain (Ohtsuka *et al.*, 2001; Sunabori *et al.*, 2008; Delgado-Esteban *et al.*, 2013; Kumamoto and Hanashima, 2017; Bernal and Arranz, 2018). Moreover, alterations of FOXG1 are the cause of a class of ASD denominated FOXG1 syndrome (Ariani *et al.*, 2008). In CYFIP1tg NPCs, *FOXG1* showed a dramatic 10-fold increase. In line with our findings that CYFIP1 overexpressing cells maintain a proliferative phenotype whereas CYFIP1-deficiency led to premature differentiation, CYFIP1ko NPCs had a lower expression of *PAX6* (4.8-fold reduction) as compared to the parental cells, whereas DCX and β-III-tubulin (*TUBB3*), had increased expression compared to the control. Conversely, *DCX* and *TUBB3* were reduced in CYFIP1tg cells compared to their isogenic control. Despite the differences in expression levels of the genes shown in **Figure 3.7 B**, the pattern of expression follows the same trend in both control and CYFIP1 manipulated cells.

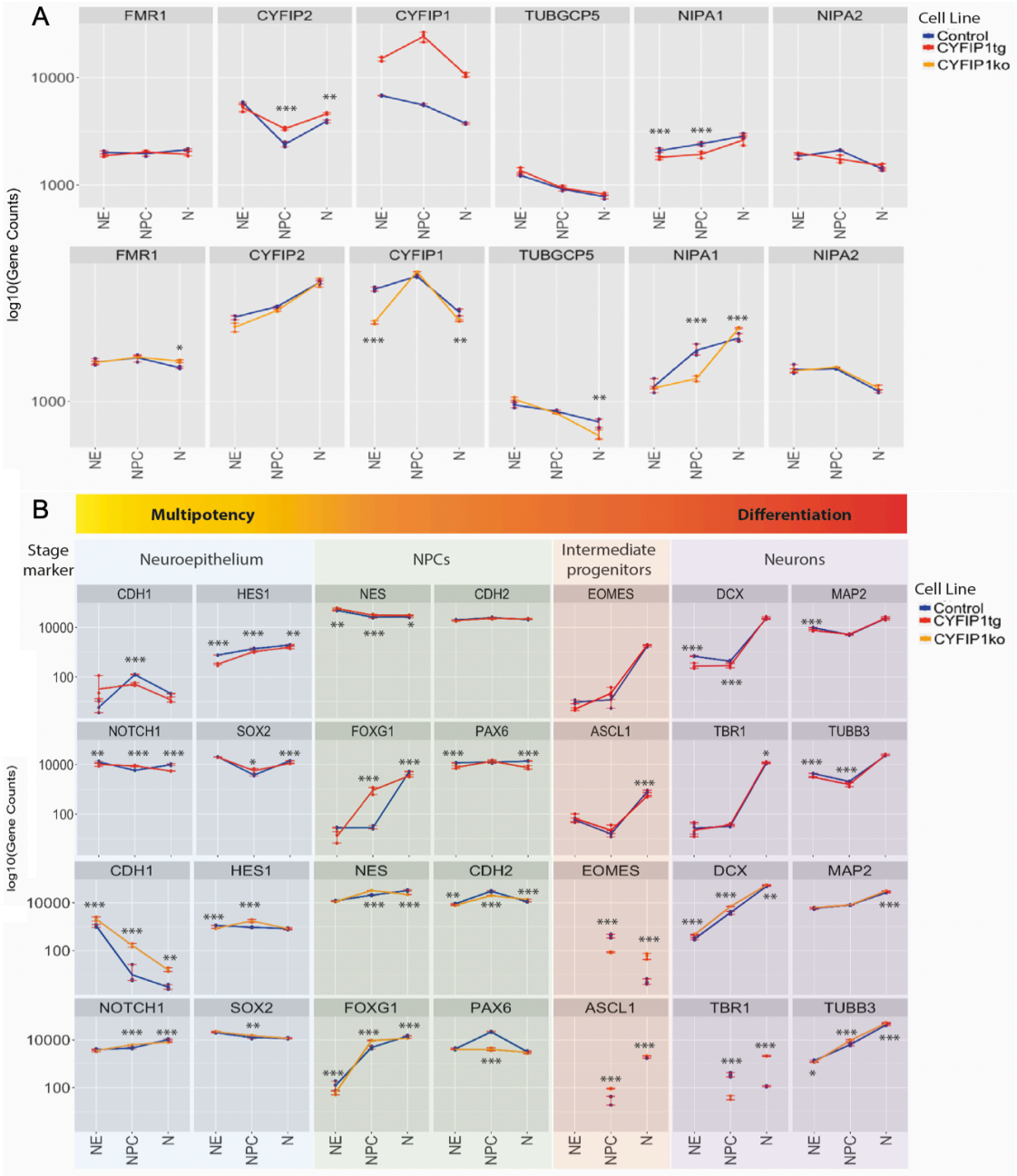


Figure 3.7. Time-course transcriptional changes in CYFIP1tg and CYFIPko during the process of cortical differentiation.

(A) Gene expression of genes in the 15q11.2 loci, CYFIP2 and FMR1 in CYFIP1tg; CYFIP1ko and control parental lines. (B) Genetic expression of key differentiation markers in the three time-points analysed. Missing data points for EOMES, ASCL1 and TBR1 are due to genes did not reach the minimum threshold of expression of at least 10 gene counts in one of the biological replicates.

(III) CYFIP1-regulated genes are enriched for FMRP targets and developmental delay disorder associated genes

CYFIP1 forms a complex with FMRP and eIF4E to regulate protein translation. In this context, FMRP recruits CYFIP1 to the 5' end of the capped target mRNAs which causes the repression of the ribosomal translation. Mutations affecting FMRP are the cause of FXS (Sitzmann *et al.*, 2018). This disorder is the most common inherited genetic condition causing a range of developmental delays, intellectual disabilities and other types of psychiatric conditions (La Fata *et al.*, 2014). I found that more than half of the FMRP-regulated genes were differentially expressed in CYFIP1 engineered cells. To determine the enrichment for FMRP target genes in CYFIP1 gain- and loss-of-function cells, the frequency of FMRP targets in CYFIP1 protein-coding DEGs was compared to that of FMRP targets in all protein-coding genes. This comparison revealed that FMRP regulated DEGs were significantly enriched in all three differentiation stages (**Table 3.1**).

Developmental delays are a classification that covers multiple disorders which include, amongst others; language and learning disorders, ADHD, ASD, and in broader definitions, SZ. A UK-wide survey of over 4,000 families identified 93 genes with a high genome-wide significance for developmental disorders (namely DDD study) (McRae *et al.*, 2017). The changes observed in these genes amongst the DEGs were highly similar to the ones seen in the FMRP-regulated genes. Fisher exact test of the DDD genes revealed that all developmental stages analysed were significantly enriched for genes causing developmental delay disorders (**Table 3.2**).

Both statistical analysis using a 5% p adj threshold for DEG significance can be found in Apendix I (**Supplementary Table 1 - 2**).

Table 3.1 Summary of the statistical enrichment for FMRP genes amongs DEGs

Cell line	Stage	FMRP targeted	No FMRP targeted	DEG	Fisher's pval
CYFIP1tg	Neuroepithelium	419	4875	p < 0.1	1.84E-18
		420	9138	p > 0.1	
	NPCs	605	9129	p < 0.1	4.08E-08
		234	5351	p > 0.1	
	Neurons	446	5959	p < 0.1	1.12E-09
		393	8058	p > 0.1	
CYFIP1ko	Neuroepithelium	325	4002	p < 0.1	4.93E-07
		514	9132	p > 0.1	
	NPCs	424	5451	p < 0.1	3.57E-12
		415	8726	p > 0.1	
	Neurons	558	7534	p < 0.1	2.65E-14
		281	6609	p > 0.1	

Fisher exact test p-values for enrichment for FMRP genes in CYFIP1tg and CYFIP1ko cells. All time-points tested were significant for both cell lines. DEG indicates the adjusted p value threshold for the DEG cut-off.

Table 3.2 Summary of the statistical enrichment for DDD genes amongst DEGs

Cell line	Stage	DDD gene	No DDD gene	DEG	Fisher's pval
CYFIP1tg	Neuroepithelium	48	5246	p < 0.1	0.00112674
		45	9513	p > 0.1	
	NPCs	69	9665	p < 0.1	0.01916738
		24	5561	p > 0.1	
	Neurons	59	6346	p < 0.1	5.9718E-05
		34	8417	p > 0.1	
CYFIP1ko	Neuroepithelium	38	4289	p < 0.1	2.71E-02
		55	9591	p > 0.1	
	NPCs	48	5827	p < 0.1	9.54E-03
		45	9096	p > 0.1	
	Neurons	67	8025	p < 0.1	2.69E-04
		26	6864	p > 0.1	

Fisher exact test p-values for enrichment for DDD genes in CYFIP1tg and CYFIP1ko cells. All time-points tested were significant for both cell lines. DEG indicates the adjusted p value threshold for the DEG cut-off.

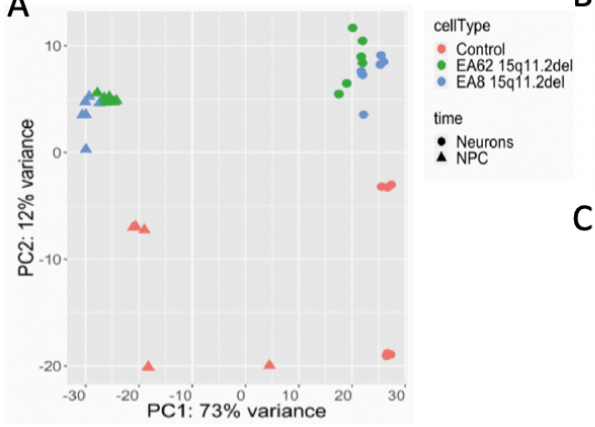
3.3.3 Transcriptomic changes in 15q11.2 deletion neural cells resemble those of CYFIP1 engineered cells

RNAseq was performed on stage-matched neural cell samples obtained from two clones of each iPSC lines of two 15q11.2 deletion carriers (namely EA8 and EA62) and control iPSC lines derived from two subjects without 15q11.2 CNV. PCA plot of iPSC samples resembled the distribution observed from the CYFIP1 engineered cells (**Figure 3.8 A**). PC1 captured the largest variance segregating the samples into two different groups:

NPCs and neurons. PC2 revealed the intrinsic variability of each individual from which each cell line was derived. Intriguingly, the differences between patient-derived lines were smaller than the two control lines.

The analysis revealed that between 5600 and 8850 genes were differentially expressed in the 15q11.2 deletion neural cells (BH padj < 0.1, **Figure 3.8 B**). For both patient cell lines, 75% of the DEGs were protein-coding genes. Similar to the CYFIP1 engineered cells, the next largest group of DEGs was the lncRNAs, representing 15% of the total DEGs (**Figure 3.8 C**). However, DEGs from only one of the developmental stages for each iPSC line was enriched for FMRP target genes. The EA8 cell line showed significant enrichment at the NPC stage (Fisher's exact p < 0.05), whereas the EA62 showed enrichment at a later stage in neuronal cells (Fisher's exact p < 0.05) (**Figure 3.8 D**). In this case, none of the differentiation stages analysed was enriched for genes involved in developmental delay disorders (Fisher's exact p > 0.05) (**Figure 3.8 D**). This analysis can be found in Apendix I using a 5% FDR threshold (**Supplementary Figure 2**).

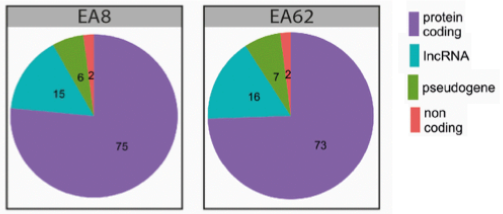
A



B

Cell line	Stage	All genes	DEGs padj < 0.1	% DEGs	Protein coding
EA8	NPCs	24222	5602	23.13	4205
	Neurons	24855	6289	25.3	4743
EA62	NPCs	24650	5586	22.66	3962
	Neurons	25314	8846	34.95	6579

C



D

Cell line	Stage	FMRP targeted	No FMRP targeted	DEG threshold	Fisher's pval
EA8	NPCs	251	3863	p < 0.1	0.04700282
		588	10346	p > 0.1	
	Neurons	222	4385	p < 0.1	0.9873096
		617	10234	p > 0.1	
EA62	NPCs	192	3676	p < 0.1	0.97102037
		647	10608	p > 0.1	
	Neurons	375	6030	p < 0.1	0.01951801
		464	8671	p > 0.1	

E

Cell line	Stage	DDD gene	No DDD gene	DEG threshold	Fisher's pval
EA8	NPCs	16	4098	p < 0.1	0.99213095
		77	10857	p > 0.1	
	Neurons	28	4579	p < 0.1	0.51361574
		65	10786	p > 0.1	
EA62	NPCs	21	3847	p < 0.1	0.78109136
		72	11183	p > 0.1	
	Neurons	46	6359	p < 0.1	0.06567867
		47	9088	p > 0.1	

Figure 3.8 Differential gene expression results for EA8 and EA62 lines.
(A) PCA plot for EA8 and EA62 and control iPSC lines. Samples clustered by developmental stage and cell line. (B) Table summarising the number of transcripts identified for each cell line and developmental stage, and the number of DEGs at 10% FDR (BH p adj < 0.1). Percentage of DEGs expressed over the total number of transcripts characterised. (C) For each cell line, the average percentage of each gene biotype identified amongst significant DEGs in each time-point. Protein coding genes represent 75% of the total DEGs in patient-derived iPSCs. (D) Table summarising the enrichment results tests for FMRP and DDD associated genes amongst the DEGs in cells carrying a deletion in the 15q11.2 region. Only one time-point for each cell line was found to be significantly enriched for FMRP target genes amongst the DEGs (Fisher exact p < 0.05), whereas none of the sets tested were enriched for DDD genes.

Whether 15q11.2 BP1-BP2 deletion affects the transcription level of the four genes within this CNV during *in vitro* cortical development had not been previously reported. To answer this question, the normalised counts of the deleted genes were plotted. All four genes showed a significant decrease in expression compared to the control cells. In all cases, the reduction was approximately 2 FC compared to the control iPSCs (**Figure 3.9 A**). Moreover, similar to that observed in the CYFIP1 engineered cells, the level of *FMR1* were comparable between control and patient-derived cells, with no significant difference between samples or time-points analysed (**Figure 3.9 A**). The expression pattern of *CYFIP2* was also similar to that observed in hESC-derived neural cultures with a higher level of transcript detected at the neuronal stage than NPC stage, a temporal trend that is opposite to that of *CYFIP1*.

Multiple cell fate determinants were found to be altered in patient-derived iPSCs throughout differentiation. Amongst the genes that play an important role in early neural development and/or neural patterning; *CDH1*, *Nestin* and *FOXG1* showed the biggest changes between patient and control cells (**Figure 3.9 B**). The changes in *CDH1* were more accentuated at the neuronal stage, where patient-derived cells showed a marked reduction in its expression. Both patient cell lines showed significant changes in transcripts expressed in postmitotic neurons, such as *DCX*, *MAP2* and *TUBB3*, at the NPC stage.

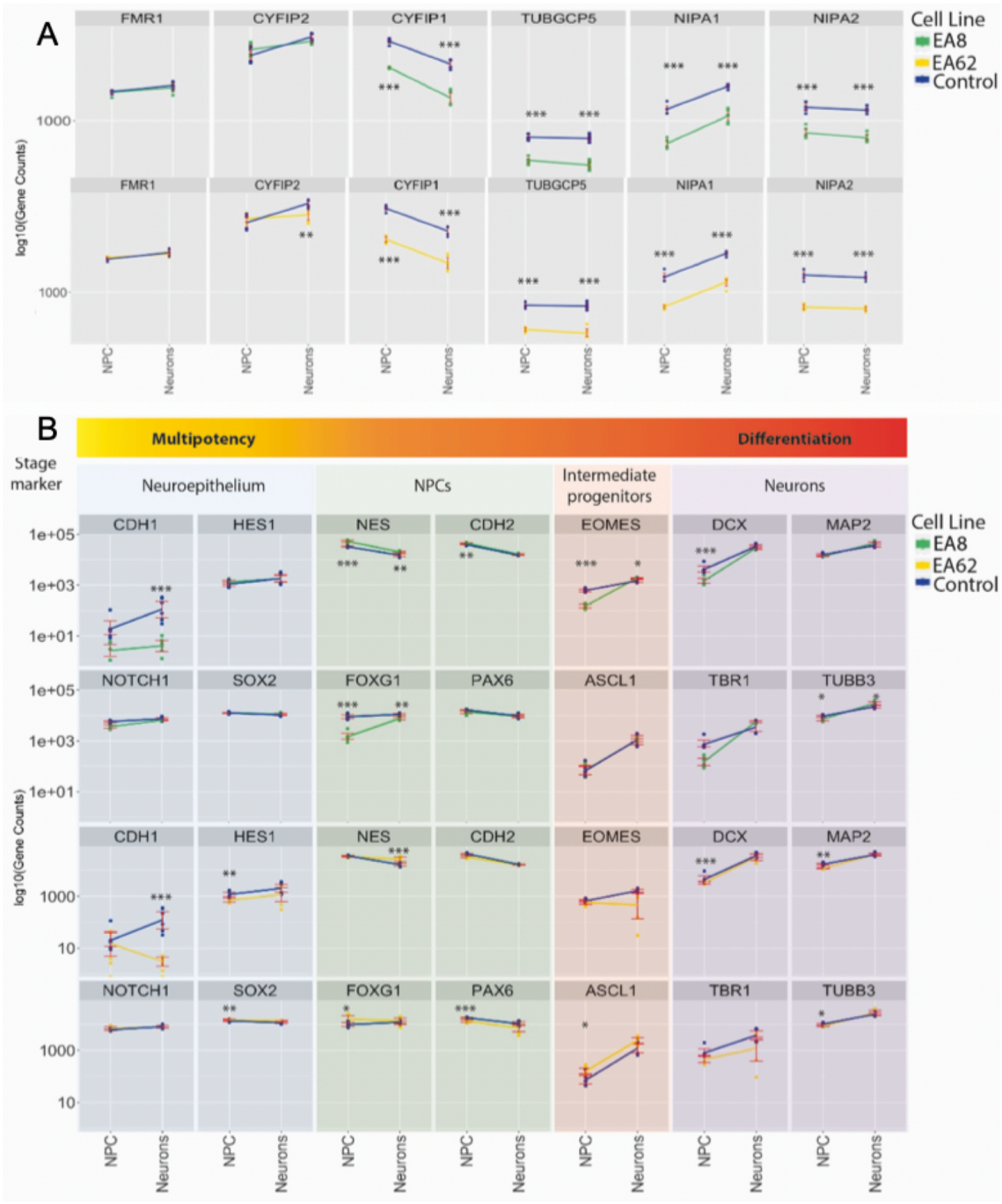


Figure 3.9 Time-course of transcriptional changes in EA8 and EA62 during the process of cortical differentiation.

(A) Plot of the normalised counts for the genes in the 15q11.2 BP1-BP2 region, CYFIP2 and FMRP1 (FMR1) genes in EA8, EA62 and control iPSCs during differentiation. RNAseq detected the effects of the deletion showing 2 FC reduction of all genes in the 15q11.2 BP1-BP2. CYFIP2 and FMRP1 were not altered. (B) Transcriptomic expression profile over time of key differentiation markers in the time-points analysed. Patient-derived cells showed similar changes in genes like CDH1, NES, DCX and TUBB3.

3.4 Discussion

This chapter describes the transcriptome analysis of PSC-derived cortical cells with altered levels of CYFIP1. DGE analysis showed that changes in CYFIP1 expression or 15q11.2 deletion have a repercussion over a large number of genes, but the cells were able to differentiate into cortical fate as their isogenic controls. A detailed analysis of the genes indicated that multiple cell fate determinant genes are affected when CYFIP1 expression is either increased or decreased in a cortical glutamatergic differentiation model.

The correction of the PSC-derived RNAseq dataset for the RIN allowed removing the variation introduced by the sample handling, without accounting for the different intrinsic expression for every sample. This confirmed not only the robustness of our experiment but also our accuracy at determining the similarity of the time-points analysed from cell lines that differentiate at different speeds.

A study carried out by Leemput and colleagues generated a complete database of *in vitro* human gene expression throughout cortical differentiation. In this study, the authors used the BrainSpan dataset for human brain developmental transcriptome to validate the model utilised was correlated to the human foetal cortex (van de Leemput *et al.*, 2014). In a similar way, the data presented in this chapter demonstrates that the PSC-derived dataset is highly correlated with the human foetal brain at weeks 8 and 9 post conception. Most importantly, this analysis reveals that the cultured cells have telencephalic characteristics and give raise to the neocortex. Samples obtained from iPSCs were more similar to ganglionic eminences, indicating a more ventral identity of the cells. This difference can be explained by the intrinsic variability in the gene expression of each cell line. It is known that cell lines generated from different individuals have a preference for generating cellular fates that are more similar to ventro-caudal regions, and, in some cases, cerebellum (Quadrato *et al.*, 2017; Pollen *et al.*, 2019).

RNAseq of CYFIP1tg and CYFIP1ko cells during differentiation revealed a significant overlap of altered genes in the early developmental stages. This could be an indication of shared mechanisms contributing to the phenotype observed in both cell lines. Despite the differences in the gene expression, the largest changes observed between developmental stages demonstrating that the cells are still able to undergo cortical differentiation. The largest proportion of DEGs were protein-coding genes, none the less, the effects of altered levels of CYFIP1 were not restricted to these genes. lncRNAs were the second most affected group of genes. Little is known about lncRNAs since they used to be ignored due to the fact that they do not contain conserved ORFs and were thought not to have any function. In recent years, research has revealed that lncRNAs play a critical

role in brain development and are necessary for cell-type specification (reviewed by (Clark and Blackshaw, 2014)). lncRNAs involved in transcriptional regulation and mRNA splicing of SZ-associated genes were found to be differentially expressed in both CYFIP1 engineered cells. This indicates that the effects of altered CYFIP1 expression are not only restricted to protein-coding genes.

The previous characterisation of CYFIP1tg and CYFIP1ko cell lines indicated changes in the expression of cortical developmental markers. The findings in this chapter validate the previous characterisation of these cell lines. At early stages; *CDH1*, *HES1* and *PAX6* were found amongst the top dysregulated genes and the pan-neuronal markers *DCX* and *TUBB3* at the mature stages. Time-course analysis showed a peak of expression of *CYFIP1* at the NPC stage, indicating the importance of the gene in this particular developmental stage. At the transcriptional level, *FMR1* was not affected by changes in CYFIP1 and remained constant across differentiation. On the other hand, *CYFIP2* showed a very characteristic expression pattern, with opposite changes to the ones observed in *CYFIP1*. The regulation of the *CYFIP2* mRNA is poorly understood, and previous studies differ in the results observed. *TUBGCP5* and *NIPA2* were not affected by changes in CYFIP1 expression; however, *NIPA1* mRNA was found to be downregulated in both CYFIP1tg and CYFIP1ko cells.

Finally, cells carrying a 15q11.2 deletion showed similar defects to the CYFIP1 engineered cells. Firstly, DEGs showed an enrichment amongst FMRP target genes, and the same effect over lncRNAs was observed. Moreover, the 15q11.2 deletion caused a 2-FC decrease of all the genes in the region. These results are in agreement with had been previously described by Das *et al* in undifferentiated cells carrying a similar CNV (Das *et al.*, 2015). Similarly to what was observed in CYFIP1 engineered cells, *FMR1* mRNA was not changed, and *CYFIP2* showed the opposite expression pattern to *CYFIP1*. The enrichment for FMRP target genes was observed in deletion carrying cells, although only in two of the stages analysed. Most importantly, pathway enrichment analysis of the DEGs in patient-derived cells showed that pathways affected in gain- and loss-of-function of CYFIP1 were conserved in these cells.

In conclusion, this chapter describes the transcriptional changes caused by altered levels of CYFIP1 identifying genes differentially expressed and affected pathways and highlight the importance of CYFIP1 during brain development. These findings were replicated in iPSCs carrying a deletion in the 15q11.2 region where similar changes in genes and pathways were observed.

Chapter 4

CYFIP1-regulated genes: biological functions and association with neuro-psychiatric disorders

4.1 Introduction

CNVs affecting chromosome 15q11.2 BP1-BP2 have been previously associated with SZ, ASD and ID. Amongst the genes affected in this region, CYFIP1 is proposed to be the causal gene for these neuropsychiatric disorders. The mechanisms by which changes in CYFIP1 dosage lead to these conditions remain unknown. Many lines of evidence have established a role for CYFIP1 in protein translation and actin dynamics in synaptic density. It is therefore not surprising to finding that CYFIP1 activities are involved in the establishment and maintenance of dendritic spines and synaptic activity (Napoli *et al.*, 2008; Oguro-Ando *et al.*, 2015; Hsiao *et al.*, 2016; Abekhoukh *et al.*, 2017). However, a role for CYFIP1 in early cortical development and neurogenesis is only emerging, and the underlying molecular mechanism remains unknown.

In the previous chapter, I presented the analysis of RNAseq data obtained from neural cultures harbouring CYFIP1 genetic manipulation and 15q11.2 deletion, with a focus on general comparison between CYFIP1 and 15q11.2 deletion mediated transcriptomic changes. It is challenging to identify relevant genes contributing to the phenotype from the thousands of DEGs. A solution would be using the approach known as pathway enrichment analysis. This analysis aims to identify groups of genes with a common function or mechanism that are significantly over-represented amongst DEGs. One such method, GSEA, uses an unfiltered (p value threshold-free) list of genes ranked on the basis of their differential expression. This approach is particularly suitable when information on most genes in the genome is available (e.g. in RNAseq experiments) (Subramanian *et al.*, 2005; Reimand *et al.*, 2019). Another approach is the use of a Fisher's exact or hypergeometric test to quantify the enrichment of a group of genes of interest compared to a background list of genes (Tavazoie *et al.*, 1999; Goeman and Buhlmann, 2007).

Using the information from GWAS, the results of the pathway enrichment analysis can be associated to a specific phenotype. GWAS have led to the identification of common and rare variants such as copy number variants (CNVs) and single nucleotide polymorphisms (SNPs) that contribute to the development of a disease (McCarroll *et al.*, 2008). The objective of GWAS is to identify genetic risk factors for complex diseases through a hypothesis-free method (Bush and Moore, 2012). More than 30 GWAS for SZ have characterised SNPs that contribute to the development of the psychiatric disease. The most recent study identified 145 SZ associated loci through meta-analysis (Pardiñas *et al.*, 2018). Big efforts are being made to identify GWAS significant loci. Now, a study from the Psychiatric Genetics Consortium has identified through meta-analysis novel ASD associated loci that overlap with SZ (Consortium, 2017). Finally, the meta-analysis of

more than 250,000 individuals has characterised genomic loci associated to variation in intelligence (Savage *et al.*, 2018).

In this chapter, I use expert-curated gene sets, to associate the DEGs into groups of biological significance explaining the phenotype observed in cells with altered levels of CYFIP1. This analysis allowed the identification of the top candidate genes that could be driving the observed phenotype in cells with gain- and loss-of-function of CYFIP1. Finally, significant gene sets will be tested in MAGMA using the summary statistics from GWAS for SZ, ASD and ID in order to determine the association of these genes for psychiatric disorders.

4.2 Results

4.2.1 Manipulation of CYFIP1 affects AKT pathway

(I) CYFIP1 gain- and loss-of-function elicit mirroring effects on biological pathways

The previous chapter showed that CYFIP1-produced DEGs exhibit large overlaps between the developmental stages examined. Moreover, over a third of the DEGs generated by CYFIP1tg and CYFIP1ko were shared at each of the time-point analysed. Gene set enrichment analysis (GSEA) is an approach that aims to identify the correlation between a ranked list of genes and a specific phenotype: in this case, cellular pathways and processes. For each dataset, protein-coding DEGs were ranked by their fold change. Using the R/Bioconductor package clusterProfiler, ranked genes were grouped based on their significant association with pathways from the Kyoto Encyclopaedia of Genes and Genomes (KEGG).

A total of 150 unique pathways (p value < 0.05) were identified to be affected amongst the three time-points analysed in CYFIP1tg and CYFIP1ko cells. The normalised enrichment scores (NES) of CYFIP1tg and CYFIP1ko cells showed, in most cases, opposite trends in multiple pathways. The NES of the pathways is a measurement that provides an overall indication whether the genes associated with a phenotype are upregulated or downregulated. Affected pathways found in CYFIP1ko data set showed positive NES values, indicating that genes have a tendency to be upregulated, whereas in CYFIP1tg negative NES values were obtained. The mirrored NES for particular pathways suggests that the gain and loss-of-function of CYFIP1 elicited opposing enrichment score. Amongst the most relevant pathways are terms involved in neural development: Wnt, Sonic Hedgehog, Notch and PI3K-AKT signalling pathways; cell adhesion: cell adhesion

molecules, ECM-receptor interaction and focal adhesion; and cellular metabolism: oxidative phosphorylation, thermogenesis and ribosome (**Figure 4.1**). The full list of significantly dysregulated pathways, including their pvalues and adjusted pvalues calculated by clusterProfiler (**Supplementary Table 3**), can be found in Apendix I. A more stringent analysis using a 5% adjusted p value threshold can be found in Apendix I (**Supplementary Figure 3**, **Supplementary Table 4**).

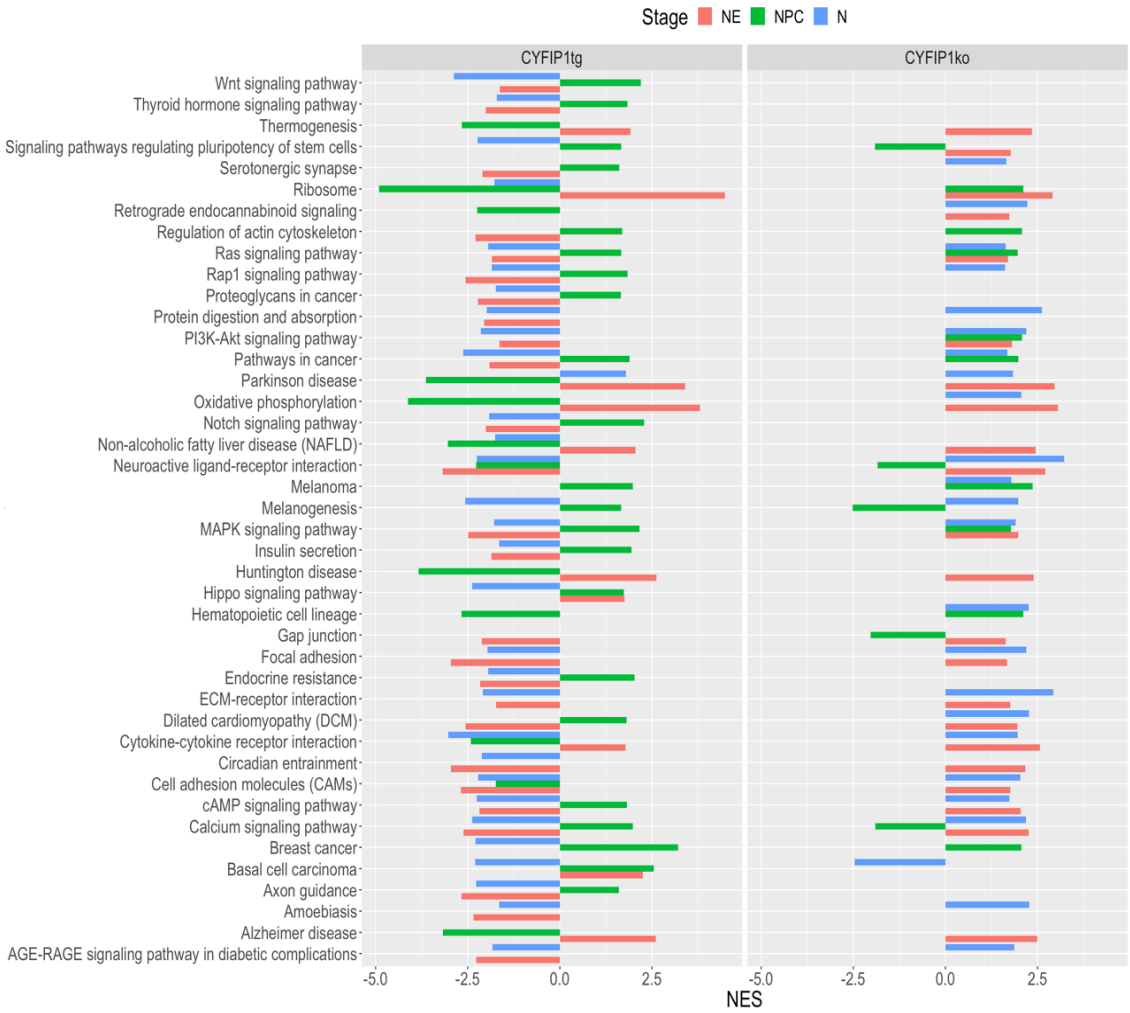


Figure 4.1 Pathways affected by altered dosage of CYFIP1.
Bar graph representing the significantly enriched pathways in CYFIP1tg and CYFIPko cells at different stages of differentiation. Normalised enrichment score for each pathway show mirroring effects in the enrichment between each cell line. For representation purposes, only those that are significantly changed in at least half of the datasets are plotted in the graph.

To identify potential candidate genes that could explain the changes of cortical neurogenesis found in CYFIP1tg and CYFIP1ko cultures, I formulated the following hypothesis: one gene can be part of multiple pathways at the same time; hence if a gene is represented in multiple enriched pathways it would have a more significant weight in the development of an abnormal phenotype. Following this hypothesis, I selected pathways with a Storey corrected p (Storey q value) < 0.05. I studied the number of occurrences in which a gene participated in the subsequent altered KEGG pathways.

The analysis of the top 10, selected as an arbitrary cut-off, altered genes revealed the contribution of three highly correlated pathways: 1) PI3K-AKT containing *PIK3R2*, *AKT3*, *PIK3CD*, *PRKCA*, *PLCB1* and *PLCB4*; 2) MAPK formed by *MAPK1* and *MAP2K1*; and 3) cAMP that consists of *ADCY2* and *ADCY9* (**Table 4.1**). Mutations affecting the 10 genes identified have been reported previously in cases of SZ or ASD (summarised in **Table 4.1**). Moreover, risk variants in the locus for *AKT3* have been identified as a candidate for SZ by the two largest GWAS in this disorder.

(II) Enrichment for AKT is independent to FMRP association

Three isoforms of AKT exist in humans. AKT3 represents more than 50% of the total AKT present in the developing brain, making it the highest expressed isoform. Translation of the AKT3 mRNA is regulated by the CYFIP1-FMRP1-eIF4E complex, along with 40 other mRNAs coding for proteins in the PI3K-AKT pathway. Since the DEGs in CYFIP1tg and CYFIP1ko cells are enriched for FMRP targets in all developmental stages, it could be that the observed enrichment for the PI3K-AKT is simply an artefact of these genes being FMRP targets.

To elucidate if the dysregulation of the PI3K-AKT is directly due to changes in CYFIP1 expression, DEGs were tested for enrichment amongst known FMRP targets using a Fisher's exact test. DEGs in the AKT pathway were significantly enriched in both gain- and loss-of-function of CYFIP1 samples. However, when DEGs in the AKT pathway were tested for enrichment amongst FMRP targets, none but the CYFIP1tg neuroepithelial cell stage showed significance (**Table 4.2 Significant enrichment for AKT in DEGs is independent from FMRP.**

Cell line	Stage	AKT genes targeted by FMRP	AKT genes not targeted by FMRP	DEG threshold	Fisher's pval
CYFIP1tg	Neuroepithelium	23	16	p < 0.1	0.0041033
		161	288	p > 0.1	
	NPCs	28	12	p < 0.1	0.0915466
		281	204	p > 0.1	
	Neurons	19	21	p < 0.1	0.37248333
		198	257	p > 0.1	
CYFIP1ko	Neuroepithelium	16	23	p < 0.1	1.29E-01
		129	290	p > 0.1	
	NPCs	24	16	p < 0.1	1.59E-01
		240	236	p > 0.1	
	Neurons	28	12	p < 0.1	2.82E-01
		297	167	p > 0.1	

). This finding indicates that altered expression of the PI3K-AKT pathway genes are likely caused by changes in CYFIP1 expression directly, rather than secondarily due to the transcriptional repression of the CYFIP1-FMRP1- eIF4E complex. This table can also be found using a 5% FDR threshold for significant DEGs in Apendix I (**Supplementary Table 5**).

	Gene	Gene name	Freq.	Rank	SZ	ASD	GWAS evidence
PI3K-AKT	*PIK3R2*	Phosphoinositide-3-Kinase Regulatory Subunit 2	80	1		(Aguayo *et al.*, 2013)	
	AKT3	AKT Serine/Threonine Kinase 3	71	2	(Afshari *et al.*, 2015; Bergeron *et al.*, 2017; K.R. *et al.*, 2017; Igolkina *et al.*, 2018)	(Aguayo *et al.*, 2013; Baek *et al.*, 2015; Alcantara *et al.*, 2017; Yeung *et al.*, 2017)	SZ[a, b*]
	PIK3CD	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Delta	67	3	(Law *et al.*, 2012; Papaleo *et al.*, 2016; Coulter *et al.*, 2017; Fregoso, Paterson and Law, 2017; Mostaid *et al.*, 2017; Sanders *et al.*, 2017; Yeung *et al.*, 2017; Karbownik *et al.*, 2018)	(Homs *et al.*, 2016; Poopal *et al.*, 2016; Coulter *et al.*, 2017; Kiyota *et al.*, 2018)	
	PRKCA	Protein Kinase C Alpha	67	4	(Carroll *et al.*, 2010; Jablensky *et al.*, 2011)	(Hussman *et al.*, 2011; Huaman *et al.*, 2015; Turner *et al.*, 2016)	
	PLCB1	Phospholipase C Beta 1	57	6	(Kolde and Vilo, 2015)(Gray *et al.*, 2009; Udawela *et al.*, 2011, 2017; Manning *et al.*, 2012; Rita Lo Vasco,	(Awadalla *et al.*, 2010; Girirajan *et al.*, 2013; St Pourcain *et al.*, 2014)	

					Cardinale and Polonia, 2012; Lo Vasco, Longo and Polonia, 2013; Afshari *et al.*, 2015; Kim *et al.*, 2015)		
	PLCB4	Phospholipase C Beta 4	50	8	(Hong *et al.*, 2013)	(Engle, K. M.; Mei, T-S.; Wasa, M.; Yu, 2008)	
MAPK	*MAPK1*	Mitogen-Activated Protein Kinase 1	58	5	(Yuan *et al.*, 2010; Law *et al.*, 2012)	(Yuan *et al.*, 2010; Satoh *et al.*, 2011; Zou *et al.*, 2011; Faridar *et al.*, 2014; Hormozdiari *et al.*, 2015; Wen, Alshikho and Herbert, 2016; Xing *et al.*, 2016; Garg *et al.*, 2017)	
	MAP2K1	Mitogen-Activated Protein Kinase Kinase 1	51	7	(Awadalla *et al.*, 2010; Yuan *et al.*, 2010; Bergen *et al.*, 2014)	(Awadalla *et al.*, 2010; Zou *et al.*, 2011; Ebert and Greenberg, 2013; Xing *et al.*, 2016)	
cAMP	*ADCY2*	Adenylate Cyclase 2	58	9	(Jajodia *et al.*, 2016; Genis-Mendoza *et al.*, 2018)	(Skafidas *et al.*, 2014)	
	ADCY9	Adenylate Cyclase 9	45	10	(Gulsuner and Mcclellan, 2014)		

Table 4.0.1 Top 10 genes contributing to CYFIP1 dysregulated pathways.
*Number of times a gene is altered in all significantly enriched KEGG pathways in CYFIP1tg and CYFIP1ko differentially expressed genes. Amongst the most dysregulated pathways are PI3K-AKT, MAPK and cAMP. The Frequency indicates the number of times a gene significantly contributes to an altered pathway. All genes have previously been implicated with SZ or ASD (or both). Only AKT3 has been successfully identified in GWAS as SZ associated gene: *a: (Ripke et al., 2014) b: (Pardiñas et al., 2018).*

Table 4.2 Significant enrichment for AKT in DEGs is independent from FMRP.

Cell line	Stage	AKT genes targeted by FMRP	AKT genes not targeted by FMRP	DEG threshold	Fisher's pval
CYFIP1tg	Neuroepithelium	23	16	p < 0.1	0.0041033
		161	288	p > 0.1	
	NPCs	28	12	p < 0.1	0.0915466
		281	204	p > 0.1	
	Neurons	19	21	p < 0.1	0.37248333
		198	257	p > 0.1	
CYFIP1ko	Neuroepithelium	16	23	p < 0.1	1.29E-01
		129	290	p > 0.1	
	NPCs	24	16	p < 0.1	1.59E-01
		240	236	p > 0.1	
	Neurons	28	12	p < 0.1	2.82E-01
		297	167	p > 0.1	

Significant differentially expressed AKT genes are not enriched amongst FRMP targets (Fisher exact p > 0.05, with exception to CYFIP1tg neuroepithelial cells.

4.2.2 Overlap DEGs between CYFIP1 mutant cells and those carrying a 15q11.2 deletion

Genes differentially expressed in EA8 and EA62 cells were found to be similar to the genes affected in the CYFIP1 gain- and loss-of-function. More than 50% of the DEGs found in the 15q11.2 deletion and the CYFIP1 datasets were shared (**Figure 4.2 A, B**). Approximately 25% of the DEGs were uniquely expressed by the neural cells with the 15q11.2 deletion. These uniquely expressed DEGs could be attributed to the intrinsic difference in genetic background of the two 15q11.2 deletion subjects, as well as specific changes caused by the other three 15q11.2 deleted genes.

Pathway enrichment analysis of the 15q11.2 DEGs identified multiple pathways involved with cell metabolism, adhesion and signalling were found enriched amongst the DEGs (**Figure 4.2 C**). Because of the substantial overlap between the CYFIP1 and patient iPSC-derived DEGs, it is very likely that the same biological processes are shared between them. Indeed, as observed in the CYFIP1ko cells, “oxidative phosphorylation” took a positive NES score. On the other hand, the pathway “neuroactive ligand-receptor interaction” showed a negative NES as seen in the CYFIP1tg cells. Other pathways like PI3K-AKT and Wnt signalling were also found to be affected in the EA8 and EA62 cells, with a tendency to be decreased. The full list of significantly dysregulated pathways, including their pvalues and adjusted pvalues calculated by clusterProfiler (**Supplementary Table 6**), can be found in Apendix I. A more stringent analysis using a 5% adjusted p value threshold can be found in Apendix I (**Supplementary Figure 4**, **Supplementary Table 7**).

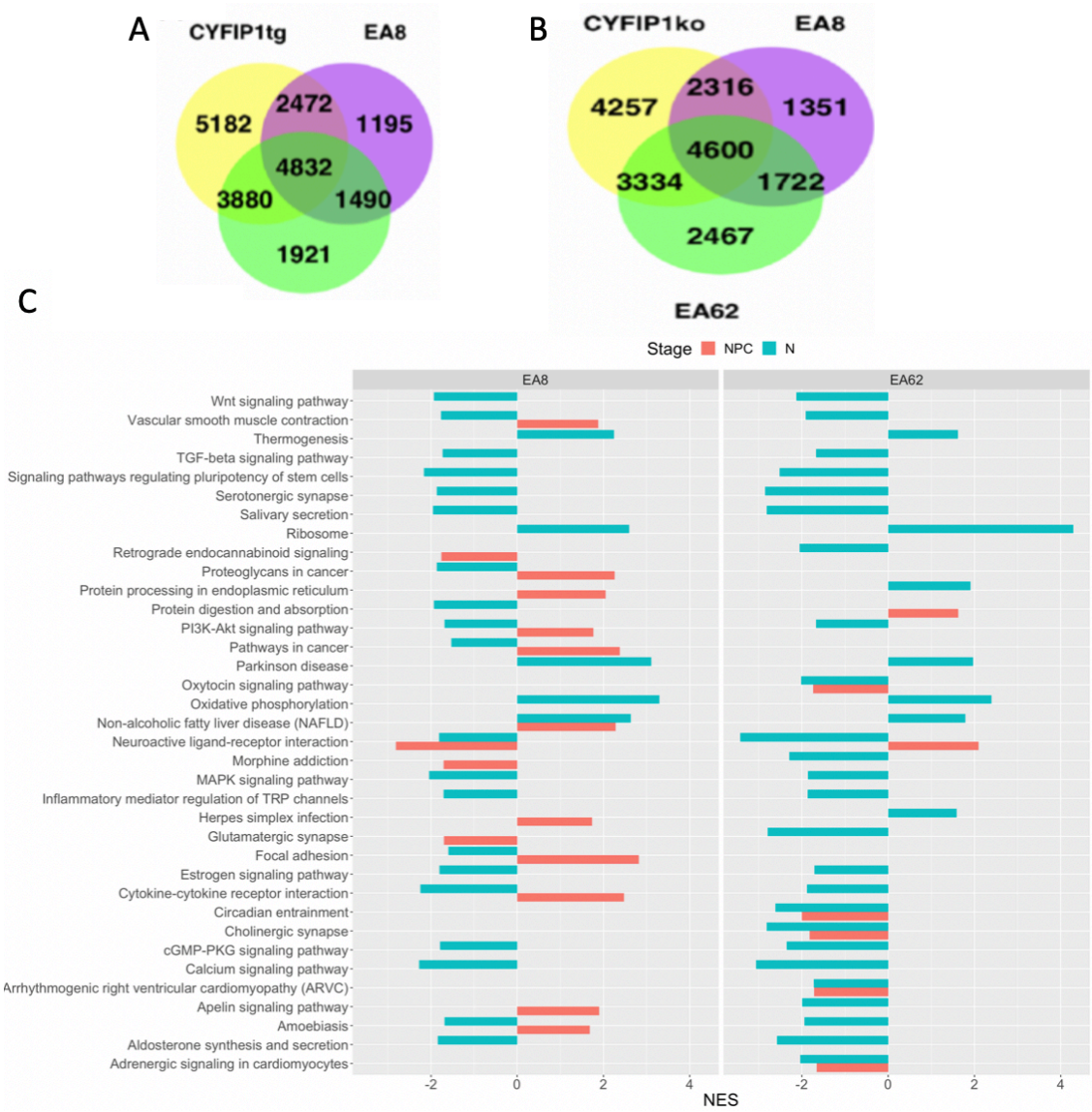


Figure 4.2 Overlapping changes at the gene expression and altered pathways between patient-derived iPSCs and CYFIP1 engineered hESCs.
(A) Bar graph representing the significantly enriched pathways in CYFIP1tg and CYFIPko cells at different stages of differentiation. Normalised enrichment score for each pathway show similar trends in the enrichment between each cell line.

4.2.3 Altered CYFIP1 expression induces changes in cholesterol metabolism and mitochondrial dynamics

(I) GO analysis validates the findings in the KEGG enrichment

In order to complement the information of the KEGG enrichment, gene ontology (GO) analysis was done with protein coding DEGs. GO categories are a more extensive curated collection of gene sets, which can reveal processes not reflected by the KEGG enrichment. This approach identified between 3 and 223 significantly enriched terms for each analysed dataset.

Because many GO terms can share a large overlap between the genes in the group, an iterative procedure was introduced to identify terms whose enrichment was due to the same subset of DEGs. Starting with the smallest significant gene set, genes constituting this set were removed the remaining gene sets, then the enrichment value was re-calculated for the reduced terms. Gene sets no longer significant (p value ≥ 0.05) were removed from the analysis and recorded that the enrichment signal was explained by the smallest gene set tested. This procedure was repeated with the next significant gene set that was not already explained. Further iterations of this cycle were performed until no more significant gene sets were left. Using this approach, the number of enriched gene sets was reduced to at most 7 independently enriched sets, each of which explained the enrichment for multiple other gene sets.

In CYFIP1 engineered cells, the independently enriched terms concerned cell proliferation and differentiation, cytoskeleton and central nervous system development (**Figure 4.3 A-B**). In addition to these, CYFIP1tg cells presented alterations in genes involved in protein synthesis, while genes affected in CYFIP1ko were involved in the extracellular matrix. Importantly, this enrichment analysis identified novel aspects associated with CYFIP1-regulated genes not previously described. The newly-identified processes included terms enriched for mitochondrial function and cholesterol metabolism. In cells carrying a 15q11.2 deletion, terms related to the mitochondria, protein synthesis, extracellular matrix and cellular differentiation were found to be affected. The overlap of affected gene sets between CYFIP1 engineered cells and 15q11.2 deletion cells reinforces the relevance of CYFIP1 in the development of neuropsychiatric disorders in the context of 15q11.2 CNVs (**Figure 4.3 C-D**).

Finally, the iterative refinement algorithm also allowed the identification of groups of terms whose enrichment was captured by a larger set, thus allowing the establishment of a hierarchical tree of functional gene sets (**Figure 4.4**). This process revealed the interconnection between the different gene sets with similar functions.

In summary, the enrichment analysis of the CYFIP1 mutant cells and 15q11.2 deletion carrying cells show the overlap between affected processes in each of the different genotypes. These pathways reflect some of the phenotypic alterations observed in these cells such as defects in neuronal differentiation or terms associated with the cytoskeleton. But most importantly, they highlight new aspects associated with CYFIP1-regulated genes that could be contributing to the observed phenotype and could have otherwise not been observed without the RNAseq analysis.

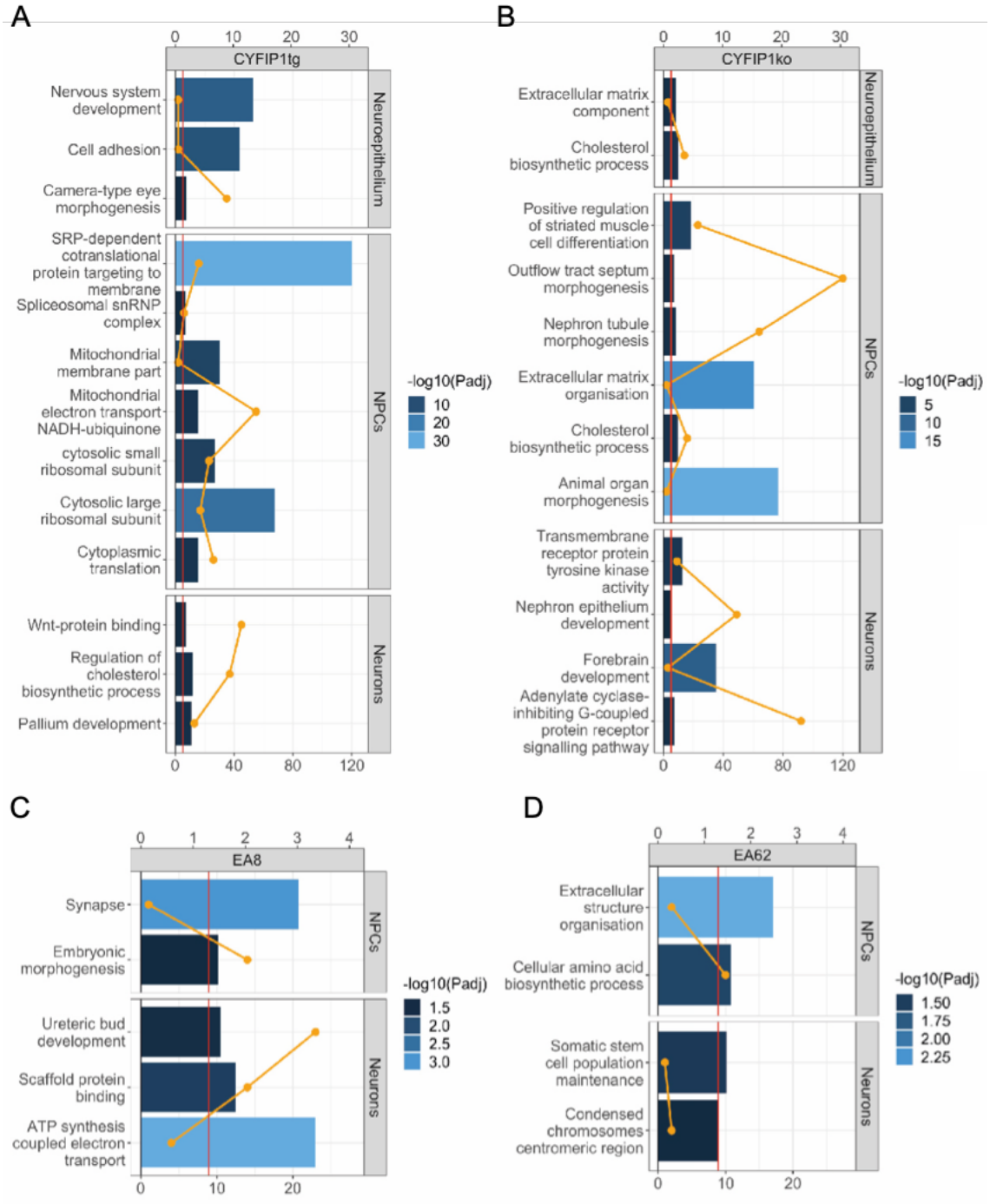


Figure 4.3 Functional gene sets in CYFIP1 engineered and 15q11.2del engineered cells. *(A-D) Individual functional gene sets formed by the smallest group of genes that capture the enrichment signal for that set in particular. Bars indicate the -log10 of the Bonferroni P adjusted, yellow dots indicate the number of gene sets that have been grouped under the functional gene set. (E) Hierarchical tree of the functional gene sets analysed indicating the relationship between parental (darker colours) and child terms (lighter colours).*

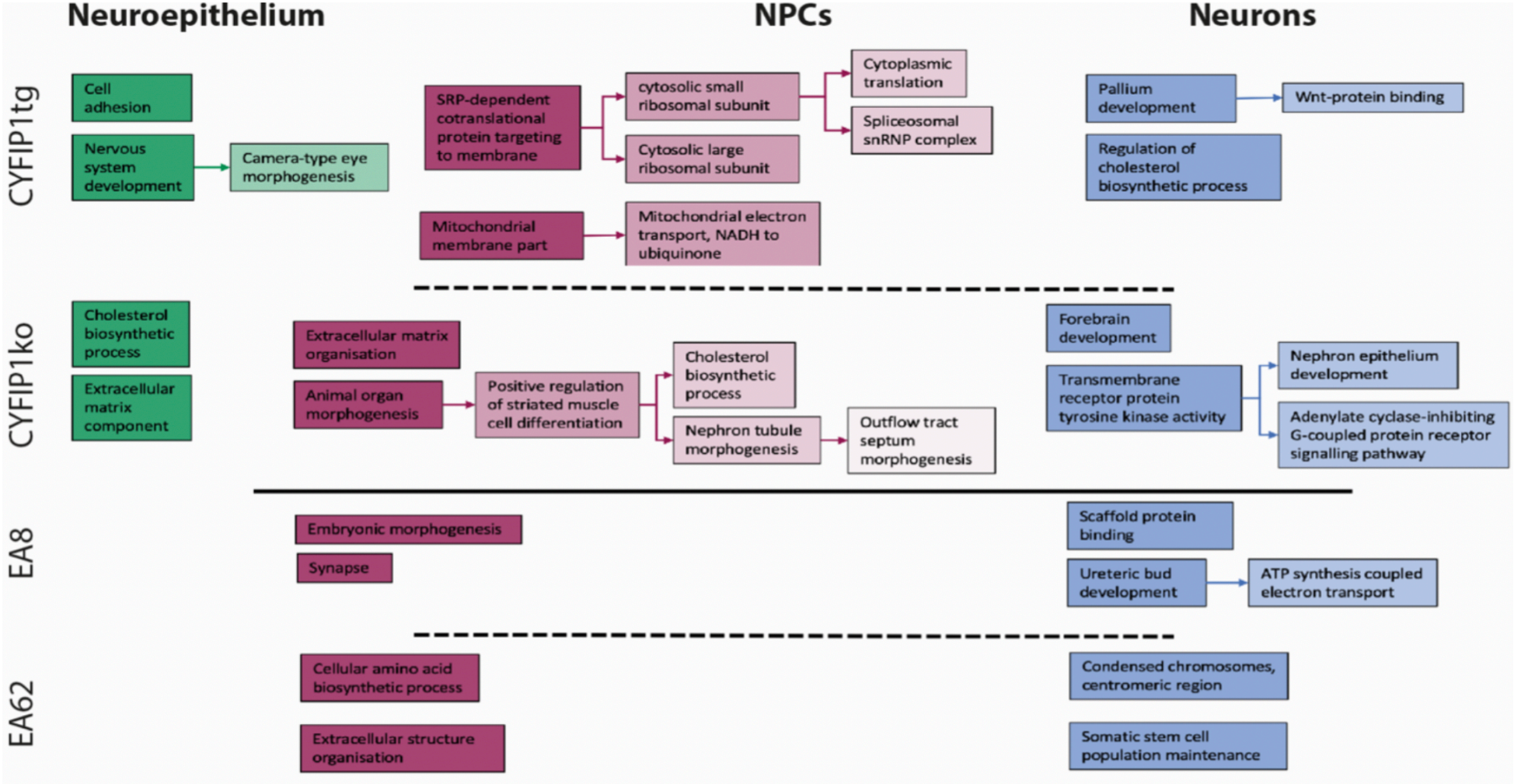


Figure 4.4 Hierarchical tree of the functional pathways.
Hierarchical structure of the significant terms with their associated parental terms.

(II) βCAT signalling and mitochondrial regulation in CYFIP1 manipulated and 15q11.2del neural cells: implications in SZ and ASD

Pathway enrichment analysis with KEGG sets and GO overrepresentation test identified gene sets belonging to the mitochondria amongst the CYFIP1-regulated genes. Mitochondria play an active role during brain development and are required for the normal functioning of the mature brain. During embryonic neural development, mitochondria regulate gene expression through the production of reactive oxygen species (ROS) (Cagalinec *et al.*, 2016; Son and Han, 2018), which in turn modulate signalling pathways including PI3K-AKT, Ras and ERK (Y., S. and R., 2001; Le Belle *et al.*, 2011). In mature neurons, neural transmission relies on accumulation of mitochondria in the synapses in order to produce ATP and regulate calcium concentrations (Evans, Derkach and Surprenant, 1992; Billups and Forsythe, 2002). In addition, mitochondria express different proteins depending on their subcellular location, which confers different functional properties to this cellular organelle (Stauch, Purnell and Fox, 2014). Therefore, changes in gene expression affecting the mitochondria can have a significant impact on brain development as well as normal synaptic activity.

The work carried out by Stauch and colleagues characterised over 1,600 proteins in mouse (Stauch, Purnell and Fox, 2014). Their analysis revealed differences in synaptic and non-synaptic mitochondria, which included the identification of proteins uniquely expressed or that showed differences in their expression in either group. Mouse gene identifiers were mapped to their human homologues using biomaRt and used as gene sets to test for enrichment amongst DEGs. In NPCs with an excess of CYFIP1, non-synaptic mitochondria appear to be differentially expressed (**Figure 4.5 A**). On the other hand, synaptic mitochondria were affected in CYFIP1-deficient neurons (**Figure 4.5 B**). No significant gene sets were identified in cells carrying a 15q11.2 deletion.

There is growing evidence that mitochondria is one of the factors involved in the development of SZ and ASD (Shao *et al.*, 2008; Siddiqui, Elwell and Johnson, 2016; Roberts, 2017). To determine if any of the genes in the significant sets could be involved in the development of psychiatric disorders, they were compared against the SFARI database for ASD genes (Abrahams *et al.*, 2013), and genes that have been identified through risk loci in the most recent GWAS for SZ (Pardiñas *et al.*, 2018). Three genes were found common in both gene sets (**Figure 4.5 C**): 1) *CTNNB1*, the gene encoding βCAT, the downstream effector in the canonical Wnt pathway and AKT signalling, which upon stabilisation induces the expression of the TCF/LEF target genes (Zhang *et al.*, 2013). 2) *TCF20*, found to be altered in both the CYFIP1tg and CYFIP1ko cells, is a gene encoding for a transcriptional coactivator responsive for signalling through oestrogen receptors

(ER) and growth factors (Gburcik *et al.*, 2005), and is shown to enhance the transcriptional activity of various transcription factors including PAX6 (Rekdal, Sjøttem and Johansen, 2000). 3) *ALDOA,* which encodes for a protein expressed across the nervous system and it is necessary for the production of ATP (Mamczur *et al.*, 2013).

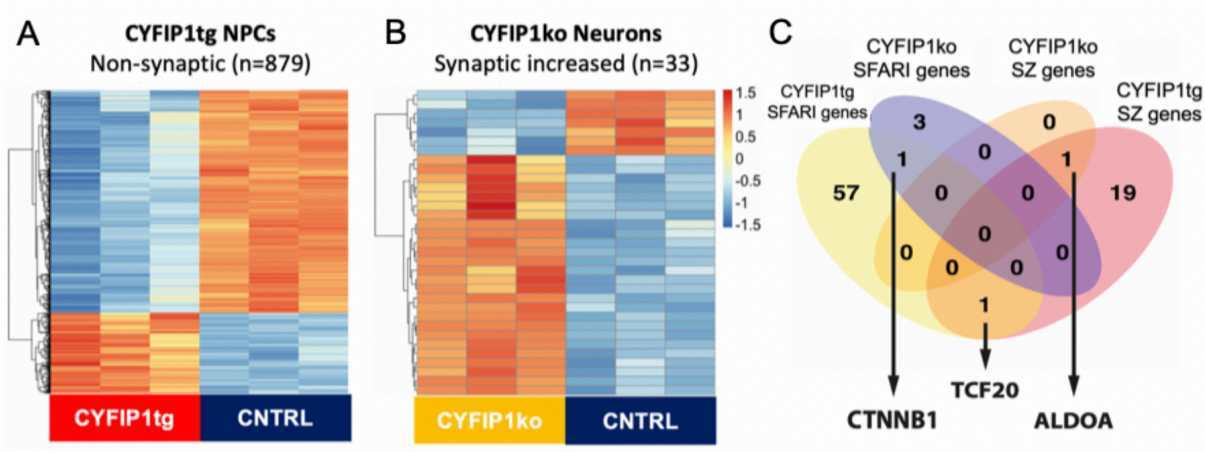


Figure 4.5 Dysregulation of mitochondrial genes in CYFIP1tg and CYFIP1ko cells associated to SZ and ASD genes.
Heatmaps comparing the gene expression significant mitochondrial gene sets generated from those identified in (Stauch, Purnell and Fox, 2014) *in CYFIP1tg NPCs (A) and CYFIP1ko neurons (B) against the control lines. (C) Shared genes in the mitochondrial gene sets with known genes in the SFARI database and GWAS SZ genes. Common genes include CTNNB1, TCF20 and ALDOA.*

Finally, graphical representation of the DEGs coding for elements of the mitochondrial electron transmission chain showed changes affecting the different subunits (**Figure 4.6)**. These changes appeared to be in opposite directions in CYFIP1tg and CYFIP1ko cells, where the affected genes were downregulated in the former and upregulated in the latter.

CYFIP1tg

CYFIP1ko

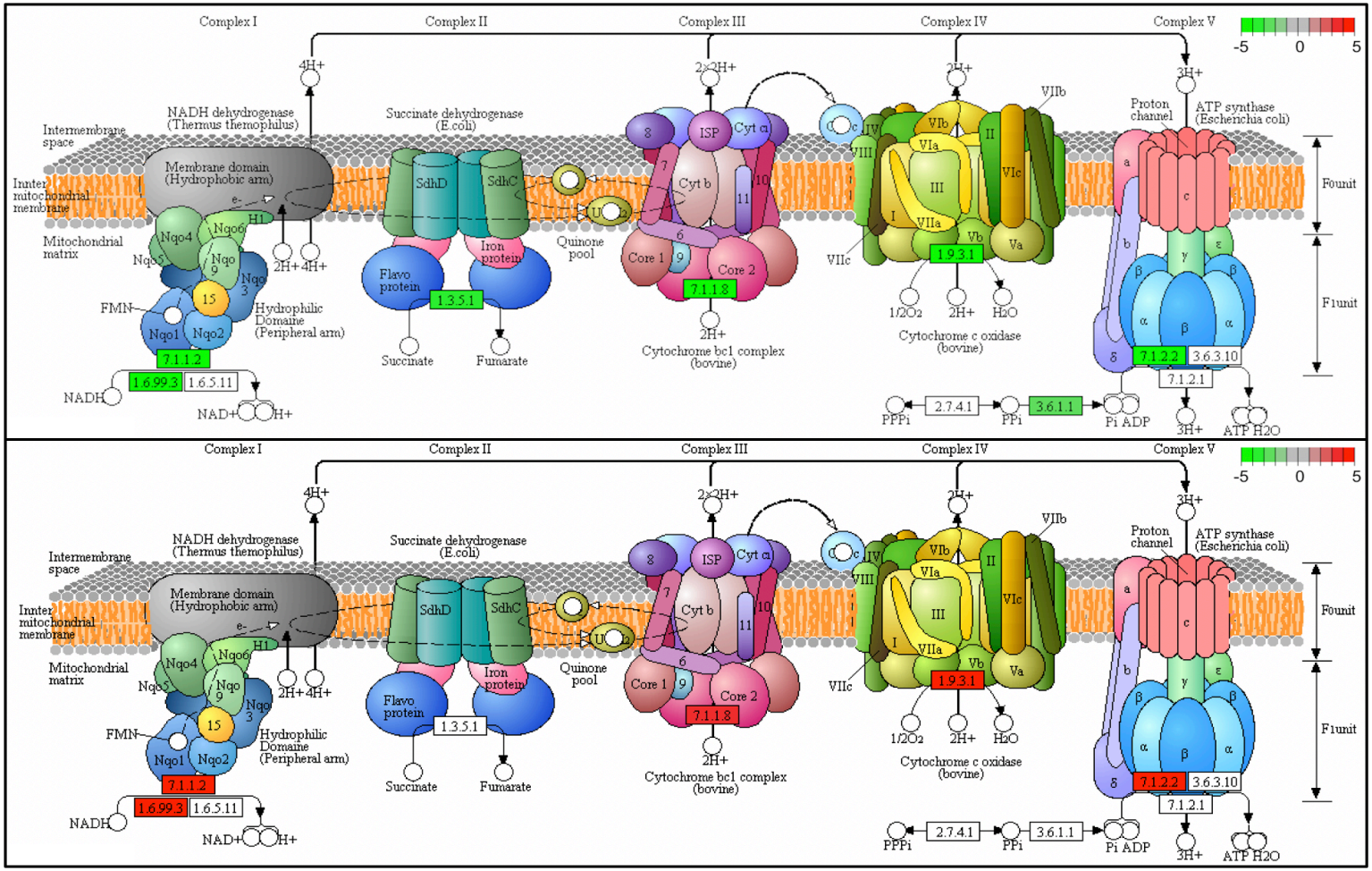


Figure 4.6 Elements of the electron transmission chain affected in CYFIP1tg and CYFIP1ko cells.
Graphical representation using R/pathview of the elements in the mitochondrial electron transport chain affected in the studied cells (green = downregulated, red = upregulated).

4.2.4 CYFIP1 regulated genes are enriched for SZ and intelligence associated genes

As shown in the previous chapter, altered expression of CYFIP1 causes dysregulation of multiple genes. To test whether genes dysregulated in CYFIP1 manipulated cells are associated with neuropsychiatric disorders, DEGs were tested for enrichment in common variant association using MAGMA and summary statistics from GWAS for SZ, ASD and intelligence (Pardiñas *et al.*, 2018; Savage *et al.*, 2018). Because the hypothesis is that CYFIP1-regulated genes contribute to causing the development of neuropsychiatric disorders, we predict no association for the genes involved in neurodegenerative disorders such as Alzheimer's disease. Therefore, the International Genomics Alzheimer's (IGAP) GWAS was included as a negative control.

Significant protein coding DEGs (Bonferroni p adj < 0.05) were grouped into three gene sets: 1) All DEGs, containing all the significant DEGs 2) Upregulated DEGs (FC > 0), with all the significantly upregulated DEGs and 3) Downregulated DEGs (FC < 0), with the significantly downregulated genes. The gene sets were analysed for association with SZ, ASD and intelligence.

Gene sets were correted for the number tests performed (for each bumber of cell types, developmental stages and GWAS analysed). For the analysed datasets, at least one stage and dataset were significantly associated with SZ or intelligence (Bonferroni p adj < 0.05, **Figure 4.7**). Nominal enrichment was found for ASD in neurons derived from patient cell line EA8 and NPCs of CYFIP1ko. However, they did not survive the test for multiple correction (Bonferroni p adj. > 0.05). As hypothesised, none of the differentially expressed gene sets showed statistical significance for IGAP even before correction for multiple testing. These results indicate that CYFIP1-regulated genes are exclusively associated with psychiatric traits and not neurodegenerative disorders.

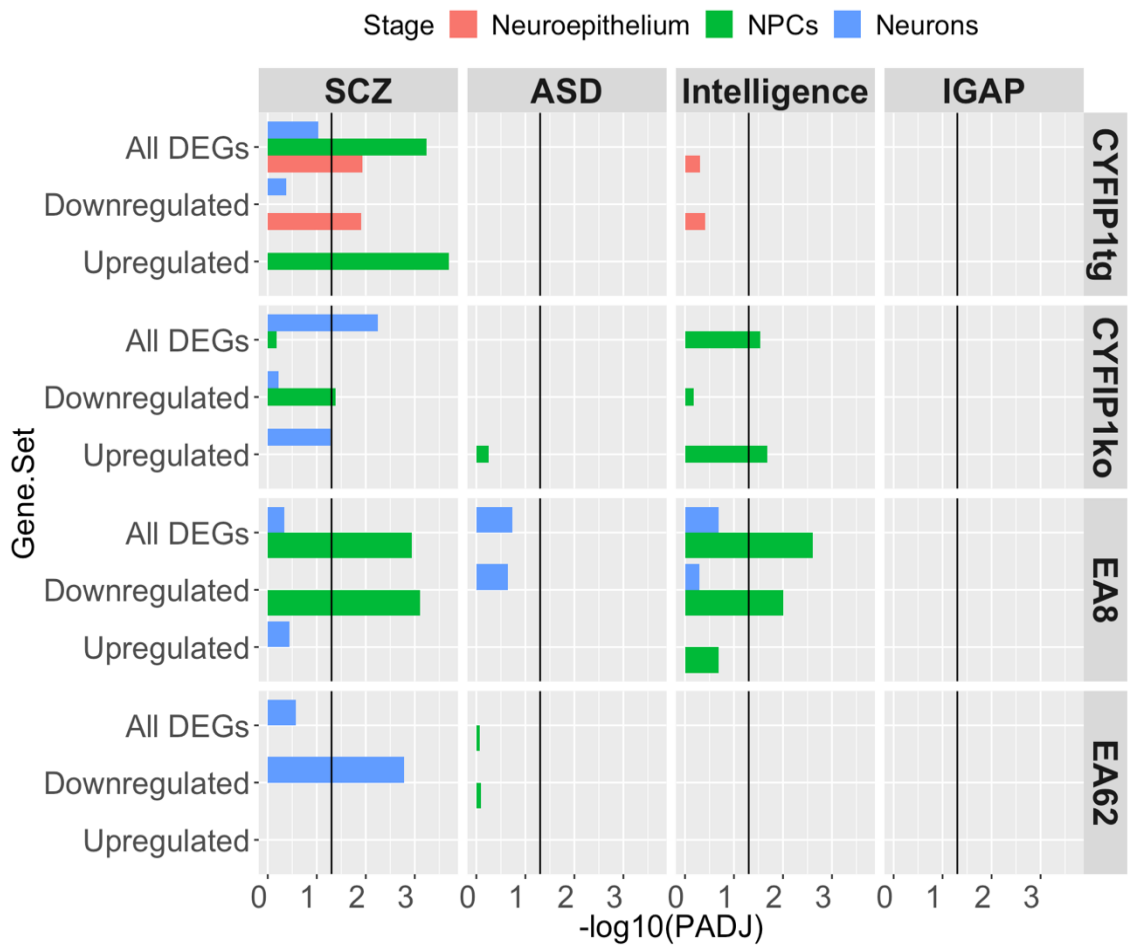


Figure 4.7 Association of CYFIP1 DEGs and neuropsychiatric disorders.
MAGMA enrichment results for GWA studies associated with schizophrenia (SCZ), autism spectrum disorders (ASD), intelligence and the international genomics Alzheimer's (IGAP). Bars indicate the -log10 of the Bonferroni P adjusted over the threshold (black line) in each dataset. No significant association was found for ASD and IGAP.

Next, I sought to identify what is the smallest group of genes that capture a genetic signal for SZ and intelligence traits. Using the iterative analysis previously described, GO enrichment was carried out for those groups of DEGs genes with a significant association for SZ or intelligence in the above paragraph. These newly-generated sets reflected the changes described in section 4.2.3, where GO associated with cell adhesion, central nervous system development and ribosomal translation were found enriched. Interestingly, terms associated with mitochondria were found amongst the genes with a significant enrichment for SZ (**Error! Reference source not found.**), suggesting a role for these in the development of psychiatric disorders. On the other hand, genes involved with intelligence were enriched for cholesterol metabolism (**Error! Reference source not found.**). This analysis revealed different functional aspects in which groups of genes can cause the same disorder.

Table 4.3 Functional sets generated from groups of genes with significant enrichment for SZ.

GROUP	CELL TYPE	STAGE	GENE SET	PVALUE	PCORR	NGENES
All DEGs	CYFIP1ko	Neurons	transmembrane receptor protein tyrosine kinase activity	1.75E-07	0.00069306	27
			adenylate cyclase-inhibiting G-protein coupled receptor signalling pathway	3.36E-06	0.01327774	22
			forebrain development	3.00E-13	1.19E-09	60
			nephron epithelium development	1.24E-05	0.04885423	23
	EA8	NPCs	synapse	2.68E-07	0.00095457	38
			embryonic morphogenesis	9.29E-06	0.0331263	27
	CYFIP1tg	Neuroepithelium	cell adhesion	2.16E-15	8.19E-12	85
			nervous system development	1.06E-17	4.01E-14	184
			camera-type eye morphogenesis	2.96E-06	0.01121383	15
		NPCs	cytoplasmic translation	2.92E-08	0.00011728	28
			mitochondrial electron transport, NADH to ubiquinone	2.75E-08	0.00011064	27
			SRP-dependent cotranslational protein targeting to membrane	9.38E-35	3.77E-31	79
			cytosolic large ribosomal subunit	1.89E-21	7.59E-18	52
			mitochondrial membrane part	5.68E-12	2.28E-08	93
			spliceosomal snRNP complex	3.97E-06	0.01595909	35
Downregulated	CYFIP1ko	NPCs	frizzled binding	1.58E-06	0.00582304	11
			positive regulation of striated muscle tissue development	4.46E-06	0.01637442	10
	CYFIP1tg	Neuroepithelium	cell adhesion	4.59E-11	1.61E-07	58
			retina morphogenesis in camera-type eye	1.03E-06	0.00362699	9
	EA8	NPCs	axon	5.29E-07	0.00158353	17
			cognition	1.54E-05	0.04610691	13
	EA62	Neurons	condensed chromosome, centromeric region	4.38E-08	0.00015979	12
			somatic stem cell population maintenance	1.99E-07	0.00072609	12
Upregulated	CYFIP1tg	NPCs	nuclear chromatin	9.15E-07	0.00359927	58
			cell adhesion	6.91E-08	0.00027154	90
			nervous system development	1.02E-15	4.01E-12	236

Significant functional gene sets generated from genes grouped by their regulation (fold change) showing a significant association with a neuropsychiatric trait.

Table 4.4 Functional gene from groups of genes with significant enrichment for intelligence.

GROUP	CELL TYPE	STAGE	GENE SET	PVALUE	PCORR	NGENES
All DEGs	CYFIP1ko	NPCs	outflow tract septum morphogenesis	4.34E-06	0.01705959	12
			cholesterol biosynthetic process	9.50E-07	0.00373475	16
			animal organ morphogenesis	9.31E-24	3.66E-20	124
			extracellular matrix organization	1.12E-19	4.41E-16	80
			positive regulation of striated muscle cell differentiation	4.53E-09	1.78E-05	16
			nephron tubule morphogenesis	2.17E-06	0.0085247	13
	EA8	NPCs	synapse	2.68E-07	0.00095457	38
			embryonic morphogenesis	9.29E-06	0.0331263	27
Downreuglated	EA8	NPCs	axon	5.29E-07	0.00158353	17
			cognition	1.54E-05	0.04610691	13
Upregulated	CYFIP1ko	NPCs	collagen trimer	2.84E-06	0.0105388	10
			cholesterol biosynthetic process	3.18E-10	1.18E-06	16
			extracellular matrix organization	3.08E-24	1.14E-20	66
			renal tubule development	2.79E-06	0.01034347	12
			platelet aggregation	8.42E-07	0.00312599	11

Significant functional gene sets generated from genes grouped by their regulation (fold change) showing a significant association with a neuropsychiatric trait.

The reduced gene sets are tested in MAGMA for association with SZ or intelligence using the complete list of DEGs in their group as a covariate. The covariate enrichment test determines whether a subset of genes is particularly enriched for a psychiatric trait as compared to the rest of the group. **Error! Reference source not found.** summarises the results of the enrichment analysis. Two functional gene sets showed nominal enrichment ($p < 0.05$) for SZ over and above that seen in DEGs as a whole: “nuclear chromatin” in CYFIP1tg NPCs and “somatic stem cell population maintenance” in 15q11.2 deletion carrier EA62 neurons. In CYFIP1ko NPCs, the functional group “nephron tubule morphogenesis” showed nominal enrichment for intelligence. The gene sets contained DNA binding proteins, cell cycle or protein biosynthesis.

Table 4.5 MAGMA enrichment results for functional gene sets and GWAS.

GWAS	REGULATION	CELL TYPE	STAGE	SET	PVALUE	P CORR	NGENES
SZ	Downregulated	EA62	Neurons	Somatic stem cell population maintenance	0.035318	0.282544	11
	Upregulated	CYFIP1tg	NPCs	Nuclear chromatin	0.045209	0.813762	55
Intelligence	All	CYFIP1ko	NPCs	Nephron tubule morphogenesis	0.013018	0.468648	13
	Upregulated	CYFIP1ko	NPCs	Rnal tubule development	0.016145	0.48435	12

Functional gene sets with nominal enrichment ($p < 0.05$) for SZ and intelligence GWAS. None of the gene sets was significant after Bonferroni correction.

4.2.5 Genes targeted by the CYFIP1-FMRP-eIF4 complex are associated with SZ

Since no significant gene sets were identified, a literature search for gene sets involved with SZ was carried out resulting in 136 gene sets. From these, 134 sets were obtained from the published study by Pocklington and colleagues analysing genome-wide CNV data (Pocklington *et al.*, 2015). These gene sets were derived in part from a CNS-related subset of the MGI mammalian phenotype database, that contains groups of genes with an experimentally tested phenotype, with a postulated relevance to SZ. Besides these, the authors included gene sets originated from previous SZ studies and curated proteomic literature. These 134 gene sets included terms involving function, development and structure of the nervous system; behaviour and some subcellular neuronal terms (e.g. post-synaptic density complex, ARC, etc.). Those gene sets were complemented with FMRP target genes, and loss-of-function (LoF) intolerant sets of genes (Darnell *et al.*, 2011; Lek *et al.*, 2016). There is extensive evidence that CNVs and mutations affecting FMRP target genes and LoF intolerant genes are more common in individuals with SZ than controls (Fromer *et al.*, 2014; Purcell *et al.*, 2014; Genovese *et al.*, 2016; Leonenko *et al.*, 2017; Curtis *et al.*, 2018).

Enrichment analysis was carried out using all the significant DEGs for the CYFIP1tg, CYFIP1ko and 15q11.2 cells. Out of the 136 gene sets tested, 44 unique terms were significantly enriched in all the datasets analysed (**Figure 4.8**). The FMRP-target genes were at the top affected group, appearing enriched in all time-points analysed for the CYFIP1 engineered cells. Multiple gene sets involved in CNS structure and morphology as well as synaptic terms were found to be significantly enriched for all the cell types analysed in different developmental stages. These terms included abnormal neuron differentiation and morphology, synaptic transmission and abnormal brain morphology and structure. CYFIP1ko NPCs showed the largest number of enriched terms where 28 out of 35 unique terms were found to be significantly enriched amongst these gene sets. Moreover, the CYFIP1ko cells showed the largest similarity to that of the 15q11.2del cells. Out of the 35 enriched terms enriched in CYFIP1ko cells, 13 of them were shared with at least one of the patient-derived cells.

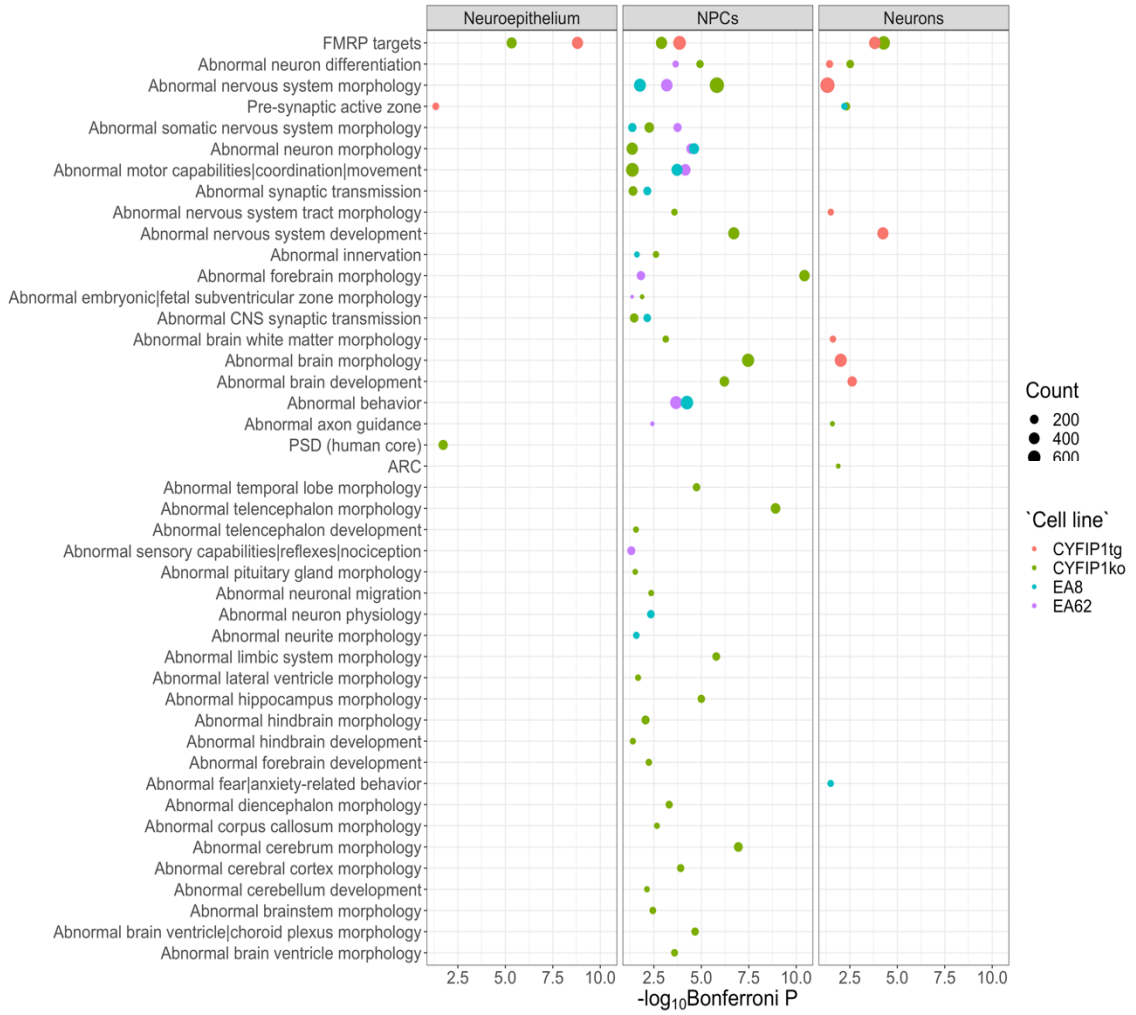


Figure 4.8 Functional enrichment for CNS-specific gene sets.
Dot plot summarising the results of the functional gene sets for CNS-specific gene sets including FMRP and LOF intolerant genes. Gene sets were ordered by the number of datasets for which they were enrich, with FMRP-target genes being the most relevant. The NPC stage captured the largest number of significant terms after Bonferroni correction. CYFIP1ko cells shared the biggest overlap with terms originated from the 15q11.2del carrying cells. The size of the dot is proportional to the number of DEGs in the gene set.

These gene sets were tested in MAGMA for enrichment in SZ-associated common variants. All the DEGs for each particular dataset were used as a covariate to determine the specific enrichment of a gene set over and above that of DEGs as a whole. No gene sets were found to be significant for EA8 cells. Gene sets were corrected for the number of tested sets and number of datasets analysed. After Bonferroni correction, only DEGs regulated by FMRP were found to be significantly enriched for SZ (**Error! Reference source not found.**).

Table 4.6 MAGMA enrichment results for CNS gene sets and GWAS for SZ.

LINE	STAGE	SET	NGENES	BETA	BETA STD	SE	P	BONFERRONI
CYFIP1tg	Neuroepithelium	FMRP targets	426	0.187	0.0281	0.0584	0.00066525	0.007983
	NPCs	FMRP targets	610	0.205	0.0365	0.047	6.5974E-06	3.95844E-05
	Neurons	FMRP targets	458	0.225	0.035	0.0546	1.8586E-05	0.000892128
CYFIP1ko	Neuroepithelium	FMRP targets	322	0.158	0.0207	0.0653	0.0077274	0.0927288
	NPCs	Abnormal neuron differentiation	120	0.238	0.0191	0.11	0.014881	1
		FMRP targets	431	0.307	0.0462	0.0583	7.3838E-08	1.5063E-05
		Abnormal hindbrain morphology	166	0.165	0.0155	0.0895	0.032724	1
		Abnormal neuron morphology	442	0.116	0.0176	0.0568	0.02101	1
	Neurons	FMRP targets	563	0.173	0.0297	0.0497	0.00025328	0.0075984

Table summarising the results for the genetic enrichment in MAGMA for CYFIP1tg and CYFIPko cells. Only genes in the FMRP-targets gene set captured significant genetic association with SZ after Bonferroni correction.

4.3 Discussion

This chapter explores the biological significance of the DEGs identified in the previous chapter and determine the association of CYFIP1-regulated genes with the development of neuropsychiatric disorders. First, DEGs were used to construct functional gene sets: groups of genes with a shared enrichment signal that exercise a function in particular. These gene sets reflected the cellular phenotype observed in the cells and act as a confirmation of the pathway enrichment analysis with KEGG pathways. Finally, altered pathways were tested in MAGMA using GWAS summary statistics for different neuropsychiatric traits to determine the association of the pathways with the diseases.

Previous transcriptomic analysis revealed that CYFIP1 knock down in NPCs caused alterations pathways involved in cell division, ATP synthesis and cytoskeleton remodelling and were enriched for FMRP targets (Nebel *et al.*, 2016). Pathway enrichment analysis of CYFIP1tg and CYFIP1ko are in agreement with these findings and identify particular signalling pathways involved in the development of the observed phenotype and linking alterations in CYFIP1 expression with developmental delays. Three pathways were characteristically represented by the top 10 most relevant genes: PI3K-AKT, MAPK and cAMP. The three pathways play a role in cell growth, differentiation and survival. During embryonic cortical development, these pathways are critical for neural progenitor cell survival and are involved in maintaining the balance between self-renewal and or differentiation.

The three pathways are known to interact between each other through multiple crosstalk and therefore, alteration in one of them could trigger compensatory mechanism in the others and reflected at the transcriptomic level. Amongst these genes, the locus

containing *AKT3* has been previously identified as a risk gene for SZ by GWAS and its mRNA is regulated by the FMRP1-CYFIP1-eIF4E complex (Darnell *et al.*, 2011; Ripke *et al.*, 2014; Skene *et al.*, 2018). This makes *AKT3* an interesting candidate for further study. The role of AKT3 in the development of the phenotype observed in CYFIP1 gain- and loss-of-function will be therefore described in Chapter 5. Moreover, the transcriptomic characterisation of cells carrying a 15q11.2 deletion indicated similar changes to those observed in the CYFIP1 engineered cell lines. More than 50% of the DEGs were shared between the patient-derived cells and the hESC derivatives with altered CYFIP1 levels.

The functional enrichment analysis identified gene sets involving cell adhesion, telencephalon development and synapse, which further confirm the findings in the phenotypical analysis. The functional sets also identified novel functions altered by the dysregulation of CYFIP1: cholesterol metabolism and mitochondria. There is growing evidence indicating that the mitochondria plays a role in the development of neuropsychiatric disorders, particularly in SZ (Shao *et al.*, 2008; Rollins *et al.*, 2017; Sullivan *et al.*, 2018). However, the molecular mechanism behind this process is still poorly understood. Little is known about the role of the cholesterol in SZ. It is known that schizophrenic patients develop altered levels of cholesterol in blood (Solberg *et al.*, 2016) and exist a positive correlation between cholesterol levels and cognitive index (Krakowski and Czobor, 2011), but whether it is a cause or a consequence of the disorder remains to be elucidated. The analysis carried out in this chapter shows that alterations in cholesterol metabolism appear in the very early stages of cortical development. This opens an opportunity to identify potential new biomarkers in the development of SZ.

Alterations affecting CYFIP1 expression have been previously associated with neuropsychiatric disorders such as SZ, ASD and intellectual disabilities. These disorders are highly polygenic, where alterations in multiple genes have been characterised for the different conditions. At the same time, CNV studies have identified the same genes across these neuropsychiatric disorders, which indicates that alteration of the same pathways can have a different phenotypical outcome (Chung, Tao and Tso, 2014). The genetic analysis of the DEGs in CYFIP1 engineered cells and 15q11.2del carrier cells show a positive enrichment for SZ and intelligence. However, trying to pinpoint a small subset of genes has not been possible. The conditional analysis of the functional gene sets did not identify any significant sets, which means that the overall group of DEGs is involved in the development of the psychiatric traits. It is possible that the highly polygenic nature of the disorders makes it difficult to characterise a small subset of genes.

Finally, the use of expert curated datasets showing association for SZ raised the attention to the CYFIP1-regulated genes that are targets of FMRP. The first evidence for the

relevance of these genes was shown in Chapter 3, where it was shown that genes regulated by FMRP were highly enriched amongst the DEGs. Despite the fact that the analysis carried out by Pocklington *et al.,* did not find a significant association between SZ and FMRP target genes, they observed a tendency for this gene set towards enrichment (Pocklington *et al.*, 2015). The difference between these results can be attributed to the sample size used to determine the enrichment for SZ. In the first case, Pocklington and colleagues were limited to CNV data of 27,771 samples (11,355 cases and 16,416 controls) whereas the analysis presented here was carried out with a sample size of more than 105,000 probands (40,675 cases and 64,643 controls) (Pocklington *et al.*, 2015; Pardiñas *et al.*, 2018). The analysis conducted in this chapter, reinforces the importance of these genes in the development of SZ upon dysregulation of *CYFIP1* expression. In both CYFIP1tg and CYFIP1ko cells, FMRP targets captured a significant genetic signal for SZ. Amongst the multiple mRNAs regulated by FMRP we can find *AKT3* and *GSK3*β, some of the key regulators of the βCAT signalling pathway.

In conclusion, the analysis of the transcriptomic profile of the CYFIP1 gain- and-loss-of-function cells as well as the patient cells carrying a deletion in the 15q11.2 confirms the phenotypical findings already characterised. Two novel functions of CYFIP1 regulated genes are identified with this analysis: cholesterol metabolism and mitochondria; and gives further evidence that the WNT/AKT/βCAT axis plays a fundamental role in the development of an abnormal development when CYFIP1 is altered. Most importantly, this chapter shows, for the first time, that CYFIP1-regulated genes generally associated with GWAS candidate genes for SZ and intelligence, and that FMRP target genes are driving the association with SZ.

Chapter 5

CYFIP1 modulates neurogenesis by regulating AKT and BCAT signalling

5.1 Introduction

Transcriptomic analysis of gain- and loss-of-function of CYFIP1 cells presented in the previous chapter indicates that amongst the multiple affected pathways, these cells show alterations in the PI3K-AKT and WNT-βCAT signalling pathways. During cortical development, the levels of transcriptionally active βCAT and total NCAD in the VZ are positively correlated, and it is mediated, in part, through AKT signalling (Zhang *et al.*, 2013).

βCAT is a downstream effector of multiple signalling pathways including AKT, WNT or Notch (see Chapter 1). In basal conditions, βCAT is actively phosphorylated in the N-terminus (p-Ser 33/36), which primes the ubiquitination of βCAT for degradation through the proteasome. This process is mediated by a destruction complex formed by adenomatous polyposis coli (APC), axis inhibition protein (AXIN), serine/threonine casein kinase 1α (CSK1α), protein phosphatase 2A (PP2A) and glycogen synthase kinase 3β (GSK3β). However, the binding of WNT to its receptor frizzled (Fz) and lipoprotein receptor-related protein 5/6 (LRP5/6) causes the recruitment of AXIN to the cell membrane. The recruitment of AXIN disrupts the destruction complex allowing the accumulation of BCAT in the cytoplasm, which later translocates to the nucleus and activating a gene response. Alternatively, AKT can phosphorylate both GSK3β and βCAT. This activity causes the stabilisation of active βCAT and increases its transcriptional activity independently from WNT signalling activation (Fang *et al.*, 2007).

The PI3K-AKT pathway plays multiple roles in cortical development. Regulation of radial glia migration, dendritic growth, cell survival and proliferation are some of the characterised functions for this signalling pathway (Jossin and Goffinet, 2007; Diez, Garrido and Wandosell, 2012; Itoh *et al.*, 2016). Genetic alterations affecting genes in the PI3K-AKT pathway have been associated with numerous brain malformations including megalencephaly, microcephaly and cortical dysplasia (Tokuda *et al.*, 2011; Jansen *et al.*, 2015), demonstrating the importance of this pathway for the normal brain development.

The AKT family is comprised of three proteins: AKT1, AKT2 and AKT3, which are encoded by three independent genes but share a large structural homology (Kumar and Madison, 2005). Three domains are present in all isoforms of AKT: 1) the pleckstrin homology (PH) domain, which allows the interaction of AKT with phosphatidylinositol (3,4,5)-triphosphate (PIP_3) and is required for its translocation to the cell membrane. 2) a protein kinase domain, which exercises the catalytic function of the protein and 3) a regulatory domain at the C-terminal tail (Elghazi *et al.*, 2007). In order to be activated, AKT needs to be recruited first to the cell membrane through the interaction between the PH domain and the PIP_3. The kinase can then be phosphorylated by other proteins that subsequently enable the phosphorylation activity of AKT.

Despite the large structural homology of around 80% of the amino acidic sequence, each AKT isoform has a distinct expression pattern and function (Kumar and Madison, 2005; Elghazi *et al.*, 2007). AKT1 regulates whole body organism growth (Cho *et al.*, 2001); AKT2 is present in insulin-responding tissues and is involved in glucose homeostasis (Cho, 2001); while AKT3 plays a role in regulating the normal brain size by controlling the cell number and size (Easton *et al.*, 2005). All AKT isoforms are expressed in the developing and adult human brain with AKT2 being expressed only in astrocytes and AKT1 and AKT3 in neuronal cells (Levenga *et al.*, 2017). AKT3 is highly expressed in both human and mouse brains, with its expression being higher during the foetal developmental stages. Studies in mice have shown that AKT3 makes up 50% - 60% of the total AKTs in the cortex and hippocampus and exhibits an increased activity of 15- to 40-fold compared to that of AKT1 and AKT2. Only AKT3 has been previously identified as a loci with genome-wide significance to be associated with SZ by the two largest SZ GWAS (Ripke *et al.*, 2014; Pardiñas *et al.*, 2018). Moreover, the mRNA of *AKT3* is one of the multiple mRNAs regulated by the CYFIP1-FMRP-eIE4F complex.

In line with the transcriptomic analysis results, preliminary western blot analysis of our-CYFIP1 gain- and loss-of-function cells throughout cortical neuron differentiation showed altered levels of NCAD and βCAT (Tamburini, unpublished). A tendency for higher levels of NCAD was observed in differentiating CYFIP1tg cells, which were accompanied by an increase in the active form of βCAT (S552) and a decrease of inactive βCAT (S33/37). These findings were observed in the opposite direction in the CYFIP1ko cells.

In light of the findings presented in Chapters 3 and 4 of this thesis, we hypothesised that the maintenance of $PAX6^{+}$ NPCs in CYFIP1tg cells could be driven by AKT, more specifically through AKT3. In order to demonstrate this hypothesis directly, I first examined by Western blot analysis the level of proteins involved in AKT-mediated activation and βCAT inhibition. Secondly, I developed a lentiviral-based CRISPR/Cas9 genome editing system for targeting *ATK3* in order to modulate its activity at the genetic level in CYFIP1tg NPCs and evaluated the effects of *AKT3* manipulation on cortical differentiation.

5.2 Results

5.2.1 CYFIP1 overexpression leads to decrease of GSK3B and AKT3 protein levels and their phosphorylation

Phosphorylation of βCAT through GSK3β promotes its degradation and inhibit its transcriptional activity, while active AKT phosphorylates GSK3β at N-terminus (p-S9) to inactivate the protein, allowing the stabilisation of βCAT (Zhang *et al.*, 2013). Changes in this tightly regulated process have effects on cellular morphology, proliferation and patterning of the CNS (Valvezan and Klein, 2012). Consistent with the transcriptomic analysis, preliminary Western blot analysis of our CYFIP1 engineered cell lines (by an ex-PhD student) confirmed a change in the NCAD-βCAT pathway, which is likely WNT-independent. However, the connection between AKT and βCAT remains to be established. I, therefore, performed further Western blot to determine the level of these two proteins.

I found that, compared to the isogenic control cells at the same differentiation stage, the total amounts of AKT3 was reduced in CYFIP1tg cells at day 20 of differentiation, while no change was observed at days 35 and 45 (**Figure 5.1**). However, the amount of pAKT3 (p-S473) was increased at day 20 in CYFIP1tg cells, leading to an increase in the ratio between p-S473 and total AKT3. Moreover, the total GSK3β was found to be significantly decreased in CYFIP1tg cells compared to the controls at all three time points analysed. Furthermore, the levels of GSK3β in control cells appeared to increase over time, whereas its levels in the CYFIP1tg were similar in this time window. The phosphorylation levels of GSK3β (the ratio between p-S9 and total GSK3β) were comparable between the control and CYFIP1 overexpressing cells across the three time-points, indicating that the little amounts of GSK3β present are being inactivated at the same level as in control cells. I also examined the level of the WNT receptor LRP5/6 and found it unchanged at all stages. These results are consistent with the previous findings and provide further evidence on increased activity of NCAD/AKT/βCAT signalling in CYFIP1tg cells.

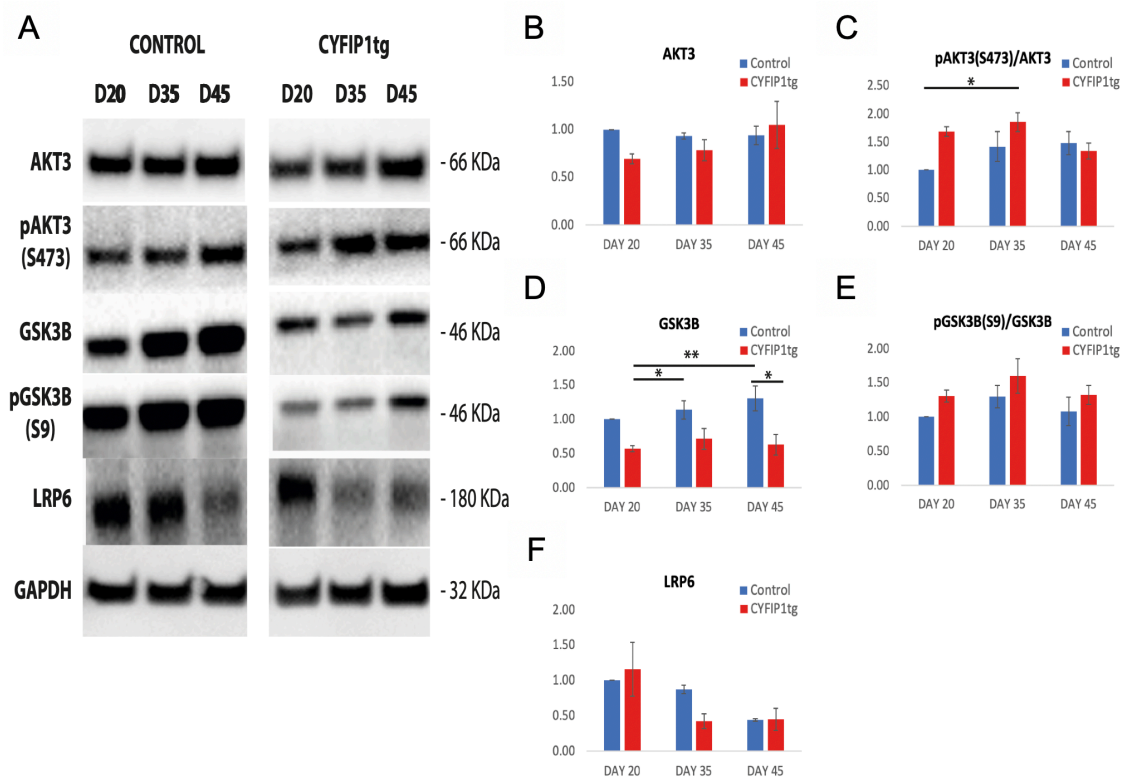


Figure 5.1 Western blot analysis of genes in the NCAD signalling pathway in CYFIP1tg cells.
(A) Western blots for AKT3, phosphorylated AKT3 (p-S473), total and phosphorylated (p-S9) GSK3B, LRP6 and GAPDH in CYFIP1tg at three time points defined as NPCs, early and mature neurons corresponding to days 20, 35 and 45 of differentiation. Bar graphs representing the quantification of AKT3 (B), p-AKT (S473) over total AKT3 (C), total GSK3B (D), p-GSK3β (S9) over total GSK3β (E) and LRP6. Data were compared by one-way ANOVA followed by Tukey post-hoc test ($p < 0.05$, ** $p < 0.01$), n = 3 independent experiments.*

5.2.2 Genetic manipulation of AKT3 in CYFIP1tg cells

(I) A lentiviral CRISPR/Cas9-based gene editing system

The results obtained in the previous section indicate that a reduced expression of GSK3β combined with an increased AKT3 kinase activity could be the leading cause for the imbalance in βCAT activity previously identified in differentiating CYFIP1tg cells. My transcriptomic analysis gathered evidence for *AKT3* as a target gene to be interrogated for validating the causal relationship. Multiple kinase inhibitors are available to reduce the AKT pathway activity. However, these are promiscuous inhibitors that are not specific for individual AKT isoforms and therefore would not be adequate to study the effect driven by AKT3. Consequently, genetic manipulation of *AKT3* was chosen as my experimental approach.

NPCs are one of the 'difficult to transfect' cell type using classic gene delivery procedures such as lipofection or calcium phosphate-based transfection. Viral vectors have been used increasingly due to their high transduction efficiency. Amongst the different viral-based vectors, lentiviral systems are able to produce stable, long-lasting genetic modifications by introducing their genetic material into the host cell (Johnston *et al.*, 1999). Another advantage of using lentiviral vectors is their ability to infect both dividing and non-dividing cells without causing a high level of cell death (Jandial *et al.*, 2008). The recently developed DNA-manipulating tool CRISPR/Cas9 has become a popular option to edit genomic DNA of mammalian cells. This system combines the use of a chimeric single guide RNA (sgRNA) to guide the nuclease Cas9 that generates a double-strand break in the area adjacent to the complementary sequence of the sgRNA. This will promote mechanisms of DNA repair, causing InDels in the targeted region and can result in the introduction of an early STOP codon causing loss of the functional protein. For these reasons, a lentiviral-based CRISPR/Cas9 system was chosen to manipulate the expression of *AKT3* in CYFIP1tg NPCs.

The method involves a second-generation lentiviral-based expression system which can carry up to four sgRNAs, a sequence coding for the spCRISPR/Cas9 and an enhanced green fluorescent protein (GFP) tag (**Figure 5.2 A**) (Kabadi *et al.*, 2014). Each sgRNA was initially cloned into an individual plasmid downstream of a distinct promoter. The four guides were then assembled into the final expression vector through golden gate assembly. To generate this construct, the sgRNAs were designed to target the first coding exon of *AKT3* containing the PH domain necessary to bind to the PIP_3 which leads to the activation of AKT (**Figure 5.2 B**) (Alessi and Cohen, 1998). The sgRNAs with the lowest off-target score were selected (**Figure 5.2 C**). This is particularly relevant since lentiviruses integrate their genetic material into the host genome generating stable integrations which can translate into potential off-targets derived from the CRISPR/Cas9 activity.

The final lentiviral donor plasmid coding for the sgRNAs-Cas9-eGFP was first validated by restriction enzyme digestion of the region containing the insert. This showed the introduction of the expected1,8 Kb sequence compared to the 500 bp sequence of the empty plasmid (**Figure 5.2 D**). Secondly, the number of guides cloned into the construct was assessed by PCR across the region containing the sgRNAs using primers as described by Kabadi and colleagues (Kabadi *et al.*, 2014). Due to the repetitive nature of the sequence inserted, the cloned sgRNA units produced a ladder of PCR amplicons with the highest being at 1,8 Kb (**Figure 5.2 E**), indicating that the four sgRNAs were successfully cloned into the Cas9 plasmid.

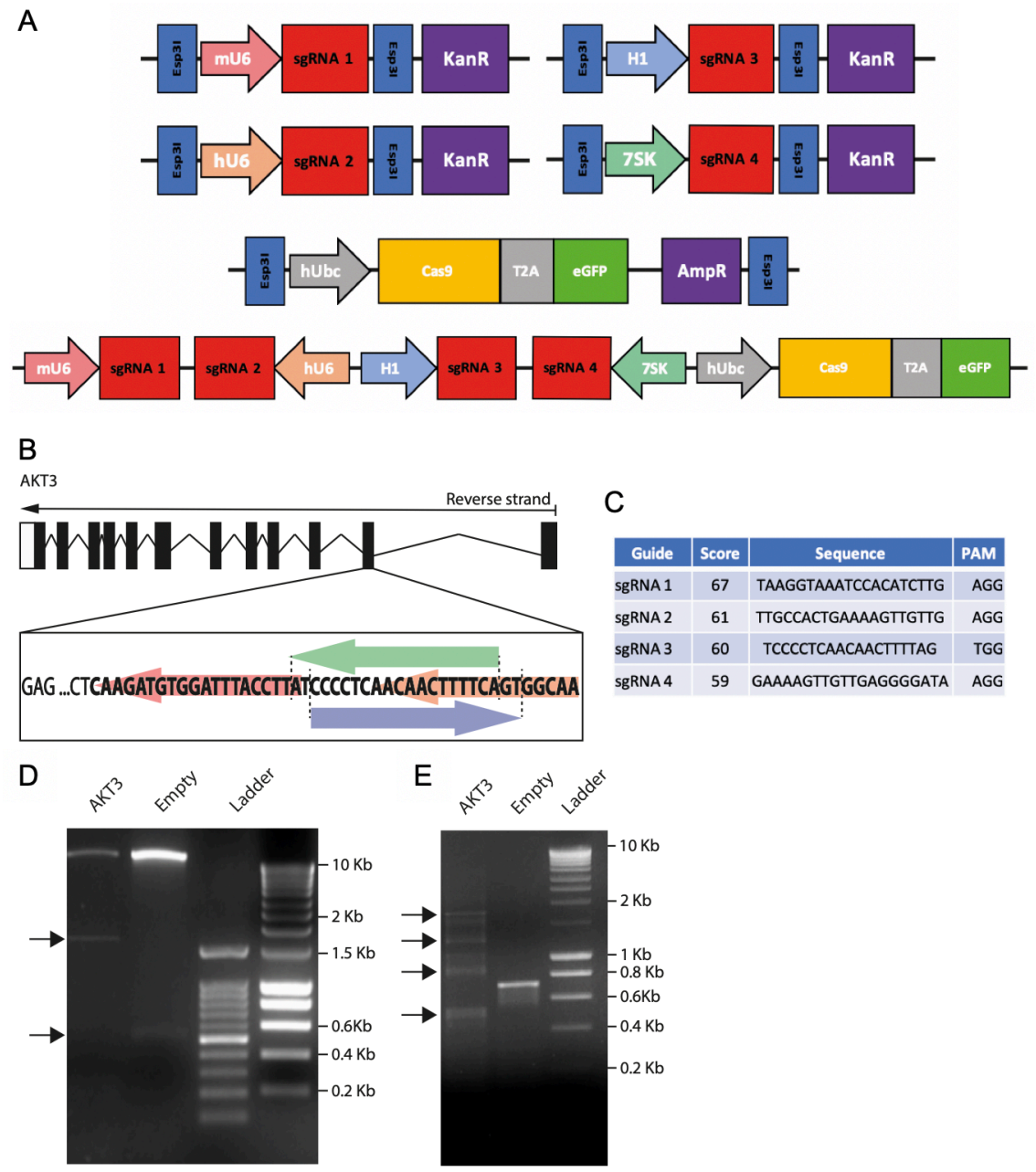


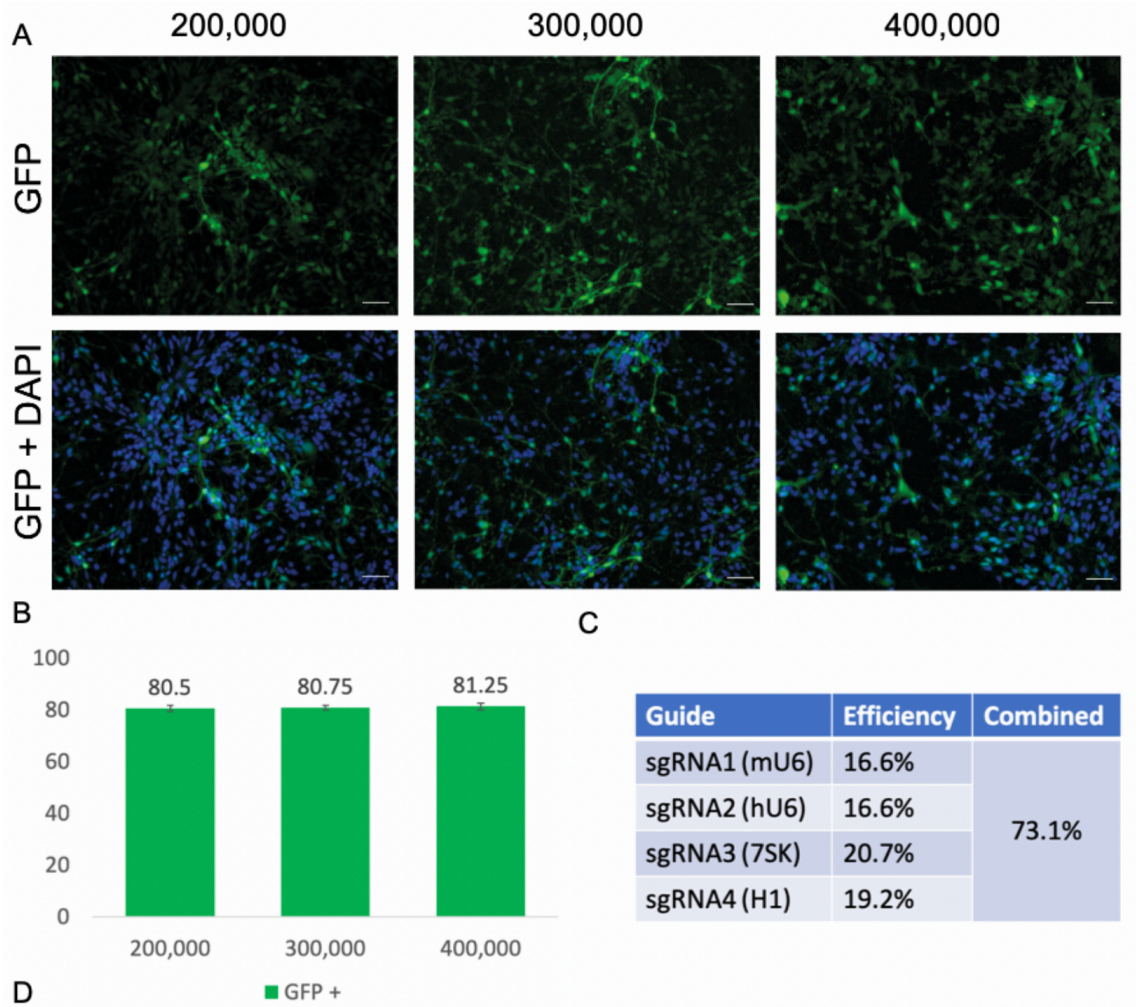
Figure 5.2 Generation of a lentiviral construct targeting AKT3.
(A) General linear structure of the lentiviral construct spanning the region containing the 4 sgRNAs, their promoters and the hUbC-Cas9-T2A-GFP, (B) AKT3 gene structure and targeted sequence depicting the sections where each sgRNA will cut, (C) sgRNA scores, sequences and PAM being cut by the Cas9. Example of DNA electrophoresis gel showing the full construct of the Cas9 targeting AKT3 compared to an empty Cas9 construct by enzymatic restriction (D) and PCR amplification (E) of the region across the sgRNAs and the Cas9-eGFP.

(II) AKT3 gene editing in hESC-derived NPCs

Control hESC cells were differentiated to cortical NPCs in order to determine the adequate number of lentiviral particles necessary for optimal transduction. At day 18 of differentiation NPCs were dissociated and plated at 200,000 per well of a 24 well plate. NPCs were transduced the following day with increasing concentrations of lentiviral particles. After the infection, the cells were grown for three days to allow a robust expression of GFP before being fixed and quantified for GFP expression. The three conditions yielded similar results, with ~80% of the cells expressing GFP (**Figure 5.3 A**). In parallel, a contiguous culture treated in the same conditions was used to extract the genomic DNA of the whole population. The sequence containing the targeted region was amplified by PCR and sequenced in order to quantify the efficiency of each sgRNA using the online tool TIDE (https://tide.deskgen.com). TIDE uses a decomposition algorithm to quantify the presence of InDels in a pool of edited and non-edited cells (Brinkman *et al.*, 2014). This allows the quantification of the efficiency of each sgRNA in a mixed population of DNA fragments formed by edited and non-edited sequences.

The analysis revealed that the cultures exposed to three increasing concentrations of lentiviral particles achieved a similar transduction efficiency of around 80% with no statistical difference (**Figure 5.3 B**). InDels were identified proximal to the predicted cutting sites targeted by all four sgRNAs at a frequency between 16 and 20%. Thus, assuming that there were no multiple cuts within one allele, 73.1% of the total sequences contained InDels in the *AKT3* coding sequence (**Figure 5.3 C**).

Before delivering the lentiviral construct to CYFIP1tg NPCs to manipulate *AKT3*, the potential top off-targets for each sgRNA were analysed. From the 5 highest-scoring off-targets, two are located in coding sequences, thus more likely to be causing unwanted phenotypic effects. Sequencing of PCR amplicons spanning these regions did not detect InDels, indicating that none of these potential off-targets were cut by the Cas9 (**Figure 5.3 D**).



Guide	Efficiency	Combined
sgRNA1 (mU6)	16.6%	73.1%
sgRNA2 (hU6)	16.6%	
sgRNA3 (7SK)	20.7%	
sgRNA4 (H1)	19.2%	

D

Guide	Sequence	Score	Mismatches	Gene name	Rank	InDels
Guide #2	GCGCCACAGAGAAGTTGTTGAGG	0.8	4MMs [1:2:8:11]	AKT1	1	NO
Guide #3	CGCCCTCATGAACTTTTCAGTAG	0.8	4MMs [1:2:9:10]	BORCS5	2	NO
Guide #4	GAGGAGTTGGCGAGGGGATACAG	0.7	4MMs [3:4:10:11]	COL7A1	3	NO
Guide #1	TAAGCTAAATGCACATATTGCAG	0.6	3MMs [5:11:17]	SNRPF	4	NO
Guide #3	CCTACTCAACAATTTTTCAGGAG	0.6	4MMs [1:3:4:13]	CCNG2	5	NO

Figure 5.3 Assessment of infectivity and editing efficiency of the lentiviral vector targeting AKT3.
(A) GFP expression in NPC cultures five days after transduction with three number of lentiviral particles tested, (B) GFP quantification by immunostaining of each of the three lentiviral densities tested, (C) individual and combined percentages of sequences edited by the sgRNAs targeting AKT3. (D) Table summarising the top off-targets of the sgRNAs against AKT3. All scale bars = 50 µm.

5.2.3 Disruption of AKT3 in CYFIPtg NPCs rescued detailed neurogenesis

To provide direct evidence that AKT3 signalling mediates CYFIP1 regulated cortical neurogenesis, I performed a viral-based AKT3 (loss-of-function) genome editing in CYFIP1tg hESC-derived NPCs, since my RNAseq analysis and Western blot indicated an increased AKT signalling in these cells. If our hypothesis proves to be correct, we would observe either a full or partial rescue of the delayed neuronal differentiation in CYFIP1tg NPCs. Lentiviruses expressing Cas9 and *AKT3*-targeting sgRNAs as described above, were delivered to the parental control and CYFIP1tg NPCs at day 18 of cortical differentiation. The number of $PAX6^+$ neural progenitors and $TBR1^+$ neurons were analysed at days 35 and 45 of differentiation. The presence of these two markers was quantified in the GFP^+ and GFP^- population of the control and CYFIP1tg cultures by double immunostaining with an anti-GFP antibody, respectively. At day 35, over 60% of the cells were GFP^+ in the control cultures, while CYFIP1tg cultures contained 67%. At day 45, the proportion of GFP^+ cells was found to decrease to 48% and 51%, respectively. This could be due to downregulation of the expression of GFP in the cells, or due to the expansion of GFP^- cells. The non-viral transduced GFP^- cell population served as an internal negative control to the GFP^+ cell population that were enriched with cells carrying disrupted AKT3.

In keeping with previous findings in our laboratory, at day 35, the non-infected GFP- control population in the CYFIP1tg cultures contained more $PAX6^+$ NPCs and less $TBR1^+$ neurons than those in the isogenic control cultures (**Figure 5.4 A-D**). A similar trend was also observed at day 45, where the number of $PAX6^+$ NPCs were significantly higher in the GFP- population of the CYFIP1tg cultures than those in the control cultures where few NPCs remained at this differentiation stage (**Figure 5.4 E, F**).

However, for both time points, the number of $PAX6^+$ NPCs in the GFP^+ population of the CYFIP1tg cultures was significantly decreased compared to those in the non-transduced cell population. Importantly, the number of CYFIP1tg GFP^+PAX6^+ cells dropped to the baseline, i.e. the level observed in the non-transduced parental cultures. I did not detect significant differences in the number of $PAX6^+$ and $TBR1^+$ cells between the GFP^+ and GFP^- population in the control cells, suggesting that *AKT3* manipulation did not influence NPC proliferation and/or differentiation of the control cells (see discussion).

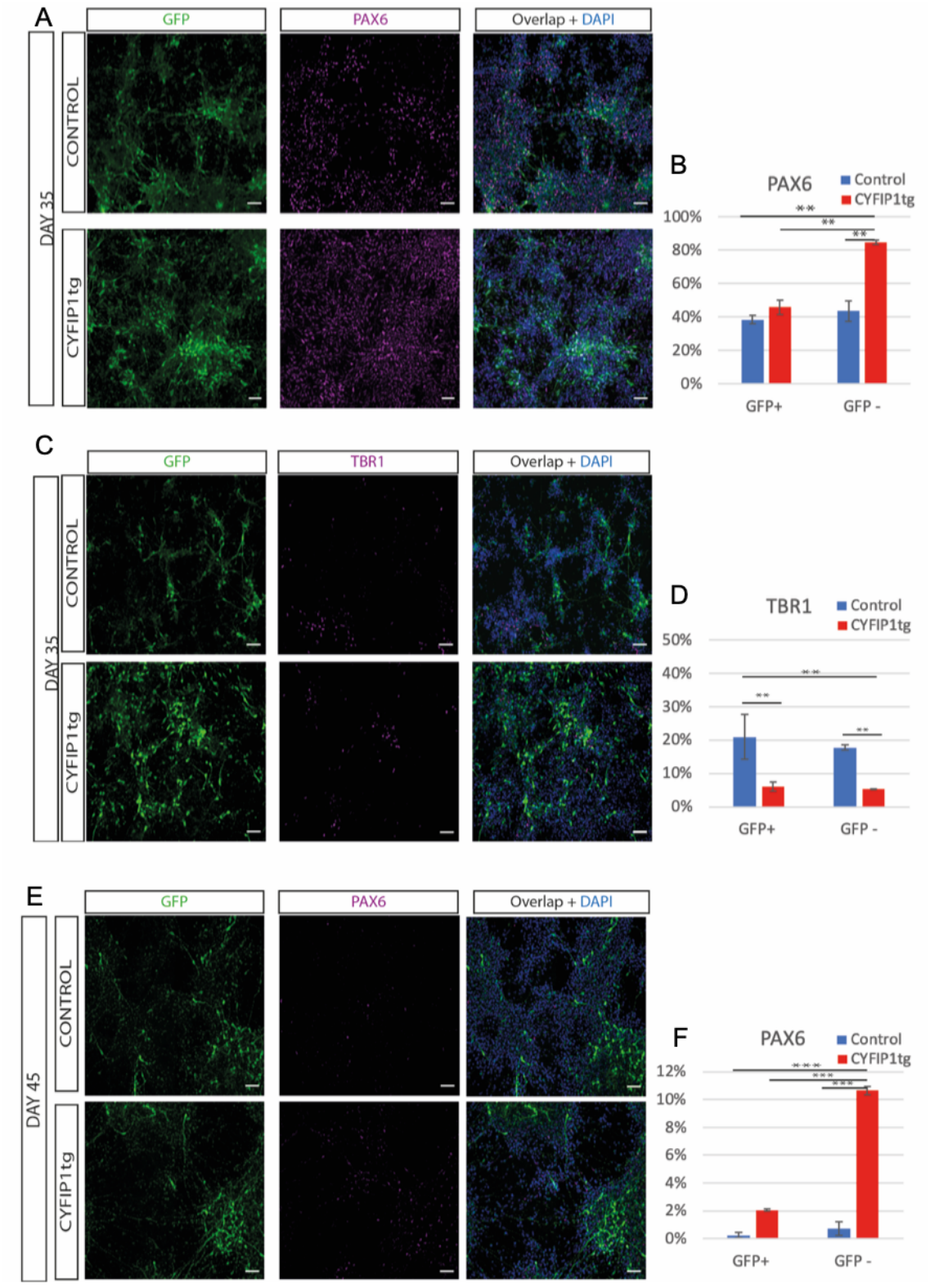


Figure 5.4 Lentiviral-based CRISPR/Cas9 targeting of AKT3 reduces PAX6 NPCs.
(A, B) Immunofluorescence and quantification of PAX6$^+$ cortical progenitor cells (magenta) overlapping with GFP (green) at day 35 and TBR1$^+$ neurons (C, D) for the same stage of differentiation. (E, F) Immunofluorescence and quantification of PAX6$^+$ cortical progenitor cells (magenta) overlapping with GFP (green) at day 45. Data were compared by one-way ANOVA followed by Tukey post-hoc test ($p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).*

To ascertain that the observed phenotypic rescue in CYFIP1tg cells was the consequence of *AKT3* manipulation, I examined the total and phosphorylated forms of AKT3 and GSK3β through western blot analysis (**Figure 5.5 A**). In the CYFIP1tg cultures, the level of total AKT3 was reduced at both day 35 and day 45. However, only day 45 data value reached statistical significance due to large variations of individual experiments (**Figure 5.5 B**). Conversely, the level of phosphorylated AKT3 (S473) was found to be significantly reduced in both CYFIP1tg and control cells following AKT3 manipulation at day 45 (**Figure 5.5 C**). For GSK3β, no difference was found at total protein level between the *AKT3* manipulated, and control condition (non-transduced) in both CYFIP1tg and parental control cells for both analysed stages (**Figure 5.5 D**). However, differences in the phosphorylated form were identified. At day 35, phosphorylated GSK3β (S9) was found to be increased in non-*AKT3* manipulated CYFIP1tg cultures compared to the control cells at the same stage of differentiation (**Figure 5.5 E**). *AKT3* targeting led to a reduction in the fraction of phosphorylated GSK3β in the CYFIP1tg comparatively to the controls at days 35 and 45 of differentiation (**Figure 5.5 E**).

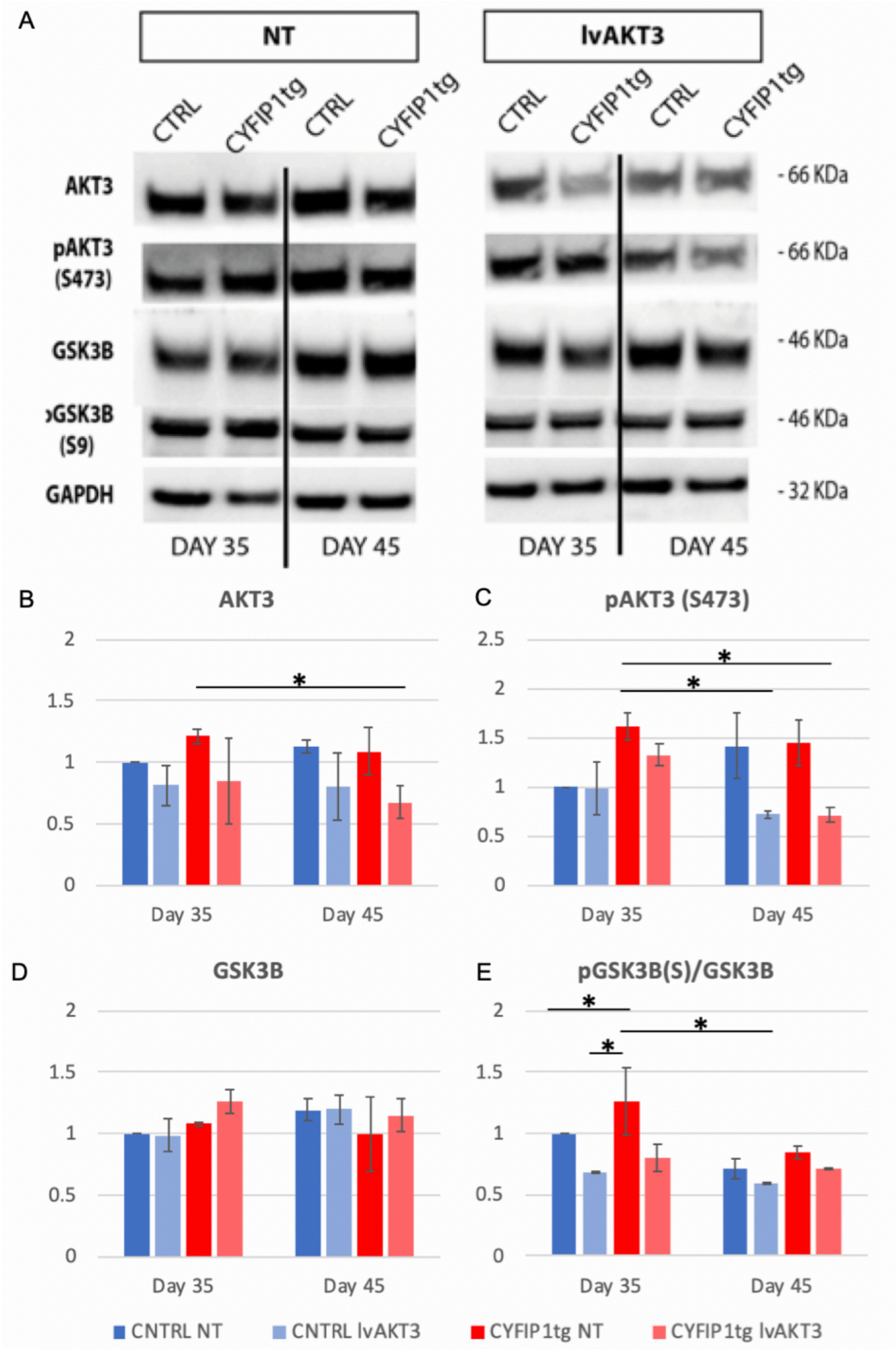


Figure 5.5 Western blot quantification of cells treated or untreated with the lentiviral CRISPR/Cas9 against AKT3.

(A) Western blots for AKT3, phosphorylated AKT3 (S473), total and phosphorylated (S9) GSK3β, and GAPDH in CYFIP1tg and controls at days 35 and 45 of differentiation with or without the presence of a lentiviral construct against AKT3.Bar graphs representing the quantification of AKT3 (B), p-AKT (S473) over total AKT3 (C), total GSK3B (D) and p-GSK3β (S9) over total GSK3β (E). NT: non-transduced. Data was compared by one-way ANOVA followed by Tukey post-hoc test ($p < 0.05$).*

5.3 Discussion

Initial phenotypic data generated by Dr Tamburini (unpublished) revealed that changes in CYFIP1 expression caused increase and decrease in NCAD and βCAT levels in both CYFIP1tg and CYFIP1ko respectively. Increased NCAD was correlated with increased levels of the active form of βCAT in CYFIP1tg cells. However, the connection between these findings was missing. The data presented in this chapter was able to characterise a possible mechanism linking the correlation of these two observations in CYFIP1tg neural cells. Firstly, the ratio of active AKT3 (p-S473)/AKT3 appeared to be higher in CYFIP1tg at day 20 of differentiation, and significantly higher at day 35. These were accompanied by a lower expression of GSK3B in CYFIP1tg throughout the process of differentiation. At the same time, the proportion of inactive GSK3B was comparable between control and CYFIP1tg. Because GSK3β phosphorylates βCAT at the N-terminus (S33/37) priming it for degradation, the excessive inactivation of GSK3β combined with a highly active form of AKT3 could explain the increased form of active βCAT (p-S552).

The development of the gene-editing tool CRISPR/Cas9 has become a true breakthrough in the field of cell biology and genome editing. This technique has been rapidly developed and improved for the use in mammalian cells. First approaches using this tool involved the use of separate plasmids coding for sgRNAs and Cas9 in order to target and edit a particular sequence, which led to a small proportion of successfully edited sequences. The success of this method is dependent on the co-expression of sgRNAs and Cas9 in the same cell. It has been shown that the use of multiple sgRNAs can improve genomic editing efficiency. Taking advantage of type IIS restriction enzymes that cut outside their recognition sequence, such as the ones used in this chapter, allows the generation of unique overhangs that allow the assembly of those complex constructs through golden gate reaction.

Kabadi and colleagues developed a multiplex lentiviral-based vector combining CRISPR/Cas9 with up to 4 sgRNAs (Kabadi *et al.*, 2014). Using this technique, the authors reported a favourable increase in the percentage of InDels present when the sgRNAs were introduced in a single construct compared to introducing them separately. This method was successfully used in this chapter to manipulate the sequence of *AKT3*. TIDE analysis proved that the designed sgRNAs were able to introduce InDels in the desired sequence without introducing unwanted modifications in other genes due to off-targets. Each guide had an individual efficiency of up to 20% and combined they manipulated more than 75% of the sequences. However, this may not be an accurate value because some of the edited sequences could have been edited by more than one guide. Nonetheless, designing overlapping sgRNAs for the same sequence make it unlikely that,

once one guide has cut and generated InDels in the sequence, the other guide can't bind to the newly edited sequence and cut again.

The strategy presented in this chapter presents multiple advantages: 1) The use of a lentiviral CRISPR/Cas9 allows the genetic manipulation of the cells at one specific developmental stage without altering the previous differentiation process, thus reducing variations between the edited and control cells (that are generally in independent cell lines). 2) The use of a vesicular stomatitis virus G (VSVG) pseudotyped viral envelope, which binds the virus to the cellular LDL receptor, allows the efficient transduction of the culture using a minimal number of viral particles. 3) The combination of multiple sgRNAS increasing the efficiency at generating InDels in mammalian cells generated by the Cas9.

The generated lentiviral particles proved to be efficient at infecting and editing the NPCs of both control and CYFIP1tg. The explored NPC marker PAX6 differed in number amongst infected and non-infected cells, being lower in the former. No differences were observed in the control cells which could indicate that altered signalling through AKT3 could be specifically playing a role in cells with an excess of CYFIP1 expression, perhaps due to CYFIP1-associated role at regulating the translation of mRNAs with FMRP. Moreover, genetic manipulation of *AKT3* did not have an effect on the number of TBR1-expressing cells. Altogether, this could indicate that the mechanism of AKT3 is restricted to promoting the self-proliferation of NPCs in CYFIP1tg. Further research needs to be developed in order to confirm this mechanism. A first approach would be increasing the expression of AKT3 in control cells and observe if this has an effect on the proliferation of $PAX6^{+}$ cells

Unfortunately, the reduced levels of GSK3β were not able to be replicated in the last part of the analysis, where the effects of lvAKT3 were assessed by Western blot. This could be explained due to biological differences existing in differentiating cells. Nonetheless, the excess of inhibited GSK3β was still observed at day 35 in CYFIP1tg. This effect was remarkably restored after applying the lentivirus targeting *AKT3*. Although no significant changes are observed at the protein level, there is a tendency to reduce the expression at the protein level, which results in the observed reduction in the phosphorylation of GSK3β. This could be due to the effective targeting of the PD domain. Even if there is not a complete KO of the gene, small mutations or frame-shift mutations are introduced in the sequence, which could alter the structure of the PD domain required for the activation of AKT; or potential structural changes in the kinase domain making it impossible for AKT3 to phosphorylate other proteins such as GSK3β.

The genetic manipulation of *AKT3* aiming to reduce the expression of the protein with a multiplex lentiviral CRISPR/Cas9, was able to reduce the total number of proliferating $PAX6^+$ cells. In this context, the disruption of AKT signalling was aimed to increase the effect of GSK3β in order to inactivate βCAT. In humans, patients with abnormal activity of AKT3 cause alterations in the cortical development, which have psychiatric repercussions. Studies in mice have associated the KO of *Akt3* to the development of psychiatric phenotypes. In this context, chronic administration of lithium, a known treatment for SZ, was able to restore the decreased levels of GSK3β (p-Ser9) rescuing the depressive-like behaviour of the mice (Bergeron *et al.*, 2017).

The PI3K-AKT pathway has been previously reported to play a role in FXS, SZ and ASD (Crespi, Stead and Elliot, 2010; Matsuda *et al.*, 2019; Telias, 2019). Moreover, pathway analysis of SZ and ASD GWAS-associated genes, have highlighted this pathway as a potential causal mechanism, amongst the multiple dysregulated processes. Others have proposed the PI3K-AKT pathway for pharmacological intervention in SZ and ASD (Enriquez-Barreto and Morales, 2016; Gross *et al.*, 2019). Protein expression of NCAD, AKT3 and GSK3β mRNAs are regulated at the ribosomal level by the complex formed by CYFIP1-FMRP-eI4F (Luo *et al.*, 2010; Darnell *et al.*, 2011; La Fata *et al.*, 2014).

Interestingly, the transcriptionally active fraction of βCAT can also be regulated by the actin cytoskeleton independently from WNT signalling. Alterations affecting the WAVE complex, of which CYFIP1 is part of, cause accumulation of F-Actin in the nucleus which in turn, promote the translocation of βCAT to the nucleus (Yamazaki *et al.*, 2016). A recent study carried out by Moraes and colleagues reported that ABI3, another component of the WAVE complex, is regulated by PI3K-AKT signalling. In this study, reduced phosphorylation of ABI3 (p-Ser342) promoting its localisation in the WAVE complex, which caused the cells to significantly increase their βCAT expression (Moraes, Zanchin and Cerutti, 2017).

In summary, the results presented in this chapter indicate that the effects of AKT signalling are reflected in the transcriptomic data of CYFIP1 engineered cells. Moreover, these results provide novel insight into the mechanism, causing the imbalance in AKT signalling in CYFIP1tg cells. Finally, the genetic manipulation of AKT3 provides independent confirmation that this kinase plays an important role in the development of the phenotype associated with CYFIP1tg cells by inactivating GSK3β and, ultimately, reinforcing the activity of BCAT in the maintenance of $PAX6^+$ NPCs.

Chapter 6

General discussion

6.1 Summary of findings

In this thesis, I used transcriptome analysis to investigate the changes induced by altered levels of CYFIP1 dosage during cortical differentiation of hPSCs. Changes in gene expression caused by excess or depletion of CYFIP1 were studied in CYFIP1tg and CYFIP1ko, respectively. Both gain- and loss-of-function of CYFIP1 affected the expression of thousands of genes, including numerous transcripts known to be involved in cortical development. Interestingly, these transcripts were not only limited to protein-coding genes but extended to lncRNAs. This novel finding may suggest a role of non-coding sequences in the development of neurodevelopmental disorders and the need for further study in this area.

Most importantly, the analysis of the biological pathways revealed that similar cellular processes were affected in CYFIP1tg and CYFIP1ko, and these had a tendency to be regulated in opposing directions. These findings were further explored in cortical derivatives of iPSCs derived from patients harbouring a deletion in the 15q11.2 BP1-BP2 region. The similarity between the cells carrying the CNV and the CYFIP1 engineered cells provides strong support to the causal role of CYFIP1 in the 15q11.2 CNV. An in-depth analysis of CYFIP1 regulated genes forming altered KEGG pathways identified *ATK3*, a GWAS risk gene for SZ and whose mRNA translation is regulated by the FMRP-CYFIP1-eIF4E complex, as a top candidate gene causing the phenotypical deficits observed in CYFIP1tg and CYFIP1ko. Follow-up experiments investigating the effect of AKT3 signalling in CYFIP1tg indicated that this pathway could be playing a role in the maintenance of $PAX6^+$ NPCs through excessive inactivation of GSK3B mediate by phosphorylation through AKT3.

Using gene ontology enrichment analysis, genes with altered expression were grouped in sets of genes that shared a similar function. These gene sets captured the altered processes affected by CYFIP1 dosage change and recapitulated previously known effects of these genes in cortical development as well as highlight novel associated functions such as mitochondrial function or cholesterol metabolism. Further analysis in MAGMA of functional gene sets characterised a signal for genetic enrichment with common variants for neuropsychiatric disorders such as SZ and ASD. I found for the first time that the genetic association with SZ in CYFIP1-regulated genes was restricted to FMRP target genes. The identification of a reduced group of genes associated with the development of SZ and a cellular pathway contributing to the cellular phenotype, could help to elucidate the neurodevelopmental origin of SZ in a CYFIP1 altered context.

6.2 Novel identified mechanisms in cells with altered expression of CYFIP1

Pathway enrichment analysis of CYFIP1tg and CYFIP1ko provided evidence that increased and decreased levels of CYFIP1 affect similar pathways in opposite directions. This analysis captured previously known aspects associated with an altered dosage of CYFIP1. However, it identified two newly-associated functions to CYFIP1-regulated genes: mitochondrial and cholesterol metabolism.

6.2.1 Mitochondrial alterations

The abundance of mitochondrial genes altered in both CYFIP1tg and CYFIP1ko indicate a role of this organelle in the phenotype observed. Indeed, mitochondria play an essential role during cortical development and in mature neurons. Alterations in the biology of the ATP-generating organelle have previously been associated with multiple neurodevelopmental disorders (see Chapter 4). A previous study identified that KD of CYFIP1 affected genes enriched for epilepsy, and amongst these, mitochondrial function appeared to be affected even though it did not reach statistical significance (Nebel *et al.*, 2016). The data presented in this thesis show, for the first time, a direct relationship between CYFIP1-regulated genes are significantly associated with mitochondrial function. Moreover, enrichment of mitochondrial genes was found in cells carrying an excess of CYFIP1 as well as in 15q11.2del carrying cells.

It is believed that an imbalance between excitation and inhibition are behind the aetiology of ASD and SZ (Marín, 2012; Canitano and Pallagrosi, 2017). There is evidence that mitochondrial defects have a repercussion in both excitatory and inhibitory neurons. Genetic or pharmacological disruption of the oxidative phosphorylation in MGE-derived interneurons impaired the process of tangential migration (Lin-Hendel *et al.*, 2016). Fast-spiking parvalbumin^{+} interneurons can be severely affected by alterations in the mitochondria. As shown by Inan and colleagues, deletion of the cytochrome oxidase 10 (*Cox10)* in mice caused the impaired assembly of the cytochrome oxidase complex causing electrophysiological abnormalities in these cells. Moreover, these mice displayed social and behavioural abnormalities, similar to those observed in SZ and ASD models (Inan *et al.*, 2016).

Defects in the cytochrome oxidase subunits affecting dopaminergic neurons have also been observed in post-mortem brains of subjects with SZ (Rice *et al.*, 2014). Alterations

in the mitochondria can also be observed in pyramidal neurons. Further studies have confirmed that layer 5/6 pyramidal neurons contain a decreased number of mitochondria (Roberts *et al.*, 2015). Moreover, alterations of mitochondrial genes and proteins have been identified in the prefrontal cortex of schizophrenic patients (Prabakaran *et al.*, 2004). However, these alterations could be secondary to other alterations presented by ASD and SZ patients, such as immune dysfunction or aberrant calcium homeostasis (Frye and Rossignol, 2011).

Current research carried out in our laboratory, validated the findings observed at the transcriptomic level. Preliminary analyses indicate that CYFIP1tg and CYFIP1ko NPCs have, respectively, significantly decreased and increased numbers of mitochondria (Tamburini *et al.,* Unpublished). The transcriptomic data pointed out differences in the different elements of the electron transmission chain. These changes are likely to be reflected in the mitochondrial membrane potential. Moreover, changes in the electron transmission chain are likely to be causing repercussion in the number of ROS generated by the mitochondria. ROS can act as signalling molecules regulating the process of neurogenesis in the developing neocortex. Study of these could elucidate the effect over neurogenesis in a CYFIP1 altered context.

6.2.2 Abnormal cholesterol metabolism

Gene ontology analysis revealed that both CYFIP1tg and CYFIP1ko present alterations in cholesterol metabolism. Cholesterol is particularly important in brain development, and it is critical for the formation of the synapses (Kelly L Wormwood, 2013). It is believed that alterations in cholesterol synthesis and metabolism can play a role in both SZ and ASD. However, no direct link between these has been identified to date. A great variety of syndromes associated with ASD, including Smith-Lemli-Opitz syndrome and FXS, present alterations in cholesterol (Lee and Tierney, 2011; Berry-Kravis *et al.*, 2015). Patients with SZ present higher levels of serum lipids, which have been typically attributed to pharmacological treatment and lifestyle factors (Paton *et al.*, 2004; Newcomer, 2005). Despite this, altered lipid profiles have been found in drug-naive patients and first-episode cases (Thakore, 2004; Schwarz *et al.*, 2008; Verma *et al.*, 2009; Kaddurah-Daouk *et al.*, 2012), which indicates a pre-existing alteration in the lipidome of these patients.

During embryonic development, the brain morphogen sonic hedgehog involved in the patterning of the telencephalon as well as regulating NPC division (Palma and Ruiz i Altaba, 2004), is modified by the covalent addition of palmitate and cholesterol before being secreted (Pepinsky *et al.*, 1998; Cooper *et al.*, 2003). Disruption of this signalling

can have an effect on the cell types generated and on gross brain structure abnormalities including a reduction in the corpus callosum, one of the most common abnormalities found in ASD neuroimaging studies (Stanfield *et al.*, 2008; Hardan *et al.*, 2009).

Our laboratory is exploring the effects of changes in CYFIP1 on the different cholesterol precursors and metabolites. Preliminary mass spectrometry analysis identified decreased concentrations of desmosterol and 7-dehydrocholesterol; and presence of 8(14)-dehydrocholesterol in CYFIP1tg neurons compared to the isogenic control. Conversely, these cells produced 27-hydroxycholesterol (27-HC), a metabolite not present in the control. This molecule is known to be an activator of the ERα (Umetani *et al.*, 2007). Previous models of cell growth in breast cancer have shown a dependency on the signalling through 27-HC and ERα (Nelson *et al.*, 2013; Wu *et al.*, 2013). Intriguingly, the analysis conducted in Chapter 4 highlighted the dysregulated gene *TCF20*, which encodes for a transcription factor that modulates the activity of ERα. Wther this molecule has an impact on the NPCs growth remains to be investigated.

The same mass spectrometry analysis found the presence of 24(S),25-epoxycholesterol (24(S),25-EC) in 15q11.2 deletion-derived neurons. 24(S),25-EC is the most abundant oxysterol in murine foetal brain development and can be synthethised parallely to cholesterol (Gill, Chow and Brown, 2008). 24(S),25-EC is known to be the main ligand to bind to the liver X receptors (LXRa and LXRb) and its hypothethised that plays a role in regulating the biosynthesis of cholesterol (Griffiths *et al.*, 2016). Research indicates that 24(S),25-EC promotes dopaminergic neurogenesis in ESCs-derived neural cultures and midbrain progenitor cells, mediated by the activation of the LXRs, indicating an activity of this molecule in the regulation of neuronal maturation (Sacchetti *et al.*, 2009; Theofilopoulos *et al.*, 2013).

Current research in our laboratory is aiming to identify the role of these two sterol precursors in cortical neural development. Preliminary data showed that increasing concentrations of 24(S),25-EC were associated to the maturation of the neural NPCs and correlated with increased numbers of $CTIP2^{+}$ cells, a marker typical of cortical layer V.

6.3 Alterations in mRNA processing

Gene set enrichment analysis of CYFIP1-regulated genes showed how genes involved in mRNA processing were contributing to the altered processes in cells with altered expression of CYFIP1. (see chapter 4). Amongst the KEGG pathways analysed, the ribosome was found to be altered in both CYFIP1tg and CYFIP1ko cells with opposite effects on each cell type. These findings were also found in cells with a 15q11.2 del, where the

enrichment score took the same direction as the CYFIP1ko cells. GO analysis showed that processes involved in mRNA splicing and ribosomal translation were found to be affected amongst the CYFIP1-regulated genes.

The process of mammalian cortical development involves a precise sequence of events. Therefore, tightly controlled spatiotemporal regulation, as well as the subcellular localisation of protein synthesis will play an essential role in the development of the neocortex (Kraushar *et al.*, 2014). In polarised cells such as the RGCs, subcellular transportation of mRNAs by FMRP to local translation sites is required for the adequate neurogenesis (Pilaz *et al.*, 2016; Pilaz and Silver, 2017). CYFIP1 has a known role in regulating the translation of FMRP targets (see chapter 1). Association of common variants resulting from GWAS and the PSC-derived transcriptomic data presented in this thesis identified FMRP-regulated mRNAs as the genes driving the association with SZ in cells with altered CYFIP1 expression. Therefore, these results have identified a subset of genes that could be directly involved in the development of the psychiatric phenotype. Moreover, they provide further evidence for the role for CYFIP1 in the development of neuropsychiatric disorders when altered in CNVs affecting the 15q11.2. However, how alterations in protein synthesis affect neural development in a CYFIP1 dysregulated context have not been studied yet.

A mass spectrometry analysis of NPCs derived from SZ patients indicated increased levels of ribosomal proteins and elongation factors required for protein synthesis (Topol *et al.*, 2015). The same study concluded that this effect is limited to NPCs and is not present in neurons. Alterations in mechanisms involved in the regulation of protein synthesis have been identified in neurodevelopmental disorders. The mRNA binding protein Hu antigen R (HuR) regulates the temporal synthesis of the Forkhead-box domain (Fox) family proteins (Kraushar *et al.*, 2014; Popovitchenko *et al.*, 2016). The Fox family proteins have a characterised role in the development of the telencephalon and alterations of their expression have been involved with ASD (Lai *et al.*, 2001; Bacon and Rappold, 2012; Araujo *et al.*, 2015). This finding is particularly interesting in an increased dosage of CYFIP1 context, as data presented in this thesis indicate a substantial increase of *FOXG1* at the mRNA level in CYFIP1tg NPCs.

Indeed, research has established a direct link between CYFIP1 and the positive regulation of protein synthesis. In a study carried out by Oguro-Ando (see Chapter 1), changes in CYFIP1 expression altered mTOR signalling, a pathway that regulates protein synthesis in the cell (Oguro-Ando *et al.*, 2015). Abekhoukh and colleagues further explored these findings. Inducing a KD of *Cyfip1* in mice primary neuron cultures, the authors identified that these effects were primarily driven by changes in affecting PI3K, a kinase involved in the activation of the AKT pathway, and resulted in reduced levels of

phosphorylated the ribosomal protein S6 (Abekhoukh *et al.*, 2017). The analysis generated in this thesis pointed out *AKT3* as one of the primary genes that could be driving the effects observed in cells with an altered dosage of CYFIP1. We hypothesised that altered AKT3 function could affect the numbers of $PAX6^{+}$ NPC population through modulation of GSK3B and BCAT activity. However, considering the abovementioned literature, it is possible that AKT3 could be contributing to the observed phenotype through a parallel mechanism. Therefore, the effects of *AKT3* manipulation should be further investigated in order to understand their role in protein synthesis in a CYFIP1 altered context.

6.4 Further directions

RNAseq is a very versatile technique. Besides the application in differential gene expression, sequencing data generated in this thesis can be used for differential isoform expression or *de novo* transcript assembly. Around 95% of the multi-exon human genes suffer alternative splicing (Pan *et al.*, 2008; Wang *et al.*, 2008). This is particularly interesting in the brain, where splicing takes place in multiple developmental stages and is required for cell-fate decisions, neuronal migration or synaptogenesis (Su, D and Tarn, 2018). An example is the switch of PTBP1 in NPCs to PTBP2 required for the differentiation to neuronal cells (Boutz *et al.*, 2007). Changes in alternative splicing can affect normal brain development causing neuropsychiatric disorders. There is evidence that abnormal splicing of *RBFOX1* is linked to the development of ASD (Conboy, 2017). Because increased CYIP1 expression in NPCs affects the spliceosomal complex (see Chapter 4), the study of the different splicing variants could give further insight into the mechanism of action of CYFIP1.

We can generate multiple neuronal types in the laboratory, which can be useful to elucidate the complex biology of neuropsychiatric disorders where multiple cell types are affected. A study carried out by Skene and colleagues identified specific brain cell types affected by SZ. In this study, the analysis of common variants resulting from SZ GWAS identified pyramidal neurons, medium spiny neurons and specific interneurons as the class of cells affected by this disorder. Each cell type had a unique characteristic gene set affected (Skene *et al.*, 2018). Nonetheless, there were shared gene sets amongst the cells, which included synaptic genes and FMRP-regulated genes, which captured an enrichment for common variants associated with SZ (Skene *et al.*, 2018). This indicated that not only cortical cells could be involved in the development of SZ, and therefore, it is justifiable to study the biology of this disorder in MSNs and interneurons.

In order to fully understand how the mutations affecting CYFIP1 expression can promote the development of an aberrant phenotype, further study of the 15q11.2 deletion-carrying iPSCs should be carried out. Current transcriptomic analysis of NPCs and neurons carrying the 15q11.2 microdeletion have highlighted similar pathways to those identified in exclusively affected *CYFIP1*. This indicates that the mechanism for which the deletion causes the observed phenotypes is due to CYFIP1. Therefore, experiments manipulating AKT3 should be carried out in 15q11.2 del to validate the results obtained in the CYFIP1 mutant cells.

The current data obtained from the lentiviral CRISPR/Cas9 construct targeting *AKT3* is in agreement with previous research carried out in our laboratory, where the addition of an AKT inhibitor in differentiating cells was able to partly reduce the number of PAX6+ cells in CYFIP1tg to levels similar of its control cells. Nonetheless, these experiments require further development. The number of cells expressing TBR1 were not changed upon the lentiviral infection. This could indicate that the time window where the treatment was applied was too late to observe an effect in this cell population. To confirm this, other postmitotic neuronal populations, such as CTIP2 or CUX1 neurons, which appear later in cortical development, should be evaluated. Another important aspect to consider is a potential susceptibility of the NPC population over a lentiviral infection. Therefore, repeating the presented experiments with the same lentiviral approach without any guides to direct the Cas9 should be tested.

Finally, multiple mutations have been associated with the development of neuropsychiatric disorders. In our laboratory, we have generated multiple datasets from *in vitro* models of neurodevelopmental disorders of these conditions. These include loss-of-function of Set binding protein 1 (SETBP1), mutation that has been associated to the development of ASD, ID and multiple congenital malformations (O'Roak *et al.*, 2012; Coe *et al.*, 2014); and iPSCs carrying a 22q11.2 deletion, a mutation associated with increased risk of SZ (Murphy, Jones and Owen, 1999; Stefansson *et al.*, 2008). Comparing the transcriptional signature of such could help to elucidate a common mechanism across the different models.

References

Abdelmoity, A. T. *et al.* (2012) '15q11.2 Proximal Imbalances Associated With a Diverse Array of Neuropsychiatric Disorders and Mild Dysmorphic Features', *Journal of Developmental & Behavioral Pediatrics*, 33(7), pp. 570–576. doi: 10.1097/DBP.0b013e31826052ae.

Abekhoukh, S. *et al.* (2017) 'New insights into the regulatory function of CYFIP1 in the context of WAVE- and FMRP-containing complexes', *Disease Models & Mechanisms*, 10(4), pp. 463–474. doi: 10.1242/dmm.025809.

Abrahams, B. S. *et al.* (2013) 'SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs)', *Molecular Autism*, 4(1), p. 36. doi: 10.1186/2040-2392-4-36.

Abranches, E. *et al.* (2009) 'Neural Differentiation of Embryonic Stem Cells In Vitro: A Road Map to Neurogenesis in the Embryo', *PLoS ONE*. Edited by G. Parise, 4(7), p. e6286. doi: 10.1371/journal.pone.0006286.

Afshari, P. *et al.* (2015) 'Characterization of a Novel Mutation in SLC1A1 Associated with Schizophrenia', *Molecular Neuropsychiatry*, 1(3), pp. 125–144. doi: 10.1159/000433599.

Aguayo, A. *et al.* (2013) 'De novo germline and postzygotic mutations in AKT3, PIK3R2 and PIK3CA cause a spectrum of related megalencephaly syndromes', *Nat Genet*, 44(8), pp. 934–940. doi: 10.1038/ng.2331.De.

Alcantara, D. *et al.* (2017) 'Mutations of AKT3 are associated with a wide spectrum of developmental disorders including extreme megalencephaly', *Brain*, 140(10), pp. 2610–2622. doi: 10.1093/brain/awx203.

Alessi, D. R. and Cohen, P. (1998) 'Mechanism of activation and function of protein kinase B.', *Current opinion in genetics & development*, 8(1), pp. 55–62. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9529606.

Anderson, B. R. *et al.* (2016) 'Identification of consensus binding sites clarifies FMRP binding determinants', *Nucleic Acids Research*, 44(14), pp. 6649–6659. doi: 10.1093/nar/gkw593.

Araujo, D. J. *et al.* (2015) 'FoxP1 orchestration of ASD-relevant signaling pathways in the striatum.', *Genes & development*. Cold Spring Harbor Laboratory Press, 29(20), pp. 2081–96. doi: 10.1101/gad.267989.115.

Arber, C. *et al.* (2015) 'Activin A directs striatal projection neuron differentiation of human

pluripotent stem cells', *Development*, 142(7), pp. 1375–1386. doi: 10.1242/dev.117093.

Ariani, F. *et al.* (2008) 'FOXG1 Is Responsible for the Congenital Variant of Rett Syndrome', *The American Journal of Human Genetics*, 83(1), pp. 89–93. doi: 10.1016/j.ajhg.2008.05.015.

Artavanis-Tsakonas, S., Rand, M. D. and Lake, R. J. (1999) 'Notch Signaling: Cell Fate Control and Signal Integration in Development', *Science*. American Association for the Advancement of Science, 284(5415), pp. 770–776. doi: 10.1126/SCIENCE.284.5415.770.

Autism Genome Project Consortium *et al.* (2007) 'Mapping autism risk loci using genetic linkage and chromosomal rearrangements', *Nature Genetics*, 39(3), pp. 319–328. doi: 10.1038/ng1985.

Awadalla, P. *et al.* (2010) 'Direct measure of the de novo mutation rate in autism and schizophrenia cohorts', *American Journal of Human Genetics*, 87(3), pp. 316–324. doi: 10.1016/j.ajhg.2010.07.019.

Backman, M. *et al.* (2005) 'Effects of canonical Wnt signaling on dorso-ventral specification of the mouse telencephalon', *Developmental Biology*, 279(1), pp. 155–168. doi: 10.1016/j.ydbio.2004.12.010.

Bacon, C. and Rappold, G. A. (2012) 'The distinct and overlapping phenotypic spectra of FOXP1 and FOXP2 in cognitive disorders', *Human Genetics*. Springer-Verlag, 131(11), pp. 1687–1698. doi: 10.1007/s00439-012-1193-z.

Baek, S. T. *et al.* (2015) 'An AKT3-FOXG1-reelin network underlies defective migration in human focal malformations of cortical development', *Nature Medicine*. Nature Publishing Group, 21(12), pp. 1445–1454. doi: 10.1038/nm.3982.

Barde, I., Salmon, P. and Trono, D. (2010) 'Production and titration of lentiviral vectors', *Current Protocols in Neuroscience*, (SUPPL.53), pp. 1–23. doi: 10.1002/0471142301.ns0100s37.

Barry, G. *et al.* (2014) 'The long non-coding RNA Gomafu is acutely regulated in response to neuronal activation and involved in schizophrenia-associated alternative splicing', *Molecular Psychiatry*, 19(4), pp. 486–494. doi: 10.1038/mp.2013.45.

Bauman, M. L. (2010) 'Medical comorbidities in autism: Challenges to diagnosis and treatment', *Neurotherapeutics*, 7(3), pp. 320–327. doi: 10.1016/j.nurt.2010.06.001.

Le Belle, J. E. *et al.* (2011) 'Proliferative Neural Stem Cells Have High Endogenous ROS

Levels that Regulate Self-Renewal and Neurogenesis in a PI3K/Akt-Dependant Manner', *Cell Stem Cell*. Elsevier Inc., 8(1), pp. 59–71. doi: 10.1016/j.stem.2010.11.028.

Ben-Shachar, S. *et al.* (2009) 'Microdeletion 15q13.3: a locus with incomplete penetrance for autism, mental retardation, and psychiatric disorders.', *Journal of medical genetics*. BMJ Publishing Group Ltd, 46(6), pp. 382–8. doi: 10.1136/jmg.2008.064378.

Bergen, S. E. *et al.* (2014) 'Genetic modifiers and subtypes in schizophrenia: Investigations of age at onset, severity, sex and family history', *Schizophrenia Research*, 154(1–3), pp. 48–53. doi: 10.1016/j.schres.2014.01.030.

Bergeron, Y. *et al.* (2017) 'Genetic Deletion of Akt3 Induces an Endophenotype Reminiscent of Psychiatric Manifestations in Mice', *Frontiers in Molecular Neuroscience*, 10(April), pp. 1–12. doi: 10.3389/fnmol.2017.00102.

Bernal, A. and Arranz, L. (2018) 'Nestin-expressing progenitor cells: function, identity and therapeutic implications', *Cellular and Molecular Life Sciences*. Springer International Publishing, 75(12), pp. 2177–2195. doi: 10.1007/s00018-018-2794-z.

Bernard, D. *et al.* (2010) 'A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression', *EMBO Journal*, 29(18), pp. 3082–3093. doi: 10.1038/emboj.2010.199.

Berry-Kravis, E. *et al.* (2015) 'Cholesterol levels in Fragile X syndrome', *American Journal of Medical Genetics Part A*. John Wiley & Sons, Ltd, 167(2), pp. 379–384. doi: 10.1002/ajmg.a.36850.

Bijlsma, E. K. *et al.* (2009) 'Extending the phenotype of recurrent rearrangements of 16p11.2: Deletions in mentally retarded patients without autism and in normal individuals', *European Journal of Medical Genetics*, 52(2–3), pp. 77–87. doi: 10.1016/j.ejmg.2009.03.006.

Billups, B. and Forsythe, I. D. (2002) 'Presynaptic Mitochondrial Calcium Sequestration Influences Transmission at Mammalian Central Synapses', *The Journal of Neuroscience*, 22(14), pp. 5840–5847. doi: 10.1523/JNEUROSCI.22-14-05840.2002.

Bleuler, E. and Jung, C. (1908) 'Komplex und Krankheitsurschen bei Dementia praecox : Zentralblatt fur Psychiatries u', *Nervenheilkunde*, 3, pp. 220–227. Available at: https://ci.nii.ac.jp/naid/10017358121/ (Accessed: 21 June 2019).

Boutz, P. L. *et al.* (2007) 'A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons', *Genes & Development*, 21(13), pp. 1636–1652. doi: 10.1101/gad.1558107.

Brem, S. *et al.* (2014) 'The neurobiological link between OCD and ADHD', *ADHD Attention Deficit and Hyperactivity Disorders*, 6(3), pp. 175–202. doi: 10.1007/s12402-014-0146-x.

Brinkman, E. K. *et al.* (2014) 'Easy quantitative assessment of genome editing by sequence trace decomposition', *Nucleic Acids Research*, 42(22), pp. e168–e168. doi: 10.1093/nar/gku936.

Brown, V. *et al.* (2001) 'Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome.', *Cell*, 107(4), pp. 477–87. doi: 10.1016/s0092-8674(01)00568-2.

Brugha, T. S. *et al.* (2014) 'The epidemiology of autism spectrum disorders in adulthood', *Adolescents and Adults with Autism Spectrum Disorders*, pp. 299–314. doi: 10.1007/978-1-4939-0506-5_15.

Brunetti-Pierri, N. *et al.* (2008) 'Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities', *Nature Genetics*, 40(12), pp. 1466–1471. doi: 10.1038/ng.279.

Brzustowicz, L. M. *et al.* (2000) 'Location of a Major Susceptibility Locus for Familial Schizophrenia on Chromosome 1q21-q22', *Science*, 288(5466), pp. 678–682. doi: 10.1126/science.288.5466.678.

Buescher, A. V. S. *et al.* (2014) 'Costs of autism spectrum disorders in the United Kingdom and the United States', *JAMA Pediatrics*, 168(8), pp. 721–728. doi: 10.1001/jamapediatrics.2014.210.

Burnside, R. D. *et al.* (2011) 'Microdeletion/microduplication of proximal 15q11.2 between BP1 and BP2: a susceptibility region for neurological dysfunction including developmental and language delay', *Human Genetics*. Springer-Verlag, 130(4), pp. 517–528. doi: 10.1007/s00439-011-0970-4.

Bush, W. S. and Moore, J. H. (2012) 'Chapter 11: Genome-Wide Association Studies', *PLoS Computational Biology*, 8(12). doi: 10.1371/journal.pcbi.1002822.

Butler, M. G. *et al.* (2004) 'Behavioral Differences Among Subjects With Prader-Willi Syndrome and Type I or Type II Deletion and Maternal Disomy', *PEDIATRICS*, 113(3), pp. 565–573. doi: 10.1542/peds.113.3.565.

Butler, M. G. *et al.* (2008) 'Array comparative genomic hybridization (aCGH) analysis in Prader–Willi syndrome', *American Journal of Medical Genetics Part A*, 146A(7), pp. 854–860. doi: 10.1002/ajmg.a.32249.

Butler, M. G. (2017) ‘Clinical and genetic aspects of the 15q11.2 BP1-BP2 microdeletion disorder’, *Journal of Intellectual Disability Research*, 61(6), pp. 568–579. doi: 10.1111/jir.12382.

Cafferkey, M. *et al.* (2014) ‘Phenotypic features in patients with 15q11.2(BP1-BP2) deletion: Further delineation of an emerging syndrome’, *American Journal of Medical Genetics Part A*. John Wiley & Sons, Ltd, 164(8), pp. 1916–1922. doi: 10.1002/ajmg.a.36554.

Cagalinec, M. *et al.* (2016) ‘Role of Mitochondrial Dynamics in Neuronal Development: Mechanism for Wolfram Syndrome’, *PLOS Biology*. Edited by C. Blackstone, 14(7), p. e1002511. doi: 10.1371/journal.pbio.1002511.

Caglayan, A. O. (2010) ‘Genetic causes of syndromic and non-syndromic autism’, *Developmental Medicine & Child Neurology*. John Wiley & Sons, Ltd (10.1111), 52(2), pp. 130–138. doi: 10.1111/j.1469-8749.2009.03523.x.

Cambray, S. *et al.* (2012) ‘Activin induces cortical interneuron identity and differentiation in embryonic stem cell-derived telencephalic neural precursors’, *Nature Communications*. Nature Publishing Group, 3(1), p. 841. doi: 10.1038/ncomms1817.

Canitano, R. and Pallagrosi, M. (2017) ‘Autism Spectrum Disorders and Schizophrenia Spectrum Disorders: Excitation/Inhibition Imbalance and Developmental Trajectories’, *Frontiers in Psychiatry*, 8, p. 69. doi: 10.3389/fpsyt.2017.00069.

Cantley, L. C. (2002) ‘The phosphoinositide 3-kinase pathway.’, *Science (New York, N.Y.)*. American Association for the Advancement of Science, 296(5573), pp. 1655–7. doi: 10.1126/science.296.5573.1655.

Cardno, A. G. and Gottesman, I. I. (2000) ‘Twin studies of schizophrenia: From bow-and-arrow concordances to Star Wars Mx and functional genomics’, *American Journal of Medical Genetics*, 97(1), pp. 12–17. doi: 10.1002/(SICI)1096-8628(200021)97:1<12::AID-AJMG3>3.0.CO;2-U.

Carroll, L. S. *et al.* (2010) ‘Evidence for rare and common genetic risk variants for schizophrenia at protein kinase C, alpha’, *Molecular Psychiatry*, 15(11), pp. 1101–1111. doi: 10.1038/mp.2009.96.

Chai, J.-H. *et al.* (2003) ‘Identification of four highly conserved genes between breakpoint hotspots BP1 and BP2 of the Prader-Willi/Angelman syndromes deletion region that have undergone evolutionary transposition mediated by flanking duplicons.’, *American journal of human genetics*. Cell Press, 73(4), pp. 898–925. doi: 10.1086/378816.

Chambers, S. M. *et al.* (2009) 'Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling.', *Nature biotechnology*, 27(3), pp. 275–80. doi: 10.1038/nbt.1529.

Chen, L. *et al.* (2003) 'The fragile X mental retardation protein binds and regulates a novel class of mRNAs containing U rich target sequences.', *Neuroscience*, 120(4), pp. 1005–17. doi: 10.1016/s0306-4522(03)00406-8.

Chen, Z. *et al.* (2010) 'Structure and control of the actin regulatory WAVE complex', *Nature*, 468(7323), pp. 533–538. doi: 10.1038/nature09623.

Cheng, X. *et al.* (2018) 'The effect of P85 on neuronal proliferation and differentiation during development of mouse cerebral cortex', *Developmental Biology*, 441(1), pp. 95–103. doi: 10.1016/j.ydbio.2018.06.016.

Chenn, A. and Walsh, C. A. (2002) 'Regulation of Cerebral Cortical Size by Control of Cell Cycle Exit in Neural Precursors', *Science*, 297(5580), pp. 365–369. doi: 10.1126/science.1074192.

Chenn, A. and Walsh, C. A. (2003) 'Increased Neuronal Production, Enlarged Forebrains and Cytoarchitectural Distortions in beta-Catenin Overexpressing Transgenic Mice', *Cerebral Cortex*, 13(6), pp. 599–606. doi: 10.1093/cercor/13.6.599.

Chesney, Edward, Goodwind, Guy M. and Fazel, Seena (2014) 'Risks of all-cause and suicide mortality in mental disorders: a meta-review', *World Psychiatry*, 13, pp. 153–160.

Chilov, D. *et al.* (2011) 'Phosphorylated β-catenin localizes to centrosomes of neuronal progenitors and is required for cell polarity and neurogenesis in developing midbrain', *Developmental Biology*. Academic Press, 357(1), pp. 259–268. doi: 10.1016/J.YDBIO.2011.06.029.

Cho, H. *et al.* (2001) 'Akt1/PKBα Is Required for Normal Growth but Dispensable for Maintenance of Glucose Homeostasis in Mice', *Journal of Biological Chemistry*, 276(42), pp. 38349–38352. doi: 10.1074/jbc.C100462200.

Cho, H. (2001) 'Insulin Resistance and a Diabetes Mellitus-Like Syndrome in Mice Lacking the Protein Kinase Akt2 (PKBbeta)', *Science*, 292(5522), pp. 1728–1731. doi: 10.1126/science.292.5522.1728.

Chong, H. Y. *et al.* (2016) 'Global economic burden of schizophrenia: a systematic review.', *Neuropsychiatric disease and treatment*, 12, pp. 357–73. doi: 10.2147/NDT.S96649.

Chow, M. L. *et al.* (2012) 'Age-Dependent Brain Gene Expression and Copy Number Anomalies in Autism Suggest Distinct Pathological Processes at Young Versus Mature Ages', *PLoS Genetics*. Edited by G. Gibson. Public Library of Science, 8(3), p. e1002592. doi: 10.1371/journal.pgen.1002592.

Christian, S. L. *et al.* (2008) 'Novel Submicroscopic Chromosomal Abnormalities Detected in Autism Spectrum Disorder', *Biological Psychiatry*, 63(12), pp. 1111–1117. doi: 10.1016/j.biopsych.2008.01.009.

Chung, B. H. Y., Tao, V. Q. and Tso, W. W. Y. (2014) 'Copy number variation and autism: New insights and clinical implications', *Journal of the Formosan Medical Association*. Elsevier Taiwan LLC, 113(7), pp. 400–408. doi: 10.1016/j.jfma.2013.01.005.

Cirillo, E. *et al.* (2018) 'From SNPs to pathways: Biological interpretation of type 2 diabetes (T2DM) genome wide association study (GWAS) results', *PLoS ONE*, 13(4), pp. 1–19. doi: 10.1371/journal.pone.0193515.

Clark, B. S. and Blackshaw, S. (2014) 'Long non-coding RNA-dependent transcriptional regulation in neuronal development and disease', *Frontiers in Genetics*, 5(JUN), pp. 1–19. doi: 10.3389/fgene.2014.00164.

Clarke, T.-K. *et al.* (2016) 'Common polygenic risk for autism spectrum disorder (ASD) is associated with cognitive ability in the general population', *Molecular Psychiatry*. Nature Publishing Group, 21(3), pp. 419–425. doi: 10.1038/mp.2015.12.

Clemson, C. M. *et al.* (2009) 'An Architectural Role for a Nuclear Noncoding RNA: NEAT1 RNA Is Essential for the Structure of Paraspeckles', *Molecular Cell*. Elsevier Ltd, 33(6), pp. 717–726. doi: 10.1016/j.molcel.2009.01.026.

Coe, B. P. *et al.* (2014) 'Refining analyses of copy number variation identifies specific genes associated with developmental delay', *Nature Genetics*. Nature Publishing Group, 46(10), pp. 1063–1071. doi: 10.1038/ng.3092.

Conboy, J. G. (2017) 'Developmental regulation of RNA processing by Rbfox proteins', *Wiley Interdisciplinary Reviews: RNA*, 8(2), p. e1398. doi: 10.1002/wrna.1398.

Consortium, T. A. S. D. W. G. of T. P. G. (2017) 'Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia', *Molecular Autism*. BioMed Central, 8(1), p. 21. doi: 10.1186/s13229-017-0137-9.

Cooper, M. K. *et al.* (2003) 'A defective response to Hedgehog signaling in disorders of cholesterol biosynthesis', *Nature Genetics*, 33(4), pp. 508–513. doi: 10.1038/ng1134.

Coulter, T. I. *et al.* (2017) 'Clinical spectrum and features of activated phosphoinositide 3-kinase δ syndrome: A large patient cohort study', *Journal of Allergy and Clinical Immunology*, 139(2), pp. 597-606.e4. doi: 10.1016/j.jaci.2016.06.021.

Courchesne, E. *et al.* (2011) 'Neuron Number and Size in Prefrontal Cortex of Children With Autism', *JAMA*, 306(18), p. 2001. doi: 10.1001/jama.2011.1638.

Craddock, N., O'Donovan, M. C. and Owen, M. J. (2005) 'The genetics of schizophrenia and bipolar disorder: Dissecting psychosis', *Journal of Medical Genetics*, 42(3), pp. 193–204. doi: 10.1136/jmg.2005.030718.

Craddock, N. and Owen, M. J. (2010) 'The Kraepelinian dichotomy – going, going … but still not gone', *British Journal of Psychiatry*, 196(2), pp. 92–95. doi: 10.1192/bjp.bp.109.073429.

Crespi, B., Stead, P. and Elliot, M. (2010) 'Comparative genomics of autism and schizophrenia', *Proceedings of the National Academy of Sciences*, 107(suppl_1), pp. 1736–1741. doi: 10.1073/pnas.0906080106.

Csernansky, J. G. *et al.* (2004) 'Abnormalities of Thalamic Volume and Shape in Schizophrenia', *American Journal of Psychiatry*, 161(5), pp. 896–902. doi: 10.1176/appi.ajp.161.5.896.

Curtis, D. (2018) 'Polygenic risk score for schizophrenia is more strongly associated with ancestry than with schizophrenia', *Psychiatric Genetics*, 28(5), pp. 85–89. doi: 10.1097/YPG.0000000000000206.

Curtis, D. *et al.* (2018) 'Weighted Burden Analysis of Exome-Sequenced Case-Control Sample Implicates Synaptic Genes in Schizophrenia Aetiology', *Behavior Genetics*. Springer US, 48(3), pp. 198–208. doi: 10.1007/s10519-018-9893-3.

Darnell, J. C. C. *et al.* (2011) 'FMRP Stalls Ribosomal Translocation on mRNAs Linked to Synaptic Function and Autism', *Cell*. Elsevier Inc., 146(2), pp. 247–261. doi: 10.1016/j.cell.2011.06.013.

Das, D. K. *et al.* (2015) 'Genetic and Morphological Features of Human iPSC-Derived Neurons with Chromosome 15q11.2 (BP1-BP2) Deletions', *Molecular Neuropsychiatry*, 1(2), pp. 116–123. doi: 10.1159/000430916.

De Rubeis, S. *et al.* (2013) 'CYFIP1 Coordinates mRNA Translation and Cytoskeleton Remodeling to Ensure Proper Dendritic Spine Formation', *Neuron*, 79(6), pp. 1169–1182. doi: 10.1016/j.neuron.2013.06.039.

Delgado-Esteban, M. *et al.* (2013) 'APC/C-Cdh1 coordinates neurogenesis and cortical size during development', *Nature Communications*, 4(1), p. 2879. doi: 10.1038/ncomms3879.

Derkach, V., Barria, A. and Soderling, T. R. (1999) 'Ca2+/calmodulin-kinase II enhances channel conductance of alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionate type glutamate receptors.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 96(6), pp. 3269–74. doi: 10.1073/pnas.96.6.3269.

Derrien, T. *et al.* (2012) 'The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression', *Genome Research*, 22(9), pp. 1775–1789. doi: 10.1101/gr.132159.111.

Diez, H., Garrido, J. J. and Wandosell, F. (2012) 'Specific roles of akt iso forms in apoptosis and axon growth regulation in neurons', *PLoS ONE*. Edited by M. Gasset, 7(4), p. e32715. doi: 10.1371/journal.pone.0032715.

Doornbos, M. *et al.* (2009) 'Nine patients with a microdeletion 15q11.2 between breakpoints 1 and 2 of the Prader–Willi critical region, possibly associated with behavioural disturbances', *European Journal of Medical Genetics*. Elsevier Masson, 52(2–3), pp. 108–115. doi: 10.1016/j.ejmg.2009.03.010.

Durinck, S. *et al.* (2005) 'BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis', *Bioinformatics*, 21(16), pp. 3439–3440. doi: 10.1093/bioinformatics/bti525.

Durinck, S. *et al.* (2009) 'Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt.', *Nature protocols*. NIH Public Access, 4(8), pp. 1184–91. doi: 10.1038/nprot.2009.97.

Easton, R. M. *et al.* (2005) 'Role for Akt3 / Protein Kinase B in Attainment of Normal Brain Size', *Molecular and Cellular Biology*, 25(5), pp. 1869–1878. doi: 10.1128/MCB.25.5.1869.

Ebert, D. H. and Greenberg, M. E. (2013) 'Activity-dependent neuronal signalling and autism spectrum disorder', *Nature*, 8(24), pp. 327–337. doi: 10.1002/bmb.20244.DNA.

Eden, S. *et al.* (2002) 'Mechanism of regulation of WAVE1-induced actin nucleation by Rac1 and Nck', *Nature*, 418(6899), pp. 790–793. doi: 10.1038/nature00859.

Elghazi, L. *et al.* (2007) 'Regulation of ?-cell mass and function by the Akt/protein kinase B signalling pathway', *Diabetes, Obesity and Metabolism*, 9(s2), pp. 147–157. doi:

10.1111/j.1463-1326.2007.00783.x.

elise Jønch, aia *et al.* (2019) 'Estimating the effect size of the 15Q11.2 BP1-BP2 deletion and its contribution to neurodevelopmental symptoms: recommendations for practice on behalf of 15q11.2 Working group Copy-number variation', *Damien Sanlaville*, 56, p. 20. doi: 10.1136/jmedgenet-2018-105879.

Emamian, E. S. *et al.* (2004) 'Convergent evidence for impaired AKT1-GSK3β signaling in schizophrenia', *Nature Genetics*, 36(2), pp. 131–137. doi: 10.1038/ng1296.

Emamian, E. S. (2012) 'AKT/GSK3 signaling pathway and schizophrenia', *Frontiers in Molecular Neuroscience*, 5(March), pp. 1–12. doi: 10.3389/fnmol.2012.00033.

Engle, K. M.; Mei, T-S.; Wasa, M.; Yu, J.-Q. (2008) 'Autism Candidate Genes via Mouse Phenomics', *J Biomed Inform*, 45(6), pp. 788–802. doi: 10.1038/jid.2014.371.

Enriquez-Barreto, L. and Morales, M. (2016) 'The PI3K signaling pathway as a pharmacological target in Autism related disorders and Schizophrenia', *Molecular and Cellular Therapies*. Molecular and Cellular Therapies, 4(1), p. 2. doi: 10.1186/s40591-016-0047-9.

Evans, R. J., Derkach, V. and Surprenant, A. (1992) 'ATP mediates fast synaptic transmission in mammalian neurons', *Nature*, 357(6378), pp. 503–505. doi: 10.1038/357503a0.

Fang, D. *et al.* (2007) 'Phosphorylation of β-catenin by AKT promotes β-catenin transcriptional activity', *Journal of Biological Chemistry*, 282(15), pp. 11221–11229. doi: 10.1074/jbc.M611871200.

Faridar, A. *et al.* (2014) 'Mapk/Erk activation in an animal model of social deficits shows a possible link to autism', *Molecular Autism*, 5(1), pp. 1–12. doi: 10.1186/2040-2392-5-57.

La Fata, G. *et al.* (2014) 'FMRP regulates multipolar to bipolar transition affecting neuronal migration and cortical circuitry', *Nature Neuroscience*. Nature Publishing Group, 17(12), pp. 1693–1700. doi: 10.1038/nn.3870.

Fernández, E., Rajan, N. and Bagni, C. (2013) 'The FMRP regulon: from targets to disease convergence', *Frontiers in Neuroscience*. Frontiers, 7, p. 191. doi: 10.3389/fnins.2013.00191.

Fink, J. K. (2003) 'The Hereditary Spastic Paraplegias', *Archives of Neurology*. American Medical Association, 60(8), p. 1045. doi: 10.1001/archneur.60.8.1045.

Fischbach, G. D. and Lord, C. (2010) ‘The simons simplex collection: A resource for identification of autism genetic risk factors’, *Neuron*. Elsevier Inc., 68(2), pp. 192–195. doi: 10.1016/j.neuron.2010.10.006.

Fish, J. L. *et al.* (2008) ‘Making bigger brains-the evolution of neural-progenitor-cell division’, *Journal of Cell Science*, 121(17), pp. 2783–2793. doi: 10.1242/jcs.023465.

Fregoso, V., Paterson, C. and Law, A. (2017) ‘Effects of PIK3CD Over-Expression on Neuronal Morphology: Implications for Schizophrenia’, *Biological Psychiatry*. Elsevier, 81(10), p. S209. doi: 10.1016/j.biopsych.2017.02.1123.

Frith, C. and Dolan, R. (1996) ‘The role of the prefrontal cortex in higher cognitive functions.’, *Brain research. Cognitive brain research*, 5(1–2), pp. 175–81. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9049084 (Accessed: 18 June 2019).

Fromer, M. *et al.* (2014) ‘De novo mutations in schizophrenia implicate synaptic networks’, *Nature*, 506(7487), pp. 179–184. doi: 10.1038/nature12929.

Fruman, D. A., Meyers, R. E. and Cantley, L. C. (1998) ‘PHOSPHOINOSITIDE KINASES’, *Annual Review of Biochemistry*. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA, 67(1), pp. 481–507. doi: 10.1146/annurev.biochem.67.1.481.

Frye, R. E. and Rossignol, D. A. (2011) ‘Mitochondrial Dysfunction Can Connect the Diverse Medical Symptoms Associated With Autism Spectrum Disorders’. Nature Publishing Group, 69(5 Part 2), pp. 41R-47R. doi: 10.1203/PDR.0b013e318212f16b.

Gaiano, N. and Fishell, G. (2002) ‘The Role of Notch in Promoting Glial and Neural Stem Cell Fates’, *Annual Review of Neuroscience*, 25(1), pp. 471–490. doi: 10.1146/annurev.neuro.25.030702.130823.

Gaiano, N., Nye, J. S. and Fishell, G. (2000) *Notch sig*, *Bao and Cepko*. Nye. Available at: https://www.cell.com/neuron/pdf/S0896-6273(00)81172-1.pdf (Accessed: 10 July 2019).

Garg, S. *et al.* (2017) ‘Autism spectrum disorder and other neurobehavioural comorbidities in rare disorders of the Ras/MAPK pathway’, *Developmental Medicine and Child Neurology*, 59(5), pp. 544–549. doi: 10.1111/dmcn.13394.

Gaugler, T. *et al.* (2014) ‘Most genetic risk for autism resides with common variation’, *Nature Genetics*. Nature Publishing Group, 46(8), pp. 881–885. doi: 10.1038/ng.3039.

Gburcik, V. *et al.* (2005) ‘SPBP Is a Phosphoserine-Specific Repressor of Estrogen

Receptor ', *Molecular and Cellular Biology*, 25(9), pp. 3421–3430. doi: 10.1128/mcb.25.9.3421-3430.2005.

Genis-Mendoza, A. *et al.* (2018) 'Comparative Analysis of Gene Expression Profiles Involved in Calcium Signaling Pathways Using the NLVH Animal Model of Schizophrenia', *Journal of Molecular Neuroscience*. Journal of Molecular Neuroscience, 64(1), pp. 111–116. doi: 10.1007/s12031-017-1013-y.

Genovese, G. *et al.* (2016) 'Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia', *Nature Neuroscience*, 19(11), pp. 1433–1441. doi: 10.1038/nn.4402.

Geschwind, D. H. and State, M. W. (2015) 'Gene hunting in autism spectrum disorder: on the path to precision medicine', *The Lancet Neurology*, 14(11), pp. 1109–1120. doi: 10.1016/S1474-4422(15)00044-7.

Giebel, B. and Wodarz, A. (2012) 'Notch Signaling: Numb Makes the Difference', *Current Biology*. Cell Press, 22(4), pp. R133–R135. doi: 10.1016/j.cub.2012.01.006.

Gill, S., Chow, R. and Brown, A. (2008) 'Sterol regulators of cholesterol homeostasis and beyond: The oxysterol hypothesis revisited and revised', *Progress in Lipid Research*, 47(6), pp. 391–404. doi: 10.1016/j.plipres.2008.04.002.

Girirajan, S. *et al.* (2013) 'Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder', *American Journal of Human Genetics*. The American Society of Human Genetics, 92(2), pp. 221–237. doi: 10.1016/j.ajhg.2012.12.016.

Gleeson, J. G. *et al.* (1998) 'Doublecortin, a brain-specific gene mutated in human X-linked lissencephaly and double cortex syndrome, encodes a putative signaling protein.', *Cell*, 92(1), pp. 63–72. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9489700 (Accessed: 21 June 2019).

Glessner, J. T. *et al.* (2009) 'Autism genome-wide copy number variation reveals ubiquitin and neuronal genes', *Nature*, 459(7246), pp. 569–573. doi: 10.1038/nature07953.

Goeman, J. J. and Buhlmann, P. (2007) 'Analyzing gene expression data in terms of gene sets: methodological issues', *Bioinformatics*, 23(8), pp. 980–987. doi: 10.1093/bioinformatics/btm051.

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.

Gottesman, I. I. and Shields, J. (1967) 'A polygenic theory of schizophrenia.', *Proceedings of the National Academy of Sciences*, 58(1), pp. 199–205. doi: 10.1073/pnas.58.1.199.

Goytain, A. *et al.* (2007) 'NIPA1 (SPG6), the Basis for Autosomal Dominant Form of Hereditary Spastic Paraplegia, Encodes a Functional Mg 2+ Transporter', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology, 282(11), pp. 8060–8068. doi: 10.1074/jbc.M610314200.

Goytain, A., Hines, R. M. and Quamme, G. A. (2008) 'Functional characterization of NIPA2, a selective Mg 2+ transporter', *American Journal of Physiology-Cell Physiology*, 295(4), pp. C944–C953. doi: 10.1152/ajpcell.00091.2008.

Gray, L. *et al.* (2009) 'Sensitivity to MK-801 in phospholipase C-B1 knockout mice reveals a specific NMDA receptor deficit', *International Journal of Neuropsychopharmacology*, 12(7), pp. 917–928. doi: 10.1017/S1461145709009961.

Greig, L. C. *et al.* (2013) 'Molecular logic of neocortical projection neuron specification, development and diversity.', *Nature reviews. Neuroscience*. NIH Public Access, 14(11), pp. 755–69. doi: 10.1038/nrn3586.

Griffiths, W. J. *et al.* (2016) 'Current trends in oxysterol research', *Biochemical Society Transactions*, 44(2), pp. 652–658. doi: 10.1042/BST20150255.

Gross, C. *et al.* (2019) 'Isoform-selective phosphoinositide 3-kinase inhibition ameliorates a broad range of fragile X syndrome-associated deficits in a mouse model', *Neuropsychopharmacology*. Springer US, 44(2), pp. 324–333. doi: 10.1038/s41386-018-0150-5.

Grove, J. *et al.* (2019) 'Identification of common genetic risk variants for autism spectrum disorder', *Nature Genetics*. Nature Publishing Group, 51(3), pp. 431–444. doi: 10.1038/s41588-019-0344-8.

Grozeva, D. *et al.* (2012) 'Independent estimation of the frequency of rare CNVs in the UK population confirms their role in schizophrenia', *Schizophrenia Research*. Elsevier B.V., 135(1–3), pp. 1–7. doi: 10.1016/j.schres.2011.11.004.

Gu, W., Zhang, F. and Lupski, J. R. (2008) 'Mechanisms for human genomic rearrangements', *PathoGenetics*, 1(1), p. 4. doi: 10.1186/1755-8417-1-4.

Gulsuner, S. and Mcclellan, J. M. (2014) 'De novo mutations in schizophrenia disrupt

genes co-expressed in fetal prefrontal cortex', *Neuropsychopharmacology*. Nature Publishing Group, 39(1), pp. 238–239. doi: 10.1038/npp.2013.219.

Gunhaga, L. *et al.* (2003) 'Specification of dorsal telencephalic character by sequential Wnt and FGF signaling', *Nature Neuroscience*, 6(7), pp. 701–707. doi: 10.1038/nn1068.

Gur, R. E. *et al.* (1999) 'Reduced Gray Matter Volume in Schizophrenia', *Archives of General Psychiatry*. American Medical Association, 56(10), p. 905. doi: 10.1001/archpsyc.56.10.905.

Haddon, C. *et al.* (1998) 'Multiple delta genes and lateral inhibition in zebrafish primary neurogenesis.', *Development (Cambridge, England)*, 125(3), pp. 359–70. Available at: https://dev.biologists.org/content/develop/125/3/359.full.pdf (Accessed: 10 July 2019).

Haijma, S. V. *et al.* (2013) 'Brain Volumes in Schizophrenia: A Meta-Analysis in Over 18 000 Subjects', *Schizophrenia Bulletin*. Narnia, 39(5), pp. 1129–1138. doi: 10.1093/schbul/sbs118.

Hannon, E. *et al.* (2018) 'Elevated polygenic burden for autism is associated with differential DNA methylation at birth', *Genome Medicine*. Genome Medicine, 10(1), pp. 1–13. doi: 10.1186/s13073-018-0527-4.

Hardan, A. Y. *et al.* (2009) 'Corpus callosum volume in children with autism', *Psychiatry Research: Neuroimaging*. Elsevier, 174(1), pp. 57–61. doi: 10.1016/J.PSCYCHRESNS.2009.03.005.

Hartfuss, E. *et al.* (2001) 'Characterization of CNS Precursor Subtypes and Radial Glia', *Developmental Biology*, 229(1), pp. 15–30. doi: 10.1006/dbio.2000.9962.

Heitzler, P. and Simpson, P. (1991) 'The choice of cell fate in the epidermis of Drosophila.', *Cell*. Elsevier, 64(6), pp. 1083–92. doi: 10.1016/0092-8674(91)90263-x.

Hemmings, B. A. and Restuccia, D. F. (2012) 'PI3K-PKB/Akt pathway.', *Cold Spring Harbor perspectives in biology*. Cold Spring Harbor Laboratory Press, 4(9), p. a011189. doi: 10.1101/cshperspect.a011189.

Henriksen, M. G., Nordgaard, J. and Jansson, L. B. (2017) 'Genetics of Schizophrenia: Overview of Methods, Findings and Limitations', *Frontiers in Human Neuroscience*. Frontiers Media SA, 11. doi: 10.3389/FNHUM.2017.00322.

Homs, A. *et al.* (2016) 'Genetic and epigenetic methylation defects and implication of the ERMN gene in autism spectrum disorders', *Translational psychiatry*, 6(7), p. e855. doi: 10.1038/tp.2016.120.

Hong, S. *et al.* (2013) 'Canonical correlation analysis for RNA-seq co-expression networks', *Nucleic Acids Research*, 41(8), pp. 1–15. doi: 10.1093/nar/gkt145.

Hormozdiari, F. *et al.* (2015) 'The discovery of integrated gene networks for autism and related disorders', *Genome Research*, 25(1), pp. 142–154. doi: 10.1101/gr.178855.114.

Howell, K. R., Floyd, K. and Law, A. J. (2017) 'PKBγ/AKT3 loss-of-function causes learning and memory deficits and deregulation of AKT/mTORC2 signaling: Relevance for schizophrenia', *PLOS ONE*. Edited by G. S. Barsh, 12(5), p. e0175993. doi: 10.1371/journal.pone.0175993.

Hsiao, K. *et al.* (2016) 'Cyfip1 Regulates Presynaptic Activity during Development', *The Journal of Neuroscience*, 36(5), pp. 1564–1576. doi: 10.1523/JNEUROSCI.0511-15.2016.

Hu, W. F., Chahrour, M. H. and Walsh, C. A. (2014) 'The Diverse Genetic Landscape of Neurodevelopmental Disorders', *Annual Review of Genomics and Human Genetics*. Annual Reviews , 15(1), pp. 195–213. doi: 10.1146/annurev-genom-090413-025600.

Huaman, M. A. *et al.* (2015) 'Genotype to phenotype relationships in autism spectrum disorders', *Nat Neurosci*, 143(5), pp. 951–959. doi: 10.1017/S0950268814002131.Tuberculosis.

Hussman, J. P. *et al.* (2011) 'A noise-reduction GWAS analysis implicates altered regulation of neurite outgrowth and guidance in autism', *Molecular Autism*. BioMed Central Ltd, 2(1), p. 1. doi: 10.1186/2040-2392-2-1.

Hyman, S. E. (2000) *The genetics of mental illness: implications for practice*. Available at: https://www.who.int/bulletin/archives/78(4)455.pdf (Accessed: 22 July 2019).

Igolkina, A. A. *et al.* (2018) 'Analysis of Gene Expression Variance in Schizophrenia Using Structural Equation Modeling', *Frontiers in Molecular Neuroscience*, 11(June), pp. 1–12. doi: 10.3389/fnmol.2018.00192.

Imayoshi, I. *et al.* (2010) 'Essential Roles of Notch Signaling in Maintenance of Neural Stem Cells in Developing and Adult Brains', *Journal of Neuroscience*, 30(9), pp. 3489–3498. doi: 10.1523/JNEUROSCI.4987-09.2010.

Imbeaud, S. *et al.* (2005) 'Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces.', *Nucleic acids research*. Oxford University Press, 33(6), p. e56. doi: 10.1093/nar/gni054.

Inan, M. *et al.* (2016) 'Energy deficit in parvalbumin neurons leads to circuit dysfunction,

impaired sensory gating and social disability', *Neurobiology of Disease*. Academic Press, 93, pp. 35–46. doi: 10.1016/J.NBD.2016.04.004.

Ingason, A. *et al.* (2011) 'Copy number variations of chromosome 16p13.1 region associated with schizophrenia', *Molecular Psychiatry*, 16(1), pp. 17–25. doi: 10.1038/mp.2009.101.

International Schizophrenia Consortium *et al.* (2008) 'Rare chromosomal deletions and duplications increase risk of schizophrenia', *Nature*. Nature Publishing Group, 455(7210), pp. 237–241. doi: 10.1038/nature07239.

Iossifov, I. *et al.* (2014) 'The contribution of de novo coding mutations to autism spectrum disorder', *Nature*. Nature Publishing Group, 515(7526), pp. 216–221. doi: 10.1038/nature13908.

Irwin, S. A. *et al.* (2001) 'Abnormal dendritic spine characteristics in the temporal and visual cortices of patients with fragile-X syndrome: a quantitative examination.', *American journal of medical genetics*, 98(2), pp. 161–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11223852 (Accessed: 15 July 2019).

Itoh, Y. *et al.* (2016) 'PDK1–Akt pathway regulates radial neuronal migration and microtubules in the developing mouse neocortex', *Proceedings of the National Academy of Sciences*, 113(21), pp. E2955–E2964. doi: 10.1073/pnas.1516321113.

Jablensky, A. *et al.* (2011) 'Polymorphisms associated with normal memory variation also affect memory impairment in schizophrenia', *Genes, Brain and Behavior*, 10(4), pp. 410–417. doi: 10.1111/j.1601-183X.2011.00679.x.

Jacquemont, S. *et al.* (2007) 'Fragile-X syndrome and fragile X-associated tremor/ataxia syndrome: two faces of FMR1', *The Lancet Neurology*, 6(1), pp. 45–55. doi: 10.1016/S1474-4422(06)70676-7.

Jajodia, A. *et al.* (2016) 'Evaluation of genetic association of neurodevelopment and neuroimmunological genes with antipsychotic treatment response in schizophrenia in Indian populations', *Molecular Genetics & Genomic Medicine*, 4(1), pp. 18–27. doi: 10.1002/mgg3.169.

Jandial, R. *et al.* (2008) 'Genetic Modification of Neural Stem Cells', *Molecular Therapy*. The American Society of Gene Therapy, 16(3), pp. 450–457. doi: 10.1038/sj.mt.6300402.

Jansen, A. *et al.* (2017) 'Gene-set analysis shows association between FMRP targets and autism spectrum disorder', *European Journal of Human Genetics*, 25(7), pp. 863–

868. doi: 10.1038/ejhg.2017.55.

Jansen, L. A. *et al.* (2015) 'PI3K/AKT pathway mutations cause a spectrum of brain malformations from megalencephaly to focal cortical dysplasia', *Brain*, 138(6), pp. 1613–1628. doi: 10.1093/brain/awv045.

Jerkovich, A. and Butler, M. (2015) 'Further phenotypic expansion of 15q11.2 BP1-BP2 microdeletion (Burnside-Butler) syndrome', *Journal of Pediatric Genetics*. NIH Public Access, 03(01), pp. 041–044. doi: 10.3233/PGE-14082.

Jiang, Y. *et al.* (2012) 'NIPA2 located in 15q11.2 is mutated in patients with childhood absence epilepsy', *Human Genetics*, 131(7), pp. 1217–1224. doi: 10.1007/s00439-012-1149-3.

Johnston, J. C. *et al.* (1999) 'Minimum requirements for efficient transduction of dividing and nondividing cells by feline immunodeficiency virus vectors.', *Journal of virology*, 73(6), pp. 4991–5000. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10233961%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC112543.

Jossin, Y. and Goffinet, A. M. (2007) 'Reelin Signals through Phosphatidylinositol 3-Kinase and Akt To Control Cortical Development and through mTor To Regulate Dendritic Growth', *Molecular and Cellular Biology*, 27(20), pp. 7113–7124. doi: 10.1128/MCB.00928-07.

Junghans, D. *et al.* (2005) 'β-catenin-mediated cell-adhesion is vital for embryonic forebrain development', *Developmental Dynamics*, 233(2), pp. 528–539. doi: 10.1002/dvdy.20365.

Kabadi, A. M. *et al.* (2014) 'Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector', *Nucleic Acids Research*, 42(19), pp. e147–e147. doi: 10.1093/nar/gku749.

Kaddurah-Daouk, R. *et al.* (2012) 'Impaired plasmalogens in patients with schizophrenia', *Psychiatry Research*. Elsevier, 198(3), pp. 347–352. doi: 10.1016/J.PSYCHRES.2012.02.019.

Kanehisa, M. and Goto, S. (2000) 'KEGG: Kyoto Encyclopedia of Genes and Genomes', *Nucleic Acids Research*, 28(1), pp. 27–30. doi: 10.1093/nar/28.1.27.

Karbownik, M. S. *et al.* (2018) 'Activation of phosphoinositide 3-kinase delta by antipsychotic drugs: Preliminary results', *Pharmacological Reports*, 70(2), pp. 346–349. doi: 10.1016/j.pharep.2017.09.005.

Kelly L Wormwood, A. G. W. (2013) 'Proteomics and Cholesterol in Autism', *Autism-Open Access*, 03(04). doi: 10.4172/2165-7890.1000112.

Kim, S. W. *et al.* (2015) 'Knockdown of phospholipase C-β1 in the medial prefrontal cortex of male mice impairs working memory among multiple schizophrenia endophenotypes', *Journal of Psychiatry and Neuroscience*, 40(2), pp. 78–88. doi: 10.1503/jpn.130285.

Kim, W.-Y. Y. *et al.* (2009) 'GSK-3 is a master regulator of neural progenitor homeostasis', *Nature Neuroscience*, 12(11), pp. 1390–1397. doi: 10.1038/nn.2408.

Kirov, G. *et al.* (2009) 'Support for the involvement of large copy number variants in the pathogenesis of schizophrenia', *Human Molecular Genetics*, 18(8), pp. 1497–1503. doi: 10.1093/hmg/ddp043.

Kirov, G. (2015) 'CNVs in neuropsychiatric disorders', *Human Molecular Genetics*, 24(R1), pp. R45–R49. doi: 10.1093/hmg/ddv253.

Kiyota, K. *et al.* (2018) 'Auto-immune disorders in a child with PIK3CD variant and 22q13 deletion', *European Journal of Medical Genetics*. Elsevier, 61(10), pp. 631–633. doi: 10.1016/j.ejmg.2018.04.008.

Klein, R. J. *et al.* (2005) 'Complement Factor H Polymorphism in Age-Related Macular Degeneration [epub ahead of print] [Record Supplied By Publisher]', *Science*, (April), pp. 385–390.

Kobayashi, K. *et al.* (1998) 'p140Sra-1 (specifically Rac1-associated protein) is a novel specific target for Rac1 small GTPase', *Journal of Biological Chemistry*, 273(1), pp. 291–295. doi: 10.1074/jbc.273.1.291.

Kolde, R. and Vilo, J. (2015) 'GOsummaries: an R Package for Visual Functional Annotation of Experimental Data', *F1000Research*, (0), pp. 1–23. doi: 10.12688/f1000research.6925.1.

Kowalczyk, T. *et al.* (2009) 'Intermediate Neuronal Progenitors (Basal Progenitors) Produce Pyramidal–Projection Neurons for All Layers of Cerebral Cortex', *Cerebral Cortex*, 19(10), pp. 2439–2450. doi: 10.1093/cercor/bhn260.

Krakowski, M. and Czobor, P. (2011) 'Cholesterol and cognition in schizophrenia: A double-blind study of patients randomized to clozapine, olanzapine and haloperidol', *Schizophrenia Research*. Elsevier B.V., 130(1–3), pp. 27–33. doi: 10.1016/j.schres.2011.04.005.

Kraushar, M. L. *et al.* (2014) 'Temporally defined neocortical translation and polysome assembly are determined by the RNA-binding protein Hu antigen R', *Proceedings of the National Academy of Sciences*, 111(36), pp. E3815–E3824. doi: 10.1073/pnas.1408305111.

Kriegstein, A., Noctor, S. and Martínez-Cerdeño, V. (2006) 'Patterns of neural stem and progenitor cell division may underlie evolutionary cortical expansion', *Nature Reviews Neuroscience*, 7(11), pp. 883–890. doi: 10.1038/nrn2008.

Kumamoto, T. and Hanashima, C. (2017) 'Evolutionary conservation and conversion of Foxg1 function in brain development', *Development Growth and Differentiation*, 59(4), pp. 258–269. doi: 10.1111/dgd.12367.

Kumar, C. C. and Madison, V. (2005) 'AKT crystal structure and AKT-specific inhibitors', *Oncogene*, 24(50), pp. 7493–7501. doi: 10.1038/sj.onc.1209087.

Kumar, R. A. *et al.* (2008) 'Recurrent 16p11.2 microdeletions in autism', *Human Molecular Genetics*, 17(4), pp. 628–638. doi: 10.1093/hmg/ddm376.

Kushima, I. *et al.* (2018) 'Comparative Analyses of Copy-Number Variation in Autism Spectrum Disorder and Schizophrenia Reveal Etiological Overlap and Biological Insights', *Cell Reports*. Elsevier, 24(11), pp. 2838–2856. doi: 10.1016/j.celrep.2018.08.022.

de la Pompa, J. L. *et al.* (1997) 'Conservation of the Notch signalling pathway in mammalian neurogenesis.', *Development (Cambridge, England)*, 124(6), pp. 1139–48. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9102301 (Accessed: 10 July 2019).

Lacey, S., Chung, J. Y. and Lin, H. (2014) 'A comparison of whole genome sequencing with exome sequencing for family-based association studies', *BMC Proceedings*, 8(S1), p. S38. doi: 10.1186/1753-6561-8-S1-S38.

Lai, C. S. L. *et al.* (2001) 'A forkhead-domain gene is mutated in a severe speech and language disorder', *Nature*. Nature Publishing Group, 413(6855), pp. 519–523. doi: 10.1038/35097076.

Lai, E. C. and Kimble, J. (2004) 'Notch signaling: control of cell communication and cell fate.', *Development (Cambridge, England)*. The Company of Biologists Ltd, 131(5), pp. 965–73. doi: 10.1242/dev.01074.

Lambert, J. C. *et al.* (2013) 'Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease.', *Nature genetics*, 45(12), pp. 1452–8. doi: 10.1038/ng.2802.

Lancaster, M. A. and Knoblich, J. A. (2014) 'Organogenesis in a dish: Modeling development and disease using organoid technologies', *Science*, 345(6194), pp. 1247125–1247125. doi: 10.1126/science.1247125.

Lander, E. and Kruglyak, L. (1995) 'Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results', *Nature Genetics*, 11(3), pp. 241–247. doi: 10.1038/ng1195-241.

Lange, C., Huttner, W. B. and Calegari, F. (2009) 'Cdk4/CyclinD1 Overexpression in Neural Stem Cells Shortens G1, Delays Neurogenesis, and Promotes the Generation and Expansion of Basal Progenitors', *Cell Stem Cell*. Cell Press, 5(3), pp. 320–331. doi: 10.1016/J.STEM.2009.05.026.

Lauritsen, M. *et al.* (1999) 'Infantile autism and associated autosomal chromosome abnormalities: a register-based study and a literature survey.', *Journal of child psychology and psychiatry, and allied disciplines*, 40(3), pp. 335–45. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10190335 (Accessed: 21 June 2019).

Lauss, M. (2018) *Package 'swamp' Type Package Title Visualization, Analysis and Adjustment of High-Dimensional Data in Respect to Sample Annotations*. Available at: https://cran.r-project.org/web/packages/swamp/swamp.pdf (Accessed: 17 June 2019).

Law, A. J. *et al.* (2012) 'Neuregulin 1-ErbB4-PI3K signaling in schizophrenia and phosphoinositide 3-kinase-p110δ inhibition as a potential therapeutic strategy', *Proceedings of the National Academy of Sciences*, 109(30), pp. 12165–12170. doi: 10.1073/pnas.1206118109.

Lawrence, M. *et al.* (2013) 'Software for Computing and Annotating Genomic Ranges', *PLoS Computational Biology*. Edited by A. Prlic. Public Library of Science, 9(8), p. e1003118. doi: 10.1371/journal.pcbi.1003118.

Lawrence, M., Gentleman, R. and Carey, V. (2009) 'rtracklayer: an R package for interfacing with genome browsers', *Bioinformatics*. Narnia, 25(14), pp. 1841–1842. doi: 10.1093/bioinformatics/btp328.

Lee, R. W. Y. and Tierney, E. (2011) 'Hypothesis: The Role of Sterols in Autism Spectrum Disorder', *Autism Research and Treatment*. Hindawi, 2011, pp. 1–7. doi: 10.1155/2011/653570.

Lee, S. H. *et al.* (2012) 'Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs', *Nature Genetics*, 44(3), pp. 247–250. doi: 10.1038/ng.1108.

Lee, S. H. *et al.* (2013) 'Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs', *Nature Genetics*. Nature Publishing Group, 45(9), pp. 984–994. doi: 10.1038/ng.2711.

Leek, J. T. *et al.* (2010) 'Tackling the widespread and critical impact of batch effects in high-throughput data', *Nature Reviews Genetics*, 11(10), pp. 733–739. doi: 10.1038/nrg2825.

van de Leemput, J. *et al.* (2014) 'CORTECON: A temporal transcriptome analysis of in vitro human cerebral cortex development from human embryonic stem cells', *Neuron*. Elsevier Inc., 83(1), pp. 51–68. doi: 10.1016/j.neuron.2014.05.013.

de Leeuw, C. A. *et al.* (2015) 'MAGMA: Generalized Gene-Set Analysis of GWAS Data', *PLOS Computational Biology*. Edited by H. Tang. Public Library of Science, 11(4), p. e1004219. doi: 10.1371/journal.pcbi.1004219.

Leggett, R. M. and Clark, M. D. (2017) 'A world of opportunities with nanopore sequencing', *Journal of Experimental Botany*, 68(20), pp. 5419–5429. doi: 10.1093/jxb/erx289.

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.

Leonenko, G. *et al.* (2017) 'Mutation intolerant genes and targets of FMRP are enriched for nonsynonymous alleles in schizophrenia', *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 174(7), pp. 724–731. doi: 10.1002/ajmg.b.32560.

Levenga, J. *et al.* (2017) 'AKT isoforms have distinct hippocampal expression and roles in synaptic plasticity', *eLife*. eLife Sciences Publications Limited, 6, pp. 1–24. doi: 10.7554/elife.30640.

Lin-Hendel, E. G. *et al.* (2016) 'Differential Mitochondrial Requirements for Radially and Non-radially Migrating Cortical Neurons: Implications for Mitochondrial Disorders', *Cell Reports*. Cell Press, 15(2), pp. 229–237. doi: 10.1016/J.CELREP.2016.03.024.

von der Lippe, C. *et al.* (2011) '15q11.2 microdeletion – Seven new patients with delayed development and/or behavioural problems', *European Journal of Medical Genetics*, 54(3), pp. 357–360. doi: 10.1016/j.ejmg.2010.12.008.

Locke, D. P. *et al.* (2004) 'BAC microarray analysis of 15q11-q13 rearrangements and the impact of segmental duplications', *Journal of Medical Genetics*, 41(3), pp. 175–182. doi: 10.1136/jmg.2003.013813.

Love, M. I., Huber, W. and Anders, S. (2014) ‘Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.’, *Genome biology*, 15(12), p. 550. doi: 10.1186/s13059-014-0550-8.

Lui, J. H., Hansen, D. V. and Kriegstein, A. R. (2011) ‘Development and Evolution of the Human Neocortex’, *Cell*. Elsevier Inc., 146(1), pp. 18–36. doi: 10.1016/j.cell.2011.06.030.

Luo, Y. *et al.* (2010) ‘Fragile X Mental Retardation Protein Regulates Proliferation and Differentiation of Adult Neural Stem/Progenitor Cells’, *PLoS Genetics*. Edited by H. Orr, 6(4), p. e1000898. doi: 10.1371/journal.pgen.1000898.

Machon, O. *et al.* (2003) ‘Role of beta-catenin in the developing cortical and hippocampal neuroepithelium.’, *Neuroscience*, 122(1), pp. 129–43. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14596855 (Accessed: 10 July 2019).

Madrigal, I. *et al.* (2012) ‘15q11.2 microdeletion and FMR1 premutation in a family with intellectual disabilities and autism’, *Gene*, 508(1), pp. 92–95. doi: 10.1016/j.gene.2012.07.023.

Makeev, G. A. (1978) ‘Some aspects of the study of early stages of schizophrenia’, *Neuroscience and Behavioral Physiology*, 7(4), pp. 388–393. doi: 10.1007/BF01181850.

Malatesta, P. *et al.* (2000) ‘Isolation of radial glial cells by fluorescent-activated cell sorting reveals a neuronal lineage.’, 127(24), pp. 5253–63. Available at: https://dev.biologists.org/content/127/24/5253.long (Accessed: 10 July 2019).

Malhotra, D. and Sebat, J. (2012) ‘CNVs: Harbingers of a Rare Variant Revolution in Psychiatric Genetics’, *Cell*. Elsevier Inc., 148(6), pp. 1223–1241. doi: 10.1016/j.cell.2012.02.039.

Mamczur, P. *et al.* (2013) ‘Nuclear localization of aldolase A correlates with cell proliferation’, *Biochimica et Biophysica Acta - Molecular Cell Research*. Elsevier B.V., 1833(12), pp. 2812–2822. doi: 10.1016/j.bbamcr.2013.07.013.

Manning, E. E. *et al.* (2012) ‘Increased adult hippocampal neurogenesis and abnormal migration of adult-born granule neurons is associated with hippocampal-specific cognitive deficits in phospholipase C-β1 knockout mice’, *Hippocampus*, 22(2), pp. 309–319. doi: 10.1002/hipo.20900.

Marín, O. (2012) ‘Interneuron dysfunction in psychiatric disorders’, *Nature Reviews Neuroscience*. Nature Publishing Group, 13(2), pp. 107–120. doi: 10.1038/nrn3155.

Marín, O. and Müller, U. (2014) 'Lineage origins of GABAergic versus glutamatergic neurons in the neocortex', *Current Opinion in Neurobiology*. NIH Public Access, 26(12), pp. 132–141. doi: 10.1016/j.conb.2014.01.015.

Marshall, C. R. *et al.* (2008) 'Structural Variation of Chromosomes in Autism Spectrum Disorder', *The American Journal of Human Genetics*, 82(2), pp. 477–488. doi: 10.1016/j.ajhg.2007.12.009.

Marshall, C. R. *et al.* (2017) 'Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects', *Nature Genetics*, 49(1), pp. 27–35. doi: 10.1038/ng.3725.

Martynoga, B. *et al.* (2005) 'Foxg1 is required for specification of ventral telencephalon and region-specific regulation of dorsal telencephalic precursor proliferation and apoptosis', *Developmental Biology*, 283(1), pp. 113–127. doi: 10.1016/j.ydbio.2005.04.005.

Masi, A. *et al.* (2017) 'An Overview of Autism Spectrum Disorder, Heterogeneity and Treatment Options', *Neuroscience Bulletin*. Springer Singapore, 33(2), pp. 183–193. doi: 10.1007/s12264-017-0100-y.

Matsuda, S. *et al.* (2019) 'Roles of PI3K/AKT/GSK3 Pathway Involved in Psychiatric Illnesses', *Diseases*, 7(1), p. 22. doi: 10.3390/diseases7010022.

Mbom, B. C., Nelson, W. J. and Barth, A. (2013) 'β-catenin at the centrosome', *BioEssays*, 35(9), pp. 804–809. doi: 10.1002/bies.201300045.

McCarroll, S. A. *et al.* (2008) 'Integrated detection and population-genetic analysis of SNPs and copy number variation', *Nature Genetics*, 40(10), pp. 1166–1174. doi: 10.1038/ng.238.

McCarroll, S. A. and Altshuler, D. M. (2007) 'Copy-number variation and association studies of human disease', *Nature Genetics*, 39(7S), pp. S37–S42. doi: 10.1038/ng2080.

McCarthy, S. E. *et al.* (2009) 'Microduplications of 16p11.2 are associated with schizophrenia', *Nature Genetics*, 41(11), pp. 1223–1227. doi: 10.1038/ng.474.

McRae, J. F. *et al.* (2017) 'Prevalence and architecture of de novo mutations in developmental disorders', *Nature*, 542(7642), pp. 433–438. doi: 10.1038/nature21062.

Mefford, H. C. *et al.* (2008) 'Recurrent Rearrangements of Chromosome 1q21.1 and Variable Pediatric Phenotypes', *New England Journal of Medicine*. Massachusetts Medical Society , 359(16), pp. 1685–1699. doi: 10.1056/NEJMoa0805384.

Mendoza, M. C. (2013) ‘Phosphoregulation of the WAVE regulatory complex and signal integration’, *Seminars in Cell & Developmental Biology*, 24(4), pp. 272–279. doi: 10.1016/j.semcdb.2013.01.007.

Miller, D. T. *et al.* (2009) ‘Microdeletion/duplication at 15q13.2q13.3 among individuals with features of autism and other neuropsychiatric disorders.’, *Journal of medical genetics*. BMJ Publishing Group Ltd, 46(4), pp. 242–8. doi: 10.1136/jmg.2008.059907.

Miller, J. A. *et al.* (2014) ‘Transcriptional landscape of the prenatal human brain’, *Nature*, 508(7495), pp. 199–206. doi: 10.1038/nature13185.

Miyata, T. *et al.* (2001) ‘Asymmetric inheritance of radial glial fibers by cortical neurons.’, *Neuron*, 31(5), pp. 727–41. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11567613 (Accessed: 21 July 2019).

Mizutani, K. *et al.* (2007) ‘Differential Notch signalling distinguishes neural stem cells from intermediate progenitors’, *Nature*, 449(7160), pp. 351–355. doi: 10.1038/nature06090.

Moises, H. W. *et al.* (1995) ‘An international two–stage genome–wide search for schizophrenia susceptibility genes’, *Nature Genetics*. Nature Publishing Group, 11(3), pp. 321–324. doi: 10.1038/ng1195-321.

Moraes, L., Zanchin, N. I. T. and Cerutti, J. M. (2017) ‘ABI3, a component of the WAVE2 complex, is potentially regulated by PI3K/AKT pathway’, *Oncotarget*. Impact Journals, 8(40). doi: 10.18632/oncotarget.18840.

Moskvina, V. *et al.* (2012) ‘Permutation-based approaches do not adequately allow for linkage disequilibrium in gene-wide multi-locus association analysis’, *European Journal of Human Genetics*, 20(8), pp. 890–896. doi: 10.1038/ejhg.2012.8.

Mostaid, M. S. *et al.* (2017) ‘Peripheral transcription of NRG-ErbB pathway genes are upregulated in treatment-resistant schizophrenia’, *Frontiers in Psychiatry*, 8(NOV), pp. 1–9. doi: 10.3389/fpsyt.2017.00225.

Mukhtar, T. and Taylor, V. (2018) ‘Untangling Cortical Complexity During Development’, *Journal of Experimental Neuroscience*. SAGE PublicationsSage UK: London, England, 12, p. 117906951875933. doi: 10.1177/1179069518759332.

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*, 56(10), pp. 940–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10530637 (Accessed: 17 June 2019).

Murphy, S. M. *et al.* (2001) 'GCP5 and GCP6: Two New Members of the Human γ-Tubulin Complex', *Molecular Biology of the Cell*. Edited by J. R. McIntosh, 12(11), pp. 3340–3352. doi: 10.1091/mbc.12.11.3340.

Murthy, S. K. *et al.* (2007) 'Detection of a novel familial deletion of four genes between BP1 and BP2 of the Prader-Willi/Angelman syndrome critical region by oligo-array CGH in a child with neurological disorder and speech impairment', *Cytogenetic and Genome Research*, 116(1–2), pp. 135–140. doi: 10.1159/000097433.

Mutch, C. A. *et al.* (2010) 'Beta-Catenin Signaling Negatively Regulates Intermediate Progenitor Population Numbers in the Developing Cortex', *PLoS ONE*. Edited by R. Linden. Public Library of Science, 5(8), p. e12376. doi: 10.1371/journal.pone.0012376.

Napoli, I. *et al.* (2008) 'The Fragile X Syndrome Protein Represses Activity-Dependent Translation through CYFIP1, a New 4E-BP', *Cell*, 134(6), pp. 1042–1054. doi: 10.1016/j.cell.2008.07.031.

Nebel, R. A. *et al.* (2016) 'Reduced CYFIP1 in human neural progenitors results in dysregulation of schizophrenia and epilepsy gene networks', *PLoS ONE*, 11(1), pp. 1–20. doi: 10.1371/journal.pone.0148039.

Need, A. C. *et al.* (2009) 'A Genome-Wide Investigation of SNPs and CNVs in Schizophrenia', *PLoS Genetics*. Edited by M. I. McCarthy, 5(2), p. e1000373. doi: 10.1371/journal.pgen.1000373.

Nelson, E. R. *et al.* (2013) '27-Hydroxycholesterol links hypercholesterolemia and breast cancer pathophysiology.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 342(6162), pp. 1094–8. doi: 10.1126/science.1241908.

Newcomer, J. W. (2005) 'Second-Generation (Atypical) Antipsychotics and Metabolic Effects', *CNS Drugs*, 19(Supplement 1), p. 1???93. doi: 10.2165/00023210-200519001-00001.

Nicholls, R. D. and Knepper, J. L. (2001) 'G <scp>ENOME</scp> O <scp>RGANIZATION</scp> , F <scp>UNCTION</scp> , <scp>AND</scp> I <scp>MPRINTING IN</scp> P <scp>RADER</scp> -W <scp>ILLI AND</scp> A <scp>NGELMAN</scp> S <scp>YNDROMES</scp>', *Annual Review of Genomics and Human Genetics*, 2(1), pp. 153–175. doi: 10.1146/annurev.genom.2.1.153.

Nicholson, K. M. and Anderson, N. G. (2002) 'The protein kinase B/Akt signalling pathway in human malignancy', *Cellular Signalling*. Pergamon, 14(5), pp. 381–395. doi: 10.1016/S0898-6568(01)00271-6.

Lo Nigro, C. *et al.* (1997) 'Point Mutations and an Intragenic Deletion in LIS1, the Lissencephaly Causative Gene in Isolated Lissencephaly Sequence and Miller-Dieker Syndrome', *Human Molecular Genetics*, 6(2), pp. 157–164. doi: 10.1093/hmg/6.2.157.

Nishioka, M. *et al.* (2018) 'Somatic mutations in the human brain: implications for psychiatric research', *Molecular Psychiatry*. Springer US. doi: 10.1038/s41380-018-0129-y.

Noctor, S. C. *et al.* (2001) 'Neurons derived from radial glial cells establish radial units in neocortex', *Nature*, 409(6821), pp. 714–720. doi: 10.1038/35055553.

Noctor, S. C. *et al.* (2002) 'Dividing precursor cells of the embryonic cortical ventricular zone have morphological and molecular characteristics of radial glia.', *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 22(8), pp. 3161–73. doi: 20026299.

Noctor, S. C. *et al.* (2004) 'Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases', *Nature Neuroscience*, 7(2), pp. 136–144. doi: 10.1038/nn1172.

O'Roak, B. J. *et al.* (2012) 'Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations', *Nature*. Nature Publishing Group, 485(7397), pp. 246–250. doi: 10.1038/nature10989.

O'Rourke, D. H. *et al.* (1982) 'Refutation of the general single-locus model for the etiology of schizophrenia.', *American journal of human genetics*, 34(4), pp. 630–49. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7102677%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1685358.

Oguro-Ando, a *et al.* (2015) 'Increased CYFIP1 dosage alters cellular and dendritic morphology and dysregulates mTOR', *Molecular Psychiatry*. Nature Publishing Group, 20(9), pp. 1069–1078. doi: 10.1038/mp.2014.124.

Ohta, H. *et al.* (2016) 'Increased Surface Area, but not Cortical Thickness, in a Subset of Young Boys With Autism Spectrum Disorder', *Autism Research*, 9(2), pp. 232–248. doi: 10.1002/aur.1520.

Ohtsuka, T. *et al.* (2001) 'Roles of the Basic Helix-Loop-Helix Genes Hes1 and Hes5 in Expansion of Neural Stem Cells of the Developing Brain', *Journal of Biological Chemistry*, 276(32), pp. 30467–30474. doi: 10.1074/jbc.M102420200.

Oikawa, T. *et al.* (2004) 'PtdIns(3,4,5)P3 binding is necessary for WAVE2-induced formation of lamellipodia', *Nature Cell Biology*, 6(5), pp. 420–426. doi: 10.1038/ncb1125.

Oliveira, G. *et al.* (2005) 'Mitochondrial dysfunction in autism spectrum disorders: a population-based study.', *Developmental medicine and child neurology*. England, 47(3), pp. 185–189. doi: 10.1017/S0012162205000332.

Owen, M. J., Sawa, A. and Mortensen, P. B. (2016) 'Schizophrenia', *The Lancet*, 388(10039), pp. 86–97. doi: 10.1016/S0140-6736(15)01121-6.

Padrick, S. B. *et al.* (2008) 'Hierarchical Regulation of WASP/WAVE Proteins', *Molecular Cell*, 32(3), pp. 426–438. doi: 10.1016/j.molcel.2008.10.012.

Pakkenberg, B. (1990) 'Pronounced reduction of total neuron number in mediodorsal thalamic nucleus and nucleus accumbens in schizophrenics.', *Archives of general psychiatry*, 47(11), pp. 1023–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2241504 (Accessed: 24 June 2019).

Palma, V. and Ruiz i Altaba, A. (2004) 'Hedgehog-GLI signaling regulates the behavior of cells with stem cell properties in the developing neocortex.', *Development (Cambridge, England)*. The Company of Biologists Ltd, 131(2), pp. 337–45. doi: 10.1242/dev.00930.

Pan, Q. *et al.* (2008) 'Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing', *Nature Genetics*, 40(12), pp. 1413–1415. doi: 10.1038/ng.259.

Papaleo, F. *et al.* (2016) 'Behavioral, Neurophysiological, and Synaptic Impairment in a Transgenic Neuregulin1 (NRG1-IV) Murine Schizophrenia Model', *Journal of Neuroscience*, 36(17), pp. 4859–4875. doi: 10.1523/JNEUROSCI.4632-15.2016.

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.

Pathania, M. *et al.* (2014) 'The autism and schizophrenia associated gene CYFIP1 is critical for the maintenance of dendritic complexity and the stabilization of mature spines.', *Translational psychiatry*, 4(February), p. e374. doi: 10.1038/tp.2014.16.

Paton, C. *et al.* (2004) 'Obesity, dyslipidaemias and smoking in an inpatient population treated with antipsychotic drugs', *Acta Psychiatrica Scandinavica*. John Wiley & Sons, Ltd (10.1111), 110(4), pp. 299–305. doi: 10.1111/j.1600-0447.2004.00372.x.

Pepinsky, R. B. *et al.* (1998) 'Identification of a palmitic acid-modified form of human Sonic hedgehog.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 273(22), pp. 14037–45. doi: 10.1074/jbc.273.22.14037.

Pers, T. H. *et al.* (2016) 'Comprehensive analysis of schizophrenia-associated loci highlights ion channel pathways and biologically plausible candidate causal genes', *Human Molecular Genetics*, 25(6), pp. 1247–1254. doi: 10.1093/hmg/ddw007.

Picinelli, C. *et al.* (2016) 'Recurrent 15q11.2 BP1-BP2 microdeletions and microduplications in the etiology of neurodevelopmental disorders', *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*. wiley, 171(8), pp. 1088–1098. doi: 10.1002/ajmg.b.32480.

Pickrell, J. K. *et al.* (2010) 'Understanding mechanisms underlying human gene expression variation with RNA sequencing', *Nature*. Nature Publishing Group, 464(7289), pp. 768–772. doi: 10.1038/nature08872.

Pilaz, L.-J. *et al.* (2009) 'Forced G1-phase reduction alters mode of division, neuron number, and laminar phenotype in the cerebral cortex.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 106(51), pp. 21924–9. doi: 10.1073/pnas.0909894106.

Pilaz, L.-J. *et al.* (2016) 'Dynamic mRNA Transport and Local Translation in Radial Glial Progenitors of the Developing Brain.', *Current biology : CB*. Elsevier, 26(24), pp. 3383–3392. doi: 10.1016/j.cub.2016.10.040.

Pilaz, L.-J. and Silver, D. L. (2017) 'Moving messages in the developing brain-emerging roles for mRNA transport and local translation in neural stem cells', *FEBS Letters*. John Wiley & Sons, Ltd, 591(11), pp. 1526–1539. doi: 10.1002/1873-3468.12626.

Pocklington, A. J. *et al.* (2015) 'Novel Findings from CNVs Implicate Inhibitory and Excitatory Signaling Complexes in Schizophrenia', *Neuron*. The Authors, 86(5), pp. 1203–1214. doi: 10.1016/j.neuron.2015.04.022.

Polderman, T. J. C. *et al.* (2015) 'Meta-analysis of the heritability of human traits based on fifty years of twin studies', *Nature Genetics*. Nature Publishing Group, 47(7), pp. 702–709. doi: 10.1038/ng.3285.

Pollen, A. A. *et al.* (2019) 'Establishing Cerebral Organoids as Models of Human-Specific Brain Evolution.', *Cell*. Elsevier Inc., 176(4), pp. 743-756.e17. doi: 10.1016/j.cell.2019.01.017.

Poopal, A. C. *et al.* (2016) 'Increased expression of the PI3K catalytic subunit p110δ underlies elevated S6 phosphorylation and protein synthesis in an individual with autism from a multiplex family', *Molecular Autism*. Molecular Autism, 7(1), pp. 1–12. doi: 10.1186/s13229-015-0066-4.

Popovitchenko, T. *et al.* (2016) 'The RNA binding protein HuR determines the differential translation of autism-associated FoxP subfamily members in the developing neocortex', *Scientific Reports*. Nature Publishing Group, 6(1), p. 28998. doi: 10.1038/srep28998.

des Portes, V. *et al.* (1998) 'A novel CNS gene required for neuronal migration and involved in X-linked subcortical laminar heterotopia and lissencephaly syndrome.', *Cell*, 92(1), pp. 51–61. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9489699 (Accessed: 21 June 2019).

Prabakaran, S. *et al.* (2004) 'Mitochondrial dysfunction in schizophrenia: evidence for compromised brain metabolism and oxidative stress', *Molecular Psychiatry*. Nature Publishing Group, 9(7), pp. 684–697. doi: 10.1038/sj.mp.4001511.

Pramparo, T. *et al.* (no date) 'Cell cycle networks link gene expression dysregulation, mutation, and brain maldevelopment in autistic toddlers'. doi: 10.15252/msb.20156108.

Purcell, S. M. *et al.* (2014) 'A polygenic burden of rare disruptive mutations in schizophrenia', *Nature*, 506(7487), pp. 185–190. doi: 10.1038/nature12975.

Quadrato, G. *et al.* (2017) 'Cell diversity and network dynamics in photosensitive human brain organoids', *Nature*. Nature Publishing Group, 545(7652), pp. 48–53. doi: 10.1038/nature22047.

Rakic, P. (1971) 'Guidance of neurons migrating to the fetal monkey neocortex', *Brain Research*. Elsevier, 33(2), pp. 471–476. doi: 10.1016/0006-8993(71)90119-3.

Rakic, P. (1972) 'Mode of cell migration to the superficial layers of fetal monkey neocortex', *The Journal of Comparative Neurology*. John Wiley & Sons, Ltd, 145(1), pp. 61–83. doi: 10.1002/cne.901450105.

Rakic, P. (1974) 'Neurons in Rhesus Monkey Visual Cortex: Systematic Relation between Time of Origin and Eventual Disposition', *Science*, 183(4123), pp. 425–427. doi: 10.1126/science.183.4123.425.

Rakic, P. (1988) 'Specification of cerebral cortical areas', *Science*, 241(4862), pp. 170–176. doi: 10.1126/science.3291116.

Rakic, P. (1995) 'A small step for the cell, a giant leap for mankind: a hypothesis of neocortical expansion during evolution.', *Trends in neurosciences*. Elsevier, 18(9), pp. 383–8. doi: 10.1016/0166-2236(95)93934-P.

Rakic, P. (2009) 'Evolution of the neocortex: a perspective from developmental biology.', *Nature reviews. Neuroscience*. NIH Public Access, 10(10), pp. 724–35. doi:

10.1038/nrn2719.

Rank, D. *et al.* (2009) 'Real-Time DNA Sequencing from Single Polymerase Molecules', *Science*, 323(January), pp. 133–138. doi: 10.1126/science.1162986.

Rees, E. *et al.* (2014) 'Analysis of copy number variations at 15 schizophrenia-associated loci', *British Journal of Psychiatry*, 204(2), pp. 108–114. doi: 10.1192/bjp.bp.113.131052.

Regad, T. *et al.* (2007) 'The neural progenitor-specifying activity of FoxG1 is antagonistically regulated by CKI and FGF', *Nature Cell Biology*. Nature Publishing Group, 9(5), pp. 531–540. doi: 10.1038/ncb1573.

Reimand, J. *et al.* (2019) 'Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap', *Nature Protocols*, 14(2), pp. 482–517. doi: 10.1038/s41596-018-0103-9.

Reiner, O. *et al.* (1993) 'Isolation of a Miller–Dicker lissencephaly gene containing G protein β-subunit-like repeats', *Nature*, 364(6439), pp. 717–721. doi: 10.1038/364717a0.

Rekdal, C., Sjøttem, E. and Johansen, T. (2000) 'The nuclear factor SPBP contains different functional domains and stimulates the activity of various transcriptional activators.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 275(51), pp. 40288–300. doi: 10.1074/jbc.M006978200.

Rice, M. W. *et al.* (2014) 'Assessment of Cytochrome C Oxidase Dysfunction in the Substantia Nigra/Ventral Tegmental Area in Schizophrenia', *PLoS ONE*. Edited by T. D. Aumann. Public Library of Science, 9(6), p. e100054. doi: 10.1371/journal.pone.0100054.

Richter, J. D. and Sonenberg, N. (2005) 'Regulation of cap-dependent translation by eIF4E inhibitory proteins', *Nature*, 433(7025), pp. 477–480. doi: 10.1038/nature03205.

Ripke, S. *et al.* (2014) 'Biological insights from 108 schizophrenia-associated genetic loci', *Nature*, 511(7510), pp. 421–427. doi: 10.1038/nature13595.

Risch, N. and Merikangas, K. (1996) 'The Future of Genetic Studies of Complex Human Diseases', *Science*, 273(5281), pp. 1516–1517. doi: 10.1126/science.273.5281.1516.

Rita Lo Vasco, V., Cardinale, G. and Polonia, P. (2012) 'Deletion of PLCB1 gene in schizophrenia-affected patients', *Journal of Cellular and Molecular Medicine*, 16(4), pp. 844–851. doi: 10.1111/j.1582-4934.2011.01363.x.

Roberts, R. C. *et al.* (2015) 'Decreased synaptic and mitochondrial density in the

postmortem anterior cingulate cortex in schizophrenia', *Schizophrenia Research*. Elsevier, 168(1–2), pp. 543–553. doi: 10.1016/J.SCHRES.2015.07.016.

Roberts, R. C. (2017) 'Postmortem studies on mitochondria in schizophrenia', *Schizophrenia Research*. Elsevier B.V., 187, pp. 17–25. doi: 10.1016/j.schres.2017.01.056.

Rollins, B. L. *et al.* (2017) 'Mitochondrial Complex I Deficiency in Schizophrenia and Bipolar Disorder and Medication Influence', *Molecular Neuropsychiatry*, 3(3), pp. 157–169. doi: 10.1159/000484348.

Rosa, A. *et al.* (2002) '1q21-q22 locus is associated with susceptibility to the reality-distortion syndrome of schizophrenia spectrum disorders', *American Journal of Medical Genetics*, 114(5), pp. 516–518. doi: 10.1002/ajmg.10526.

Rosenfeld, J. A. *et al.* (2013) 'Estimates of penetrance for recurrent pathogenic copy-number variations', *Genetics in Medicine*. Nature Publishing Group, 15(6), pp. 478–481. doi: 10.1038/gim.2012.164.

De Rubeis, S. *et al.* (2013) 'CYFIP1 coordinates mRNA translation and cytoskeleton remodeling to ensure proper dendritic spine formation.', *Neuron*. Elsevier, 79(6), pp. 1169–82. doi: 10.1016/j.neuron.2013.06.039.

De Rubeis, S. *et al.* (2014) 'Synaptic, transcriptional and chromatin genes disrupted in autism', *Nature*. Nature Publishing Group, 515(7526), pp. 209–215. doi: 10.1038/nature13772.

Rüdin, E. (1916) 'Zur Vererbung und Neuentstehung der Dementia praecox: mit... Tab'.

Sacchetti, P. *et al.* (2009) 'Liver X Receptors and Oxysterols Promote Ventral Midbrain Neurogenesis In Vivo and in Human Embryonic Stem Cells', *Cell Stem Cell*, 5(4), pp. 409–419. doi: 10.1016/j.stem.2009.08.019.

Sadler, T. W. (2005) 'Embryology of neural tube development', *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*. John Wiley & Sons, Ltd, 135C(1), pp. 2–8. doi: 10.1002/ajmg.c.30049.

Saki, M., Toulany, M. and Rodemann, H. P. (2013) 'Acquired resistance to cetuximab is associated with the overexpression of Ras family members and the loss of radiosensitization in head and neck cancer cells', *Radiotherapy and Oncology*. Elsevier, 108(3), pp. 473–478. doi: 10.1016/J.RADONC.2013.06.023.

Sanders, A. R. *et al.* (2017) 'Transcriptome sequencing study implicates immune-related

genes differentially expressed in schizophrenia: New data and a meta-analysis', *Translational Psychiatry*, 7(4), pp. 1–10. doi: 10.1038/tp.2017.47.

Sanders, S. J. *et al.* (2015) 'Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci', *Neuron*. Cell Press, 87(6), pp. 1215–1233. doi: 10.1016/J.NEURON.2015.09.016.

Satoh, Y. *et al.* (2011) 'ERK2 Contributes to the Control of Social Behaviors in Mice', *Journal of Neuroscience*, 31(33), pp. 11953–11967. doi: 10.1523/JNEUROSCI.2349-11.2011.

Savage, J. E. *et al.* (2018) 'Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence', *Nature Genetics*, 50(7), pp. 912–919. doi: 10.1038/s41588-018-0152-6.

Schenck, a. *et al.* (2001) 'A highly conserved protein family interacting with the fragile X mental retardation protein (FMRP) and displaying selective interactions with FMRP-related proteins FXR1P and FXR2P', *Proceedings of the National Academy of Sciences*, 98(15), pp. 8844–8849. doi: 10.1073/pnas.151231598.

Schenck, A. *et al.* (2001) 'A highly conserved protein family interacting with the fragile X mental retardation protein (FMRP) and displaying selective interactions with FMRP-related proteins FXR1P and FXR2P', *Proceedings of the National Academy of Sciences*, 98(15), pp. 8844–8849. doi: 10.1073/pnas.151231598.

Schijven, D. *et al.* (2018) 'Comprehensive pathway analyses of schizophrenia risk loci point to dysfunctional postsynaptic signaling', *Schizophrenia Research*, 199, pp. 195–202. doi: 10.1016/j.schres.2018.03.032.

Schwab, S. G. *et al.* (2000) 'A genome-wide autosomal screen for schizophrenia susceptibility loci in 71 families with affected siblings: support for loci on chromosome 10p and 6', *Molecular Psychiatry*. Nature Publishing Group, 5(6), pp. 638–649. doi: 10.1038/sj.mp.4000791.

Schwab, S. G. and Wildenauer, D. B. (2013) 'Genetics of psychiatric disorders in the GWAS era: An update on schizophrenia', *European Archives of Psychiatry and Clinical Neuroscience*, 263(SUPPL.2). doi: 10.1007/s00406-013-0450-z.

Schwarz, E. *et al.* (2008) 'High Throughput Lipidomic Profiling of Schizophrenia and Bipolar Disorder Brain Tissue Reveals Alterations of Free Fatty Acids, Phosphatidylcholines, and Ceramides', *Journal of Proteome Research*. American Chemical Society, 7(10), pp. 4266–4277. doi: 10.1021/pr800188y.

Sebat, D. M. and J. (2013) ‘CNVs: Harbinger of a Rare Variant Revolution in Psychiatric Genetics’, *Cell*, 148(6), pp. 1223–1241. doi: 10.1016/j.cell.2012.02.039.CNVs.

Sebat, J. *et al.* (2007) ‘Strong Association of De Novo Copy Number Mutations with Autism’, *Science*. American Association for the Advancement of Science, 316(5823), pp. 445–449. doi: 10.1126/science.1138659.

Selemon, L. D., Rajkowska, G. and Goldman-Rakic, P. S. (1995) ‘Abnormally high neuronal density in the schizophrenic cortex. A morphometric analysis of prefrontal area 9 and occipital area 17.’, *Archives of general psychiatry*, 52(10), pp. 805–18; discussion 819-20. Available at: https://www.ncbi.nlm.nih.gov/pubmed/7575100?dopt=Abstract&holding=npg (Accessed: 24 June 2019).

Selemon, L. D. and Zecevic, N. (2015) ‘Schizophrenia: a tale of two critical periods for prefrontal cortical development’, *Translational Psychiatry*. Nature Publishing Group, 5(8), pp. e623–e623. doi: 10.1038/tp.2015.115.

Shao, L. *et al.* (2008) ‘Mitochondrial involvement in psychiatric disorders’, *Annals of Medicine*, 40(4), pp. 281–295. doi: 10.1080/07853890801923753.

Sharon, D. *et al.* (2013) ‘A single-molecule long-read survey of the human transcriptome’, *Nature Biotechnology*, 31(11), pp. 1009–1014. doi: 10.1038/nbt.2705.

Shih, R. A., Belmonte, P. L. and Zandi, P. P. (2004) ‘A review of the evidence from family, twin and adoption studies for a genetic contribution to adult psychiatric disorders’, *International Review of Psychiatry*, 16(4), pp. 260–283. doi: 10.1080/09540260400014401.

Siddiqui, M. F., Elwell, C. and Johnson, M. H. (2016) ‘Mitochondrial Dysfunction in Autism Spectrum Disorders’, *Autism-Open Access*, 6(4), pp. 1–7. doi: 10.4172/2165-7890.1000190.

Sitzmann, A. F. *et al.* (2018) ‘Rare FMR1 gene mutations causing fragile X syndrome: A review’, *American Journal of Medical Genetics Part A*, 176(1), pp. 11–18. doi: 10.1002/ajmg.a.38504.

Skafidas, E. *et al.* (2014) ‘Predicting the diagnosis of autism spectrum disorder using gene pathway analysis’, *Molecular Psychiatry*. Nature Publishing Group, 19(4), pp. 504–510. doi: 10.1038/mp.2012.126.

Skene, N. G. *et al.* (2018) ‘Genetic identification of brain cell types underlying schizophrenia’, *Nat Genet*. doi: 10.1038/s41588-018-0129-5.

Smart, I. H. M. *et al.* (2002) 'Unique Morphological Features of the Proliferative Zones and Postmitotic Compartments of the Neural Epithelium Giving Rise to Striate and Extrastriate Cortex in the Monkey', *Cerebral Cortex*, 12(1), pp. 37–53. doi: 10.1093/cercor/12.1.37.

Smith, J. L. and Schoenwolf, G. C. (1997) 'Neurulation: Coming to closure', *Trends in Neurosciences*, 20(97), pp. 510–517. doi: 10.1016/S0166-2236(97)01121-1.

Solberg, D. K. *et al.* (2016) 'Lipid profiles in schizophrenia associated with clinical traits: A five year follow-up study', *BMC Psychiatry*. BMC Psychiatry, 16(1), pp. 1–9. doi: 10.1186/s12888-016-1006-3.

Son, G. and Han, J. (2018) 'Roles of mitochondria in neuronal development', *BMB Reports*, 51(11), pp. 549–556. doi: 10.5483/BMBRep.2018.51.11.226.

St Pourcain, B. *et al.* (2014) 'Variability in the common genetic architecture of social-communication spectrum phenotypes during childhood and adolescence', *Molecular Autism*, 5(1), pp. 1–12. doi: 10.1186/2040-2392-5-18.

Stanfield, A. C. *et al.* (2008) 'Towards a neuroanatomy of autism: A systematic review and meta-analysis of structural magnetic resonance imaging studies', *European Psychiatry*. Elsevier Masson, 23(4), pp. 289–299. doi: 10.1016/J.EURPSY.2007.05.006.

Stauch, K. L., Purnell, P. R. and Fox, H. S. (2014) 'Quantitative Proteomics of Synaptic and Nonsynaptic Mitochondria: Insights for Synaptic Mitochondrial Vulnerability', *Journal of Proteome Research*, 13(5), pp. 2620–2636. doi: 10.1021/pr500295n.

Stefansson, H. *et al.* (2008) 'Large recurrent microdeletions associated with schizophrenia', *Nature*. Macmillan Publishers Limited. All rights reserved, 455(7210), pp. 232–236. doi: 10.1038/nature07229.

Steffen, A. *et al.* (2004) 'Sra-1 and Nap1 link Rac to actin assembly driving lamellipodia formation', *The EMBO Journal*, 23(4), pp. 749–759. doi: 10.1038/sj.emboj.7600084.

Straub, R. E. *et al.* (1995) 'A potential vulnerability locus for schizophrenia on chromosome 6p24–22: evidence for genetic heterogeneity', *Nature Genetics*. Nature Publishing Group, 11(3), pp. 287–293. doi: 10.1038/ng1195-287.

Su, C.-H., D, D. and Tarn, W.-Y. (2018) 'Alternative Splicing in Neurogenesis and Brain Development.', *Frontiers in molecular biosciences*. Frontiers Media SA, 5, p. 12. doi: 10.3389/fmolb.2018.00012.

Subramanian, A. *et al.* (2005) 'Gene set enrichment analysis: a knowledge-based

approach for interpreting genome-wide expression profiles.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 102(43), pp. 15545–50. doi: 10.1073/pnas.0506580102.

Sullivan, C. R. *et al.* (2018) 'Defects in Bioenergetic Coupling in Schizophrenia', *Biological Psychiatry*. Elsevier Inc, 83(9), pp. 739–750. doi: 10.1016/j.biopsych.2017.10.014.

Sunabori, T. *et al.* (2008) 'Cell-cycle-specific nestin expression coordinates with morphological changes in embryonic cortical neural progenitors', *Journal of Cell Science*, 121(8), pp. 1204–1212. doi: 10.1242/jcs.025064.

Sztainberg, Y. and Zoghbi, H. Y. (2016) 'Lessons learned from studying syndromic autism spectrum disorders', *Nature Neuroscience*, 19(11), pp. 1408–1417. doi: 10.1038/nn.4420.

Takahashi, S. *et al.* (2003) 'Significant linkage to chromosome 22q for exploratory eye movement dysfunction in schizophrenia', *American Journal of Medical Genetics*, 123B(1), pp. 27–32. doi: 10.1002/ajmg.b.10046.

Takenawa, T. and Suetsugu, S. (2007) 'The WASP–WAVE protein network: connecting the membrane to the cytoskeleton', *Nature Reviews Molecular Cell Biology*, 8(1), pp. 37–48. doi: 10.1038/nrm2069.

Tandon, R., Nasrallah, H. A. and Keshavan, M. S. (2009) 'Schizophrenia, "just the facts" 4. Clinical features and conceptualization', *Schizophrenia Research*. Elsevier B.V., 110(1–3), pp. 1–23. doi: 10.1016/j.schres.2009.03.005.

Tavazoie, S. *et al.* (1999) 'Systematic determination of genetic network architecture', *Nature Genetics*, 22(3), pp. 281–285. doi: 10.1038/10343.

Telias, M. (2019) 'Molecular Mechanisms of Synaptic Dysregulation in Fragile X Syndrome and Autism Spectrum Disorders', *Frontiers in Molecular Neuroscience*, 12(March), pp. 1–12. doi: 10.3389/fnmol.2019.00051.

Thakore, J. H. (2004) 'Metabolic disturbance in first-episode schizophrenia', *British Journal of Psychiatry*. Cambridge University Press, 184(S47), pp. s76–s79. doi: 10.1192/bjp.184.47.s76.

Thapar, A. and Cooper, M. (2013) 'Copy number variation: What is it and what has it told us about child psychiatric disorders?', *Journal of the American Academy of Child and Adolescent Psychiatry*. American Academy of Child and Adolescent Psychiatry, 52(8), pp. 772–774. doi: 10.1016/j.jaac.2013.05.013.

Thapar, A., Cooper, M. and Rutter, M. (2017) 'Neurodevelopmental disorders', *The Lancet Psychiatry*. Elsevier, 4(4), pp. 339–346. doi: 10.1016/S2215-0366(16)30376-5.

Theofilopoulos, S. *et al.* (2013) 'Brain endogenous liver X receptor ligands selectively promote midbrain neurogenesis', *Nature Chemical Biology*, 9(2), pp. 126–133. doi: 10.1038/nchembio.1156.

Tokuda, S. *et al.* (2011) 'A novel Akt3 mutation associated with enhanced kinase activity and seizure susceptibility in mice', *Human Molecular Genetics*, 20(5), pp. 988–999. doi: 10.1093/hmg/ddq544.

Topol, A. *et al.* (2015) 'Increased abundance of translation machinery in stem cell–derived neural progenitor cells from four schizophrenia patients', *Translational Psychiatry*. Nature Publishing Group, 5(10), pp. e662–e662. doi: 10.1038/tp.2015.118.

Toulopoulou, T. *et al.* (2019) 'Polygenic risk score increases schizophrenia liability through cognition-relevant pathways', *Brain*, 142(2), pp. 471–485. doi: 10.1093/brain/awy279.

Tsai, M. *et al.* (2010) 'Modification Complexes', *Science*, 329(August), pp. 689–693.

Turner, T. N. *et al.* (2016) 'Genome Sequencing of Autism-Affected Families Reveals Disruption of Putative Noncoding Regulatory DNA', *American Journal of Human Genetics*. The American Society of Human Genetics, 98(1), pp. 58–74. doi: 10.1016/j.ajhg.2015.11.023.

Twyman, R. M. (2009) 'Single-Nucleotide Polymorphism (SNP) Analysis', in *Encyclopedia of Neuroscience*. Elsevier, pp. 871–875. doi: 10.1016/B978-008045046-9.00866-4.

Udawela, M. *et al.* (2011) 'Phospholipase C beta 1 expression in the dorsolateral prefrontal cortex from patients with schizophrenia at different stages of illness', *Australian and New Zealand Journal of Psychiatry*, 45(2), pp. 140–147. doi: 10.3109/00048674.2010.533364.

Udawela, M. *et al.* (2017) 'Isoform specific differences in phospholipase C beta 1 expression in the prefrontal cortex in schizophrenia and suicide', *npj Schizophrenia*, 3(1), p. 19. doi: 10.1038/s41537-017-0020-x.

Uher, R. and Zwicker, A. (2017) 'Etiology in psychiatry: embracing the reality of poly-gene-environmental causation of mental illness', *World Psychiatry*. John Wiley & Sons, Ltd, 16(2), pp. 121–129. doi: 10.1002/wps.20436.

Umetani, M. *et al.* (2007) '27-Hydroxycholesterol is an endogenous SERM that inhibits the cardiovascular effects of estrogen', *Nature Medicine*, 13(10), pp. 1185–1192. doi: 10.1038/nm1641.

Valvezan, A. J. and Klein, P. S. (2012) 'GSK-3 and Wnt Signaling in Neurogenesis and Bipolar Disorder', *Frontiers in Molecular Neuroscience*, 5(January), pp. 1–13. doi: 10.3389/fnmol.2012.00001.

Vance, K. W. *et al.* (2014) 'The long non-coding RNA paupar regulates the expression of both local and distal genes', *EMBO Journal*, 33(4), pp. 296–311. doi: 10.1002/embj.201386225.

Vanlerberghe, C. *et al.* (2015) '15q11.2 microdeletion (BP1–BP2) and developmental delay, behaviour issues, epilepsy and congenital heart disease: A series of 52 patients', *European Journal of Medical Genetics*. Elsevier Masson SAS, 58(3), pp. 140–147. doi: 10.1016/j.ejmg.2015.01.002.

Lo Vasco, V. R., Longo, L. and Polonia, P. (2013) 'Phosphoinositide-specific Phospholipase C β1 gene deletion in bipolar disorder affected patient', *Journal of Cell Communication and Signaling*, 7(1), pp. 25–29. doi: 10.1007/s12079-012-0182-2.

Vassos, E. *et al.* (2017) 'An Examination of Polygenic Score Risk Prediction in Individuals With First-Episode Psychosis', *Biological Psychiatry*. Elsevier, 81(6), pp. 470–477. doi: 10.1016/j.biopsych.2016.06.028.

Verma, S. K. *et al.* (2009) 'Metabolic Risk Factors in Drug-Naive Patients With First-Episode Psychosis', *The Journal of Clinical Psychiatry*. Physicians Postgraduate Press, Inc., 70(7), pp. 997–1000. doi: 10.4088/JCP.08m04508.

Vita, A. *et al.* (2012) 'Progressive loss of cortical gray matter in schizophrenia: a meta-analysis and meta-regression of longitudinal MRI studies.', *Translational psychiatry*. Nature Publishing Group, 2(11), p. e190. doi: 10.1038/tp.2012.116.

Walsh, T. *et al.* (2008) 'Rare Structural Variants Disrupt Multiple Genes in Neurodevelopmental Pathways in Schizophrenia', *Science*. American Association for the Advancement of Science, 320(5875), pp. 539–543. doi: 10.1126/science.1155174.

Wang, E. T. *et al.* (2008) 'Alternative isoform regulation in human tissue transcriptomes', *Nature*, 456(7221), pp. 470–476. doi: 10.1038/nature07509.

Wang, K., Li, M. and Bucan, M. (2007) 'Pathway-Based Approaches for Analysis of Genomewide Association Studies', *The American Journal of Human Genetics*, 81(6), pp. 1278–1283. doi: 10.1086/522374.

Wang, Z., Gerstein, M. and Snyder, M. (2009) 'RNA-Seq: a revolutionary tool for transcriptomics', *Nature Reviews Genetics*, 10(1), pp. 57–63. doi: 10.1038/nrg2484.

Weiss, L. A. *et al.* (2008) 'Association between Microdeletion and Microduplication at 16p11.2 and Autism', *New England Journal of Medicine*, 358(7), pp. 667–675. doi: 10.1056/NEJMoa075974.

Weissman, J. R. *et al.* (2008) 'Mitochondrial disease in autism spectrum disorder patients: A cohort analysis', *PLoS ONE*. Edited by R. Schiffmann, 3(11), pp. 1–6. doi: 10.1371/journal.pone.0003815.

Wen, Y., Alshikho, M. J. and Herbert, M. R. (2016) 'Pathway network analyses for autism reveal multisystem involvement, major overlaps with other diseases and convergence upon MAPK and calcium signaling', *PLoS ONE*, 11(4), pp. 1–23. doi: 10.1371/journal.pone.0153329.

White, M. J. *et al.* (2019) 'Strategies for Pathway Analysis Using GWAS and WGS Data', *Current Protocols in Human Genetics*, 100(1), pp. 1–17. doi: 10.1002/cphg.79.

Wray, N. R., Goddard, M. E. and Visscher, P. M. (2007) 'Prediction of individual genetic risk to disease from genome-wide association studies', *Genome Research*, 17(10), pp. 1520–1528. doi: 10.1101/gr.6665407.

Wu, Q. *et al.* (2013) '27-Hydroxycholesterol Promotes Cell-Autonomous, ER-Positive Breast Cancer Growth', *Cell Reports*, 5(3), pp. 637–645. doi: 10.1016/j.celrep.2013.10.006.

Xiao, Y. *et al.* (2013) 'Similar and Different Gray Matter Deficits in Schizophrenia Patients and Their Unaffected Biological Relatives', *Frontiers in Psychiatry*. Frontiers, 4, p. 150. doi: 10.3389/fpsyt.2013.00150.

Xie, H. *et al.* (2014) 'Functional Study of NIPA2 Mutations Identified from the Patients with Childhood Absence Epilepsy', *PLoS ONE*. Edited by M. Seaman. Public Library of Science, 9(10), p. e109749. doi: 10.1371/journal.pone.0109749.

Xing, L. *et al.* (2016) 'Layer specific and general requirements for ERK/MAPK signaling in the developing neocortex', *eLife*, 5(FEBRUARY2016), pp. 1–29. doi: 10.7554/eLife.11123.

Xu, B. *et al.* (no date) 'Strong association of de novo copy number mutations with sporadic schizophrenia', *nature.com*. Available at: https://www.nature.com/ng/journal/v40/n7/abs/ng.162.html (Accessed: 23 June 2019).

Y., G., S., E. and R., P.-K. (2001) ‘Neuregulin induces sustained reactive oxygen species generation to mediate neuronal differentiation’, *Cellular and Molecular Neurobiology*, 21(6), pp. 753–769. doi: 0.1023/A:1015108306171.

Yamazaki, S. *et al.* (2016) ‘Nuclear F-actin enhances the transcriptional activity of β-catenin by increasing its nuclear localization and binding to chromatin’, *Histochemistry and Cell Biology*. Springer Berlin Heidelberg, 145(4), pp. 389–399. doi: 10.1007/s00418-016-1416-9.

Yeung, K. S. *et al.* (2017) ‘Identification of mutations in the PI3K-AKT-mTOR signalling pathway in patients with macrocephaly and developmental delay and/or autism’, *Molecular Autism*. Molecular Autism, 8(1), pp. 1–11. doi: 10.1186/s13229-017-0182-4.

Yi, H. *et al.* (2018) ‘Detecting hidden batch factors through data-adaptive adjustment for biological effects’, *Bioinformatics*. Edited by Z. Bar-Joseph, 34(7), pp. 1141–1147. doi: 10.1093/bioinformatics/btx635.

Ylisaukko-oja, T. *et al.* (2006) ‘Search for autism loci by combined analysis of Autism Genetic Resource Exchange and Finnish families’, *Annals of Neurology*, 59(1), pp. 145–155. doi: 10.1002/ana.20722.

Yoo, H. (2015) ‘Genetics of Autism Spectrum Disorder: Current Status and Possible Clinical Applications’, *Experimental Neurobiology*. Korean Society for Brain and Neural Science, 24(4), p. 257. doi: 10.5607/en.2015.24.4.257.

Yoon, K.-J. *et al.* (2014) ‘Modeling a Genetic Risk for Schizophrenia in iPSCs and Mice Reveals Neural Stem Cell Deficits Associated with Adherens Junctions and Polarity’, *Cell Stem Cell*. Elsevier Inc., 15(1), pp. 79–91. doi: 10.1016/j.stem.2014.05.003.

Young, K. A. *et al.* (2000) ‘Reduced number of mediodorsal and anterior thalamic neurons in schizophrenia’, *Biological Psychiatry*. Elsevier, 47(11), pp. 944–953. doi: 10.1016/S0006-3223(00)00826-X.

Yuan, P. *et al.* (2010) ‘Altered levels of extracellular signal-regulated kinase signaling proteins in postmortem frontal cortex of individuals with mood disorders and schizophrenia’, *Journal of Affective Disorders*. Elsevier B.V., 124(1–2), pp. 164–169. doi: 10.1016/j.jad.2009.10.017.

Yue, Y. *et al.* (2016) ‘Regional Abnormality of Grey Matter in Schizophrenia: Effect from the Illness or Treatment?’, *PloS one*. Public Library of Science, 11(1), p. e0147204. doi: 10.1371/journal.pone.0147204.

Zecevic, N., Chen, Y. and Filipovic, R. (2005) ‘Contributions of cortical subventricular

zone to the development of the human cerebral cortex', *The Journal of Comparative Neurology*, 491(2), pp. 109–122. doi: 10.1002/cne.20714.

Zhang, J. *et al.* (2010) 'Cortical neural precursors inhibit their own differentiation via N-cadherin maintenance of beta-catenin signaling.', *Developmental cell*. NIH Public Access, 18(3), pp. 472–9. doi: 10.1016/j.devcel.2009.12.025.

Zhang, J. *et al.* (2013) 'AKT activation by N-cadherin regulates beta-catenin signaling and neuronal differentiation during cortical development', *Neural Development*, 8(1), p. 7. doi: 10.1186/1749-8104-8-7.

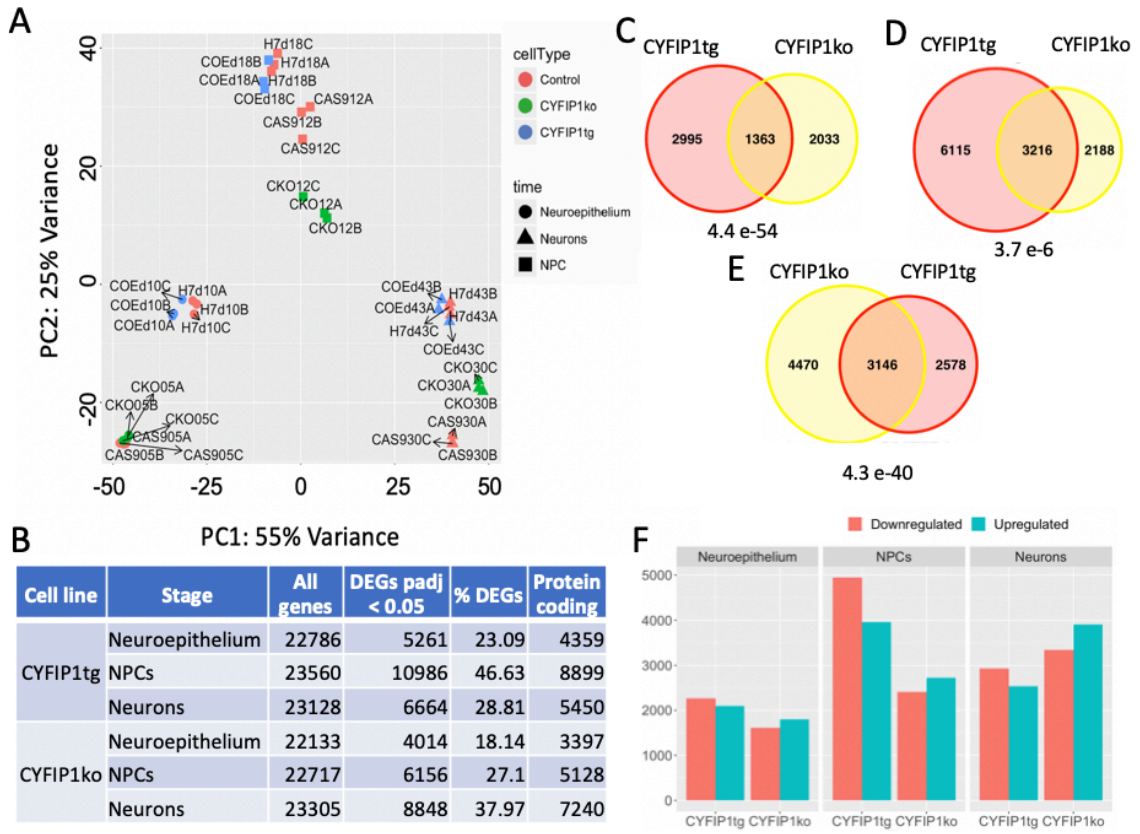
Zheng, Y. *et al.* (2010) 'Protein tyrosine kinase 6 directly phosphorylates AKT and promotes AKT activation in response to epidermal growth factor.', *Molecular and cellular biology*. American Society for Microbiology (ASM), 30(17), pp. 4280–92. doi: 10.1128/MCB.00024-10.

Zou, H. *et al.* (2011) 'Association of upregulated Ras/Raf/ERK1/2 signaling with autism', *Genes, Brain and Behavior*, 10(5), pp. 615–624. doi: 10.1111/j.1601-183X.2011.00702.x.

van der Zwaag, B. *et al.* (2009) 'A co-segregating microduplication of chromosome 15q11.2 pinpoints two risk genes for autism spectrum disorder', *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 9999B(4), p. n/a-n/a. doi: 10.1002/ajmg.b.31055.

Appendix I:

Supplementary tables and figures



Cell line	Stage	All genes	DEGs padj < 0.05	% DEGs	Protein coding
CYFIP1tg	Neuroepithelium	22786	5261	23.09	4359
	NPCs	23560	10986	46.63	8899
	Neurons	23128	6664	28.81	5450
CYFIP1ko	Neuroepithelium	22133	4014	18.14	3397
	NPCs	22717	6156	27.1	5128
	Neurons	23305	8848	37.97	7240

Supplementary Figure 1. Differential gene expression in CYFIP1tg and CYFIP1ko cells. *(A) PCA plot of the CYFIP1tg and CYFIP1ko and their isogenic control parental lines. Samples clustered by stage of differentiation and cell line. (B) Number of expressed genes and significant DEGs at 10% FDR after DESeq2 analysis comparing each CYFIP1tg and CYFIP1ko time-points with their respective isogenic control parental line. The percentage of DEGs indicates the proportion of differentially expressed genes over the total amount of transcripts sequenced for that sample. Overlapping DEGs between CYFIP1tg and CYFIP1ko for neuroepithelium (C), NPCs (D) and neurons (E). Neuroepithelial cells and NPCs shared a significant overlap of DEGs (hypergeometric p-value overall genes tested for that developmental stage). (F) The directionality of the fold change of the genes commonly expressed between CYFIP1tg and CYFIP1ko.*

Supplementary Table 1. Summary of the statistical enrichment for FMRP genes amongs DEGs

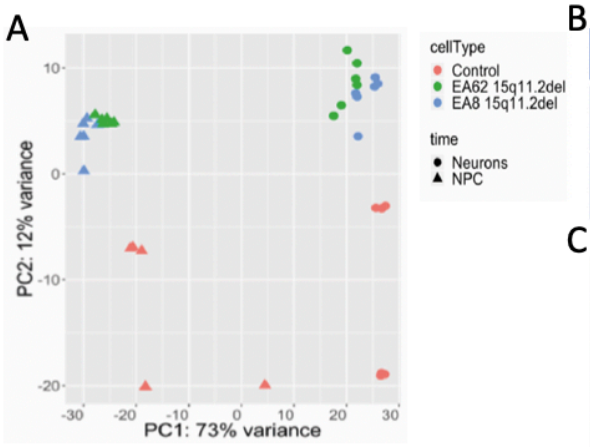
Cell line	Stage	DEG	FMRP targeted	No FMRP targeted	Fisher's pval
CYFIP1tg	Neuroepithelium	pval < 0.05	360	3999	5.79E-18
		pval > 0.05	479	10014	
	NPCs	pval < 0.05	561	8338	5.11E-08
		pval > 0.05	278	6142	
	Neurons	pval < 0.05	400	5050	1.33E-11
		pval > 0.05	439	8967	
CYFIP1ko	Neuroepithelium	pval < 0.05	282	3115	2.40E-10
		pval > 0.05	557	10019	
	NPCs	pval < 0.05	373	4755	1.24E-10
		pval > 0.05	466	9422	
	Neurons	pval < 0.05	512	6728	2.03E-14
		pval > 0.05	327	7415	

Fisher exact test p-values for enrichment for FMRP genes in CYFIP1tg and CYFIP1ko cells. All time-points tested were significant for both cell lines. DEG indicates the adjusted p value threshold for the DEG cut-off.

Supplementary Table 2. Summary of the statistical enrichment for DDD genes amongst DEGs

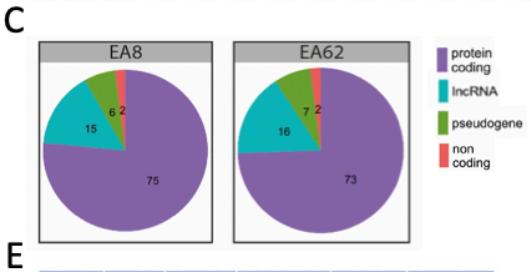
Cell line	Stage	DDD gene	No DDD gene	DEG	Fisher's pval
CYFIP1tg	Neuroepithelium	40	4319	pval < 0.05	0.03083576
		53	10440	pval > 0.05	
	NPCs	63	8836	pval < 0.05	0.04647156
		30	6390	pval > 0.05	
	Neurons	54	5396	pval < 0.05	0.00026202
		39	9367	pval > 0.05	
CYFIP1ko	Neuroepithelium	31	3366	pval < 0.05	3.08E-02
		62	10514	pval > 0.05	
	NPCs	40	5088	pval < 0.05	4.65E-02
		53	9835	pval > 0.05	
	Neurons	62	7178	pval < 0.05	2.62E-04
		31	7711	pval > 0.05	

Fisher exact test p-values for enrichment for DDD genes in CYFIP1tg and CYFIP1ko cells. All time-points tested were significant for both cell lines. DEG indicates the adjusted p value threshold for the DEG cut-off.



B

Cell line	Stage	All genes	DEGs padj < 0.05	% DEGs	Protein coding
EA8	NPCs	24222	4338	17.91	3239
	Neurons	24855	4937	19.86	3666
EA62	NPCs	24650	4320	17.53	3045
	Neurons	25314	7170	28.32	5306



D

Cell line	Stage	FMRP targeted	No FMRP targeted	DEG	Fisher's pval
EA8	NPCs	218	3021	pval < 0.05	0.00087366
		621	11188	pval > 0.05	
	Neurons	181	3485	pval < 0.05	0.93962577
		658	11134	pval > 0.05	
EA62	NPCs	155	2890	pval < 0.05	0.90038451
		684	11394	pval > 0.05	
	Neurons	321	4985	pval < 0.05	0.00570907
		518	9716	pval > 0.05	

E

Cell line	Stage	DDD gene	No DDD gene	DEG	Fisher's pval
EA8	NPCs	15	3224	pval < 0.05	0.9227572
		78	11731	pval > 0.05	
	Neurons	21	3645	pval < 0.05	0.64123409
		72	11720	pval > 0.05	
EA62	NPCs	11	3034	pval < 0.05	0.9878516
		82	11996	pval > 0.05	
	Neurons	38	5268	pval < 0.05	0.10480915
		55	10179	pval > 0.05	

Supplementary Figure 2. Differential gene expression results for EA8 and EA62 lines. *(A) PCA plot for EA8 and EA62 and control iPSC lines. Samples clustered by developmental stage and cell line. (B) Table summarising the number of transcripts identified for each cell line and developmental stage, and the number of DEGs at 10% FDR (BH p adj < 0.05). Percentage of DEGs expressed over the total number of transcripts characterised. (C) For each cell line, the average percentage of each gene biotype identified amongst significant DEGs in each time-point. Protein coding genes represent 75% of the total DEGs in patient-derived iPSCs. (D) Table summarising the enrichment results tests for FMRP and DDD associated genes amongst the DEGs in cells carrying a deletion in the 15q11.2 region. Only one time-point for each cell line was found to be significantly enriched for FMRP target genes amongst the DEGs (Fisher exact p < 0.05), whereas none of the sets tested were enriched for DDD genes.*

Supplementary Table 3. Significantly enriched KEGG pathways in CYFIP1tg and CYFIP1ko cells.

Description	set-Size	enrichmentScore	NES	pvalue	p.adjust	qvalues	Cell Type	Stage
Thermogenesis	77	0.280122524	2.34499199	0.00160256	0.02581166	0.02122669	CYFIP1ko	NE
Huntington disease	69	0.300693986	2.39221952	0.00162866	0.02581166	0.02122669	CYFIP1ko	NE
Alzheimer disease	63	0.323571971	2.4907692	0.00165563	0.02581166	0.02122669	CYFIP1ko	NE
Oxidative phosphorylation	57	0.408577626	3.05456974	0.00167504	0.02581166	0.02122669	CYFIP1ko	NE
Non-alcoholic fatty liver disease (NAFLD)	52	0.34171171	2.44456758	0.00170068	0.02581166	0.02122669	CYFIP1ko	NE
Parkinson disease	49	0.422820958	2.96069823	0.00170358	0.02581166	0.02122669	CYFIP1ko	NE
Ribosome	47	0.424202375	2.9102571	0.0017094	0.02581166	0.02122669	CYFIP1ko	NE
Neuroactive ligand-receptor interaction	27	0.48458751	2.71137536	0.00176991	0.02581166	0.02122669	CYFIP1ko	NE
Cytokine-cytokine receptor interaction	21	0.518655897	2.56622927	0.00181488	0.02581166	0.02122669	CYFIP1ko	NE
Calcium signaling pathway	31	0.383626866	2.25484769	0.0069808	0.07843137	0.06449948	CYFIP1ko	NE
Circadian entrainment	22	0.429238538	2.16651208	0.00720721	0.07843137	0.06449948	CYFIP1ko	NE
Rheumatoid arthritis	20	0.446130704	2.1517142	0.00735294	0.07843137	0.06449948	CYFIP1ko	NE
Cardiac muscle contraction	26	0.392255098	2.15265424	0.008881	0.08744364	0.07191089	CYFIP1ko	NE
MAPK signaling pathway	71	0.246428037	1.97619111	0.00978793	0.08948963	0.07359345	CYFIP1ko	NE
cAMP signaling pathway	40	0.319837625	2.04465116	0.01415929	0.12082596	0.09936345	CYFIP1ko	NE
Dilated cardiomyopathy (DCM)	22	0.387165455	1.95415499	0.01621622	0.12972973	0.10668563	CYFIP1ko	NE
Hepatitis C	25	-0.33950731	-1.9155401	0.02	0.15058824	0.12383901	CYFIP1ko	NE
Ubiquitin mediated proteolysis	34	-0.274603726	-1.7992499	0.02540416	0.17483154	0.14377594	CYFIP1ko	NE
Human T-cell leukemia virus 1 infection	55	0.250599205	1.82312688	0.02595156	0.17483154	0.14377594	CYFIP1ko	NE
PI3K-Akt signaling pathway	74	0.218558691	1.80442083	0.0273752	0.17520129	0.14408001	CYFIP1ko	NE
Signaling pathways regulating pluripo-tency of stem cells	28	0.314801831	1.77141441	0.03041145	0.18536502	0.15243834	CYFIP1ko	NE

ECM-receptor interaction	25	0.32941138	1.7614362	0.03442029	0.20026351	0.16469038	CYFIP1ko	NE
Retrograde endocannabinoid signaling	43	0.263675148	1.73658628	0.03664922	0.20396085	0.16773096	CYFIP1ko	NE
Cell adhesion molecules (CAMs)	32	0.297342353	1.76358544	0.04006969	0.20401417	0.16777481	CYFIP1ko	NE
Ras signaling pathway	54	0.235212152	1.69932006	0.04310345	0.20401417	0.16777481	CYFIP1ko	NE
Gap junction	22	0.324848314	1.63961931	0.04324324	0.20401417	0.16777481	CYFIP1ko	NE
Pyrimidine metabolism	20	0.332151382	1.60198533	0.04411765	0.20401417	0.16777481	CYFIP1ko	NE
Focal adhesion	62	0.217621885	1.67096826	0.0446281	0.20401417	0.16777481	CYFIP1ko	NE
PI3K-Akt signaling pathway	136	0.378905265	2.08219278	0.00195313	0.18893528	0.18019998	CYFIP1ko	NPC
Melanogenesis	50	-0.578303138	-2.5136655	0.00208768	0.18893528	0.18019998	CYFIP1ko	NPC
Pathways in cancer	206	0.332543249	1.97752897	0.00394477	0.22445437	0.21407686	CYFIP1ko	NPC
Melanoma	35	0.573386102	2.37150431	0.00575816	0.22445437	0.21407686	CYFIP1ko	NPC
Ribosome	75	0.435054174	2.11719907	0.00792079	0.22445437	0.21407686	CYFIP1ko	NPC
Transcriptional misregulation in cancer	68	0.461744491	2.19433027	0.00795229	0.22445437	0.21407686	CYFIP1ko	NPC
Regulation of actin cytoskeleton	86	0.408844915	2.07852525	0.00984252	0.22445437	0.21407686	CYFIP1ko	NPC
Breast cancer	72	0.428196065	2.06176565	0.00992064	0.22445437	0.21407686	CYFIP1ko	NPC
Gap junction	46	-0.476570841	-2.0291365	0.01234568	0.24828532	0.23680601	CYFIP1ko	NPC
Ras signaling pathway	84	0.389610478	1.95715309	0.01388889	0.25138889	0.23976608	CYFIP1ko	NPC
Hematopoietic cell lineage	24	0.570672095	2.11524973	0.01912046	0.28116505	0.26816556	CYFIP1ko	NPC
Signaling pathways regulating pluripo-tency of stem cells	66	-0.41007941	-1.910115	0.02061856	0.28116505	0.26816556	CYFIP1ko	NPC
Calcium signaling pathway	65	-0.408111685	-1.8996618	0.02066116	0.28116505	0.26816556	CYFIP1ko	NPC
Neuroactive ligand-receptor interaction	66	-0.395084621	-1.8402706	0.02268041	0.28116505	0.26816556	CYFIP1ko	NPC
MAPK signaling pathway	126	0.328075161	1.77648304	0.02330097	0.28116505	0.26816556	CYFIP1ko	NPC
Gastric cancer	74	0.38915721	1.88223253	0.02584493	0.29237078	0.2788532	CYFIP1ko	NPC
Viral myocarditis	23	0.48890745	1.79156077	0.0440613	0.46912328	0.44743366	CYFIP1ko	NPC

PI3K-Akt signaling pathway	170	0.33871449	2.19159499	0.00111359	0.02377822	0.01880111	CYFIP1ko	N
Focal adhesion	110	0.364669883	2.19157524	0.0011655	0.02377822	0.01880111	CYFIP1ko	N
Neuroactive ligand-receptor interaction	93	0.555973648	3.22412885	0.00118343	0.02377822	0.01880111	CYFIP1ko	N
Calcium signaling pathway	79	0.389733141	2.18800277	0.00119474	0.02377822	0.01880111	CYFIP1ko	N
Glutamatergic synapse	69	0.40028534	2.16935856	0.00122549	0.02377822	0.01880111	CYFIP1ko	N
Morphine addiction	60	0.462352333	2.43213814	0.00123153	0.02377822	0.01880111	CYFIP1ko	N
Hypertrophic cardiomyopathy (HCM)	51	0.435535163	2.15707865	0.001287	0.02377822	0.01880111	CYFIP1ko	N
Protein digestion and absorption	47	0.542369373	2.62349987	0.00129366	0.02377822	0.01880111	CYFIP1ko	N
Amoebiasis	46	0.475261728	2.27900991	0.00129702	0.02377822	0.01880111	CYFIP1ko	N
ECM-receptor interaction	44	0.616011531	2.92751251	0.00130719	0.02377822	0.01880111	CYFIP1ko	N
Systemic lupus erythematosus	28	0.601495361	2.53319834	0.00140253	0.02377822	0.01880111	CYFIP1ko	N
Nicotine addiction	25	0.624170055	2.52187857	0.00141844	0.02377822	0.01880111	CYFIP1ko	N
Complement and coagulation cascades	23	0.627917119	2.4710745	0.0014245	0.02377822	0.01880111	CYFIP1ko	N
Retrograde endocannabinoid signaling	91	0.386103052	2.22244979	0.0023753	0.03223411	0.02548707	CYFIP1ko	N
Oxidative phosphorylation	70	0.377810343	2.0622897	0.00245098	0.03223411	0.02548707	CYFIP1ko	N
Dilated cardiomyopathy (DCM)	55	0.442760174	2.26428569	0.00249377	0.03223411	0.02548707	CYFIP1ko	N
GABAergic synapse	53	0.421280259	2.12614334	0.00252525	0.03223411	0.02548707	CYFIP1ko	N
Hematopoietic cell lineage	21	0.591772318	2.25940815	0.00287356	0.0346424	0.02739128	CYFIP1ko	N
MAPK signaling pathway	151	0.301004799	1.90323864	0.00337458	0.03854124	0.03047403	CYFIP1ko	N
Cell adhesion molecules (CAMs)	58	0.390088065	2.03129869	0.00494438	0.05147364	0.04069951	CYFIP1ko	N
Melanogenesis	56	0.386425041	1.98376434	0.00498132	0.05147364	0.04069951	CYFIP1ko	N
Cytokine-cytokine receptor interaction	70	0.359421207	1.96191202	0.00612745	0.0584559	0.04622029	CYFIP1ko	N
Estrogen signaling pathway	61	0.371253888	1.94745992	0.00619579	0.0584559	0.04622029	CYFIP1ko	N
Rheumatoid arthritis	27	0.495523212	2.06925785	0.00699301	0.06086957	0.04812874	CYFIP1ko	N
Regulation of lipolysis in adipocytes	28	0.504969075	2.12667778	0.00701262	0.06086957	0.04812874	CYFIP1ko	N

Pathways in cancer	277	0.244484665	1.68546085	0.00862069	0.0719496	0.05688957	CYFIP1ko	N
Leukocyte transendothelial migration	49	0.39377812	1.92986217	0.01023018	0.08222033	0.06501049	CYFIP1ko	N
Parkinson disease	83	0.325087312	1.83429265	0.01079137	0.08363309	0.06612755	CYFIP1ko	N
AGE-RAGE signaling pathway in diabetic complications	59	0.3578329	1.86758138	0.01243781	0.09306914	0.0735885	CYFIP1ko	N
Platelet activation	54	0.361661523	1.82604425	0.01643489	0.11887906	0.09399605	CYFIP1ko	N
cAMP signaling pathway	100	0.293626953	1.73206834	0.02112676	0.13929603	0.11013948	CYFIP1ko	N
Hedgehog signaling pathway	26	-0.603591178	-3.0256453	0.02120141	0.13929603	0.11013948	CYFIP1ko	N
DNA replication	27	-0.560455752	-2.8674221	0.02439024	0.13929603	0.11013948	CYFIP1ko	N
Taste transduction	21	0.494958969	1.88977128	0.02442529	0.13929603	0.11013948	CYFIP1ko	N
TGF-beta signaling pathway	50	0.35842284	1.76080564	0.02448454	0.13929603	0.11013948	CYFIP1ko	N
Homologous recombination	26	-0.489880787	-2.4556447	0.02473498	0.13929603	0.11013948	CYFIP1ko	N
Fanconi anemia pathway	26	-0.490257343	-2.4575323	0.02473498	0.13929603	0.11013948	CYFIP1ko	N
Lysine degradation	29	-0.360794077	-1.8723688	0.0248227	0.13929603	0.11013948	CYFIP1ko	N
Bile secretion	26	0.454652106	1.88108881	0.02503477	0.13929603	0.11013948	CYFIP1ko	N
Cocaine addiction	31	0.424895847	1.82359228	0.02797203	0.1517011	0.119948	CYFIP1ko	N
Base excision repair	20	-0.440716185	-2.0887427	0.02866242	0.1517011	0.119948	CYFIP1ko	N
Basal cell carcinoma	38	-0.424705467	-2.4590089	0.03225807	0.16666667	0.13178107	CYFIP1ko	N
Ras signaling pathway	115	0.269628687	1.63414471	0.03333333	0.16821705	0.13300694	CYFIP1ko	N
Human papillomavirus infection	168	0.238387167	1.54064549	0.03674833	0.17777518	0.14056441	CYFIP1ko	N
Leishmaniasis	26	0.434460951	1.79754942	0.03755216	0.17777518	0.14056441	CYFIP1ko	N
Melanoma	36	0.397353198	1.79216434	0.03768506	0.17777518	0.14056441	CYFIP1ko	N
Dopaminergic synapse	80	0.302991264	1.70295647	0.03956835	0.18268789	0.14444883	CYFIP1ko	N
Rap1 signaling pathway	106	0.272070682	1.6214681	0.04084014	0.18463147	0.14598559	CYFIP1ko	N
Phospholipase D signaling pathway	76	0.302671961	1.67522934	0.04227053	0.18719807	0.14801497	CYFIP1ko	N

Olfactory transduction	27	0.419539415	1.75195673	0.04475525	0.19423776	0.15358116	CYFIP1ko	N
Nucleotide excision repair	24	-0.349264258	-1.7300767	0.04590164	0.19527328	0.15439992	CYFIP1ko	N
Relaxin signaling pathway	74	0.295261812	1.63734239	0.04796163	0.19527328	0.15439992	CYFIP1ko	N
Alcoholism	74	0.295625249	1.6393578	0.04796163	0.19527328	0.15439992	CYFIP1ko	N
Serotonergic synapse	52	0.331792326	1.65383777	0.04859335	0.19527328	0.15439992	CYFIP1ko	N
MAPK signaling pathway	83	-0.286685598	-2.4899417	0.00155039	0.00982987	0.00437767	CYFIP1tg	NE
Regulation of actin cytoskeleton	68	-0.284362812	-2.2858903	0.0015873	0.00982987	0.00437767	CYFIP1tg	NE
Axon guidance	70	-0.326585487	-2.6684555	0.00158983	0.00982987	0.00437767	CYFIP1tg	NE
Focal adhesion	74	-0.35724039	-2.9535614	0.00159236	0.00982987	0.00437767	CYFIP1tg	NE
Proteoglycans in cancer	65	-0.282267193	-2.2293211	0.00159236	0.00982987	0.00437767	CYFIP1tg	NE
Rap1 signaling pathway	63	-0.329906854	-2.5640272	0.00162075	0.00982987	0.00437767	CYFIP1tg	NE
Neuroactive ligand-receptor interaction	48	-0.458519423	-3.1832102	0.00169492	0.00982987	0.00437767	CYFIP1tg	NE
Dopaminergic synapse	42	-0.41584325	-2.7634514	0.0017301	0.00982987	0.00437767	CYFIP1tg	NE
Cell adhesion molecules (CAMs)	37	-0.423983957	-2.6827732	0.00175131	0.00982987	0.00437767	CYFIP1tg	NE
Platelet activation	37	-0.351021395	-2.2211	0.00175131	0.00982987	0.00437767	CYFIP1tg	NE
Circadian entrainment	37	-0.467774964	-2.9598623	0.00175131	0.00982987	0.00437767	CYFIP1tg	NE
Relaxin signaling pathway	38	-0.365538032	-2.3446685	0.00175131	0.00982987	0.00437767	CYFIP1tg	NE
Calcium signaling pathway	39	-0.402926459	-2.617061	0.00175439	0.00982987	0.00437767	CYFIP1tg	NE
Arrhythmogenic right ventricular cardi-omyopathy (ARVC)	30	-0.481960468	-2.8053889	0.00175747	0.00982987	0.00437767	CYFIP1tg	NE
Endocrine resistance	31	-0.368476636	-2.1605911	0.00176367	0.00982987	0.00437767	CYFIP1tg	NE
Aldosterone synthesis and secretion	31	-0.369298994	-2.1654131	0.00176367	0.00982987	0.00437767	CYFIP1tg	NE
AGE-RAGE signaling pathway in diabetic complications	31	-0.387300162	-2.2709643	0.00176367	0.00982987	0.00437767	CYFIP1tg	NE
Dilated cardiomyopathy (DCM)	31	-0.436174681	-2.5575438	0.00176367	0.00982987	0.00437767	CYFIP1tg	NE
Glutamatergic synapse	35	-0.435786831	-2.6834243	0.00176991	0.00982987	0.00437767	CYFIP1tg	NE

ErbB signaling pathway	32	-0.353287147	-2.098029	0.0017762	0.00982987	0.00437767	CYFIP1tg	NE
Hypertrophic cardiomyopathy (HCM)	28	-0.417960435	-2.3578	0.00178253	0.00982987	0.00437767	CYFIP1tg	NE
Osteoclast differentiation	26	-0.411115738	-2.262742	0.00178571	0.00982987	0.00437767	CYFIP1tg	NE
Leukocyte transendothelial migration	26	-0.379576241	-2.0891516	0.00178571	0.00982987	0.00437767	CYFIP1tg	NE
Inflammatory mediator regulation of TRP channels	26	-0.443431804	-2.4406066	0.00178571	0.00982987	0.00437767	CYFIP1tg	NE
Adherens junction	34	-0.37604482	-2.2730023	0.00179856	0.00982987	0.00437767	CYFIP1tg	NE
Long-term potentiation	24	-0.471585187	-2.4927807	0.00179856	0.00982987	0.00437767	CYFIP1tg	NE
Serotonergic synapse	24	-0.396734412	-2.0971224	0.00179856	0.00982987	0.00437767	CYFIP1tg	NE
Amphetamine addiction	24	-0.456412063	-2.4125761	0.00179856	0.00982987	0.00437767	CYFIP1tg	NE
Salivary secretion	22	-0.412783745	-2.1050526	0.00184502	0.00982987	0.00437767	CYFIP1tg	NE
Amoebiasis	20	-0.480448297	-2.3463191	0.00189036	0.00982987	0.00437767	CYFIP1tg	NE
Basal cell carcinoma	25	0.393140841	2.24757866	0.00226757	0.01002506	0.0044646	CYFIP1tg	NE
Proteasome	32	0.485781098	3.08327614	0.0022779	0.01002506	0.0044646	CYFIP1tg	NE
Ribosome biogenesis in eukaryotes	35	0.42657685	2.81207104	0.00228833	0.01002506	0.0044646	CYFIP1tg	NE
Pyrimidine metabolism	41	0.363367201	2.57702806	0.00231482	0.01002506	0.0044646	CYFIP1tg	NE
Parkinson disease	46	0.44895098	3.39315674	0.00240964	0.01002506	0.0044646	CYFIP1tg	NE
Oxidative phosphorylation	49	0.494253685	3.80196653	0.00246914	0.01002506	0.0044646	CYFIP1tg	NE
Ribosome	55	0.557156134	4.48294325	0.0025	0.01002506	0.0044646	CYFIP1tg	NE
Spliceosome	56	0.338412528	2.75159318	0.0025	0.01002506	0.0044646	CYFIP1tg	NE
RNA transport	57	0.392977967	3.23325219	0.00250627	0.01002506	0.0044646	CYFIP1tg	NE
Alzheimer disease	59	0.314054434	2.60490695	0.00260417	0.01015625	0.00452303	CYFIP1tg	NE
Huntington disease	71	0.296166419	2.6205934	0.00268097	0.01020075	0.00454284	CYFIP1tg	NE
Thermogenesis	75	0.212754075	1.91922921	0.0027933	0.0103751	0.00462049	CYFIP1tg	NE
cAMP signaling pathway	57	-0.293203141	-2.1773428	0.00331675	0.0113906	0.00507274	CYFIP1tg	NE

Adrenergic signaling in cardiomyocytes	45	-0.315477137	-2.1374663	0.00339559	0.0113906	0.00507274	CYFIP1tg	NE
Oxytocin signaling pathway	45	-0.341966726	-2.3169424	0.00339559	0.0113906	0.00507274	CYFIP1tg	NE
Parathyroid hormone synthesis, secretion and action	33	-0.357526331	-2.1383929	0.00355872	0.0113906	0.00507274	CYFIP1tg	NE
Vascular smooth muscle contraction	25	-0.368605354	-1.982709	0.00356506	0.0113906	0.00507274	CYFIP1tg	NE
Gap junction	28	-0.376137716	-2.1218695	0.00356506	0.0113906	0.00507274	CYFIP1tg	NE
Phosphatidylinositol signaling system	29	-0.364571534	-2.0760905	0.00357782	0.0113906	0.00507274	CYFIP1tg	NE
Protein digestion and absorption	22	-0.403057352	-2.0554514	0.00369004	0.01151292	0.00512721	CYFIP1tg	NE
mRNA surveillance pathway	31	0.327648869	2.06784747	0.0045977	0.01406356	0.00626312	CYFIP1tg	NE
FoxO signaling pathway	38	-0.338462915	-2.1710007	0.00525394	0.01551879	0.0069112	CYFIP1tg	NE
GnRH signaling pathway	27	-0.378027701	-2.107565	0.00527241	0.01551879	0.0069112	CYFIP1tg	NE
Notch signaling pathway	20	-0.411847831	-2.0113016	0.00567108	0.01638311	0.00729612	CYFIP1tg	NE
Pathways in cancer	154	-0.180598138	-1.9096238	0.0057971	0.01644269	0.00732265	CYFIP1tg	NE
Thyroid hormone signaling pathway	42	-0.301946097	-2.0065574	0.00692042	0.01917758	0.00854062	CYFIP1tg	NE
Autophagy - animal	36	-0.324543141	-2.0244282	0.00702988	0.01917758	0.00854062	CYFIP1tg	NE
Bacterial invasion of epithelial cells	28	-0.336097042	-1.8959919	0.00713013	0.01917758	0.00854062	CYFIP1tg	NE
Prostate cancer	30	-0.332877301	-1.9376076	0.00878735	0.02323434	0.01034728	CYFIP1tg	NE
Non-alcoholic fatty liver disease (NAFLD)	50	0.263171028	2.05525631	0.00966184	0.02512077	0.01118739	CYFIP1tg	NE
Estrogen signaling pathway	37	-0.309244333	-1.9567542	0.01050788	0.026	0.01157895	CYFIP1tg	NE
Phospholipase D signaling pathway	36	-0.308875779	-1.9266986	0.01054482	0.026	0.01157895	CYFIP1tg	NE
Cholinergic synapse	36	-0.303108093	-1.8907211	0.01054482	0.026	0.01157895	CYFIP1tg	NE
Purine metabolism	64	0.217578223	1.84618444	0.01066667	0.026	0.01157895	CYFIP1tg	NE
Ras signaling pathway	61	-0.24127429	-1.8498301	0.01141925	0.0274062	0.01220519	CYFIP1tg	NE
Chemokine signaling pathway	38	-0.307391008	-1.9716964	0.01225919	0.02864413	0.01275649	CYFIP1tg	NE
Hepatitis C	27	-0.327837376	-1.8277459	0.01230229	0.02864413	0.01275649	CYFIP1tg	NE

Renal cell carcinoma	26	-0.33728563	-1.8563881	0.0125	0.02867647	0.0127709	CYFIP1tg	NE
T cell receptor signaling pathway	22	-0.356125723	-1.8161165	0.01291513	0.02919942	0.01300379	CYFIP1tg	NE
Choline metabolism in cancer	32	-0.309609497	-1.8386452	0.01420959	0.03133238	0.01395369	CYFIP1tg	NE
Insulin secretion	25	-0.345200429	-1.8568152	0.01426025	0.03133238	0.01395369	CYFIP1tg	NE
MicroRNAs in cancer	51	-0.260360403	-1.847728	0.01517707	0.03288364	0.01464454	CYFIP1tg	NE
Cytokine-cytokine receptor interaction	35	0.270619419	1.78397171	0.01601831	0.0342309	0.01524453	CYFIP1tg	NE
Fluid shear stress and atherosclerosis	43	-0.271692249	-1.8102759	0.01929825	0.04068279	0.01811784	CYFIP1tg	NE
Human cytomegalovirus infection	53	-0.24782923	-1.7873247	0.02003339	0.04166945	0.01855725	CYFIP1tg	NE
Gastric acid secretion	24	-0.346914919	-1.8337785	0.02158273	0.0443014	0.01972937	CYFIP1tg	NE
Tight junction	40	-0.266082939	-1.7328702	0.02464789	0.04875762	0.02171392	CYFIP1tg	NE
RNA degradation	34	0.267092836	1.74501221	0.02466368	0.04875762	0.02171392	CYFIP1tg	NE
ECM-receptor interaction	31	-0.295451092	-1.7324002	0.02469136	0.04875762	0.02171392	CYFIP1tg	NE
JAK-STAT signaling pathway	25	-0.324876405	-1.7474933	0.02673797	0.05213904	0.02321981	CYFIP1tg	NE
Neurotrophin signaling pathway	38	-0.273052026	-1.751436	0.02802102	0.0539664	0.02403362	CYFIP1tg	NE
Hippo signaling pathway	62	0.206503132	1.7538407	0.02864583	0.05449695	0.0242699	CYFIP1tg	NE
Glycerophospholipid metabolism	21	-0.352990217	-1.7566981	0.02962963	0.05568942	0.02480096	CYFIP1tg	NE
PI3K-Akt signaling pathway	88	-0.185712954	-1.6394804	0.03301887	0.06132076	0.02730884	CYFIP1tg	NE
Chagas disease (American trypanosomiasis)	23	-0.318856308	-1.6559358	0.03683241	0.06759831	0.03010451	CYFIP1tg	NE
Inositol phosphate metabolism	21	-0.332864394	-1.6565395	0.04074074	0.07390181	0.03291174	CYFIP1tg	NE
Insulin signaling pathway	40	-0.252983609	-1.6475606	0.04225352	0.07494457	0.03337612	CYFIP1tg	NE
Wnt signaling pathway	60	-0.213941659	-1.6272938	0.04227642	0.07494457	0.03337612	CYFIP1tg	NE
TGF-beta signaling pathway	29	0.266614521	1.63800292	0.04514673	0.07913359	0.03524168	CYFIP1tg	NE
C-type lectin receptor signaling pathway	23	-0.312746597	-1.6242058	0.04604052	0.07980356	0.03554005	CYFIP1tg	NE
Metabolic pathways	674	-0.134565585	-1.9391519	0.00116414	0.0315427	0.02406844	CYFIP1tg	NPC

Thermogenesis	152	-0.265746356	-2.6579116	0.00141844	0.0315427	0.02406844	CYFIP1tg	NPC
Alzheimer disease	114	-0.343699166	-3.1735193	0.00144509	0.0315427	0.02406844	CYFIP1tg	NPC
Ribosome	118	-0.525110637	-4.9089177	0.00144718	0.0315427	0.02406844	CYFIP1tg	NPC
Huntington disease	129	-0.399228795	-3.8294793	0.00144718	0.0315427	0.02406844	CYFIP1tg	NPC
Non-alcoholic fatty liver disease (NAFLD)	97	-0.349521818	-3.0354469	0.00148368	0.0315427	0.02406844	CYFIP1tg	NPC
Parkinson disease	97	-0.418835675	-3.6374079	0.00148368	0.0315427	0.02406844	CYFIP1tg	NPC
Oxidative phosphorylation	90	-0.490753099	-4.1175118	0.0015361	0.0315427	0.02406844	CYFIP1tg	NPC
Cardiac muscle contraction	44	-0.35227261	-2.2429681	0.00163666	0.0315427	0.02406844	CYFIP1tg	NPC
Proteasome	34	-0.451158355	-2.5943003	0.00167224	0.0315427	0.02406844	CYFIP1tg	NPC
Systemic lupus erythematosus	28	-0.402666364	-2.1413872	0.00169779	0.0315427	0.02406844	CYFIP1tg	NPC
Hematopoietic cell lineage	26	-0.520142725	-2.6719777	0.00171527	0.0315427	0.02406844	CYFIP1tg	NPC
Taste transduction	22	0.528592451	2.75135316	0.00226244	0.0315427	0.02406844	CYFIP1tg	NPC
Basal cell carcinoma	40	0.365746124	2.54528827	0.00240385	0.0315427	0.02406844	CYFIP1tg	NPC
Lysine degradation	32	0.346116669	2.14634928	0.00244499	0.0315427	0.02406844	CYFIP1tg	NPC
Notch signaling pathway	34	0.357058556	2.28343187	0.00247525	0.0315427	0.02406844	CYFIP1tg	NPC
Melanoma	38	0.295795188	1.97976134	0.00248756	0.0315427	0.02406844	CYFIP1tg	NPC
Retrograde endocannabinoid signaling	94	-0.260058033	-2.2463067	0.00296296	0.0315427	0.02406844	CYFIP1tg	NPC
Breast cancer	81	0.34564814	3.20628463	0.00296736	0.0315427	0.02406844	CYFIP1tg	NPC
Neuroactive ligand-receptor interaction	80	-0.280003602	-2.2666872	0.00298954	0.0315427	0.02406844	CYFIP1tg	NPC
Wnt signaling pathway	93	0.222341005	2.19333775	0.00301205	0.0315427	0.02406844	CYFIP1tg	NPC
Gastric cancer	79	0.278661507	2.53850546	0.0030303	0.0315427	0.02406844	CYFIP1tg	NPC
MAPK signaling pathway	152	0.178455447	2.16027357	0.003367	0.03318841	0.02532418	CYFIP1tg	NPC
Amino sugar and nucleotide sugar metabolism	25	-0.390759555	-1.9589557	0.00347826	0.03318841	0.02532418	CYFIP1tg	NPC
Cytokine-cytokine receptor interaction	66	-0.322704927	-2.4130019	0.00460123	0.04141548	0.03160179	CYFIP1tg	NPC

Phagosome	59	-0.289373799	-2.0742856	0.00470219	0.04141548	0.03160179	CYFIP1tg	NPC
Insulin secretion	45	0.267767977	1.94313008	0.00515464	0.04371898	0.03335946	CYFIP1tg	NPC
Protein processing in endoplasmic reticulum	113	-0.227713082	-2.0967714	0.00578871	0.04604403	0.03513358	CYFIP1tg	NPC
Calcium signaling pathway	84	0.20887536	1.97784864	0.0058309	0.04604403	0.03513358	CYFIP1tg	NPC
Rap1 signaling pathway	109	0.176793132	1.83478585	0.00623053	0.04755971	0.0362901	CYFIP1tg	NPC
Regulation of actin cytoskeleton	116	0.160929191	1.69546001	0.00645161	0.04765869	0.03636563	CYFIP1tg	NPC
Pathways in cancer	277	0.127896847	1.88676265	0.00803213	0.05719281	0.04364057	CYFIP1tg	NPC
Endocrine resistance	59	0.248760479	2.02253215	0.00824176	0.05719281	0.04364057	CYFIP1tg	NPC
Hepatocellular carcinoma	94	0.187428342	1.84200036	0.00917431	0.06179169	0.04714972	CYFIP1tg	NPC
Ras signaling pathway	115	0.160128045	1.66781773	0.00983607	0.06377054	0.04865966	CYFIP1tg	NPC
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	41	0.26574324	1.85303174	0.01002506	0.06377054	0.04865966	CYFIP1tg	NPC
Purine metabolism	99	-0.215446467	-1.8826267	0.0119403	0.07390077	0.05638946	CYFIP1tg	NPC
Dilated cardiomyopathy (DCM)	49	0.240398125	1.80745134	0.01278772	0.07706286	0.05880228	CYFIP1tg	NPC
Thyroid hormone signaling pathway	71	0.211811038	1.83448313	0.01457726	0.08411696	0.06418485	CYFIP1tg	NPC
Cushing syndrome	88	0.193593517	1.87686065	0.0148368	0.08411696	0.06418485	CYFIP1tg	NPC
cAMP signaling pathway	99	0.179945384	1.81742106	0.01506024	0.08411696	0.06418485	CYFIP1tg	NPC
Proteoglycans in cancer	120	0.155171517	1.65762317	0.01602564	0.08578921	0.06546086	CYFIP1tg	NPC
Axon guidance	117	0.152078585	1.60321981	0.01618123	0.08578921	0.06546086	CYFIP1tg	NPC
Inflammatory mediator regulation of TRP channels	59	0.220476435	1.79257043	0.01648352	0.08578921	0.06546086	CYFIP1tg	NPC
Ferroptosis	27	-0.34930952	-1.8224196	0.01873935	0.09536248	0.07276567	CYFIP1tg	NPC
Ribosome biogenesis in eukaryotes	46	-0.27634604	-1.7896398	0.01948052	0.09697911	0.07399923	CYFIP1tg	NPC
Signaling pathways regulating pluripotency of stem cells	77	0.185751974	1.6628844	0.02194357	0.10691656	0.08158193	CYFIP1tg	NPC

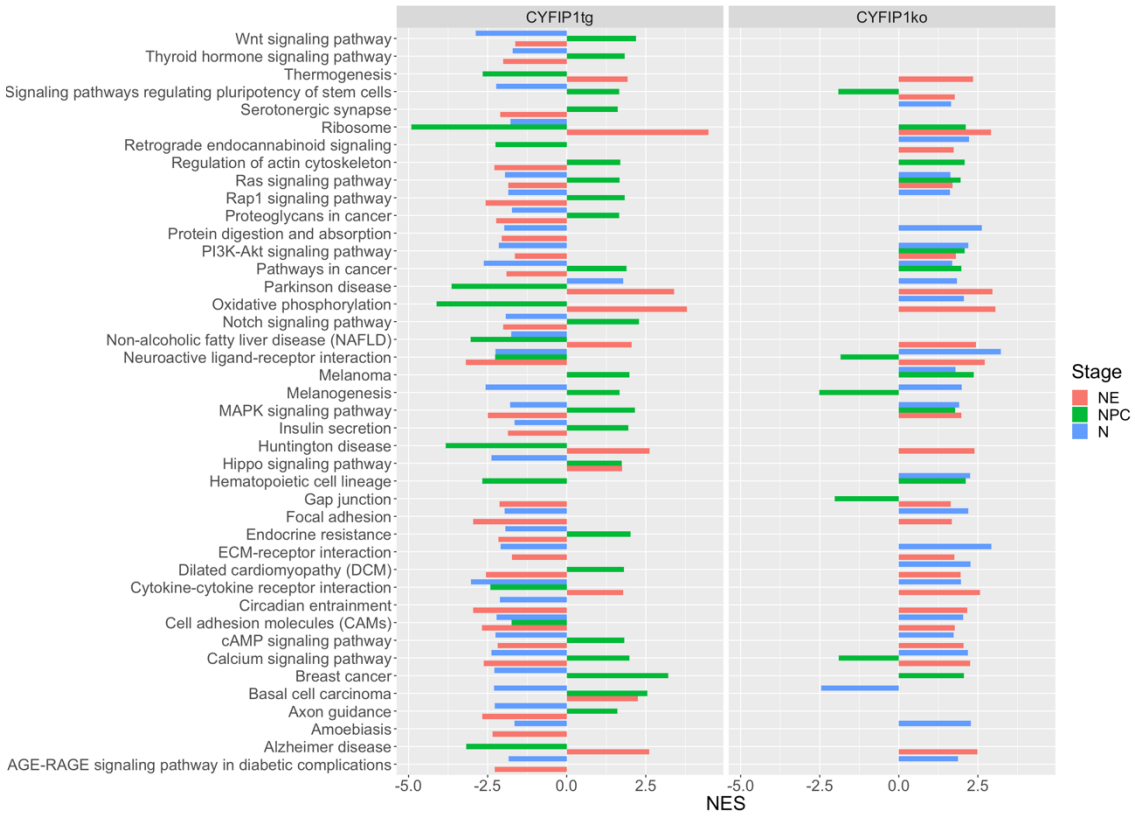
NF-kappa B signaling pathway	38	-0.28884542	-1.739592	0.025	0.11927083	0.09100877	CYFIP1tg	NPC
Cell adhesion molecules (CAMs)	45	-0.270677958	-1.738834	0.02605863	0.12123529	0.09250774	CYFIP1tg	NPC
Hippo signaling pathway	83	0.184020589	1.73794953	0.02647059	0.12123529	0.09250774	CYFIP1tg	NPC
Salivary secretion	37	0.253322919	1.68520503	0.02962963	0.13304285	0.10151741	CYFIP1tg	NPC
Cholinergic synapse	65	0.196374185	1.64454301	0.03047091	0.13418922	0.10239214	CYFIP1tg	NPC
Adrenergic signaling in cardiomyocytes	80	0.178761961	1.63627597	0.03303303	0.14272763	0.10890732	CYFIP1tg	NPC
Melanogenesis	62	0.200550467	1.66630594	0.03399433	0.14416116	0.11000116	CYFIP1tg	NPC
Legionellosis	20	-0.357523798	-1.6767958	0.03546099	0.14764668	0.11266076	CYFIP1tg	NPC
Olfactory transduction	21	0.331575261	1.68195483	0.03854875	0.15763686	0.1202837	CYFIP1tg	NPC
Pancreatic secretion	40	0.229410478	1.59650577	0.04086539	0.16417848	0.12527523	CYFIP1tg	NPC
Serotonergic synapse	57	0.201212432	1.60995828	0.04189944	0.16543055	0.12623062	CYFIP1tg	NPC
Neurotrophin signaling pathway	68	0.179997148	1.53403453	0.04558405	0.17692092	0.13499826	CYFIP1tg	NPC
Arachidonic acid metabolism	20	0.330943375	1.65167798	0.04794521	0.17692092	0.13499826	CYFIP1tg	NPC
Bile secretion	31	0.268354167	1.61721707	0.04834606	0.17692092	0.13499826	CYFIP1tg	NPC
Type II diabetes mellitus	28	0.275312603	1.60369036	0.04842615	0.17692092	0.13499826	CYFIP1tg	NPC
Fluid shear stress and atherosclerosis	76	-0.199001689	-1.5825794	0.04867257	0.17692092	0.13499826	CYFIP1tg	NPC
Pathways in cancer	186	-0.229955126	-2.6261617	0.00152439	0.02060185	0.01388889	CYFIP1tg	N
PI3K-Akt signaling pathway	113	-0.223618612	-2.1472949	0.00161031	0.02060185	0.01388889	CYFIP1tg	N
Hippo signaling pathway	80	-0.277089297	-2.3764761	0.00161551	0.02060185	0.01388889	CYFIP1tg	N
Axon guidance	89	-0.25534368	-2.2692678	0.00161812	0.02060185	0.01388889	CYFIP1tg	N
MicroRNAs in cancer	69	-0.293958802	-2.3981257	0.00162866	0.02060185	0.01388889	CYFIP1tg	N
Cushing syndrome	67	-0.27990106	-2.2563086	0.00163934	0.02060185	0.01388889	CYFIP1tg	N
Breast cancer	67	-0.284073223	-2.2899408	0.00163934	0.02060185	0.01388889	CYFIP1tg	N
Signaling pathways regulating pluripo-tency of stem cells	72	-0.268905887	-2.2312684	0.00164474	0.02060185	0.01388889	CYFIP1tg	N

Chemokine signaling pathway	51	-0.283650838	-2.0366583	0.00165838	0.02060185	0.01388889	CYFIP1tg	N
Calcium signaling pathway	52	-0.328051698	-2.3761781	0.00166113	0.02060185	0.01388889	CYFIP1tg	N
Wnt signaling pathway	66	-0.359915957	-2.8754081	0.00166113	0.02060185	0.01388889	CYFIP1tg	N
Neuroactive ligand-receptor interaction	55	-0.303844924	-2.249912	0.00166389	0.02060185	0.01388889	CYFIP1tg	N
cAMP signaling pathway	58	-0.297934679	-2.2546518	0.00168634	0.02060185	0.01388889	CYFIP1tg	N
Melanogenesis	46	-0.377713144	-2.5682326	0.00170358	0.02060185	0.01388889	CYFIP1tg	N
Cytokine-cytokine receptor interaction	38	-0.480846588	-3.0252619	0.00173611	0.02060185	0.01388889	CYFIP1tg	N
Human papillomavirus infection	129	-0.223284352	-2.241239	0.00312989	0.03184258	0.02146691	CYFIP1tg	N
Human T-cell leukemia virus 1 infection	69	-0.261463733	-2.1330299	0.00325733	0.03184258	0.02146691	CYFIP1tg	N
Cell adhesion molecules (CAMs)	40	-0.346302253	-2.2130686	0.0034965	0.03184258	0.02146691	CYFIP1tg	N
JAK-STAT signaling pathway	34	-0.390229133	-2.3370687	0.00357782	0.03184258	0.02146691	CYFIP1tg	N
Basal cell carcinoma	34	-0.383399097	-2.2961638	0.00357782	0.03184258	0.02146691	CYFIP1tg	N
Peroxisome	29	0.379349581	2.29410737	0.00464037	0.03933267	0.02651641	CYFIP1tg	N
Focal adhesion	71	-0.237710623	-1.9602429	0.00655738	0.05106139	0.03442341	CYFIP1tg	N
Circadian entrainment	37	-0.340668909	-2.1161168	0.00688468	0.05106139	0.03442341	CYFIP1tg	N
Parathyroid hormone synthesis, secretion and action	37	-0.323674299	-2.0105522	0.00688468	0.05106139	0.03442341	CYFIP1tg	N
Notch signaling pathway	22	-0.38006031	-1.9231203	0.00724638	0.0515942	0.03478261	CYFIP1tg	N
Aldosterone synthesis and secretion	31	-0.347905129	-2.0257989	0.00883392	0.06046196	0.04076087	CYFIP1tg	N
Hepatocellular carcinoma	74	-0.231765421	-1.9380549	0.00980392	0.06046196	0.04076087	CYFIP1tg	N
Influenza A	45	-0.288454653	-1.9407175	0.01025641	0.06046196	0.04076087	CYFIP1tg	N
ECM-receptor interaction	32	-0.354463411	-2.0938486	0.01061947	0.06046196	0.04076087	CYFIP1tg	N
Longevity regulating pathway - multiple species	23	-0.379176777	-1.9636378	0.01073345	0.06046196	0.04076087	CYFIP1tg	N
Protein digestion and absorption	25	-0.370772331	-1.9780844	0.01084991	0.06046196	0.04076087	CYFIP1tg	N
Thyroid hormone synthesis	22	-0.37056327	-1.8750648	0.01086957	0.06046196	0.04076087	CYFIP1tg	N

Ras signaling pathway	79	-0.227964592	-1.9470908	0.01134522	0.06119542	0.04125534	CYFIP1tg	N
Measles	38	-0.304833271	-1.9178684	0.01215278	0.06191612	0.04174121	CYFIP1tg	N
Endocrine resistance	30	-0.336132496	-1.9424484	0.01230229	0.06191612	0.04174121	CYFIP1tg	N
Vascular smooth muscle contraction	34	-0.333679358	-1.998394	0.01252236	0.06191612	0.04174121	CYFIP1tg	N
cGMP-PKG signaling pathway	50	-0.263981537	-1.8739812	0.01328904	0.06393104	0.04309958	CYFIP1tg	N
Longevity regulating pathway	31	-0.325111871	-1.8930772	0.01413428	0.06551343	0.04416636	CYFIP1tg	N
MAPK signaling pathway	110	-0.18737678	-1.785946	0.01435407	0.06551343	0.04416636	CYFIP1tg	N
Arginine and proline metabolism	21	0.343008351	1.84302378	0.0173913	0.07491583	0.05050505	CYFIP1tg	N
Rap1 signaling pathway	76	-0.218645622	-1.8464195	0.01788618	0.07491583	0.05050505	CYFIP1tg	N
Protein processing in endoplasmic reticulum	62	-0.23081666	-1.7920363	0.01836394	0.07491583	0.05050505	CYFIP1tg	N
AMPK signaling pathway	48	-0.260254257	-1.8149607	0.01842546	0.07491583	0.05050505	CYFIP1tg	N
FoxO signaling pathway	54	-0.24447082	-1.7912832	0.01851852	0.07491583	0.05050505	CYFIP1tg	N
Pancreatic secretion	26	-0.33499013	-1.8226297	0.0198556	0.0782217	0.05273373	CYFIP1tg	N
Ribosome	61	-0.230699132	-1.7737497	0.02047782	0.0782217	0.05273373	CYFIP1tg	N
ErbB signaling pathway	37	-0.299301928	-1.8591595	0.02065405	0.0782217	0.05273373	CYFIP1tg	N
Acute myeloid leukemia	29	-0.313839631	-1.8073006	0.02276708	0.0844279	0.05691769	CYFIP1tg	N
Prostate cancer	42	-0.266339778	-1.7531825	0.02401372	0.08622837	0.05813149	CYFIP1tg	N
AGE-RAGE signaling pathway in diabetic complications	39	-0.288156532	-1.8330204	0.02422145	0.08622837	0.05813149	CYFIP1tg	N
Parkinson disease	39	0.255678283	1.78879319	0.0259434	0.09054754	0.06104329	CYFIP1tg	N
Non-alcoholic fatty liver disease (NAFLD)	39	-0.275481816	-1.7523941	0.02768166	0.09475646	0.06388076	CYFIP1tg	N
Thyroid hormone signaling pathway	54	-0.23377089	-1.7128828	0.02861953	0.09611842	0.06479893	CYFIP1tg	N
Adherens junction	40	-0.269637331	-1.7231361	0.02972028	0.09796685	0.06604507	CYFIP1tg	N
Proteoglycans in cancer	87	-0.196099662	-1.7365521	0.03069467	0.09933911	0.06697019	CYFIP1tg	N

Chronic myeloid leukemia	35	-0.283139085	-1.7133239	0.03197158	0.10162395	0.06851053	CYFIP1tg	N
Fc gamma R-mediated phagocytosis	31	-0.292201435	-1.7014447	0.03356891	0.10482921	0.07067138	CYFIP1tg	N
Human cytomegalovirus infection	65	-0.21013152	-1.6825257	0.03453947	0.10600045	0.07146098	CYFIP1tg	N
Endometrial cancer	29	-0.289562416	-1.667496	0.03677758	0.1109561	0.07480186	CYFIP1tg	N
Transcriptional misregulation in cancer	57	-0.221209402	-1.6692795	0.04145937	0.1215847	0.08196721	CYFIP1tg	N
NF-kappa B signaling pathway	22	-0.324006176	-1.6394841	0.04166667	0.1215847	0.08196721	CYFIP1tg	N
Oxytocin signaling pathway	54	-0.221304693	-1.6215407	0.04377104	0.12501756	0.0842815	CYFIP1tg	N
Amoebiasis	32	-0.278886816	-1.6474106	0.04424779	0.12501756	0.0842815	CYFIP1tg	N
Insulin secretion	25	-0.309590975	-1.6516795	0.04520796	0.12573463	0.08476492	CYFIP1tg	N

clusterProfiler GSEA results for CYFIP1tg and CYFIP1ko cells at three developmental stages using a 10% FDR cut-off threshold for DEGs.NE = neuroepithelium, NPC = neural progenitor cells, N = neurons.



Supplementary Figure 3. Pathways affected by altered dosage of CYFIP1.
Bar graph representing the significantly enriched pathways in CYFIP1tg and CYFIPko cells at different stages of differentiation. Normalised enrichment score for each pathway show mirroring effects in the enrichment between each cell line. For representation purposes, only those that are significantly changed in at least half of the datasets are plotted in the graph.

Supplementary Table 4. Significantly enriched KEGG pathways in CYFIP1tg and CYFIP1ko cells

Description	setSize	enrichmentScore	NES	pvalue	p.adjust	qvalues	Cell Type	Stage
Thermogenesis	77	0.280122524	2.34499199	0.00160256	0.02581166	0.02122669	CYFIP1ko	NE
Huntington disease	69	0.300693986	2.39221952	0.00162866	0.02581166	0.02122669	CYFIP1ko	NE
Alzheimer disease	63	0.323571971	2.4907692	0.00165563	0.02581166	0.02122669	CYFIP1ko	NE
Oxidative phosphorylation	57	0.408577626	3.05456974	0.00167504	0.02581166	0.02122669	CYFIP1ko	NE
Non-alcoholic fatty liver disease (NAFLD)	52	0.34171171	2.44456758	0.00170068	0.02581166	0.02122669	CYFIP1ko	NE
Parkinson disease	49	0.422820958	2.96069823	0.00170358	0.02581166	0.02122669	CYFIP1ko	NE
Ribosome	47	0.424202375	2.9102571	0.0017094	0.02581166	0.02122669	CYFIP1ko	NE
Neuroactive ligand-receptor interaction	27	0.48458751	2.71137536	0.00176991	0.02581166	0.02122669	CYFIP1ko	NE
Cytokine-cytokine receptor interaction	21	0.518655897	2.56622927	0.00181488	0.02581166	0.02122669	CYFIP1ko	NE
Calcium signaling pathway	31	0.383626866	2.25484769	0.0069808	0.07843137	0.06449948	CYFIP1ko	NE
Circadian entrainment	22	0.429238538	2.16651208	0.00720721	0.07843137	0.06449948	CYFIP1ko	NE
Rheumatoid arthritis	20	0.446130704	2.1517142	0.00735294	0.07843137	0.06449948	CYFIP1ko	NE
Cardiac muscle contraction	26	0.392255098	2.15265424	0.00888099	0.08744364	0.07191089	CYFIP1ko	NE

MAPK signaling pathway	71	0.246428037	1.9761911	0.00978793	0.08948963	0.07359345	CYFIP1ko	NE
cAMP signaling pathway	40	0.319837625	2.04465116	0.01415929	0.12082596	0.09936345	CYFIP1ko	NE
Dilated cardiomyopathy (DCM)	22	0.387165455	1.95415499	0.01621622	0.12972973	0.10668563	CYFIP1ko	NE
Hepatitis C	25	-0.33950731	-1.9155401	0.02	0.15058824	0.12383901	CYFIP1ko	NE
Ubiquitin mediated proteolysis	34	-0.274603726	-1.7992499	0.02540416	0.17483154	0.14377594	CYFIP1ko	NE
Human T-cell leukemia virus 1 infection	55	0.250599205	1.82312688	0.02595156	0.17483154	0.14377594	CYFIP1ko	NE
PI3K-Akt signaling pathway	74	0.218558691	1.80442083	0.0273752	0.17520129	0.14408001	CYFIP1ko	NE
Signaling pathways regulating pluripotency of stem cells	28	0.314801831	1.77141441	0.03041145	0.18536502	0.15243834	CYFIP1ko	NE
ECM-receptor interaction	25	0.32941138	1.7614362	0.03442029	0.2002635	0.16469038	CYFIP1ko	NE
Retrograde endocannabinoid signaling	43	0.263675148	1.73658628	0.03664921	0.20396085	0.16773096	CYFIP1ko	NE
Cell adhesion molecules (CAMs)	32	0.297342353	1.76358544	0.04006969	0.20401417	0.16777481	CYFIP1ko	NE
Ras signaling pathway	54	0.235212152	1.69932006	0.04310345	0.20401417	0.16777481	CYFIP1ko	NE
Gap junction	22	0.324848314	1.63961931	0.04324324	0.20401417	0.16777481	CYFIP1ko	NE
Pyrimidine metabolism	20	0.332151382	1.60198533	0.04411765	0.20401417	0.16777481	CYFIP1ko	NE

Focal adhesion	62	0.217621885	1.67096826	0.0446281	0.20401417	0.16777481	CYFIP1ko	NE
PI3K-Akt signaling pathway	136	0.378905265	2.08219278	0.00195313	0.18893528	0.18019998	CYFIP1ko	NPC
Melanogenesis	50	-0.578303138	-2.5136655	0.00208768	0.18893528	0.18019998	CYFIP1ko	NPC
Pathways in cancer	206	0.332543249	1.97752897	0.00394477	0.22445437	0.21407686	CYFIP1ko	NPC
Melanoma	35	0.573386102	2.37150431	0.00575816	0.22445437	0.21407686	CYFIP1ko	NPC
Ribosome	75	0.435054174	2.11719907	0.00792079	0.22445437	0.21407686	CYFIP1ko	NPC
Transcriptional misregulation in cancer	68	0.461744491	2.19433027	0.00795229	0.22445437	0.21407686	CYFIP1ko	NPC
Regulation of actin cytoskeleton	86	0.408844915	2.07852525	0.00984252	0.22445437	0.21407686	CYFIP1ko	NPC
Breast cancer	72	0.428196065	2.06176565	0.00992063	0.22445437	0.21407686	CYFIP1ko	NPC
Gap junction	46	-0.476570841	-2.0291365	0.01234568	0.24828532	0.23680601	CYFIP1ko	NPC
Ras signaling pathway	84	0.389610478	1.95715309	0.01388889	0.25138889	0.23976608	CYFIP1ko	NPC
Hematopoietic cell lineage	24	0.570672095	2.11524973	0.01912046	0.28116505	0.26816556	CYFIP1ko	NPC
Signaling pathways regulating pluripotency of stem cells	66	-0.41007941	-1.910115	0.02061856	0.28116505	0.26816556	CYFIP1ko	NPC
Calcium signaling pathway	65	-0.408111685	-1.8996618	0.02066116	0.28116505	0.26816556	CYFIP1ko	NPC

Neuroactive ligand-receptor interaction	66	-0.395084621	-1.8402706	0.02268041	0.28116505	0.26816556	CYFIP1ko	NPC
MAPK signaling pathway	126	0.328075161	1.77648304	0.02330097	0.28116505	0.26816556	CYFIP1ko	NPC
Gastric cancer	74	0.38915721	1.88223253	0.02584493	0.29237078	0.2788532	CYFIP1ko	NPC
Viral myocarditis	23	0.48890745	1.79156077	0.0440613	0.46912328	0.44743366	CYFIP1ko	NPC
PI3K-Akt signaling pathway	170	0.33871449	2.19159499	0.00111359	0.02377822	0.01880111	CYFIP1ko	N
Focal adhesion	110	0.364669883	2.19157524	0.0011655	0.02377822	0.01880111	CYFIP1ko	N
Neuroactive ligand-receptor interaction	93	0.555973648	3.22412884	0.00118343	0.02377822	0.01880111	CYFIP1ko	N
Calcium signaling pathway	79	0.389733141	2.18800277	0.00119474	0.02377822	0.01880111	CYFIP1ko	N
Glutamatergic synapse	69	0.40028534	2.16935856	0.00122549	0.02377822	0.01880111	CYFIP1ko	N
Morphine addiction	60	0.462352333	2.43213814	0.00123153	0.02377822	0.01880111	CYFIP1ko	N
Hypertrophic cardiomyopathy (HCM)	51	0.435535163	2.15707865	0.001287	0.02377822	0.01880111	CYFIP1ko	N
Protein digestion and absorption	47	0.542369373	2.62349987	0.00129366	0.02377822	0.01880111	CYFIP1ko	N
Amoebiasis	46	0.475261728	2.27900991	0.00129702	0.02377822	0.01880111	CYFIP1ko	N
ECM-receptor interaction	44	0.616011531	2.92751251	0.00130719	0.02377822	0.01880111	CYFIP1ko	N

Systemic lupus erythematosus	28	0.601495361	2.53319833	0.00140252	0.02377822	0.01880111	CYFIP1ko	N
Nicotine addiction	25	0.624170055	2.52187857	0.00141844	0.02377822	0.01880111	CYFIP1ko	N
Complement and coagulation cascades	23	0.627917119	2.4710745	0.0014245	0.02377822	0.01880111	CYFIP1ko	N
Retrograde endocannabinoid signaling	91	0.386103052	2.22244979	0.0023753	0.03223411	0.02548707	CYFIP1ko	N
Oxidative phosphorylation	70	0.377810343	2.0622897	0.00245098	0.03223411	0.02548707	CYFIP1ko	N
Dilated cardiomyopathy (DCM)	55	0.442760174	2.26428569	0.00249377	0.03223411	0.02548707	CYFIP1ko	N
GABAergic synapse	53	0.421280259	2.12614334	0.00252525	0.03223411	0.02548707	CYFIP1ko	N
Hematopoietic cell lineage	21	0.591772318	2.25940815	0.00287356	0.0346424	0.02739128	CYFIP1ko	N
MAPK signaling pathway	151	0.301004799	1.90323864	0.00337458	0.03854123	0.03047403	CYFIP1ko	N
Cell adhesion molecules (CAMs)	58	0.390088065	2.03129869	0.00494438	0.05147364	0.04069951	CYFIP1ko	N
Melanogenesis	56	0.386425041	1.98376434	0.00498132	0.05147364	0.04069951	CYFIP1ko	N
Cytokine-cytokine receptor interaction	70	0.359421207	1.96191202	0.00612745	0.0584559	0.04622029	CYFIP1ko	N
Estrogen signaling pathway	61	0.371253888	1.94745992	0.00619579	0.0584559	0.04622029	CYFIP1ko	N
Rheumatoid arthritis	27	0.495523212	2.06925785	0.00699301	0.06086957	0.04812874	CYFIP1ko	N

Regulation of lipolysis in adipocytes	28	0.504969075	2.12667778	0.00701262	0.06086957	0.04812874	CYFIP1ko	N
Pathways in cancer	277	0.244484665	1.68546085	0.00862069	0.0719496	0.05688957	CYFIP1ko	N
Leukocyte transendothelial migration	49	0.39377812	1.92986217	0.01023018	0.08222033	0.06501049	CYFIP1ko	N
Parkinson disease	83	0.325087312	1.83429265	0.01079137	0.08363309	0.06612755	CYFIP1ko	N
AGE-RAGE signaling pathway in diabetic complications	59	0.3578329	1.86758138	0.01243781	0.09306914	0.0735885	CYFIP1ko	N
Platelet activation	54	0.361661523	1.82604425	0.01643489	0.11887906	0.09399605	CYFIP1ko	N
cAMP signaling pathway	100	0.293626953	1.73206834	0.02112676	0.13929603	0.11013948	CYFIP1ko	N
Hedgehog signaling pathway	26	-0.603591178	-3.0256453	0.02120141	0.13929603	0.11013948	CYFIP1ko	N
DNA replication	27	-0.560455752	-2.8674221	0.02439024	0.13929603	0.11013948	CYFIP1ko	N
Taste transduction	21	0.494958969	1.88977128	0.02442529	0.13929603	0.11013948	CYFIP1ko	N
TGF-beta signaling pathway	50	0.35842284	1.76080564	0.02448454	0.13929603	0.11013948	CYFIP1ko	N
Homologous recombination	26	-0.489880787	-2.4556447	0.02473498	0.13929603	0.11013948	CYFIP1ko	N
Fanconi anemia pathway	26	-0.490257343	-2.4575323	0.02473498	0.13929603	0.11013948	CYFIP1ko	N
Lysine degradation	29	-0.360794077	-1.8723688	0.0248227	0.13929603	0.11013948	CYFIP1ko	N

Bile secretion	26	0.454652106	1.88108881	0.02503477	0.13929603	0.11013948	CYFIP1ko	N
Cocaine addiction	31	0.424895847	1.82359228	0.02797203	0.1517011	0.119948	CYFIP1ko	N
Base excision repair	20	-0.440716185	-2.0887427	0.02866242	0.1517011	0.119948	CYFIP1ko	N
Basal cell carcinoma	38	-0.424705467	-2.4590089	0.03225806	0.16666667	0.13178107	CYFIP1ko	N
Ras signaling pathway	115	0.269628687	1.63414471	0.03333333	0.16821705	0.13300694	CYFIP1ko	N
Human papillomavirus infection	168	0.238387167	1.54064549	0.03674833	0.17777518	0.14056441	CYFIP1ko	N
Leishmaniasis	26	0.434460951	1.79754942	0.03755216	0.17777518	0.14056441	CYFIP1ko	N
Melanoma	36	0.397353198	1.79216434	0.03768506	0.17777518	0.14056441	CYFIP1ko	N
Dopaminergic synapse	80	0.302991264	1.70295647	0.03956835	0.18268789	0.14444883	CYFIP1ko	N
RaEA8 signaling pathway	106	0.272070682	1.6214681	0.04084014	0.18463147	0.14598559	CYFIP1ko	N
Phospholipase D signaling pathway	76	0.302671961	1.67522934	0.04227053	0.18719807	0.14801496	CYFIP1ko	N
Olfactory transduction	27	0.419539415	1.75195673	0.04475524	0.19423776	0.15358116	CYFIP1ko	N
Nucleotide excision repair	24	-0.349264258	-1.7300767	0.04590164	0.19527328	0.15439992	CYFIP1ko	N
Relaxin signaling pathway	74	0.295261812	1.63734239	0.04796163	0.19527328	0.15439992	CYFIP1ko	N

Alcoholism	74	0.295625249	1.6393578	0.04796163	0.19527328	0.15439992	CYFIP1ko	N
Serotonergic synapse	52	0.331792326	1.65383777	0.04859335	0.19527328	0.15439992	CYFIP1ko	N
MAPK signaling pathway	83	-0.286685598	-2.4899417	0.00155039	0.00982987	0.00437767	CYFIP1tg	NE
Regulation of actin cytoskeleton	68	-0.284362812	-2.2858903	0.0015873	0.00982987	0.00437767	CYFIP1tg	NE
Axon guidance	70	-0.326585487	-2.6684555	0.00158983	0.00982987	0.00437767	CYFIP1tg	NE
Focal adhesion	74	-0.35724039	-2.9535614	0.00159236	0.00982987	0.00437767	CYFIP1tg	NE
Proteoglycans in cancer	65	-0.282267193	-2.2293211	0.00159236	0.00982987	0.00437767	CYFIP1tg	NE
RaEA8 signaling pathway	63	-0.329906854	-2.5640272	0.00162075	0.00982987	0.00437767	CYFIP1tg	NE
Neuroactive ligand-receptor interaction	48	-0.458519423	-3.1832102	0.00169492	0.00982987	0.00437767	CYFIP1tg	NE
Dopaminergic synapse	42	-0.41584325	-2.7634514	0.0017301	0.00982987	0.00437767	CYFIP1tg	NE
Cell adhesion molecules (CAMs)	37	-0.423983957	-2.6827732	0.00175131	0.00982987	0.00437767	CYFIP1tg	NE
Platelet activation	37	-0.351021395	-2.2211	0.00175131	0.00982987	0.00437767	CYFIP1tg	NE
Circadian entrainment	37	-0.467774964	-2.9598623	0.00175131	0.00982987	0.00437767	CYFIP1tg	NE
Relaxin signaling pathway	38	-0.365538032	-2.3446685	0.00175131	0.00982987	0.00437767	CYFIP1tg	NE

Calcium signaling pathway	39	-0.402926459	-2.617061	0.00175439	0.00982987	0.00437767	CYFIP1tg	NE
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	30	-0.481960468	-2.8053889	0.00175747	0.00982987	0.00437767	CYFIP1tg	NE
Endocrine resistance	31	-0.368476636	-2.1605911	0.00176367	0.00982987	0.00437767	CYFIP1tg	NE
Aldosterone synthesis and secretion	31	-0.369298994	-2.1654131	0.00176367	0.00982987	0.00437767	CYFIP1tg	NE
AGE-RAGE signaling pathway in diabetic complications	31	-0.387300162	-2.2709643	0.00176367	0.00982987	0.00437767	CYFIP1tg	NE
Dilated cardiomyopathy (DCM)	31	-0.436174681	-2.5575438	0.00176367	0.00982987	0.00437767	CYFIP1tg	NE
Glutamatergic synapse	35	-0.435786831	-2.6834243	0.00176991	0.00982987	0.00437767	CYFIP1tg	NE
ErbB signaling pathway	32	-0.353287147	-2.098029	0.0017762	0.00982987	0.00437767	CYFIP1tg	NE
Hypertrophic cardiomyopathy (HCM)	28	-0.417960435	-2.3578	0.00178253	0.00982987	0.00437767	CYFIP1tg	NE
Osteoclast differentiation	26	-0.411115738	-2.262742	0.00178571	0.00982987	0.00437767	CYFIP1tg	NE
Leukocyte transendothelial migration	26	-0.379576241	-2.0891516	0.00178571	0.00982987	0.00437767	CYFIP1tg	NE
Inflammatory mediator regulation of TRP channels	26	-0.443431804	-2.4406066	0.00178571	0.00982987	0.00437767	CYFIP1tg	NE

Adherens junction	34	-0.37604482	-2.2730023	0.00179856	0.00982987	0.00437767	CYFIP1tg	NE
Long-term potentiation	24	-0.471585187	-2.4927807	0.00179856	0.00982987	0.00437767	CYFIP1tg	NE
Serotonergic synapse	24	-0.396734412	-2.0971224	0.00179856	0.00982987	0.00437767	CYFIP1tg	NE
Amphetamine addiction	24	-0.456412063	-2.4125761	0.00179856	0.00982987	0.00437767	CYFIP1tg	NE
Salivary secretion	22	-0.412783745	-2.1050526	0.00184502	0.00982987	0.00437767	CYFIP1tg	NE
Amoebiasis	20	-0.480448297	-2.3463191	0.00189036	0.00982987	0.00437767	CYFIP1tg	NE
Basal cell carcinoma	25	0.393140841	2.24757866	0.00226757	0.01002506	0.0044646	CYFIP1tg	NE
Proteasome	32	0.485781098	3.08327614	0.0022779	0.01002506	0.0044646	CYFIP1tg	NE
Ribosome biogenesis in eukaryotes	35	0.42657685	2.81207104	0.00228833	0.01002506	0.0044646	CYFIP1tg	NE
Pyrimidine metabolism	41	0.363367201	2.57702806	0.00231481	0.01002506	0.0044646	CYFIP1tg	NE
Parkinson disease	46	0.44895098	3.39315674	0.00240964	0.01002506	0.0044646	CYFIP1tg	NE
Oxidative phosphorylation	49	0.494253685	3.80196653	0.00246914	0.01002506	0.0044646	CYFIP1tg	NE
Ribosome	55	0.557156134	4.48294325	0.0025	0.01002506	0.0044646	CYFIP1tg	NE
Spliceosome	56	0.338412528	2.75159317	0.0025	0.01002506	0.0044646	CYFIP1tg	NE

RNA transport	57	0.392977967	3.23325219	0.00250627	0.01002506	0.0044646	CYFIP1tg	NE
Alzheimer disease	59	0.314054434	2.60490694	0.00260417	0.01015625	0.00452303	CYFIP1tg	NE
Huntington disease	71	0.296166419	2.6205934	0.00268097	0.01020075	0.00454284	CYFIP1tg	NE
Thermogenesis	75	0.212754075	1.91922921	0.0027933	0.0103751	0.00462049	CYFIP1tg	NE
cAMP signaling pathway	57	-0.293203141	-2.1773428	0.00331675	0.0113906	0.00507274	CYFIP1tg	NE
Adrenergic signaling in cardiomyocytes	45	-0.315477137	-2.1374663	0.00339559	0.0113906	0.00507274	CYFIP1tg	NE
Oxytocin signaling pathway	45	-0.341966726	-2.3169424	0.00339559	0.0113906	0.00507274	CYFIP1tg	NE
Parathyroid hormone synthesis, secretion and action	33	-0.357526331	-2.1383929	0.00355872	0.0113906	0.00507274	CYFIP1tg	NE
Vascular smooth muscle contraction	25	-0.368605354	-1.982709	0.00356506	0.0113906	0.00507274	CYFIP1tg	NE
Gap junction	28	-0.376137716	-2.1218695	0.00356506	0.0113906	0.00507274	CYFIP1tg	NE
Phosphatidylinositol signaling system	29	-0.364571534	-2.0760905	0.00357782	0.0113906	0.00507274	CYFIP1tg	NE
Protein digestion and absorption	22	-0.403057352	-2.0554514	0.00369004	0.01151292	0.00512721	CYFIP1tg	NE
mRNA surveillance pathway	31	0.327648869	2.06784747	0.0045977	0.01406356	0.00626312	CYFIP1tg	NE
FoxO signaling pathway	38	-0.338462915	-2.1710007	0.00525394	0.01551879	0.0069112	CYFIP1tg	NE

GnRH signaling pathway	27	-0.378027701	-2.107565	0.00527241	0.01551879	0.0069112	CYFIP1tg	NE
Notch signaling pathway	20	-0.411847831	-2.0113016	0.00567108	0.01638311	0.00729612	CYFIP1tg	NE
Pathways in cancer	154	-0.180598138	-1.9096238	0.0057971	0.01644269	0.00732265	CYFIP1tg	NE
Thyroid hormone signaling pathway	42	-0.301946097	-2.0065574	0.00692042	0.01917758	0.00854062	CYFIP1tg	NE
Autophagy - animal	36	-0.324543141	-2.0244282	0.00702988	0.01917758	0.00854062	CYFIP1tg	NE
Bacterial invasion of epithelial cells	28	-0.336097042	-1.8959919	0.00713012	0.01917758	0.00854062	CYFIP1tg	NE
Prostate cancer	30	-0.332877301	-1.9376076	0.00878735	0.02323434	0.01034728	CYFIP1tg	NE
Non-alcoholic fatty liver disease (NAFLD)	50	0.263171028	2.05525631	0.00966184	0.02512077	0.01118739	CYFIP1tg	NE
Estrogen signaling pathway	37	-0.309244333	-1.9567542	0.01050788	0.026	0.01157895	CYFIP1tg	NE
Phospholipase D signaling pathway	36	-0.308875779	-1.9266986	0.01054482	0.026	0.01157895	CYFIP1tg	NE
Cholinergic synapse	36	-0.303108093	-1.8907211	0.01054482	0.026	0.01157895	CYFIP1tg	NE
Purine metabolism	64	0.217578223	1.84618444	0.01066667	0.026	0.01157895	CYFIP1tg	NE
Ras signaling pathway	61	-0.24127429	-1.8498301	0.01141925	0.0274062	0.01220519	CYFIP1tg	NE
Chemokine signaling pathway	38	-0.307391008	-1.9716964	0.01225919	0.02864413	0.01275649	CYFIP1tg	NE

Hepatitis C	27	-0.327837376	-1.8277459	0.01230228	0.02864413	0.01275649	CYFIP1tg	NE
Renal cell carcinoma	26	-0.33728563	-1.8563881	0.0125	0.02867647	0.0127709	CYFIP1tg	NE
T cell receptor signaling pathway	22	-0.356125723	-1.8161165	0.01291513	0.02919942	0.01300379	CYFIP1tg	NE
Choline metabolism in cancer	32	-0.309609497	-1.8386452	0.01420959	0.03133238	0.01395369	CYFIP1tg	NE
Insulin secretion	25	-0.345200429	-1.8568152	0.01426025	0.03133238	0.01395369	CYFIP1tg	NE
MicroRNAs in cancer	51	-0.260360403	-1.847728	0.01517707	0.03288364	0.01464454	CYFIP1tg	NE
Cytokine-cytokine receptor interaction	35	0.270619419	1.78397171	0.01601831	0.0342309	0.01524453	CYFIP1tg	NE
Fluid shear stress and atherosclerosis	43	-0.271692249	-1.8102759	0.01929825	0.04068279	0.01811784	CYFIP1tg	NE
Human cytomegalovirus infection	53	-0.24782923	-1.7873247	0.02003339	0.04166945	0.01855724	CYFIP1tg	NE
Gastric acid secretion	24	-0.346914919	-1.8337785	0.02158273	0.0443014	0.01972937	CYFIP1tg	NE
Tight junction	40	-0.266082939	-1.7328702	0.02464789	0.04875762	0.02171392	CYFIP1tg	NE
RNA degradation	34	0.267092836	1.74501221	0.02466368	0.04875762	0.02171392	CYFIP1tg	NE
ECM-receptor interaction	31	-0.295451092	-1.7324002	0.02469136	0.04875762	0.02171392	CYFIP1tg	NE
JAK-STAT signaling pathway	25	-0.324876405	-1.7474933	0.02673797	0.05213904	0.02321981	CYFIP1tg	NE

Neurotrophin signaling pathway	38	-0.273052026	-1.751436	0.02802102	0.0539664	0.02403362	CYFIP1tg	NE
Hippo signaling pathway	62	0.206503132	1.7538407	0.02864583	0.05449695	0.0242699	CYFIP1tg	NE
Glycerophospholipid metabolism	21	-0.352990217	-1.7566981	0.02962963	0.05568942	0.02480096	CYFIP1tg	NE
PI3K-Akt signaling pathway	88	-0.185712954	-1.6394804	0.03301887	0.06132075	0.02730884	CYFIP1tg	NE
Chagas disease (American trypanosomiasis)	23	-0.318856308	-1.6559358	0.03683241	0.06759831	0.03010451	CYFIP1tg	NE
Inositol phosphate metabolism	21	-0.332864394	-1.6565395	0.04074074	0.07390181	0.03291174	CYFIP1tg	NE
Insulin signaling pathway	40	-0.252983609	-1.6475606	0.04225352	0.07494457	0.03337612	CYFIP1tg	NE
Wnt signaling pathway	60	-0.213941659	-1.6272938	0.04227642	0.07494457	0.03337612	CYFIP1tg	NE
TGF-beta signaling pathway	29	0.266614521	1.63800292	0.04514673	0.07913359	0.03524168	CYFIP1tg	NE
C-type lectin receptor signaling pathway	23	-0.312746597	-1.6242058	0.04604052	0.07980356	0.03554005	CYFIP1tg	NE
Metabolic pathways	674	-0.134565585	-1.9391519	0.00116414	0.0315427	0.02406844	CYFIP1tg	NPC
Thermogenesis	152	-0.265746356	-2.6579116	0.00141844	0.0315427	0.02406844	CYFIP1tg	NPC
Alzheimer disease	114	-0.343699166	-3.1735193	0.00144509	0.0315427	0.02406844	CYFIP1tg	NPC
Ribosome	118	-0.525110637	-4.9089177	0.00144718	0.0315427	0.02406844	CYFIP1tg	NPC

Huntington disease	129	-0.399228795	-3.8294793	0.00144718	0.0315427	0.02406844	CYFIP1tg	NPC
Non-alcoholic fatty liver disease (NAFLD)	97	-0.349521818	-3.0354469	0.00148368	0.0315427	0.02406844	CYFIP1tg	NPC
Parkinson disease	97	-0.418835675	-3.6374079	0.00148368	0.0315427	0.02406844	CYFIP1tg	NPC
Oxidative phosphorylation	90	-0.490753099	-4.1175118	0.0015361	0.0315427	0.02406844	CYFIP1tg	NPC
Cardiac muscle contraction	44	-0.35227261	-2.2429681	0.00163666	0.0315427	0.02406844	CYFIP1tg	NPC
Proteasome	34	-0.451158355	-2.5943003	0.00167224	0.0315427	0.02406844	CYFIP1tg	NPC
Systemic lupus erythematosus	28	-0.402666364	-2.1413872	0.00169779	0.0315427	0.02406844	CYFIP1tg	NPC
Hematopoietic cell lineage	26	-0.520142725	-2.6719777	0.00171527	0.0315427	0.02406844	CYFIP1tg	NPC
Taste transduction	22	0.528592451	2.75135316	0.00226244	0.0315427	0.02406844	CYFIP1tg	NPC
Basal cell carcinoma	40	0.365746124	2.54528826	0.00240385	0.0315427	0.02406844	CYFIP1tg	NPC
Lysine degradation	32	0.346116669	2.14634928	0.00244499	0.0315427	0.02406844	CYFIP1tg	NPC
Notch signaling pathway	34	0.357058556	2.28343187	0.00247525	0.0315427	0.02406844	CYFIP1tg	NPC
Melanoma	38	0.295795188	1.97976134	0.00248756	0.0315427	0.02406844	CYFIP1tg	NPC
Retrograde endocannabinoid signaling	94	-0.260058033	-2.2463067	0.00296296	0.0315427	0.02406844	CYFIP1tg	NPC

Breast cancer	81	0.34564814	3.20628463	0.00296736	0.0315427	0.02406844	CYFIP1tg	NPC
Neuroactive ligand-receptor interaction	80	-0.280003602	-2.2666872	0.00298954	0.0315427	0.02406844	CYFIP1tg	NPC
Wnt signaling pathway	93	0.222341005	2.19333775	0.00301205	0.0315427	0.02406844	CYFIP1tg	NPC
Gastric cancer	79	0.278661507	2.53850546	0.0030303	0.0315427	0.02406844	CYFIP1tg	NPC
MAPK signaling pathway	152	0.178455447	2.16027357	0.003367	0.03318841	0.02532418	CYFIP1tg	NPC
Amino sugar and nucleotide sugar metabolism	25	-0.390759555	-1.9589557	0.00347826	0.03318841	0.02532418	CYFIP1tg	NPC
Cytokine-cytokine receptor interaction	66	-0.322704927	-2.4130019	0.00460123	0.04141548	0.03160179	CYFIP1tg	NPC
Phagosome	59	-0.289373799	-2.0742856	0.00470219	0.04141548	0.03160179	CYFIP1tg	NPC
Insulin secretion	45	0.267767977	1.94313008	0.00515464	0.04371898	0.03335946	CYFIP1tg	NPC
Protein processing in endoplasmic reticulum	113	-0.227713082	-2.0967714	0.00578871	0.04604403	0.03513358	CYFIP1tg	NPC
Calcium signaling pathway	84	0.20887536	1.97784864	0.0058309	0.04604403	0.03513358	CYFIP1tg	NPC
RaEA8 signaling pathway	109	0.176793132	1.83478585	0.00623053	0.04755971	0.0362901	CYFIP1tg	NPC
Regulation of actin cytoskeleton	116	0.160929191	1.69546001	0.00645161	0.04765869	0.03636563	CYFIP1tg	NPC
Pathways in cancer	277	0.127896847	1.88676265	0.00803213	0.05719281	0.04364057	CYFIP1tg	NPC

Endocrine resistance	59	0.248760479	2.02253215	0.00824176	0.05719281	0.04364057	CYFIP1tg	NPC
Hepatocellular carcinoma	94	0.187428342	1.84200036	0.00917431	0.06179169	0.04714971	CYFIP1tg	NPC
Ras signaling pathway	115	0.160128045	1.66781773	0.00983607	0.06377054	0.04865966	CYFIP1tg	NPC
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	41	0.26574324	1.85303174	0.01002506	0.06377054	0.04865966	CYFIP1tg	NPC
Purine metabolism	99	-0.215446467	-1.8826267	0.0119403	0.07390077	0.05638946	CYFIP1tg	NPC
Dilated cardiomyopathy (DCM)	49	0.240398125	1.80745134	0.01278772	0.07706286	0.05880228	CYFIP1tg	NPC
Thyroid hormone signaling pathway	71	0.211811038	1.83448313	0.01457726	0.08411696	0.06418485	CYFIP1tg	NPC
Cushing syndrome	88	0.193593517	1.87686065	0.0148368	0.08411696	0.06418485	CYFIP1tg	NPC
cAMP signaling pathway	99	0.179945384	1.81742106	0.01506024	0.08411696	0.06418485	CYFIP1tg	NPC
Proteoglycans in cancer	120	0.155171517	1.65762316	0.01602564	0.08578921	0.06546085	CYFIP1tg	NPC
Axon guidance	117	0.152078585	1.60321981	0.01618123	0.08578921	0.06546085	CYFIP1tg	NPC
Inflammatory mediator regulation of TRP channels	59	0.220476435	1.79257043	0.01648352	0.08578921	0.06546085	CYFIP1tg	NPC
Ferroptosis	27	-0.34930952	-1.8224196	0.01873935	0.09536248	0.07276567	CYFIP1tg	NPC

Ribosome biogenesis in eukaryotes	46	-0.27634604	-1.7896398	0.01948052	0.09697911	0.07399923	CYFIP1tg	NPC
Signaling pathways regulating pluripotency of stem cells	77	0.185751974	1.6628844	0.02194357	0.10691656	0.08158193	CYFIP1tg	NPC
NF-kappa B signaling pathway	38	-0.28884542	-1.739592	0.025	0.11927083	0.09100877	CYFIP1tg	NPC
Cell adhesion molecules (CAMs)	45	-0.270677958	-1.738834	0.02605863	0.12123529	0.09250774	CYFIP1tg	NPC
Hippo signaling pathway	83	0.184020589	1.73794953	0.02647059	0.12123529	0.09250774	CYFIP1tg	NPC
Salivary secretion	37	0.253322919	1.68520503	0.02962963	0.13304285	0.10151741	CYFIP1tg	NPC
Cholinergic synapse	65	0.196374185	1.64454301	0.03047091	0.13418922	0.10239214	CYFIP1tg	NPC
Adrenergic signaling in cardiomyocytes	80	0.178761961	1.63627597	0.03303303	0.14272763	0.10890732	CYFIP1tg	NPC
Melanogenesis	62	0.200550467	1.66630594	0.03399433	0.14416116	0.11000116	CYFIP1tg	NPC
Legionellosis	20	-0.357523798	-1.6767958	0.03546099	0.14764668	0.11266076	CYFIP1tg	NPC
Olfactory transduction	21	0.331575261	1.68195483	0.03854875	0.15763686	0.1202837	CYFIP1tg	NPC
Pancreatic secretion	40	0.229410478	1.59650577	0.04086538	0.16417848	0.12527523	CYFIP1tg	NPC
Serotonergic synapse	57	0.201212432	1.60995828	0.04189944	0.16543055	0.12623062	CYFIP1tg	NPC
Neurotrophin signaling pathway	68	0.179997148	1.53403453	0.04558405	0.17692092	0.13499826	CYFIP1tg	NPC

Arachidonic acid metabolism	20	0.330943375	1.65167798	0.04794521	0.17692092	0.13499826	CYFIP1tg	NPC
Bile secretion	31	0.268354167	1.61721707	0.04834606	0.17692092	0.13499826	CYFIP1tg	NPC
Type II diabetes mellitus	28	0.275312603	1.60369036	0.04842615	0.17692092	0.13499826	CYFIP1tg	NPC
Fluid shear stress and atherosclerosis	76	-0.199001689	-1.5825794	0.04867257	0.17692092	0.13499826	CYFIP1tg	NPC
Pathways in cancer	186	-0.229955126	-2.6261617	0.00152439	0.02060185	0.01388889	CYFIP1tg	N
PI3K-Akt signaling pathway	113	-0.223618612	-2.1472949	0.00161031	0.02060185	0.01388889	CYFIP1tg	N
Hippo signaling pathway	80	-0.277089297	-2.3764761	0.00161551	0.02060185	0.01388889	CYFIP1tg	N
Axon guidance	89	-0.25534368	-2.2692678	0.00161812	0.02060185	0.01388889	CYFIP1tg	N
MicroRNAs in cancer	69	-0.293958802	-2.3981257	0.00162866	0.02060185	0.01388889	CYFIP1tg	N
Cushing syndrome	67	-0.27990106	-2.2563086	0.00163934	0.02060185	0.01388889	CYFIP1tg	N
Breast cancer	67	-0.284073223	-2.2899408	0.00163934	0.02060185	0.01388889	CYFIP1tg	N
Signaling pathways regulating pluripotency of stem cells	72	-0.268905887	-2.2312684	0.00164474	0.02060185	0.01388889	CYFIP1tg	N
Chemokine signaling pathway	51	-0.283650838	-2.0366583	0.00165837	0.02060185	0.01388889	CYFIP1tg	N
Calcium signaling pathway	52	-0.328051698	-2.3761781	0.00166113	0.02060185	0.01388889	CYFIP1tg	N

Wnt signaling pathway	66	-0.359915957	-2.8754081	0.00166113	0.02060185	0.01388889	CYFIP1tg	N
Neuroactive ligand-receptor interaction	55	-0.303844924	-2.249912	0.00166389	0.02060185	0.01388889	CYFIP1tg	N
cAMP signaling pathway	58	-0.297934679	-2.2546518	0.00168634	0.02060185	0.01388889	CYFIP1tg	N
Melanogenesis	46	-0.377713144	-2.5682326	0.00170358	0.02060185	0.01388889	CYFIP1tg	N
Cytokine-cytokine receptor interaction	38	-0.480846588	-3.0252619	0.00173611	0.02060185	0.01388889	CYFIP1tg	N
Human papillomavirus infection	129	-0.223284352	-2.241239	0.00312989	0.03184258	0.02146691	CYFIP1tg	N
Human T-cell leukemia virus 1 infection	69	-0.261463733	-2.1330299	0.00325733	0.03184258	0.02146691	CYFIP1tg	N
Cell adhesion molecules (CAMs)	40	-0.346302253	-2.2130686	0.0034965	0.03184258	0.02146691	CYFIP1tg	N
JAK-STAT signaling pathway	34	-0.390229133	-2.3370687	0.00357782	0.03184258	0.02146691	CYFIP1tg	N
Basal cell carcinoma	34	-0.383399097	-2.2961638	0.00357782	0.03184258	0.02146691	CYFIP1tg	N
Peroxisome	29	0.379349581	2.29410737	0.00464037	0.03933267	0.02651641	CYFIP1tg	N
Focal adhesion	71	-0.237710623	-1.9602429	0.00655738	0.05106139	0.03442341	CYFIP1tg	N
Circadian entrainment	37	-0.340668909	-2.1161168	0.00688468	0.05106139	0.03442341	CYFIP1tg	N
Parathyroid hormone synthesis, secretion and action	37	-0.323674299	-2.0105522	0.00688468	0.05106139	0.03442341	CYFIP1tg	N

Notch signaling pathway	22	-0.38006031	-1.9231202	0.00724638	0.0515942	0.03478261	CYFIP1tg	N
Aldosterone synthesis and secretion	31	-0.347905129	-2.0257989	0.00883392	0.06046196	0.04076087	CYFIP1tg	N
Hepatocellular carcinoma	74	-0.231765421	-1.9380549	0.00980392	0.06046196	0.04076087	CYFIP1tg	N
Influenza A	45	-0.288454653	-1.9407175	0.01025641	0.06046196	0.04076087	CYFIP1tg	N
ECM-receptor interaction	32	-0.354463411	-2.0938486	0.01061947	0.06046196	0.04076087	CYFIP1tg	N
Longevity regulating pathway - multiple species	23	-0.379176777	-1.9636378	0.01073345	0.06046196	0.04076087	CYFIP1tg	N
Protein digestion and absorption	25	-0.370772331	-1.9780844	0.01084991	0.06046196	0.04076087	CYFIP1tg	N
Thyroid hormone synthesis	22	-0.37056327	-1.8750648	0.01086957	0.06046196	0.04076087	CYFIP1tg	N
Ras signaling pathway	79	-0.227964592	-1.9470908	0.01134522	0.06119542	0.04125534	CYFIP1tg	N
Measles	38	-0.304833271	-1.9178684	0.01215278	0.06191612	0.0417412	CYFIP1tg	N
Endocrine resistance	30	-0.336132496	-1.9424484	0.01230228	0.06191612	0.0417412	CYFIP1tg	N
Vascular smooth muscle contraction	34	-0.333679358	-1.998394	0.01252236	0.06191612	0.0417412	CYFIP1tg	N
cGMP-PKG signaling pathway	50	-0.263981537	-1.8739812	0.01328904	0.06393104	0.04309958	CYFIP1tg	N
Longevity regulating pathway	31	-0.325111871	-1.8930772	0.01413428	0.06551343	0.04416636	CYFIP1tg	N

MAPK signaling pathway	110	-0.18737678	-1.785946	0.01435407	0.06551343	0.04416636	CYFIP1tg	N
Arginine and proline metabolism	21	0.343008351	1.84302378	0.0173913	0.07491582	0.05050505	CYFIP1tg	N
RaEA8 signaling pathway	76	-0.218645622	-1.8464195	0.01788618	0.07491582	0.05050505	CYFIP1tg	N
Protein processing in endoplasmic reticulum	62	-0.23081666	-1.7920363	0.01836394	0.07491582	0.05050505	CYFIP1tg	N
AMPK signaling pathway	48	-0.260254257	-1.8149607	0.01842546	0.07491582	0.05050505	CYFIP1tg	N
FoxO signaling pathway	54	-0.24447082	-1.7912832	0.01851852	0.07491582	0.05050505	CYFIP1tg	N
Pancreatic secretion	26	-0.33499013	-1.8226297	0.0198556	0.0782217	0.05273373	CYFIP1tg	N
Ribosome	61	-0.230699132	-1.7737497	0.02047782	0.0782217	0.05273373	CYFIP1tg	N
ErbB signaling pathway	37	-0.299301928	-1.8591595	0.02065404	0.0782217	0.05273373	CYFIP1tg	N
Acute myeloid leukemia	29	-0.313839631	-1.8073006	0.02276708	0.0844279	0.05691769	CYFIP1tg	N
Prostate cancer	42	-0.266339778	-1.7531825	0.02401372	0.08622837	0.05813149	CYFIP1tg	N
AGE-RAGE signaling pathway in diabetic complications	39	-0.288156532	-1.8330204	0.02422145	0.08622837	0.05813149	CYFIP1tg	N
Parkinson disease	39	0.255678283	1.78879318	0.0259434	0.09054754	0.06104329	CYFIP1tg	N
Non-alcoholic fatty liver disease (NAFLD)	39	-0.275481816	-1.7523941	0.02768166	0.09475645	0.06388076	CYFIP1tg	N

Thyroid hormone signaling pathway	54	-0.23377089	-1.7128828	0.02861953	0.09611842	0.06479893	CYFIP1tg	N
Adherens junction	40	-0.269637331	-1.7231361	0.02972028	0.09796685	0.06604507	CYFIP1tg	N
Proteoglycans in cancer	87	-0.196099662	-1.7365521	0.03069467	0.09933911	0.06697019	CYFIP1tg	N
Chronic myeloid leukemia	35	-0.283139085	-1.7133239	0.03197158	0.10162395	0.06851053	CYFIP1tg	N
Fc gamma R-mediated phagocytosis	31	-0.292201435	-1.7014447	0.0335689	0.10482921	0.07067138	CYFIP1tg	N
Human cytomegalovirus infection	65	-0.21013152	-1.6825257	0.03453947	0.10600045	0.07146098	CYFIP1tg	N
Endometrial cancer	29	-0.289562416	-1.667496	0.03677758	0.1109561	0.07480186	CYFIP1tg	N
Transcriptional misregulation in cancer	57	-0.221209402	-1.6692795	0.04145937	0.1215847	0.08196721	CYFIP1tg	N
NF-kappa B signaling pathway	22	-0.324006176	-1.6394841	0.04166667	0.1215847	0.08196721	CYFIP1tg	N
Oxytocin signaling pathway	54	-0.221304693	-1.6215407	0.04377104	0.12501756	0.0842815	CYFIP1tg	N
Amoebiasis	32	-0.278886816	-1.6474106	0.04424779	0.12501756	0.0842815	CYFIP1tg	N
Insulin secretion	25	-0.309590975	-1.6516795	0.04520796	0.12573463	0.08476492	CYFIP1tg	N

clusterProfiler GSEA results for CYFIP1tg and CYFIP1ko cells at three developmental stages using a 5% FDR cut-off threshold for DEGs.NE = neuroepithelium, NPC = neural progenitor cells, N = neurons.

Supplementary Table 5. Significant enrichment for AKT in DEGs is independent from FMRP.

Cell line	Stage	AKT genes targeted by FMRP	AKT genes not targeted by FMRP	DEG threshold	Fisher's pval
CYFIP1tg	Neuroepithelium	19	20	pval < 0.05	0.01492832
		135	314	pval > 0.05	
	NPCs	26	14	pval < 0.05	0.10961212
		260	225	pval > 0.05	
	Neurons	19	21	pval < 0.05	0.13700181
		170	285	pval > 0.05	
CYFIP1ko	Neuroepithelium	11	28	pval < 0.05	3.83E-01
		104	315	pval > 0.05	
	NPCs	21	19	pval < 0.05	2.91E-01
		222	254	pval > 0.05	
	Neurons	28	12	pval < 0.05	7.28E-02
		264	200	pval > 0.05	

Significant differentially expressed AKT genes are not enriched amongst FRMP targets (Fisher exact p > 0.05), with exception to CYFIP1tg neuroepithelial

Supplementary Table 6. Significantly enriched KEGG pathways in EA8 and EA62 cells.

Description	setSize	enrichmentScore	NES	pvalue	p.adjust	qvalues	Patient	Stage
Neuroactive ligand-receptor interaction	48	-0.463717924	-2.8111068	0.00146199	0.05333333	0.03859649	EA8	NPC
Chemokine signaling pathway	29	0.369105734	2.21986947	0.00284091	0.05333333	0.03859649	EA8	NPC
Hepatocellular carcinoma	40	0.37966562	2.56473406	0.00288184	0.05333333	0.03859649	EA8	NPC
Cytokine-cytokine receptor interaction	36	0.383166008	2.47527629	0.00295858	0.05333333	0.03859649	EA8	NPC
Fluid shear stress and atherosclerosis	44	0.425487794	2.96339171	0.00296736	0.05333333	0.03859649	EA8	NPC
Proteoglycans in cancer	48	0.3115418	2.25990775	0.00314465	0.05333333	0.03859649	EA8	NPC
Focal adhesion	55	0.365093297	2.81603982	0.00333333	0.05333333	0.03859649	EA8	NPC
Pathways in cancer	120	0.232646242	2.38064403	0.004329	0.06060606	0.04385965	EA8	NPC
Non-alcoholic fatty liver disease (NAFLD)	23	0.421098142	2.28364294	0.0052356	0.06515416	0.04715104	EA8	NPC
Protein processing in endoplasmic reticulum	36	0.317031635	2.04804411	0.0147929	0.13423404	0.09714305	EA8	NPC
Vascular smooth muscle contraction	23	0.346519629	1.87919876	0.01570681	0.13423404	0.09714305	EA8	NPC
Leukocyte transendothelial migration	27	0.340265095	2.00576265	0.01902174	0.13423404	0.09714305	EA8	NPC
Relaxin signaling pathway	36	0.285223651	1.84256256	0.02071006	0.13423404	0.09714305	EA8	NPC
Human T-cell leukemia virus 1 infection	43	0.272870517	1.8780169	0.02089552	0.13423404	0.09714305	EA8	NPC
AGE-RAGE signaling pathway in diabetic complications	28	0.328970977	1.96837641	0.02222222	0.13423404	0.09714305	EA8	NPC
Herpes simplex infection	25	0.306725182	1.73929436	0.024	0.13423404	0.09714305	EA8	NPC
Apoptosis	33	0.305192379	1.93538011	0.0252809	0.13423404	0.09714305	EA8	NPC
Apelin signaling pathway	34	0.295599973	1.90188857	0.02549575	0.13423404	0.09714305	EA8	NPC
p53 signaling pathway	20	0.342230565	1.74672843	0.02583979	0.13423404	0.09714305	EA8	NPC
Biosynthesis of amino acids	21	-0.395569621	-1.786649	0.02614379	0.13423404	0.09714305	EA8	NPC

Glucagon signaling pathway	25	0.299556902	1.69864641	0.02666667	0.13423404	0.09714305	EA8	NPC
PI3K-Akt signaling pathway	74	0.204517361	1.76568706	0.02816901	0.13423404	0.09714305	EA8	NPC
Gastric cancer	31	0.294463342	1.81114286	0.02816901	0.13423404	0.09714305	EA8	NPC
Regulation of actin cytoskeleton	53	0.241740682	1.84324605	0.02912621	0.13423404	0.09714305	EA8	NPC
Amoebiasis	20	0.329336569	1.6809181	0.03100775	0.13423404	0.09714305	EA8	NPC
Kaposi sarcoma-associated herpesvirus infection	34	0.289111484	1.86014166	0.03116147	0.13423404	0.09714305	EA8	NPC
Epstein-Barr virus infection	36	0.275348927	1.77877123	0.03254438	0.1349989	0.09769658	EA8	NPC
Retrograde endocannabinoid signaling	39	-0.312659707	-1.7622096	0.03479576	0.13918306	0.10072458	EA8	NPC
Necroptosis	29	0.29529102	1.77593426	0.03693182	0.14263323	0.10322142	EA8	NPC
Morphine addiction	26	-0.34962333	-1.7068979	0.03930818	0.14675052	0.10620104	EA8	NPC
Glutamatergic synapse	33	-0.322513197	-1.7026306	0.04334365	0.15280899	0.11058545	EA8	NPC
MicroRNAs in cancer	41	0.260666884	1.75871288	0.04464286	0.15280899	0.11058545	EA8	NPC
TNF signaling pathway	29	0.262623759	1.57946738	0.04545455	0.15280899	0.11058545	EA8	NPC
Gap junction	23	0.302191401	1.63880386	0.04712042	0.15280899	0.11058545	EA8	NPC
Cellular senescence	30	0.252685008	1.54825544	0.04775281	0.15280899	0.11058545	EA8	NPC
Signaling pathways regulating pluripotency of stem cells	47	-0.392105498	-2.1616211	0.00133333	0.04375	0.03212074	EA8	N
Calcium signaling pathway	44	-0.424068873	-2.279199	0.00134409	0.04375	0.03212074	EA8	N
Cytokine-cytokine receptor interaction	38	-0.437110644	-2.2457355	0.00136986	0.04375	0.03212074	EA8	N
Insulin resistance	26	-0.479526575	-2.1007556	0.00147059	0.04375	0.03212074	EA8	N
Proteasome	26	0.543681319	2.93062708	0.00310559	0.04837398	0.03551562	EA8	N
Ribosome	30	0.440292397	2.59656482	0.003125	0.04837398	0.03551562	EA8	N
MAPK signaling pathway	66	-0.3395161	-2.0437904	0.00382166	0.04837398	0.03551562	EA8	N
Non-alcoholic fatty liver disease (NAFLD)	58	0.343684586	2.63130924	0.00444444	0.04837398	0.03551562	EA8	N

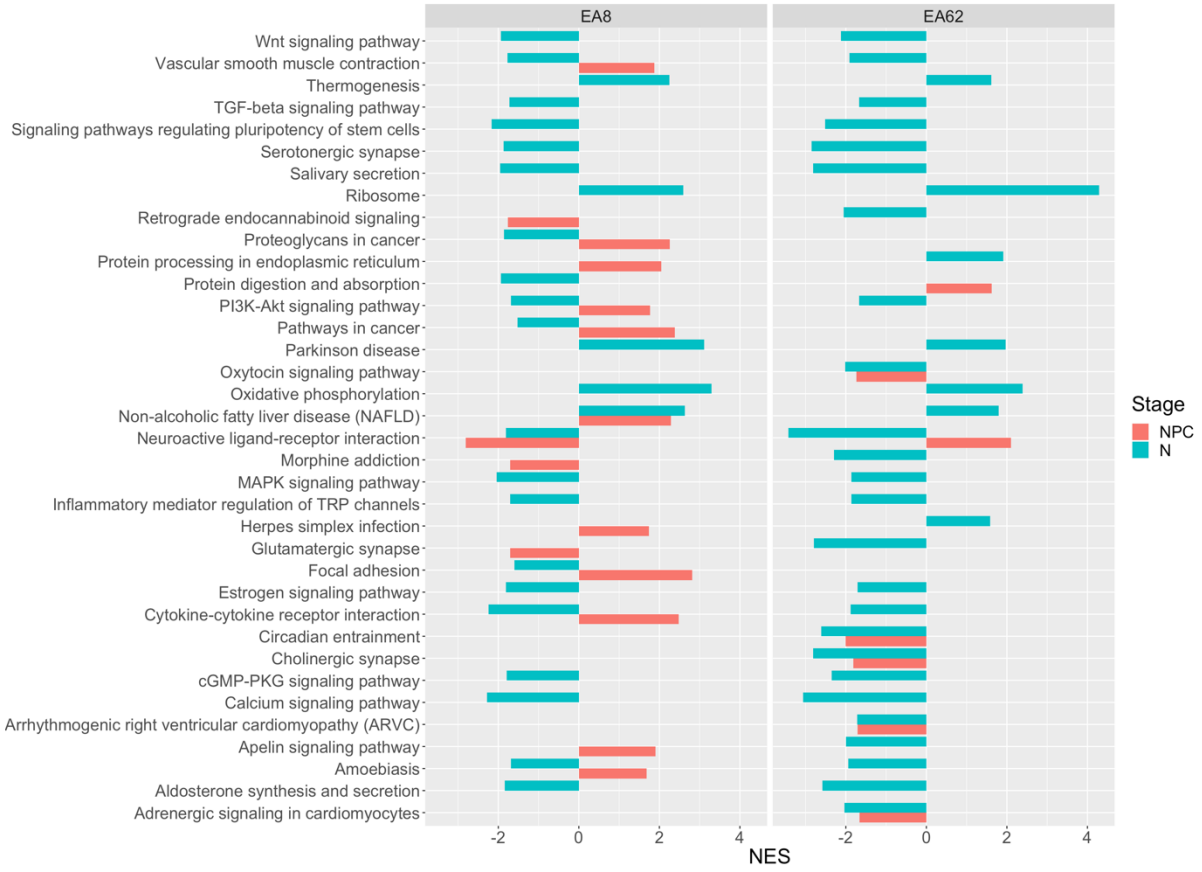
Parkinson disease	68	0.385000969	3.10858323	0.00454545	0.04837398	0.03551562	EA8	N
Oxidative phosphorylation	71	0.401011464	3.29847665	0.00458716	0.04837398	0.03551562	EA8	N
Salivary secretion	21	-0.483471337	-1.9556194	0.00466563	0.04837398	0.03551562	EA8	N
Huntington disease	84	0.312187854	2.67445282	0.00487805	0.04837398	0.03551562	EA8	N
Tight junction	33	-0.431439497	-2.0935394	0.00560224	0.05128205	0.03765069	EA8	N
Protein digestion and absorption	21	-0.478164388	-1.9341531	0.00777605	0.06149871	0.04515164	EA8	N
Serotonergic synapse	26	-0.426184545	-1.8670698	0.00882353	0.06149871	0.04515164	EA8	N
ECM-receptor interaction	29	-0.417631803	-1.8907893	0.0089153	0.06149871	0.04515164	EA8	N
Neuroactive ligand-receptor interaction	63	-0.30614823	-1.8130352	0.00895141	0.06149871	0.04515164	EA8	N
Alzheimer disease	77	0.247491922	2.09112002	0.00930233	0.06149871	0.04515164	EA8	N
Thermogenesis	93	0.252580814	2.24546402	0.01010101	0.06165803	0.04526861	EA8	N
Proteoglycans in cancer	55	-0.322586262	-1.8621071	0.01036269	0.06165803	0.04526861	EA8	N
Aldosterone synthesis and secretion	27	-0.415113306	-1.8424144	0.01186944	0.06630758	0.04868226	EA8	N
Thyroid hormone synthesis	20	-0.462186844	-1.8360167	0.01269841	0.06630758	0.04868226	EA8	N
HIF-1 signaling pathway	26	-0.412705094	-1.8080177	0.01323529	0.06630758	0.04868226	EA8	N
Estrogen signaling pathway	29	-0.398874689	-1.8058682	0.01337296	0.06630758	0.04868226	EA8	N
Wnt signaling pathway	44	-0.359599552	-1.9327024	0.01478495	0.07037634	0.0516695	EA8	N
Hippo signaling pathway	46	-0.340718085	-1.8539334	0.02005348	0.09114284	0.06691602	EA8	N
TGF-beta signaling pathway	28	-0.384740713	-1.7278348	0.02067947	0.09114284	0.06691602	EA8	N
Vascular smooth muscle contraction	32	-0.369482765	-1.7699825	0.02269504	0.0964539	0.07081534	EA8	N
Amoebiasis	29	-0.372926944	-1.6883922	0.02377415	0.09755598	0.07162447	EA8	N
PI3K-Akt signaling pathway	84	-0.265323766	-1.6832107	0.0250941	0.09803156	0.07197364	EA8	N
cGMP-PKG signaling pathway	44	-0.332829628	-1.7888249	0.02553763	0.09803156	0.07197364	EA8	N
Inflammatory mediator regulation of TRP channels	27	-0.384807018	-1.7079048	0.02818991	0.10483123	0.07696588	EA8	N
RaEA8 signaling pathway	46	-0.301918769	-1.6428165	0.04010695	0.1446281	0.10618427	EA8	N

Ras signaling pathway	51	-0.286399181	-1.6139168	0.04623514	0.16169666	0.11871581	EA8	N
Focal adhesion	60	-0.272556336	-1.5999691	0.04755784	0.16169666	0.11871581	EA8	N
Pathways in cancer	120	-0.220968385	-1.5270215	0.04929577	0.16294992	0.11963594	EA8	N
Dilated cardiomyopathy (DCM)	29	-0.441914641	-2.3809315	0.00172117	0.10829103	0.10211654	EA62	NPC
Neuroactive ligand-receptor interaction	36	0.333929641	2.09728007	0.00238095	0.10829103	0.10211654	EA62	NPC
Hypertrophic cardiomyopathy (HCM)	27	-0.432774477	-2.2924067	0.00338409	0.10829103	0.10211654	EA62	NPC
Circadian entrainment	20	-0.427820141	-2.0017598	0.01243339	0.29840142	0.2813873	EA62	NPC
Cholinergic synapse	22	-0.369453306	-1.818082	0.02581756	0.39706712	0.37442733	EA62	NPC
Protein digestion and absorption	27	0.291876769	1.62317538	0.02919708	0.39706712	0.37442733	EA62	NPC
Carbon metabolism	24	-0.337873172	-1.7214683	0.0359589	0.39706712	0.37442733	EA62	NPC
Oxytocin signaling pathway	29	-0.322313732	-1.7365501	0.03614458	0.39706712	0.37442733	EA62	NPC
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	28	-0.319966142	-1.7123642	0.03722504	0.39706712	0.37442733	EA62	NPC
Adrenergic signaling in cardiomyocytes	35	-0.286952768	-1.6587018	0.04835924	0.44918585	0.42357438	EA62	NPC
Neuroactive ligand-receptor interaction	80	-0.494714207	-3.4288045	0.00139276	0.03297975	0.02436177	EA62	N
Calcium signaling pathway	66	-0.464697595	-3.0577766	0.00139665	0.03297975	0.02436177	EA62	N
Signaling pathways regulating pluripotency of stem cells	51	-0.413353765	-2.512028	0.00143885	0.03297975	0.02436177	EA62	N
Cholinergic synapse	43	-0.487359259	-2.8131237	0.00147929	0.03297975	0.02436177	EA62	N
Glutamatergic synapse	46	-0.477066832	-2.792278	0.00148368	0.03297975	0.02436177	EA62	N
Circadian entrainment	38	-0.470222065	-2.6096199	0.00151057	0.03297975	0.02436177	EA62	N
Serotonergic synapse	31	-0.555206257	-2.8489446	0.00153374	0.03297975	0.02436177	EA62	N
Salivary secretion	22	-0.623031325	-2.8163113	0.0015748	0.03297975	0.02436177	EA62	N
cGMP-PKG signaling pathway	51	-0.3866922	-2.3500007	0.0028777	0.03297975	0.02436177	EA62	N
Dopaminergic synapse	49	-0.382845678	-2.2851848	0.00291971	0.03297975	0.02436177	EA62	N

Oxidative phosphorylation	46	0.33421707	2.39002457	0.00304878	0.03297975	0.02436177	EA62	N
Aldosterone synthesis and secretion	32	-0.496061364	-2.5770896	0.0030722	0.03297975	0.02436177	EA62	N
Protein processing in endoplasmic reticulum	48	0.260286586	1.91113025	0.00311526	0.03297975	0.02436177	EA62	N
Parkinson disease	50	0.265198439	1.96489021	0.003125	0.03297975	0.02436177	EA62	N
Morphine addiction	26	-0.477228722	-2.2933663	0.00312989	0.03297975	0.02436177	EA62	N
Insulin secretion	24	-0.519168227	-2.410852	0.00318471	0.03297975	0.02436177	EA62	N
Ribosome	53	0.562641072	4.28951074	0.00327869	0.03297975	0.02436177	EA62	N
Cortisol synthesis and secretion	20	-0.484088179	-2.0939965	0.00481541	0.04574639	0.03379235	EA62	N
Rheumatoid arthritis	21	0.381799195	1.95440368	0.00546448	0.04723757	0.03489386	EA62	N
Peroxisome	27	0.341468716	1.98323202	0.00552486	0.04723757	0.03489386	EA62	N
Wnt signaling pathway	45	-0.364788258	-2.1243988	0.00591716	0.04818259	0.03559194	EA62	N
GABAergic synapse	24	-0.451808191	-2.0980535	0.00636943	0.04923697	0.0363708	EA62	N
Spliceosome	56	0.254210307	1.96788691	0.00662252	0.04923697	0.0363708	EA62	N
Retrograde endocannabinoid signaling	56	-0.326769888	-2.0511439	0.00714286	0.05089286	0.03759398	EA62	N
Oxytocin signaling pathway	61	-0.313489868	-2.0110147	0.00852273	0.05740181	0.04240208	EA62	N
Adrenergic signaling in cardiomyocytes	43	-0.353005757	-2.0376116	0.00887574	0.05740181	0.04240208	EA62	N
Parathyroid hormone synthesis, secretion and action	38	-0.361534414	-2.0064295	0.00906344	0.05740181	0.04240208	EA62	N
cAMP signaling pathway	61	-0.309494434	-1.9853843	0.00994318	0.06072443	0.04485646	EA62	N
Basal cell carcinoma	22	0.345479171	1.80840606	0.01089918	0.06263736	0.04626952	EA62	N
Long-term depression	28	-0.422303655	-2.0797387	0.01098901	0.06263736	0.04626952	EA62	N
Non-alcoholic fatty liver disease (NAFLD)	47	0.247616865	1.79072878	0.0125	0.06895161	0.05093379	EA62	N
Apelin signaling pathway	46	-0.340218114	-1.9913008	0.01335312	0.07135571	0.05270967	EA62	N
RNA degradation	30	0.291319181	1.76425658	0.01392758	0.07217017	0.0533113	EA62	N

Long-term potentiation	30	-0.397263131	-1.9992855	0.0155521	0.07645875	0.05647923	EA62	N
Amoebiasis	26	-0.403245909	-1.9378351	0.01564945	0.07645875	0.05647923	EA62	N
MAPK signaling pathway	94	-0.253168226	-1.8592088	0.01735648	0.08244326	0.06089991	EA62	N
Vascular smooth muscle contraction	38	-0.344636389	-1.9126495	0.01812689	0.08377562	0.06188411	EA62	N
Inflammatory mediator regulation of TRP channels	34	-0.351864518	-1.8621905	0.02173913	0.09782609	0.07226304	EA62	N
Platelet activation	34	-0.345013996	-1.8259351	0.02329193	0.0995889	0.07356521	EA62	N
Gastric acid secretion	25	-0.403767201	-1.9044188	0.02340094	0.0995889	0.07356521	EA62	N
Mitophagy - animal	24	0.296773803	1.61438414	0.02406417	0.0995889	0.07356521	EA62	N
Cytokine-cytokine receptor interaction	51	-0.308891779	-1.877193	0.02446043	0.0995889	0.07356521	EA62	N
Adherens junction	33	-0.342886258	-1.7991482	0.02936631	0.11678229	0.08626577	EA62	N
Pancreatic secretion	26	-0.375334953	-1.8037065	0.03286385	0.12700535	0.09381743	EA62	N
PI3K-Akt signaling pathway	93	-0.227534144	-1.671921	0.03342246	0.12700535	0.09381743	EA62	N
Thermogenesis	85	0.18309286	1.61553254	0.03971119	0.14762204	0.10904675	EA62	N
Estrogen signaling pathway	46	-0.291850727	-1.7082059	0.04154303	0.15114591	0.11164979	EA62	N
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	23	-0.373805745	-1.7200503	0.04388715	0.15634796	0.11549249	EA62	N
TGF-beta signaling pathway	28	-0.339908623	-1.673964	0.0455259	0.15887611	0.11736001	EA62	N
Cell adhesion molecules (CAMs)	40	-0.297886505	-1.6915642	0.04661654	0.15942857	0.1177681	EA62	N
Herpes simplex infection	55	0.206002135	1.58532278	0.04966887	0.16653681	0.12301888	EA62	N

clusterProfiler GSEA results for EA8 and EA62 cells at three developmental stages using a 5% FDR cut-off threshold for DEGs. NPC = neural progenitor cells, N = neurons.



Supplementary Figure 4. Overlapping altered pathways between patient-derived iPSCs and CYFIP1 engineered hESCs.

(A) Bar graph representing the significantly enriched pathways in CYFIP1tg and CYFIPko cells at different stages of differentiation. Normalised enrichment score for each pathway show similar trends in the enrichment between each cell line.

Supplementary Table 7. Significantly enriched KEGG pathways in EA8 and EA62 cells

Description	setSize	enrichmentScore	NES	pvalue	p.adjust	qvalues	Patient	Stage
Neuroactive ligand-receptor interaction	48	-0.463717924	-2.8111068	0.00146199	0.05333333	0.03859649	EA8	NPC
Chemokine signaling pathway	29	0.369105734	2.21986947	0.00284091	0.05333333	0.03859649	EA8	NPC
Hepatocellular carcinoma	40	0.37966562	2.56473406	0.00288184	0.05333333	0.03859649	EA8	NPC
Cytokine-cytokine receptor interaction	36	0.383166008	2.47527629	0.00295858	0.05333333	0.03859649	EA8	NPC
Fluid shear stress and atherosclerosis	44	0.425487794	2.96339171	0.00296736	0.05333333	0.03859649	EA8	NPC
Proteoglycans in cancer	48	0.3115418	2.25990775	0.00314465	0.05333333	0.03859649	EA8	NPC
Focal adhesion	55	0.365093297	2.81603982	0.00333333	0.05333333	0.03859649	EA8	NPC
Pathways in cancer	120	0.232646242	2.38064403	0.004329	0.06060606	0.04385965	EA8	NPC
Non-alcoholic fatty liver disease (NAFLD)	23	0.421098142	2.28364294	0.0052356	0.06515416	0.04715104	EA8	NPC
Protein processing in endoplasmic reticulum	36	0.317031635	2.04804411	0.0147929	0.13423404	0.09714305	EA8	NPC
Vascular smooth muscle contraction	23	0.346519629	1.87919876	0.01570681	0.13423404	0.09714305	EA8	NPC
Leukocyte transendothelial migration	27	0.340265095	2.00576265	0.01902174	0.13423404	0.09714305	EA8	NPC

Relaxin signaling pathway	36	0.285223651	1.84256256	0.02071006	0.13423404	0.09714305	EA8	NPC
Human T-cell leukemia virus 1 infection	43	0.272870517	1.8780169	0.02089552	0.13423404	0.09714305	EA8	NPC
AGE-RAGE signaling pathway in diabetic complications	28	0.328970977	1.96837641	0.02222222	0.13423404	0.09714305	EA8	NPC
Herpes simplex infection	25	0.306725182	1.73929436	0.024	0.13423404	0.09714305	EA8	NPC
Apoptosis	33	0.305192379	1.93538011	0.0252809	0.13423404	0.09714305	EA8	NPC
Apelin signaling pathway	34	0.295599973	1.90188857	0.02549575	0.13423404	0.09714305	EA8	NPC
p53 signaling pathway	20	0.342230565	1.74672843	0.02583979	0.13423404	0.09714305	EA8	NPC
Biosynthesis of amino acids	21	-0.395569621	-1.786649	0.02614379	0.13423404	0.09714305	EA8	NPC
Glucagon signaling pathway	25	0.299556902	1.69864641	0.02666667	0.13423404	0.09714305	EA8	NPC
PI3K-Akt signaling pathway	74	0.204517361	1.76568706	0.02816901	0.13423404	0.09714305	EA8	NPC
Gastric cancer	31	0.294463342	1.81114286	0.02816901	0.13423404	0.09714305	EA8	NPC
Regulation of actin cytoskeleton	53	0.241740682	1.84324605	0.02912621	0.13423404	0.09714305	EA8	NPC
Amoebiasis	20	0.329336569	1.6809181	0.03100775	0.13423404	0.09714305	EA8	NPC

Kaposi sarcoma-associated herpesvirus infection	34	0.289111484	1.86014166	0.03116147	0.13423404	0.09714305	EA8	NPC
Epstein-Barr virus infection	36	0.275348927	1.77877123	0.03254438	0.1349989	0.09769658	EA8	NPC
Retrograde endocannabinoid signaling	39	-0.312659707	-1.7622096	0.03479576	0.13918306	0.10072458	EA8	NPC
Necroptosis	29	0.29529102	1.77593426	0.03693182	0.14263323	0.10322142	EA8	NPC
Morphine addiction	26	-0.34962333	-1.7068979	0.03930818	0.14675052	0.10620104	EA8	NPC
Glutamatergic synapse	33	-0.322513197	-1.7026306	0.04334365	0.15280899	0.11058545	EA8	NPC
MicroRNAs in cancer	41	0.260666884	1.75871288	0.04464286	0.15280899	0.11058545	EA8	NPC
TNF signaling pathway	29	0.262623759	1.57946738	0.04545455	0.15280899	0.11058545	EA8	NPC
Gap junction	23	0.302191401	1.63880386	0.04712042	0.15280899	0.11058545	EA8	NPC
Cellular senescence	30	0.252685008	1.54825544	0.04775281	0.15280899	0.11058545	EA8	NPC
Signaling pathways regulating pluripotency of stem cells	47	-0.392105498	-2.1616211	0.00133333	0.04375	0.03212074	EA8	N
Calcium signaling pathway	44	-0.424068873	-2.279199	0.00134409	0.04375	0.03212074	EA8	N

Cytokine-cytokine receptor interaction	38	-0.437110644	-2.2457355	0.00136986	0.04375	0.03212074	EA8	N
Insulin resistance	26	-0.479526575	-2.1007556	0.00147059	0.04375	0.03212074	EA8	N
Proteasome	26	0.543681319	2.93062708	0.00310559	0.04837398	0.03551562	EA8	N
Ribosome	30	0.440292397	2.59656482	0.003125	0.04837398	0.03551562	EA8	N
MAPK signaling pathway	66	-0.3395161	-2.0437904	0.00382166	0.04837398	0.03551562	EA8	N
Non-alcoholic fatty liver disease (NAFLD)	58	0.343684586	2.63130924	0.00444444	0.04837398	0.03551562	EA8	N
Parkinson disease	68	0.385000969	3.10858323	0.00454545	0.04837398	0.03551562	EA8	N
Oxidative phosphorylation	71	0.401011464	3.29847665	0.00458716	0.04837398	0.03551562	EA8	N
Salivary secretion	21	-0.483471337	-1.9556194	0.00466563	0.04837398	0.03551562	EA8	N
Huntington disease	84	0.312187854	2.67445282	0.00487805	0.04837398	0.03551562	EA8	N
Tight junction	33	-0.431439497	-2.0935394	0.00560224	0.05128205	0.03765069	EA8	N
Protein digestion and absorption	21	-0.478164388	-1.9341531	0.00777605	0.06149871	0.04515164	EA8	N
Serotonergic synapse	26	-0.426184545	-1.8670698	0.00882353	0.06149871	0.04515164	EA8	N

ECM-receptor interaction	29	-0.417631803	-1.8907893	0.0089153	0.06149871	0.04515164	EA8	N
Neuroactive ligand-receptor interaction	63	-0.30614823	-1.8130352	0.00895141	0.06149871	0.04515164	EA8	N
Alzheimer disease	77	0.247491922	2.09112002	0.00930233	0.06149871	0.04515164	EA8	N
Thermogenesis	93	0.252580814	2.24546402	0.01010101	0.06165803	0.04526861	EA8	N
Proteoglycans in cancer	55	-0.322586262	-1.8621071	0.01036269	0.06165803	0.04526861	EA8	N
Aldosterone synthesis and secretion	27	-0.415113306	-1.8424144	0.01186944	0.06630758	0.04868226	EA8	N
Thyroid hormone synthesis	20	-0.462186844	-1.8360167	0.01269841	0.06630758	0.04868226	EA8	N
HIF-1 signaling pathway	26	-0.412705094	-1.8080177	0.01323529	0.06630758	0.04868226	EA8	N
Estrogen signaling pathway	29	-0.398874689	-1.8058682	0.01337296	0.06630758	0.04868226	EA8	N
Wnt signaling pathway	44	-0.359599552	-1.9327024	0.01478495	0.07037634	0.0516695	EA8	N
Hippo signaling pathway	46	-0.340718085	-1.8539334	0.02005348	0.09114284	0.06691602	EA8	N
TGF-beta signaling pathway	28	-0.384740713	-1.7278348	0.02067947	0.09114284	0.06691602	EA8	N
Vascular smooth muscle contraction	32	-0.369482765	-1.7699825	0.02269504	0.0964539	0.07081534	EA8	N

Amoebiasis	29	-0.372926944	-1.6883922	0.02377415	0.09755598	0.07162447	EA8	N
PI3K-Akt signaling pathway	84	-0.265323766	-1.6832107	0.0250941	0.09803156	0.07197364	EA8	N
cGMP-PKG signaling pathway	44	-0.332829628	-1.7888249	0.02553763	0.09803156	0.07197364	EA8	N
Inflammatory mediator regulation of TRP channels	27	-0.384807018	-1.7079048	0.02818991	0.10483123	0.07696588	EA8	N
RaEA8 signaling pathway	46	-0.301918769	-1.6428165	0.04010695	0.1446281	0.10618427	EA8	N
Ras signaling pathway	51	-0.286399181	-1.6139168	0.04623514	0.16169666	0.11871581	EA8	N
Focal adhesion	60	-0.272556336	-1.5999691	0.04755784	0.16169666	0.11871581	EA8	N
Pathways in cancer	120	-0.220968385	-1.5270215	0.04929577	0.16294992	0.11963594	EA8	N
Dilated cardiomyopathy (DCM)	29	-0.441914641	-2.3809315	0.00172117	0.10829103	0.10211654	EA62	NPC
Neuroactive ligand-receptor interaction	36	0.333929641	2.09728007	0.00238095	0.10829103	0.10211654	EA62	NPC
Hypertrophic cardiomyopathy (HCM)	27	-0.432774477	-2.2924067	0.00338409	0.10829103	0.10211654	EA62	NPC
Circadian entrainment	20	-0.427820141	-2.0017598	0.01243339	0.29840142	0.2813873	EA62	NPC
Cholinergic synapse	22	-0.369453306	-1.818082	0.02581756	0.39706712	0.37442733	EA62	NPC

Protein digestion and absorption	27	0.291876769	1.62317538	0.02919708	0.39706712	0.37442733	EA62	NPC
Carbon metabolism	24	-0.337873172	-1.7214683	0.0359589	0.39706712	0.37442733	EA62	NPC
Oxytocin signaling pathway	29	-0.322313732	-1.7365501	0.03614458	0.39706712	0.37442733	EA62	NPC
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	28	-0.319966142	-1.7123642	0.03722504	0.39706712	0.37442733	EA62	NPC
Adrenergic signaling in cardiomyocytes	35	-0.286952768	-1.6587018	0.04835924	0.44918585	0.42357438	EA62	NPC
Neuroactive ligand-receptor interaction	80	-0.494714207	-3.4288045	0.00139276	0.03297975	0.02436177	EA62	N
Calcium signaling pathway	66	-0.464697595	-3.0577766	0.00139665	0.03297975	0.02436177	EA62	N
Signaling pathways regulating pluripotency of stem cells	51	-0.413353765	-2.512028	0.00143885	0.03297975	0.02436177	EA62	N
Cholinergic synapse	43	-0.487359259	-2.8131237	0.00147929	0.03297975	0.02436177	EA62	N
Glutamatergic synapse	46	-0.477066832	-2.792278	0.00148368	0.03297975	0.02436177	EA62	N
Circadian entrainment	38	-0.470222065	-2.6096199	0.00151057	0.03297975	0.02436177	EA62	N
Serotonergic synapse	31	-0.555206257	-2.8489446	0.00153374	0.03297975	0.02436177	EA62	N

Salivary secretion	22	-0.623031325	-2.8163113	0.0015748	0.03297975	0.02436177	EA62	N
cGMP-PKG signaling pathway	51	-0.3866922	-2.3500007	0.0028777	0.03297975	0.02436177	EA62	N
Dopaminergic synapse	49	-0.382845678	-2.2851848	0.00291971	0.03297975	0.02436177	EA62	N
Oxidative phosphorylation	46	0.33421707	2.39002457	0.00304878	0.03297975	0.02436177	EA62	N
Aldosterone synthesis and secretion	32	-0.496061364	-2.5770896	0.0030722	0.03297975	0.02436177	EA62	N
Protein processing in endoplasmic reticulum	48	0.260286586	1.91113025	0.00311526	0.03297975	0.02436177	EA62	N
Parkinson disease	50	0.265198439	1.96489021	0.003125	0.03297975	0.02436177	EA62	N
Morphine addiction	26	-0.477228722	-2.2933663	0.00312989	0.03297975	0.02436177	EA62	N
Insulin secretion	24	-0.519168227	-2.410852	0.00318471	0.03297975	0.02436177	EA62	N
Ribosome	53	0.562641072	4.28951074	0.00327869	0.03297975	0.02436177	EA62	N
Cortisol synthesis and secretion	20	-0.484088179	-2.0939965	0.00481541	0.04574639	0.03379235	EA62	N
Rheumatoid arthritis	21	0.381799195	1.95440368	0.00546448	0.04723757	0.03489386	EA62	N
Peroxisome	27	0.341468716	1.98323202	0.00552486	0.04723757	0.03489386	EA62	N

Wnt signaling pathway	45	-0.364788258	-2.1243988	0.00591716	0.04818259	0.03559194	EA62	N
GABAergic synapse	24	-0.451808191	-2.0980535	0.00636943	0.04923697	0.0363708	EA62	N
Spliceosome	56	0.254210307	1.96788691	0.00662252	0.04923697	0.0363708	EA62	N
Retrograde endocannabinoid signaling	56	-0.326769888	-2.0511439	0.00714286	0.05089286	0.03759398	EA62	N
Oxytocin signaling pathway	61	-0.313489868	-2.0110147	0.00852273	0.05740181	0.04240208	EA62	N
Adrenergic signaling in cardiomyocytes	43	-0.353005757	-2.0376116	0.00887574	0.05740181	0.04240208	EA62	N
Parathyroid hormone synthesis, secretion and action	38	-0.361534414	-2.0064295	0.00906344	0.05740181	0.04240208	EA62	N
cAMP signaling pathway	61	-0.309494434	-1.9853843	0.00994318	0.06072443	0.04485646	EA62	N
Basal cell carcinoma	22	0.345479171	1.80840606	0.01089918	0.06263736	0.04626952	EA62	N
Long-term depression	28	-0.422303655	-2.0797387	0.01098901	0.06263736	0.04626952	EA62	N
Non-alcoholic fatty liver disease (NAFLD)	47	0.247616865	1.79072878	0.0125	0.06895161	0.05093379	EA62	N
Apelin signaling pathway	46	-0.340218114	-1.9913008	0.01335312	0.07135571	0.05270967	EA62	N
RNA degradation	30	0.291319181	1.76425658	0.01392758	0.07217017	0.0533113	EA62	N

Long-term potentiation	30	-0.397263131	-1.9992855	0.0155521	0.07645875	0.05647923	EA62	N
Amoebiasis	26	-0.403245909	-1.9378351	0.01564945	0.07645875	0.05647923	EA62	N
MAPK signaling pathway	94	-0.253168226	-1.8592088	0.01735648	0.08244326	0.06089991	EA62	N
Vascular smooth muscle contraction	38	-0.344636389	-1.9126495	0.01812689	0.08377562	0.06188411	EA62	N
Inflammatory mediator regulation of TRP channels	34	-0.351864518	-1.8621905	0.02173913	0.09782609	0.07226304	EA62	N
Platelet activation	34	-0.345013996	-1.8259351	0.02329193	0.0995889	0.07356521	EA62	N
Gastric acid secretion	25	-0.403767201	-1.9044188	0.02340094	0.0995889	0.07356521	EA62	N
Mitophagy - animal	24	0.296773803	1.61438414	0.02406417	0.0995889	0.07356521	EA62	N
Cytokine-cytokine receptor interaction	51	-0.308891779	-1.877193	0.02446043	0.0995889	0.07356521	EA62	N
Adherens junction	33	-0.342886258	-1.7991482	0.02936631	0.11678229	0.08626577	EA62	N
Pancreatic secretion	26	-0.375334953	-1.8037065	0.03286385	0.12700535	0.09381743	EA62	N
PI3K-Akt signaling pathway	93	-0.227534144	-1.671921	0.03342246	0.12700535	0.09381743	EA62	N
Thermogenesis	85	0.18309286	1.61553254	0.03971119	0.14762204	0.10904675	EA62	N

Estrogen signaling pathway	46	-0.291850727	-1.7082059	0.04154303	0.15114591	0.11164979	EA62	N
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	23	-0.373805745	-1.7200503	0.04388715	0.15634796	0.11549249	EA62	N
TGF-beta signaling pathway	28	-0.339908623	-1.673964	0.0455259	0.15887611	0.11736001	EA62	N
Cell adhesion molecules (CAMs)	40	-0.297886505	-1.6915642	0.04661654	0.15942857	0.1177681	EA62	N
Herpes simplex infection	55	0.206002135	1.58532278	0.04966887	0.16653681	0.12301888	EA62	N

clusterProfiler GSEA results for EA8 and EA62 cells at three developmental stages using a 5% FDR cut-off threshold for DEGs. NPC = neural progenitor cells, N = neurons.