THE UNIVERSITY
of EDINBURGH

Expression of genes in the *16p11.2* locus during human fetal cortical neurogenesis

Sarah E. Morson

For the degree of Doctor of Philosophy

The University of Edinburgh

2019

Declaration

I declare that this thesis is an original report of my own research, has been written by me and has not been submitted for any previous degree with the following exceptions: part of the analysis in chapter three was included in a BSc Honours dissertation by Erin Boyle, and the bioinformatics analysis was performed by Yifei Yang, a fellow PhD student; both of these instances are clearly stated in the text.

Work from all three results chapters has been included in a publicly available manuscript which is included in the appendix.

Signed: …… Date: …….

Acknowledgements

This thesis could not have been completed without the help, support and guidance of so many people. Firstly, I wish to thank my supervisors Tom Pratt and David Price for all their help, ever since I was a summer student and throughout my PhD. They have taught me how to be a scientist, to ask questions and design experiments, interpret my results and supported me at every stage. This is also true of my thesis chair, John Mason, who has always supported me throughout my PhD. Thank you all of you.

Thank you to my parents for your unwavering support throughout this. For asking questions about my thesis and encouraging my scientific interests from as early as I can remember. And to my brothers for your encouragement even though we all know you both have no idea what I do.

Kathy and Mike, I could not have managed in the lab without you, thank you so much for teaching me everything and Viv and Louise your help with histology was invaluable. Also, I am grateful to Yifei Yang for his assistance with the bioinformatics analysis, which provided the foundation to my PhD project.

My friends have been the most incredible support system throughout my PhD and I couldn't have made it through without you: Tian, Idoia, Farhanna, Rosie, Rana, Calvin, Zrinko, Kai, Jonothon, Tiago, Laura and Laura, Caroline, Sandy, Jorge, Alazne, Michaela, Alex, Matilda, Jack, Connor and Ailwyn. Thank you all for always being there, making me laugh, getting drunk with me and listening to me complain incessantly about my antibodies and thesis. I'm so grateful for our friendships.

I wish to thank the Node for providing me with my internship that refuelled my passion for developmental biology. Finally thank you to my funding council

EASTBIO for your scholarship that allowed me to fulfil my dreams of completing a PhD.

Abstract

The process of the brain developing from a single fertilized egg to the most sophisticated known organ requires precise spatial and temporal control to produce the necessary correct brain size and architecture. A particular region of interest is the cerebral cortex, responsible for higher functions such as language, reasoning and conscious thought. Its expansion in size and complexity from smaller mammals, such as mice, to humans is thought to contribute to our higher functions. However, a caveat of this increased complexity is the increased challenge of generating such a complex structure, and the potential for subtle changes during neurodevelopment to manifest in neurodevelopmental disorders such as Autism Spectrum Disorders (ASD).

ASD is a spectrum disorder diagnosed early in childhood based on a range of diagnostic criteria. It is frequently characterised by impaired social interaction, repetitive behaviour and delayed development. While ASD patients share some symptoms, the genetic underpinnings of ASD are highly heterogeneous, with mutations to many single genes or larger genetic regions implicated as ASD risk factors.

The 593kbp *16p11.2* locus encompasses 29 protein coding genes and its copy number variation (CNV) by heterozygous microduplication or microdeletion is implicated in around 1% of ASD cases, with many patients born with macrocephaly (deletion) or microcephaly (duplication), potentially indicating a possible problem with generating the correct number of brain cells during development. This suggests the hypothesis that some of the *16p11.2* region genes are involved in neural proliferation in early corticogenesis, and changes to the levels of these genes may affect proliferation, contributing to the *16p11.2* patient phenotype. This hypothesis is supported by a *16p11.2* deletion mouse model which exhibits ASD-like symptoms and altered proliferation in the cortex during embryonic

development. Given that the *16p11.2* CNV's 1% autism incidence makes it the most frequent aetiology of ASD, this region is a promising area of study to understand how genetic dysregulation during critical prenatal cortical neurogenesis can contribute to the ASD phenotype.

Despite its strong association with ASD, very little is known about the majority of the *16p11.2* genes, especially regarding brain development. In this thesis, focussing on the developing human neocortex, we aimed to identify which, if any, of these 29 genes were expressed in progenitor cells and describe their expression pattern during critical stages of cortical neurogenesis.

We first used a bioinformatics approach to identify *16p11.*2 genes expressed in progenitors to narrow down from the 29 genes in the region. We analysed a publicly available single-cell RNA sequencing (scRNA-seq) dataset from the proliferative zones, the ventricular zone and subventricular zone, of the 16-18 gestational week (GW) human fetal cortex. We identified six genes as being highly expressed in the cortical progenitor cells, and two as being significantly higher expressed in progenitors compared to post mitotic cells: *KIF22* and *ALDOA*. We described their protein expression pattern *in vivo* at key stages of human fetal cortex development.

We showed KIF22 protein to be expressed in the germinative zones, and its expression to be restricted to proliferating cells, suggesting a role for this protein in proliferation. We showed KIF22 protein levels to vary with the cell cycle, increasing from G1 through S and G2 phases to peak in mitosis. This suggests that changing the KIF22 protein level, as in the microduplication or microdeletion patients, will affect cell cycle and proliferation manifesting in changes to cortical size and architecture contributing to the *16p11.2* phenotype.

ALDOA protein was shown to be present throughout the cortex, although higher in the proliferating regions. We demonstrated that ALDOA is predominantly localised to cytoplasm, and its protein levels or sub-cellular localisation do not change in proliferating or non-proliferating cells. ALDOA has a critical role in energy metabolism, and we can hypothesise that due to its expression throughout the cortex any changes to the ALDOA protein by the *16p11.2* CNV will induce a wide range of effects on brain development

In conclusion we have identified two genes highly expressed in progenitors and expressed at much lower levels in post-mitotic cells from the *16p11.2* locus. These genes provide interesting targets for future studies to elucidate the mechanism by which they mediate proliferation and the effects of manipulating their protein levels. This is outwith the scope of this PhD thesis however a range of new techniques are emerging such as cerebral organoids, which can be easily manipulated. These will be a powerful tool to address the hypotheses produced by the descriptive work of this PhD thesis.

Lay Summary

The nine-month process that starts with a single fertilised egg and ends with a newborn human is a truly remarkable one. We are particularly interested in the development of the human brain during this nine-month period; to understand the mechanisms that precisely control the formation of the brain. Over many years and using many different techniques, models and approaches, scientists from research groups around the world are unravelling the multitude of processes which control brain development. Our work is particularly focussed on the region of the brain called the cerebral cortex. This is the part of the brain responsible for thought, language and much of our information processing. The cerebral cortex in humans is particularly enlarged when compared to small mammals. It is often postulated that it is our enlarged cerebral cortex is what makes us unique as humans.

Sometimes however, there are disruptions to the complex process of brain development. These disruptions can lead to neurodevelopmental disorders: perhaps one of the most well-known, and the subject of this thesis, is Autism Spectrum Disorder (ASD). Symptoms of ASD vary between patients (as does severity), however the disorder is typically characterised by impaired social behaviour, repetitive and obsessive behaviours, and delayed learning and development. Research to identify genetic mutations that may lead to ASD has identified hundreds of possible genetic changes. Genetic mutations associated with ASD are extremely diverse. For example, the most common genetic mutation associated with ASD, the *16p11.2* copy number variation (CNV), has only been found in 1% of ASD patients.

Our research and this thesis focus on this most common genetic cause of ASD, the *16p11.2* CNV. The *16p11.2* region is located on the 16th chromosome of humans and contains 29 genes. The *16p11.2* CNV is a heterozygous mutation, meaning that of the pair of chromosomes, one of the

two chromosomes is affected by the mutation whilst the other is unaffected. This heterozygous copy number variation can occur as either a duplication or a deletion. In the duplication, the region is duplicated on one chromosome resulting in three copies of each gene. In the deletion, the region is deleted from one chromosome resulting in only one copy of each gene.

Around 1% of ASD patients have the *16p11.2* CNV, making it the most frequent known genetic cause of autism. Research is focused on understanding how the changes to gene level, or dosage, during development, can cause ASD in patients. One clue of this disruption is evident at birth: many *16p11.2* CNV patients are born with changes to head size. In patients with the duplication, many are born with a smaller than average head - known as microcephaly. Conversely, many patients with the deletion are born with a larger than average head - macrocephaly. These changes to head size (indicative of changes to brain size) suggest that the *16p11.2* CNV alters brain development in some way.

The final cerebral cortex should possess both correct overall cell number and proportions of the different cell types. This requires a balance of cells dividing (proliferating) to produce correct final cell number, and cells to specialise (differentiation) to produce correct proportion of cell types. In *16p11.2* CNV patients, the changes to brain size suggest that something may have altered this balance. Therefore, it is possible these patients may have altered final cell number and proportions which could contribute to their ASD.

We wanted to investigate how the *16p11.2* CNV could impact human brain development, and this balance of proliferation and differentiation. As the *16p11.2* region contains 29 genes, they are likely to have many effects throughout the body at different stages of life. Therefore, the first step for our research was to identify individual genes from this region which were active during the proliferation stage of pre-natal cortical development, in a cell class

known as progenitor cells. The majority of neurons in the final cerebral cortex are descended from early progenitor cells, therefore it is likely that any alteration to progenitors during early cortical development will have many effects on the cerebral cortex.

The first aim of our study was to identify genes from the *16p11.2* region which are switched on in proliferating progenitor cells and are then switched off as these cells differentiate into neurons and stop dividing. To identify these genes, we used a computational approach. We used a previously published dataset which provided us with information about whether a gene was switched on or off in different cell types in the developing brain. From this we were able to show that two of the *16p11.2* region genes were switched on and expressed at high levels in cells which were proliferating, and switched off in cells which had stopped dividing to become neurons. These two genes were *KIF22* and *ALDOA.*

The gene eventually encodes a protein, and it is this protein that exerts an effect within the cell. Therefore, we wanted to study the roles of KIF22 and ALDOA proteins in human cerebral cortex development. This thesis is a descriptive work: we aimed to describe the normal expression pattern of these two proteins in the human cerebral cortex during the late-first and early-second trimester. By understanding the normal roles that genes and proteins play within development, we can later consider how mutations to them may adversely affect development.

We started by examining the KIF22 protein and where it is found in the cortex. We found it to be exclusively expressed in the proliferating progenitor cells, and to be absent from cells which had stopped dividing and differentiated into neurons. We also found that the amount of KIF22 protein present in these progenitor cells varies as cells go through the preparatory stages before they divide – it is highest when the cells undergo division.

This finding tells us that this KIF22 protein likely has an important role in controlling divisions of progenitor cells. Altering the amount of KIF22 protein as in the *16p11.2* CNV may cause changes to progenitor proliferation. For example, they could prematurely become neurons which could reduce the overall number of cells in the final cerebral cortex.

We next looked at the second protein identified: ALDOA. Unlike KIF22, we found this protein to be present in both dividing progenitors and neurons that had ceased dividing. This suggests ALDOA protein to have many different roles in these different cell types and we think that altering it will affect many aspects of brain development.

Our research has studied the *16p11.2* copy number variation, which is known to be mutated in 1% of autism patients. We questioned which of the genes from the *16p11.2* region are expressed in actively dividing progenitor cells of the developing human cerebral cortex, and found two. We would next like to study the effects of changing the level at which these genes are expressed in the developing brain. However, we cannot do this in the developing human brain so we will need to use other models. To answer these questions we have proposed from our findings in this study, future work could make use of “mini-brains” grown in the laboratory from stem cells, which would allow us to model and alter some aspects of human brain development.

Table of Contents

Table of figures

List of Abbreviations

ABC - Avidin-Biotin Complex
ADHD - Attention Deficiency Hyperactivity Disorder
ADP - Adenosine Diphosphate
AIP - Apical Intermediate Progenitor
ANOVA - Analysis of Variance
ARGC - Apical Radial Glial Cell
ARRAY CGH - Array Comparative Genomic Hybridisation
ASD - Autism Spectrum Disorder
ATP - Adenosine Triphosphate
BIP - Basal Intermediate Progenitor
BLAST - Basic Local Alignment Search Tool
BP = Base Pair
BRGC - Basal Radial Glia Cell
CDK - Cyclin Dependent Kinase
CDNA - Complementary DNA
CGE - Caudal Ganglionic Eminence
CH - Cortical Hem
CN - Caudate Nucleus
CNV - Copy Number Variation
CP - Cortical Plate
DAB - 3,3'-Diaminobenzidine
DAPI - 4′,6-Diamidino-2-Phenylindole
DD - Double Distilled
DEG - Differentially Expressed Genes
DH - Dorsal Hippocampus
DNA - Deoxyribose Nucleic Acid
DSM - Diagnostic and Statistical Manual
FAA - Formalin-Acetic-Acid-Alcohol
FIJI - FIJI Is Just ImageJ
FoSTeS = Fork Stalling and Template Switching

GABA - Gamma-Aminobutyric Acid

GE - Ganglionic Eminence

GO - Gene Ontology

GW - Gestational Weeks

H&E - Haematoxylin and Eosin

HCx - Hippocampal Cortex

HDBR - Human Developmental Biology Resource

hIPSCs - Human Induced Pluripotent Stem Cells

HIV - Human Immunodeficiency Virus

IC - Internal Capsule

ID - Intellectual Disability

IF - Immunofluorescence

IHC - Immunohistochemistry

IKNM - Interkinetic Nuclear Migration

IP - Intermediate Progenitor Cell

IPSCs - Induced Pluripotent Stem Cells

IZ - Intermediate Zone

kpb - Kilobase Pairs

LGE - Lateral Ganglion Eminence

LV - Lateral Ventricle

MGE - Medial Ganglion Eminence

mM - Millimolar

mm - Millimetre

MRI - Magnetic Resonance Imaging

mRNA Messenger RNA

MST - Mitotic Somal Translocation

MZ - Marginal Zone

NADH - Nicotinamide Adenine Dinucleotide

NAHR - Non-Allelic Homologous Recombination

NEC - Neural Epithelial Cells

NFR - Nuclear Fast Red

NGS - Next Generation Sequencing

NHEJ - Non-Homologous End Joining

NPCs - Neural Progenitor Cells

OCT - Optimal Cutting Temperature Compound

OPCs - Oligodendrocyte Precursor Cells

oRGC - Outer Radial Glial Cell

oSVZ - Outer Subventricular Zone

PBS - Phosphobuffered Saline

PCR - Polymerase Chain Reaction

PCW - Post Conceptual Weeks

PFA - Paraformaldehyde

pH3 - Phospho-Histone Three

PP - Preplate

RGC - Radial Glial Cell

RIE - Royal Edinburgh Infirmary

RNA - Ribonuclear Acid

ROI - Region of Interest

RPKM - Reads Per Kilobase Of Transcript Per Million Mapped Reads

RT - Room Temperature

scRNA-seq - Single-Cell RNA Sequencing

SP - Subplate

SVZ - Subventricular Zone

TBE - Tris-Borate-EDTA

TBS - Tris Buffered Saline

tSNE - T-Distributed Stochastic Neighbour Embedding

UTR - Untranslated Region

VE - Ventricular Edge

VH - Ventral Hippocampus

vRGC - Ventral Radial Glial Cell

VZ - Ventricular Zone

μl - Microliter

μm - Micrometre

1 Introduction

1.1 The history of the human cerebral cortex

The earliest recorded reference to the brain from 1700BC was written in the Edwin Smith Surgical Papyrus describing the correlation between brain injury site with specific injury symptoms. During Roman times the physician Galen overturned the previous views of Aristotle that the heart was the epicentre of human intellect. Through careful dissections he deduced the brain to be the site of human cognition, memories and sensations (Gross, 2009). Galen's deductions of the brain remained unchallenged for nearly 1500 years and were not substantially expanded until the 1600s when Thomas Willis pioneered research of the brain and nervous systems (Feindel, 1962). Willis was the first to argue that it was the cerebral cortex that provides higher cognitive functions in humans through many careful comparative neuroanatomy dissections. With these he highlighted the enlarged cortical size and convoluted gyri of humans compared to other species and proposed the advanced human intellect was due to these differences (Molnár, 2004).

From there much understanding of the particular functions of individual cortical regions came observations of patients having incurred brain damage. Perhaps a most famous example is that of Phineas Gage, the railroad worker whose frontal lobe was damaged by a rod in 1848 which provided understanding of the how connections from the prefrontal cortex drive executive functions including behaviour and personality (Macmillan, 2001). Triggering movement of particular muscle sets through electrical stimulation coupled with lesions to specific cortical regions inducing defects to motor functions allowed researchers to build maps of the motor cortex (Fritsch and Hitzig, 1870; Ferrier, 1886). It is these, and many other, insights of researchers and anatomists throughout history, and constantly ongoing research, that has presented the cerebral cortex as a vastly complex brain region responsible for so much of our higher level cognitive functions including our consciousness, language, movement and ability to perceive and interpret the world around us.

The human cerebral cortex accounts for 40% of the brain's mass (Saladin, 2011). It is a 3-4mm thick stratified tissue structure containing around 15 billion neurons (Roberts, 1992; Saladin, 2011) divided into functional regions which can be identified and characterised by their regional specific cytoarchitecture. This variation of cellular composition, or cytoarchitecture, in different cortical regions was first noted in the 1860s by Theodor Meynert who developed histological staining of brain slices for visualisation.

This histological staining method allowed for the work of Korbinian Brodmann who comprehensively described the cytoarchitecture of the cerebral cortex (Brodmann, 1910). His work described 52 areas, identified by the arrangement of cells in each region, which were later linked to specific functions including the auditory and visual systems.

1.2 The structure of the cerebral cortex

The developed, adult cerebral cortex is a highly folded structure comprising two cerebral hemispheres separated by the medial longitudinal fissure. The cortex is organised into a 6-layed structure with each layer characterised by different neuronal populations and neuronal connections. All of the cortical layers contain neurons one of two types; the majority (70-80%) are excitatory glutamatergic neurons and the minority (20-30%) are GABAergic inhibitory interneurons (DeFelipe and Farinas, 1992). Correct balance of the numbers and connectivity of these two cell types is critical to correct cortical function.

1.3 Developmental terminology

For clarification, there are two different terms to describe the progression of human fetal development: post-conceptual weeks (pcw) and gestational weeks (GW). While the work described in this thesis uses pcw, the bioinformatics dataset used, and other descriptions in the literature, use GW. The difference between them is approximately two weeks: gestational weeks are counted from the date of the mother's last menstrual period and includes

about two weeks where she was not pregnant. Post-conceptual weeks (pcw) is the actual number of weeks since the conception and is more accurate and is determined by ultrasonography. It is important to consider the slight differences in these timescales throughout this thesis, particularly when we describe the dataset used for bioinformatics analysis.

1.4 Mice as models for the human cerebral cortex

In this thesis, and particularly the introduction, we are describing human fetal cortex development. However, much of our knowledge of human development is derived from work in non-human models, in particular mice. Perhaps the most physically obvious difference between mice and human cortices (besides the size), is the smooth brains of the lisencephalic mice compared to the folded cortices of gyrencephalic humans and other species including primates and ferrets. However simply increasing the size of the brain does not scale to increased volume or neuron number (reviewed in (Herculano-Houzel, Manger and Kaas, 2014) Cortical folding is a fascinating topic in itself (reviewed here (Llinares-Benadero and Borrell, 2019)), however in humans it does not begin until after 20pcw, an age past the scope of this thesis so any schematics of human brains are shown as predominantly smooth. The folded brain of humans allows for a considerable increase in neuron number in relation to brain size, and it is the presence of a second fibre layer, the outer sub-ventricular zone (oSVZ) which acts as a secondary proliferative hub in humans (and primates) taking over from the ventricular zone (VZ) later in development as the primary proliferative hub. It is thought that the presence of this secondary proliferative hub in the larger brained-humans and primates contributes in part to the massive increase of the neocortical size in these species (Lui, Hansen and Kriegstein, 2011). The germinative regions of the primate brain is divided into distinct compartments: The ventricular zone (VZ) and the sub-ventricular zone (SVZ). In primates the SVZ is divided into two parts – the inner sub-

ventricular zone (iSVZ) and the outer sub-ventricular zone (oSVZ) with these two layers separated by an inner fibre layer (IFL) (Smart *et al.*, 2002). No oSVZ has been described in rodents (Smart *et al.*, 2002).

The differences in cortical structure between mice and humans are summarised in Figure 1 and the rest of this thesis will refer to human brain development.

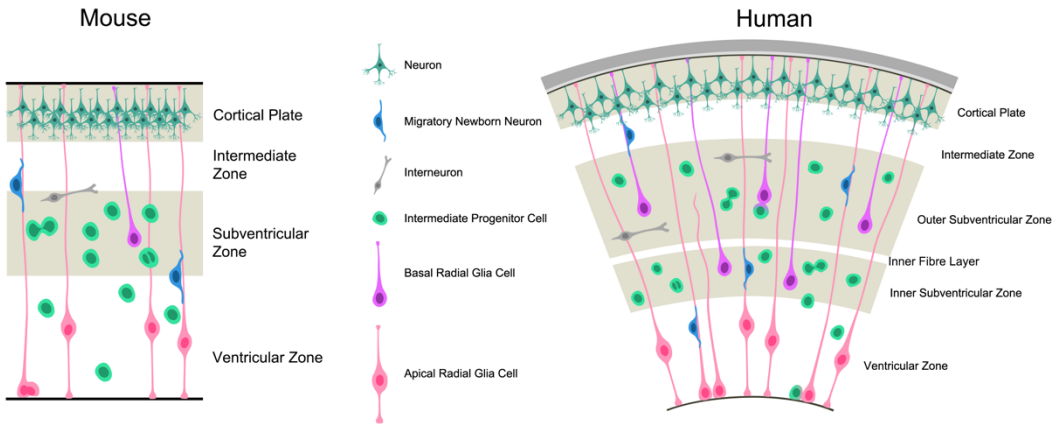


Figure 1: Comparison of mouse and human cortical development

A comparison of mouse and human cortex development. The primary difference is the outer subventricular zone in humans which acts as a secondary proliferative hub contributing to a large increase in neuron number.

1.5 Forebrain development

Gastrulation begins around 3 weeks after fertilisation (3pcw) in humans and is the very beginning of the development of the nervous system. Here the embryo is organised into three germ layers: the outer ectoderm layer, the middle mesoderm layer and the inner endoderm layer. It is the ectoderm on the outside that forms the neural plate, the precursor to the nervous system. This neural plate is a flat sheet of neuroepithelial cells that folds inwards dorsally to form the neural tube, the process is termed neurulation and occurs during 3-4pcw. Following neural tube closure, distinct swellings appear in the rostral part of the tube. These bulges result in formation of three primary vesicles.

1.5.1 Embryonic forebrain structures

Initially there are three primary vesicles which form the embryonic brain: the prosencephalon (forebrain), mesencephalon (midbrain) and the rhombencephalon (hindbrain) (Vaage, 1969; Martinez and Puelles, 2000; Price *et al.*, 2011) The prosencephalon then subdivides into the telencephalon which will go on to form the cerebrum, and the more caudal diencephalon (Figure 2).

The dorsal part of the telencephalon (the pallium) will eventually form the cerebral cortex, the region of study for this thesis, while the ventral part (the subpallium) will form the basal ganglia (Sur and Rubenstein, 2005).

The vast majority of the developing cortex comprises the 6 layered neocortex, however there are also the phylogenetically older allocortex areas: the paleocortex and the archicortex which includes the hippocampal cortex (N Filimonoff, 1947; Goulas *et al.*, 2019). These older areas typically

comprise of three or four layers in adults. This thesis will focus on development of the 6-layered neocortex.

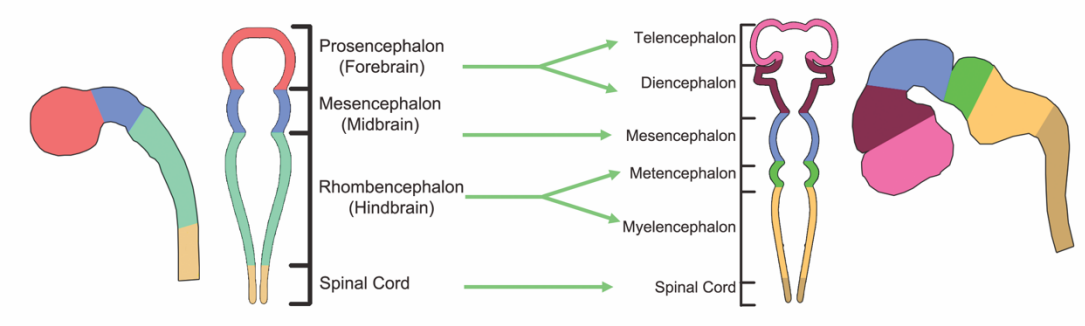


Figure 2: Formation of the three primary vesicles

The three primary vesicles: The prosencephalon (forebrain) which goes on to form the telencephalon and diencephalon. The mesencephalon (midbrain). The rhombencephalon (hindbrain) which forms the metencephalon and the myelencephalon. There is also shown on this figure the spinal cord.

1.6 Development of the human cortex

Development of the human neocortex is a vastly complex procedure that requires precise temporal and spatial regulation to ensure correct cortical size and architecture. The process is summarised in Figure 3 (based on diagram from (Budday, Steinmann and Kuhl, 2015)) and is described in detail throughout this section.

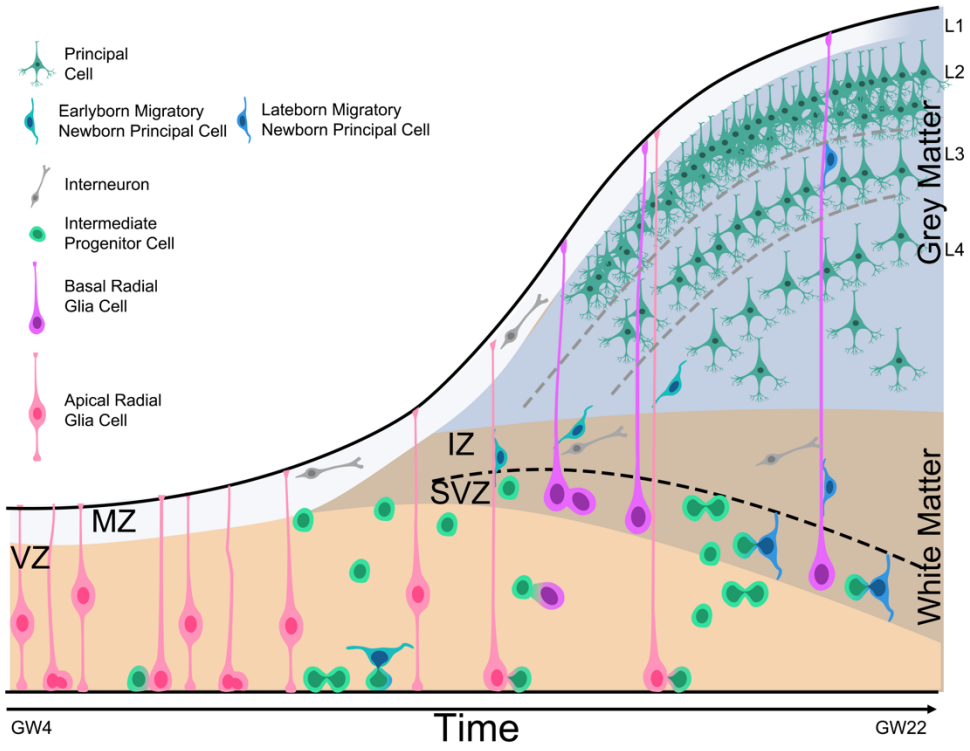


Figure 3: Process of human cortex development over time

Schematic showing the process of cortical development until the end of neurogenesis. VZ = ventricular zone, MZ = marginal zone, SVZ = subventricular zone, IZ = intermediate zone. GW = gestational weeks. The cortex forms in an inside out manner with the earlier born neurons (teal green) migrating to the innermost layers while the late born neurons (dark blue) migrate to the outer layers.

1.6.1 Cortical development occurs in waves

Cortical development (corticogenesis) can be divided into three main stages as visualised in Figure 4 which will be described in greater detail in this chapter. These are: **proliferation**, characterised by rapid expansion of cell numbers, **neurogenesis** and **gliogenesis**. Ensuring the cells switch to undertake different types of cell division is critical to ensuring both the correct number of neurons at the end of development and correct proportions of neuronal types for circuitry.

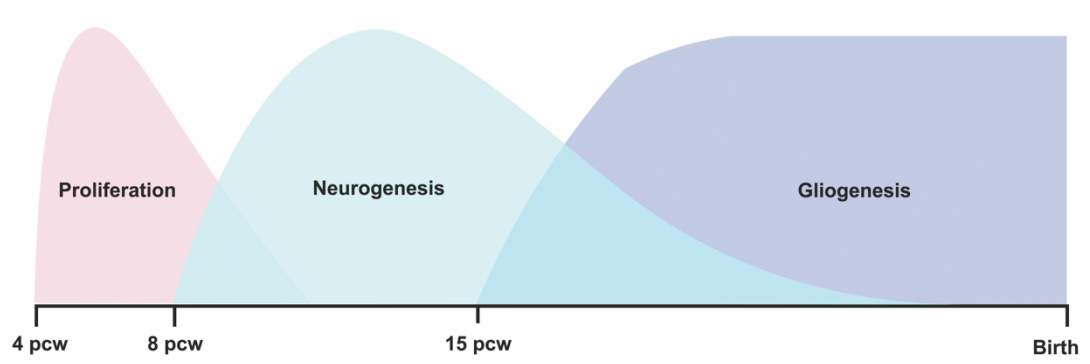


Figure 4: Stages of human cortical development

Schematic showing human cortical development divided into the main stages. Graph is not to scale, the numbers of cells generated at each stage is not proportional to the graph. Y axis indicates which stage the majority of cells are in.

1.6.2 Germinative zones in human cortex development

The complete cortex is made up of a variety of distinct cell types, which can be sorted into two basic classes of excitatory neurons and inhibitory interneurons. The progenitor cells which create these separate populations are located in spatially specific germinal zones within the developing cortex (Goulas *et al.*, 2019). A schematic of the developing cortex (based on a 12pcw brain) is shown in Figure 5 with the germinative zones shown in green. The excitatory glutamatergic neurons are born in the cortex itself, in the VZ and SVZ, and migrate out radially (indicated by pink arrow) to their appropriate layer in an inside-out manner (Rakic, 1974; Bystron, Blakemore and Rakic, 2008). Their counterparts, the inhibitory interneurons, are born in the sub-pallial ganglionic eminences (GE) and they tangentially migrate up to the neocortex through the IZ and CP to their final position and integration into circuitry (dark blue arrows) (Porteus *et al.*, 1994; Bystron, Blakemore and Rakic, 2008). In this thesis we are focusing on the germinative zones of the cortex: the VZ and SVZ.

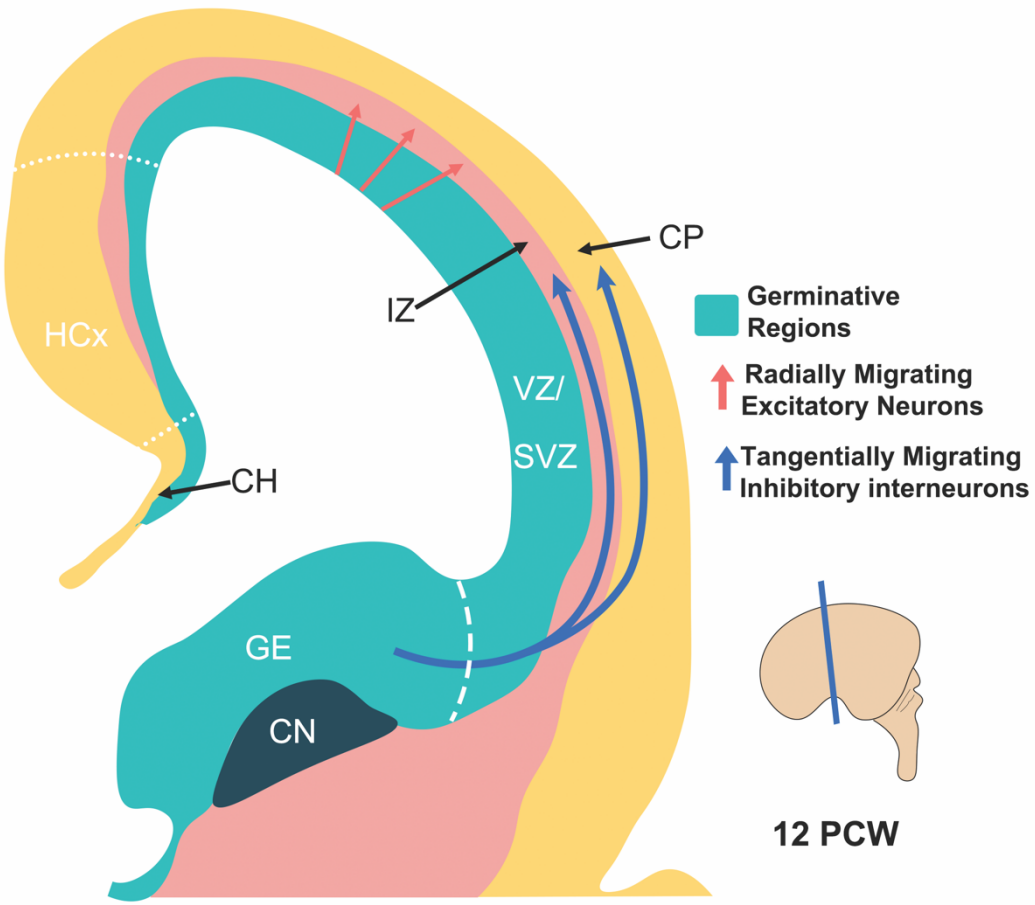


Figure 5: Schematic of a coronal section showing the different germinative regions of the 12pcw human fetal cortex

The VZ and SVZ (green) are the primary birth site of excitatory neurons which migrate radially out through the intermediate zone (IZ) to their final destination of the cortical plate (CP) where they form the six cortical layers in an inside out manner.

Inhibitory interneurons are mostly born in the ganglionic eminences (GE) of the subpallium of the brain where they tangentially migrate through the IZ and CP to their final position in the cortex where they integrate into cortical circuitry.

Other structures shown are the cortical hem (CH), the hippocampal cortex (HCx) which comprise the medial pallium and the caudate nucleus (CN).

1.6.3 Initiation of corticogenesis in pre-natal brain development

Following closure of the neural tube, the first wave of corticogenesis, the proliferative wave, is initiated at 4-6pcw by the division of highly polarized, actively mitotic neuroepithelial cells (NECs). These symmetric divisions occur at the ventricular margin and are characterised by the process of interkinetic nuclear migration (IKNM). In this process the nuclei of the cells “bounce” up and down through the cytoplasm during the cell cycle: they move from the basal surface, where they undergo DNA synthesis during S phase, down to the apical ventricular margin where they divide producing two identical daughter cells and they migrate back up towards the basal abventricular surface to repeat the process (Figure 6) (Sauer and Walker, 1959; Bystron, Blakemore and Rakic, 2008).

These NECs are the earliest progenitors in the cortex and the origin of all neocortical neurons. The NECs undergo proliferative symmetric divisions producing proliferative daughter cells thus rapidly increasing the committed neural precursor population and resulting in rapid increase of cortical thickness (Rakic, 1995).

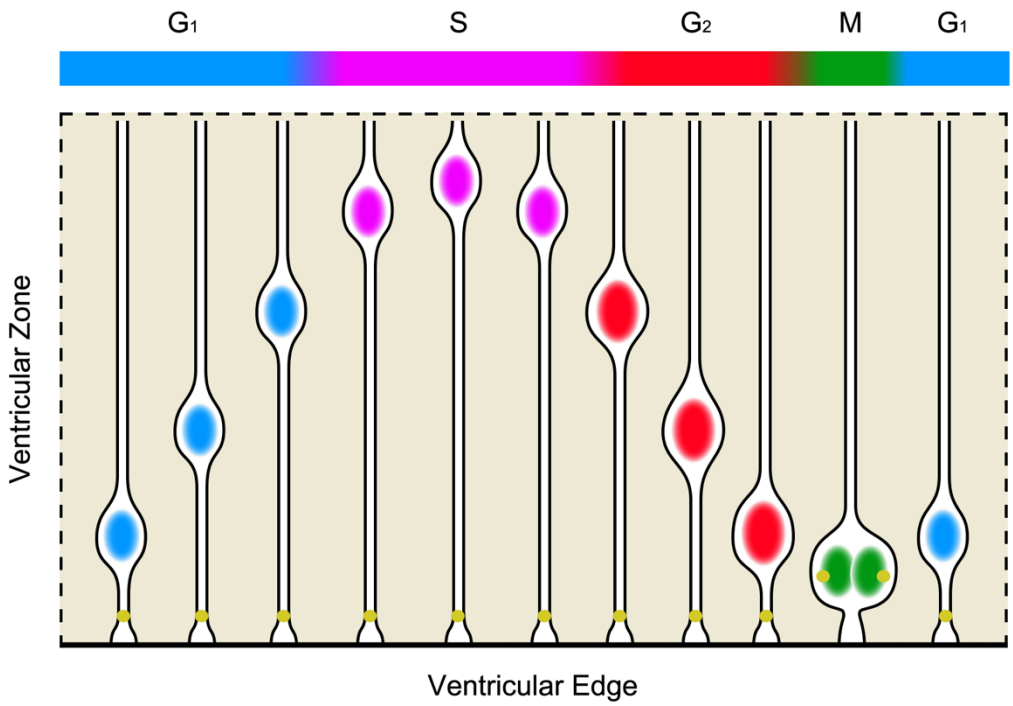


Figure 6: Interkinetic nuclear migration

Schematic showing interkinetic nuclear migration (IKNM) of an aRGC during cortical development. As the cells go through the cell cycle their nucleus migrates within the VZ allowing for a much greater number of cells to be present and creating a pseuo-stratified layer. The cell nucleus moves away from the ventricular edge as it goes through G1 phase transitioning to S phase as it reaches the basal surface of the VZ. Entering G2 phase it migrates back down towards the ventricular edge where it undergoes mitosis producing two daughter cells. In aRGCs the centrosome (yellow circle) is located near the ventricular edge and the nucleus moves towards it for cells to undergo mitosis.

1.6.4 Onset of cortical neurogenesis

Next is the continuation of proliferative corticogenesis and the onset of the neurogenic wave of cortical development. This begins when the proliferative NECs switch from a symmetric to asymmetric division and the establishment of the radial glial system. The process of neurogenesis is highly complex and described in some detail below, the compartments and cell types in the cortex are shown in Figure 7.

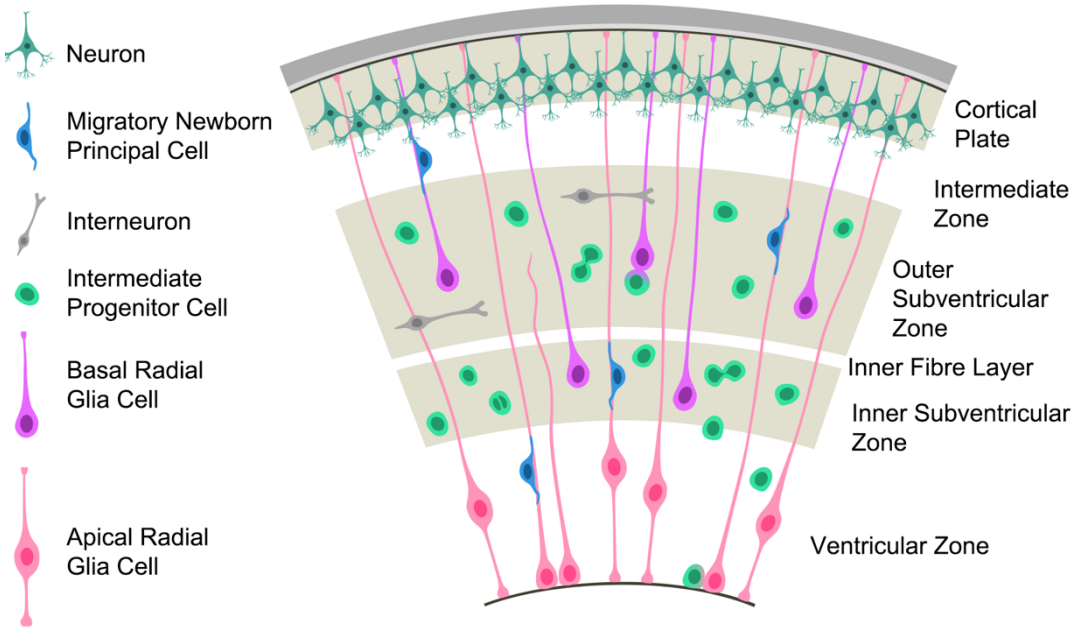


Figure 7: Schematic of the human cortex during neurogenesis

Schematic showing the different compartments of the human cerebral cortex during the neurogenesis stage of development.

The ventricular zone harbours proliferative apical radial glia cells and their daughter intermediate progenitor cells. The subventricular zone contains basal radial glia cells which also produce intermediate progenitor cells. These proliferating cells will also produce neurons which migrate outwards along the radial glia fibres to their final position in the cortical plate.

1.6.4.1 Establishment of the radial glial system

The NECs gradually express the transcription factor *PAX6* and adopt an apical radial glial cell (aRGC) identity (Walther and Gruss, 1991; Stoykova and Gruss, 1994; Kriegstein and Alvarez-Buylla, 2009; Manuel *et al.*, 2015). These aRGCs exhibit an elongated morphology and span the cortical wall with their basal process extending to the pial surface and their shorter apical process connecting to the ventricular edge. Similar to their NEC precursors, these aRGCs also exhibit apical to basal polarity and undergo IKNM and divide at the ventricular edge (Götz and Huttner, 2005). Unlike NECs however, the aRGCs do not migrate the entire length of their fibres which span the width of the cortex, only through the ventricular zone.

1.6.4.2 Key functions of aRGCs during corticogenesis

aRGCs have several key functions during cortical development, the two most relevant are described:

1) Through their asymmetric divisions producing one identical daughter cell and one fate-restricted progenitor or neuron they act to contribute to both generation of neurons, ensuring correct final neuron number and ensuring maintenance of a sufficient progenitor pool to ensure correct cortical size.
2) Their basal processes span the cortex to the pial surface and provide a scaffold for cells such as neurons to migrate to the outer layers of the cortex (Rakic, 1988).

1.6.5 Progenitor division, diversity and differentiation

The proliferative zones can be considered the factories of the cortex, responsible for production of cellular number and variety and are shown in Figure 5. In humans the first proliferative zone, the ventricular zone (VZ), emerges with the onset of neurogenesis and cells undergo IKNM through it (see Figure 6). The cell bodies of aRGC cells are confined to the VZ with their centrosomes located in the apical region of their fibre (Chenn *et al.*, 1998). As neurogenesis progresses, dividing cells from the VZ congregate at the basal edge of the VZ thus forming the second proliferative hub of the cortex, the subventricular zone (SVZ) (Smart, 1973; Smart *et al.*, 2002; Bystron, Blakemore and Rakic, 2008). As neurogenesis progresses, more progenitors switch from the proliferative to neurogenic divisions as shown in Figure 8 with the timescales for human and mouse described in the legend (Picco *et al.*, 2018). Maintenance of this balance of proliferation and differentiation is essential for correct cortical formation.

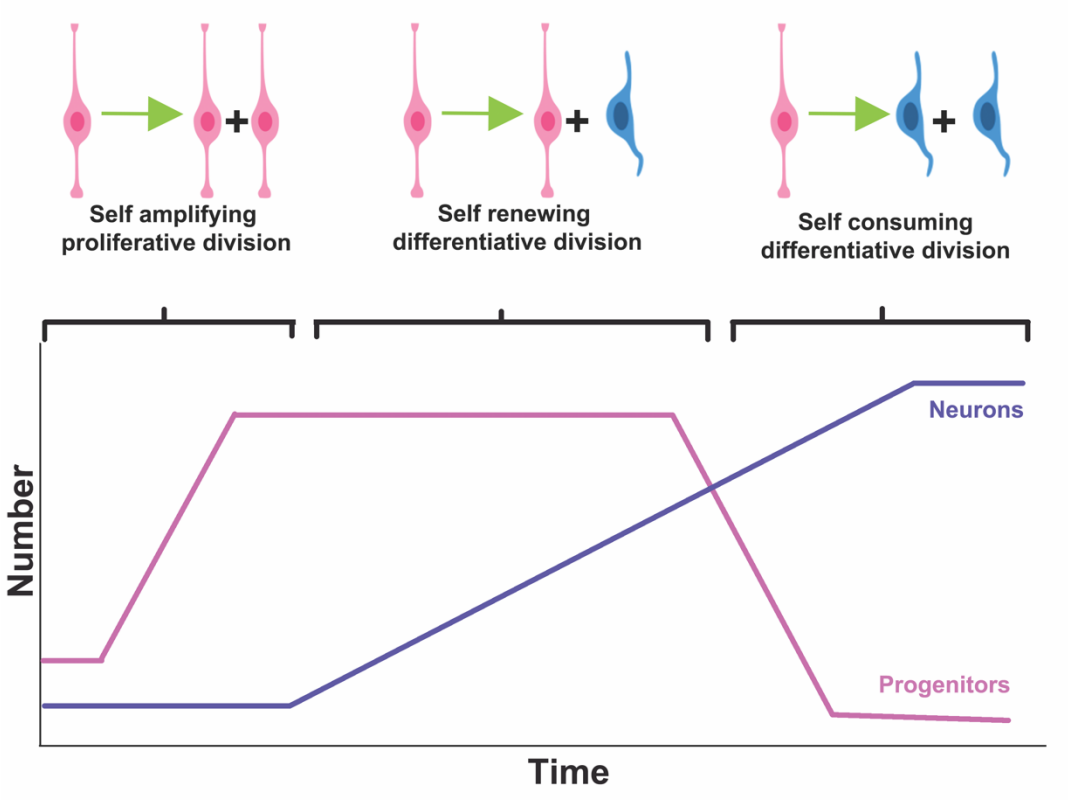


Figure 8: The balance of proliferation and differentiation

As neurogenesis proceeds the types of division cells undergo changes accordingly, initially there is huge amounts of proliferation to build cortical numbers followed by generation of neurons to increase numbers of neurons in the final complete cortex. To give some scale to the time axis; in humans, neurogenesis starts around E40 and continues for approximately 80 days while in mouse it occurs from around E11 until E19.

1.6.5.1 Types of progenitor division

There are a variety of types of division progenitor cells can undergo during cortex development. Cells undergoing correct division at the appropriate time is essential for correct cortical formation. The division types are as follows (Götz and Huttner, 2005; Florio and Huttner, 2014) and summarised in Figure 9 (based on figure from (Florio and Huttner, 2014)):

Symmetric divisions:

1) **Symmetric proliferative** – cell produces two identical proliferative daughter cells, the same type as the parent cell. This is a self-amplifying division.
2) **Symmetric differentiative** – cell produces two identical proliferative daughter cells, a different type to the parent cell. This is a self-consuming division
3) **Symmetric neurogenic** – cell produces two identical post-mitotic daughter cells. This is a self-consuming division.

Asymmetric divisions:

1) **Asymmetric bi-differentiative** – cell produces two non-identical proliferative daughter cells. This is a self-consuming division.
2) **Differentiative/neurogenic** – cell produces one proliferative cell (different to the parent cell) and one neuronal cell. This is a self-consuming division.
3) **Differentiative** – cell produces two proliferating daughter cells, one of the same type as the parent cell and the other different. This is a self-renewing division.

4) **Neurogenic** – cell produces one proliferating daughter cell of the same type as the parent cell and one neuronal cell. This is a self-renewing division.

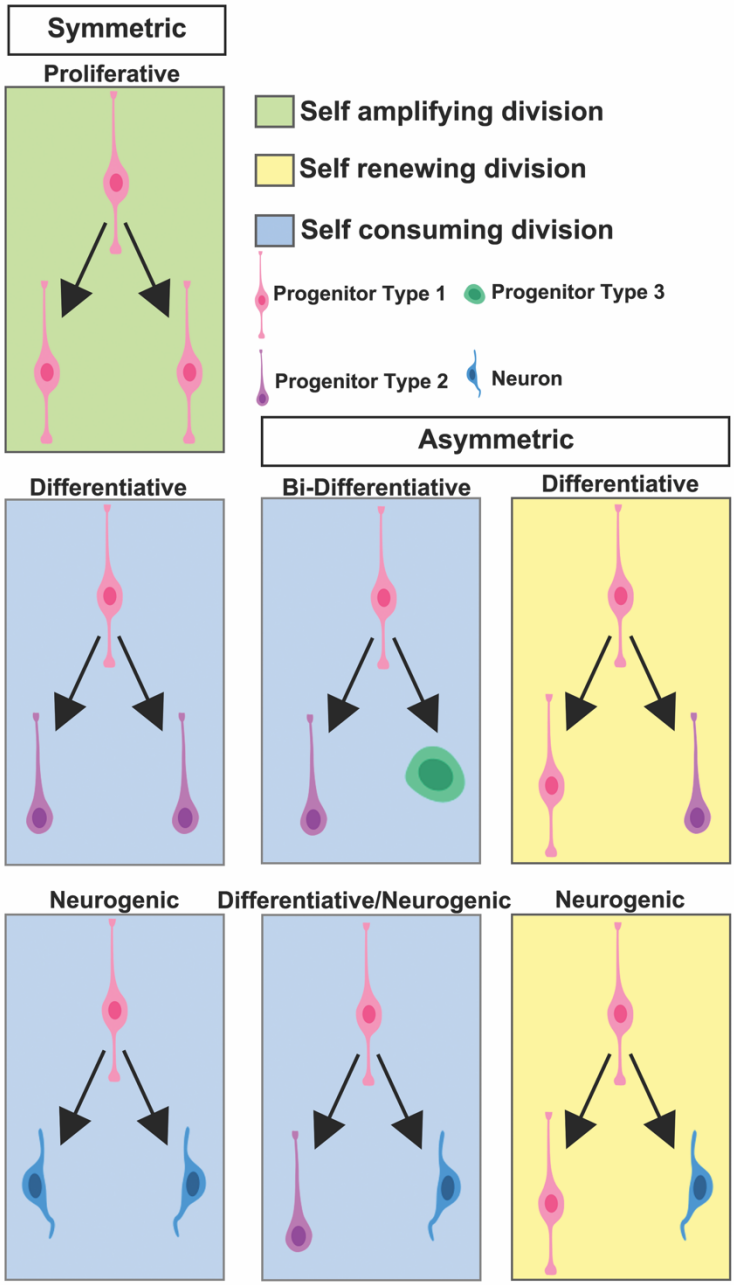


Figure 9: Different types of progenitor divisions in the developing cortex

Schematic showing the different types of progenitor divisions; however, it should be noted that not all types of cell produced by each kind of division are shown, only an example. Based on figure from (Florio and Huttner, 2014)

1.6.5.2 Different progenitor subtypes

After establishment of the radial glial system and formation of the proliferative VZ and SVZ, progenitor cells in the developing human cerebral cortex can be divided into two main classes; apical progenitors and basal progenitors.

1.6.5.2.1 Apical progenitors

The first apical progenitor cell class to appear is the described above **neuroepithelial cell (NEC)** which undergoes predominantly proliferative symmetric division.

Next is the emergence of **apical radial glial cells (aRGCs)** (also called ventral radial glial cells (vRGCs), the cell bodies of which reside in the ventricular zone where they undergo IKNM during G1 and G2 phases of their cell cycle with mitosis occurring at the apical surface (Figure 6) (Götz and Huttner, 2005). These aRGCs express the transcription factor *PAX6* and astroglial markers such as *GFAP* (Levitt and Rakic, 1980; Bystron, Blakemore and Rakic, 2008). As shown in Figure 9, aRGCs have a huge capacity for many different types of division however their proliferation and differentiation must be tightly controlled to produce the required number and subtypes of neurons.

One daughter type produced by aRGC division is **apical intermediate progenitors (aIPs)**. These cells are similar to their aRGC parent cells in that they are located in the ventricular zone, express *PAX6* and undergo IKNM to divide at the apical surface (Gal *et al.*, 2006; Tyler and Haydar, 2013). In other aspects they differ from their parent cell: most strikingly that they lose self-renewing potential and can only perform a self-consuming, symmetric neurogenic division producing a pair of neurons (Gal *et al.*, 2006; Tyler and Haydar, 2013; Florio and Huttner, 2014).

1.6.5.2.2 Basal progenitors

Basal radial glia cells (bRGCs), also known as **outer radial glia cells (oRGCs)** are born from differentiative divisions of their mother aRGC from which they inherit their basal process. Located in the SVZ they lack the apical connection of aRGCs and undergo asymmetric self-renewing divisions ensuring a continuous pool of undifferentiated bRGCs as well as sufficient neuron production (Hansen *et al.*, 2010; LaMonica *et al.*, 2012).

In contrast to aRGCs, which divide by undergoing IKNM, bRGCs divide by the process of mitotic somal translocation (MST). During MST, in the hour before cytokinesis the cell soma rapidly ascends its radial fibre towards the pial surface (Hansen *et al.*, 2010; Wang *et al.*, 2011). Conversely to aRGCs, the centrosome in bRGCs is located in the fibre on the pial side of the nucleus hence the movement in this direction during MST (Ostrem *et al.*, 2014).

Basal intermediate progenitors (bIPs) are born from either bRGCs or aRGC divisions, they migrate to the SVZ losing their apical and basal processes and begin expressing *TBR2* (Miyata *et al.*, 2004; Noctor *et al.*, 2004; Englund *et al.*, 2005). These bIPs can undergo two different types of division.

1) The first type is similar to that of aIPCs, undergoing a self-consuming, symmetric neurogenic division to produce two neurons (Miyata *et al.*, 2004; Noctor *et al.*, 2004).
2) The second type increases the number of bIPs, they undergo multiple symmetric proliferative divisions producing two daughter bIPs before undergoing a final self-consuming, symmetric neurogenic division (Noctor *et al.*, 2004).

1.6.6 The cell cycle in cerebral cortex development

1.6.6.1 The cell cycle

The cell cycle is the process by which a cell replicates its DNA and divides its organelles and cytoplasm to produce two daughter cells. It comprises four distinct phases: G1 phase, S phase, G2 phase and M phase (mitosis) (in addition there is G0 phase which is cell cycle exit) and is shown in Figure 10.

G1 Phase is a period of immense cellular growth and G1 phase has a highly variable duration. During G1 the cell grows in size and increases organelle number.
S Phase is the stage of DNA replication, during this phase all chromosomal DNA is replicated.
G2 Phase is the second growth phase during which there is rapid cell growth and protein synthesis.
M Phase comprises chromosome separation and cytokinesis. M phrase is a highly regulated, complex sequence of events divided into six phases each with key events:

- Prophase: chromatin condenses, and the nucleolus disappears
- Prometaphase: the nuclear membrane breaks down; the chromosomes form kinetochores and kinetochore microtubules attach to the kinetochores.
- Metaphase: The chromosomes line up on the metaphase plate of the cell.
- Anaphase: Chromosomes are split at the centromeres and move to opposite poles of the cell.
- Telophase: Nuclear envelope re-assembles around each set of chromatids and the chromosomes unfold to chromatin.
- Cytokinesis: The cytoplasm of the cell cleaves in two, separating the cell into two daughter cells.

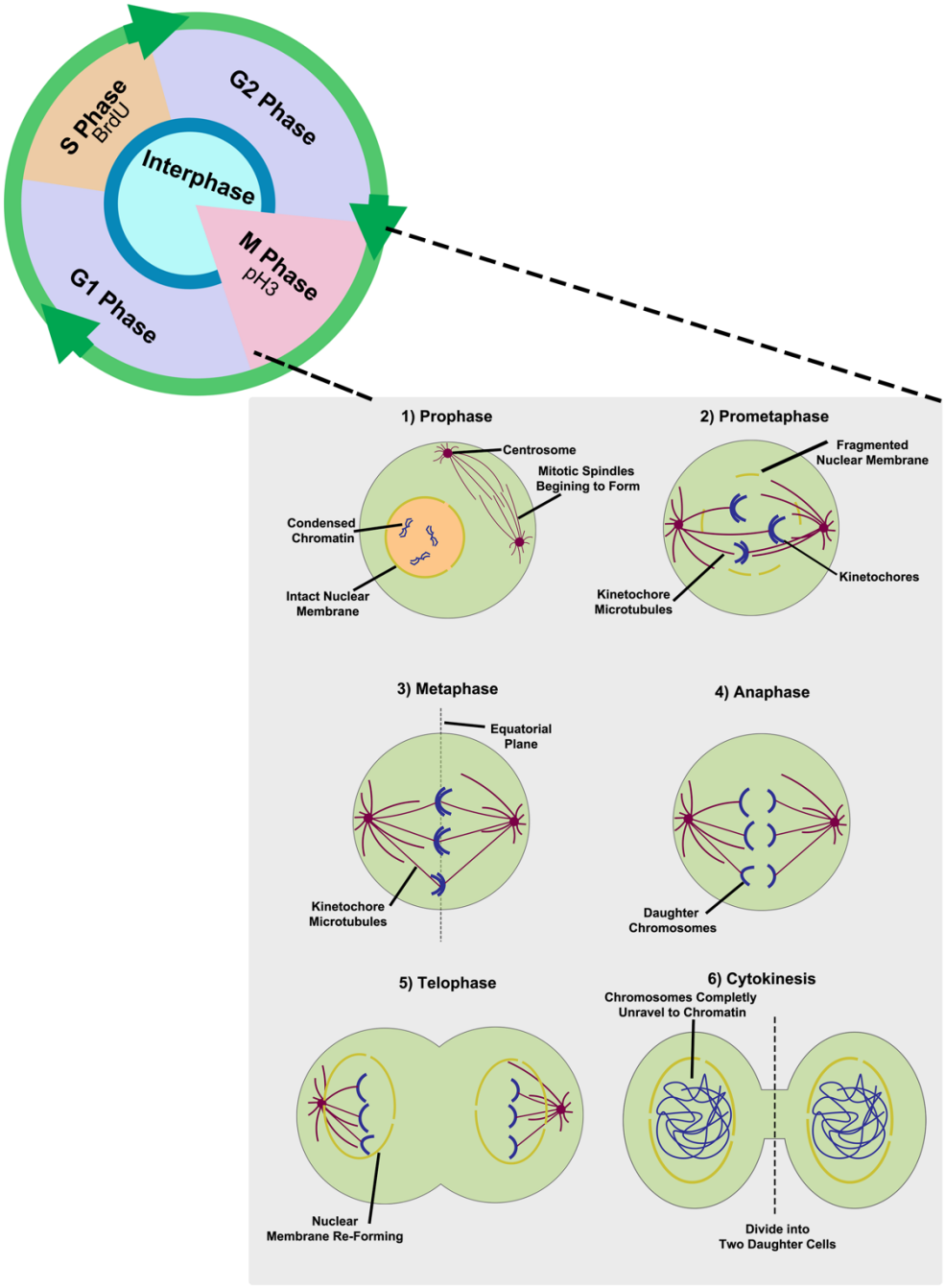


Figure 10: The cell cycle and M phase

Interphase of the cell cycle comprises G1, S and G2 phase before entering M phase where cells divide to produce two daughter cells. S phase can be marked with BrdU – a thymidine analogue taken up by cells in S phase and then stained with an antibody while M phase cells can be identified with a phospho-histone 3 (pH3) antibody.

1.6.6.2 The cell cycle in corticogenesis

To ensure the precise temporal and spatial control required to produce the complete cortex, strict regulation of progenitor division in the proliferative regions of the cortex is essential. Major controlling factors of neuron number and final brain size is cell cycle length, exit from the cell cycle and differentiation to neurons. Any changes to this carefully orchestrated balance of proliferation and differentiation can have many cascading effects leading to a wide range of neurodevelopmental disorders.

1.6.6.2.1 Cell cycle length and neural progenitor fate

As neuroepithelial cells transition to radial glial cells and these radial glial cells switch from neurogenic to proliferative divisions with neural development progression, there in an increase in the cell cycle length of these cells from (in mice) 10.2 hours (E12) to 18 hours (E16) (Takahashi, Nowakowski and Caviness, 1995). As neurogenesis progresses there are increasing proportion of progenitor cells undergoing neurogenic divisions correlating with increasing cell cycle length (Takahashi, Nowakowski and Caviness, 1995). Strikingly this is predominantly due to increased G1 length (Takahashi, Nowakowski and Caviness, 1995) leading to the “cell cycle length hypothesis” proposed initially by Calegari and Huttner from studies in mice cortex development (Calegari and Huttner, 2003). This hypothesis proposes that lengthening the time of G1 phase is sufficient to induce the switch between proliferative and neurogenic fates.

In this model of how cell cycle length determines progenitor division type, summarised in Figure 11, there is a threshold of exposure to a cell fate determinant (e.g. a differentiation factor) that cells must reach to make a neurogenic vs differentiative division. The exposure to this factor can be increased by increasing the cell cycle length, thus increasing the chance the threshold will be reached. If it is not reached (red cell) then the cell will

undergo a proliferative division, if it is reached or surpassed (green cell) then the cell will undergo a neurogenic division.

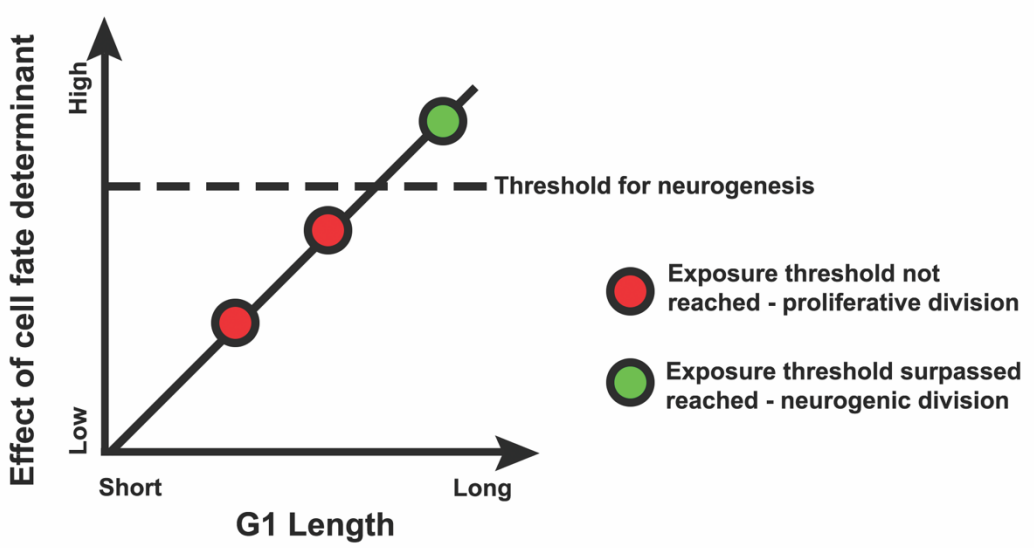


Figure 11: Summary of the cell cycle length hypothesis

The cell cycle length hypothesis states that cell fate is dependent on the cell being exposed to a certain amount of a differentiative factor. The longer the cell cycle, specifically G1 phase, the more exposure the cell has and the greater the chance of it reaching this threshold. If G1 phase is relatively short, it does not reach the threshold and undergoes a proliferative division. However, as neurogenesis progresses and the cell cycle and therefore G1 length increases, the cell is exposed to the factor for longer and eventually it will surpass the threshold and take the neurogenic fate.

This was evidenced by knockdown of known components of the G1 checkpoint mechanisms in mice, cdk4/cyclinD1, to artificially induce lengthening of the G1 phase which was sufficient to induce premature switch of progenitors to neurogenic divisions (Calegari and Huttner, 2003). In addition, in 2009, it was shown that shortening G1 produced the opposite effect; inhibiting neurogenesis and increasing the progenitor pool resulting in a larger brain size (Lange, Huttner and Calegari, 2009). This fits with the model (Figure 11) which proposed that reduced G1 length reduces exposure to factors required for the neurogenic switch and thus the exposure threshold for a neurogenic division is not met.

In summary, the decision of neural progenitor cells to remain proliferating or to exit the cell cycle to form neurons is tightly controlled by a range of factors, including the cell-cycle length and the factors which govern this. These factors act to govern the number of cell divisions and final cell number and type. Any disruption to this process will have dramatic effects on final cortex size, structure and function.

1.6.7 Neurogenesis and the formation of cortical layers

The layers of the human cerebral cortex form in an inside-out manner, with early born neurons from the germinative zones of the cerebral cortex migrating radially out to form the innermost layer six of the cortex and subsequently born neurons migrating past them to form the outer layers. This is illustrated in Figure 3.

These first born Cajal-Retzius cells (appearing around 8pcw in humans) migrate tangentially from their birth-site in the cortical hem (Takiguchi-Hayashi *et al.*, 2004; Bystron, Blakemore and Rakic, 2008) to form the transient **preplate (PP)** at the pial surface of the cortex (Marin-Padilla, 1983; Luskin and Shatz, 1985; Ayala, Shu and Tsai, 2007). With this first wave of

neuronal migration the **Intermediate Zone (IZ)** emerges; a sparse area comprising mainly of axon tracts with few cell bodies, separating the germinative VZ and SVZ from the post-mitotic neurons (Ayala, Shu and Tsai, 2007; Bystron, Blakemore and Rakic, 2008; Hoerder-Suabedissen and Molnár, 2015)

Subsequent waves of neurons migrating from the germinative zones split this preplate into the outer **Marginal Zone (MZ)** which harbours those first born Cajal-Retzius cells which will eventually become layer I, and the transient **Subplate (SP)** (Ayala, Shu and Tsai, 2007; Bystron, Blakemore and Rakic, 2008; Hoerder-Suabedissen and Molnár, 2015). As waves of neuronal migration continue these neurons fill the space between the MZ and SP to form the cortical plate.

1.6.7.1 Migration

As described above, neuronal migration resulting in neurons arriving in their appropriate position is essential for correct cortical formation and occurs in successive waves. In the developing cortex there are two main forms of neuronal migration: radial and tangential as shown in Figure 5.

1.6.7.1.1 Radial migration

During radial migration, excitatory neurons born in the cortex migrate out from their germinal zones (Noctor *et al.*, 2004; Ayala, Shu and Tsai, 2007) to their final position in the cortex. This process accounts for the vast majority of neurons in the cortex (80-90%) (Ayala, Shu and Tsai, 2007). Initially neurons are able to migrate by somal translocation, in which they extend a long process to the pial surface and pull themselves up through the cortex to the appropriate position (Nadarajah *et al.*, 2001). As neurogenesis progresses and perhaps as the distance required for travel increases with the thickening cortex, neurons rely more on locomotion for transport, in which they attach to

the apical-basal spanning processes of the radial glial cells still in the germinative zones and migrate up them to their final position (see Figure 3) (Rakic, 1972, 1990; Marín and Rubenstein, 2001; Nadarajah *et al.*, 2001; Ayala, Shu and Tsai, 2007).

1.6.7.1.2 Tangential migration

The second, and much smaller in number, migratory stream of cortical neurons is the tangential migration of inhibitory interneurons. Born in the ganglionic eminences they migrate tangentially up through the CP and IZ (see Figure 5) where they then migrate radially to their final position in the cortex for integration into cortical circuitry. It is thought that any disruption to migration will affect the inhibitory-excitatory balance of neuronal circuitry causing a range of neurodevelopmental disorders such as Autism Spectrum Disorder (ASD). However in human cortex development the birth-place of interneurons is thought to be more complex than the mouse and is an area of conflicting research opinions, with some studies suggesting that some cortical interneurons may arise from the dorsal, rather than ventral, telencephalon (Zecevic, Hu and Jakovcevski, 2011). One 2015 study identified evidence in favour of this theory - cells expressing Nkx2.1 were found in the germinative zones of the human and monkey fetal cortex (Radonjić *et al.*, 2014). However several additional studies have been unable to corroborate these findings (Hansen *et al.*, 2013; Ma *et al.*, 2013). Thus, this still remains a topic of controversy and will likely require detailed long-term slice culture experiments to fully resolve.

At this point of development, the neurogenesis wave begins to decrease as the rate of the gliogenic wave increases.

1.6.7.2 Gliogenesis

Glia cells comprise at least 50% of the human brain (Rowitch and Kriegstein, 2010), and act as a non-neuronal support system for the brain. They have many diverse roles and are essential for ensuring correct neurological function through maintaining homeostasis, neuronal protection and support and formation of myelin (Jessen and Mirsky, 1980; Kessaris, Pringle and Richardson, 2001; Rowitch and Kriegstein, 2010). The incorporation of these glial cells by the process of gliogenesis is essential for the neuronal maturation of the cerebral cortex.

The process of gliogenesis in human cerebral cortex can be divided into early, intermediate and late gliogenesis. **Early gliogenesis**, beginning with early neurogenesis, features the establishment of the radial glial system. During **intermediate gliogenesis**, oligodendrocytes and fibrous astrocytes are generated and incorporated into the expanding white matter. Lastly, **late gliogenesis** in which first lamina astrocytes and grey matter protoplasmic astrocytes are generated (Marín-Padilla, 2011).

The switch of progenitor cells in the cortex from producing neurons to producing glial cells must be tightly controlled to ensure gliogenesis occurs at correct times in relation to neurogenesis. In early neurogenesis, NECs do not respond to cytokines, and the promotors of the astrocytic genes, *GFAP* and S100B are methylated causing repression of astrocyte development. When neurogenesis commences with emergence of RGCs, these RGCs become capable of responding to the cytokines as the astrocytic gene promotors are demethylating, giving them the potential to generate astrocytes earlier than required. To prevent early generation of astrocytes, extrinsic signals such as FGF2 repress gliogenesis ensuring that RGCs keep producing neurons. Gliogenesis is also repressed by Neuregulin 1 which activates the nuclear signalling pathway which involves *ERBB2*, the pathway which also

antagonizes astrogenesis-promoting signals. When neurogenesis ends, *ERBB2* expression increases coinciding with the transformation of RGCs to astrocytes (Molne *et al.*, 2000; Takizawa *et al.*, 2001; Temple, 2001; Hermanson, Jepsen and Rosenfeld, 2002; Fox and Kornblum, 2005; Sardi *et al.*, 2006; Britsch, 2007; Sanosaka *et al.*, 2008; Rowitch and Kriegstein, 2010; Marín-Padilla, 2011).

This is a description of one of the possible mechanisms by which the neuron to glial switch is controlled, others include bHLH factors, *NOTCH* signalling, cytokines secreted by neurons, BMPS and *OLIG2* (Mabie, Mehler and Kessler, 1999; Mehler, 2002; Barnabe-Heider *et al.*, 2005; Fox and Kornblum, 2005; Marshall, Novitch and Goldman, 2005; Sanosaka *et al.*, 2008; Chenn, 2009; Hirabayashi *et al.*, 2009; Namihira *et al.*, 2009; Rowitch and Kriegstein, 2010; Wilkinson, Dennis and Schuurmans, 2013). This variety of mechanisms for regulating gliogenesis allow for transition of these RGCs into a variety of glial cell types. The two principal types of neuroepithelium-derived glial cells are oligodendrocytes and astrocytes.

1.6.7.2.1 Oligodendrogenesis

The key role of oligodendrocytes is to create the myelin sheath, essential for axonal insulation and conduction of action potentials (Barres and Barde, 2000; de Castro and Bribian, 2005). Oligodendrocytes arise during neural development, starting at 10pcw and continuing until adulthood. During development oligodendrocytes are born in discreate anatomical niches which causes them to adopt specific migration patterns resulting in them migrating to highly regionally specific parts of the brain (de Castro and Bribian, 2005).

During development, oligodendrocyte progenitors become oligodendrocyte precursor cells (OPCs) which transform into pre-oligodendrocytes, all three of these cell types within the oligodendrocyte lineage have the capacity for self-renewal. This self-renewing capacity is lost as pre-oligodendrocyte cells become immature oligodendrocytes before reaching their final adult form as mature oligodendrocytes (de Castro and Bribian, 2005; Rowitch and Kriegstein, 2010). The process of axonal myelination by oligodendrocytes continues long after birth, extending well into adulthood before it is completed.

1.6.7.2.2 Astrogenesis

Astrocytes are the stars of the glial system: they have a variety of roles in the nervous system including maintaining extracellular ion balance and water balance, structural support and modulating synaptic transmission to highlight a few and their dysfunction has been implicated in Fragile X and Rett Syndrome autism models (Rowitch and Kriegstein, 2010; Sofroniew and Vinters, 2010; Molofsky and Deneen, 2015). Unlike oligodendrocytes, astrocytes are rather enigmatic and the stages of astrogenesis are not so clearly understood as the individual stages lack specific markers. However they are still known to arise from RGCs undertaking the gliogenic switch

producing astrocyte precursors which then migrate through the cortex to their terminal location and transform to mature astrocytes (Rowitch and Kriegstein, 2010; Molofsky and Deneen, 2015).

1.6.7.3 Termination of the radial glial system

As pre-natal development draws to a close, neurogenesis has long ended, and neuronal and glial migration is complete making the radial fibres redundant. The remains of the proliferative RGC populations undergo final terminations into glial cells. Some proliferative populations do remain into adulthood and are an area of intense research in regenerative medicine. This includes the hippocampus where there are conflicting studies with one showing undetectable levels of adult hippocampal neurogenesis (Sorrells *et al.*, 2018) while another suggested that hippocampal neurogenesis persisted long into adulthood (Boldrini *et al.*, 2018). The extent of this neurogenesis is controversial and while it has been shown to occur in several species (reviewed in (Gage, 2002)), its occurrence in humans has not been extensively proven (reviewed in (Kempermann *et al.*, 2018)). Interesting as it is this topic far outside the scope of this thesis which will focus on neurogenesis in the human fetal cortex.

1.7 Autism Spectrum Disorder

1.7.1 Introduction to Autism Spectrum Disorder

First described in 1943 by Leo Kanner (Kanner, 1943), Autism Spectrum Disorder (hereafter abbreviated to ASD) is a neurodevelopmental disorder characterised by impaired social interaction and communication, repetitive and restricted behaviours and interests with typical (although not exclusive) early childhood onset of symptoms (WHO, 2019) which continue to persist without remission.

1.7.1.1 The umbrella of autism

As stated in the name, ASD is a spectrum of disorders, and the autism umbrella term is used to encompass five main subtypes of autistic disorders.

1) Pervasive developmental disorder
2) Asperger syndrome
3) Rett syndrome
4) Childhood disintegrative disorder
5) Autism disorder

The terminology of these is often changing and a source of disagreement among clinicians, as of the time of writing in the UK these are now classed under the broader autism diagnosis as per DSM-5 (Association and Force, 2013). In this thesis the individual subtypes of autism will not be discussed in detail and will be referred to as ASD.

1.7.1.2 Clinical features of ASD

Typically noticed in the first three years of life, the clinical features of ASD are highly variable between individuals however there is a convergence on impaired social communication. The behaviours most commonly associated with ASD, the so called **core symptoms**, are impaired social interests, impaired communication skills and highly rigid, specific interests (Nazeer and Ghaziuddin, 2012). As well as these stereotyped social impairments there are many **secondary symptoms** associated with ASD including hyperactivity, aggression, with some patients with ASD also being diagnosed with intellectual disability and epilepsy (Nazeer and Ghaziuddin, 2012).

1.7.1.3 Prevalence of ASD

The estimates for prevalence of autism does vary drastically depending on factors such as diagnostic criteria used, age of children, and geographical location. In general for the UK it is estimated to be around 1% of the population (Newschaffer *et al.*, 2007; Park *et al.*, 2016). While there has been attention focussed onto the increased prevalence of ASD within the last 20 years, this increase is likely to changes in diagnostic criteria and earlier diagnosis, however increased risk factors cannot be ruled out (Newschaffer *et al.*, 2007; Fisch, 2012; Park *et al.*, 2016).

The prevalence of ASD diagnoses is highly skewed between the sexes affecting four times as many boys than girls, however it is suspected that there is severe under-recognition in girls and the true number is much greater (Park *et al.*, 2016).

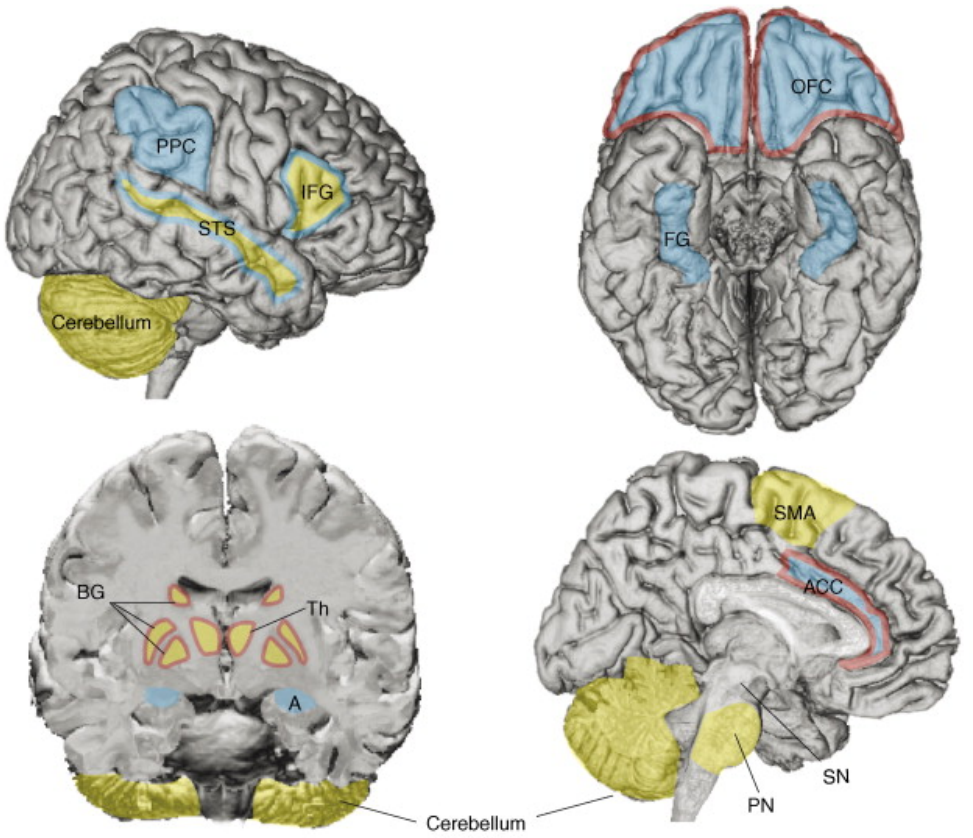
1.7.1.4 Heritability of ASD

ASD is a highly hereditary disorder, even in early case studies of the disorder the presence of a second autistic child in a family containing one was 50-100 times greater than expected by chance (Rutter, 1968). Twin studies further solidified this link, monozygotic twins have autism concordance rates of up to 70% (Folstein and Rosen-Sheidley, 2001).

Coupled with studies that showed autistic-like traits (although not reaching the threshold for diagnosis) were more prevalent in parents of children affected by ASD than those unaffected (Folstein and Rosen-Sheidley, 2001) (although it is important to note that in non-severe cases the diagnosis of ASD is “soft” and dependent upon the opinion of the clinician, and lack of diagnosis does not mean absence) this suggested some genetic underpinnings for ASD. The search for the genetic cause of autism is, and has long been, an area of intense research and will be summarised later in this chapter.

1.7.2 Neuroanatomy of ASD

While ASD commonly presents as a variable set of behavioural phenotypes, there have also been neuroanatomical changes identified in ASD patients encompassing structure, functions and connectivity, however it is important to note that none of these neuroanatomical changes have been identified as causative to ASD. There are many areas of the brain that are linked to the functions implicated by autism highlighting the complexity of the disorder. They are shown in Figure 12 and include regions throughout the brain including the thalamus, cerebellum and cortex. While it is likely they all to contribute to ASD in different ways through their different functions, in this thesis we will focus on the cerebral cortex.



Social impairment	Communication deficits	Repetitive behaviors
OFC – Orbitofrontal cortex ACC – Anterior cingulate cortex FG – Fusiform gyrus STS – Superior temporal sulcus A – Amygdala mirror neuron regions IFG – Inferior frontal gyrus PPC – Posterior parietal cortex	IFG- Inferior frontal gyrus (Broca's area) STS – Superior temporal sulcus SMA – Supplementary motor area BG – Basal ganglia SN – Substantia nigra Th – Thalamus PN – Pontine nuclei cerebellum	OFC – Orbitofrontal cortex ACC – Anterior cingulate cortex BG – Basal ganglia Th – Thalamus

TRENDS in Neurosciences

Figure 12: Brain regions implicated in autism

Shown are the brain regions implicated in the three core behaviours altered in ASD. Image from (Amaral, Schumann and Nordahl, 2008).

Interestingly a leading theory of the neuropathology of ASD, and one of the most replicated findings as well as a described feature in the original 1943 Kanner paper, is that of altered brain growth. This suggests that ASD begins much earlier than the onset of clinical symptoms. Many patients exhibit accelerated brain growth in early post-natal life, as evidenced by rapid increase in head circumference (a proxy for brain size) peaking to 10% greater than normal from 18 months to two years (Courchesne, Carper and Akshoomoff, 2003; Amaral, Schumann and Nordahl, 2008; Sacco, Gabriele and Persico, 2015). A key issue in MRI studies of ASD is the small sample sizes and age variability coupled with the fact that they tend to focus on higher-functioning individuals (most likely as they are willing to enter a scanner) (Amaral, Schumann and Nordahl, 2008; Lord and Jones, 2012). Regardless, MRI studies still provide valuable targets for identifying neuroanatomical changes in ASD patients.

1.7.3 Genetics of ASD

Autism was originally reasoned to be a disorder of neuropathological origin, a view which shifted in in the 1950s to consider it a psychiatric disorder stemming from deficiencies in parenting; so-called “refrigerator mothers”. This view has been relatively discarded and we circle back to the present day idea of autism as genetic neurodevelopmental disorder with a neurobiological explanation (Folstein and Rosen-Sheidley, 2001). While rogue theories do continue to persist surrounding the causes of autism, with one of the most harmful being the 1998 study linking the MMR vaccine to autism which was unequivocally refuted by a number of studies (DeStefano, 2007; Taylor, Swerdfeger and Eslick, 2014), the general consensus is for autism as a genetic disorder.

This section will discuss a little of the genetics of autism. There is some standard terminology used throughout which, before diving into the technicalities, is clearly defined in Table 1.

Table 1: Table of terminology

Term	Definition	Example
Syndrome	Collection of uncorrelated symptoms associated with a specific disease/disorder	Down's syndrome
Monogenic syndrome	A disorder arising from a mutation to a single gene	Fragile X
Polygenic syndrome	A disorder arising from the combined mutation of multiple genes	*16p11.2* microduplication and microdeletion
Idiopathic	Disorder which arises spontaneously and/or has no known genetic basis	Idiopathic epilepsy
De novo disorder	Disorder that arises for the first time in a family - often during meiosis	Angelman Syndrome
Inherited disorder	Disorder that has been passed down from parents	Cystic fibrosis

1.7.3.1 Genetic Studies of ASD

Autism is a disease with a complex and highly heterogeneous nature, while it tends to converge on a rather limited set of symptoms, there is no single genetic cause, making it a particularly intriguing area for study. Instead it can be considered to be multi-factorial arising from many different genetic risk factors (Packer, 2016). Even the most common genetic mutation associated with ASD, the *16p11.2* CNV, accounts for only 1% of ASD cases highlighting just how vast the autism aetiological landscape is (Abrahams and Geschwind, 2008). To try to further understand the genetics of ASD, in 2008 the Autism Gene Database was created, providing a database of causative and candidate ASD genes. There are many different types genetic changes (summarised in Table 1), which are associated with autism: the mutation of a single gene or mutation to a larger region, whether it occurs *de novo* or is inherited and whether the genetic cause is known or not.

1.7.3.1.1 Monogenic causes of ASD

The “traditional” identification of autism genes involved identifying a patient with a syndrome of which the features included autism, thus linking the mutation in that gene to autism. Within the past 20 years things have moved rapidly as genome sequencing has become easier and cheaper, as well as datasets becoming more widely available and larger cohort availability. This has allowed the field to expand from studying a small number of monogenic disorders, to consider the vast genetic landscape of autism. The advent of next generation sequencing (NGS) allows for unbiased interrogation into the genome of a patient with autism to identify mutations that may be associated with the autism diagnosis. Uploading this to databases will allow multiple researchers to correlate many cases worldwide to identify frequently occurring mutations in ASD patients.

Around 5-10% of ASD patients have been shown to have monogenic ASD; where a single gene mutation produces a syndrome, which includes autism as a symptom – also classing it as syndromic ASD (Benger, Kinali and Mazarakis, 2018; Wisniowiecka-Kowalnik and Nowakowska, 2019). The most common single-mutation (monogenic) syndrome associated with ASD is Fragile X syndrome and other monogenic disorders with ASD as a frequent symptom include Rett Syndrome, Angelman Syndrome and tuberous sclerosis (Sztainberg and Zoghbi, 2016; Benger, Kinali and Mazarakis, 2018; Wisniowiecka-Kowalnik and Nowakowska, 2019). Some of these, like Fragile X syndrome, are typically inherited (Saldarriaga *et al.*, 2014), others such as Angelman Syndrome frequently occur *de novo* (Clayton-Smith and Laan, 2003). These and other monogenic syndromes with high ASD prevalence are summarised in Table 2 however this is by no means an exhaustive list of genes, nor non-ASD clinical features. These single-gene mutations occur at a very low prevalence in the population and individually account for no more than 1% of ASD, however being able to link specific gene mutations to ASD in these patients provides a pathway to begin to understand the genetics of autism.

Table 2: Most common causes of monogenic ASD

Syndrome	Gene	Occurrence of ASD in patients with this mutation	Other clinical features
Fragile X	*FMR1*	30-60% (males only)	ID, epilepsy, hyperactivity
Rett Syndrome	*MECP2*	~60% (females only)	Microcephaly, regression, epilepsy, ID
MECP2 Duplication	*MECP2*	>90% (males only)	Epilepsy, ID, recurrent respiratory infections, GI dysfunction
Angelman Syndrome	*UBE3A*	~30%	Microcephaly, ID, epilepsy, excessive laughter and smiling
Tuberous sclerosis	*TSC1* *TSC2*	~50%	ADHD, epilepsy, benign tumours in many tissues
Timothy	*CACNA1C*	60%	ID, developmental delay, congenital heart malformations
Phelan-McDermid	*SHANK3*	75%	ID, schizophrenia,

			impulsivity, epilepsy
Neurofibromatosis Type 1	*NF1*	18%	ID, anxiety, macrocephaly, seizures

1.7.3.2 Copy number variations in autism

The search for the underlying genetic causes of ASD gained a new avenue of research with the use of array comparative genomic hybridisation (array CGH), a tool to detect microduplication and microdeletions in the chromosomes that would not be detected by traditional karyotyping. This identified *de novo* copy number variations (CNVs) as being strongly associated with ASD (Sebat *et al.*, 2007). A CNV is the duplication or deletion of a greater than 1000bp region of a chromosome. CNVs are relatively common in the human genome and have important functions for increasing biodiversity and be either inherited or occur *de novo* (Yoo, 2015).

The significance of CNVs in ASD came from studies which identified a greater occurrence of CNVs in idiopathic ASD patients (in whom there was no inherited autism-implicated genetic mutation) compared to controls (Vicari *et al.*, 2019) suggesting that presence of a CNV is a significant risk for ASD (Sebat *et al.*, 2007). CNVs can be classed as recurrent or nonrecurrent. **Recurrent CNVs** occur when the region of DNA is flanked by two almost identical (95-97% sequence homology) low copy repeats. This similarity makes the region predisposed to non-allelic homologous recombination (NAHR) resulting in duplications or deletions with reciprocal changes to gene dosage (Wisniowiecka-Kowalnik and Nowakowska, 2019). **Nonrecurrent CNVs** come from fork stalling and template switching (FoSTeS) and non-homologous end joining (NHEJ). This thesis will be focusing on the recurrent CNVs. As these CNVs contain many genes identifying which contribute to

the ASD phenotype is a challenge. Some of the recurrent CNVs most commonly identified in ASD patients are shown in Table 3. In addition, many of these CNVs have incomplete penetrance for ASD suggesting that the individual's genetic background will impact susceptibility for ASD.

Table 3: CNVs most frequently associated with ASD

Locus	**CNV**	**Occurrence of ASD in patients with this CNV**	**Other Clinical Features**	**Transmission**
1q21.1	Deletion	unknown	ID, schizophrenia, microcephaly, cataracts	Inherited or *de novo*
1q21.1	Duplication	8-50% (Brunetti-Pierri *et al.*, 2008; Mefford *et al.*, 2008; Aldinger, 2009)	Macrocephaly, ID, ADHD, heart defects	Inherited or *de novo*
2q37	Deletion	~25%	ID, seizures, hypotonia	*De novo* (very very rarely inherited)
3q29	Deletion	~30% (Pollak *et al.*, 2018)	Mild-moderate ID, anxiety, schizophrenia, microcephaly	More commonly *de novo*, occasionally inherited

7q11.23	Duplication	~20-30% (Mervis *et al.*, 2015; Klein-Tasman and Mervis, 2018)	ID, ADHD, macrocephaly, seizures	75% *de novo*, 25% inherited
15q11q13	Duplication	~80% (Hogart *et al.*, 2010)	ID, ataxia, epilepsy, behavioural difficulties	85% *de novo*, 15% maternal inherited
15q13.3	Deletion	~30%	ID (mild to moderate), epilepsy, heart defects	75% inherited, 25% *de novo*
16p11.2	Deletion	~25% (Niarchou *et al.*, 2019)	Macrocephaly, obesity, ADHD, ID (mild to moderate)	Mostly *de novo,* occasionally inherited
16p11.2	Duplication	~25 % (Niarchou *et al.*, 2019)	ID (mild to moderate), ADHD, low body weight, microcephaly	Mostly inherited by sometimes *de novo*
16p12.1	Deletion	unknown	ID (mild to moderate), epilepsy, microcephaly, hypotonia	Inherited
16p13.1	Deletion	~67% (Allach El	ID, schizophrenia,	90% inherited

		Khattabi *et al.*, 2018)	epilepsy, microcephaly	
17p11.2	Deletion	90% (Laje *et al.*, 2010)	ID (mild to moderate), disrupted sleep,	Mostly *de novo*, occasionally inherited
17p11.2	Duplication	~80% (Molina *et al.*, 2008)	ID (mild to moderate), heart defects, ADHD,	Mostly *de novo*, occasionally inherited
17q12	Deletion	~66% (Moreno-De-Luca *et al.*, 2010)	Schizophrenia, ID (mild to moderate), epilepsy, bi-polar disorder	Mostly *de novo*, occasionally inherited
17q21.31	Deletion	unknown	Epilepsy, hypotonia, cheerful disposition, ID (mild to moderate)	Mostly *de novo*, occasionally inherited
17q21.31	Duplication	unknown	ID (mild to moderate), microcephaly (Natacci *et al.*, 2016)	Unknown, so few cases
22q11.2	Deletion	~18% (Ousley *et al.*, 2017)	ID, ADHD, schizophrenia, autoimmune disorders	90% *de novo*, 10% inherited

22q11.2	Duplication	~14-25% (Wenger *et al.*, 2016)	ID, delayed growth, hypotonia, microcephaly	70% inherited, 30% *de novo*

1.7.4 Autism as a synaptic disorder

Despite its vast and somewhat confusing genetic landscape, ASD has a rather convergent profile of symptoms. Autism has historically be considered to be a synaptic disorder, not least of which because some genes first identified and most consistently reported in ASD patients such as *SHANK3*, *NRXN* and *NLGN2* are genes with direct synaptic roles (Zoghbi and Bear, 2012; Yoo, 2015; Packer, 2016). The synaptic hypothesis of ASD is outwith the scope of this thesis and will not be discussed.

1.7.5 Disruptions to neurogenesis in ASD

Despite the more traditional view of ASD as a synaptic disorder, there is a complimentary view that ASD begins much earlier during neurodevelopment, during neurogenesis and migration during cortical development (Casanova, 2014; Casanova and Casanova, 2014; Reiner *et al.*, 2016). As stated earlier, one of the hallmarks of ASD is early brain overgrowth, implying changes to proliferation earlier in neurogenesis (Amaral, Schumann and Nordahl, 2008). This section will summarise the evidence that the underpinnings of ASD begin early in neurodevelopment, the topic of investigation in this thesis.

1.7.5.1 Evidence from patient brains

Some of the most compelling evidence for changes to proliferation and neuron number in ASD patients comes from study of post-mortem brains. One small study in 2011 using post-mortem brain tissue from 7 boys with ASD and 6 controls showed a 67% increase in neuron numbers in the prefrontal cortex (Courchesne *et al.*, 2011), implying that excess neurogenesis may be occurring in ASD brain. This was further evidenced in 2012, a study showed differential expression of genes involved in controlling cell cycle and neural patterning in brains from young ASD patients compared to controls (Chow *et al.*, 2012). In addition in 2014 Stoner et al showed disruption to cortical lamination in ten out of eleven ASD post-mortem brains (Stoner *et al.*, 2014).

1.7.5.2 Alternative models for studying the neurodevelopmental basis of ASD

While a powerful tool, a key issue with post-mortem tissue is its availability. However, researchers are developing methods to circumvent this each of which provides their own advantages and caveats. The more traditional method from which some of our early understandings of ASD (and almost every other disease) have emerged are animal models. The somewhat newer approach is to use cells cultured from ASD patients and control individuals, induced pluripotent stem cells (iPSCs). iPSCs are an exceptionally powerful tool to allow us to understand ASD at the molecular level, and as a tool it is only growing with the rise of cerebral organoids, a 3-D assembly of cells in a dish that accurately recapitulate some aspects of human cortical development.

1.7.5.2.1 Animal models and the neurodevelopmental hypothesis of ASD

Mouse models have provided a powerful tool to understand the roles of autism associated genes play in neurodevelopment. As described above, ASD patients often present with brain overgrowth (macrocephaly), however microcephaly does also present in some ASD cases, typically more frequently in cases with more severe ASD and ID (Gilbert and Man, 2017). Table 4 summarises some ASD genes and CNVs implicated in disruption to neural progenitors and what we have learnt from the mouse model, this is by no means an exhaustive list.

Table 4: Studies using mice have provided insight as to how many ASD implicated genes may contribute to altered brain size

Gene/CNV	Macro/Microcephaly	Biological Function	Mouse Knockdown Model
PTEN	Macrocephaly	Tumour suppressor and negative regulator of PI3K signalling	Brain enlargement, seizures, increased progenitor proliferation (Chen *et al.*, 2015; Gilbert and Man, 2017)
CHD8	Macrocephaly (in 80% of *CHD8* patients (Bernier *et al.*, 2014))	Chromatin remodelling by binding to beta-catenin and regulator of Wnt signalling	Increased brain weight, ASD like behaviours (Katayama *et al.*, 2016)

		(Bernier *et al.*, 2014)	
WDFY3	Macrocephaly	Selected macroautophagy of aggregated proteins (Filimonenko *et al.*, 2010)	Enlarged cerebral cortex, migration defects of cortical projection neurons, increased symmetric progenitor divisions (Orosco *et al.*, 2014)
Ankrd11	Microcephaly	Regulates transcription by histone binding (Gallagher *et al.*, 2015)	ASD-like behaviour, reduced progenitor proliferation (Gallagher *et al.*, 2015)
WDR62	Microcephaly	Accumulates to spindle pole during mitosis	Microcephaly, decreased neural progenitor population (Chen *et al.*, 2014)
1q21.1 Deletion	Microcephaly	In particular the gene *Rbm8a*	Microcephaly, decreased progenitor and

			disrupted cortical lamination (Mao *et al.*, 2015)
22q11.2 Deletion	Macro and microcephaly reported (Coppinger *et al.*, 2009)	32 genes in this region (Meechan *et al.*, 2011) in particular the gene *Ranbp1*	Decreased intermediate progenitor production (Meechan *et al.*, 2015)
16p11.2 Deletion	Macrocephaly	This CNV is the focus of this thesis	Enlarged brain size (Horev *et al.*, 2011) and increased RGC proliferation (Pucilowska *et al.*, 2015)

While mice provide a powerful tool to study genetic mutations contributing to ASD, they, like any model, have some caveats. Firstly, there are fundamental differences in the cortical structure of mice and humans; the outer SVZ (see Figure 1). Another is that they can only be easily used to study syndromic ASD, where there is a known genetic cause such as FMR1. In addition most identified gene mutations associated with ASD are heterozygous which rarely induces ASD symptoms instead requiring a homozygous mutation suggesting that mice cannot accurately recapitulate the roles of these genes in ASD (Ey, Leblond and Bourgeron, 2011). Therefore, it is necessary to use other methods in conjunction with mouse models to unravel the genetics of ASD.

1.7.5.2.2 What we have learned from cell culture systems about ASD and neurogenesis

One powerful approach to study both idiopathic ASD and to take a different approach to studying syndromic ASD is the use of human induced pluripotent stem cells (hIPSCs). These can be taken from ASD patients from skin biopsies or blood and differentiated to neural progenitor cells (NPCs) to understand the effects of ASD genes on early proliferation. This system provides researchers with a 2-D system to study neurogenesis, or it can be transformed into a 3-D organoid system; cortical organoids recapitulate some aspects of human cortical development. They are both powerful systems and can be utilised in different ways depending on the questions of the research. This proved a particularly powerful tool in a 2017 study that showed NPCs derived from idiopathic ASD individuals showed faster proliferation than controls suggesting a proliferative defect that could underlie part of the ASD phenotype (Marchetto *et al.*, 2017). Regarding syndromic ASD, studies have also showed using hiPSCs derived from Rett syndrome patients to have delayed neuronal maturation (Kim, Hysolli and Park, 2011) and studies with hiPSCs from Williams Syndrome (associated with ASD and schizophrenia) patients showed incomplete neural commitment and altered cell cycle genes (Lalli *et al.*, 2016).

In conclusion there are many tools available that can provide powerful insights into the genetic landscape of ASD. As discussed above there are many genetic disorders with ASD as an associated symptom. This thesis is focusing on the *16p11.2* CNV.

1.8 The *16p11.2* CNV

The *16p11.2* CNV has been implicated in ~1% of ASD cases making it tied with *15q11-q13* duplication as the most frequent aetiologies of ASD (Cook and Scherer, 2008). This frequency makes it a promising area of research into how its genetic changes contribute to ASD and the focus of this thesis.

1.8.1 The *16p11.2* locus

The *16p11.2* locus is a 593kpb region at the p11.2 region of human chromosome 16 (BP4-BP5). It encompasses 29 protein coding genes. The region is flanked by two homologous 147kpb sequences that arose after the evolutionary divergence of humans from primates, generating a misalignment hotspot for this region (Nuttle *et al.*, 2016). The structure of the region is shown in Figure 13 and the non-identical genes (*BOLA2, GIYD2, SULT1A4 and SULT1A3* are repeated at either end) within the region are summarised in Table 5.

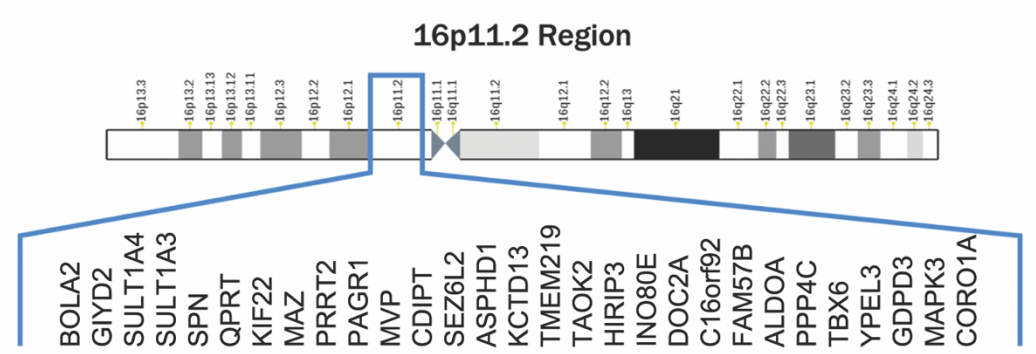


Figure 13: The 16p11.2 region

The *16p11.2* locus comprises 29 protein coding genes and is located on the p arm of the 16th chromosome in humans.

Table 5: Summary of the *16p11.2* locus genes

Gene Name	Function	Expressed in the Human Brain?	Associated Diseases
SPN	Protein is a highly sialylated glycoprotein required for antigen specific T cell activation	Low levels	Wiskott-Aldrich syndrome
QPRT	Key enzyme in the catabolism of quinolinate	Yes	High quinolinate in the brain associated to epilepsy, Alzheimer's and Huntington's
KIF22	Microtubule motors which move chromosomes during mitosis as well as spindle formation	Yes	Spondyloepimetaphyseal dysplasia
MAZ	Nucleic acid binding and chromatin binding	Yes	
PRRT2	Transmembrane protein	Yes	Paroxysmal kinesigenic choreoathetosis
PAGR1	Epigenetic transcriptional activation	Yes	
MVP	Gene encodes vaults for nucleo-cytoplasmic transport	Yes	Evidence of its upregulation in epilepsy (Balan *et al.*, 2013)

CDIPT	Involved in biosynthesis of phosphatidylinositol	Yes	
SEZ6L2	Encodes a seizure related protein on the cell surface	Yes	Epilepsy
ASPHD1	Dioxygenase activity	Yes	
KCTD13	Voltage-gated potassium channels	Yes	
TMEM219	IGFBP3 specific cell death receptor	Yes	
TAOK2	Encodes a protein kinase	Yes	Mice have dose dependent changes in head size (Richter *et al.*, 2018)
HIRIP3	Complexes with HIRA to function in chromatin and histone metabolism	Yes	
INO80E	Involved in transcriptional regulation	Yes	
DOC2A	Involved in Ca2+ dependent neurotransmitter release	Yes	
C16orf92	Protein coding gene		
FAM57B	Gene encodes a transmembrane protein	Yes	
ALDOA	Encodes a glycolytic enzyme catalyses	Yes	ALDOA deficiency

	the reversible conversion of fructose-1,6-bisphosphate to glyceraldehyde 3-phosphate and dihydroxyacetone phosphate		
PPP4C	Involved in microtubule organisation	Yes	
TBX6	Involved in fate determination of axial fate	Yes	
YPEL3	Proliferation of myeloid precursor cells	Yes	spondylocostal dysostosis
GDPD3	Glycerophospholipid biosynthesis	Very low levels	
MAPK3	Encodes member of the ERK family which has many roles in proliferation and cell cycle	Yes	
CORO1A	Implicated in T-cell mediated immunity	Very low levels	immunodeficiency

1.8.2 *16p11.2* Copy Number Variation

Large, recurrent CNVs are implicated in many neuropsychiatric disorders including ASD, epilepsy, ID and schizophrenia (Girirajan and Eichler, 2010; Kirov, Rees and Walters, 2015). CNVs are not causative to any of these disorders; while carrying a particular CNV will increase the risk of developing the associated disorder it is not guaranteed. Even among siblings with the same CNV, there are incomplete penetrance of phenotype and high phenotypic variability of symptoms (Levy *et al.*, 2011; Giaroli *et al.*, 2014; Qureshi *et al.*, 2014). The homozygous deletion or duplication (Figure 14) of the *16p11.2* locus can be inherited but it predominantly occurs *de novo* during meiosis (Shinawi *et al.*, 2010) and is implicated in a wide range of neurodevelopmental and neuropsychiatric disorders (Kumar *et al.*, 2008; Bijlsma *et al.*, 2009; Rosenfeld *et al.*, 2010; Shinawi *et al.*, 2010; Hanson *et al.*, 2015).

The *16p11.2* deletion was first described in 2007 in male monozygotic twins who presented with seizures, ID and aortic valve abnormalities (Ghebranious *et al.*, 2007). Its implication in ASD was proposed shortly after by Sebat et al who identified its *de novo* deletion in a screen of simplex families (a simplex family is a family where only one member has ASD) (Sebat *et al.*, 2007). Since then a range of studies have rapidly implicated the *16p11.2* CNV in a range of neurodevelopmental and neuropsychiatric disorders (Kumar *et al.*, 2008; Bijlsma *et al.*, 2009; Rosenfeld *et al.*, 2010; Shinawi *et al.*, 2010; Hanson *et al.*, 2015). The clinical features of this CNV extend beyond the neuropsychiatric - including changes to body mass, and are highly variable in their presentation with incomplete penetrance of CNV; 46.8% for the deletion and 27.2% for the deletion (Rosenfeld *et al.*, 2013).

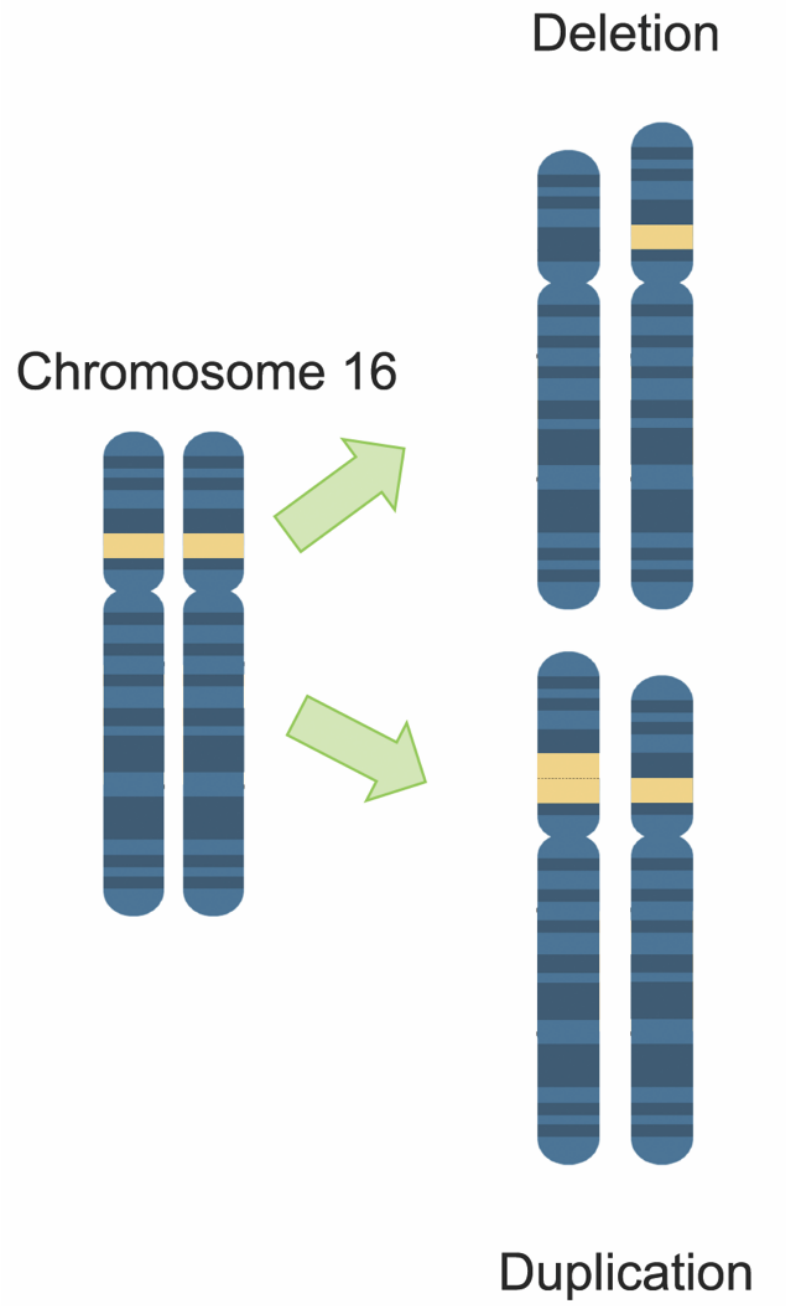


Figure 14: The *16p11.2* copy number variation

This CNV can result in the heterozygous duplication or deletion of the *16p11.2* locus of chromosome 16

1.8.2.1 Dosage changes in the *16p11.2* CNV

The *16p11.2* CNV is heterozygous not homozygous and therefore will cause a change in gene dosage rather than complete ablation of the gene. Altering the gene dosage will alter mRNA levels and subsequent protein levels (Figure 15). This expression change is the potential driving pathogenic mechanism in CNVs (Blaker-Lee *et al.*, 2012). This change in gene dosage has been confirmed in mice (Horev *et al.*, 2011; Blumenthal *et al.*, 2014) and human lymphoblasts (Blumenthal *et al.*, 2014) showing that gene dosage disruption predictably impacts mRNA expression levels; increasing to 1.5 times that of the control in the microduplication and decreasing to 0.5 times that of the control in the microdeletion.

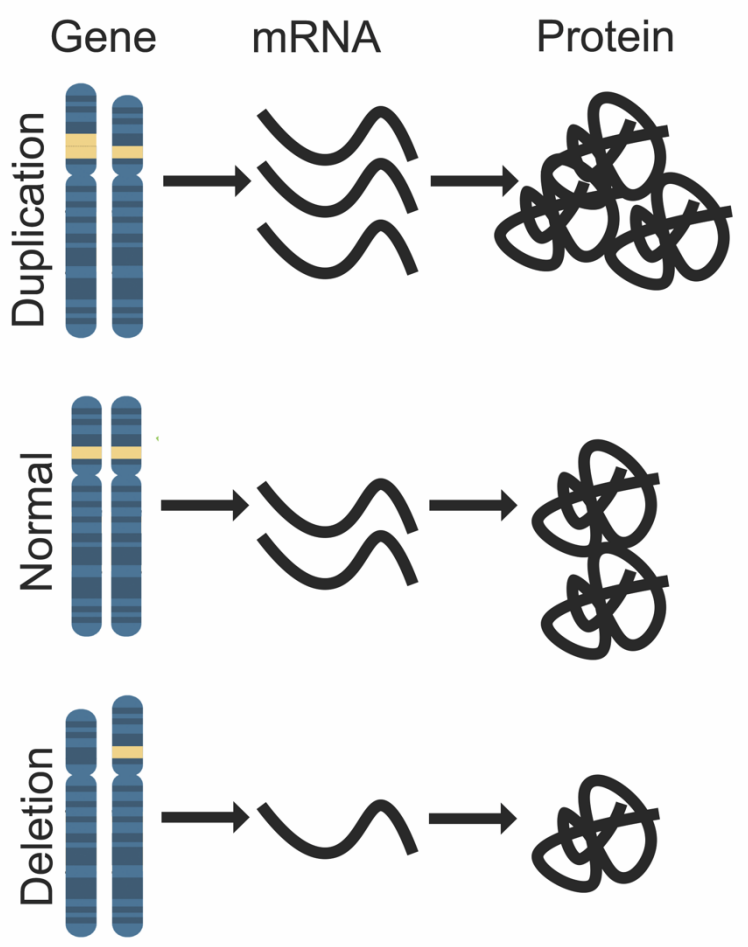


Figure 15: Effect of changes of gene dosage on transcription and translation

Altering the numbers of copies in the gene by heterozygous deletion will result in less mRNA being transcribed and subsequently less protein being translated. Conversely, in a heterozygous duplication more mRNA will be transcribed, and more protein translated.

1.8.2.2 Clinical features associated with the *16p11.2* CNV

There is growing evidence that the opposing duplication and deletion of the *16p11.2* region produce mirrored phenotypes. This section will briefly discuss the various clinical features associated with the duplication and deletion.

One of the most striking features in many patients that appears well before the onset of clinical neuropsychiatric symptoms is the changes to head size at birth, a clear example of the mirrored phenotype of this CNV. Both McCarthy et al in 2008 and Shinawi et al in 2010 showed increased head size (macrocephaly) in deletion patients and decreased head size (microcephaly) in duplication patients (McCarthy *et al.*, 2009; Shinawi *et al.*, 2010). As suggested earlier in this chapter changes to head size are indicative of altered brain size which is often due to altered neural proliferation. Continuing the mirror phenotype the deletion is associated with a highly penetrant form of obesity while duplication patients are often underweight (Walters *et al.*, 2010; Maillard *et al.*, 2015).

Regarding neuropsychiatric disorders, their occurrence also differs between microdeletion and microduplication patients. Meta-analysis of several studies evidenced that the microduplication to be strongly associated with multiple psychiatric disorders including schizophrenia, bi-polar disorder and autism. The microdeletion is strongly associated with autism, developmental delays but not schizophrenia (McCarthy *et al.*, 2009; Hanson *et al.*, 2015). In addition, both the microdeletion and the microduplication, but more strongly the microduplication, are associated with ADHD (Shinawi *et al.*, 2010) and both are heavily associated with speech delays and motor delays (Shinawi *et al.*, 2010).

Epilepsy and seizure disorders are also frequent in both the microduplication and the microdeletion (Shinawi *et al.*, 2010). In addition, many microdeletion

patients share facial features including broad forehead, hypertelorism and micrognathia and while microduplication patients present with facial dysmorphia, there is no consistent pattern (Shinawi *et al.*, 2010). Several studies have also identified altered brain structures in a subset of patients, including corpus callosum hypoplasia (Bedoyan *et al.*, 2010), enlarged ventricles with thinned corpus callosum and thinned periventricular white matter (Filges *et al.*, 2014), cerebral cortical atrophy (Ghebranious *et al.*, 2007) and partial left temporal lobe agenesis (Bijlsma *et al.*, 2009).

1.8.3 *16p11.2* CNV and the neurogenic basis of autism

The changes in head size in *16p11.2* patients implicate neurogenic defects during prenatal development that may contribute to the patient phenotype. There is range of evidence from different models supporting this hypothesis.

Murine models generated to harbour the *16p11.2* CNV exhibit autism-like behaviours with opposing changes to the size of some brain regions in the duplication and deletion in one study (Horev *et al.*, 2011). Notably this has been contradicted by later studies showing reversed head size phenotypes (Pucilowska *et al.*, 2015). However, this may well be due to the difference in human and mouse brain development (Figure 1) or it may be that the different backgrounds of the mice strains used induce this difference. Despite this, the murine model shows clear disruption to cortex development, supporting the contribution of altered neurogenesis in *16p11.2* pathogenesis. The deletion mouse shows increased mitosis of apical and basal progenitors, with increased IP mitosis increasing neuron production mid-neurogenesis and the progenitor pool prematurely being depleted causing microcephaly (Pucilowska *et al.*, 2015).

In addition, while not thoroughly investigating changes to neural progenitors such as analysing changes to cell cycle length, a 2017 study from Deshpande et al deriving cortical neurons from *16p11.2* patient iPSCs

showed altered neuron size which mirrored the opposing patients brain size as well as diminished synapse production (Deshpande *et al.*, 2017). Therefore, it is possible that changes in cell size may explain the patient head size phenotypes rather than changes to cell numbers.

1.8.3.1 Effect of individual *16p11.2* genes on neurogenesis

The *16p11.2* CNV harbours 29 genes, to understand how this region contributes to the patient phenotype, studies to understand the individual roles of these genes during brain development are essential. These are summarised in Table 6 and suggest that changes to at least some of the *16p11.2* genes are likely to contribute to neurogenic causes of ASD.

Table 6: Evidence implicating some of the *16p11.2* locus genes in altered proliferation

Gene	Model	Effect	Reference
QPRT	Knockdown in SH-SY5Y neuroblastoma cells	Altered neuronal differentiation	(Haslinger *et al.*, 2018)
MAZ	Knockdown in murine isolated NPCs	Inhibition of gliogenesis	(B. Liu *et al.*, 2016)
PRRT2	In utero electroporation knockdown in mouse embryos	Delayed neuronal migration	(Y.-T. Liu *et al.*, 2016)
KCTD13	Knockdown in zebrafish	Altered head size (note this finding has been disputed)	(Golzio *et al.*, 2012)
PPP4C	Deleted in the mouse cortex	Spindle misorientation causing progenitors to prematurely	(Xie *et al.*, 2013)

		switch to neurogenic divisions	
MAPK3	Deletion of Erk protein early in mouse neurogenesis	Elongated cell cycle, with neurogenic differentiation	(Pucilowska *et al.*, 2012)

1.8.3.2 The *16p11.2* phenotype is likely polygenic

It is unlikely to be one single gene from the *16p11.2* region causative for the neuroanatomical phenotype, instead the phenotype is likely to be polygenic with the phenotype stemming from the altered dosage of several *16p11.2* genes. This is supported by several lines of evidence including a study systematically knocking down each gene in zebrafish embryos reporting neuroanatomical phenotypes for the vast majority of genes, although 50% knockdown only induced phenotypes in two genes; *ALDOA* and *KIF22* (Blaker-Lee *et al.*, 2012). In addition systematic deletion of all combinations of the genes in drosophila showed complex interactions between all of the genes to contribute to the phenotype (Iyer *et al.*, 2018).

1.8.4 Summary of evidence

Evidence indicates that multiple *16p11.2* transcript levels are affected by the *16p11.2* CNV and the pathology of the *16p11.2* CNV could stem from altered dosage of either one of them. Altered proliferation during cortex development is observed in mouse models, however it is unknown which of the proteins produced by *16p11.2* CNV genes are expressed by progenitor cells in the developing human cerebral cortex and are therefore candidates for regulating neurogenesis.

1.9 Aims of Thesis

This thesis has the following aims:

1) To perform an unbiased screen of an existing scRNA-seq dataset from the developing human cerebral cortex to identify genes from the *16p11.2* locus whose transcripts are expressed in progenitors, in particular genes whose mRNA levels are then reduced as cells become post-mitotic. This bioinformatics analysis was performed by Yifei Yang.

2) To describe the mRNA expression of these genes expressed in progenitors using sections of the developing human fetal cerebral cortex during the late first and early second trimester.

3) Having identified genes highly expressed in cortical progenitors and down regulated as cells become postmitotic we aim to describe their protein expression during development and their changes during the cell cycle.

2 Materials and Methods

2.1 Human tissue

Human tissue was obtained either from the Human Developmental Biology Resource (hereafter referred to as HDBR) or the Royal Edinburgh Infirmary (hereafter referred to as RIE).

2.1.1 Samples obtained from HDBR

Samples obtained from HDBR were collected and processed by the staff at HDBR. The processes used are described here but is work performed by HDBR staff.

2.1.1.1 Collection

All samples were collected from women >16 years old undergoing termination of pregnancy. Donation was voluntary and written consent given by donors only after they received counselling regarding termination of pregnancy. Following collection all information was anonymised.

2.1.1.2 Staging of fetuses

Staging of foetus was estimated based on patient information and ultrasound prior to termination procedure and then confirmed upon arrival at HRBR based on length measurements and other developmental guides e.g. facial and hand development.

2.1.1.3 Dissection and fixation

Small slits were cut in the skull of the head and fixed overnight in formalin or methacarn to preserve tissue morphology. The brain was then dissected out and cut into pieces depending on size of sample and further fixed for 24-48 hours dependent on tissue size.

2.1.1.4 Embedding

Following fixation and washing brain pieces were embedded in paraffin using a processer resulting in paraffin blocks.

2.1.1.5 Sectioning

The paraffin blocks where then sectioned at 8μm using a wax microtome and mounted onto superfrost plus slides where they were stored at room temperature prior to shipping to Edinburgh for use in this project.

2.1.2 Samples obtained from RIE

Collection of fetal samples from Edinburgh was coordinated by Richard Anderson and day to day collection and coordination was by Anne Saunderson and Norma Forson. They performed all clinical duties regarding consent and collection and brought the specimen to RIE for collection by researchers.
Ethical approval for this study was obtained from Lothian Research Ethics Committee (study code LREC 08/S1101/1), and women gave informed written consent. Human fetuses were obtained after elective termination of pregnancy, and all fetuses used in this study were morphologically normal.

Gestational age was determined by ultrasound scan, and confirmed (for second trimester fetuses) by direct measurement of foot length
Patients deemed suitable for the study were required to be <16 years old, without any known infection (HIV, Hepatitis B/C/E or chlamydia) and without any identified fetal/pregnancy abnormality. Patients were also excluded from study if nursing staff had any concerns about their emotional state, they had learning disabilities, or if the patients was not fluent in English when there were no adequate translation services available at the time. Patients were approached by research midwifes and provided a patient information sheet, if they wished to proceed with sample donation, they signed a consent form with the understanding they could change their mind at any time. Following collection all information was anonymised.

2.1.2.1 Dissection and fixation

Dissection of specimens from RIE was performed by Sarah. Briefly the head was removed and transported back to Hugh Robson Building on ice. There, small slits were cut in the skull in the soft membrane where the four unfused skull plates meet, and the specimen placed in 4% paraformaldehyde (hereafter termed 4% PFA) overnight at 4ºC to preserve brain morphology.

The next day the sample was washed 10x15 minutes in 1 x PBS on a rocker to remove PFA. The sample was then placed in a 10cm petri dish, the skin removed, and skull plates removed to expose the brain. Cutting the brainstem (or as close to possible) allowed removal of the brain from the skull which was then placed in PFA at 4ºC overnight. Skull fragments and any leftover tissue were returned to RIE to be respectfully disposed of with the rest of the sample

The following day the brain sample was again washed 10 x 15 minutes with 1 x PBS on a rocked to remove all PFA. The brain was then divided down the

middle and cut into 8 pieces (as shown in Figure 16). These sections were transferred to individually labelled specimen jars with PFA and fixed at 4ºC for 5 days.

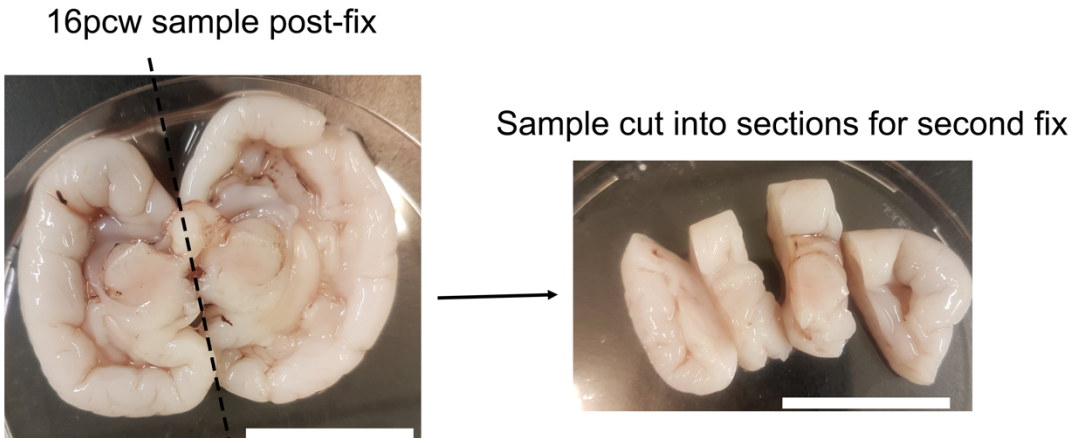


Figure 16: How tissue was divided up for sectioning

We cut down the middle of the sample and then divided each hemisphere into four blocks for secondary fixation. Scale bar = 3cm

2.1.2.2 Cryoprotection and embedding

After the final fixation step, brain sections were washed 5 x 15 minutes in 1 x PBS on a rocker at room temperature. Following this the PBS in each specimen jar was replaced with 30% sucrose/1 x PBS and left at 4ºC for 3 days to cryoprotect the tissue.

Following cryoprotection each section was transferred to a mould (Thermofisher scientific) and covered with a 50:50 mix of OCT and 30% sucrose/1 x PBS and allowed to equilibrate for 30 minutes at room temperature. Samples were then flash frozen on dry ice and transferred to -80ºC for long term storage until sectioning.

2.1.2.3 Sectioning

Samples were transported on dry ice from -80ºC storage to the cryostat. OCT was used to attach the tissue block to the chuck and sections were coronally cut at 14μm and mounted on Superfrost positively charged slides. Slides were allowed to air dry for 2 hours before storage at -20ºC.

2.1.3 Haematoxylin and Eosin Staining

For the H&E staining cryostat sections were used and allowed to warm to room temperature for 30 minutes. The staining was performed on adjacent sections to those used for *in* situ hybridisation and all developmental ages stained simultaneously. To try to reduce tissue loss, a secondary fixation step was performed using Formalin-acetic-acid-alcohol (FAA). Here sections were immersed in FAA for 30 seconds at RT and then rinsed under running water for two minutes before being immersed in Haematoxylin for two minutes. Sections were then rinsed under running water for two minutes and then immersed in water-based Eosin and then rinsed in running water for one then immersed in potassium aluminium for two minutes. Next slides were rinsed for two minutes under running water and then dehydrated by immersion in increasing alcoholic concentrations: 70%, 90%, 95% followed by two changes of 100%. Sections were then immersed in xylene and mounted onto coverslips with the DPX mounting medium.

2.2 *In situ* hybridisation

2.2.1 Probes used

The probes were used in this thesis are for *KIF22, ALDOA, HIRIP3, PAGR1, MAZ and SPN*. All were designed and generated in house.

2.2.2 Probe preparation

The pGEMT easy kit was used for creation of plasmids. All cDNA used was converted from RNA extracted from the NAS2 cell line, with the exception of the *KIF22* and *ALDOA* plasmids which used purchased Image cDNA (source bioscience). *KIF22* and *ALDOA* probes were made with the help of Kathy Howe.

2.2.2.1 Primer design

The following primers were designed using primer3 software to target all known splice variants of each gene according to BLAST and are shown in Table 7. Oligos were ordered dry from Thermo Fisher scientific. Upon arrival they were resuspended in RNase-free ddH_2O to a primer concentration of 25μM.

Table 7: Primers used for *in situ* hybridisation probes

Gene	Forward Primer	Reverse Primers
KIF22	CGA CGA CGG ATG GTG CTA AT	GAG ACC CAG GAT GTT TGC CT
ALDOA	CTG TCA CTG GGA TCA CCT T	GTG ATG GAC TTA GCA TTC AC
PAGR1	ATG ATG AGC CAG TGA CAC CA	TCT GCC TCT CCC TTC AAG TG
HIRIP3	TGG TGC CCA TCG AAA CTA CA	TGG CCC AAA ATA CAG GAG GT
MAZ	CAC GAG GAG AAA GTG CCA TG	GAG AGA AGA GGA CCG TCG AG
SPN	TGA TCT TTC CCA GGG CAC AT	GAG AGA GAG AGA GCG CAC AT

2.2.2.2 Extraction of human cDNA

For amplification of target DNA sequence human cDNA was required. This was obtained from the NAS2 hIPSC line gifted from John Mason. Briefly cells were cultured at 37ºC in Stem Max IPS Brew XF in 6 well plates until 70-80% confluency. RNA was extracted using the RNeasy kit from Qiagen and nanodropped to determine RNA concentration.

RNA was then converted to cDNA as follows: RNA was diluted to 2μg/13μl in RNAse free water and 2μl random primers (500ng/μl) added and then the mixture heated at 70ºC for 5 minutes then rapidly cooled on ice. 10μl of Promega M-MLV reverse transcription mix was added and the mixture heated

at 37ºC for 60 minutes to produce cDNA at a final concentration of 80ng/μl. cDNA was then diluted to 5ng/μl in ddH_2O and stored at -20ºC.

2.2.2.3 Amplification of target DNA sequence

Primer mix was made as shown in Table 8

Table 8: Primer mix used for PCR

Forward Primer	**Reverse Primer**	**ddH_20**
10μl of 25μM primer	10μl of 25μM primer	80μl

The target sequence was amplified by PCR prior to transfection into pGEMT Easy vectors using the following reagents in Table 9.

Table 9: Reagents used for PCR amplification of target sequence

Reagent	**Volume**
ddH_20	16.3μl
DNTP mix	2.5μl
Promega PCR buffer	5μl
Taq polymerase	0.2μl
Human cDNA	1.5μl

The PCR mix was then run on the PCR cycle shown in Table 10.

Table 10: PCR cycle times and temperatures

Temperature	Time	
96ºC	2 minutes	
96ºC	30 seconds	28 Cycles
59ºC	30 seconds	
72ºC	30 seconds	

PCR product was then cleaned up using the QIAGEN gel extraction kit and nanodropped to confirm concentration.

2.2.2.4 DNA clean up

To clean up DNA the QIAgen gel extraction kit was used. 200ul of DNA was used and 200μl of isopropanol and 600μl of buffer QC were added. The mixture was added to a QIAquick spin column and placed into a 2ml collection tube. This was then centrifuged at 13000rpm (speed used throughout) for 60 seconds to bind DNA to column and the flow through discarded. 500μl of buffer QG was added to column and centrifuged for 60 seconds and flow through discarded. Then 750μl of buffer PE was added, allowed to stand for 5 minutes then column centrifuged for 60 seconds and flow through discarded. The column was then centrifuged for 60 second, flow through discarded and centrifuged for another 60 seconds to dry the filter. Column was placed into a 1.5ml centrifuge tube and 20ul of sterile ddH_20 added directly to the centre of the column and allowed to stand for 5 minutes before being centrifuged for 60 seconds. Eluted DNA was collected in the 1.5ml tube and stored at -20ºC.

2.2.2.5 Agarose Gel

After amplification and clean up the mix was run on a 1.5% agarose gel (Table 11) for 30 minutes to confirm a band indicating successful amplification.

Table 11: Reagents used for a 1.5% agarose gel

TBE Buffer	Agarose	Voltage	Syber Safe
50ml	0.75g	260mV	1.5μl

The reagents for x5 stock TBE buffer are shown in Table 12, this was diluted to x1 in ddH_20 prior to use.

Table 12: Reagents for x5 stock TBE buffer

Reagent	Volume
Tris base	54g
Boric acid	27.5g
0.5M EDTA	20ml
H_20	Make up to 1l
HCl	pH to 8.3

2.2.2.6 Ligation

Ligation was performed as per the instructions in the pGEMT easy kit. A 1:1 ration of insert:vector ratio was used for all plasmids. The following in Table 13 was used for the ligation reaction:

Table 13: Reagents and volumes used for ligation reaction

Component	Volume	Positive control	Background
2x ligation buffer	5μl	5μl	5μl
Vector	1μl	1μl	1μl
PCR product	Appropriate volume depending on insert size	-	-
Control DNA	-	2μl	-
T4 ligase	1μl	1μl	1μl
ddH_20	To final volume of 10μl	1μl	3μl

Solutions were thoroughly mixed, and ligation allowed to occur overnight at 4ºC.

2.2.2.6.1 Ampicillin Plates

Ampicillin plates were used to grow transformed cells on. 250μl of ampicillin was added to 250ml of autoclaved agar and poured into petri dishes. As the vector contained an ampicillin resistant insert this ensured only bacteria which had taken up the plasmid could grow. Plates were allowed to set and then stored upside down at 4ºC to prevent condensation contamination. Just

prior to use plates were coated with 100μl XGAL/BTP solution (Sigma) and incubated at 37ºC for 30 minutes.

2.2.2.7 Transformation

Transformations required the pGEMT easy vector to be heat shocked into JM109 competent cells (Promega) as per the instructions. The reaction reagents used are shown in Table 14.

Table 14: Reagents for transformation of competent cells

Component	Volume (μl)
Ligation mix	2
Competent Cells	50

Cells were thawed on ice and following addition of reagents the reaction mixture was incubated on ice for 20 minutes before being heat shocked for 45 seconds in a 42ºC water bath then returned to ice for 2 minutes.

950μl of SOC medium (Invitrogen) was added to transformation mix and incubated 1.5 hours at 37ºC shaking at ~150rpm.

100μl of each culture was streaked onto plates using sterile rods and incubated at 37ºC overnight.

2.2.2.8 Identification of transformed colonies

The next day the effectiveness of the ligation was confirmed as background plates were all blue and control were all white. To preserve transformed colonies a grid system was used. A grid was drawn and placed under a new dish (see Figure 17) and one white colony streaked to each square of the grid using a sterile P200 pipette tip.

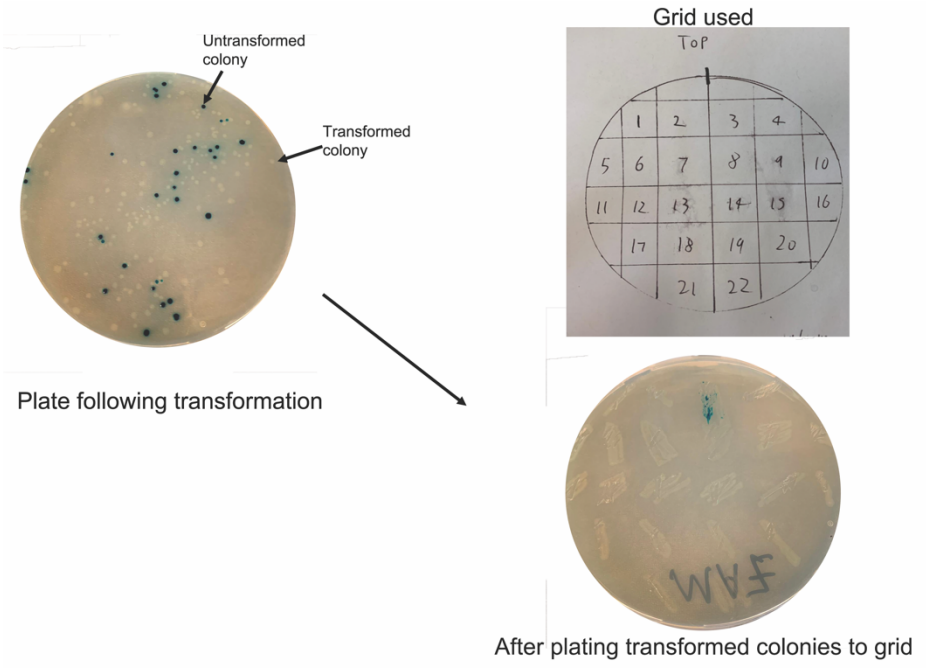


Figure 17: Schematic of colony selection and streaking onto grid

Blue colonies had not taken up the plasmid so were not used; white colonies had taken up the plasmid so where used for streaking to the grid plate.

2.2.2.9 Plasmid mini-preps

To set up for the mini-preps, a dot of each bacterial colony was added to 4ml of sterile LB broth and incubated overnight at 37ºC on a shaker ~160rpm. Next day tubes were removed from shaker and the mini-prep performed with the Qiagen mini-prep kit. Cultures were centrifuged at 13000rpm at room temperature (RT) using a table top centrifuge for 3 minutes. The supernatant was removed, and pellet resuspended in 250μl of buffer P1 with added RNase A and LyseBlue. 250μl of buffer P2 was added and tube thoroughly mixed by inverting 6 times until the solution turned blue. 350μl buffer P3 was added immediately by inverting the tube 6 times until the solution turned clear. Mixture was centrifuged at RT at 13000rmp (standard speed for this protocol) for 10 minutes. The supernatant was carefully removed and transferred to a QIAprep spin column and the pellet discarded. Column was centrifuged for 60 seconds and flow through discarded. 500μl of buffer PB was added and column centrifuged for 60 seconds to wash the DNA then flow through discarded. 750μl of buffer PE was added and centrifuged for 60 seconds and flow through discarded. Column was centrifuged for 60 seconds and flow through discarded then centrifuged for another 60 seconds and flow through discarded to thoroughly dry column. The column was transferred to a new 1.5ml centrifuge tube and 40μl of sterile ddH_20 added directly to the centre of the column and left for 5 minutes. Column was then centrifuged for 60 seconds and then 5ul of the DNA run on a 1.5% agarose gel to confirm DNA. 1μl of samples that produced a band were used for nanodrop and the two with the strongest concentration of DNA used for sequencing.

2.2.2.10 Sequencing of plasmids

Sequencing of the vector was Sanger Sequencing performed by Edinburgh Genomics. Forward and reverse sequencing was performed separately, and two samples were sequenced for each gene. The reagents provided to Edinburgh Genomics are shown in Table 15 and were provided in a 0.2ml strip tube with the DNA and primer combined to final volume of 6μl.

Table 15: Reagents provided for sequencing

Reagent	**Forward Sequencing**		**Reverse Sequencing**	
	Volume	**Concentration**	**Volume**	**Concentration**
Clean DNA	5μl	300-400ng	5μl	300-400ng
Primer	1μl	6.4μM	1μl	6.4μM

Sanger sequencing standard Big Dye reaction was performed by Edinburgh Genomics using the following Big Dye master mix (Table 16) and then the Big Dye reaction mix (Table 17).

Table 16: Reaction components for BigDye master mix

Reaction Components	**1 x reaction**
5x sequencing buffer	2μl
BigDye	1μl
Molecular grade water	1μl
Total Volume	4μl

Table 17: Reagents for BigDye Reaction

Reaction Components	1 x reaction
Big Dye master mix	4μl
DNA	5μl
Primer at 6.4μM	1μl
Total Volume	10μl

The following PCR reaction in Table 18 was performed by Edinburgh Genomics to amplify the DNA prior to sequencing:

Table 18: PCR settings for BigDye Reaction

Temperature	Time	
96ºC	60 seconds	
96ºC	10 seconds	Repeat for 24 cycles
50ºC	5 seconds	
60ºC	60 minutes	
12ºC	Hold	

The BigDye reaction mix was then column purified, dried in a vacuum, resuspended in HiDi and denatured prior to sequencing run on the 3730XL sequencing machine.

Sequencing data was then returned and compared to BLAST to ensure DNA region covered all splice variants of the gene and matched to the reference genome.

2.2.2.11 Amplification of sequenced plasmid

Following confirmation of sequencing results the best colony (best meaning clearest and longest sequencing result) of each gene was selected from the

grid and a dot added to 150ml of LB broth with 150μl of Ampicillin added and incubated at 37ºC overnight shaking at ~160rmp.

2.2.2.12 Midi-prep of plasmid

The next day cultures were removed, and the plasmid extracted using the Qiagen HiSpeed plasmid midi kit. Cells harvested from culture were spun down at 6000 x g for 15 minutes at 4ºC. All supernatant was removed, and the pellet resuspended in 6ml of chilled buffer P1 (with RNase solution A and LyseBlue reagent already added to buffer) by vortexing. 6ml buffer P2 was added and the mixture inverted 6 times until the solution turned blue and then allowed to incubate for 5 minutes at RT. 6ml of chilled buffer P3 was added and the solution inverted 6 times until the solution turned clear. The lysate was transferred to the barrel of a QIAfilter cartridge and incubated at RT for 10 minutes. During this incubation a HiSpeed tip was equilibrated with 4ml of buffer QBT and the column allowed to empty by gravity flow. After the 10 minutes incubation the plunger was fitted into the QIAfilter cartridge and the lysate filtered into the previously equilibrated HiSpeed tip which was then washed with 20ml buffer QC by gravity flow. The bound DNA was then eluted into a 15ml falcon tube using 5ml buffer QF. 3.5ml of isopropanol was added, mixed and incubated for 5 minutes at RT to precipitate the DNA.
The elute-isopropanol mix was then transferred into the syringe with a QIA precipitator, the plunger inserted, and the mixture filtered with constant pressure on the plunger. The QIA precipitator was then removed from the syringe, 2ml of 70% ethanol added, the QIA precipitator re-attached and the mixture filtered. To dry the membrane the QIA precipitator was removed, plunger removed, precipitator reattached and pressing air through the filter. This was repeated several times to thoroughly the membrane and then the membrane dried with absorbent paper. The QIA precipitator was attached to a new 5ml syringe, 1ml of buffer TE was inserted to the syringe, the plunger inserted, and mixture filtered into a clean 1.5ml centrifuge tube. QIA

precipitator and plunger were removed, eluted mixture transferred back to the syringe and re-filtered into the same 1.5ml tube.

DNA was then nanodropped and stored at -20ºC until required.

2.2.2.13 Plasmid linearization

To determine the appropriate enzyme for linearization of the plasmid it was first necessary to determine the orientation of the plasmid in the vector using the sequencing data, from this the correct restriction enzyme for linearization can be identified and this information is shown in Table 19:

Table 19: Orientation of plasmids and restriction enzymes used for digest

Plasmid	Orientation in vector	Restriction enzyme
KIF22	5'-3'	NCO1
ALDOA	3'-5'	SAL1
HIRIP3	3'-5'	SACII
MAZ	3'-5'	SACII
PAGR1	5'-3'	PST1
SPN	5'-3'	PST1

The reaction mixture for the digest is shown in Table 20, restriction enzymes and buffers are from New England Biolabs and the digest mixture was incubated at 37ºC for three hours

Table 20: Reagents used for digest

Reagent	Volume
Plasmid DNA	Required volume for concentration of 10μg
Enzyme	5μl
BSA (1000x)	2μl
10x Buffer	20μl
ddH_20	To final volume of 200μl

2μl of the digested plasmid was run for 30 minutes on a 1.5% agarose gel with 2μl of the unlinearised plasmid to confirm the digest was successful. The DNA was then cleaned up using the QIAgen gel extraction kit and eluted with 20μl of sterile ddH_20 and stored at -20ºC.

2.2.2.14 Transcription and DIG labelling of Plasmid

To make a DIG labelled anti-sense probe the reagents used are shown in Table 21 and were made up in 1.5ml tubes and incubated for two hours at 37ºC. Due to the different plasmid orientation in the vector different RNA polymerases were used and are shown Table 22.

Table 21: Reagents used for transcription and DIG labelling

Reagent	Volume
Linearized probe	14μl
10x transcription buffer	2μl
DIG labelling mix (Roche)	2μl
RNA polymerase	2μl

Table 22: Polymerases used for each of the probes

Plasmid	RNA polymerase
KIF22	SP6
ALDOA	T7
HIRIP3	SP6
MAZ	SP6
PAGR1	T7
SPN	T7

Following incubation 2μl of DNAse1, RNAse free (10units/μl) was added and the mixture incubated at 37ºC for 15 minutes. Then 2μl of 0.2M EDTA (pH8) was added to stop the reaction. Then 2.5μl of 4M LiCl and 75μl ice cold EtOH were added and the probe left at -20ºC overnight to precipitate.

Next day the probe was spun at 13000rpm at room temperature using a tabletop centrifuge (used throughout this protocol) for 15 minutes and then supernatant removed. 100μl of RT 70% EtOH was added and the tube spun for another 15 minutes. All supernatant was carefully removed, and the tube covered with pierced parafilm and allowed to dry in a fume hood for 1-2 hours.

The pellet was then resuspended in 50μl sterile ddH_2O and 2μl of the probe run for 20 minutes on a 1.5% agarose gel and quickly imaged to prevent RNA degredation to confirm successful transcription. Probe was then stored at -20ºC until needed.

2.2.3 *In situ* hybridisation staining

Frozen cryostat tissue collected from RIE was used for all *in situ* hybridisation stains.

2.2.3.1 Hybridisation

Three sections, rostral, middle, caudal, were used for each age. Preliminary optimisation runs included *PAX6* as a positive control (which has been previously validated by the lab). A no probe control was used as a negative control during optimisation. Slides were defrosted and allowed to stand at RT for 15 minutes and then were placed into a 37ºC oven for 4 hours to prevent tissue falling off prior to hybridisation.

A sealed plastic box was lined with two sheets of white Tork roll which was then soaked in 100ml of 50% formamide/1x Salt. Reagents for 10x salt are shown in Table 23 and 1x salt was made from diluting 10x stock solution. The box was then incubated at 65ºC until required for hybridisation

Table 23: Reagents for 10x salt solution

Reagent	Amount
NaCl	22.8g
Tris HCl (pH7.5)	2.8g
$NaH_2PO_4 \bullet 2H_2O$	1.56g
NaH_2PO_4	1.42g
0.5M EDTA (pH8)	20ml
ddH_2O	Make up to 200ml
Autoclave and store	

The probes were thawed quickly and diluted in hybridisation mix (reagents for hybridisation mix shown in Table 24) using 1.5ml RNAse free screw cap tubes. Optimisation was required for each probe with test concentrations of 1:1000, 1:2000. 1:4000 and 1:8000 used to find optimum concentration. Hybridisation mix and probes were vortexed vigorously and the diluted probes denatured at 95ºC for 10 minutes using a hot block before being returned to ice.

A hydrophobic pen was used to delineate sections, probe mix was vortexed and 300µl added to each slide and carefully distributed across the tissue sections using a p200 pipette tip (note no coverslips were used to minimise tissue damage). Slides were then placed in the pre-heated hybridisation box and incubated at 65ºC overnight.

Table 24: Reagents for hybridisation mix

Reagent	Amount
10x Salt	1ml
Deionized formamide	5ml
50% dextran sulfate	2ml
tRNA (50mg/ml)	200µl (tRNA denatures at 85ºC for 10 mins before use)
100x Denhardts	100µl
RNAse free water	To total volume of 10ml

2.2.3.2 Post hybridisation washes

Following hybridisation, slides were transferred from hybridisation box to a glass colpin jar where they were washed in wash buffer (pre-warmed to 65ºC), reagents shown in Table 25, for 30 minutes at 65ºC. Washes were repeated another 2 x 30 minutes at 65ºC with fresh wash buffer used each time.

Table 25: Reagents for wash buffer

Reagent	Amount
20x SCC	15ml
50% formamide	150ml
0.1% Tween-20	300µl
ddH_2O	135ml

Next 3x 30 minutes washes were performed with 1x MABT (reagents for buffer in Table 26 at RT on a rocker.

Table 26: Reagents for 1x MABT

Reagent	Amount
ddH_2O	600ml
0.5M Maleic acid	11.64g
NaOH	pH to 7.5 after addition of Maleic acid
0.75M NaCl	8.76g
0.5% Tween 20	1ml

2.2.3.3 Blocking and antibody staining

A box was lined with Tork roll soaked with ddH_2O to ensure a humidified environment. Slides were removed from MABT and a hydrophobic pen used to re-delineate the edges of sections as much of the pen was lost during hybridisation. The sections were blocked using 200µl blocking solution (see Table 27) per slide and incubated in the humidified box at RT for 1 hour.

Table 27: Reagents for blocking solution

Reagent	Amount
Heat inactivated sheep serum	1ml
1x MABT	4ml
Blocking reagent	0.1g

Following blocking 200µl of anti-DIG antibody diluted 1:1500 in blocking solution was added to each slide to detect RNA expression. Slides were returned to the humidified box and incubated at 4ºC overnight.

2.2.3.4 Post antibody washes and colour reaction

The next day slides were moved to a glass colpin jar and washed 5x 20 minutes in 1x MABT at RT on a rocker. Sections were then equilibrated in pre-staining buffer (Table 28) for 2x 10 minutes at room temperature on a rocker.

Table 28: Reagents for pre-stain buffer

Reagent	**Amount**
5M NaCl	6ml
1M $MgCl_2$	15ml
1M Tris (pH 9.5)	30ml
0.1% Tween-20	300µl
ddH_2O	250ml

Following equilibration of slides staining buffer was used to visualise RNA expression. Staining buffer reagents are shown in Table 29, with the nitro-blue tetrazolium and 5-bromo-4-chloro-3”-indolyphosphate (NBT/BCIP) added directly before reaction. Staining reaction was incubated in the dark in the humidified box at RT until colour developed (different development times were required for different probes ranging from 6 hours to 4 days).

Table 29: Table of reagents for staining buffer

Reagent	Amount
5M NaCl	0.2ml
1M Tris (pH 9.5)	1ml
ddH_2O	8.3ml
PVA	0.9g
Heat to 85ºC with constant stirring, when dissolved allow to cool	
1M $MgCl_2$	0.45ml
0.1% Tween-20	9µl
NBT/BCIP	180µl

After the colour developed, slides were transferred to a colpin jar and washed many times in 1x PBS to stop the colour reaction and left overnight in 1x PBS at 4ºC. Slides where then mounted with coverslips using Aquatex mounting medium, allowed to dry and stored at 4ºC until imaging.

2.2.4 Imaging of *in situ* hybridisations

Imaging of *in situ* hybridisation was performed using a Zeiss Axioscan z.1 slide scanner, access to which was kindly provided by the Nolan lab. Slides were cleaned and loaded into the scanner trays and imaging was calibrated using the ZEN software provided by Zeiss. ROI was delineated manually around each section and the focal plane chosen for each section. Scanning was performed at x20 objective allowing for single cell resolution and tiling was performed by the ZEN software. Images were viewed and exported from the ZEN software. Individual regions at high magnification were selected and stitched together and saturation adjusted using Abode Photoshop.

2.3 Immunohistochemistry

For immunohistochemistry all sections used were paraffin sections provided by the HDBR except the PAX6 immunofluoresence which used cryostat sections.

2.3.1 Immunofluorescence

2.3.1.1 Dewaxing and antigen retrieval of sections

Sections were dewaxed by immersion for 15 minutes in 2 changes of xylene. They were then rehydrated with 5 minutes in each of the following: 2 changes of 100% IMS, 95% IMS, 90% IMS, 70% IMS and then placed in a glass colpin jar under running water for 10 minutes.

Antigen retrieval was used by washing sections for 10 minutes in 10mM sodium citrate (diluted from 100mM stock, see Table 30 before being heated in a microwave for 5 minutes on high, topped up with ddH_2O to ensure buffer concentration was not altered by evaporation and then heated for another 5 minutes on high and topped up again. The slides were kept in solution and placed on ice to cool for a minimum of half an hour.

Table 30: Reagents for 100mM stock sodium citrate solution

Reagent	**Amount**
Sodium citrate	29.4g
ddH_2O	Up to 1 litre
NaOH	To pH6

Sections were then washed 3x 15 minutes in 1x TBS (diluted from 10x stock solution of TBS pH7.6 see Table 31 at RT on a rocker.

Table 31: Reagents for 10x TBS stock solution

Reagent	Blocking
NaCl	87.6g
Trizma base	60.5g
ddH_2O	Up to 1litre volume
Conc HCl	pH to 7.6

2.3.1.2 Blocking and antibody incubation

A hydrophobic pen was used to delineate sections and then sections were incubated with 200μl of blocking solution comprising 20% heat-inactivated serum in TBS for 1 hour at RT in the previously described humidified box. Next sections were incubated with primary antibody diluted to the required concentration in blocking solution overnight at 4ºC in the humidified box.

2.3.1.3 Post antibody washes and secondary incubation

Next day sections were washed 3x15 minutes in TBS at RT on a rocker. Sections were then incubated for 1 hour at RT in the dark humidified box with 200μl of the appropriate fluorescent secondary antibody diluted 1:200 in blocking solution.
From this point on all sections were kept in the dark by wrapping colpin jars in foil. Sections were washed 3x 15 minutes at RT on a rocker in TBS. Next the sections were incubated with 200μl of DAPI diluted 1:1000 in ddH_2O for 10 minutes at RT in the humidified box. Sections were then washed 1x 15 minutes in PBS before being mounted in H1400 Vectashield hardset mounting medium, sealed with nail varnish and kept at 4ºC until imaging.

2.3.1.4 Imaging of sections

Sections were imaged using a Leica DM5500B epifluorescence microscope with a DFC360FX camera. Confocal images were obtained using a Nikon AIR FILM microscope and analysed in ImageJ. Details of imaging and analysis are provided in the relevant chapter.

2.3.2 Immunocytochemistry

2.3.2.1 Dewaxing, peroxidase blocking and antigen retrieval of sections

Sections were dewaxed by immersion for 15 minutes in two changes of xylene, and then 5 minutes in two changes of 100% IMS. To ensure blocking of endogenous peroxidases sections were transferred to a glass colpin jar and incubated with 3% H_2O_2 in methanol at RT for 10 minutes. Sections were then washed in running tap water for 10 minutes.

Antigen retrieval was used by washing sections for 10 minutes in 10mM sodium citrate (diluted from 100mM stock) before being heated in a microwave for 5 minutes on high, topped up with ddH_2O to ensure buffer concentration was not altered by evaporation and then heated for another 5 minutes on high and topped up again. The slides were kept in solution and placed on ice to cool for a minimum of half an hour.

Sections were then washed in 1x TBS for 3x 15 minutes at RT on a rocker.

2.3.2.2 Blocking and antibody incubation

Sections were delineated with a hydrophobic pen and then incubated with 200µl blocking solution (10% heat inactivated serum diluted in 1x TBS) in a humidified box at RT for 1 hour. Next sections were incubated with primary antibody diluted to the required concentration in blocking solution overnight at 4ºC in the humidified box.

2.3.2.3 Post antibody washes, secondary antibody incubation and amplification

Next day sections were transferred to a glass colpin jar and washed 3x 15 minutes at RT on a rocker in 1x TBS. Next sections were incubated with 200µl of the appropriate biotinylated secondary antibody diluted 1:200 in blocking solution for 1 hour at RT in the humidified box. While this incubation was performed, Avidin-Biotin Complex (ABC) solution was made up in 1x TBS at least 30 minutes prior to use to allow the complex to form.

After blocking sections were moved to glass colpin jar and washed 3x 15 minutes in 1x TBS at RT on a rocker. Next each section was incubated with 200µl of the ABC solution in the humidified box at RT for 30 minutes. Next sections were moved back to the glass colpin jar and washed 3x 15 minutes in TBS at RT on a rocker.

2.3.2.4 Colour reaction

Colour was developed using the 3,3'-Diaminobenzidine (DAB) kit. The solution was made up using ddH_2O and pipetted onto slides. Due to the low concentration of antibody used development took longer than expected and sections were incubated with the solution for 30 minutes at RT until colour

developed. When staining a range of ages sections were all incubated for the same amount of time.

After colour developed sections were washed in running tap water for at least 10 minutes.

2.3.2.5 Counterstaining with NFR

To enable visualisation of cytoarchitecture sections were then counterstained with Nuclear Fast Red (NFR) (vector labs). NFR solution was kept in the dark from day of opening and filtered before use. NFR was applied to sections for 30 sections at RT and then washed off in running water until only nuclear staining remained.

2.3.2.6 Mounting of sections

Sections were then dehydrated by immersion for 1.5 minutes in each of the following concentrations: 70% IMS, 90% IMS, 95% IMS, 2x changes of 100% IMS and 2 changes of Xylene. Sections were then mounted using DPX mountant and allowed to dry. Slides were then stored at RT until imaging.

2.3.2.7 Imaging of sections

Sections were imaged using a Leica DMNB microscope with an attached Leica DFC480 camera and analysed in ImageJ. Specific details of analysis method are provided in the relevant chapter.

3 Identification of progenitor-enriched *16p11.2* transcripts by bioinformatics analysis and *in situ* hybridisation

3.1 Introduction

As described in the introductory chapter one (1.8), there are 29 protein-coding genes within the *16p11.2* locus. This thesis focuses on the neurogenesis hypothesis of ASD, the idea that changes during neural development impact cortical architecture and function leading to manifestation of ASD in some patients in early life. This is strongly evidenced by the mirrored head size phenotype of many *16p11.2* patients, macrocephaly in the deletion and microcephaly in duplication patients. Changes to head size are a proxy for changes to brain growth, and changes to head size are one of the most replicated findings when investigating the neuropathology of ASD (Courchesne, Carper and Akshoomoff, 2003; Amaral, Schumann and Nordahl, 2008; Sacco, Gabriele and Persico, 2015). This suggests that final cell numbers are altered and disruptions to the balance of proliferation and differentiation early in neurodevelopment is a potential cause. Therefore, studying the effects of the *16p11.2* CNV, the most frequent aetiology of ASD, and the roles of each of its genes in neurogenesis is a particularly important avenue of research.

Very little is known about the roles of many of the *16p11.2* locus genes in neurogenesis, especially in the developing human cerebral cortex. It is unknown which of the proteins produced by the *16p11.2* locus genes are expressed by progenitor cells in the developing human cerebral cortex and are therefore candidates for regulating neurogenesis, and the potential disruptive effects on neurogenesis when their level is altered as a result of the *16p11.2* CNV.

Here we aim to identify candidate genes from the locus: genes which will potentially specifically affect neurogenesis in the developing human cerebral cortex, by identifying *16p11.2* locus genes which are highly expressed in cerebral cortex progenitors and are downregulated as cells become post-mitotic. Genes that are expressed in progenitors and then switched off (or

downregulated) as cells exit proliferation are likely to have important proliferative functions and therefore are ideal candidates for our study.

To identify candidate genes we made use of a previously published single cell RNA-sequencing (scRNA-seq) dataset of the germinative VZ and SVZ of the 16-18GW human cerebral cortex (Pollen *et al.*, 2015). This provided us with progenitor-enriched candidate transcripts encoded by the *16p11.2* locus for investigation.

3.1.1 Chapter aims.

The aims of this chapter were as follows:

1) Use bioinformatics analysis of sc-RNAseq data (Pollen *et al.*, 2015) from the developing human cerebral cortex to perform an unbiased screen and identify *16p11.2* locus transcripts enriched in progenitors, and ideally down-regulated in post-mitotic cells. This analysis was performed by Yifei Yang.

2) *In vivo* validation of identified progenitor enriched transcripts by *in situ* hybridisation on sections of human fetal cortex.

3) By combining the results of these two approaches we will select candidates for further analysis.

3.2 Materials and methods

All bioinformatics analysis was performed by fellow PhD student Yifei Yang.

3.2.1 scRNA-seq dataset

The scRNA-seq dataset of 16-18GW human cortical tissue was generated and published by Pollen et al in 2015 (Pollen *et al.*, 2015). Briefly the VZ and SVZ of 16-18GW (three brains were used in total) human cerebral cortex samples were microdissected, dissociated and sequenced with data from 393 cells included in the final published dataset.

3.2.2 scRNA-seq analysis

The Pollen et al (Pollen *et al.*, 2015) dataset was used to identify candidate genes. Prior to dataset publication the reads were aligned, we normalized the RPKMs as log(x+1) to give zero reads a meaningful value for log transformation. Analysis was performed by Yifei Yang using R studio. To determine genes with significant changes a Wilcox test by *FindAllMarkers* in Seurat package was used. Monocle2 R package was used to order cells in pseudotime. To identify cell-cycle phase specific transcripts we used function CellCycleScoring from Serurat R package.

3.2.3 Fetal samples used

For the *in situ* hybridisation the cryostat tissue obtained from The Royal Edinburgh Infirmary (RIE) was prepared as described in 2.1.2, with the exception of the 12pcw sample which was obtained from HDBR as a PFA fixed brain and processed in the same way as RIE samples. One sample of each age was used, and the ages of tissue were used for the *in situ* hybridisations shown in this chapter are explained in Table 32. The rostral

region of the 12pcw sample was also used for the PAX6 immunofluorescence.

Table 32: Table of ages used for this chapter

Age	Rostral	Middle	Caudal
12pcw	✓	✓	✓
14pcw	✓	✓	✓
15pcw	✓	✓	✓
16pcw	Damaged & unusable	✓	✓

3.2.4 Immunofluorescence of PAX6

The immunofluorescence stain of PAX6 was performed in the same way as described in 2.3.1, however it was performed on cryostat tissue rather than paraffin. Therefore, we made the slight change to heat the tissue at 37ºC for several hours before proceeding the protocol and the dewaxing in xylene step was not required. The PAX6 antibody used here was the rabbit polyclonal PAX6 antibody purchased from Biolegend used at a concentration of 1:500 with an Alexa Fluor Goat anti-rabbit 488 secondary antibody. Imaging was performed with the Nikon confocal microscope as described in 2.2.4.

3.3 Results

3.3.1 Identification of genes expressed in progenitors

We used the scRNA-seq dataset to identify candidate genes of interest from the *16p11.2* locus whose transcripts are enriched in progenitors.

3.3.1.1 Dimensional reduction to cluster cells

First, we performed dimensional reduction of the scRNA-seq dataset to separate the 393 cells into three distinct classes based on their transcriptome similarity. This is shown in Figure 18, the three classes were subsequently identified as the three cardinal cell classes: progenitors, post-mitotic/principal cells and interneurons.

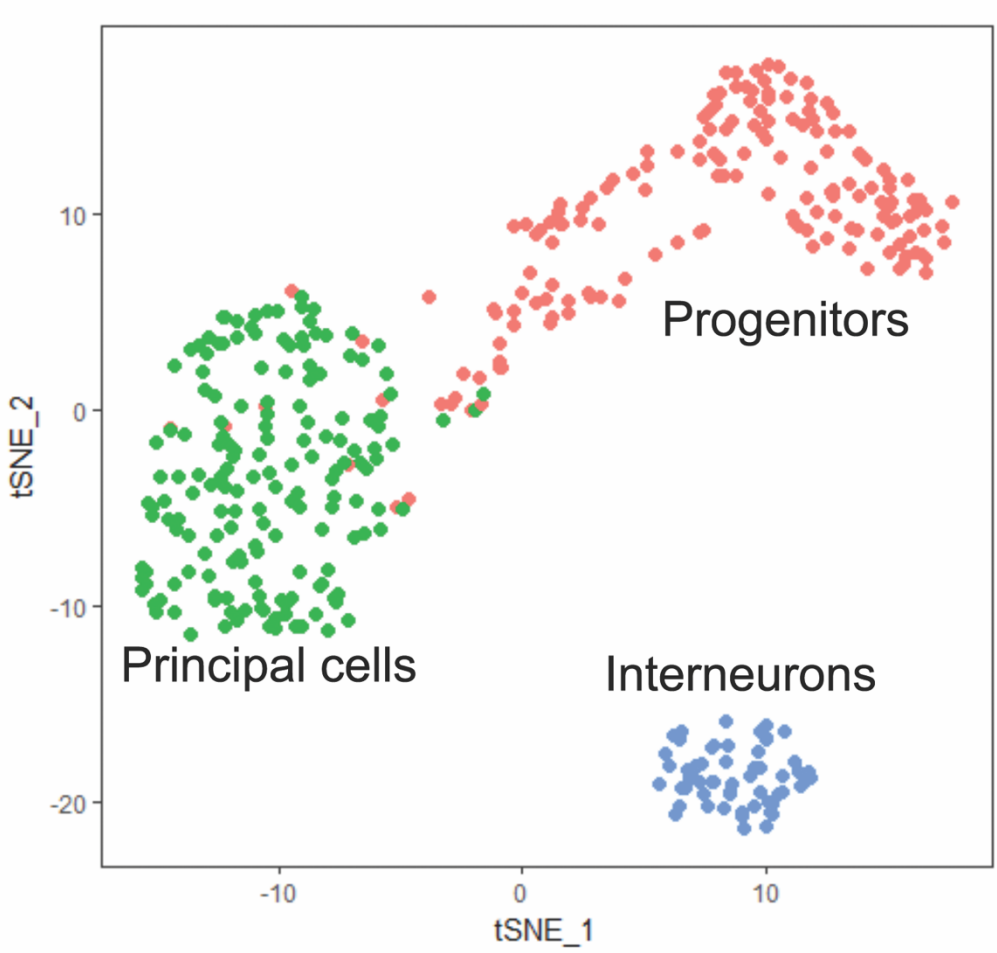


Figure 18: tSNE plot separating cells into three clusters

Dimensional reduction allowed classification of each cell into one of three classes: progenitors = red, principal cells = green and interneurons = blue. Each dot on the tSNE plot represents an individual cell.

The following genes were used to identify the three cardinal cell classes. They are shown overlaid on the tSNE plot for progenitors (Figure 19), principal cells (Figure 20) and interneurons (Figure 21).

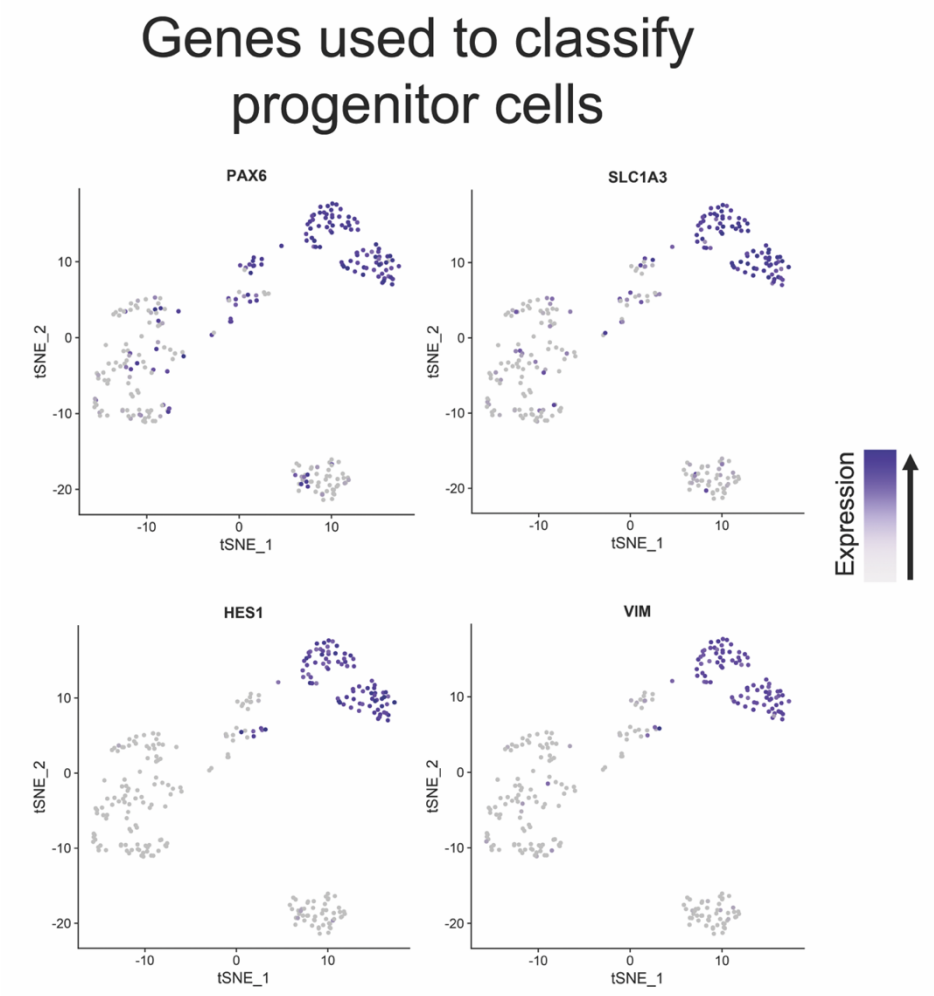


Figure 19: The four genes used to identify progenitor cells overlaid to the tSNE plot

Four genes were used to identify progenitor cells: *PAX6, SLC1A3, HES1* and *VIM*. Each dot represents an individual cell and highest expression is indicated by dark purple, cells with no expression are grey. *PAX6* is a transcription factor expressed by neural progenitor cells. *SLC1A3* expression is downregulated by basal progenitors. *HES1* is expressed in neural progenitors and its absence accelerates neurogenesis (Kageyama, Ohtsuka and Kobayashi, 2008). *VIM* encodes for an intermediate filament protein.

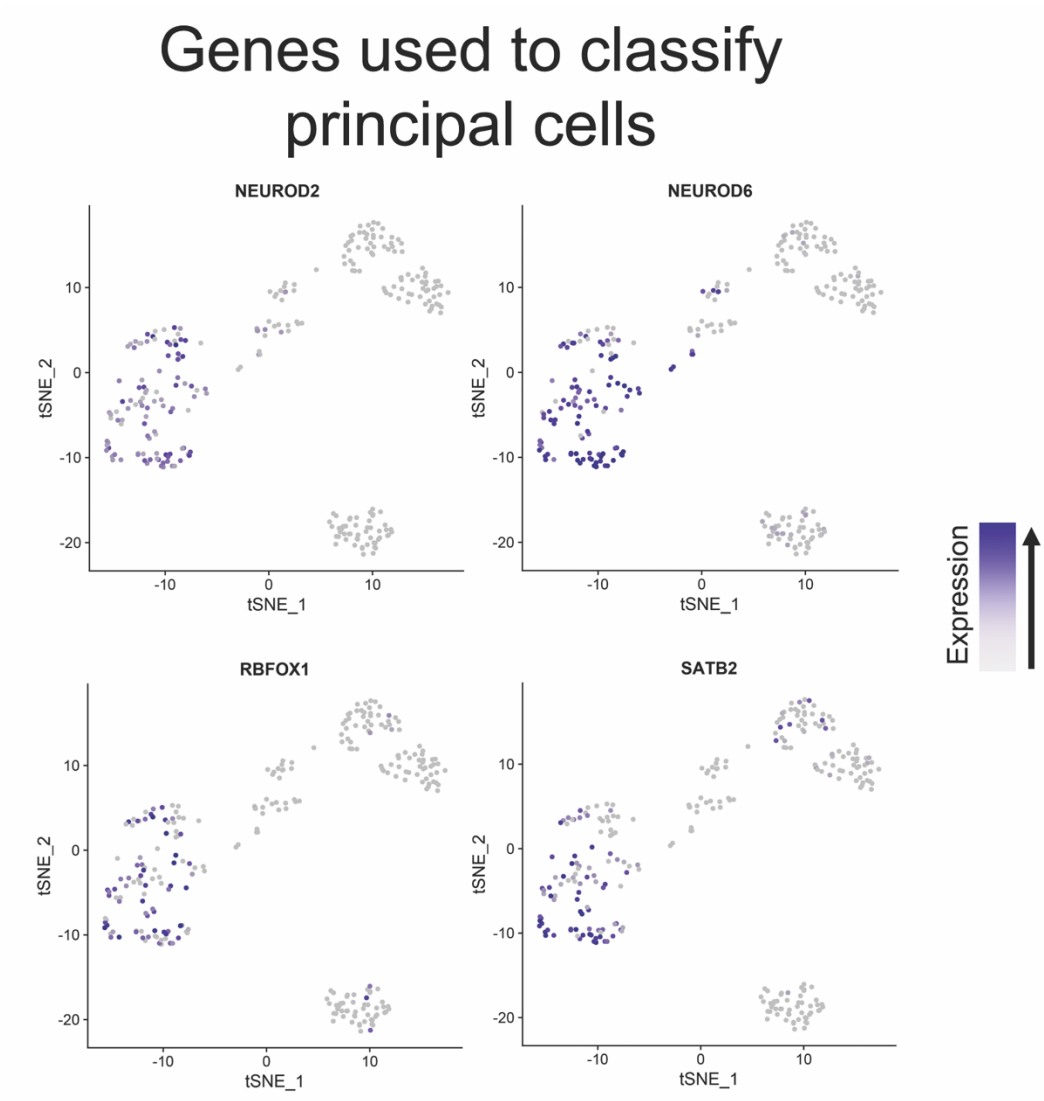


Figure 20: The four genes used to identify principal cells overlaid to the tSNE plot

Four genes were used to identify principal cells: *NEUROD2, NEUROD6, RBFOX1* and *SATB2*. Each dot represents an individual cell and highest expression is indicated by dark purple, cells with no expression are grey. *NEUROD2* is expressed in cortical projection neurons during cortical neurogenesis. *NEUROD6* is involved in neuronal differentiation. *RBFOX1* is expressed in neurons and heart tissue. *SATB2* is expressed in upper layer neurons.

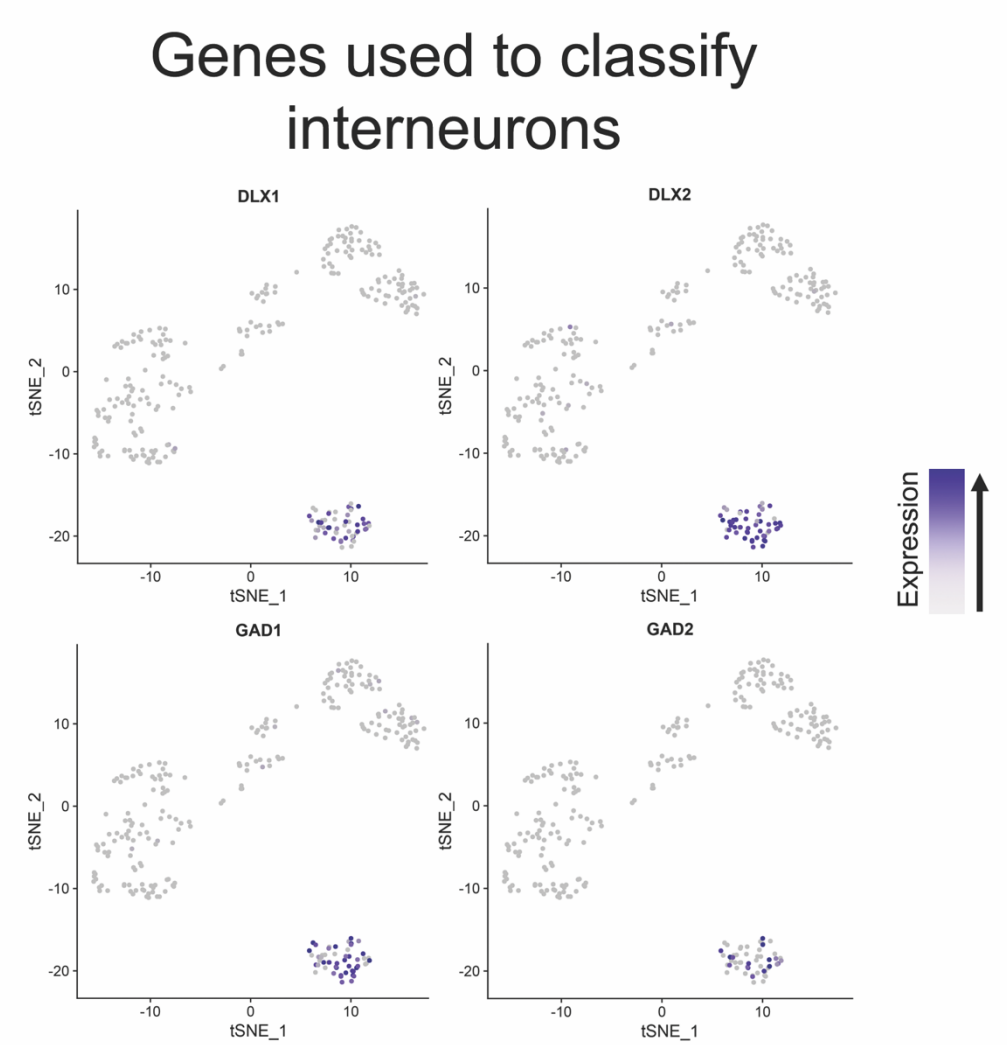


Figure 21: The four genes used to identify interneurons overlaid to the tSNE plot

Four genes were used to identify interneuron cells: *DLX1, DLX2, GAD1* and GAD2 Each dot represents an individual cell and highest expression is indicated by dark purple, cells with no expression are grey. *DLX1* and *DLX2* promote GABA-ergic interneuron development. *GAD1* and *GAD2* are expressed by GABA-egric interneurons.

3.3.1.2 Pseudotime analysis of gene expression

Next, we wanted to identify how the expression pattern of each of the *16p11.2* locus genes changed in pseudotime as cells changed from progenitors towards postmitotic cells. Here, pseudotime acts to provide us a relative measure of transcript expression level at different “time-points” through the process of progenitor cells differentiating into post-mitotic cells (Campbell and Yau, 2018). This technique is particularly useful for our work as it allows us to recapitulate temporal dynamics of transcript levels through changing cell fates despite only having access to tissue from one time point.

We used the monocle2 R package to order the cells in pseudotime using the normalized expression of levels of selected differentially expressed genes (DEGs) as input to order the cells (Trapnell *et al.*, 2014; Qiu *et al.*, 2017) moving from the progenitor state (left) to the post-mitotic stage (right). We plotted the average expression of each *16p11.2* transcript at each pseudotime point on the Y axis and the result is shown in Figure 22.

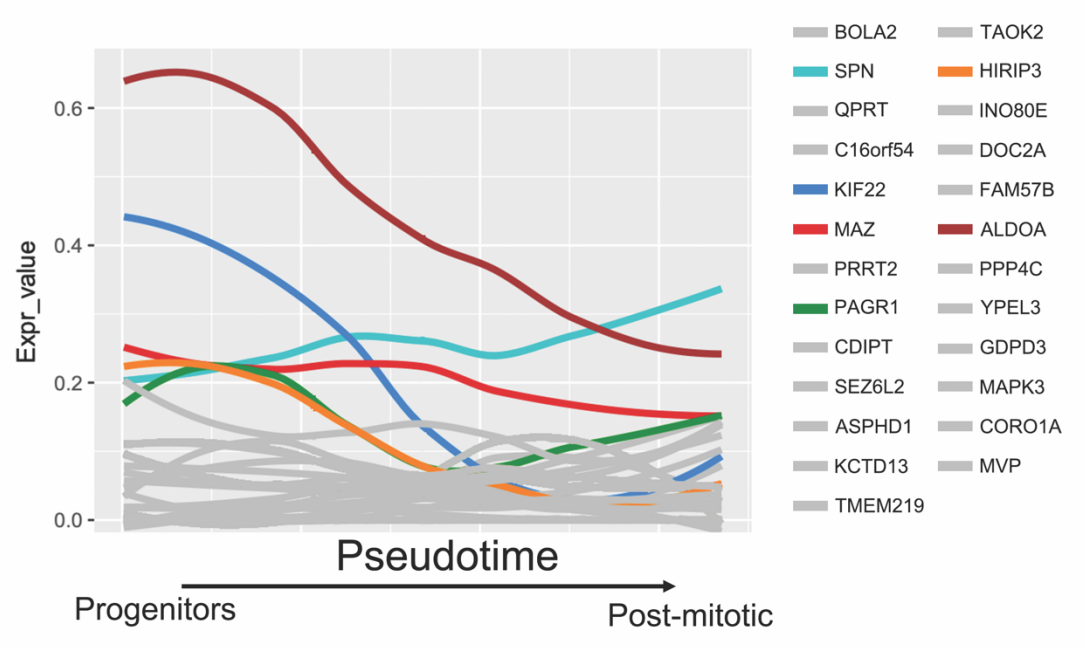


Figure 22: Changing mRNA expression levels of *16p11.2* genes across pseudotime

Pseudotime analysis showing how mRNA expression levels of each of the *16p11.2* locus genes change as cell state moves through pseudotime from progenitor to post-mitotic. In this figure the six genes which we selected for initial study are shown in colour, with the colour for each gene shown in the key. The other genes are all coloured grey and we did not study them beyond this initial pseudotime analysis

From this pseudotime analysis we identified two genes, *KIF22* (blue line) and *ALDOA* (maroon line) which were notable for having high mRNA expression in progenitors which then declined as cells became post-mitotic. We used a Wilcox test which identified *ALDOA* and *KIF22* as the only *16p11.2* transcripts that were significantly higher in progenitor than neuronal populations ($p<<0.05$). Four other genes, *HIRIP3* (orange line), *MAZ* (red line), *PAGR1* (green line) and *SPN* (turquoise line), were not significantly enriched in progenitors but their transcripts were expressed in progenitors at higher levels than the remaining *16p11.2* transcripts (shown as grey lines), many of which were barely expressed at all.

To obtain a little insight into what the functions of these genes may be in progenitors we looked up their GO terms; Gene Ontology provides information about the functions of the products of these genes. GO terms of these genes are shown in Table 33.

Table 33: GO terms associated with each of the six candidate *16p11.2* locus genes

Gene	GO Term
ALDOA	Fructose-bisphosphate aldolase activity Protein binding Actin binding Cadherin binding
KIF22	DNA binding Microtubule motor activity ATP binding ATPase activity
HIRIP3	Chromatin assembly or disassembly
PAGR1	Protein binding Oestrogen receptor binding Histone H3-K4 methylation (chromatin remodelling)
MAZ	DNA binding Nucleic acid binding
SPN	Heat shock protein binding Transmembrane signalling receptor activity

3.3.2 Expression of the candidate genes in the clustered populations

To add to the findings of the pseudotime analysis, we wanted to visualise the expression levels of each of these genes in the cells of each of the three cardinal cell classes. To do this we mapped the expression levels of each of the genes as shown in Figure 23 onto the previously shown tSNE plot. These tSNE plots have the advantage that they allow us to visualise heterogeneity in transcript levels between single cells which wasn't possible from the pseudotime analysis.

This data showed *ALDOA* and *KIF22* to be expressed at the highest (although variable cell to cell) levels, with *KIF22* restricted almost completely to progenitor cells. *ALDOA* is strongly expressed in progenitors but also expressed at lower levels in the other two cardinal cell classes.
MAZ, HIRIP3 and *PAGR1* appear to be expressed at lower levels and in fewer cells and also their expression is scattered throughout the cell classes, although *HIRIP3* does appear predominantly in the progenitor class. *SPN* expression is also scattered throughout the cell classes, appearing to be expressed in more principal cells and interneurons that any of the other six genes in line with the pseudotime analysis.
For all the genes (although less so for ALDOA), all six of these transcripts are expressed in a subset of progenitors and at varying levels within those progenitors.

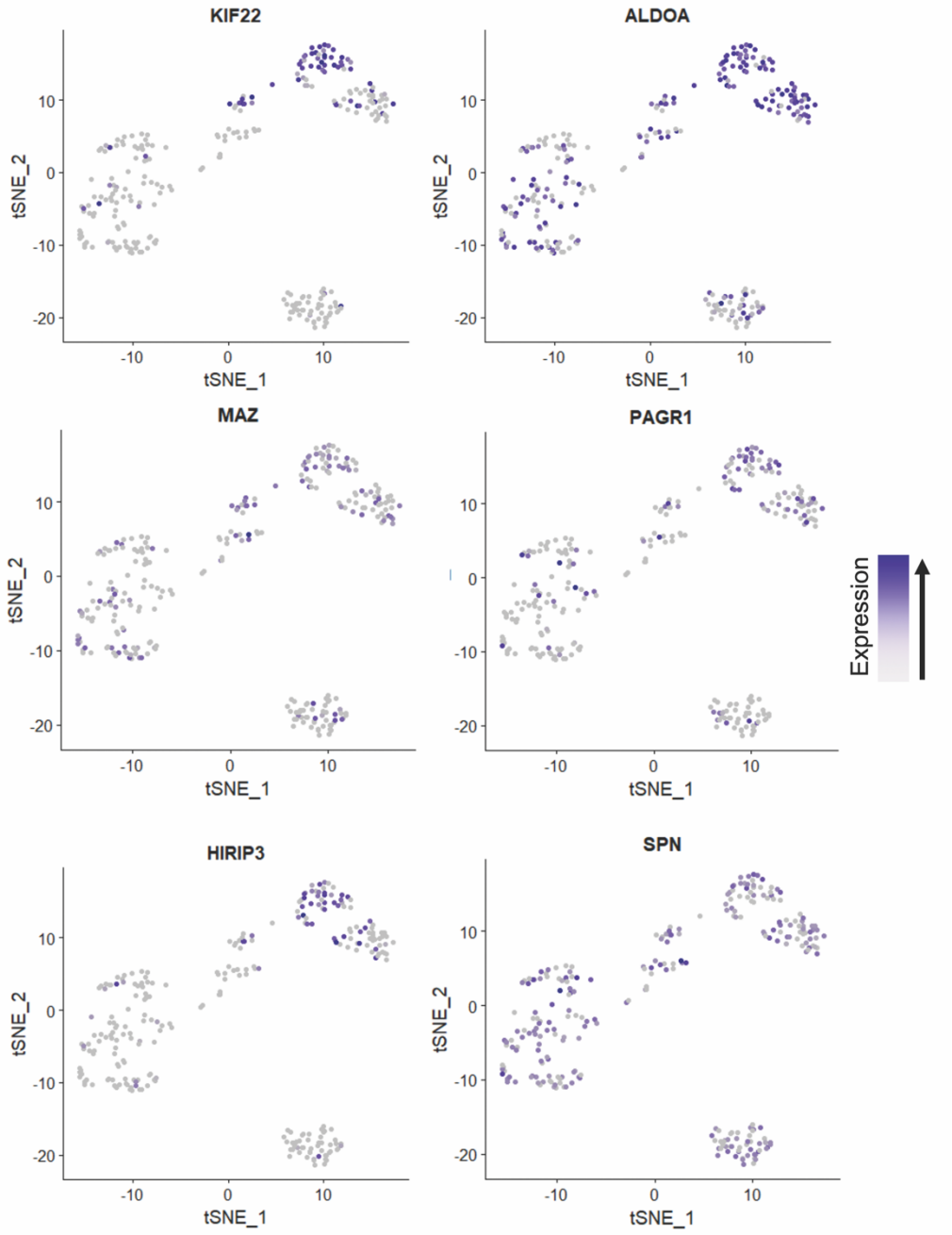


Figure 23: Gradient plots of the six candidate genes

The gradient plots show the expression levels of each of the six candidate genes overlaid to the tSNE plot indicating the expression level of the gene in each cell; grey is lowest and dark purple is highest. Each dot represents an individual cell.

With our initial scRNA-seq analysis we have identified two highly promising candidate genes and four other potentially interesting genes. To build on this foundation we wanted to next take a complementary *in situ* hybridisation approach. This allowed us to visualise the transcripts identified; gaining more information about their spatial expression *in vivo* and to validate the results of our scRNA-seq experiment by an independent method.

3.3.3 mRNA expression of the six candidate genes using *in situ* hybridisation of human fetal cortex sections

To build on our bioinformatics foundations, our next step was to visualise the expression of our six candidate genes *in vivo* by performing *in situ* hybridisation for mRNA of each of these genes in sections of human fetal cerebral cortex at several ages. Due to the limited previous research into these genes in the brain, particularly the human brain, it was necessary to design probes from scratch (see 2.2.2 for details of cloning) as there were none publicly available.

3.3.3.1 Design and sequencing of *in situ* probes

This section will briefly describe the genomic regions identified for use in the *in situ* probes. For all six probes we used human RNA which was converted to cDNA (except *KIF22* and *ALDOA* see 2.2.2). We designed primer pairs to amplify sequences selected from the cDNA by PCR and the amplified PCR products were then cloned into a pGEMT easy vector and their identity confirmed by sequencing and BLAST. For each gene we ensured the probes were designed to detect all known splice variants annotated on ENSEMBLE.

3.3.3.1.1 *KIF22*

KIF22 has three transcript variants, the structure of the three variants and the region common to all variants which was used for the probes are shown in Figure 24.

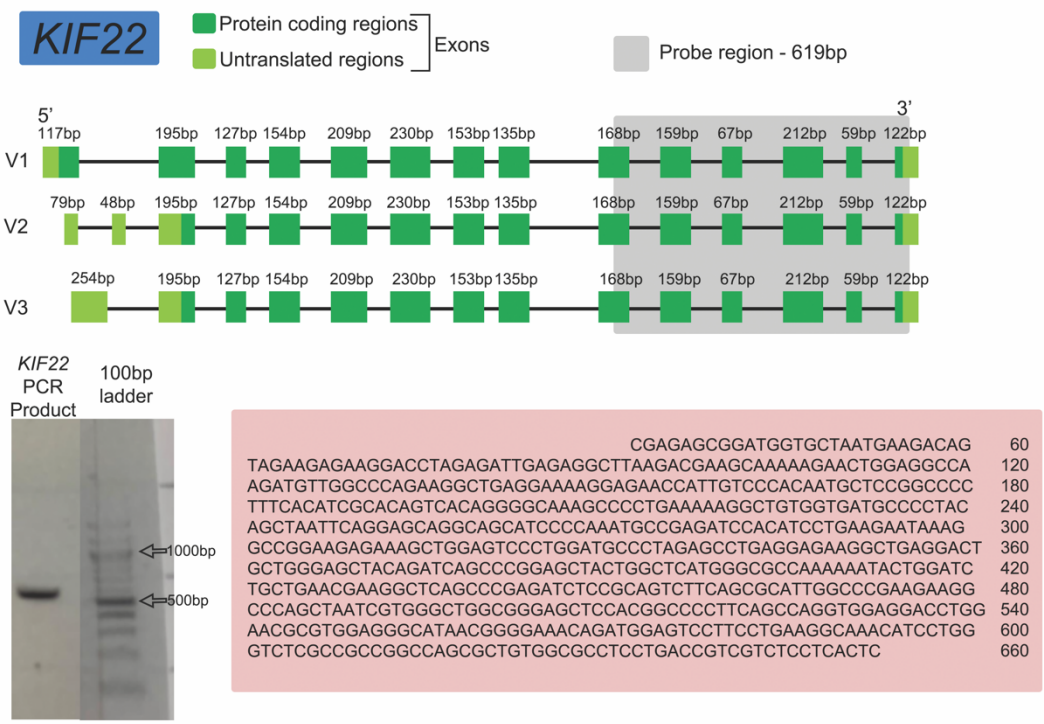


Figure 24: Gene structure of KIF22 and region used for probe

Schematic showing the three variants of the *KIF22* gene and the 619bp region spanning 1443-2062 of the KIF22 splice variant one region. Exons are shown in green: dark green = protein coding, light green = untranslated region (UTR). Black line = introns.

Gel showing *KIF22* PCR fragment estimated to be ~600bp on the left side of an agarose gel, on the right side there is the 100bp ladder run simultaneously with 500bp and 1000bp indicated by arrows.

Pink rectangle contains the sequence of the PCR product which confirmed that the probe made from it would detect all three splice variants.

3.3.3.1.2 *ALDOA*

ALDOA has four transcript variants, the structure of the four variants and the region common to all variants used for the probes are shown in Figure 25.

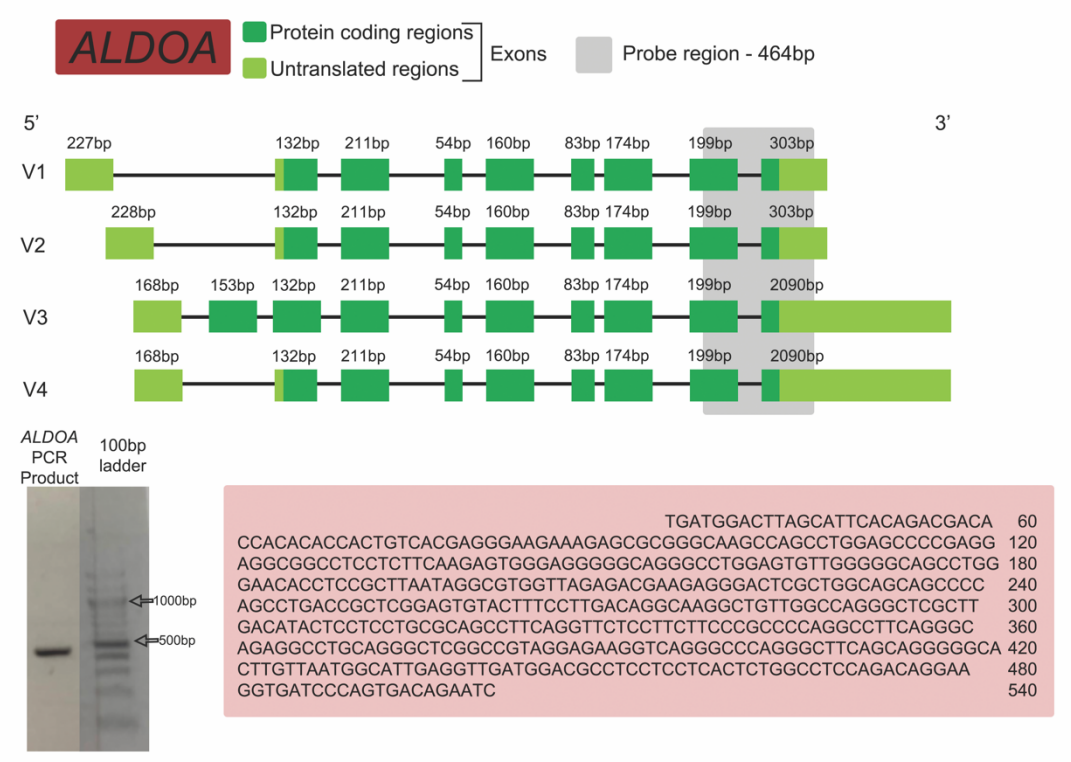


Figure 25: Structure of *ALDOA* gene and probe region

Schematic showing the four variants of the *ALDOA* gene and the 464bp region spanning 1798-2262 of the *ALDOA* splice variant one region. Exons are shown in green: dark green = protein coding, light green = untranslated region (UTR). Black line = introns.

Gel showing *ALDOA* PCR fragment estimated to be <500bp on the left side of an agarose gel, on the right side there is the 100bp ladder run simultaneously with 500bp and 1000bp indicated by arrows.

Pink rectangle contains the sequence of the PCR product which confirmed that the probe made from it would detect all four splice variants.

3.3.3.1.3 *HIRIP3*

HIRIP3 has two known transcript variants, the structure of the two variants and the region common to all variants used for the probes are shown in Figure 26.

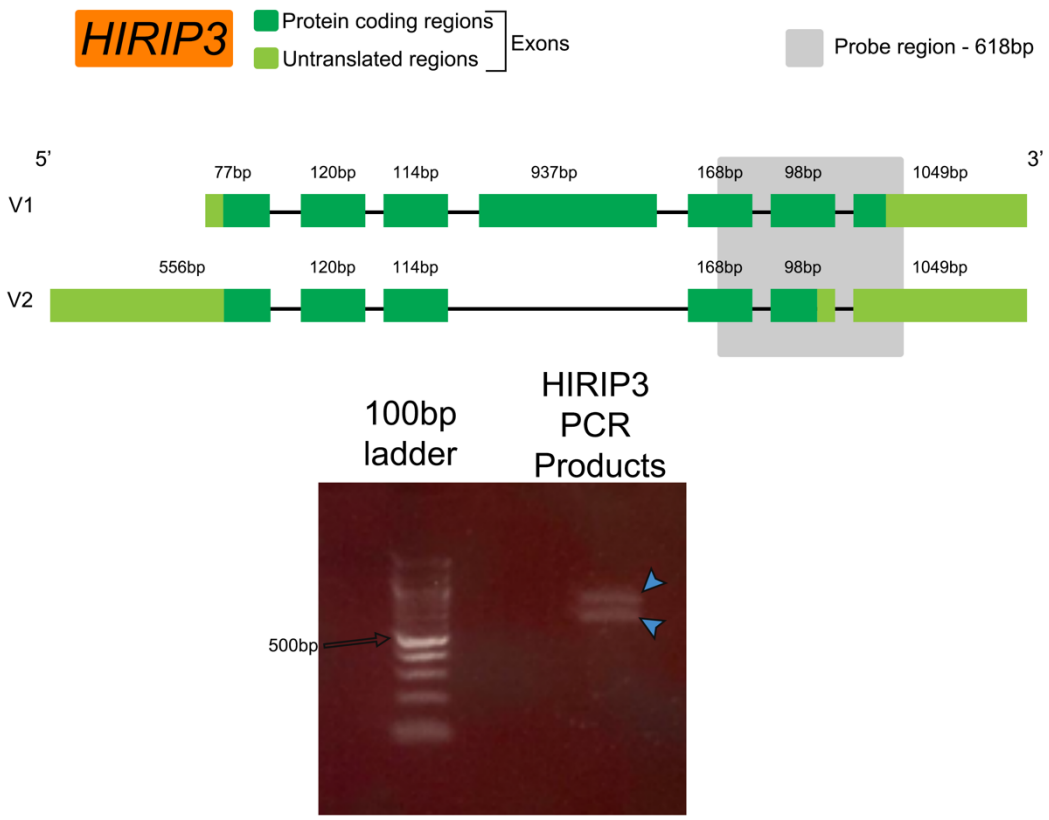


Figure 26: Structure of *HIRIP3* gene and probe region

Schematic showing the two variants of the *HIRIP3* gene and the 618bp region spanning 1978 - 2416 of the *HIRIP3* splice variant one region. Exons are shown in green: dark green = protein coding, light green = untranslated region (UTR). Black line = introns.
Gel showing *HIRIP3* PCR fragments, unlike the other genes, two bands estimated to be between 600-800bp were identified as indicated by the blue arrows on the right side of the gel. On the left side there is the 100bp ladder run simultaneously with 500bp indicated by arrow.

Our primers were carefully designed to amplify a 618bp sequence that was common to both HIRIP3 splice variants. This would ensure that our probe would detect all splice variants of HIRIP3. However, upon running the amplified PCR product on a 1.5% agarose gel we observed two bands as seen in Figure 26 which, based on their position relative to the ladder we estimated one to be 600-700bp and the other to be 700-800bp in length.

This result was unexpected, as we were predicting to only see one band as we had for the other five genes. This result of two bands meant that we could not be sure any probe created from these PCR products would be specific to *HIRIP3* and detect all splice variants. There were two possible explanations for this: it could be due to off target effects, where our primers also amplified an unknown region of a different gene, or the presence of another HIRIP3 splice variant which had not been previously predicted by BLAST.

To confirm which of these possibilities was the case here, we cloned each of the two PCR products into the pGEMT easy vector and sequenced the DNA. When we compared the sequences to BLASTN, our shorter band (estimated to be 600-700bp on the DNA ladder) showed 99% similarity for both *HIRIP3* splice variants one and two as seen in Figure 27 making it the most appropriate PCR product to use for generating the *HIRIP3* probe. The longer sequence only showed to have 77% similarity to the *HIRIP3* splice variants one and two and no similarity to any other genes. It also matched 99% to the chromosome 16 reference genomic sequence confirming that the transcript is encoded by the *HIRIP3* gene as seen in Figure 27. Therefore, we can strongly conclude that we have identified a new *HIRIP3* splice variant that was not predicted by BLAST.

Shorter PCR Product Sequence

```
                                                caaggagcgcctgagtatcctccgggcagaactg     60
gaagcgctaggcatgaagggtacccctccctagggaagtgtcgggccctgaaggagcag                      120
agggaggaggcagctgaggtggcctccttggatgttgcgaacatcatcagtggctcgggc                     180
cggccacgcagacgtacagcctggaacccttaggagaagcagcacccccaggggagctg                      240
taccgacggaccctggactcagatgaagagcggccccgtcccgcacccccagactggtca                     300
catatgcgtggcatcatcagcagtgatggcgagagtaactgagctctgccacccccagga                     360
gggacccttgatacatgtacaaagcatacatagcacccctgccctgtgtctgtggaaca                      420
gaagcagcttccttcagagaagactgcagctcccaaggacacaagctgttgggatgctac                     480
ttctcagcttcacgctgtccctttaaggtgtttatttttttaagactcaataaaggagtgt                     540
ttttaatcacctcatcaaatttggtcccccattctcacctcctgtattttgggccaa                        600
```

Longer PCR Product Sequence

```
                                                gctccacaaggagcgcctgagtatcctccgggcagaactggaa     60
gcgctaggcatgaagggtgaggcccggcgtgccctgggggctgcaggagagggccttcct                    120
gctggggaggagagaatcaccagcttcttctgccccaggtacccctccctagggaagtg                     180
tcgggccctgaaggagcagagggaggaggcagctgaggtggcctccttggatgttgcgaa                    240
catcatcagtggctcgggtaaggggtttcctctcactgcccaggagcagataggtggggt                    300
tggaccttctcagacatctgatctctcgtttccccttcccctaggccggccacgcagac                     360
gtacagcctggaacccttaggagaagcagcacccccaggggagctgtaccgacggaccc                     420
tggactcagatgaagagcggccccgtcccgcacccccagactggtcacatatgcgtggca                    480
tcatcagcagtgatggcgagagtaactgagctctgccacccccaggagggacccttgata                    540
catgtacaaagcatacatagcacccctgccctgtgtctgtggaacagaagcagcttcct                     600
tcagagaagactgcagctcccaaggacacaagctgttgggatgctacttctcagcttcac                    720
gctgtccctttaaggtgtttatttttttaagactcaataaaggagtgtttttaatcacctc                    780
atcaaatttgggtcccccccattctcacctcctgtattttgggccaa                                 840
```

Figure 27: Sequence alignments of *HIRIP3* PCR products

The two *HIRIP3* PCR products visualised to the HIRIP3 genome within the region targeted by the PCR primers. In the top figure the overlapping aligned sequences between the shorter and longer PCR products are shown with a blue background while those that differ are shown with a pink background.

Below, the pink box shows the sequence of the shorter PCR product, which matched to both the reference genome and the two known HIRIP3 splice variants. Based on its high specificity for both known *HIRIP3* splice variants we decided it was the most appropriate of the two PCR products to use to generate the *HIRIP3* probe.

In the green box is the sequence of the longer PCR product, this matched to the reference genome however only 77% of it matched to known HIRIP3 splice variants and to the shorter PCR product. The sequence regions which matched these are shown in blue while the sequence regions which did not match are indicated in red.

3.3.3.1.4 *MAZ*

MAZ has four transcript variants, the structure of the four variants and the region common to all variants used for the probes are shown in Figure 28.

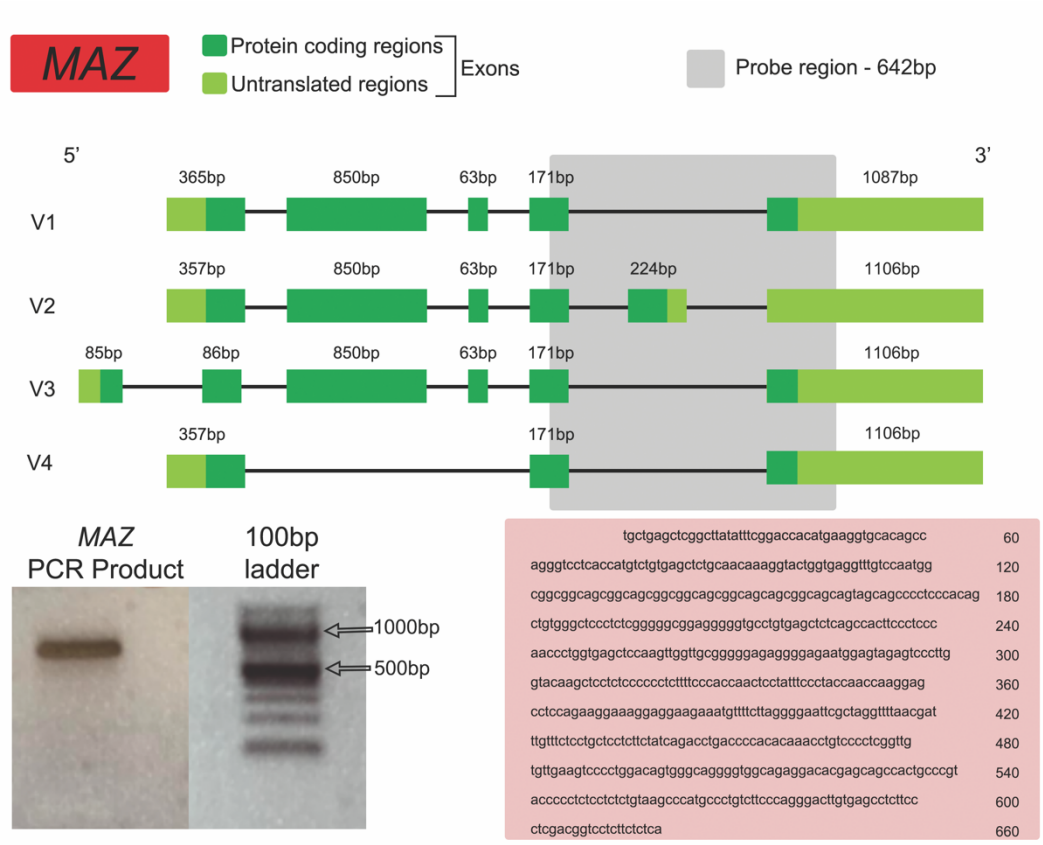


Figure 28:Structure of *MAZ* gene and probe region

Schematic showing the four variants of the *MAZ* gene and the 642bp region spanning 1335 - 1976 of the *MAZ* splice variant one region. Exons are shown in green: dark green = protein coding, light green = untranslated region (UTR). Black line = introns.
Gel showing *MAZ* PCR fragment estimated to be ~600bp on the left side of an agarose gel, on the right side there is the 100bp ladder run simultaneously with 500bp and 1000bp indicated by arrows.
Pink rectangle contains the sequence of the PCR product which confirmed that the probe made from it would detect all four splice variants.

3.3.3.1.5 *PAGR1*

PAGR1 has only one splice variant; its structure and the region used for the probe is shown in Figure 29.

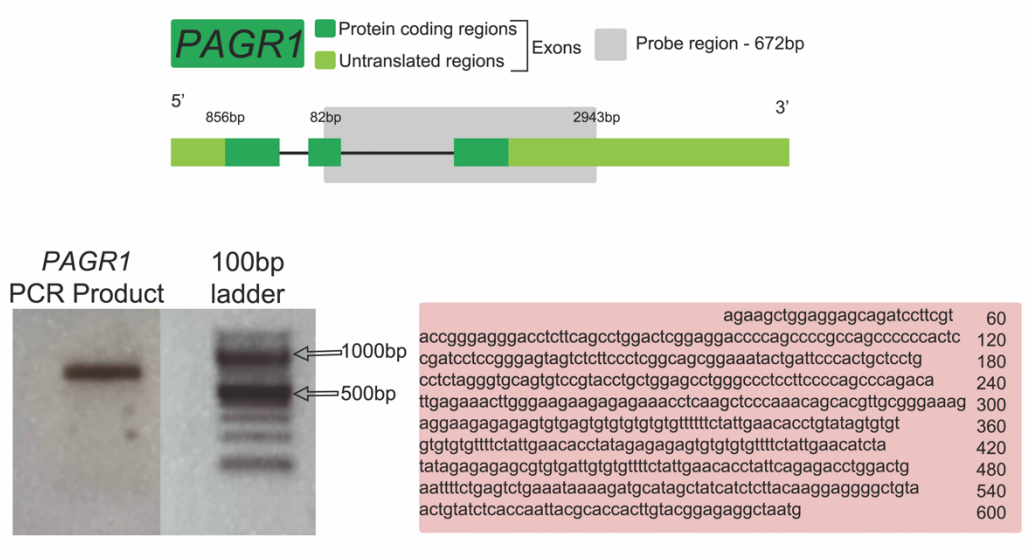


Figure 29: *PAGR1* gene structure and region used for probe

Schematic showing the *PAGR1* gene and the 672bp region spanning 886 – 1557 of the gene. Exons are shown in green: dark green = protein coding, light green = untranslated region (UTR). Black line = introns.

Gel showing *PAGR1* PCR fragment estimated to be ~600bp on the left side of an agarose gel, on the right side there is the 100bp ladder run simultaneously with 500bp and 1000bp indicated by arrows.

Pink rectangle contains the sequence of the PCR product which confirmed that the probe made from it would detect PAGR1 mRNA.

3.3.3.1.6 *SPN*

SPN has two transcript variants, the structure of the two variants and the region common to all variants used for the probes are shown in Figure 30.

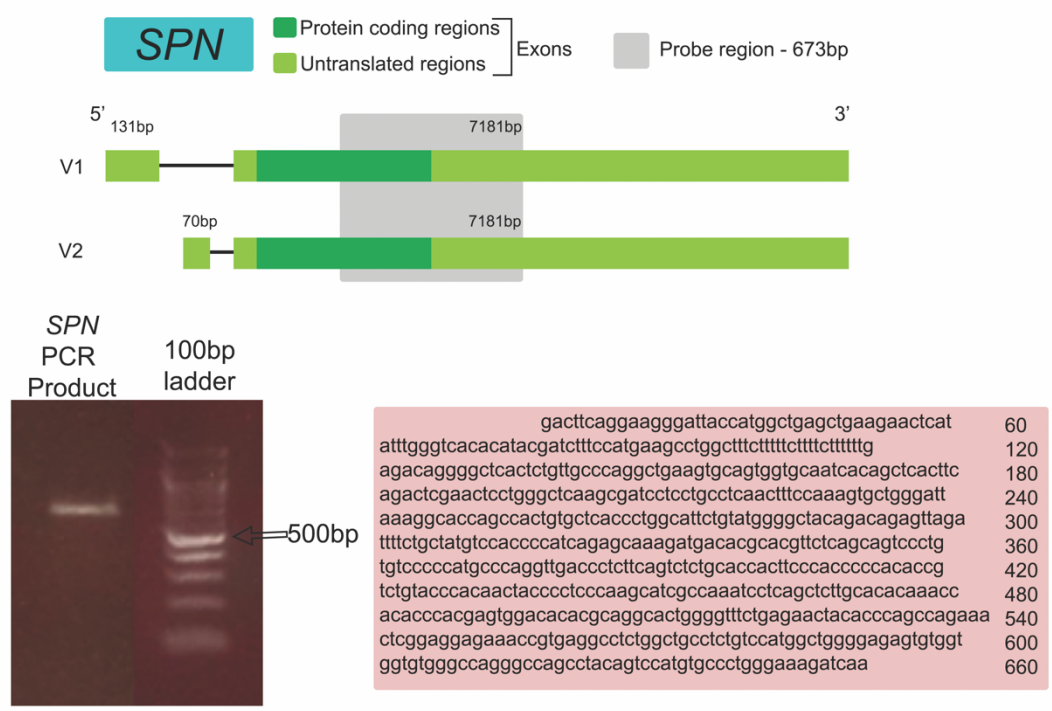


Figure 30: *SPN* gene structure and region used for probe

Schematic showing the *SPN* gene and the 673bp region spanning 2280 - 2952 of the gene. Exons are shown in green: dark green = protein coding, light green = untranslated region (UTR). Black line = introns.

Gel showing *SPN* PCR fragment estimated to be ~600bp on the left side of an agarose gel, on the right side there is the 100bp ladder run simultaneously with 500bp indicated by an arrow.

Pink rectangle contains the sequence of the PCR product which confirmed that the probe made from it would detect both splice variants.

Having confirmed all the cloned regions, it was possible to make DIG-labelled RNA probes for *in situ* hybridisation as detailed in the methods section 2.2.2.14.

3.3.4 Optimisation of processing human cortex tissue

As human fetal cerebral cortex samples had not previously been used in our lab some optimisation of the tissue preparation was required.

To best preserve the morphology of the tissue it was fixed overnight still in the skull, this was determined to be the best method, as not fixing and dissecting at RIE resulted in significantly greater tissue damage, especially during transport method consisting of a half hour bike ride, as shown in Figure 31.

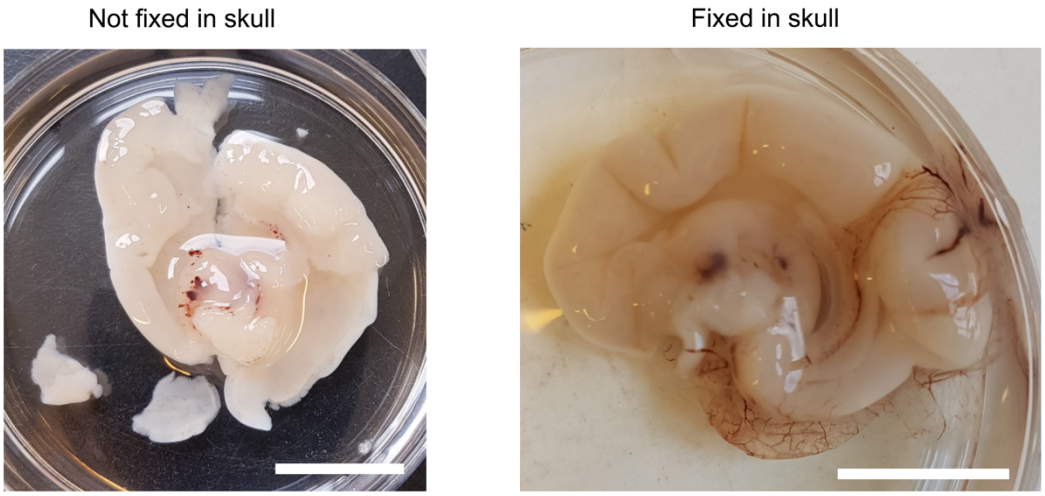


Figure 31: Comparison of fixation techniques for preserving brain morphology

The fixing tissue overnight in the head as shown right, proved to be much better for preserving brain morphology and reducing damage than only fixing brain after removing from skull (left). Scale bar = 2cm.

Due to tissue size, complete fixation of the whole intact brain was not optimal here as it would result in over fixed tissue outside and under-fixed inside. Therefore, the tissue was cut into sections as described in 2.1.2.1 to ensure complete fixation.

3.3.5 *In situ* hybridisation of the six candidate genes

We performed *in situ* hybridisation of our six candidate genes as described in the methods. We used four ages of tissue and one slide from rostral, middle and caudal with the exception of the 16pcw sample where the rostral brain tissue was damaged beyond use during dissection.

3.3.5.1 H&E Staining to determine cytoarchitecture

Prior to performing the *in situ* hybridisation, we first wanted to be able to determine the different compartments of the telencephalic wall based on cellular density. This also had the advantage of, as none of the samples from RIE were karyotyped, allowing us to confirm each sample used had approximately normal brain structure for its age (Bayer and Altman, 2002, 2005). H&E was performed as described in the methods 2.1.3. In Figure 32, Figure 33, Figure 34 and Figure 35 we can see schematics of the region of the brain sections used in part a-a*, and low magnification image of the entire brain sections in part b-b*. On the low magnification we have indicated various anatomical locations of the brain such as the ganglionic eminences and where possible the different lobules. The high magnification images of the cortex were always taken from the ventral telencephalon.

At high power (c-c*) we can determine the ventricular zone as the highly dense pseudostratified layer at the apical edge of the cortical section. Basally from this ventricular zone we can determine the less dense sub-ventricular zone, however based on cell density we cannot easily differentiate between the inner and outer sub-ventricular zone. Further towards the pial edge the SVZ transforms into the intermediate zone as indicated by sparse purple nuclei and very dense pink fibres. Next, we can see the IZ become the cortical plate/subplate regions, this is characterized by a change to dense

purple nuclei indicating the dense populations of neurons in their final position of the cortical plate that will go on the form the cortical layers. Although not indicated in the figures, adjacent to the pial surface we can observe a thin layer with very few nuclei, this is the marginal zone which mostly comprises of RGC pial end-feet and fibres, however, will eventually form layer 1 of the cortex. The nuclear densities used to determine each cortical region are summarised in Table 34.

Table 34: Summary of nuclear density in each cortical region

Cortical Region	**Cell Nuclear Density**
Ventricular zone	Very high
Sub-ventricular zone	Moderate
Intermediate zone	Low
Sub-plate/Cortical plate	High

These H&E stains act to provide a guide for the different cortical regions when studying the *in situ* hybridisation data.

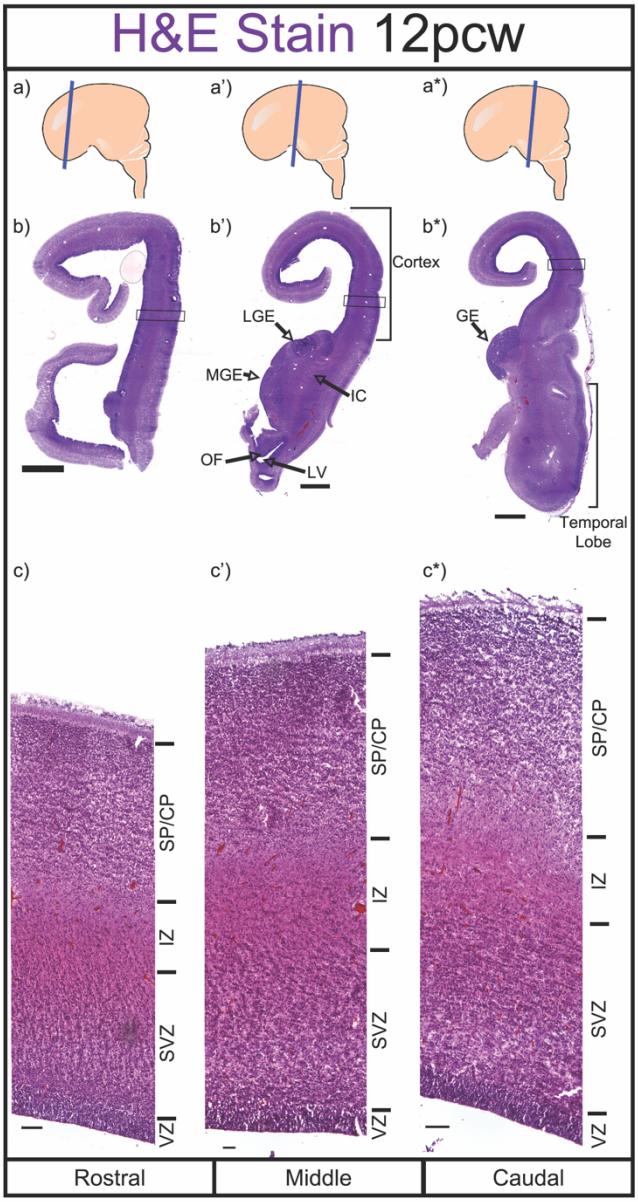


Figure 32: 12pcw H&E Stain

12pcw brain sections stained with H&E stains. a-a* show schematic of brain location taken for section. At low power (b-b*) scale bar = 2mm and high power (c-c*) scale bar = 100μm. Purple stain indicates cell nucleus and pink stain shows extracellular matrix and cytoplasm. b-b*) regions of brain indicated: LGE = lateral ganglionic eminence, MGE = medial ganglionic eminence, IC = internal capsule, LV = lateral ventricle, OF = orbitofrontal neuroepithelium and subventricular zone, GE = ganglionic eminence.

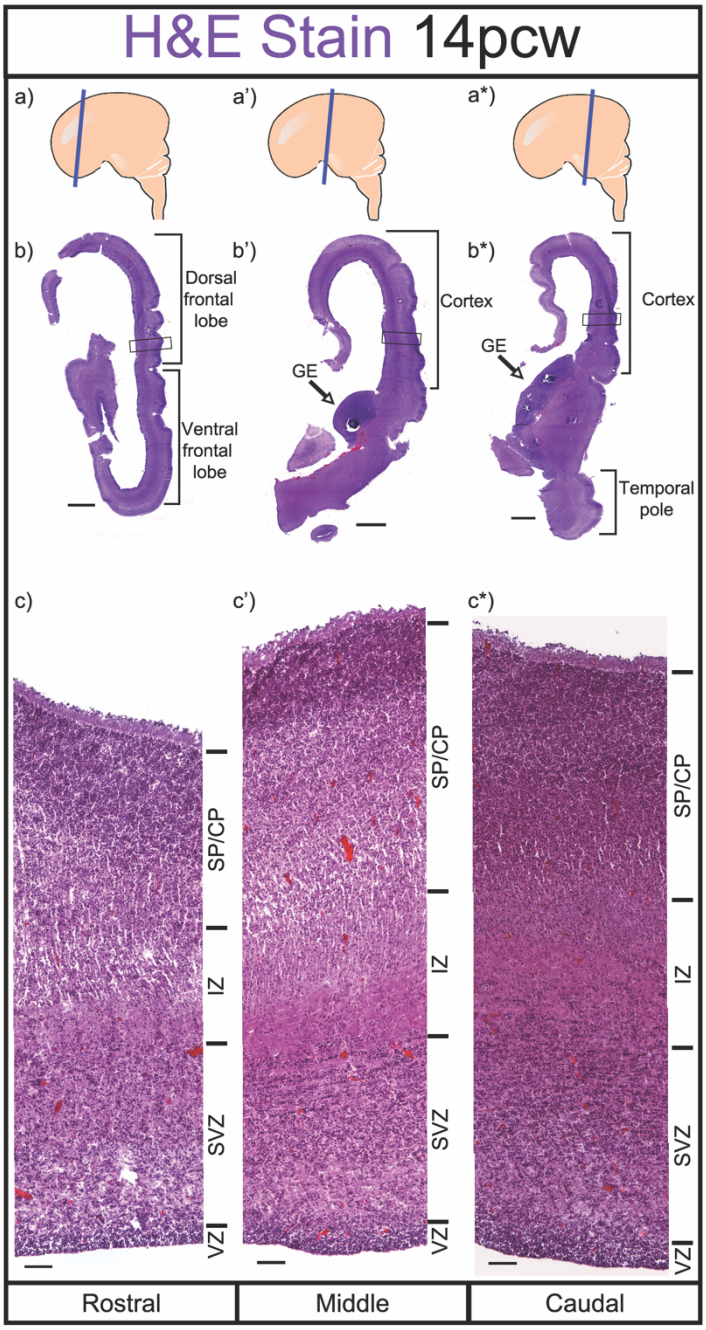


Figure 33: H&E Stain at 14pcw

14pcw brain sections stained with H&E stains. a-a* show schematic of brain location taken for section. At low power (b-b*) scale bar = 2mm and high power (c-c*) scale bar = 100μm. Purple stain indicates cell nucleus and pink stain shows extracellular matrix and cytoplasm.

b-b*) regions of brain indicated: GE= ganglionic eminence

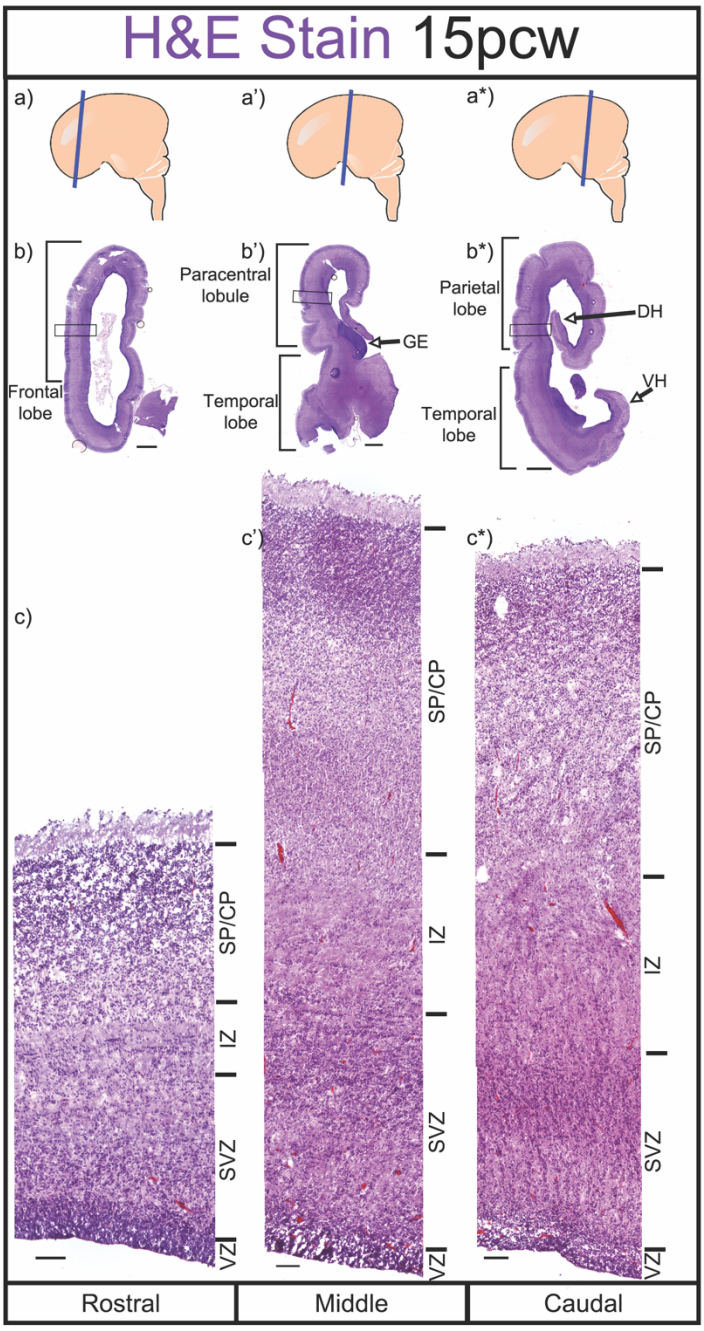


Figure 34: 15pcw H&E stain

15pcw brain sections stained with H&E stains. a-a* show schematic of brain location taken for section. At low power (b-b*) scale bar = 2mm and high power (c-c*) scale bar = 100μm. Purple stain indicates cell nucleus and pink stain shows extracellular matrix and cytoplasm. b-b*) regions of brain indicated: GE= ganglionic eminence, DH = dorsal hippocampus, VH = ventral hippocampus

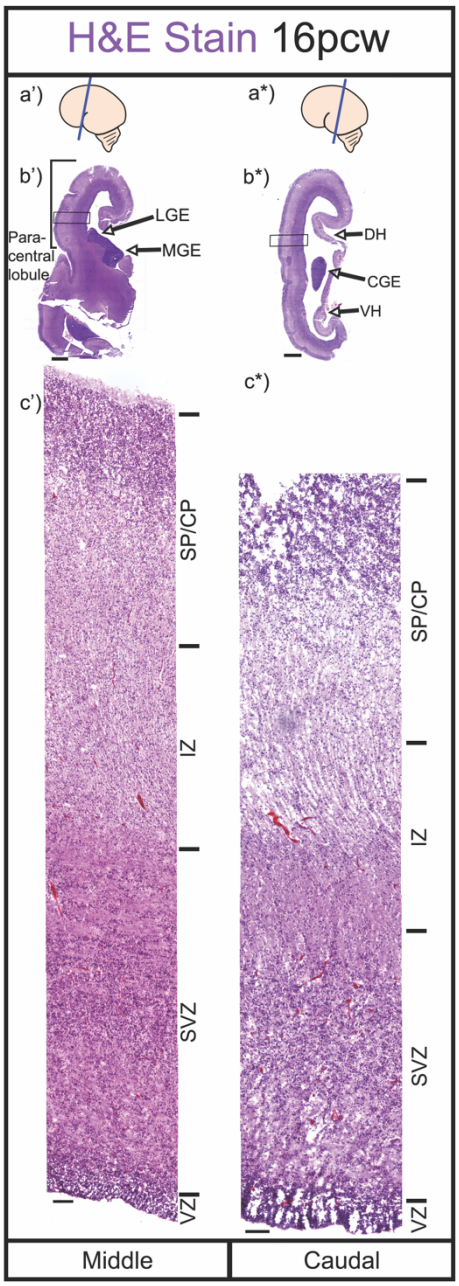


Figure 35: H&E stain at 16pcw

15pcw brain sections stained with H&E stains. a-a* show schematic of brain location taken for section. At low power (b-b*) scale bar = 2mm and high power (c-c*) scale bar = 100μm. Purple stain indicates cell nucleus and pink stain shows extracellular matrix and cytoplasm. b-b*) regions of brain indicated: DH = dorsal hippocampus, VH = ventral hippocampus, LGE = lateral ganglionic eminence, MGE = medial ganglionic eminence, CGE = caudal ganglionic eminence

3.3.5.2 Confirmation of the germinative zones using PAX6

As the focus of this thesis is the proliferative germinative zones of the cortex, the VZ and SVZ, we wanted to be sure we had correctly identified them. Therefore, we stained a 12pcw brain section for PAX6 protein, the transcription factor expression by glial cells in both the VZ and SVZ. Our result in Figure 36 showed that PAX6 protein is strongly expressed in almost all cells of the VZ and is expressed in a subset of cells in the SVZ. There is a change in nuclear density between the SVZ and IZ which is also identifiable by ablation of PAX6 staining. There is also no PAX6 stain present in the SP/CP.

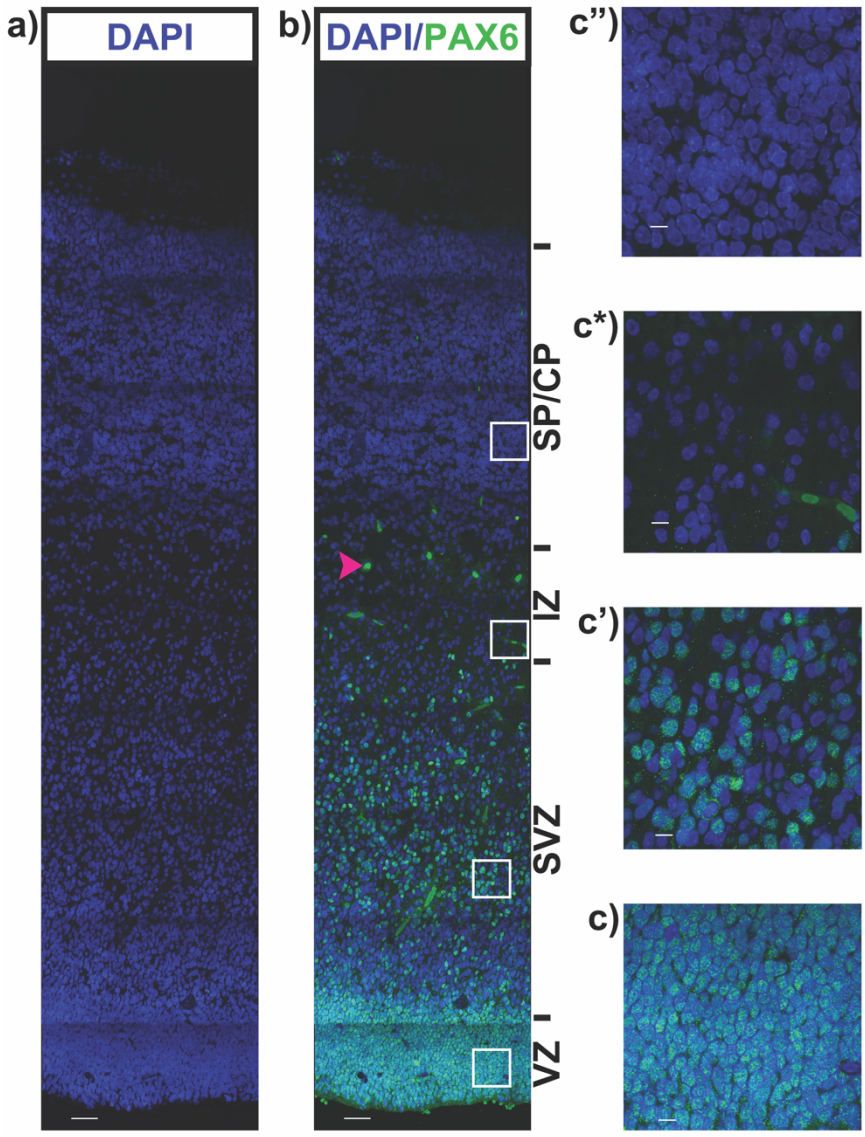


Figure 36: PAX6 protein expression at 12pcw

PAX6 protein at 12pcw. a) the nuclear DAPI stain allowing us to see nuclear cytoarchitecture of the telencephalic wall. b) PAX6 (green) stain overlaid to DAPI (blue). Non-specific antibody binding to blood vessels is indicated by pink arrow. White boxes indicate higher region shown for higher magnification. a, b) scale bars = 50μm. c-c'') high magnification images for the c) VZ, c') SVZ, c*) IZ and c'') SP/CP. Scale bars = 10μm.

3.3.5.3 *KIF22*

The first transcript we stained for was *KIF22*. The scRNA-seq data predicted the mRNA to be expressed highly in progenitor cells and greatly decreased in post-mitotic cells as shown in Figure 37. We stained rostral-middle-caudal sections at four ages as described in the methods (with the exception of 16pcw where the rostral part of the brain was damaged beyond use) and the results are shown for 12pcw (Figure 38), 14pcw (Figure 39), 15pcw (Figure 40) and 16pcw (Figure 41). Our *in situ* hybridisation produced staining for mRNA first confirming that the *KIF22* mRNA is expressed in brain tissue at the ages studied.

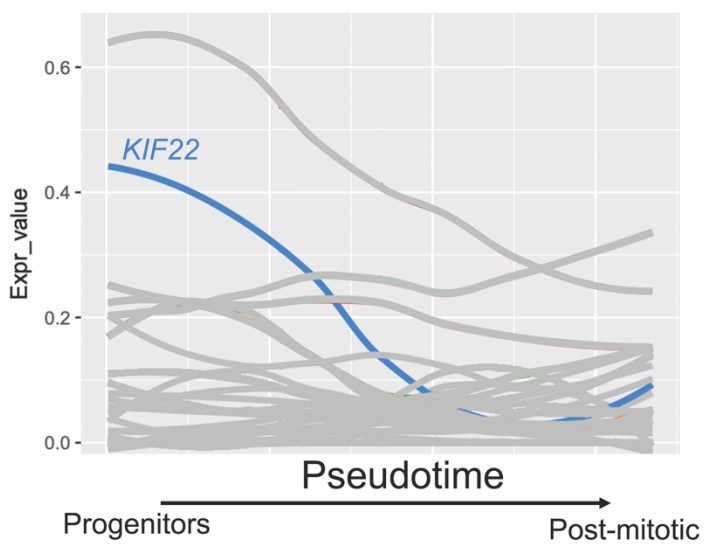


Figure 37: *KIF22* mRNA expression through pseudotime

Pseudotime analysis as described above showing only *KIF22* mRNA as cells move from progenitor to post-mitotic expression profile.

At lower magnification (all *KIF22* figures parts b-b*) we can see the expression of *KIF22* mRNA across the entire section, it is expressed at all ages studied across the rostral-middle-caudal axis. It looks to be restricted to

the germinative zones of the cortex with a clear blue band visible around the innermost edge of the cortex. Additionally, extremely strong *KIF22* mRNA staining is observed at the ganglionic eminence of several sections (specifically Figure 38 part b' and b*, Figure 39 part b' and b*, Figure 39 part b' and Figure 41 part b'). In some sections we can observe a small region of strong highly localised *KIF22* mRNA staining (Figure 38 parts b' and b*), this is the region around the lateral ventricle.

We then studied the *KIF22* mRNA expression at higher magnification, (all *KIF22* figures part c-c*) examining across the cortex as indicated by the box in part b-b*. Using the H&E sections above as a guide for cortical lamination we observed at all ages and rostral-middle-caudal sections *KIF22* mRNA staining to be most prominent in the germinative VZ and SVZ of all sections while there is very few or complete absence of expressing cells in the IZ and SP/CP. We also note the "speckled" appearance of the stain suggesting that *KIF22* mRNA is present in a subset of cells within the germinative zones. This similarity of *KIF22* mRNA expression *in vivo* to our result of the scRNA-seq analysis, coupled with its restrictive expression to germinative zones *in vivo* make it a good candidate for further protein studies.

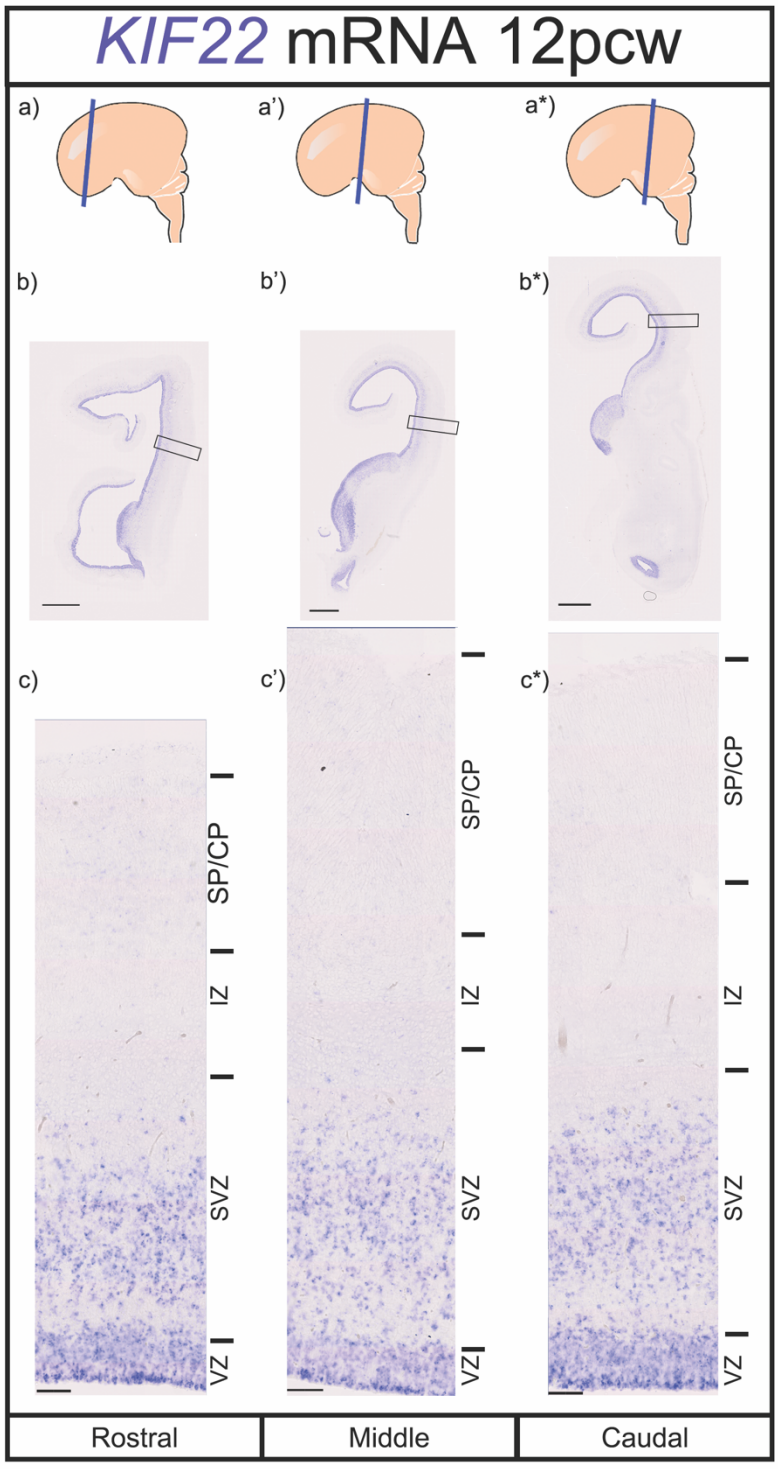


Figure 38: *in situ* hybridisation of *KIF22* at 12pcw

12pcw brain sections stained with a *KIF22* probe. a-a* show schematic of brain location taken for section. At low power (b-b*) scale bar = 2mm and high power (c-c*) scale bar = 100μm.

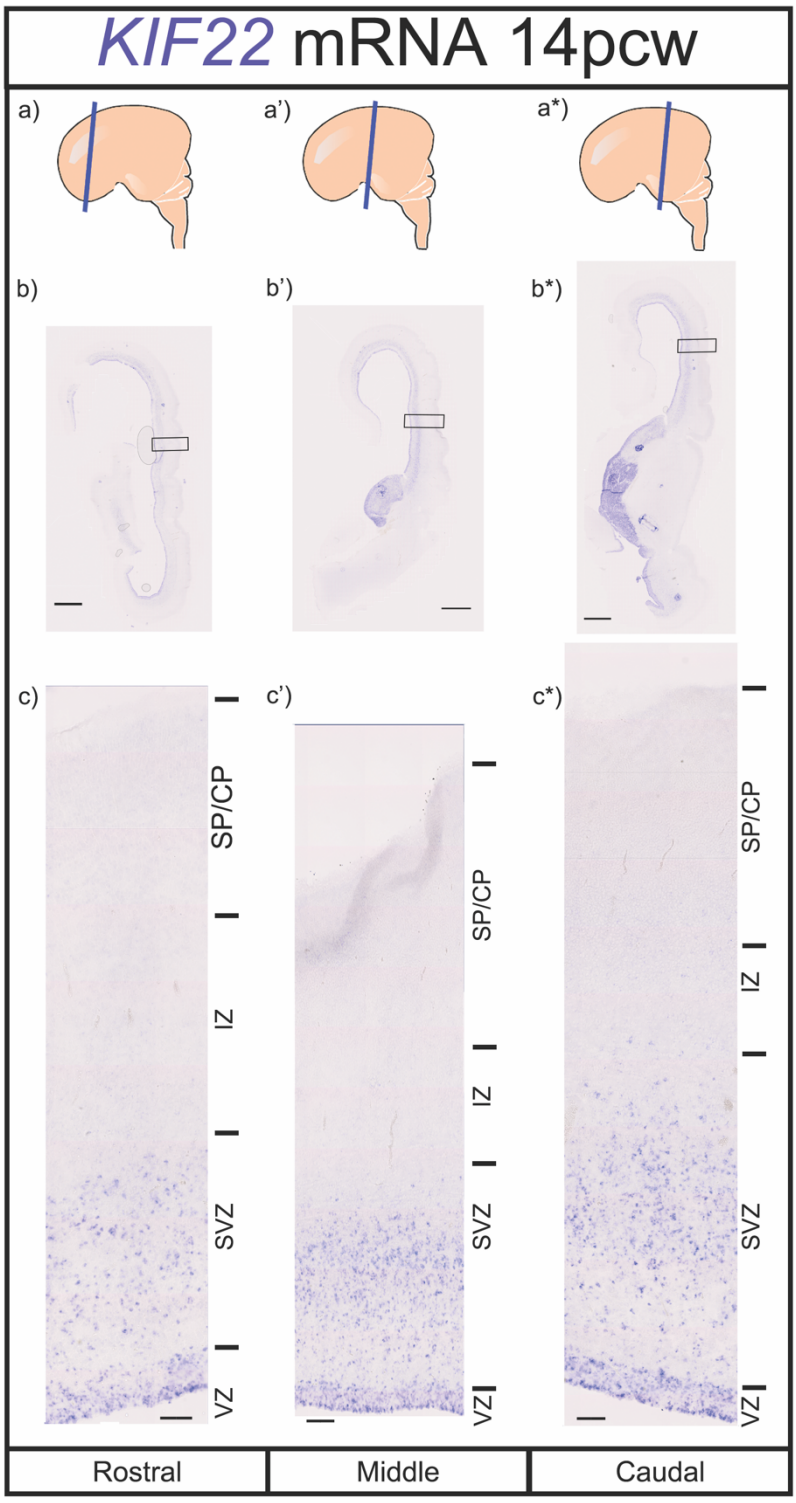


Figure 39: *in situ* hybridisation with *KIF22* probe at 14pcw

14pcw brain sections stained with a *KIF22* probe. a-a* show schematic of brain location taken for section. At low power (b-b*) scale bar = 2mm and high power (c-c*) scale bar = 100μm.

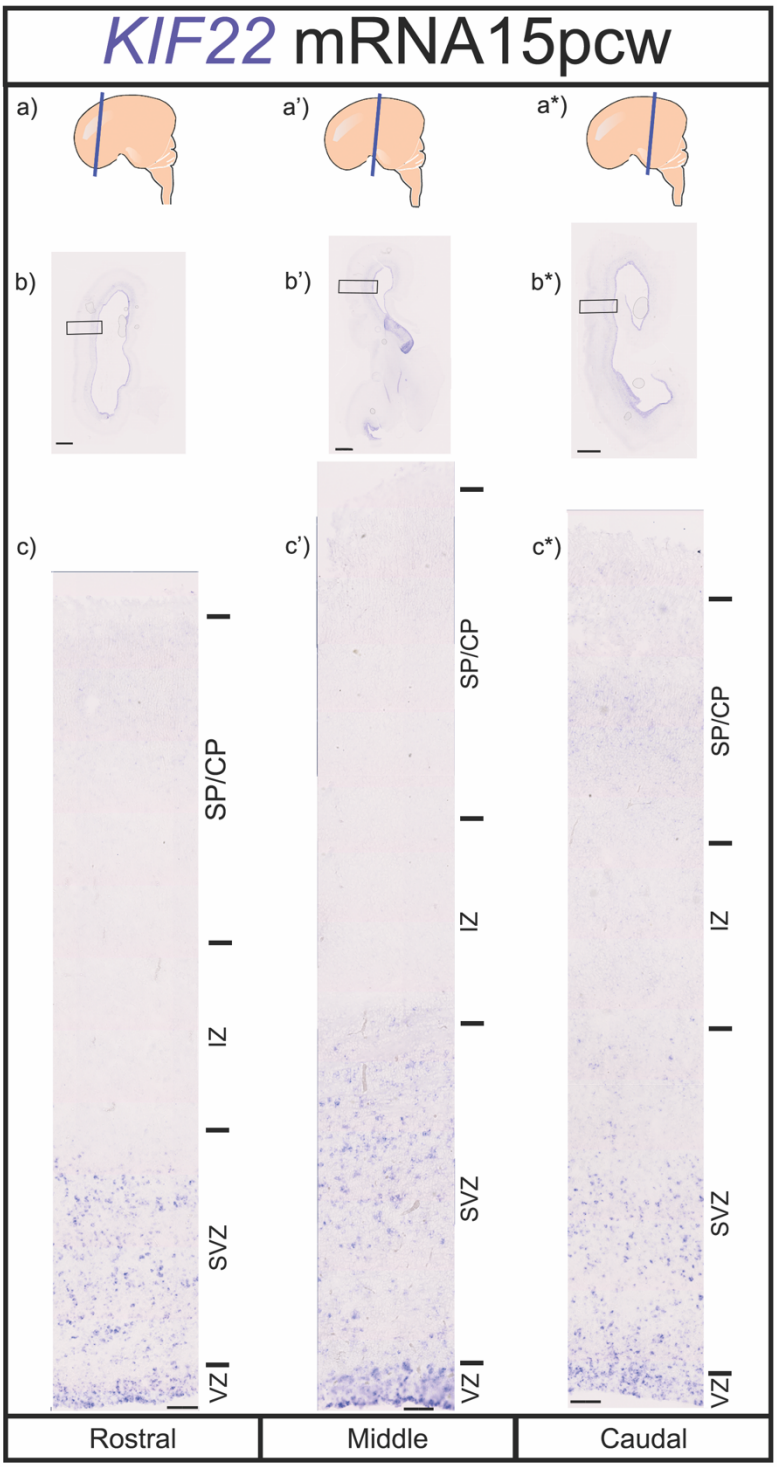


Figure 40: *in situ* hybridisation with *KIF22* probe at 15pcw

15pcw brain sections stained with a *KIF22* probe. a-a* show schematic of brain location taken for section. At low power (b-b*) scale bar = 2mm and high power (c-c*) scale bar = 100μm.

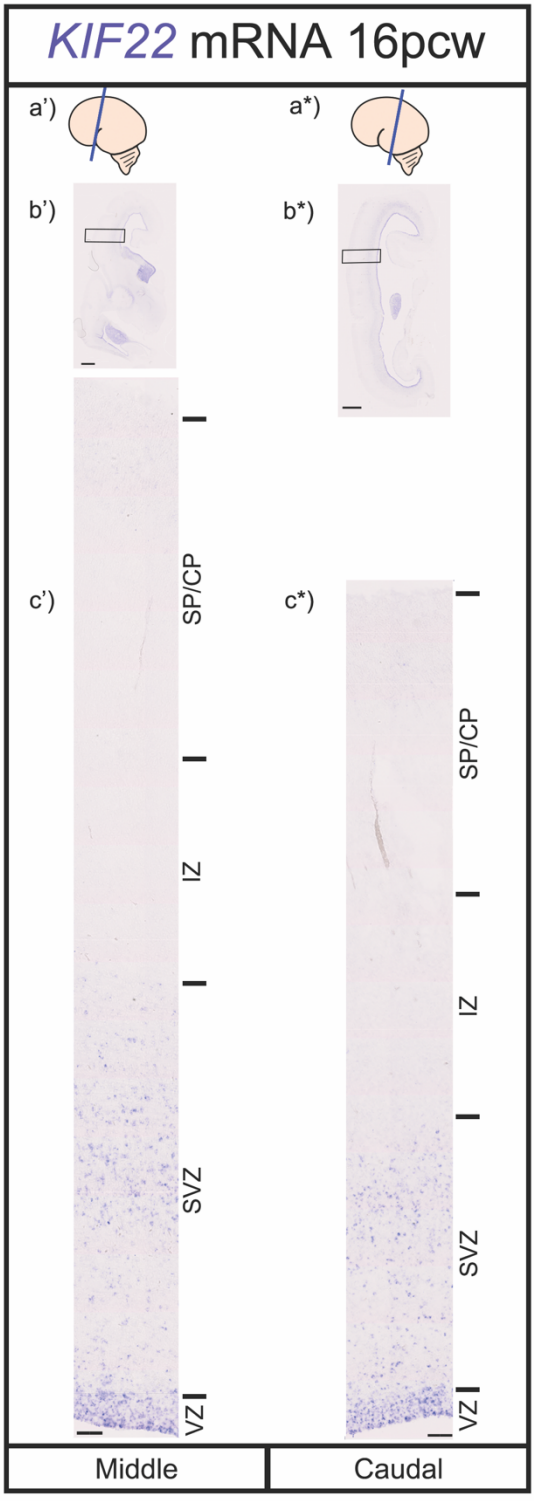


Figure 41: *in situ* hybridisation with *KIF22* probe at 16pcw

16pcw brain sections stained with a *KIF22* probe. a-a* show schematic of brain location taken for section. At low power (b-b*) scale bar = 2mm and high power (c-c*) scale bar = 100μm.

3.3.5.4 *ALDOA*

The second probe used was for *ALDOA* mRNA. The scRNA-seq analysis showed this gene to be highest expressed in progenitors and decreasing significantly but still expressed, albeit at lower levels, as cells become post-mitotic as shown in Figure 42. As previously described, we stained four ages and data is shown for 12pcw (Figure 43), 14pcw (Figure 44), 15pcw (Figure 45) and 16pcw (Figure 46) across the rostral-middle-caudal axis with the exception of 16pcw where the rostral part of the brain was damaged beyond use.

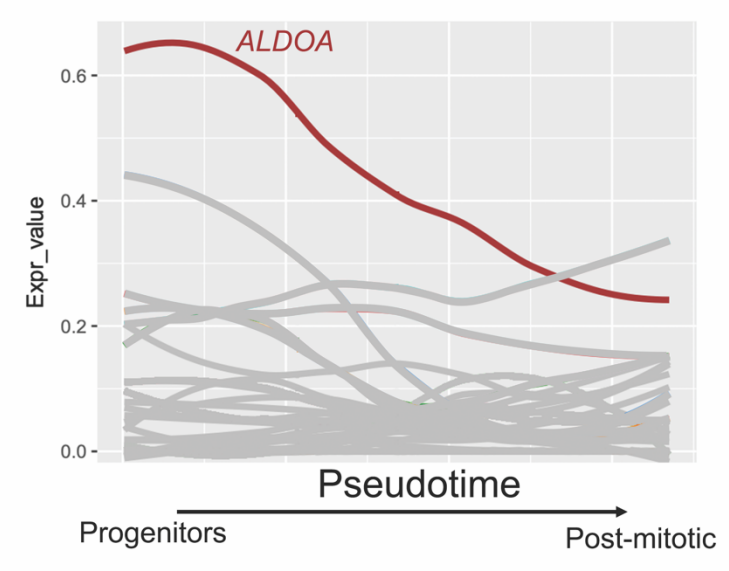


Figure 42: *ALDOA* mRNA expression through pseudotime

Pseudotime analysis as described above showing only *ALDOA* mRNA as cells move from progenitor to post-mitotic expression profile.

Our staining at low magnification (all *ALDOA* figures parts b-b*) show *ALDOA* mRNA staining to be present in all sections at all ages. Like *KIF22* there is a strong staining band around the inside, the germinative zones of

the cortex. However, unlike *KIF22*, *ALDOA* mRNA staining is also present throughout the rest of the cortex. We also can observe high *ALDOA* staining in the ganglionic eminences (particularly in Figure 43 parts b' and b*, Figure 44 parts b' and b*, Figure 45 part b' and Figure 46 part b'). As with the *KIF22* mRNA we can also see high *ALDOA* mRNA expression in the ventricular zone surrounding the lateral ventricle (Figure 43 part b' and b* and Figure 45 part b').

Looking at the higher magnification across the cortex (as indicated by the black box on all *ALDOA* figures part b-b*) we can see the expression of *ALDOA* mRNA in the different parts of the cortex (seen in all *ALDOA* figures parts c-c*). We observe at all ages and rostral-middle-caudal sections *ALDOA* staining to be most prominent in the proliferative ventricular and sub-ventricular zones, however there are also substantial numbers of *ALDOA* expressing cells in SP/CP and some in the IZ (however as shown by H&E there are very few nuclei in the IZ). This matches what was predicted by the scRNA-seq data and we also note that, particularly in the VZ, *ALDOA* seems to be present in many more cells than *KIF22* matching the gradient plot shown in Figure 23.

This similarity of *ALDOA* mRNA expression *in vivo* to that shown by the scRNA-seq analysis, coupled with its high expression in germinative zones *in vivo* make it a good candidate for further protein studies.

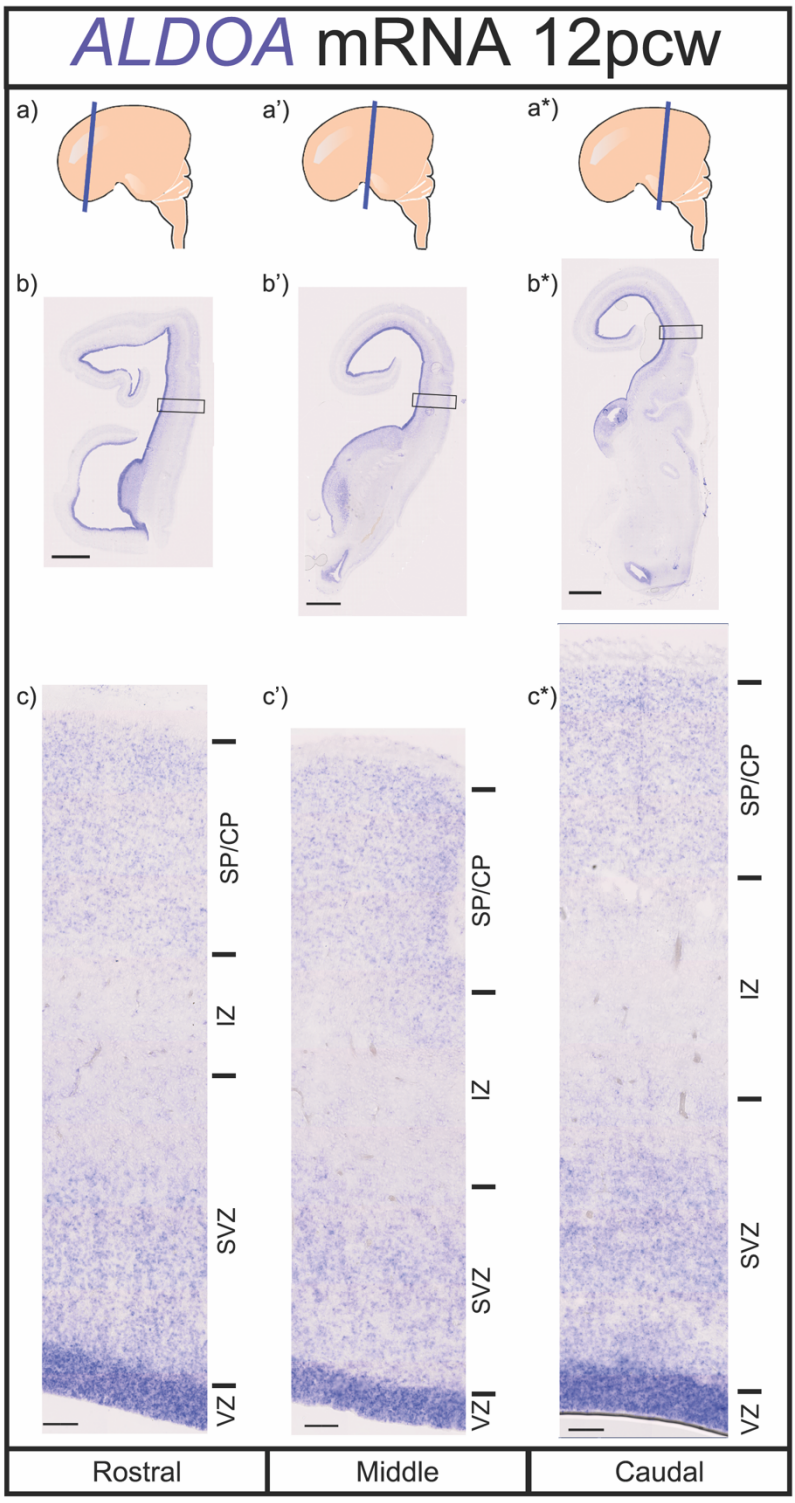


Figure 43: *in situ* hybridisation with *ALDOA* probe at 12pcw

12pcw brain sections stained with an *ALDOA* probe. a-a* show schematic of brain location taken for section. At low power (b-b*) scale bar = 2mm and high power (c-c*) scale bar = 100μm.

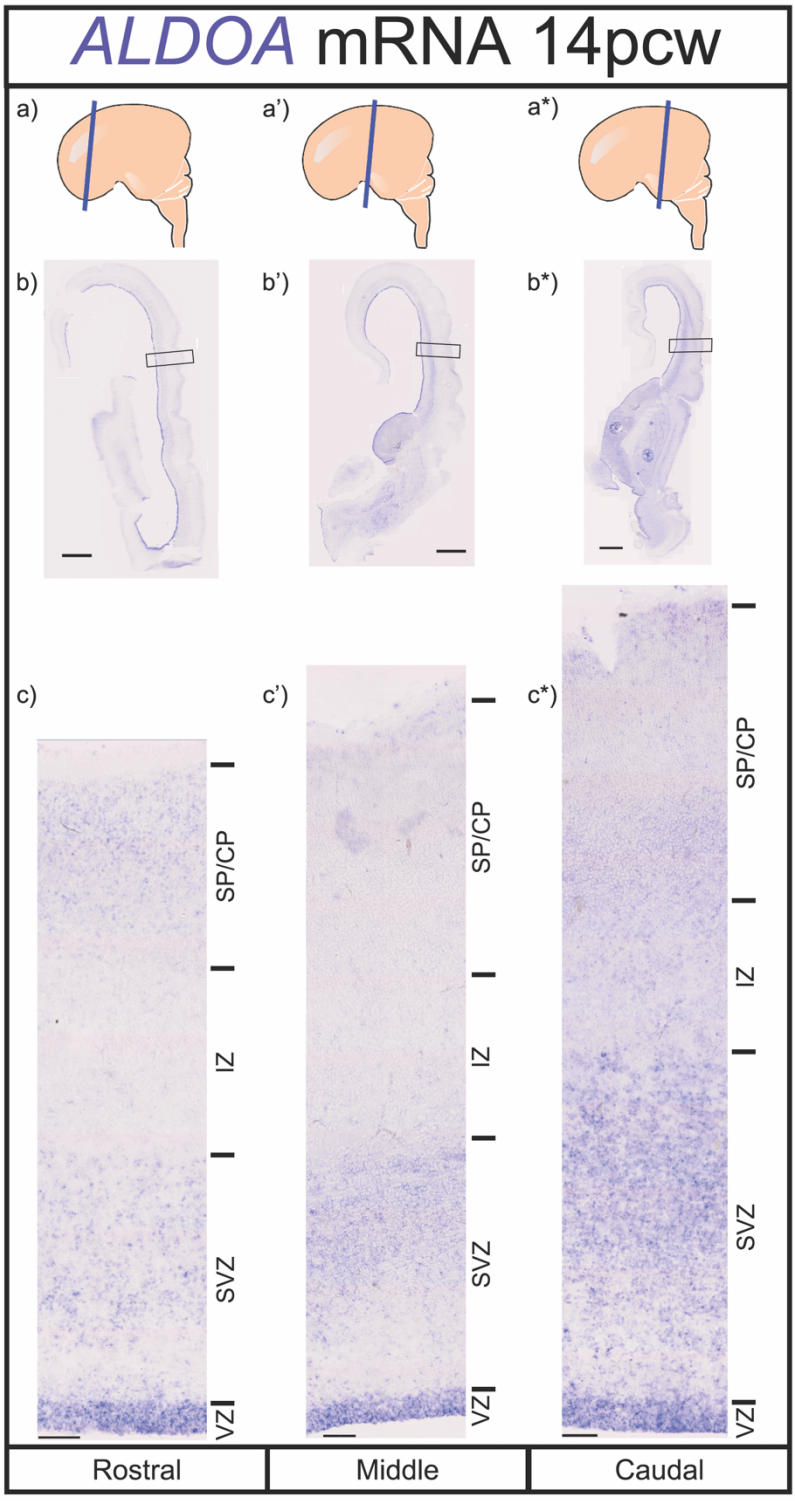


Figure 44: *in situ* hybridisation with *ALDOA* probe at 14pcw

14pcw brain sections stained with an *ALDOA* probe. a-a* show schematic of brain location taken for section. At low power (b-b*) scale bar = 2mm and high power (c-c*) scale bar = 100μm.

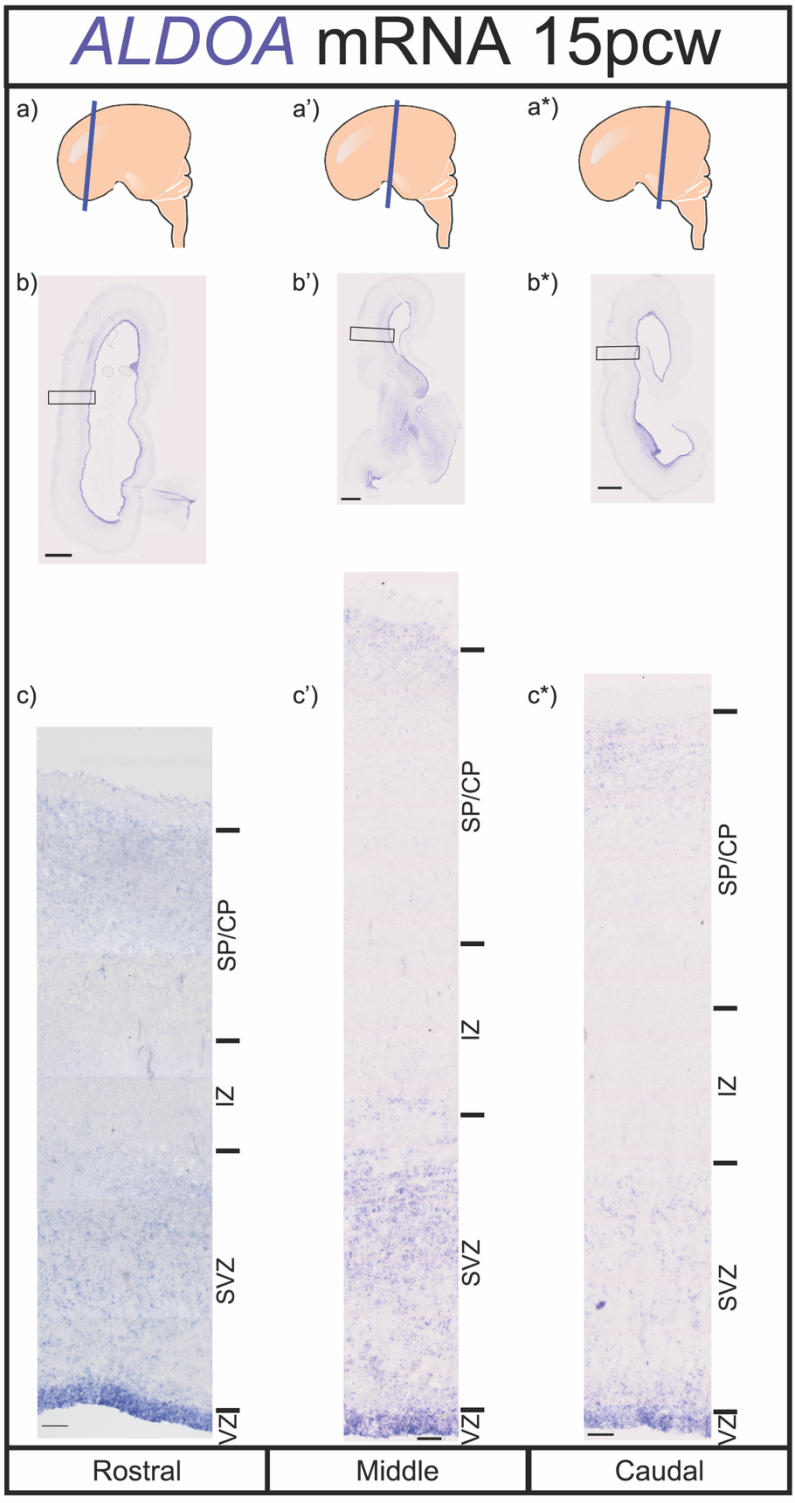


Figure 45: *in situ* hybridisation with *ALDOA* probe at 15pcw

15pcw brain sections stained with an *ALDOA* probe. a-a* show schematic of brain location taken for section. At low power (b-b*) scale bar = 2mm and high power (c-c*) scale bar = 100μm.

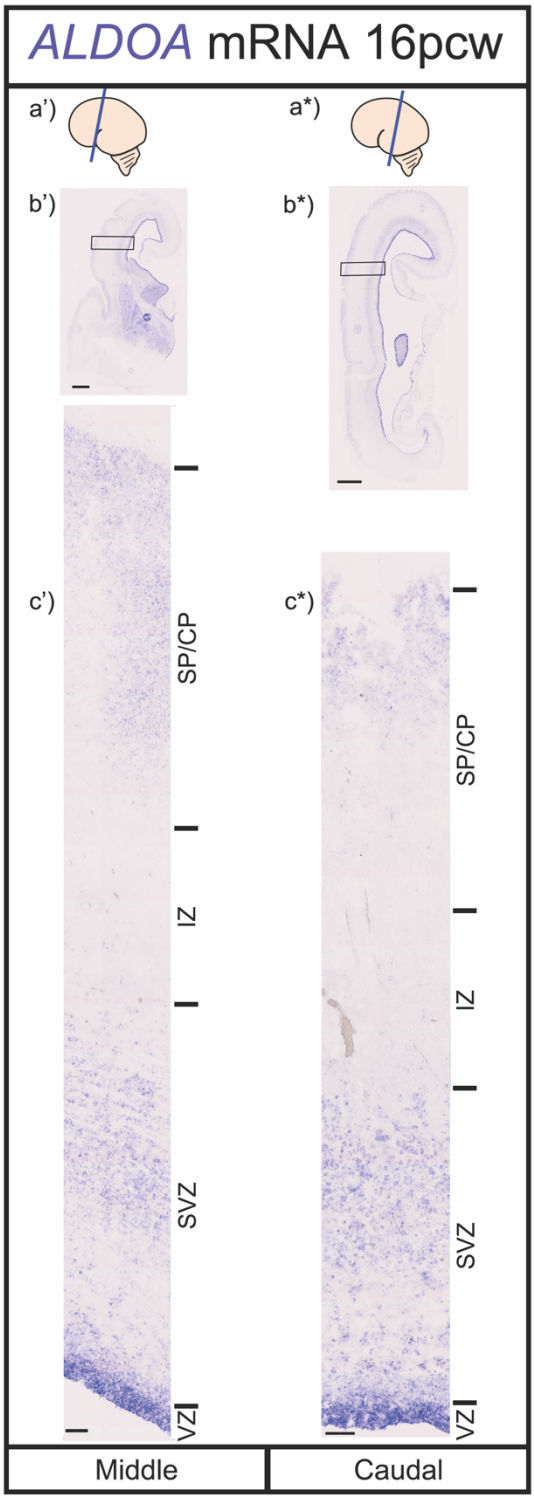


Figure 46: *in situ* hybridisation with *ALDOA* probe at 16pcw

16pcw brain sections stained with an *ALDOA* probe. a-a* show schematic of brain location taken for section. At low power (b-b*) scale bar = 2mm and high power (c-c*) scale bar = 100μm.

3.3.5.5 *HIRIP3*

The third gene used for *in situ* hybridisation was *HIRIP3*. This gene was shown by the scRNA-seq analysis to be expressed in progenitor cells, although at lower levels than *ALDOA* and *KIF22* and to decrease as cells move to post-mitotic state (as shown in Figure 47) but not significantly. As before we stained rostral-middle-caudal sections at four ages as described in methods, except 16pcw rostral. The results are shown for 12pcw (Figure 48), 14pcw (Figure 49), 15pcw (Figure 50) and 16pcw Figure 51). Our *in situ* hybridisation staining was successful confirming that the *HIRIP3* mRNA is expressed in brain tissue at the ages studied.

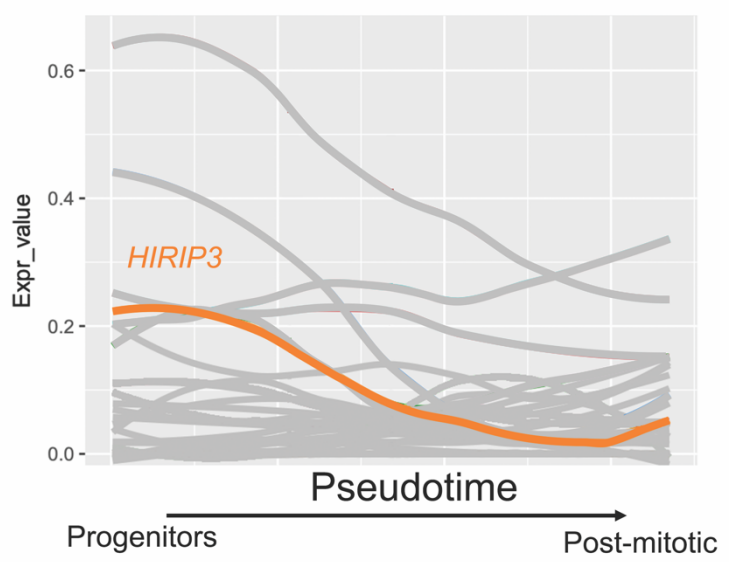


Figure 47: *HIRIP3* mRNA expression through pseudotime

Pseudotime analysis as described above showing only *HIRIP3* mRNA as cells move from progenitor to post-mitotic expression profile.

First looking at lower magnification (all *HIRIP3* figures parts b-b*) we see *HIRIP3* mRNA expression across the entire section and it is expressed at all ages studied across the rostral-middle-caudal axis. Unlike the *KIF22* and

ALDOA mRNA stains there is not the clear band of staining around the ventricular zone, the increased staining that can be seen here is likely due to increased cell density in the ventricular zone. We also do not observe the very strong staining in the ganglionic eminences with the exception of 14pcw caudal (Figure 49 part b*).

Moving to *HIRIP3* mRNA expression at higher magnification (all *HIRIP3* figures part c-c*) we looked at expression across the cortex as indicated by the box in parts b-b*. Using the H&E sections above as a guide to the cortical lamination we observed *HIRIP3* to be expressed (when taking varying cell density into account), rather uniformly across the cortex at all ages and locations in the rostral-middle-caudal axis. It appears to be more strongly expressed in the ventricular zone, however that may be an artefact of the high nuclear density in this region. We also observe a speckled pattern of staining suggesting that *HIRIP3* mRNA is in a subset of cells. This staining shows that, as demonstrated by the scRNA-seq data, *HIRIP3* mRNA is not restricted to, or significantly higher expressed in the germinative zones of the human fetal cortex *in vivo*. This lack of progenitor enrichment from bioinformatics analysis and strong staining in germinative zones *in vivo* does not make *HIRIP3* a good candidate for further protein studies.

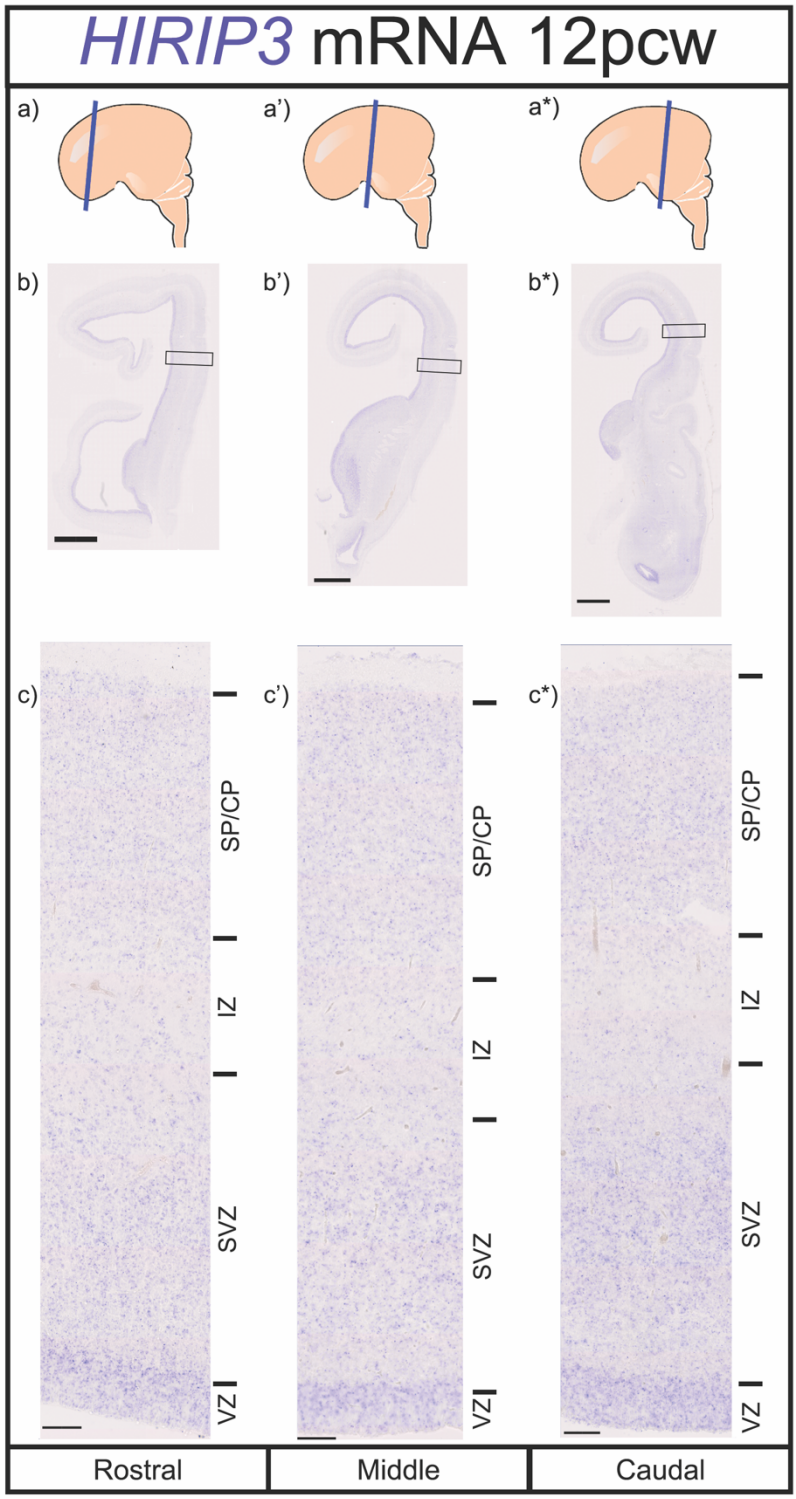


Figure 48: *in situ* hybridisation with *HIRIP3* probe at 12pcw

12pcw brain sections stained with a *HIRIP3* probe. a-a* show schematic of brain location taken for section. At low power (b-b*) scale bar = 2mm and high power (c-c*) scale bar = 100μm.

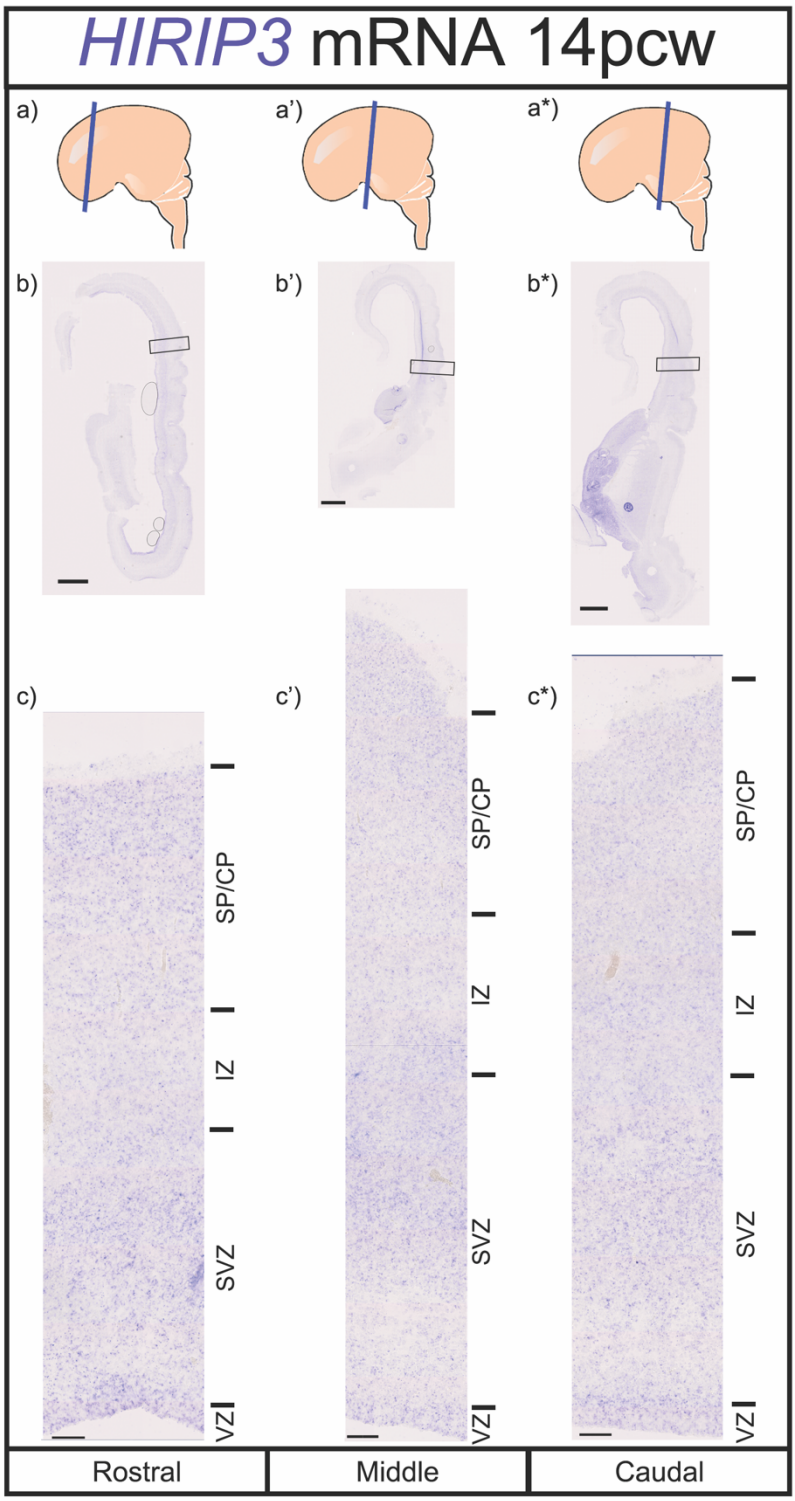


Figure 49: *in situ* hybridisation with *HIRIP3* probe at 14pcw

14pcw brain sections stained with a *HIRIP3* probe. a-a* show schematic of brain location taken for section. At low power (b-b*) scale bar = 2mm and high power (c-c*) scale bar = 100μm.

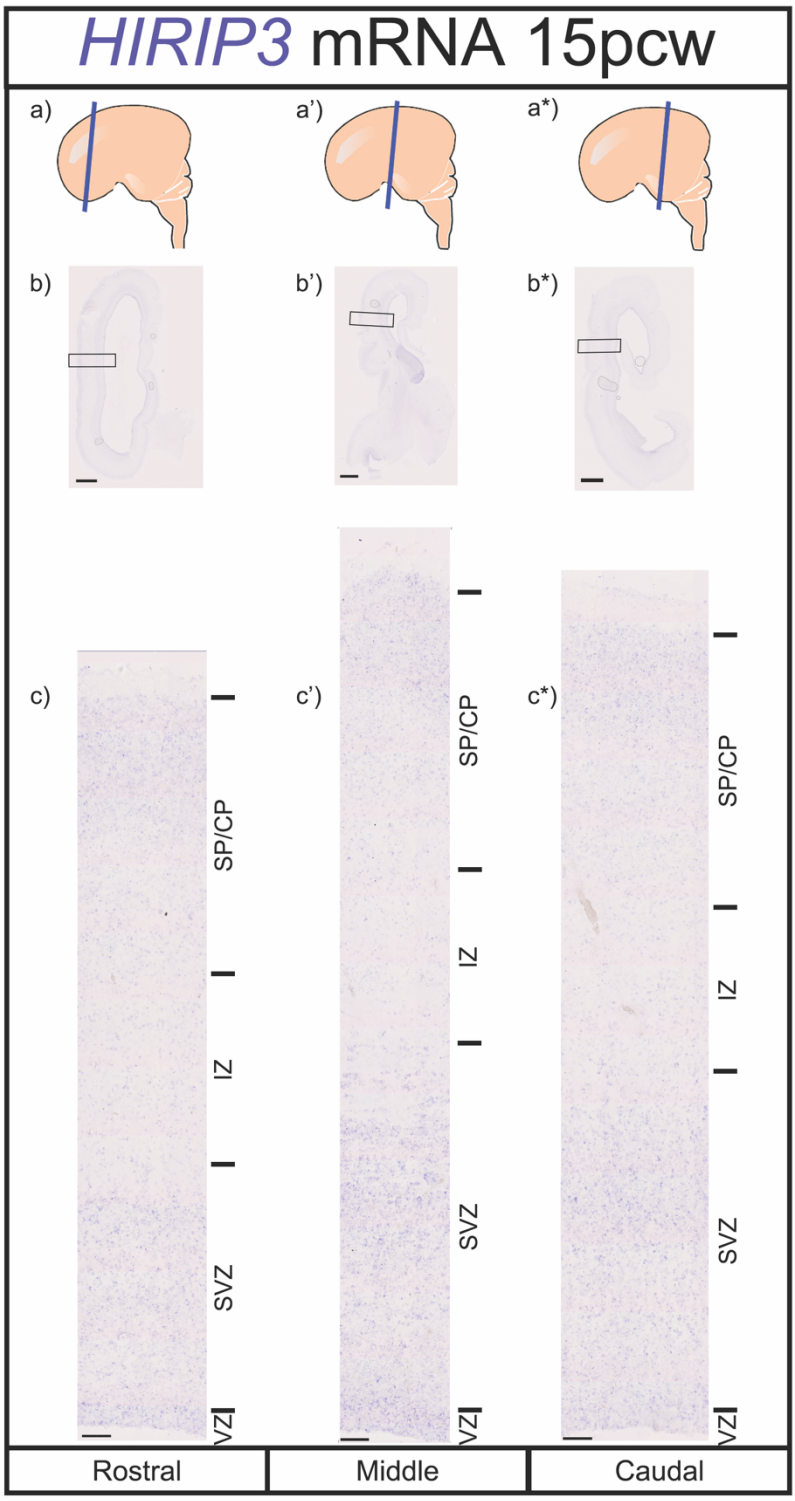


Figure 50: *in situ* hybridisation with *HIRIP3* probe at 15pcw

15pcw brain sections stained with a *HIRIP3* probe. a-a* show schematic of brain location taken for section. At low power (b-b*) scale bar = 2mm and high power (c-c*) scale bar = 100μm.

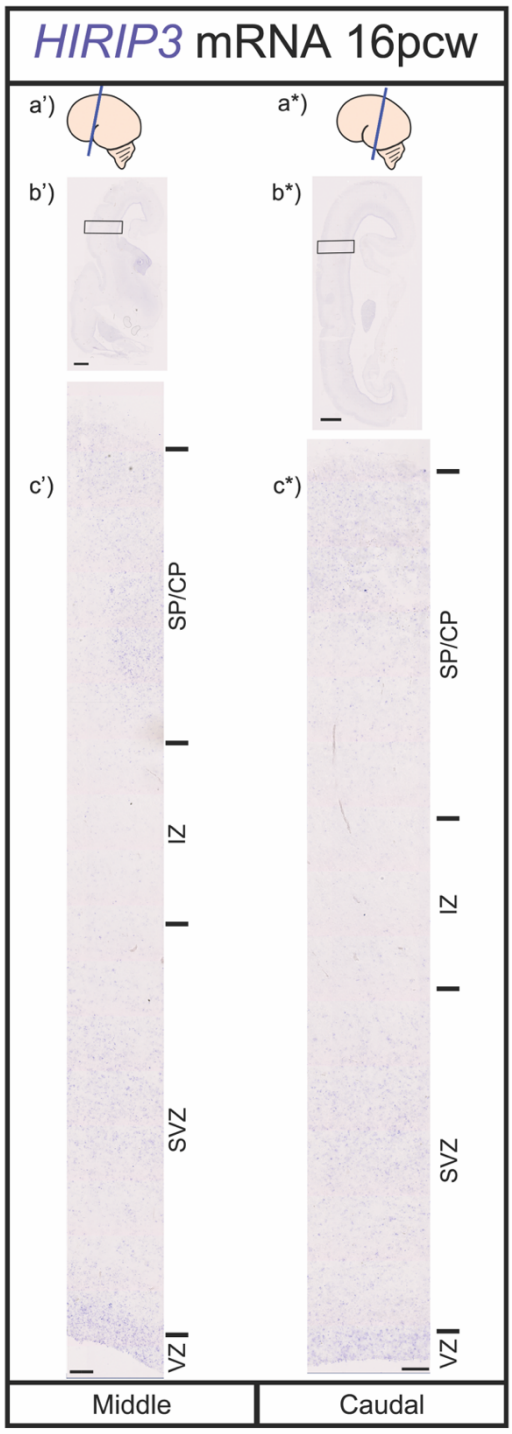


Figure 51: *in situ* hybridisation with *HIRIP3* probe at 16pcw

16pcw brain sections stained with a *HIRIP3* probe. a-a* show schematic of brain location taken for section. At low power (b-b*) scale bar = 2mm and high power (c-c*) scale bar = 100μm.

3.3.5.6 *PAGR1*

The fourth gene we studied mRNA expression of was *PAGR1*. *PAGR1* was shown by the scRNA-seq analysis to be expressed in progenitor cells at about the same level as *HIRIP3* and at lower levels than *ALDOA* and *KIF22*. It then decreased midway through the pseudotime and then increased again, although to lower levels than progenitors, as cells became post-mitotic as shown in Figure 52. *PAGR1* mRNA did not change significantly between progenitor and post-mitotic cell classes. Again, we stained rostral-middle-caudal sections at four stages, except 16pcw rostral. Our results for the *PAGR1 in situ* hybridisation are shown for 12pcw (Figure 53), 14pcw (Figure 54), 15pcw (Figure 55) and 16pcw (Figure 56). Our *in situ* hybridisation staining was successful despite requiring a long incubation time of four days and confirmed that *PAGR1* mRNA is expressed in human brain tissue at the ages studied.

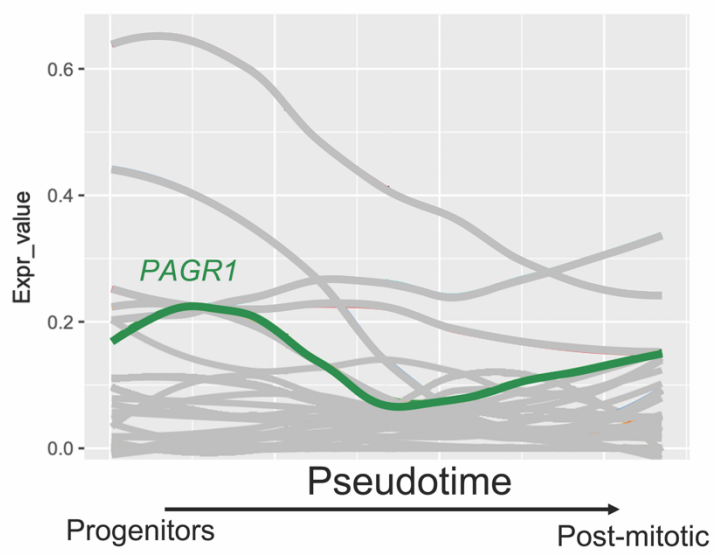


Figure 52: *PAGR1* mRNA expression through pseudotime

Pseudotime analysis as described above showing only *PAGR1* mRNA as cells move from progenitor to post-mitotic expression profile

First looking at lower magnification (all *PAGR1* figures parts b-b*) we can see *PAGR1* appears to be uniformly expressed at low levels across the entire section at all ages and across the rostral-middle-caudal axis. Accounting for cell density *PAGR1* does not seem to show particularly high expression in the ganglionic eminences at any ages.

Considering now *PAGR1* at higher magnification (all *PAGR1* figures part c-c*), we looked at its mRNA expression across the cortex as indicated by the boxes in parts b-b* and again used the H&E sections as a guide for cortical lamination. We observed some variation in staining intensity at some ages, most notable at 14pcw (Figure 54) where the middle has noticeably less staining than the rostral or caudal sections. Overall, it appears *PAGR1* mRNA is not restricted to any particular compartment of the cortex and its expression is rather uniform across the cortex, with particularly noticeable staining in the SP/CP compared to *HIRIP3.* This is perhaps indicative of the slight increase of *PAGR1* expression in post-mitotic cells seen by the scRNA-seq analysis. As with *HIRIP3* staining we also observe a speckled staining pattern of *PAGR1* mRNA suggesting that it is only expressed in a subset of cells.

This lack of progenitor enrichment *in vivo* or from bioinformatics analysis does not make *PAGR1* a good candidate for further studies.

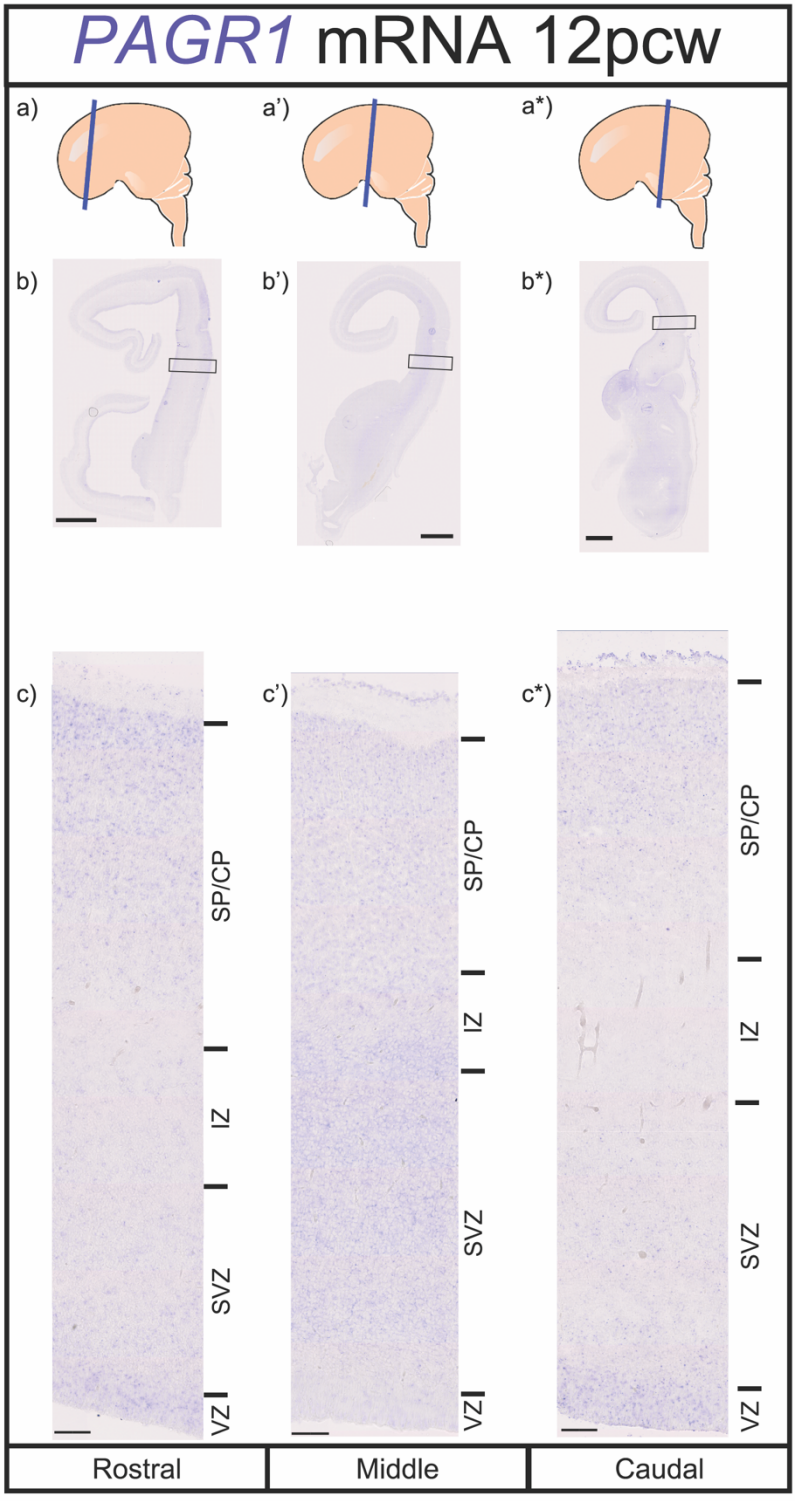


Figure 53: *in situ* hybridisation with *PAGR1* probe at 12pcw

12pcw brain sections stained with a *PAGR1* probe. a-a* show schematic of brain location taken for section. At low power (b-b*) scale bar = 2mm and high power (c-c*) scale bar = 100μm.

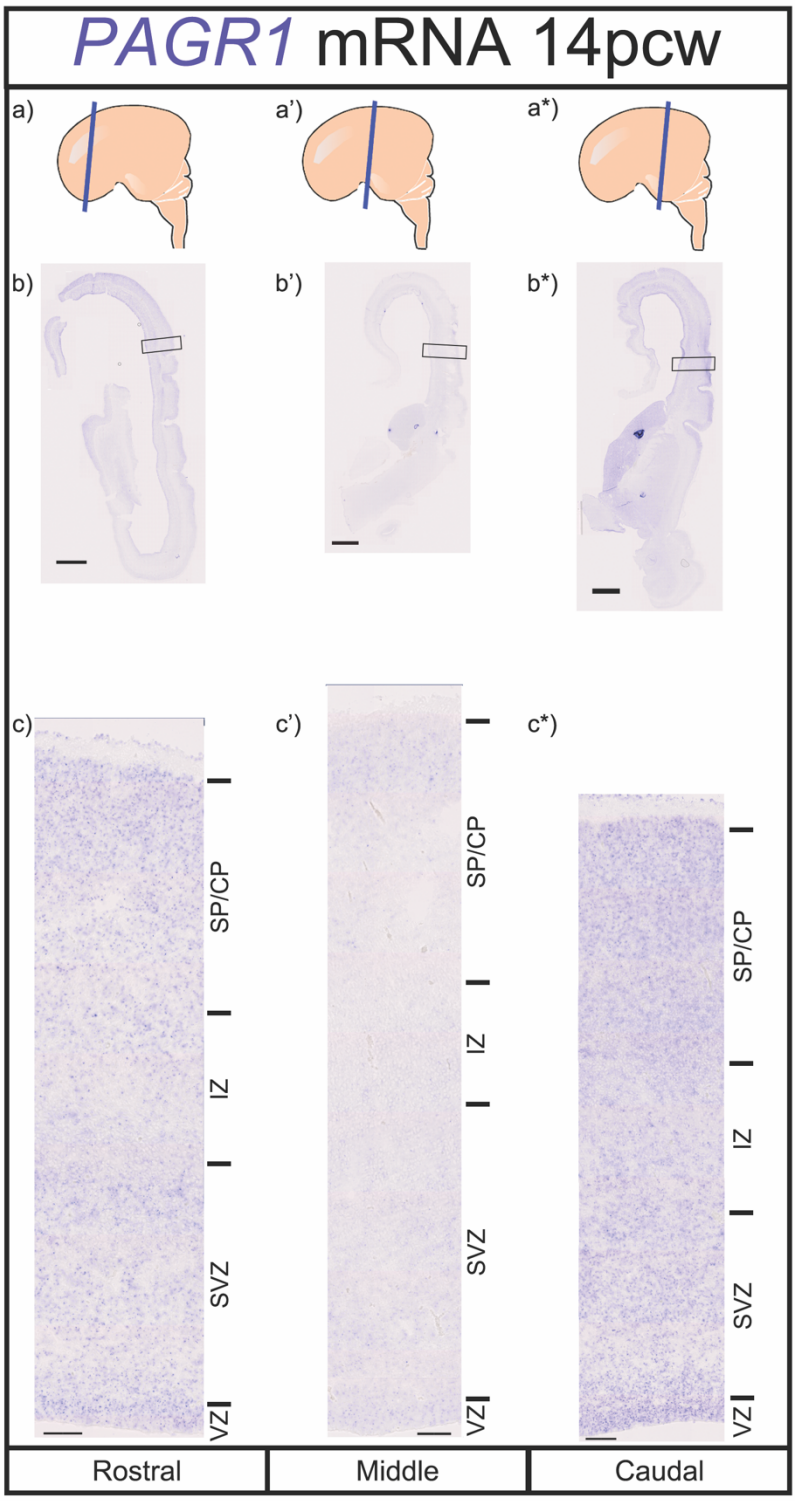


Figure 54: *in situ* hybridisation with *PAGR1* probe at 14pcw

14pcw brain sections stained with a *PAGR1* probe. a-a* show schematic of brain location taken for section. At low power (b-b*) scale bar = 2mm and high power (c-c*) scale bar = 100μm.

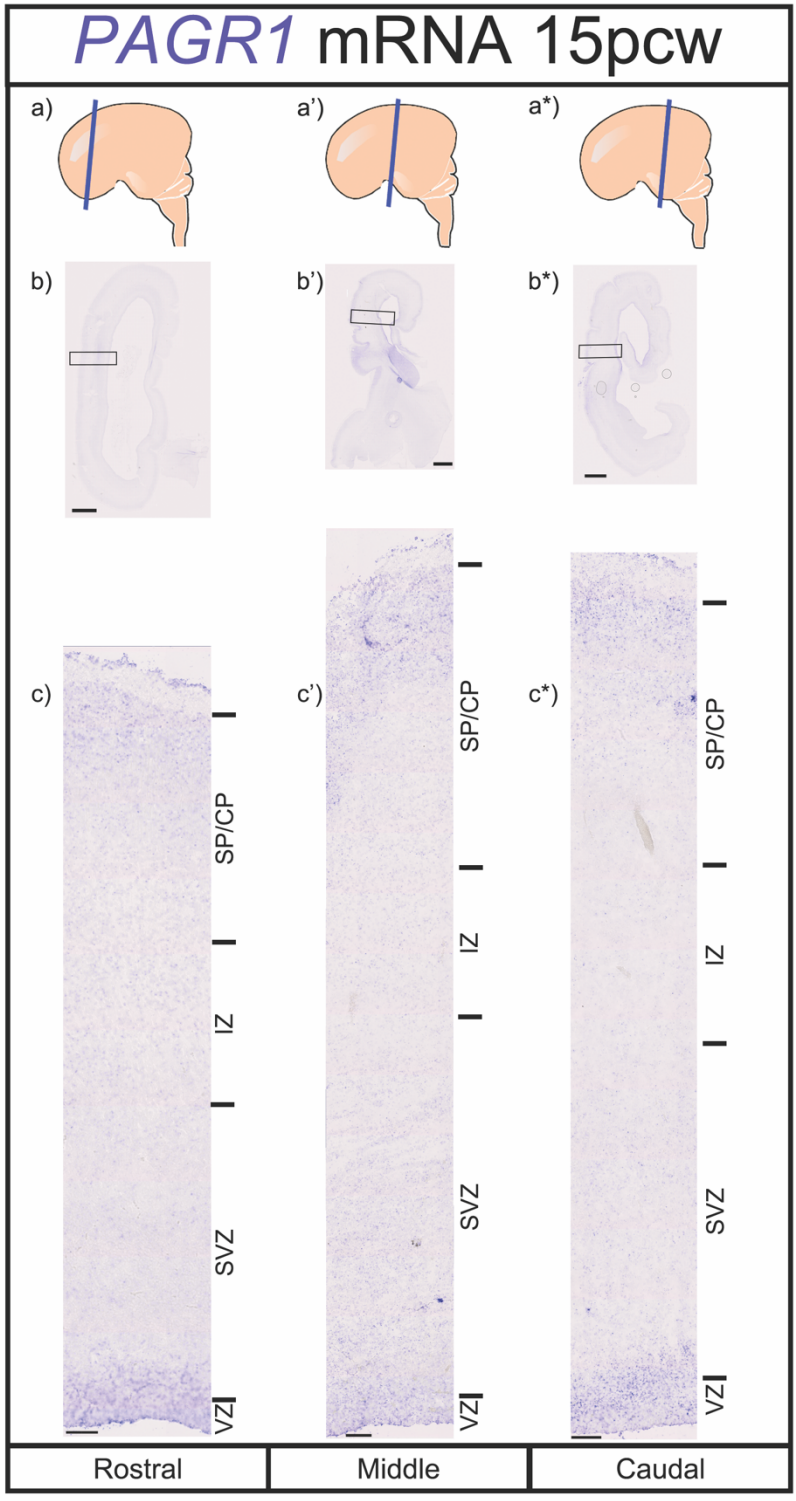


Figure 55: *in situ* hybridisation with *PAGR1* probe at 15pcw

15pcw brain sections stained with a *PAGR1* probe. a-a* show schematic of brain location taken for section. At low power (b-b*) scale bar = 2mm and high power (c-c*) scale bar = 100μm.

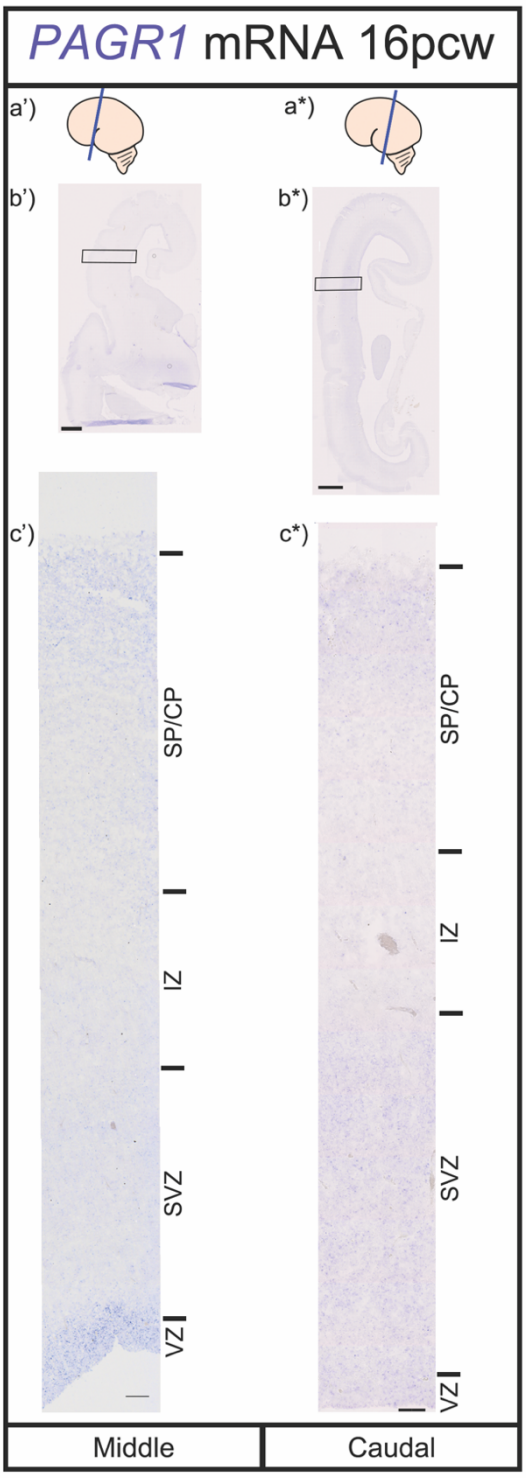


Figure 56: *in situ* hybridisation with *PAGR1* probe at 16pcw

16pcw brain sections stained with a *PAGR1* probe. a-a* show schematic of brain location taken for section. At low power (b-b*) scale bar = 2mm and high power (c-c*) scale bar = 100μm.

3.3.5.7 *MAZ*

The fifth gene studied *in vivo* using *in situ* hybridisation was *MAZ*. *MAZ* mRNA was shown by the scRNA-seq analysis to be expressed in progenitor cells, and to only decrease very slightly, and not significantly, as cells move to the post-mitotic state as shown in Figure 57. As with all genes, before we stained rostral-middle-caudal sections at four ages except 16pcw rostral. Our *in situ* hybridisation results are shown for 12pcw (Figure 58), 14pcw (Figure 59), 15pcw (Figure 60) and 16pcw (Figure 61). Our *in situ* hybridisation produced staining for mRNA allowing us to first confirm that the *MAZ* mRNA is expressed in brain tissue at the ages studied.

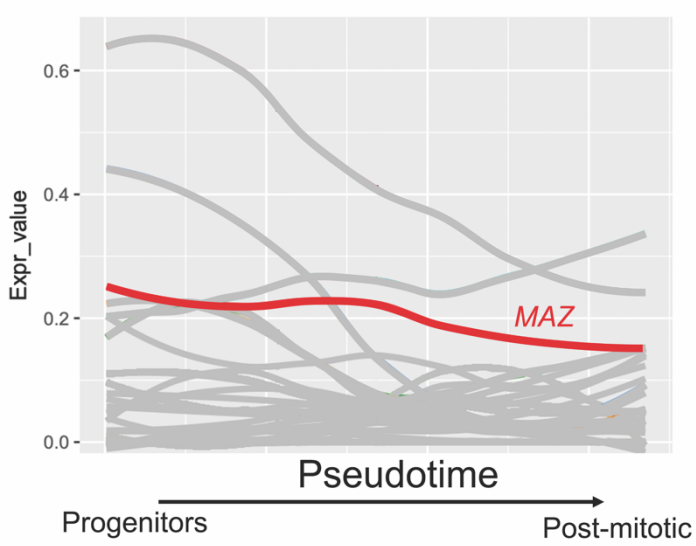


Figure 57: *MAZ* mRNA expression through pseudotime

Pseudotime analysis as described above showing only *MAZ* mRNA as cells move from progenitor to post-mitotic expression profile

Looking at lower magnification (all *MAZ* figures parts b-b*), we can see *MAZ* mRNA expression to be expressed across the whole section and expressed at all ages across the rostral-middle-caudal axis. For *MAZ* mRNA its expression pattern somewhat mimics that of the purple nuclear part of the

H&E stains above suggesting that its expression pattern mimics cell nuclear density patterns across the cortex.

Looking now at *MAZ* mRNA expression at higher magnification (all *MAZ* figures parts c-c*), we examined its expression across the cortex as shown by the boxes in parts b-b* and we used the H&E sections above as a guide for cortical lamination. We saw *MAZ* mRNA to be expressed in a very similar pattern to that of cell nuclear density; high in the VZ, lower in the SVZ, very low in the IZ and increasing again in the SP/CP. This is the same at all ages and across the rostral-middle-caudal axis and suggests that, as predicted by the scRNA-seq *MAZ* mRNA expression is relatively consistent across the cortex. The speckled staining pattern again suggests that *MAZ* mRNA is only expressed by some cells, matching the gradient plot shown in Figure 23. The lack of *MAZ* mRNA enrichment in progenitors as shown by the scRNA-seq analysis coupled with its very general staining pattern *in vivo* means it is not a good candidate for further protein studies into progenitor enriched genes from the *16p11.2* locus.

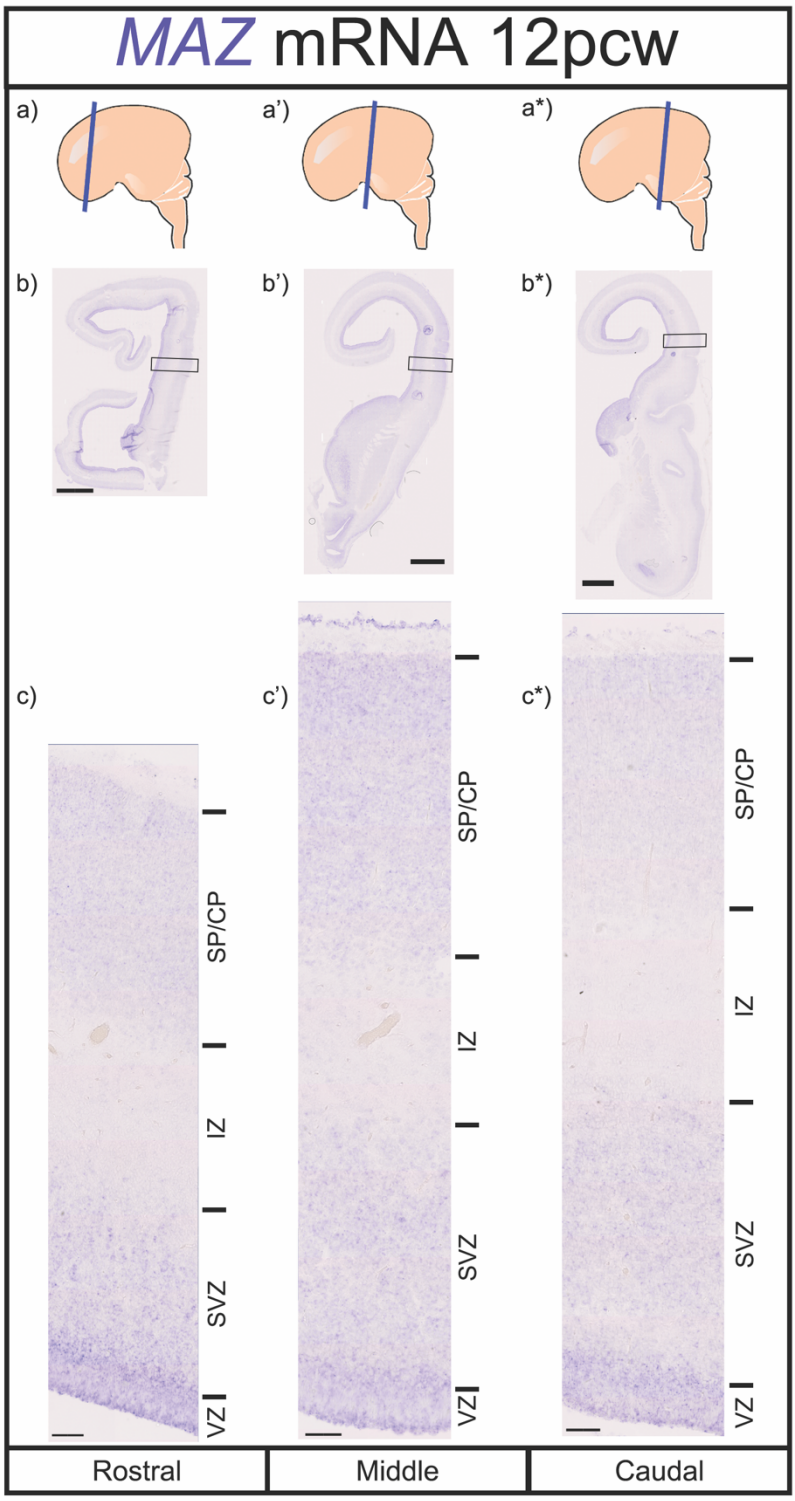


Figure 58: *in situ* hybridisation with *MAZ* probe at 12pcw

12pcw brain sections stained with a *MAZ* probe. a-a* show schematic of brain location taken for section. At low power (b-b*) scale bar = 2mm and high power (c-c*) scale bar = 100μm.

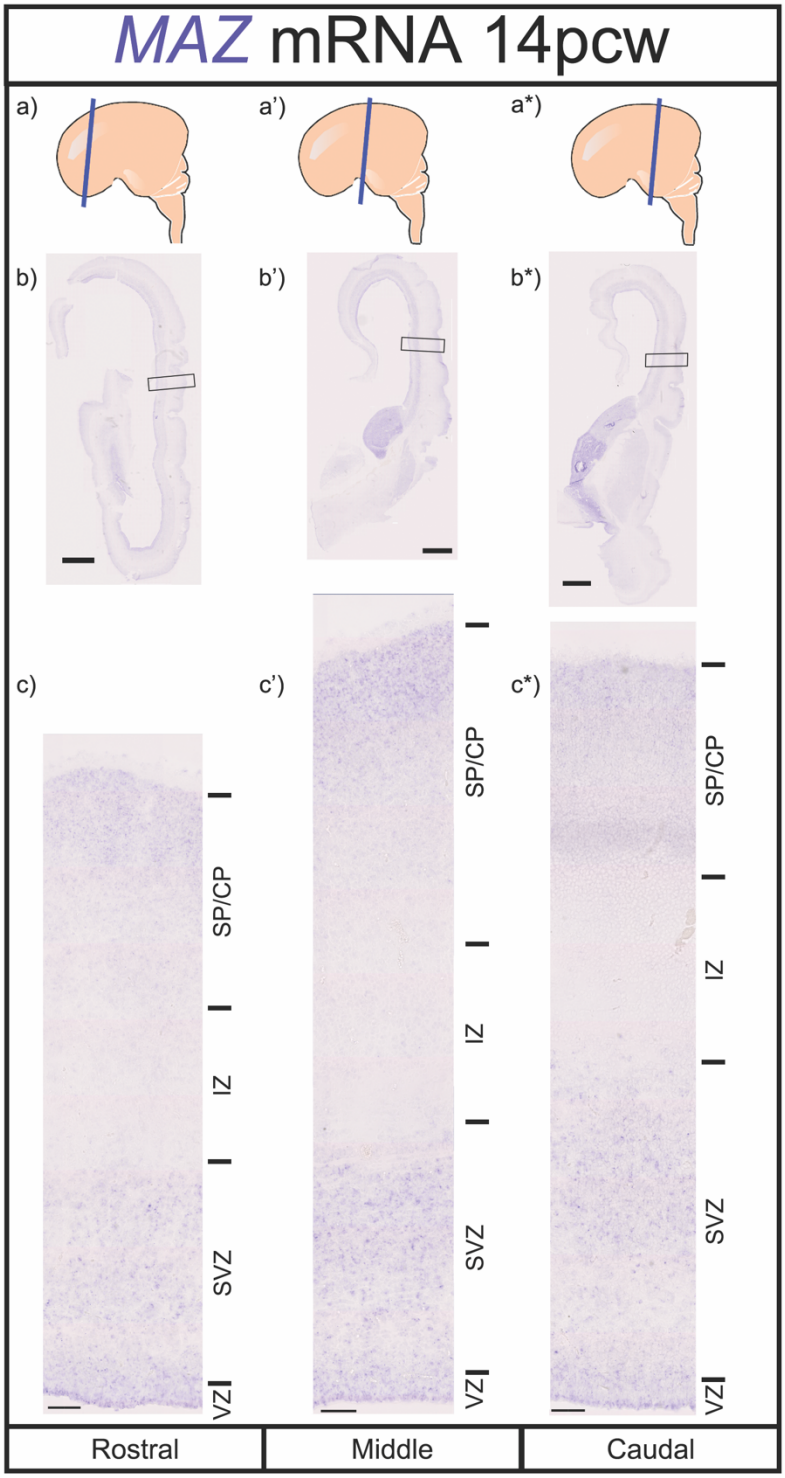


Figure 59: *in situ* hybridisation with *MAZ* probe at 14pcw

14pcw brain sections stained with a *MAZ* probe. a-a* show schematic of brain location taken for section. At low power (b-b*) scale bar = 2mm and high power (c-c*) scale bar = 100μm.

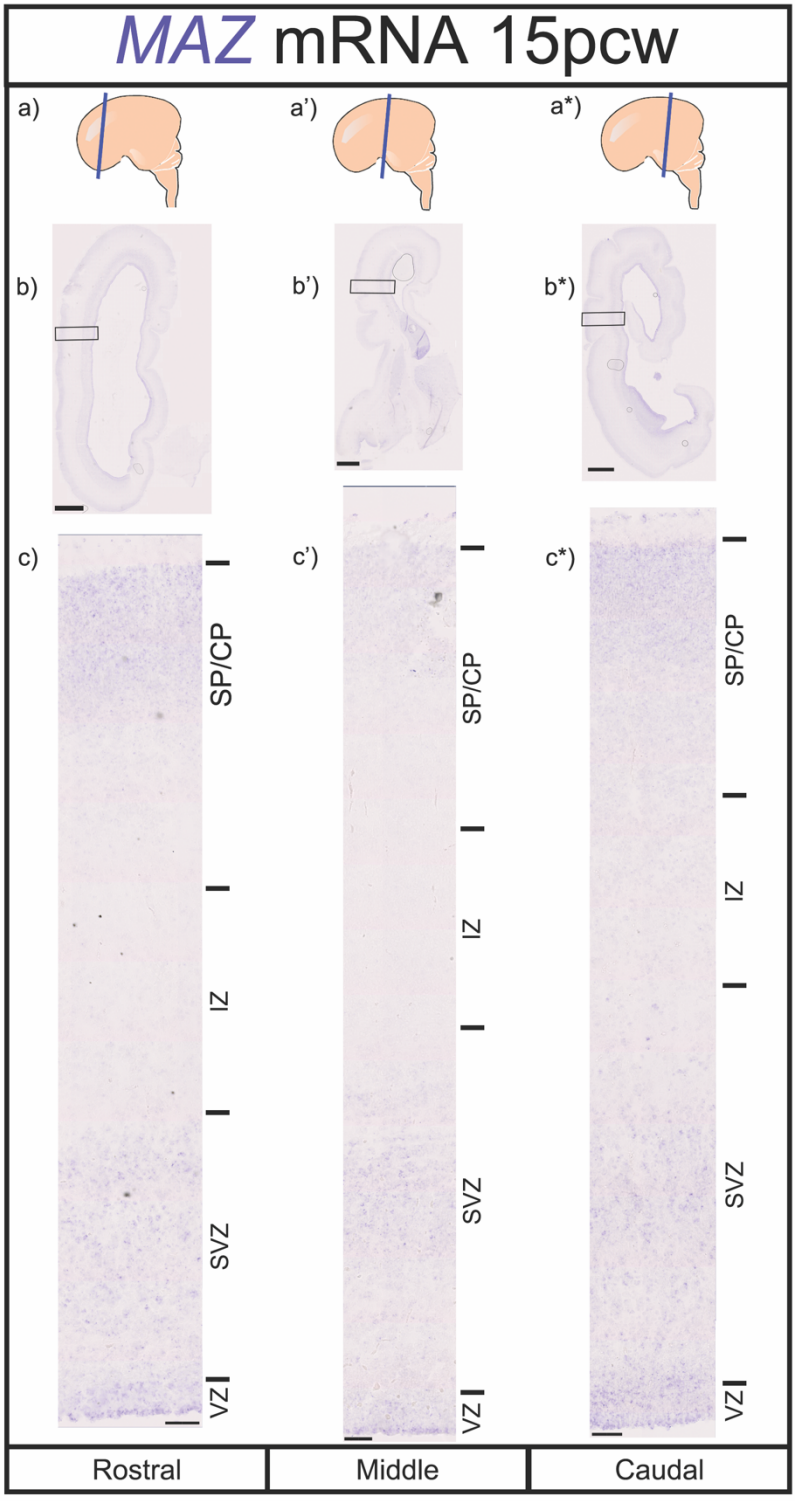


Figure 60: *in situ* hybridisation with *MAZ* probe at 15pcw

15pcw brain sections stained with a *MAZ* probe. a-a* show schematic of brain location taken for section. At low power (b-b*) scale bar = 2mm and high power (c-c*) scale bar = 100μm.

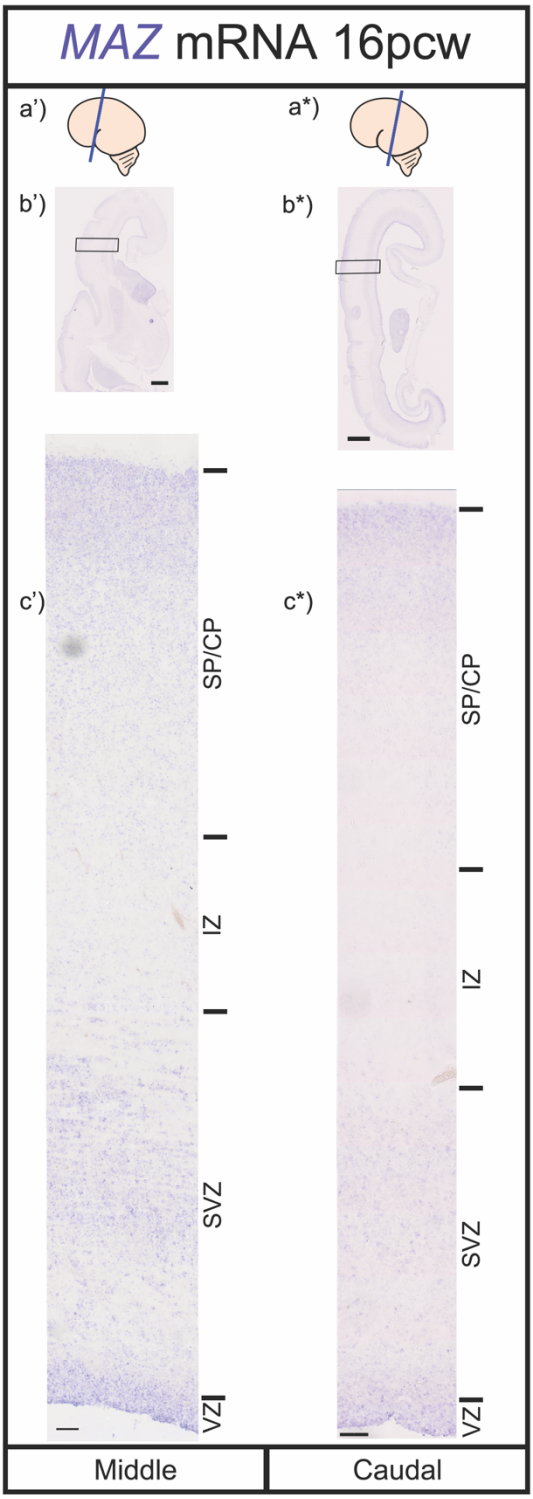


Figure 61: *in situ* hybridisation with *MAZ* probe at 16pcw

16pcw brain sections stained with a *MAZ* probe. a-a* show schematic of brain location taken for section. At low power (b-b*) scale bar = 2mm and high power (c-c*) scale bar = 100μm.

3.3.5.8 *SPN*

The final gene studied for *in situ* hybridisation was *SPN.* Unlike the other five genes, *SPN* was shown by the scRNA-seq analysis to be expressed in progenitor cells at similar levels to *HIRIP3, PAGR1* and *MAZ* and then increase as cells become post-mitotic as shown in Figure 62. Our *in situ* hybridisation staining produced staining for *SPN* mRNA allowing us to first confirm that it is expressed in human brain tissue at the ages studied: 12pcw (Figure 63), 14pcw (Figure 64), 15pcw (Figure 65) and 16pcw (Figure 66).

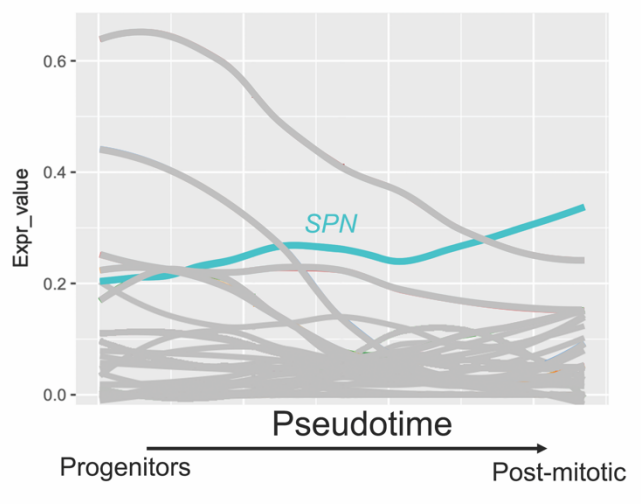


Figure 62: *SPN* mRNA expression through pseudotime

Pseudotime analysis as described above showing only *SPN* mRNA as cells move from progenitor to post-mitotic expression profile.

At lower magnification (all *SPN* figures parts b-b*) we can see *SPN* mRNA expression across the whole section and it is expressed at all ages studied across the rostral-middle-caudal axis. Most notable is the clear band of staining we can see around the outer edge of the cortex; this is particularly evident at 12pcw (Figure 63). In addition to this there is staining across the rest of the width of the cortex.

At higher magnification (all *SPN* figures part c-c*) we examined expression across the width of the cortex as shown by the boxes in parts b-b* and used H&E sections above as a guide to the cortical lamination. Here we saw staining that very closely matched that seen in the scRNA-seq analysis: *SPN* mRNA staining in all regions of the cortex, but particularly (especially when compared to stains such as *MAZ*) increased staining in the SP/CP. This is seen at all ages and across the rostral-middle-caudal axis. This mirrors our result from the scRNA-seq analysis and shows that *in vivo*, *SPN* is expressed in by cells in the proliferative regions and its level increases in the post-mitotic regions. This result means that *SPN* is not a good candidate for further protein studies into progenitor enriched genes in the *16p11.2* locus.

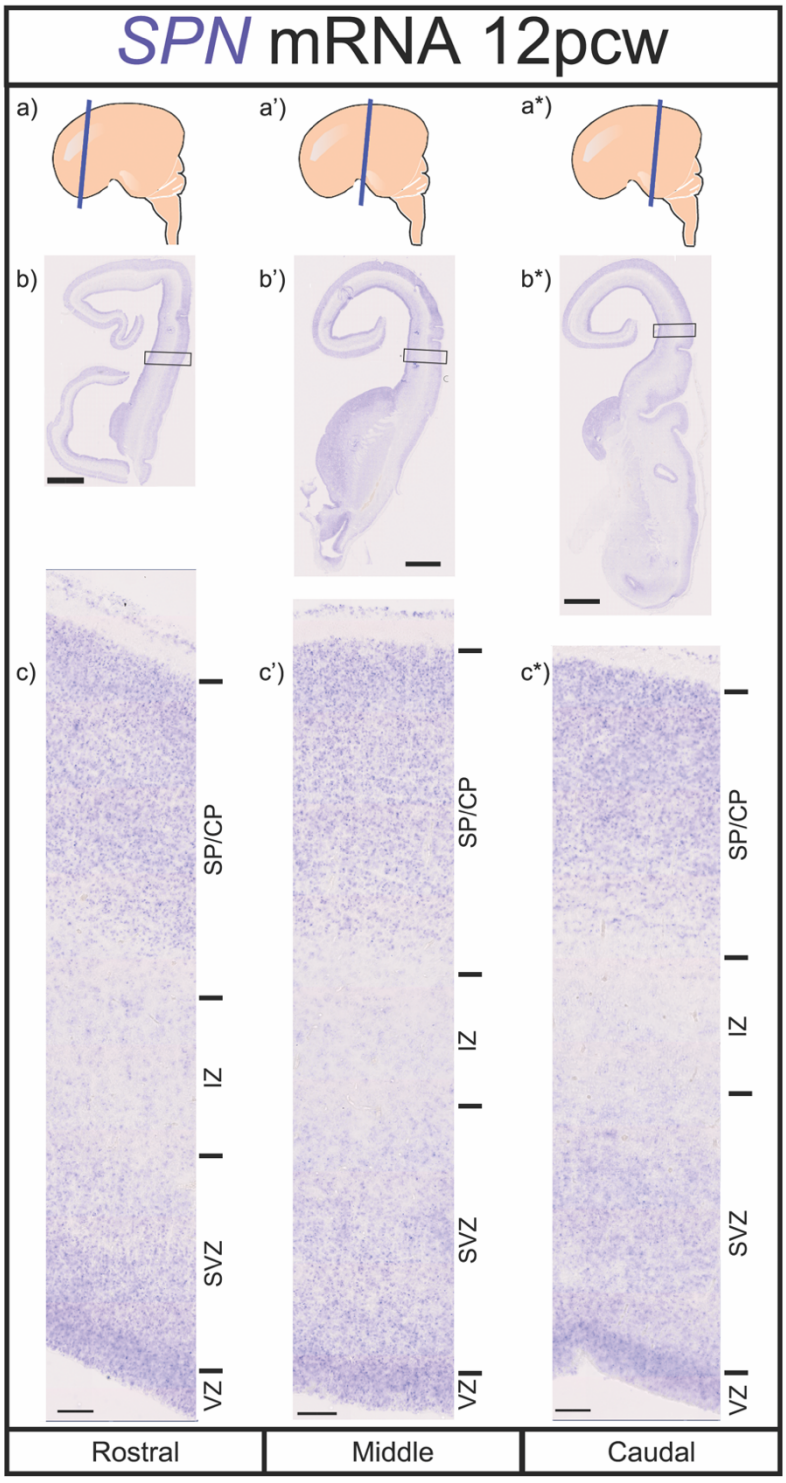


Figure 63: *in situ* hybridisation with *SPN* probe at 12pcw

12pcw brain sections stained with a *SPN* probe. a-a* show schematic of brain location taken for section. At low power (b-b*) scale bar = 2mm and high power (c-c*) scale bar = 100μm.

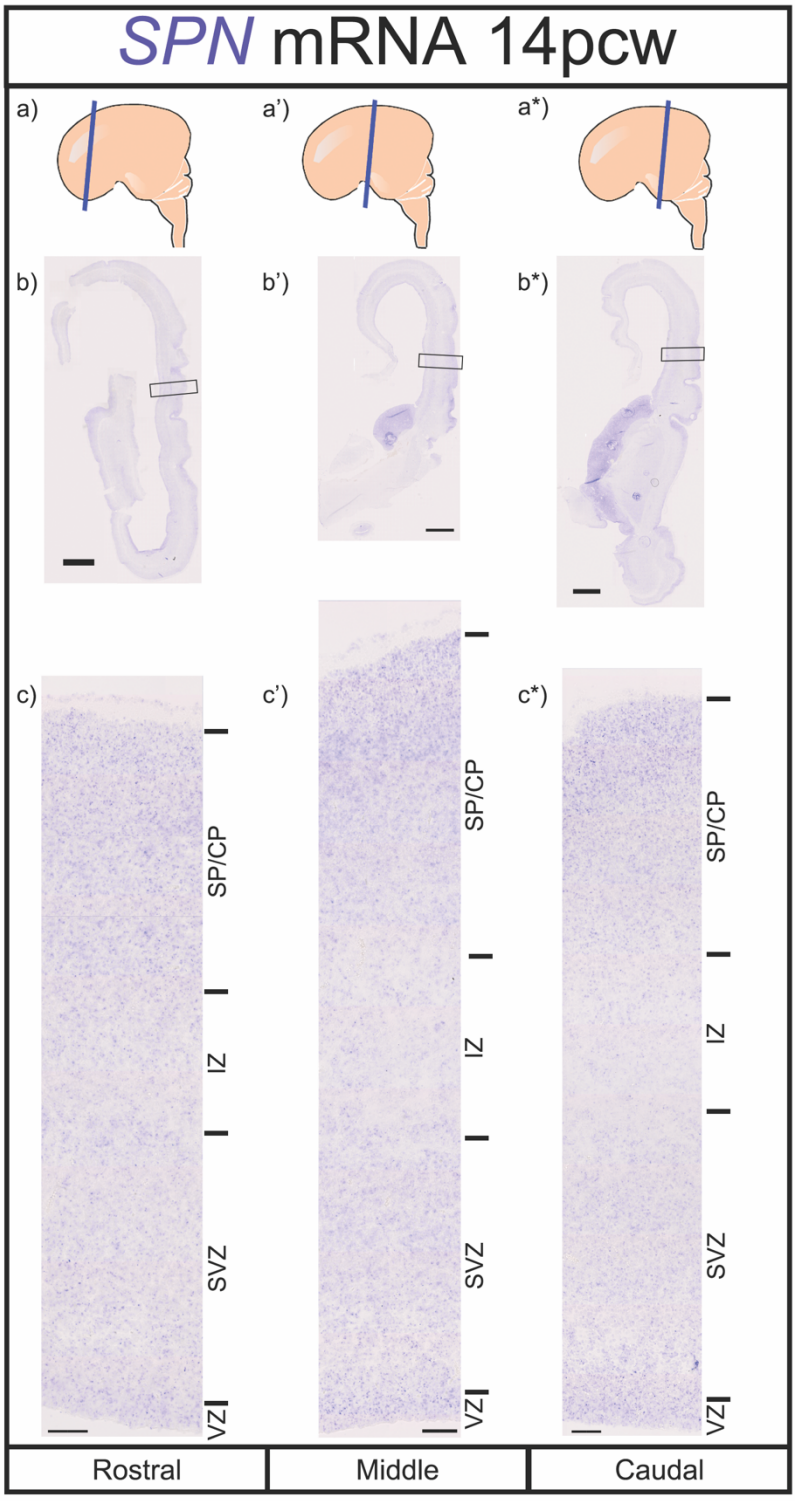


Figure 64: *in situ* hybridisation with *SPN* probe at 14pcw

14pcw brain sections stained with a *SPN* probe. a-a* show schematic of brain location taken for section. At low power (b-b*) scale bar = 2mm and high power (c-c*) scale bar = 100μm.

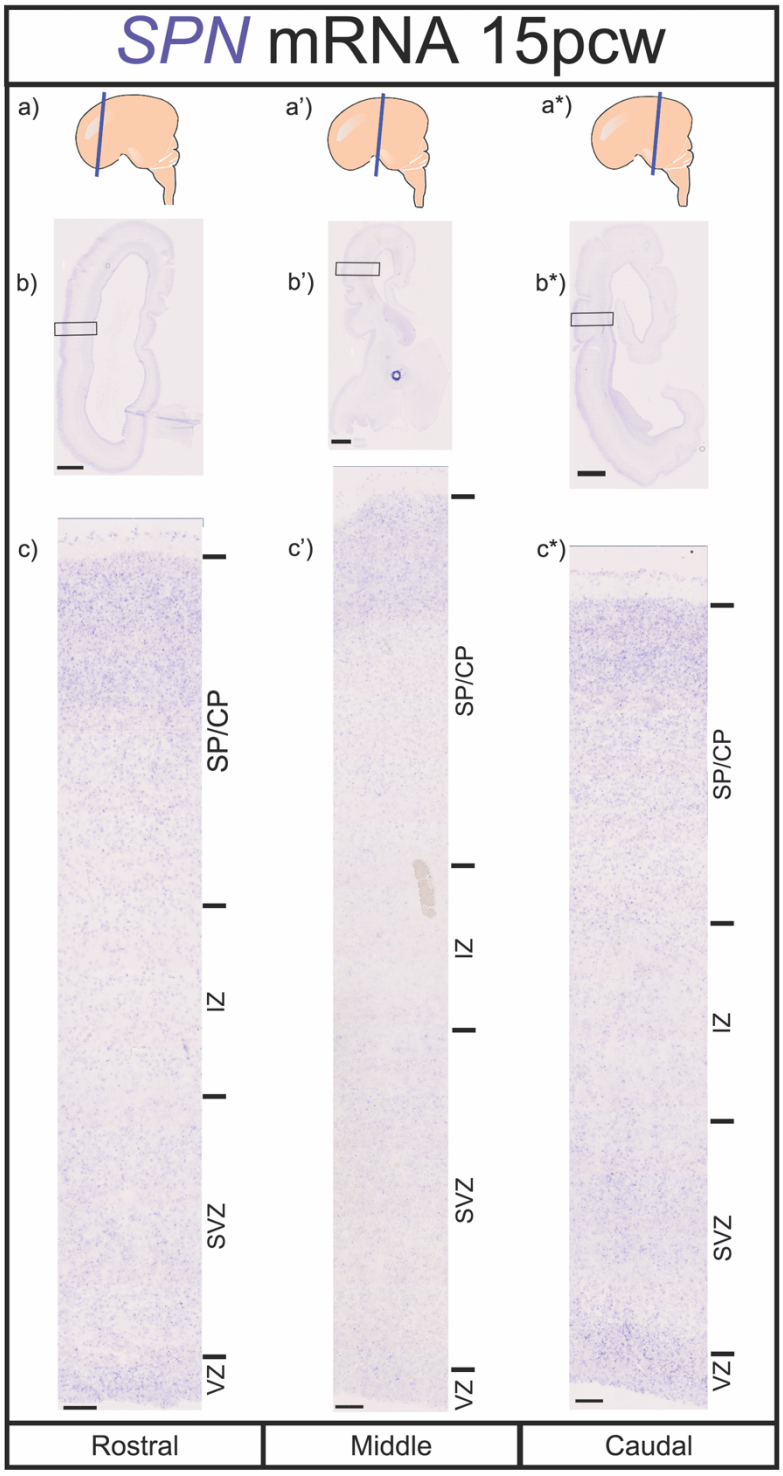


Figure 65: *in situ* hybridisation with *SPN* probe at 15pcw

15pcw brain sections stained with a *SPN* probe. a-a* show schematic of brain location taken for section. At low power (b-b*) scale bar = 2mm and high power (c-c*) scale bar = 100μm.

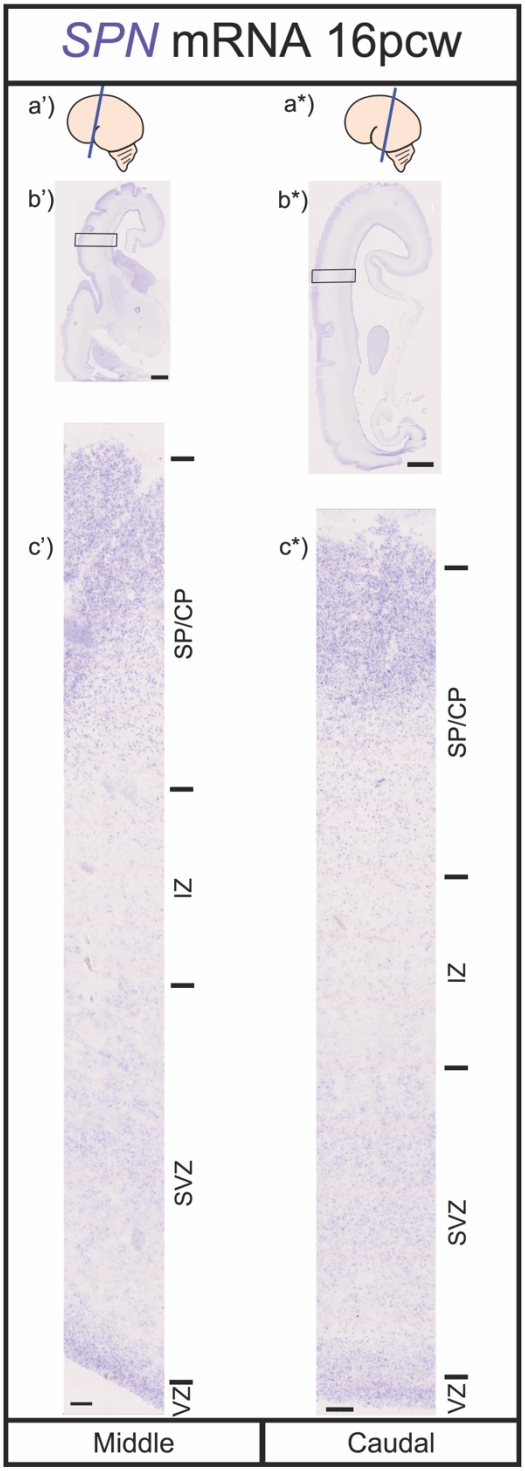


Figure 66: *in situ* hybridisation with *SPN* probe at 16pcw

16pcw brain sections stained with a *SPN* probe. a-a* show schematic of brain location taken for section. At low power (b-b*) scale bar = 2mm and high power (c-c*) scale bar = 100μm.

3.3.6 Selection of candidate genes for further studies

Based on the results from our unbiased screen using the scRNA-seq dataset to present six candidate genes from the *16p11.2* locus whose transcripts were expressed in progenitors, and the complimentary *in vivo* expression study of the mRNA of these six genes in human fetal cerebral cortex sections, we identified two candidate genes. These two candidate genes showed significantly increased mRNA expression in progenitors compared to post-mitotic cells in the scRNA-dataset, and high mRNA expression in proliferative regions *in vivo*. The two candidate genes used for future study were *ALDOA* and *KIF22.*

3.3.6.1 Cell cycle variability of target genes

We wanted to consider what these genes may be important for. Considering the literature that cell cycle disruption has been implicated in the *16p11.2* mouse model (Pucilowska *et al.*, 2015), and that disruption to cell cycle can alter brain size and architecture we wanted to see if mRNA expression of these genes changed with the cell cycle. Therefore, we returned to the scRNA-seq data, and this analysis was performed by Yifei Yang.

3.3.6.1.1 *KIF22* mRNA fluctuates with cell cycle

First, we asked if *KIF22* mRNA expression level was related to cell-cycle phase. We used the expression of cell-cycle phase specific transcripts as described in methods (3.2.2) to divide the cells into three classes, G1/S, G2/M and post-mitotic neurons. We then compared *KIF22* transcript levels between these three groups using a violin plot as shown in Figure 67. This analysis showed us that the majority of cells in G2/M phase expressed higher levels of *KIF22* mRNA (red plot), cells in G1/S expressed lower levels (blue plot) and the vast majority of post-mitotic cells expressed low levels of *KIF22* (green plot).

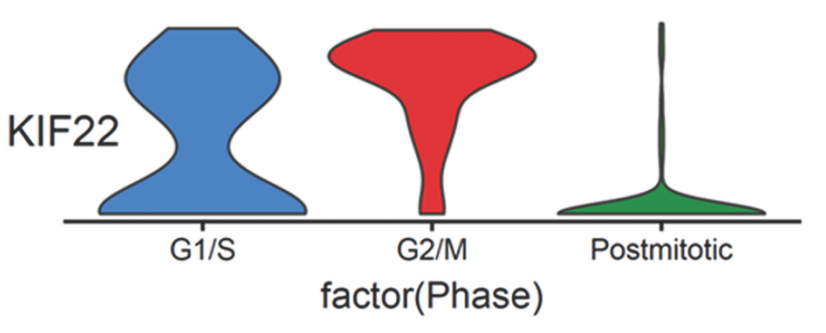


Figure 67: *KIF22* mRNA at different cell cycle stages

Violin plot showing *KIF22* mRNA distribution at three different cell cycle stages: G1/S, G2/M and postmitotic.

3.3.6.1.2 *ALDOA* mRNA does not fluctuate with cell cycle

We performed the same analysis for *ALDOA* transcripts. Comparison of *ALDOA* transcript levels between the three cell cycle state groups is shown in Figure 68. In contrast to *KIF22*, there is no clear difference in the partitioning of *ALDOA* mRNA expression level between different phases of the cell cycle.

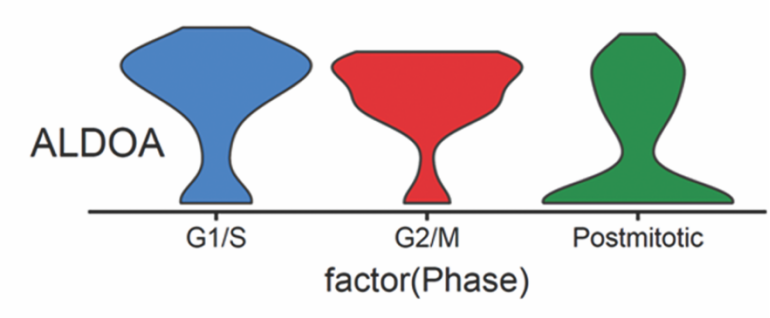


Figure 68: *ALDOA* mRNA at different cell cycle stages

Violin plot showing *ALDOA* mRNA distribution at three different cell cycle stages: G1/S, G2/M and postmitotic.

3.3.7 Conclusions of results

The results of this chapter are summarised here:

We performed an unbiased screen of previously published scRNA-seq data from the VZ and SVZ of human fetal cerebral cortex and identified six genes whose transcripts were expressed in progenitors. We then used *in situ* hybridisation using sections of human fetal cortex to examine expression patterns of these genes *in vivo*. The results of this are briefly summarised in Table 35.

From this we identified the two best candidates for further studies: *KIF22* and *ALDOA.* They were determined as the best candidates as their mRNA was highly expressed in progenitors and significantly downregulated as cells became post-mitotic suggesting a progenitor-specific role for these genes. *KIF22* mRNA was tightly restricted to germinative zones *in vivo* and we found it to vary with the cell cycle. *ALDOA* mRNA was not restricted to the germinative zones *in vivo*, however its expression was much stronger there than its expression in the post-mitotic IZ and SP/CP and its transcript did not vary with the cell cycle.

Table 35: Summary of results from scRNA-seq analysis and *in situ* hybridisation

Gene	scRNA-seq analysis	Expression *in vivo*	Candidate for protein studies?
KIF22	Significantly enriched in progenitors compared to post-mitotic cells	Present in the germinative VZ and SVZ, very low or absent from IZ and SP/CP	Yes
ALDOA	Significantly enriched in progenitors compared to post-mitotic cells, still expressed in post-mitotic cells but lower	Strong staining in the germinative VZ and SVZ, lower but substantial staining in the IZ and SP/CP	Yes
HIRIP3	Higher expression in progenitors than post-mitotic, but not significantly changed	Expressed throughout the width of the cortex, not especially high in germinative zones	No
PAGR1	Higher expression in progenitors than post-mitotic, but not significantly changed	Expressed throughout the width of the cortex, not especially high in germinative zones	No
MAZ	Very slightly higher expression in progenitors than post-mitotic, but not significantly changed	Expressed across the width of the cortex, staining pattern matches nuclear density of H&E staining	No

SPN	Expressed in progenitors and increases as cells become post-mitotic	Expressed across the width of the cortex, strong staining in the post-mitotic SP/CP region	No

3.4 Discussion

3.4.1 The *16p11.2* transcript expression during human neurogenesis

The *16p11.2* CNV is a polygenic mutation strongly associated with neurodevelopmental disorders including Autism Spectrum Disorder. This chapter has used an unbiased screen of existing human fetal scRNA-seq data identified a number of the 29 *16p11.2* genes as being expressed in progenitor cells of the cerebral cortex and confirmed their expression *in vivo* using *in situ* hybridisation. We identified *ALDOA* and *KIF22* as significantly enriched in progenitors and four other transcripts; *HIRIP3*, *PAGR1*, *MAZ* and *SPN* as also expressed in progenitors but at lower levels and not significantly down regulated as cells become post-mitotic. We have also confirmed this expression pattern *in vivo* to validate our two candidate choices as the best two for further studies.

This section of the chapter will discuss the results we obtained and the caveats of this work.

3.4.2 Analysis of scRNA-seq data

All technical analysis of the scRNA-seq dataset was performed by Yifei Yang. This used the previously published dataset generated from single cell RNA sequencing of micro-dissected VZ and SVZ regions from three 16-18GW human fetal cortices (Pollen *et al.*, 2015).

From this dataset we were able to group cells based on their transcripts to determine the expression levels of the *16p11.2* genes in each of the cell classes. This proved to be a powerful tool for producing candidates, allowing us to identify progenitor enriched genes.

When we were considering which human fetal scRNA-seq dataset to use for this study our decision to use the one generated by Pollen et al was made with the following considerations. In this study we are interested in proliferating cells, this cell class is found within the germinative zones of the developing brain, therefore using this dataset which uses cells exclusively from the germinative zones focuses our analysis. In addition, it uses ages available to us in the UK; 16-18GW (roughly equivalent to 14-16pcw), which allowed us to perform complementary *in vivo* experiments on tissue obtained here. One limitation of this dataset could be the low cell number: 393 cells, and it can be postulated that analysing datasets with more cells could produce a stronger result. However, our conclusions of this chapter do not rely solely on the results of the scRNA-seq analysis, we follow this analysis with an *in vivo in situ* hybridisation study providing two different approaches to identify our top two candidate genes.

3.4.3 Fetal Samples used

We used human fetal samples collected from the RIE (with the exception of the 12pcw sample which was obtained from HDBR in Newcastle) for the H&E stains and *in situ* hybridisations shown in this chapter. Screening of patients was performed by research midwife nurses as described in the methods, however in using this tissue we did make some assumptions. As there was no karyotyping of our samples, we had no way to know if the samples would have been affected by any neurodevelopmental disorders having proceeded to term. In particular there was no way to know if they could have been affected by the *16p11.2* CNV. Based on its prevalence at only 3 in 100,000 of the population we determined the risk of this to be negligible.

Another potential caveat is the lack of replicates for each age. This is a general problem associated with working with human tissue and one we encounter throughout this thesis. We are able to compensate for this to some

extent by comparing rostral-middle-caudal of each age and using a wide range of ages, and as this chapter is a purely qualitative expression study in this chapter it does not affect statistical power. Therefore, the tissue samples used were sufficient for this descriptive gene expression study.

3.4.4 Using H&E staining and nuclear density to determine cortical lamination

We used H&E staining on adjacent sections to those used for the *in situ* hybridisation to guide our determination of the cortical architecture. This allowed us a rudimentary guide to identify the different parts of the cortex based on their nuclear cell density. In addition, our IF stain for PAX6 protein allowed us to more clearly identify the VZ and SVZ, the regions of interest for this study. This method was sufficient for this as we were not looking to precisely identify individual regions although ideally, we would have used other region-specific markers to determine the different parts on adjacent sections. In addition for our indication of key anatomical locations at low magnification, many of these are approximate based on atlases (Bayer and Altman, 2002, 2005) and are only indicative. To be sure of the various regions we would need to stain with individual markers such as NKX2.1 for the medial ganglionic eminence, and *COUP-TFII* for the caudal ganglionic eminence. However, while we notice mRNA expression of some of our genes, particularly *KIF22* and *ALDOA*, their potential role in interneurons is outwith the scope of this thesis.

If possible we would have used layer specific markers for some of the regions that were more difficult to delineate by eye such as the boundary between the inner and outer SVZ could have been determined using a oSVZ specific marker such as *HOPX* (Pollen *et al.*, 2015). In addition it can be difficult to determine the boundary between the IZ and SP (and we combined the SP and CP for our data as these post-mitotic regions was not our focus),

we could have better identified these using *GAP-43* to identify the SP at our time-points of interest and *MAP2* to determine the cortical plate (Bayatti *et al.*, 2008). However, as this thesis is focused primarily on the germinative zones, the VZ and SVZ, this does not significantly impact our findings.

One issue we encountered when staining sections for H&E was the fragility of the cryostat sections, particularly the highly nuclear dense VZ. Despite secondary fixing in FAA as described in methods we noticed large amounts of damage to the sections, however this was not detrimental enough to prevent identifying individual cortical regions.

3.4.5 *In situ* hybridisation results

Our *in situ* hybridisation data provides a detailed atlas of the mRNA expression of our six candidate genes across key developmental timepoints. It validates the candidates selected by our unbiased screen of the scRNA-seq data and allowed us to identify the top two candidate genes for further studies. Having identified two genes: *ALDOA* and *KIF22* as having mRNA enriched in progenitors and being significantly downregulated as cells become postmitotic, and four other genes *HIRIP3, PAGR1, MAZ,* and *SPN* as having mRNA expressed in progenitors but at lower levels and not significantly down regulated as cells became post-mitotic. Our *in vivo* results confirmed that, for our study looking to identify *16p11.2* locus progenitor enriched transcripts, *HIRIP3, PAGR1, MAZ* and *SPN* while their mRNA were expressed in the progenitor regions of the brain, it was not at a particularly high level compared to other regions excluding them as candidates for this study and therefore they will not be discussed further in this thesis.

3.4.6 Candidates for further study

Out of our unbiased bioinformatics screen and follow-up *in situ* hybridisation, we identified two strong candidate genes from the *16p11.2* locus for further study: *ALDOA* and *KIF22*. These two genes emerging as our study genes is particularly interesting, as they were previously identified as the only two

genes from the *16p11.2* locus to induce head size changes in zebrafish when their mRNA levels were reduced 50% as in the *16p11.2* CNV patients (Blaker-Lee *et al.*, 2012). Therefore, there is already some evidence for these two genes as important in proliferation and brain growth during neural development, and altering their dosage affects neural development, albeit in the slightly smaller and simpler zebrafish brain.

It is important to consider that while we can learn a lot from mRNA expression, it is the protein that exerts effect in the cell. Therefore, having achieved the aim of this chapter to identify progenitor enriched genes from the *16p11.2* locus, in the next chapters of this thesis we will investigate *ALDOA* and *KIF22* at the protein level.

4 KIF22 protein expression is restricted to proliferating cells and varies with the cell cycle in human fetal corticogenesis

4.1 Introduction

From our work in chapter three we identified two candidate genes from the *16p11.2* locus whose transcripts were enriched in progenitors and downregulated in post-mitotic cells: *KIF22* and *ALDOA*. These progenitor-enriched genes were of particular interest to us as they are likely to specifically affect neurogenesis. In this chapter we will study further the first of these two genes *KIF22*.

From the scRNA-seq analysis we can summarise using a violin plot (Figure 69) the numbers of cells expressing different levels of *KIF22* mRNA in the different cardinal cell classes to show that *KIF22* is expressed predominantly in progenitors.

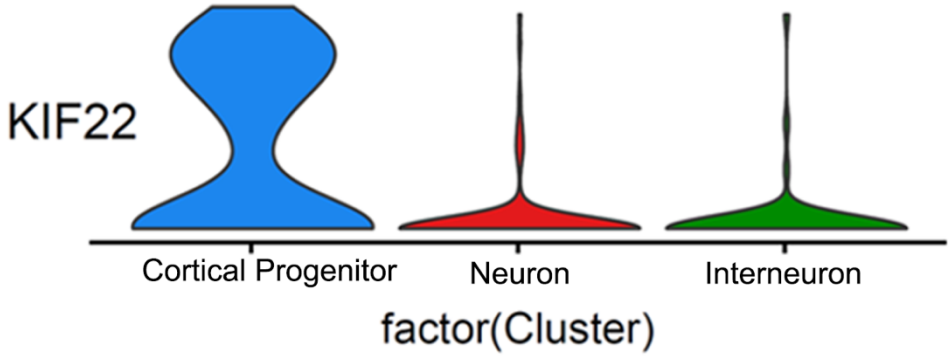


Figure 69: *KIF22* mRNA expression

Violin plot showing the distribution of *KIF22* in different cells types. Cortical progenitors = blue, post-mitotic neurons = red and interneurons = green.

We next wanted to study KIF22 protein expression during human cerebral cortex development to understand more about its role in neurogenesis. In this introduction we will briefly summarise what is known about the KIF22 protein from the literature and outline the aims of this chapter.

4.1.1 Kinesin family

The *KIF22* gene encodes a protein which is a member of the kinesin-like protein family (Tokai *et al.*, 1996). The first of the kinesin family (Kinesin-1) was discovered from the giant axon of the squid in 1985 (Allen *et al.*, 1982; Brady, Lasek and Allen, 1982; Vale, Reese and Sheetz, 1985). This protein, when isolated, was found to comprise of two main parts: two identical heavy chains which act as the motor and associate with the second part; two light chains responsible for binding the motor to the cargo for transportation (Vale, Reese and Sheetz, 1985). Since the discovery of this first kinesin a whole superfamily of kinesins has been identified comprising 15 families and organised into three main classes (Hirokawa *et al.*, 2009):

1) N-kinesins with their motor domain in the amino terminal region which drive microtubule plus end-directed motilities.
2) C-kinesins with their motor domain in the carboxy-terminal region which drive microtubule minus end-directed motilities.
3) M-kinesins with their motor domain in the middle which depolymerize microtubules.

There are a vast number of kinesin genes identified since the discovery of the first, the families of all 45 mouse kinesin genes are shown in Table 36.

Table 36: Kinesins ordered by family

Family	Class	Members
Kinesin 1	N-Kinesin	KIF5A KIF5B KIF5C
Kinesin 2	N-Kinesin	KIF3A KIF3B KIF3C KIF17
Kinesin 3	N-Kinesin	KIF1A KIF1B KIF1C KIF13A KIF13B KIF14 KIF16A KIF16B
Kinesin 4	N-Kinesin	KIF4A KIF4B KIF7 KIF21A KIF21B KIF27
Kinesin 5	N-Kinesin	KIF11
Kinesin 6	N-Kinesin	KIF20A KIF20B KIF23
Kinesin 7	N-Kinesin	KIF10
Kinesin 8	N-Kinesin	KIF18A KIF18B KIF19A KIF19B

Kinesin 9	N-Kinesin	KIF6 KIF9
Kinesin 10	N-Kinesin	KIF22
Kinesin 11	N-Kinesin	KIF26A KIF26B
Kinesin 12	N-Kinesin	KIF12 KIF15
Kinesin 13	M-Kinesin	KIF2A KIF2B KIF2C KIF24
Kinesin 14A	C-Kinesin	KIFC1
Kinesin 14B	C-Kinesin	KIFC2 KIFC3 KIF25

4.1.1.1 Kinesin Structure

While there is some variability in structure between members of the kinesin superfamily, there is conservation in the overall structure which is shown in Figure 70. The prototypical form of the kinesin-1 protein was identified in 1989 from electron microscopy studies of bovine brain kinesin. It was shown to comprise of two identical heads (which act as the motor domain) which each connect to a stalk, a long alpha helix coiled coil domain. The coiled coil ends in a tail which associates with the light chain which binds to the cargo. Each of the coiled coils of each head intertwine to direct dimerization of the heavy chain (Hirokawa *et al.*, 1989).

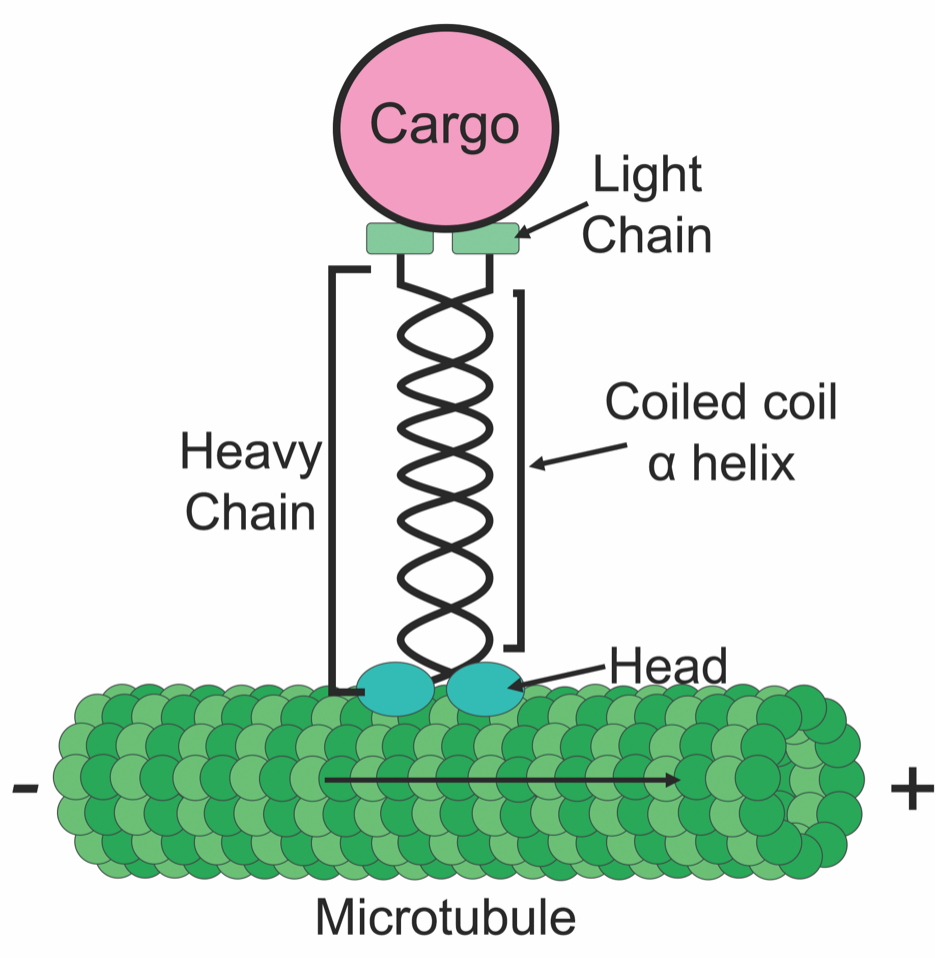


Figure 70: Schematic of a kinesin protein

Diagram showing a N-kinesin. Arrow on microtubule indicates direction of travel towards plus end of microtubule. The heavy (124 kDa) chain comprises the two identical polypeptide chains dimerize to form the coiled coil α helix stalk (the neck) and the globular heads which act as the motor domain which hydrolyses ATP to move the protein along the microtubules. The light (64 kDa) chain binds to the cargo for transport along the microtubule.

4.1.1.2 Kinesin movement

Kinesin protein move along the microtubule by “walking”. The exact details of this movement are unknown, however there are two current leading theories: The “hand-over-hand” mechanism and the “inchworm” mechanism, these are shown in Figure 71 (based on diagram in (Hua, Chung and Gelles, 2002)). While it is unknown exactly by which method the kinesin protein walks, mounting evidence suggests the “hand-over-hand” mechanism as the most probable (Yildiz *et al.*, 2004).

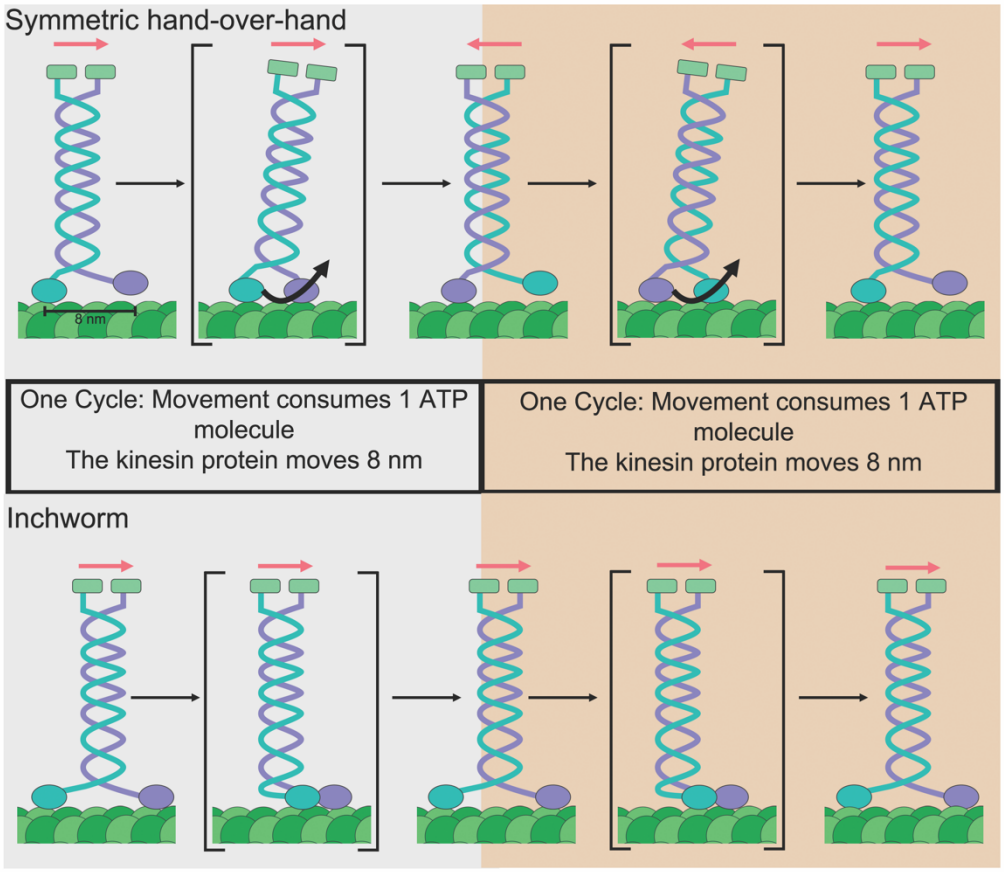


Figure 71: Proposed mechanisms of kinesin movement

There are two proposed mechanisms of kinesin movement:

Symmetric hand-over-hand: in this mechanism the heads alternately move past each other. One head is tightly bound to the microtubule while the other uses 1 ATP molecule to move forward (black arrow). The coiled domain moves 180º around its axis each step (pink arrow). This is the favoured proposed mechanism.

Inchworm: In this mechanism the leading head always stays in the lead (here the purple head) and the heads never swap places. The coiled domain does not rotate (pink arrow). This mechanism would also use 1 ATP molecule per cycle of movement.

4.1.2 KIF22

Our kinesin of interest for this thesis is KIF22, the kinesin-like microtubule-based motor which binds to chromosomes and microtubules throughout mitosis. The KIF22 protein was first identified in 1996 by Tokai et al (Tokai *et al.*, 1996) as a **ki**nesin-like **D**NA-binding protein in humans, hence its protein's alternative name as KID (throughout this thesis it is referred to only as KIF22). Over the years, KIF22 has proved itself a multi-functional protein, with many domains that are proposed to facilitate a wide range of its roles. These are shown in Figure 72, and its roles can be separated into two main classes: its functions in mediating the mitotic spindle as a microtubule motor, and its other transcriptional roles. It is known that KIF22 has many roles in mitosis through its role as a microtubule-based motor protein, however it also has other, transcriptional, roles which make it interesting to study its expression during human fetal cortex development.

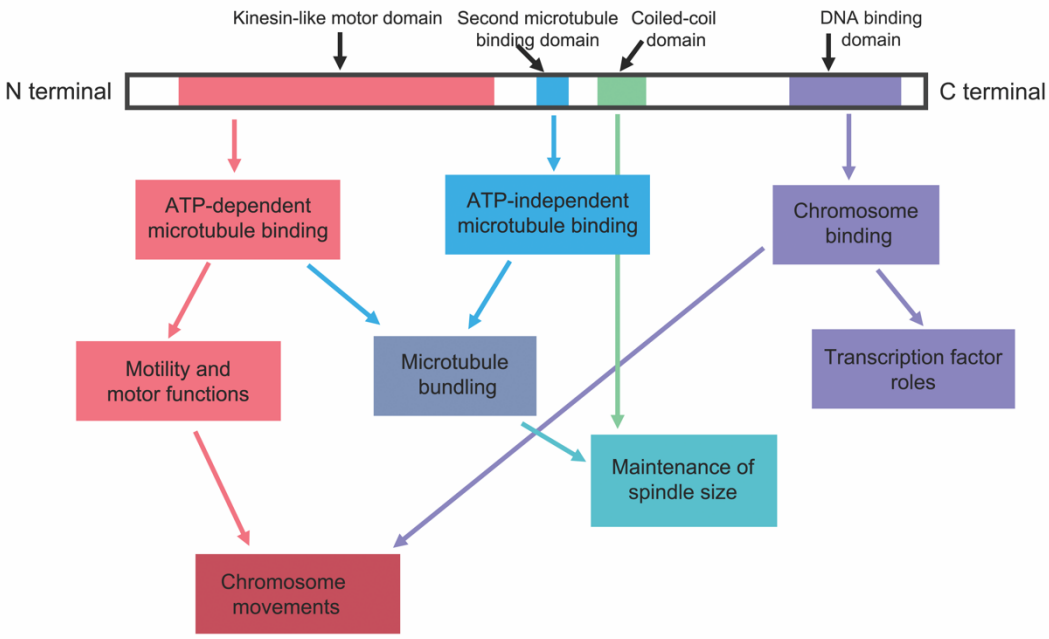


Figure 72: KIF22 domains and associated functions

The motor domain (pink) harbours an ATP loop allowing for ATP hydrolysis and is also capable of bundling microtubules

The second microtubule binding domain (blue) does not require ATP for bundling microtubules.

The coiled-coil domain (green) plays a role in maintaining the spindle

The DNA binding domain (purple) allows for chromosome binding and some transcriptional repressor roles such as repression of CDC25C.

It is shown that some of KIF22 cellular functions require more than one domain e.g. spindle size maintenance involves both the coiled coil domain and the second microtubule binding domain.

4.1.3 KIF22 as a microtubule motor

Originally identified as localised to chromosomes during mitosis (Tokai *et al.*, 1996), in 2003 three papers by the labs of Toyoshima and Yamamoto provided greater insight to this protein (Ohsugi *et al.*, 2003; Shiroguchi *et al.*, 2003; Yajima *et al.*, 2003) including its role as a plus end-directed microtubule-based motor. KIF22 is present throughout the entirety of mitosis, although its location varies at different points as shown in Table 37. The modulation of the many KIF22 functions and locations has been shown to be regulated by CDK1-mediated phosphorylation of Thr463 amino acid (Ohsugi *et al.*, 2003). It has been previously shown that, in order for KIF22 to switch between its opposing motor actions on chromosomes during the metaphase-anaphase transition, an alteration of KIF22's microtubule and chromosomal affinities must happen. This happens because CDK1-CyclinB dependent phosphorylation of Thr463 activates the second microtubule binding domain and suspends generation of the polar ejection force as cells enter anaphase (Soeda, Yamada-Nomoto and Ohsugi, 2016).

The known function of KIF22 protein in mediating spindle microtubule stability (Tokai-Nishizumi *et al.*, 2005) is of particular interest to our work. The mitotic spindle is known to control whether a cell undergoes a symmetric or asymmetrical division, therefore contributing to maintaining the balance of proliferation and differentiation in neural progenitors (Sun and Hevner, 2014).

Table 37: Localisation of KIF22 throughout the cell cycle

Point in cell cycle	Location	Purpose if known	Reference
Interphase	Nucleus	unknown	(Tokai *et al.*, 1996)
Prometaphase	Spindles and the length of chromosomes	unknown	(Tokai *et al.*, 1996)
Metaphase	Metaphase plate	To provide force to move chromosome arms to the spindle equator (polar ejection force)	(Funabiki and Murray, 2000)
Anaphase	Moves with the chromosomes to the spindle poles	Involved in chromosome compaction	(Ohsugi *et al.*, 2008)
Telophase and cytokinesis	Spindle pole-proximal side of the chromosomes	unknown	(Shiroguchi *et al.*, 2003)

4.1.4 The transcriptional role of KIF22

As its second role which involved transcriptional regulation, KIF22 protein acts to regulates the expression of the cell-cycle regulator CDC25C. During the cell cycle, in order for cells to enter mitosis they require to reach a threshold level of activated cell cycle regulator proteins; CDK1-CyclinB. This complex accumulates during G2 phase, but it is kept inactive as it is phosphorylated. As cells begin to transition from G2 to M phase, the phosphatase cell cycle regulator protein CDC25C is activated, which then is able to dephosphorylate the CDK1-CyclinB complex and render it active allowing the cell to enter mitosis. KIF22 is involved in regulating this process of driving the cell into mitosis through its transcriptional regulation of CDC25C.

Evidence from a study in cancer tissue in 2014 helped to reveal the feedback loop by which KIF22 and the CDK1-CyclinB complex are linked (Yu *et al.*, 2014). They showed that inhibition of KIF22 supresses cancer cell proliferation and reducing KIF22 levels caused cells to exhibit a faster S to G2/M progression but to have a slower mitotic exit (Yu *et al.*, 2014). In their model; KIF22 transcriptionally and translationally represses CDC25C, a primary driver of mitotic entry (Bulavin *et al.*, 2003), and the inhibition of KIF22 increased CDK1 activity by transcriptionally upregulating CDC25C expression as shown in Figure 73 (adapted from (Yu *et al.*, 2014)).

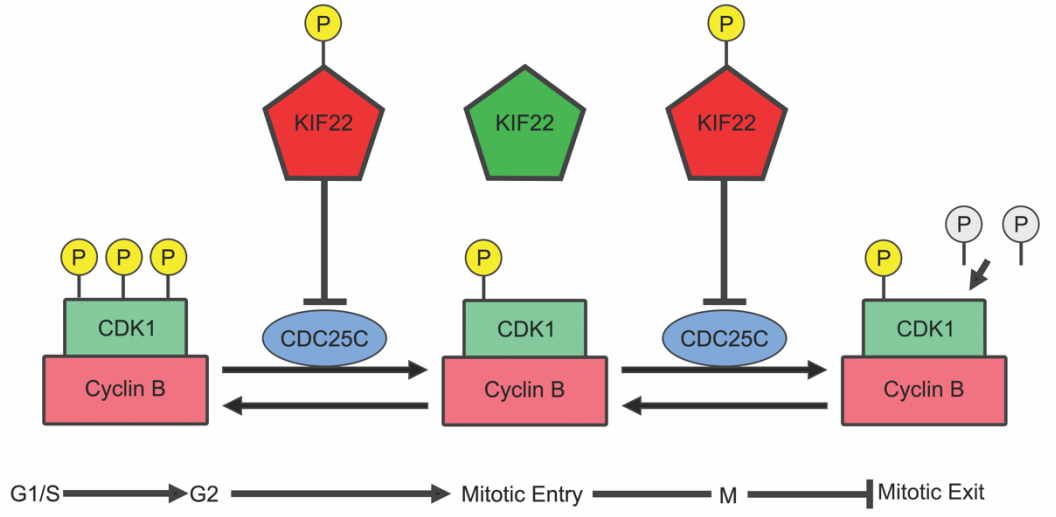


Figure 73: Schematic of KIF22's transcriptional regulation of cell cycle regulators

Schematic showing the result from cancer cell lines that evidenced a transcriptional regulatory role for KIF22 in the cell cycle. Here during the transition from G1/S phase to G2 phase the CDK1-CyclinB complex is highly phosphorylated by CDC25C which allows it to phosphorylate KIF22 which transcriptionally represses CDC25C, a primary mitotic entry driver. This repression results in the dephosphorylation of the CDK-CyclinB complex and subsequently KIF22 dephosphorylation which no longer represses CDC25C allowing for mitotic entry. As CDC25C is no longer repressed, the CDK1-CyclinB complex phosphorylates CDC25C enhancing its phosphatase activity creating a feedback loop to drive cells to mitosis. However, this activated CDK1-CyclinB complex also phosphorylates KIF22 allowing it to transcriptionally repress CDC25C reducing CDK1-CyclinB activity and triggering mitotic exit.

4.1.5 KIF22 as a cell cycle regulator

The multiple roles of the KIF22 protein in cell division, transcriptional and as a motor protein, make it an intriguing candidate: one of the key determinants of NPCs undergoing a neurogenic vs proliferative division is increased cell cycle length. Most of the knowledge about KIF22's functions are from studies in cell culture systems or cancer tissues, and nothing is known about its roles in neural development. While a KIF22 depleted mouse has been previously generated, no investigation has been performed into any changes to neural development in these mice and half died in utero (Ohsugi *et al.*, 2008).

As there is much previous evidence as to KIF22's importance in the cell cycle and mitosis, it can be hypothesised that any alteration to its normal level by changes to gene dosage has potential to disrupt neural progenitor proliferation and differentiation. However, as nothing is currently known about KIF22 in human neural development it is important to first describe its normal expression in the human fetal cerebral cortex.

4.1.6 Chapter Aims

This chapter has the following aims:

1) To characterise the anatomical expression pattern of the KIF22 protein during human fetal cortical development using high-level immunohistochemistry.
2) To establish whether the restriction of *KIF22* mRNA to progenitor cells is also true for the KIF22 protein.
3) To further investigate whether, like *KIF22* mRNA, KIF22 protein levels vary with cell cycle phase.

4.2 Methods

4.2.1 Fetal tissue used

The fetal tissue for this chapter was paraffin embedded tissue from the HDBR. Tissue was used for both colourmetric immunohistochemistry (IHC) and immunofluorescence (IF). The ages used are shown in Table 38. Details of tissue preparation is provided in 2.1.1.

Table 38: Fetal ages used in this chapter

Age	Rostral	Middle	Caudal
12pcw sample one	IHC and IF	IHC and IF	IHC and IF
12pcw sample two	IF	IF	IF
14pcw	IHC and IF	IHC and IF	IHC and IF
16pcw	IHC	IHC	IHC

4.2.2 Colourmetric immunohistochemistry of KIF22

The colourmetric immunohistochemistry (referred to as IHC hereafter) was performed as described in 2.3.2. The antibody used for this IHC stain was the KIF22/KID polyclonal antibody anti-KID (Invitrogen #PA5-29490) at a concentration determined to optimal of 1:5000. Sections where then counterstained with nuclear fast red (NFR) to produce a pink stain in cell nuclei.

4.2.2.1 Cell Counting for immunohistochemistry

Cell counts were performed manually in FIJI with the help of undergraduate honours student Erin Boyle. Having taken individual high magnification images across the width of the cortex as described in the methods chapter, we used FIJI software to stitch them together using the stitching plugin (Preibisch, Saalfeld and Tomancak, 2009). Having stitched the images, we overlaid a grid of regular counting boxes (34x88μm) onto the images extending across the entire width as shown in Figure 74a and aligned with the ventricular edge. To ensure cells were not double counted, we used an inclusion/exclusion system to systematically determine if cells on the borders of grids should be included or not as shown in Figure 74b. We used the ImageJ cell counting plugin to count cells which were denoted as either $KIF22^{+}$ (brown) or $KIF22^{-}$ (red) with examples shown in Figure 74b, green arrows indicating $KIF22^{+}$ cells and blue arrows indicating $KIF22^{-}$ cells. The distinction between the regions (VZ, SVZ, IZ and SP/CP) was determined anatomically by cell nuclear density. The count for each box was averaged with other boxes in the region to provide the final value. The size and number of bins used was optimised to make sure to include the ventricular edge, using larger and fewer boxes excluded this detail.

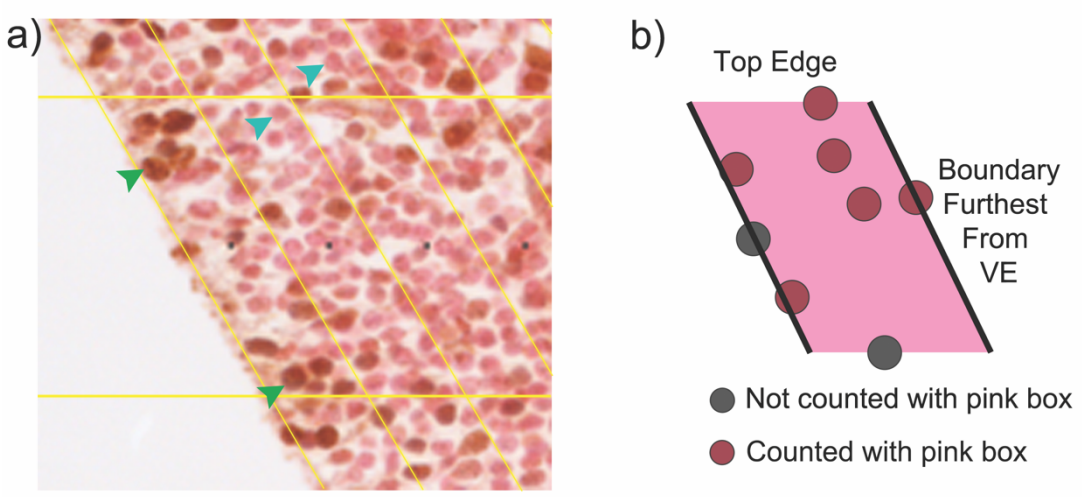


Figure 74: Counting method for KIF22 IHC

A = showing the grid overlaid across the cortical width for counting aligned with the ventricular edge. Example $KIF22^{+}$ cells are indicated by green arrow and example $KIF22^{-}$ cells are indicated by blue arrows.

B= inclusion/exclusion criteria for cells on the boundaries of grid boxes. Cell nuclei which has >50% of their stain in the a box (pink cells) were included for that box count, where there was not a clear box and the cell nuclei sat on a boundary line , cells on the top edge were included (pink cells) and on the bottom edge were not included (grey cells), cells spanning the boundary furthest from the VE were included in the box.

4.2.2.2 Data analysis for IHC

First, we identified which boxes corresponded to which region based on cellular nuclear density of the nuclear fast red (NFR) counterstain. The ventricular edge was considered to be the first box of the counting grid. The $KIF22^{+}$ and $KIF22^{-}$ counts for each box were averaged for the respective regions and displayed as graphs showing both the percentage of KIF22 expression cells and the numbers of both $KIF22^{+}$ and $KIF22^{-}$ cells. To combine our data, we presented the data two ways: first by averaging the values for each age but keeping the rostral-middle-caudal values separate, and secondly by averaging the values for rostral-middle-caudal and keeping the ages separate. No statistical tests were performed on the IHC data as one fetal sample was used for each age.

4.2.3 Immunofluorescence of KIF22 and KI67

For our immunofluorescence study of KIF22 the tissue from HDBR was again used and the IF protocol performed as described in the methods. The antibodies used for this part are shown in Table 39.

Table 39: Antibodies used for IF

Antibody	Concentration	Vendor	Secondary Antibody
KIF22 rabbit polyclonal	1:2000	Invitrogen	Alexa Fluor Goat anti-rabbit 568
KI67 mouse IgG1 monoclonal	1:750	Novus Biologicals	Alexa Fluor Goat anti-mouse 488 IgG1

4.2.3.1 Cell Counting for immunofluorescence

For the IF study of KIF22 we only performed high-magnification imaging of tissue from the germinative zones: the VZ and SVZ. The images were then stitched together in FIJI and a counting grid (20x145µm per box) overlaid onto the image in the same style as the IHC. For cell counts performed in FIJI, the same inclusion/exclusion criteria were used as for IHC. For determining intensity, cells were randomly selected on the DAPI channel, the nucleus outlined and the fluorescent intensity of KIF22 and KI67 recorded. The counts from the individual boxes were then combined to give final values.

4.2.3.2 Data analysis and statistics for IF

First, we determined cells as expressing KI67, KIF22, both or neither. We performed this individually for 12pcw (two brains were used, and the results kept independent) and 14pcw (one brain used) with individual rostral-middle-caudal sections. As we saw no difference in expression quantification between the ages nor anatomical locations, we then pooled the data. We did this by averaging the rostral-middle-caudal values for each age to use as a single n value and treating the two 12pcw and one 14pcw values to give n=3 as explained in Table 40 to perform an ANOVA with a post-hoc test to determine statistical significance.

Table 40: Summary of data handling

Sample	n-value
12pcw brain one - Rostral	First n value
12pcw brain one - Middle	
12pcw brain one - Caudal	
12pcw brain two - Rostral	Second n value
12pcw brain two - Middle	
12pcw brain two - Caudal	
14pcw brain - Rostral	Third n value
14pcw brain - Middle	
14pcw brain - Caudal	

Next, we examined KIF22 and KI67 levels by recording fluorescent intensity as described above. Cells were determined as being $KI67^-$ if their expression level fell below that of a background reading taken from recording mean fluorescence intensity of several clearly $KI67^-$ cells and averaging those values. To transform the data prior to statistical testing to produce normal distribution, we wanted to log transform it, however as some values = 0 it was necessary to do a +1(log) transformation where we added a value of one to every data point prior to the log transformation. To show KIF22 levels in $KI67^+$ cells vs $KI67^-$ cells we created boxplots from the individual values in each sample (two 12pcw samples treated individually) and performed an unpaired t-test to determine statistical significance for each of the rostral-middle-caudal sections with the clear caveat that any result is derived from pseudo replication. As again we saw no difference between ages nor anatomical location, we then averaged the values to pool the untransformed data, however for the average values were greater than 0.1 we were able to perform log transformation rather than +1(log), as described above in Table 40 with each age (two 12pcw and one 14pcw) considered a separate n value.

All analysis was performed, and graphs generated in GraphPad prism version 6.0, with the exception of the correlation graphs which were created in R studio.

4.3 Results

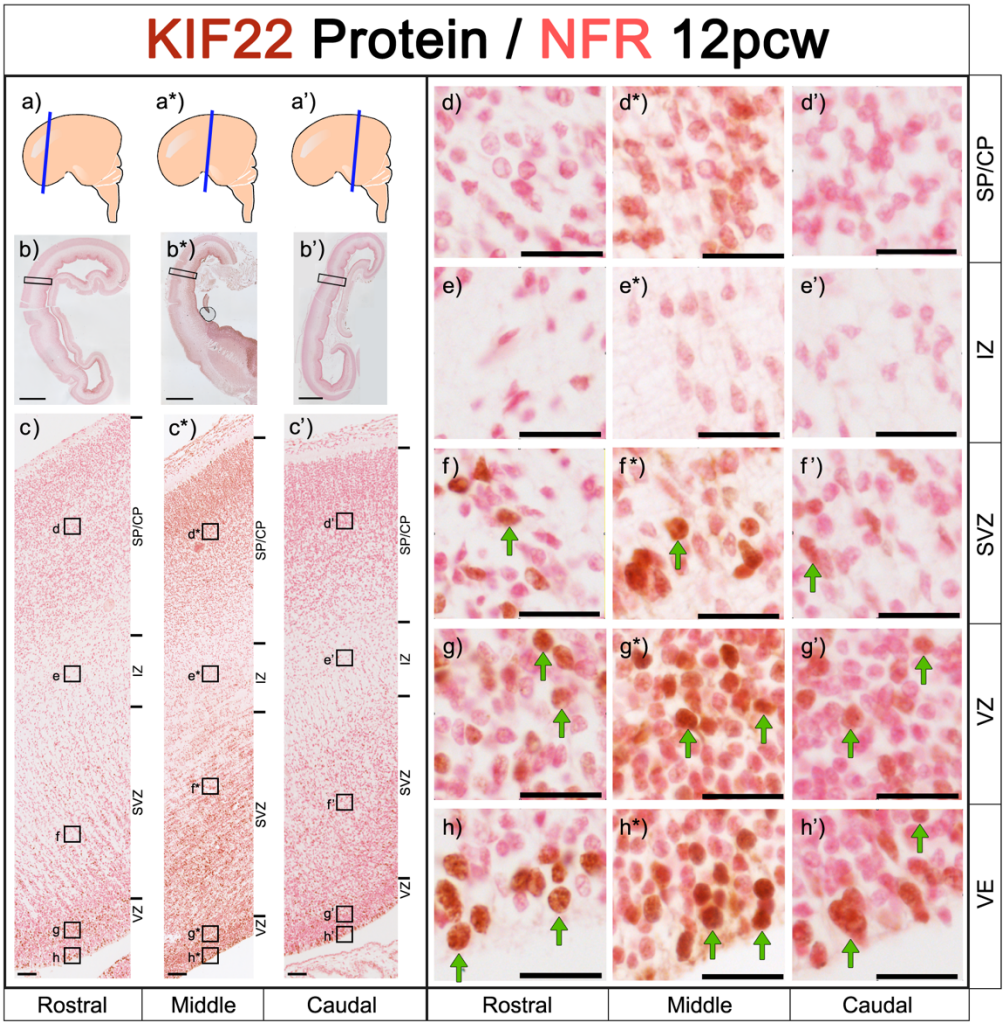
4.3.1 KIF22 protein is expressed in the germinative zones of the 12, 14 and 16pcw cortex

Here we characterize KIF22 protein expression during human corticogenesis. As described in the methods, coronal cortex sections samples from the rostral-middle-caudal axis were immunostained for the KIF22 protein and counterstained with NFR to show cytoarchitecture. $KIF22^{+}$ (brown) and $KIF22^{-}$ (red) cells were counted for each region in the telencephalic wall (VE, VZ, SVZ, IZ and SP/CP) (see 4.2.2.1 for sampling details), and the lamination was identified by cell nuclear density (Bayer and Altman, 2002, 2005).

We investigated three developmental stages across the rostral-middle-caudal axis and saw similar expression patterns. The data is shown for 12pcw (Figure 75), 14pcw (Figure 76), 16pcw (Figure 77). We use schematics (a, a*, a') to indicate the rostral-middle-caudal region of the brain sections were taken from. Next, we show low magnification image of the entire section stained (b, b*, b') on which the cortical region used for high magnification imaging and analysis is indicated by a black box. At higher magnification (c, c*, c'), we can see that KIF22 protein is most abundant in the germinative zones, in a similar pattern seen in the *in situ* hybridisation stains in chapter three. This can be seen more clearly at higher magnification (d-h') of the boxed region shown on (c, c*, c') where green arrows indicate examples of $KIF22^{+}$ cells. Looking qualitatively at these images it appears as though KIF22 protein is predominantly present in cells within the germinative zones of the cerebral cortex. We sought to confirm this with quantification, for clarity we presented the data two different ways: first as a percentage of total cells which express KIF22 (i-k), in this way across the rostral-middle-caudal axis, we can see that KIF22 expressing cells are most abundant in the VE, followed by the VZ and SVZ with the IZ and SP/CP presenting a very low to

complete absence of KIF22. The second way we presented the data (i*-k*), shows the total number of cells split to indicate how many are $KIF22^{+}$ and how many are $KIF22^{-}$. This provides the same result, showing the highest abundance of $KIF22^{+}$ cells in the VE followed by the VZ and SVZ. This way of presenting the data also clearly shows us that even in the most KIF22 abundant region, only a subset of cells expresses KIF22 protein. In addition we can also see high KIF22 protein expression in the ganglionic eminence, notably in Figure 76 b* and Figure 77 b*, suggesting KIF22 is not restricted to the dorsal cortical germinative zones.

Our observation of KIF22 cortical expression is the same across all ages and rostral-middle-caudal ages studied showing that within the ages looked at, KIF22 protein remains restricted to the germinative zones.


KIF22 Protein / NFR 12pcw
a)
a*)
a')
b)
b*)
b')
c)
c*)
c')
d)
d*)
d')
e)
e*)
e')
f)
f*)
f')
g)
g*)
g')
h)
h*)
h')
SP/CP
IZ
SVZ
VZ
VE
Rostral
Middle
Caudal

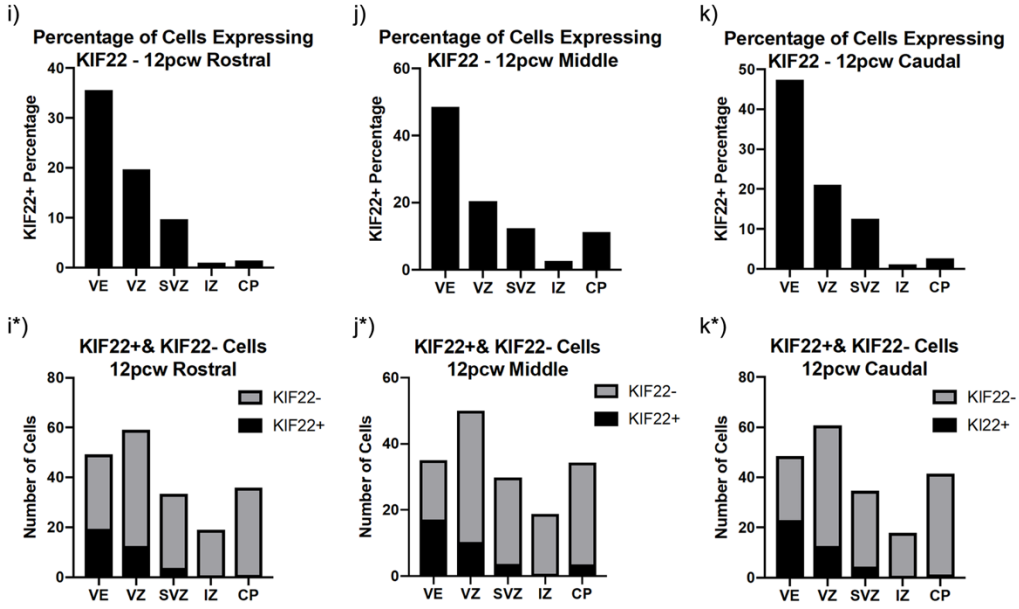
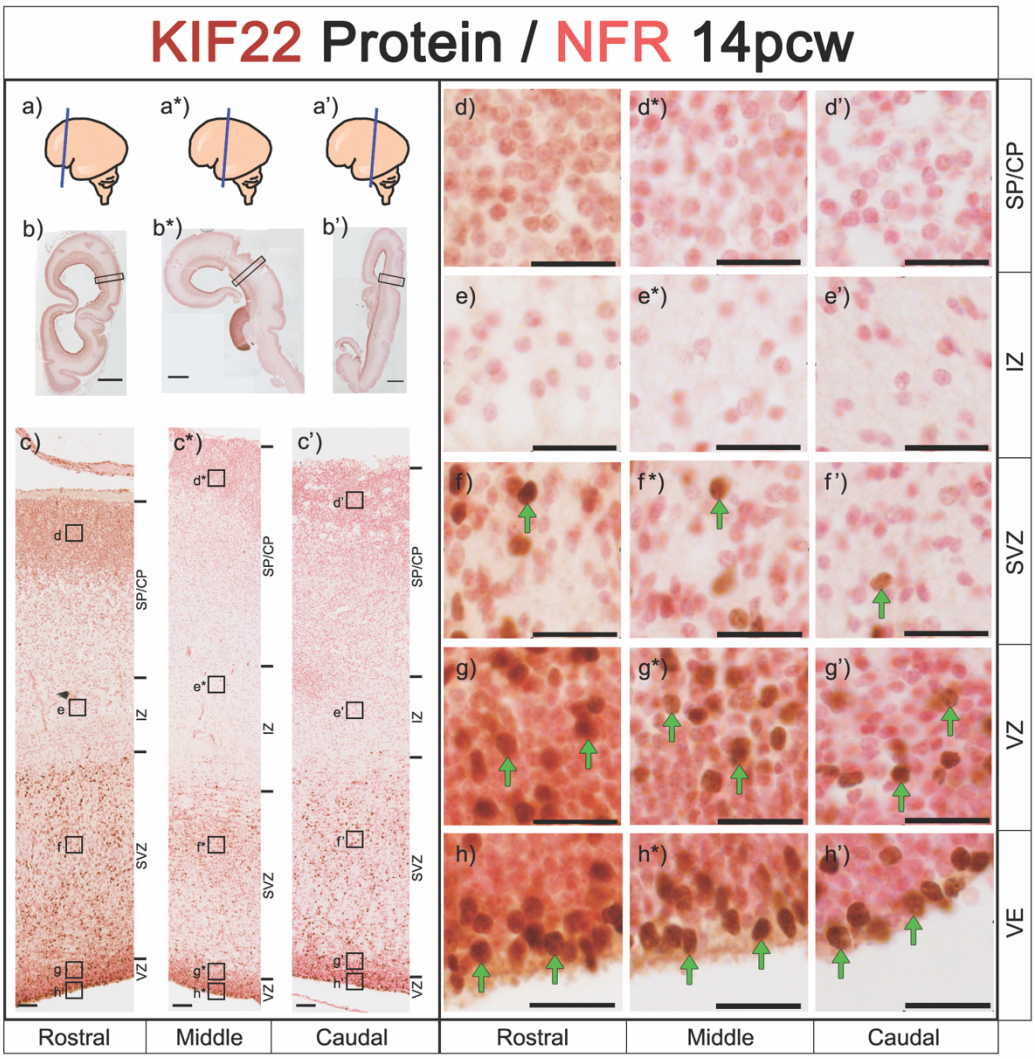

i)
Percentage of Cells Expressing KIF22 - 12pcw Rostral
KIF22+ Percentage
j)
Percentage of Cells Expressing KIF22 - 12pcw Middle
KIF22+ Percentage
k)
Percentage of Cells Expressing KIF22 - 12pcw Caudal
KIF22+ Percentage
VE VZ SVZ IZ CP
i*)
KIF22+& KIF22- Cells 12pcw Rostral
j*)
KIF22+& KIF22- Cells 12pcw Middle
k*)
KIF22+& KIF22- Cells 12pcw Caudal
Number of Cells
KIF22-
KIF22+
KI22+

Figure 75: KIF22 protein expression at 12pcw

a, a*, a') showing brain regions sectioned. b, b*, b') low magnification image of the whole brain section, scale bars = 2mm. c, c*, c') sections spanning the rostral-middle-caudal axis showing KIF22 protein expression in the telencephalic wall, scale bars = 100μm. d-h') high magnification images of different cortical zones rostral-caudal. $KIF22^{+}$ cells in brown and examples indicated by green arrows, $KIF22^{-}$ cells in red pink, scale bars = 25μm. i-k) quantification of percentage of cells expressing KIF22. i*-k* = quantification separating the cell population in each zone into $KIF22^{+}$ and $KIF22^{-}$ cells.


KIF22 Protein / NFR 14pcw
a)
a*)
a')
b)
b*)
b')
c)
c*)
c')
d)
d*)
d')
e)
e*)
e')
f)
f*)
f')
g)
g*)
g')
h)
h*)
h')
SP/CP
IZ
SVZ
VZ
VE
Rostral
Middle
Caudal

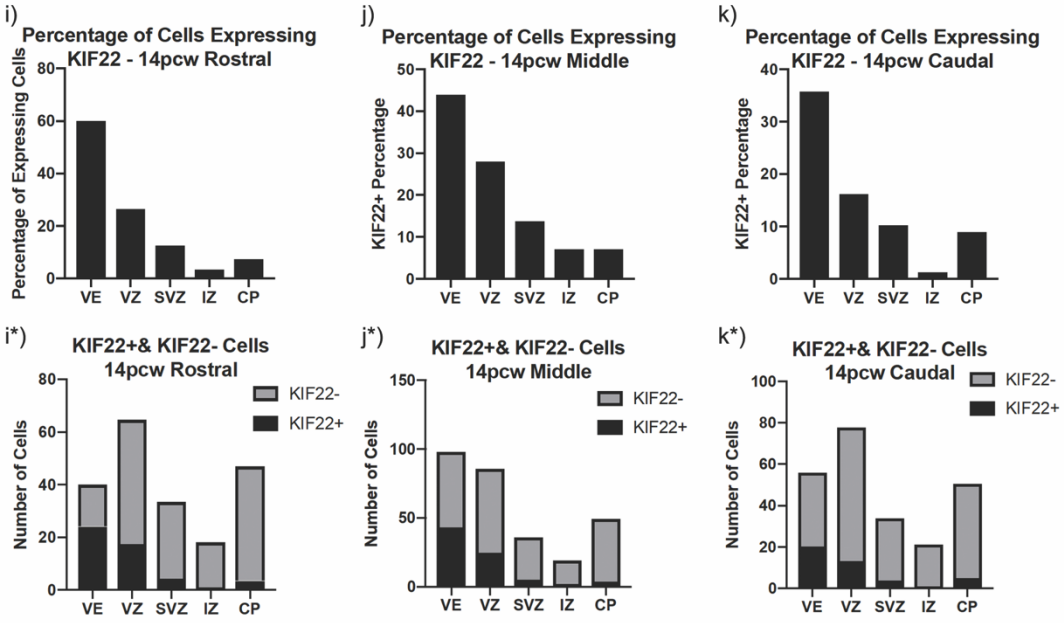
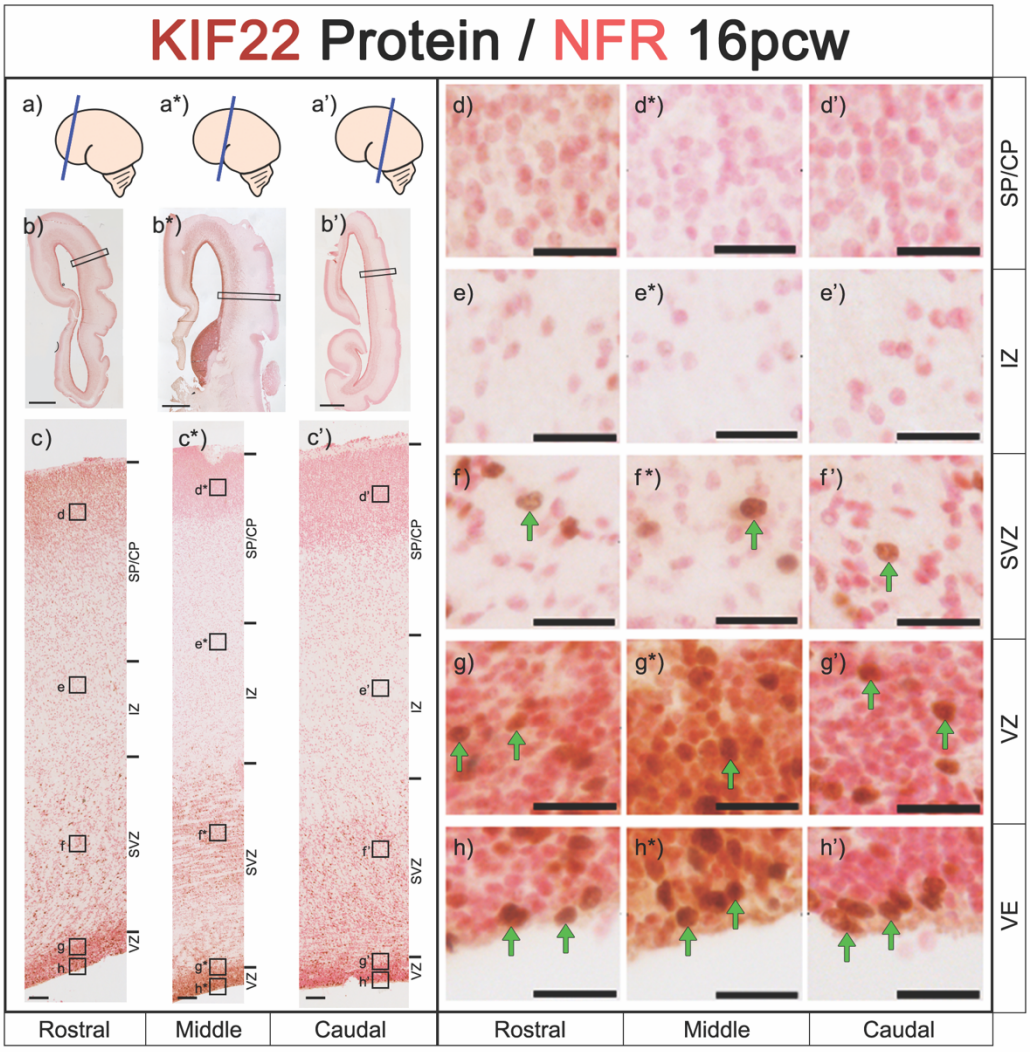

i)
Percentage of Cells Expressing KIF22 - 14pcw Rostral
Percentage of Expressing Cells
j)
Percentage of Cells Expressing KIF22 - 14pcw Middle
KIF22+ Percentage
k)
Percentage of Cells Expressing KIF22 - 14pcw Caudal
KIF22+ Percentage
VE VZ SVZ IZ CP
i*)
KIF22+& KIF22- Cells 14pcw Rostral
j*)
KIF22+& KIF22- Cells 14pcw Middle
k*)
KIF22+& KIF22- Cells 14pcw Caudal
Number of Cells
KIF22-
KIF22+

Figure 76: KIF22 protein expression at 14pcw

a, a*, a') showing brain regions sectioned. b, b*, b') low magnification image of the whole brain section, scale bars = 2mm. c, c*, c') sections spanning the rostral-middle-caudal axis showing KIF22 protein expression in the telencephalic wall, scale bars = 100μm. d-h') high magnification images of different cortical zones rostral-caudal. $KIF22^{+}$ cells in brown and examples indicated by green arrows, $KIF22^{-}$ cells in red pink, scale bars = 25μm. i-k) quantification of percentage of cells expressing KIF22. i*-k* = quantification separating the cell population in each zone into $KIF22^{+}$ and $KIF22^{-}$ cells.


KIF22 Protein / NFR 16pcw
a)
a*)
a')
b)
b*)
b')
c)
c*)
c')
d)
d*)
d')
e)
e*)
e')
f)
f*)
f')
g)
g*)
g')
h)
h*)
h')
SP/CP
IZ
SVZ
VZ
VE
Rostral
Middle
Caudal
Rostral
Middle
Caudal

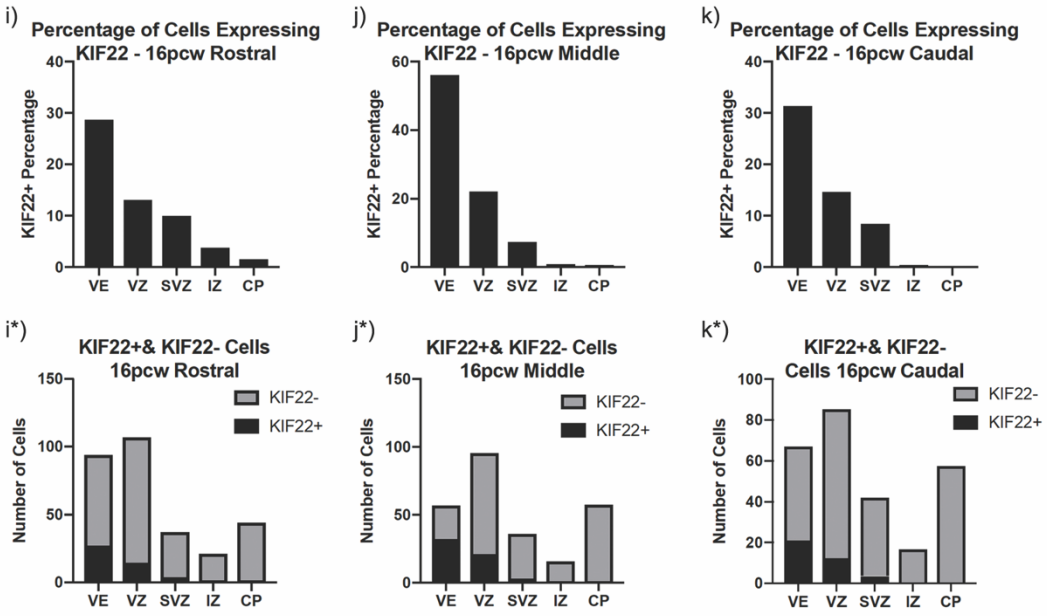

i)
Percentage of Cells Expressing KIF22 - 16pcw Rostral
KIF22+ Percentage
VE VZ SVZ IZ CP
j)
Percentage of Cells Expressing KIF22 - 16pcw Middle
KIF22+ Percentage
VE VZ SVZ IZ CP
k)
Percentage of Cells Expressing KIF22 - 16pcw Caudal
KIF22+ Percentage
VE VZ SVZ IZ CP
i*)
KIF22+& KIF22- Cells 16pcw Rostral
KIF22-
KIF22+
Number of Cells
VE VZ SVZ IZ CP
j*)
KIF22+& KIF22- Cells 16pcw Middle
KIF22-
KIF22+
Number of Cells
VE VZ SVZ IZ CP
k*)
KIF22+& KIF22- Cells 16pcw Caudal
KIF22-
KIF22+
Number of Cells
VE VZ SVZ IZ CP

Figure 77: KIF22 protein expression at 16pcw

a, a*, a') showing brain regions sectioned. b, b*, b') low magnification image of the whole brain section, scale bars = 2mm. c, c*, c') sections spanning the rostral-middle-caudal axis showing KIF22 protein expression in the telencephalic wall, scale bars = 100μm. d-h') high magnification images of different cortical zones rostral-caudal. $KIF22^{+}$ cells in brown and examples indicated by green arrows, $KIF22^{-}$ cells in red pink, scale bars = 25μm. i-k) quantification of percentage of cells expressing KIF22. i*-k* = quantification separating the cell population in each zone into $KIF22^{+}$ and $KIF22^{-}$ cells.

We next pooled the KIF22$^+$ count data as shown in Figure 78 to compare between all anatomical regions (Figure 78 a) and all ages (Figure 78b). This concisely summarises the individual values from Figure 75, Figure 76 and Figure 77. We found that the percentage of KIF22$^+$ cells in the VE (40-50%) was consistently higher than in other regions, followed by the VZ (20-30%) and SVZ (10%) with even fewer cells (<10%) in the IZ and SP/CP. This result describes KIF22 protein as predominantly restricted to a subset of cells in the germinative zones of the developing cortex at all stages studied.

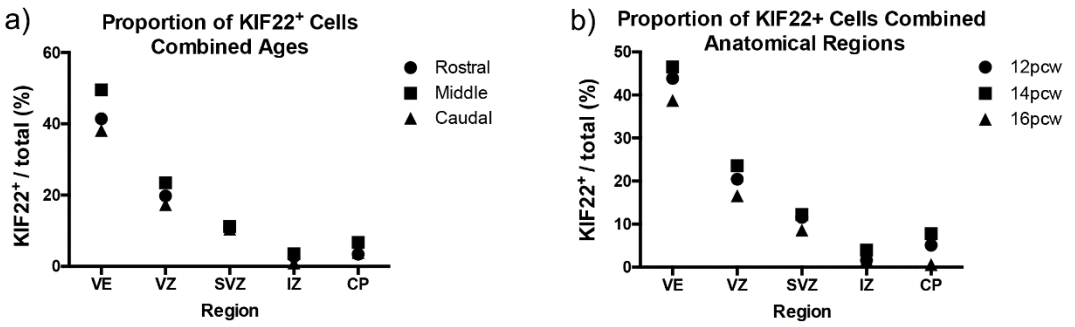


Figure 78: Pooled values for KIF22 protein

Pooled values for our percentage of KIF22$^+$ cells.

a) values combined for individual ages with rostral (circle), middle (square) and caudal (triangle) separate.

b) values combined for the anatomical regions with 12pcw (circle), 14pcw (square) and 16pcw (triangle) kept separate.

4.3.2 KIF22 protein expression is restricted to proliferating cells

Our results above show us that KIF22 protein expression is almost exclusively restricted to a subset of cells within the germinative zones of the human fetal cerebral cortex. From the scRNA-seq data we expect these to be progenitor cells (Figure 69), however we wanted to identify *in vivo* what these cells are. To determine what these KIF22 positive cells were, we performed double immunofluorescence for KIF22 and KI67. The KI67 protein is expressed in all proliferating cells and switched off as cells become post-mitotic (Scholzen and Gerdes, 2000; Miller *et al.*, 2018) making it an ideal tool to identify proliferating cells in our tissue. As our IHC analysis showed that the $KIF22^{+}$ cells were predominantly located in the VE, VZ and SVZ (Figure 78), we examined these regions for analysis.

We performed double immunofluorescence as seen in Figure 79 for KI67 (green) and KIF22 (red) with all cell nuclei stained blue with DAPI at 12pcw (a) and 14pcw (b), and a high magnification region of KIF22 and KI67 expression cells can be seen in Figure 79 c.

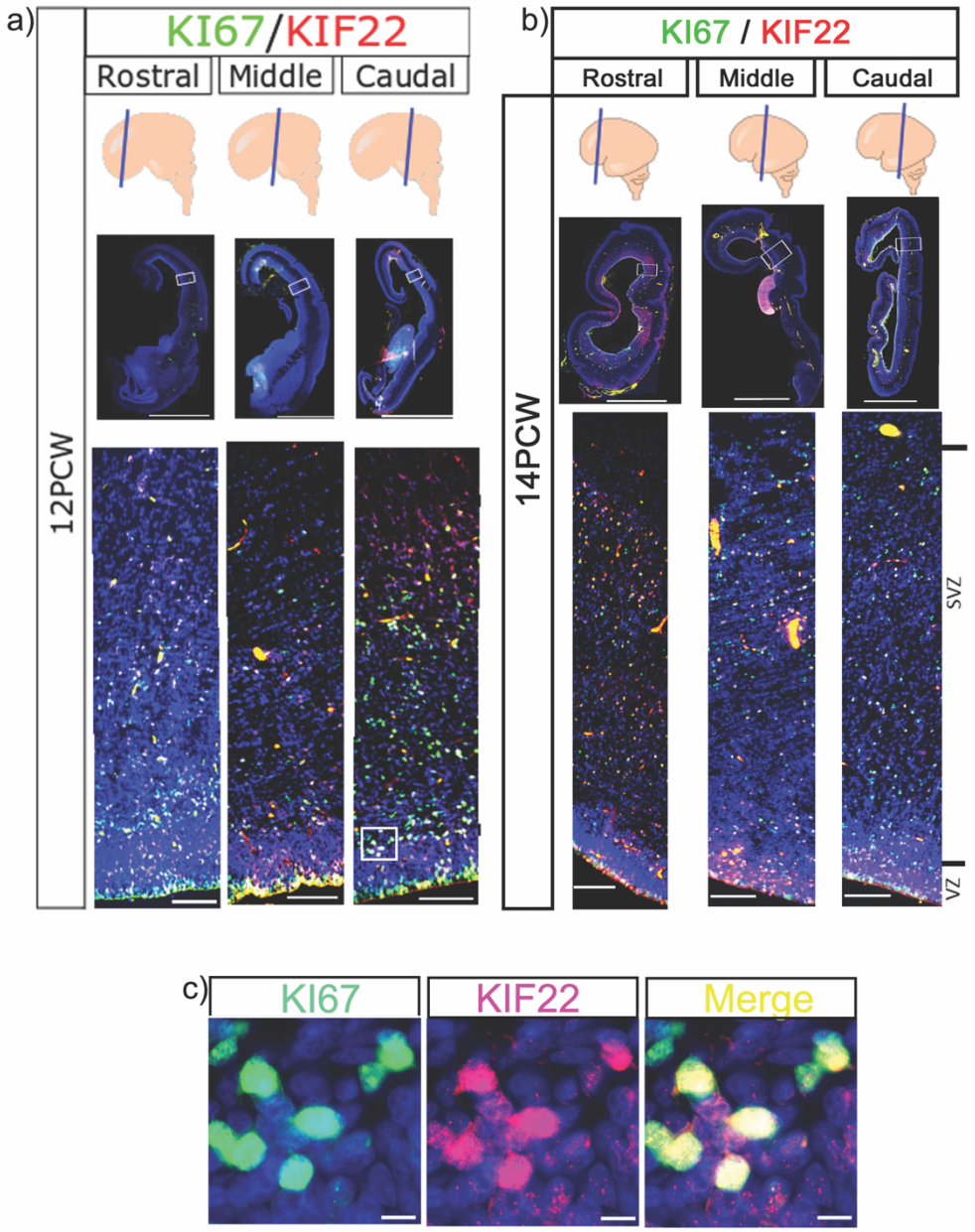


Figure 79: Immunofluorescence of KIF22 and KI67 in the cortex

KI67 = green, KIF22 = red, DAPI = blue

a) KIF22 and KI67 at 12pcw, low magnification scale bars = 4mm, high magnification scale bars = 100µm. White box on high magnification denotes region for c.

b) KIF22 and KI67 at 14pcw, low magnification scale bars = 4mm, high magnification scale bars = 100µm.

c) High magnification of KIF22/KI67 expression cells, scale bars = 10µm

The first analysis we performed was to count if cells expressed KI67 and KIF22 (see methods 4.2.3.1 for details of counting method) and the results of this analysis are shown in Figure 80. We performed this for two 12pcw brains (Figure 80 a and b) and one 14pcw brain (Figure 80 c). Cell counts at all ages for $KIF22^{+}/KI67^{+}$ labelled cells show that the majority (80-90%) of $KI67^{+}$ cells also express KIF22 at both ages and across the rostral-middle-caudal axis. We then combined the data for anatomical locations and ages (d) (see methods 4.2.3.2 for details) and performed statistical analysis which revealed significantly more $KIF22^{+}/KI67^{+}$ cells than $KIF22^{+}/KI67^{-}$ and $KIF22^{-}/KI67^{+}$ cells (One way ANOVA $p<0.0001$, $F = 604.4$, Tukey post hoc test). This analysis confirms to us that KIF22 protein expression is restricted to proliferating cells within the germinative zones.

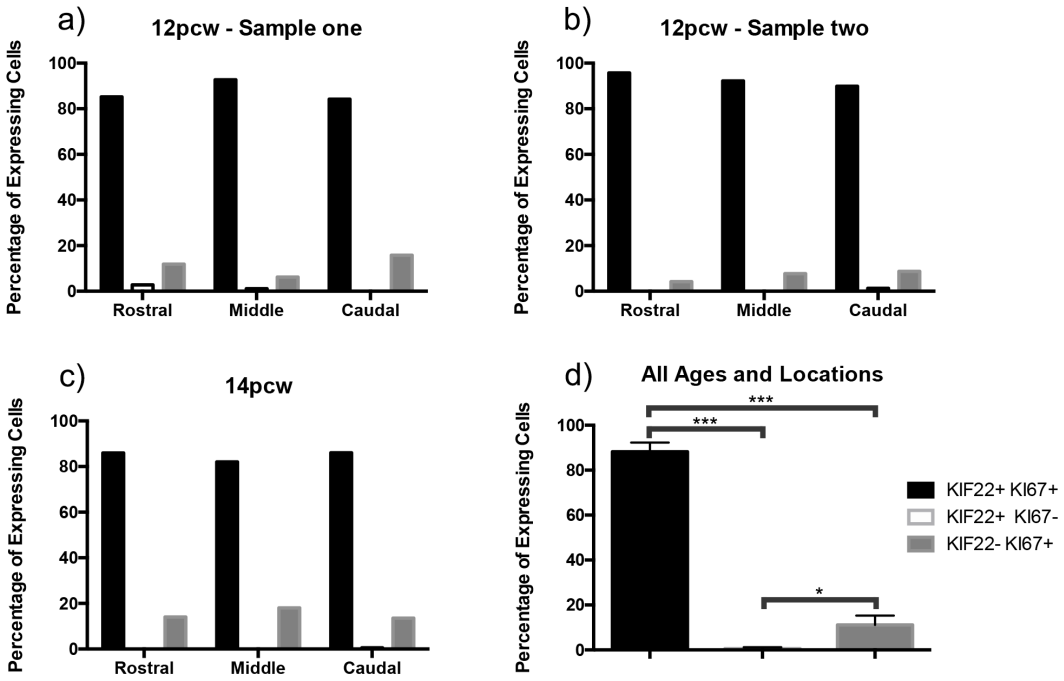


Figure 80: KIF22 and KI67 cell count analysis

Quantification of KIF22 and KI67 expression cells in the cortex. a) percentage of cells expressing KIF22, KI67 or both at 12pcw in the first brain sample. b) percentage of cells expressing KIF22, KI67 or both at 12pcw in the second brain sample. c) percentage of cells expressing KIF22, KI67 or both at 14pcw. d) Combined data of percentage of cells expressing KIF22, KI67, or both with one-way ANOVA followed by post-hoc Tukey test performed.

4.3.2.1 KIF22 protein is variable but higher in proliferating cells

Next, we considered the importance of the protein level, beyond whether it is a binary present or absence. From the scRNA-seq analysis in (3.3.2), we see that within the progenitor population there is variability in the mRNA expression level. To see if this was the same for the protein, we quantified nuclear KIF22 fluorescence level at the single-cell level (Figure 81) at 12pcw (brain one= a-a*, brain two = b-b*) and 14pcw across the rostral-middle-caudal axis, again acknowledging that the statistics for the individual values arise from pseudo replication. Combining the data from the two ages and locations (see methods 4.2.3.2 for pooling of data), our result showed significantly higher KIF22 fluorescence in KI67^{+} cells (Figure 81 d). Thus, we are able to confirm that KIF22^{+} cells are proliferating, that KI67^{+} cells have higher KIF22 protein levels and that those levels are variable.

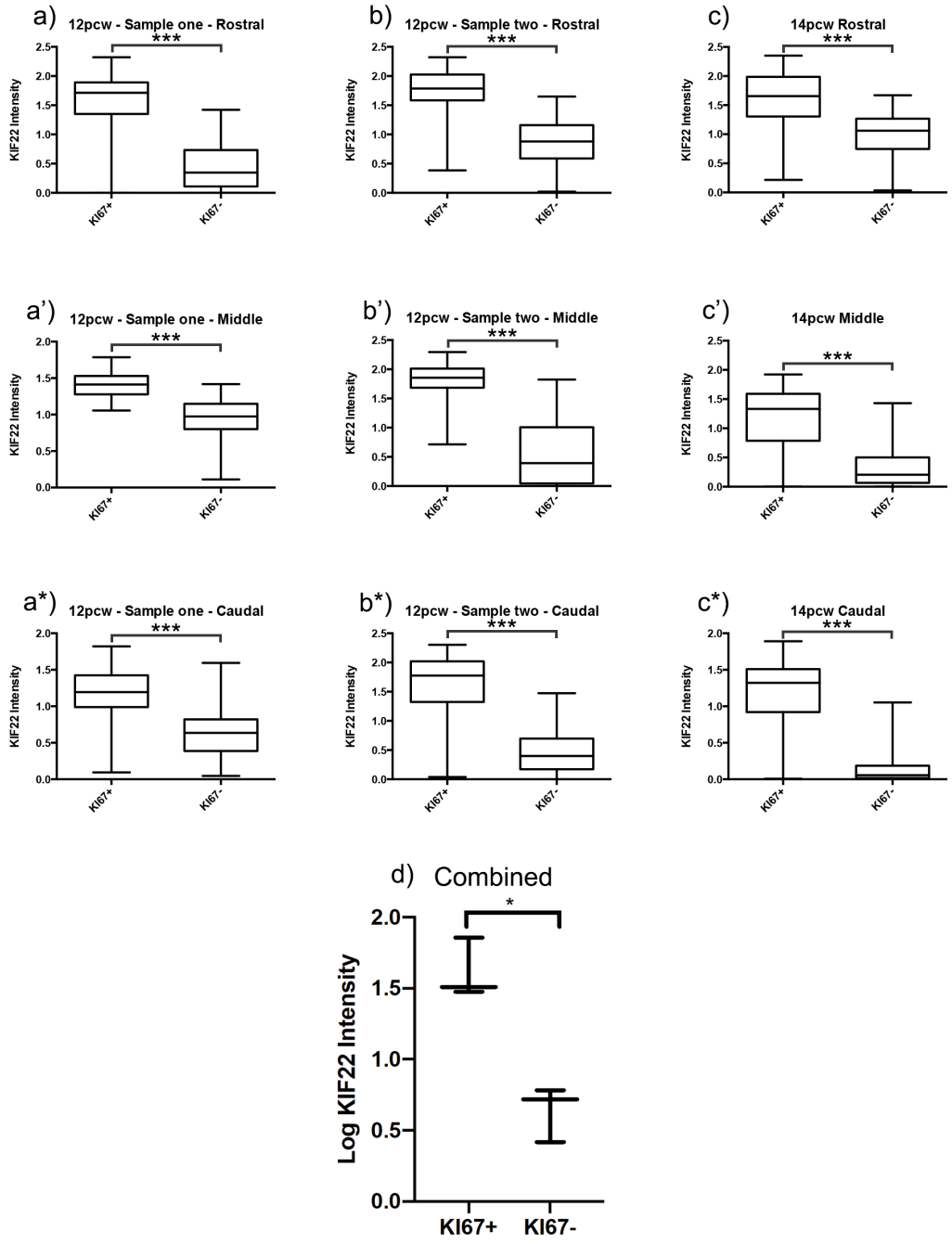

a)
12pcw - Sample one - Rostral
***
KIF22 Intensity
KI67+
KI67-
b)
12pcw - Sample two - Rostral
***
c)
14pcw Rostral
***
a')
12pcw - Sample one - Middle
***
b')
12pcw - Sample two - Middle
***
c')
14pcw Middle
***
a*)
12pcw - Sample one - Caudal
***
b*)
12pcw - Sample two - Caudal
***
c*)
14pcw Caudal
***
d)
Combined
*
Log KIF22 Intensity
KI67+
KI67-

Figure 81: KIF22 intensity in proliferating vs non proliferating cells

a, a', a*) 12pcw brain sample one quantification of KIF22 fluorescence intensity in $KI67^{+}/KI67^{-}$ cells (raw data transformation = +1(log), unpaired t-test with Welch's correction p=<0.0001).
b, b', b*) 12pcw brain sample two quantification of KIF22 fluorescence intensity in $KI67^{+}/KI67^{-}$ cells (raw data transformation = +1(log), unpaired t-test with Welch's correction p=<0.0001).
c, c', c*) 14pcw quantification of KIF22 fluorescence intensity in $KI67^{+}/KI67^{-}$ cells (raw data transformation = +1(log), unpaired t-test with Welch's correction p=<0.0001).
d) quantification of KIF22 fluorescence intensity in $KI67^{+}/KI67^{-}$ cells, 12 and 14pcw combined (raw data transformation =(log), paired t-test, p=0.0122.

4.3.3 KIF22 protein levels vary with the cell cycle

From the scRNA-seq analysis (3.3), and the variable KIF22 protein levels in KI67^{+} cells, we can now hypothesise that KIF22 protein levels change throughout the cell cycle. We wanted to test this hypothesis and see how they changed during the cell cycle at the single-cell level, so we returned again to the double immunofluorescence staining of KIF22 and KI67 in Figure 79. We quantified the nuclear immunofluorescence intensity of KIF22 and KI67 in 300-400 individual cells in two 12pcw brains and one 14pcw brain across the rostral-middle-caudal axis (see methods 4.2.3.1 for details of sampling procedure). KI67 protein levels vary during the cell cycle as shown in Figure 82: they are lowed in G1 phase, increasing through S and G2 phase to peak in mitosis (Scholzen and Gerdes, 2000; Miller *et al.*, 2018).

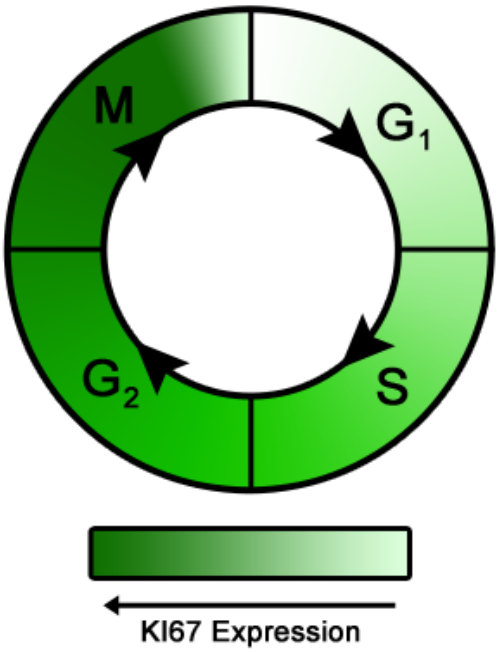


Figure 82: KI67 levels throughout the cell cycle

Schematic showing how KI67 protein levels vary throughout the cell cycle to peak in mitosis

We found a strong correlation between KI67 and KIF22 in our three samples across the rostral-middle caudal axis and the data for all of these is shown in Figure 82. We found a strong correlation between KIF22 and KI67 fluorescence intensity in both 12pcw brain samples (a, a' and a* and b, b' and b*) (Brain one rostral R^2=0.8095, middle R^2=0.8139, caudal R^2=0.6691. Brain two rostral R^2=0.7489, middle R^2=0.7447, caudal R^2=0.7763). We also saw the same strong positive correlation at 14pcw (c, c' and c*) (rostral R^2=0.7465, middle R^2=0.6668, caudal R^2=0.643). These data demonstrate that the correlation between KI67 and KIF22 is consistent between ages and rostral-middle-caudal location. Combining all values of KIF22/KI67 nuclear intensity (Figure 82 d) showed us that KIF22 levels strongly correlated with KI67 levels (R^2=0.7236)

Although KIF22 expressing cells were scattered throughout the VE, VZ and SVZ, there was a general trend for the cells expressing the highest levels of KIF22 to be closest to the apical surface (yellow coloured dots on scatterplots in Figure 82) with lower expressing cells tending to be further from the apical surface (purple coloured dots on scatterplots in Figure 82) however there are still some high KIF22/KI67 blue dots noticeable. During interkinetic nuclear migration, radial glial cell nuclei move to the apical surface to perform mitosis, so this spatial distribution suggests KIF22 is expressed at high levels by apical radial glial cells undergoing mitosis at the apical surface of the VZ. Basal radial glial cells in the SVZ however do not undergo IKNM, instead they undergo mitotic somal translocation as described in (1.6.3), and the presence of high KIF22/KI67 in SVZ cells suggests KIF22 to also have a role in the division of bRGCs and IPCs,

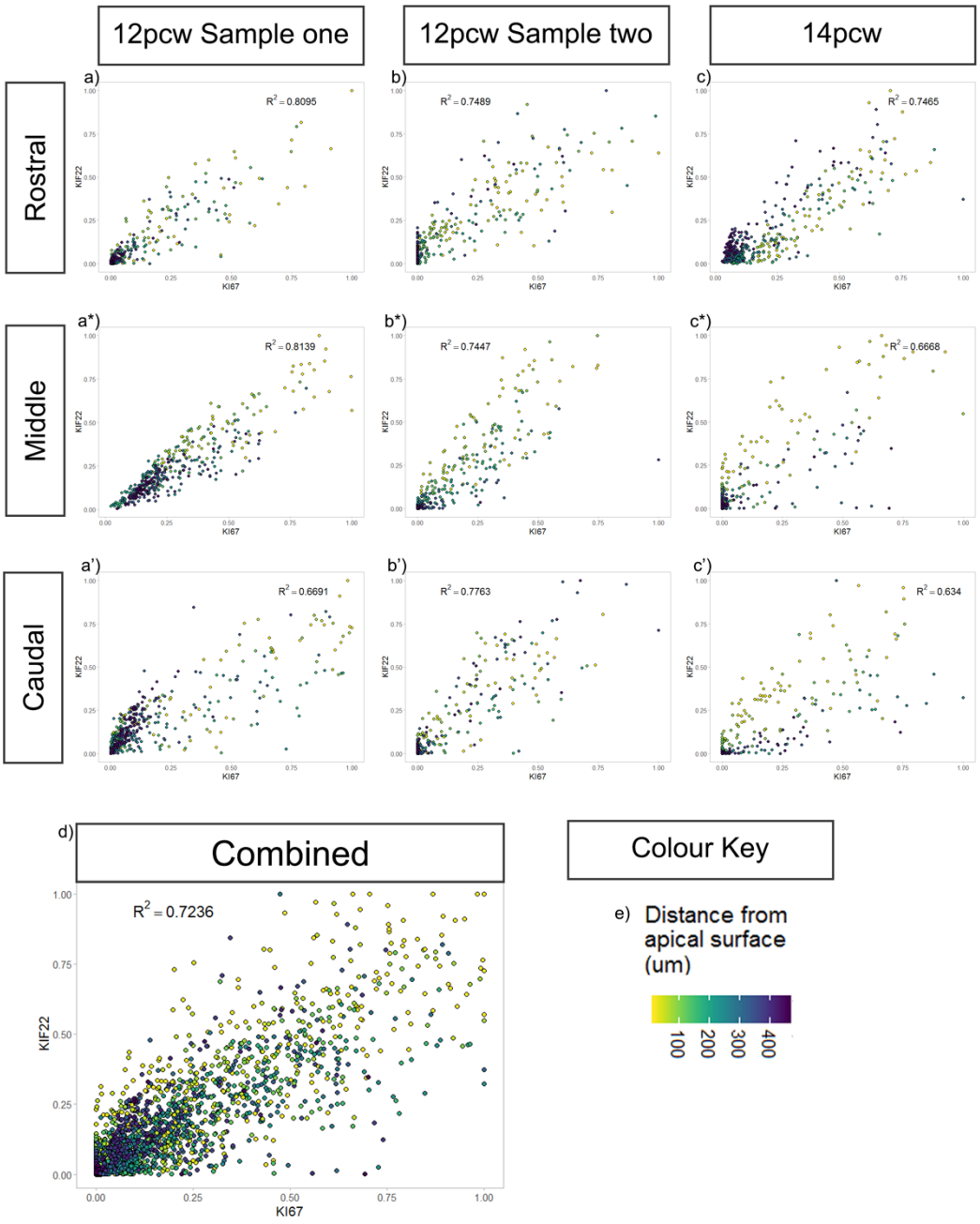

12pcw Sample one
12pcw Sample two
14pcw
Rostral
Middle
Caudal
a)
b)
c)
a*)
b*)
c*)
a')
b')
c')
R2 = 0.8095
R2 = 0.7489
R2 = 0.7465
R2 = 0.8139
R2 = 0.7447
R2 = 0.6668
R2 = 0.6691
R2 = 0.7763
R2 = 0.634
KIF22
KI67
d)
Combined
R2 = 0.7236
Colour Key
e) Distance from apical surface (um)
100
200
300
400

Figure 83: Quantification of KI67 and KIF22 levels

a) Quantification of KIF22 and KI67 fluorescence intensity in 12pcw brain sample one with distance from apical surface indicated by dot colour.
b) Quantification of KIF22 and KI67 fluorescence intensity in 12pcw brain sample two with distance from apical surface indicated by dot colour.
c) Quantification of KIF22 and KI67 fluorescence intensity in 14pcw with distance from apical surface indicated by dot colour.
d) Intensity correlations of KIF22 and KI67 nuclear fluorescence intensity for combined data of three brain samples rostral-middle-caudal with distance from apical surface indicated by dot colour.
e) Colour key showing distance from apical surface: yellow dots closest to apical surface, purple dots are furthest away

We wanted to confirm that the strong correlation we observed between KIF22 and KI67 fluorescence intensity was not simply a result of the nucleus size changing with the cell cycle. To do this we confirmed that KIF22 nor KI67 protein levels did not correlate with nuclear size by DAPI staining. At all samples and ages studied we did not see any correlation at 12pcw (sample one = Figure 84, sample two = Figure 85) nor at 14pcw (Figure 86).

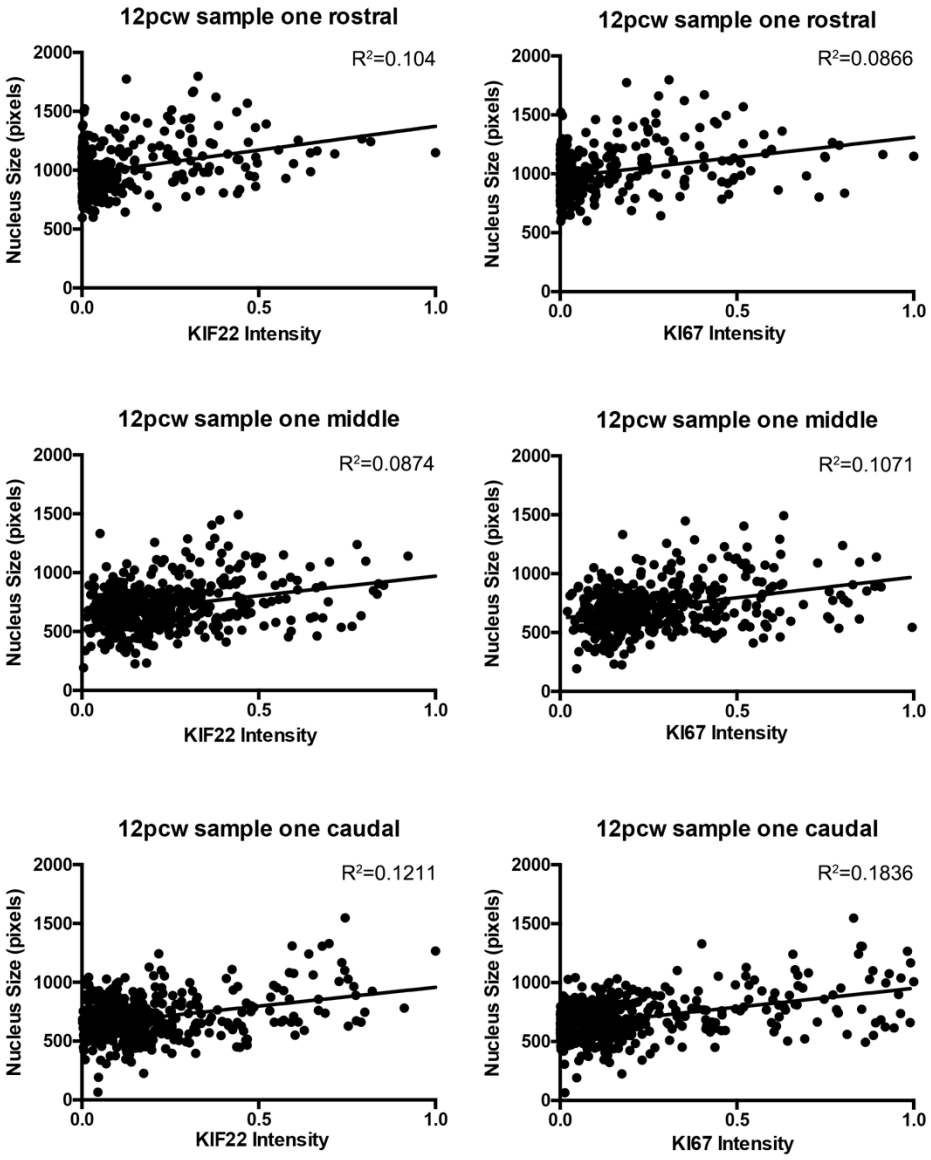


Figure 84: No correlation between nucleus size and fluorescence intensity in 12pcw sample one

Quantification showing that there is no correlation between KIF22 fluorescence intensity and nuclear size nor KI67 fluorescence intensity and nuclear size across the rostral-middle-caudal axis in the first 12pcw sample.

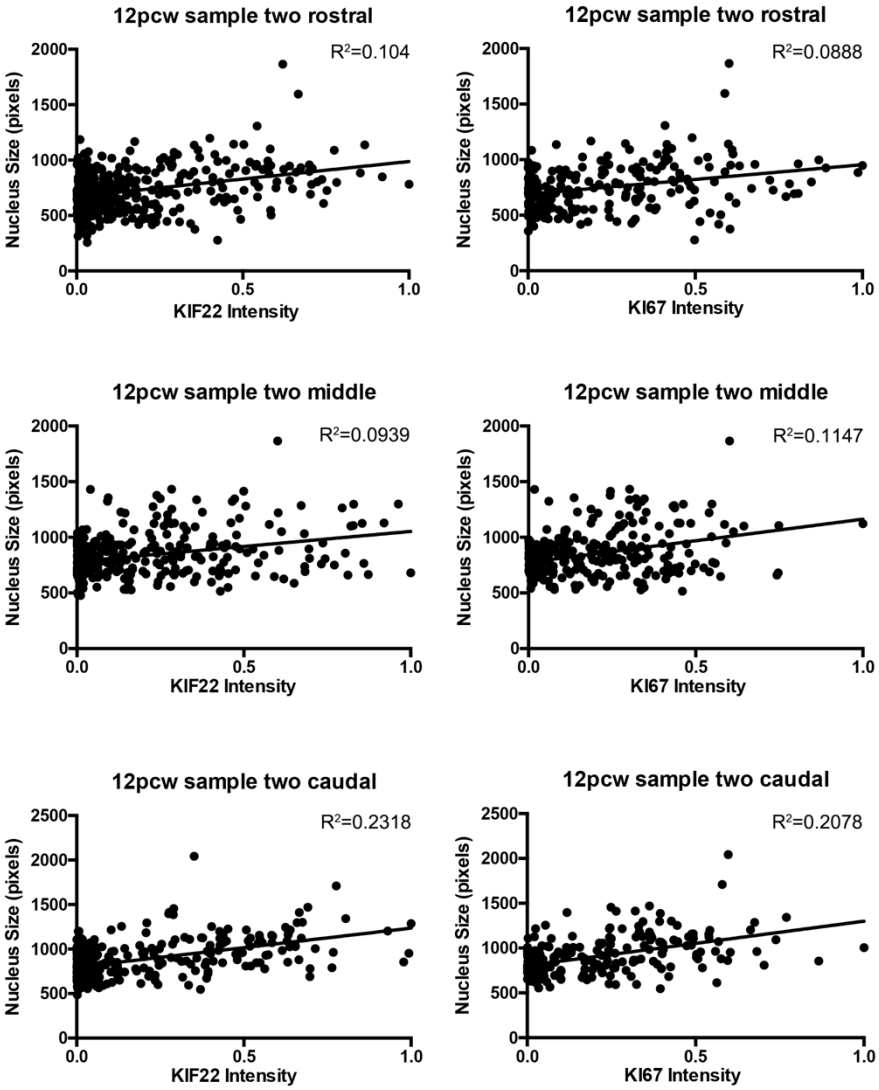


Figure 85:No correlation between nucleus size and fluorescence intensity in 12pcw sample two

Quantification showing that there is no correlation between KIF22 fluorescence intensity and nuclear size nor KI67 fluorescence intensity and nuclear size across the rostral-middle-caudal axis in the second 12pcw sample.

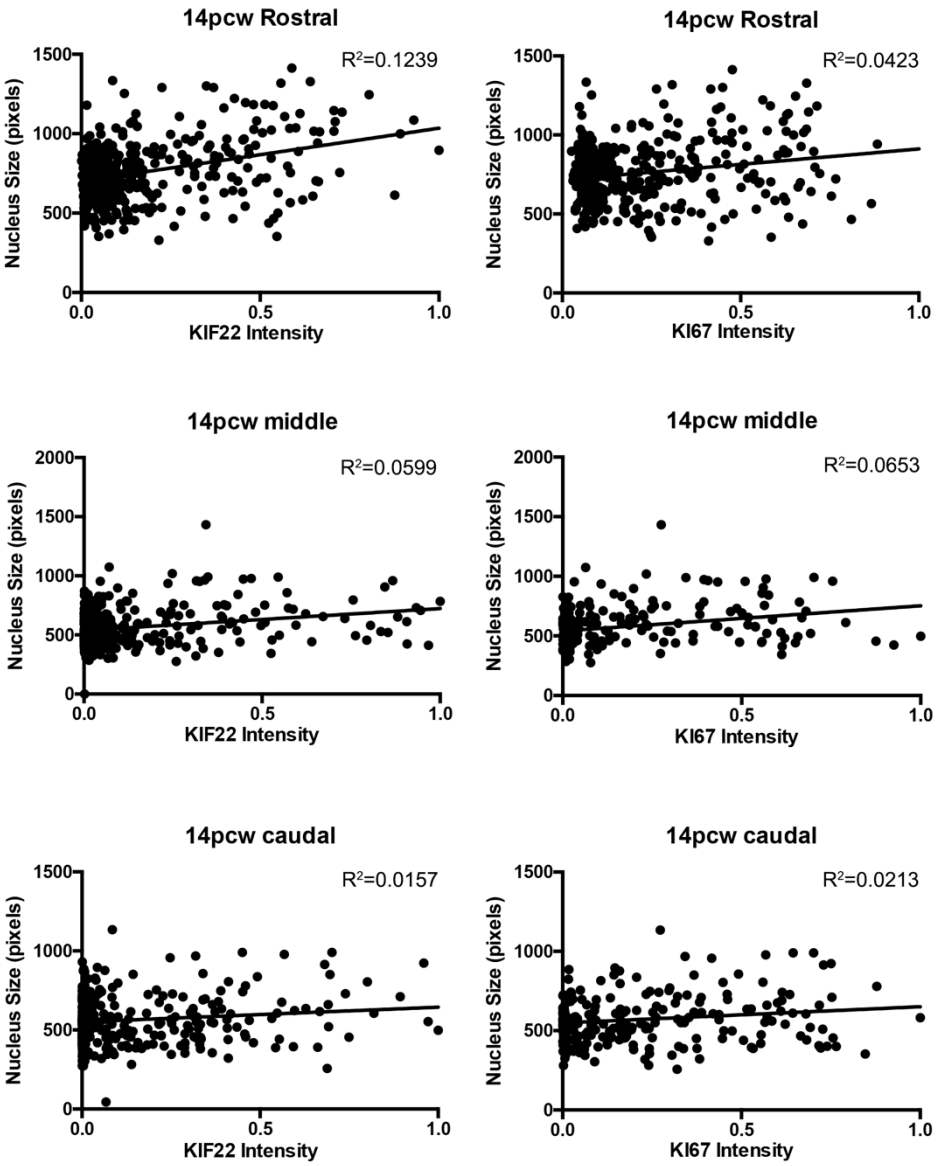


Figure 86: No correlation between nucleus size and fluorescence intensity at 14pcw

Quantification showing that there is no correlation between KIF22 fluorescence intensity and nuclear size nor KI67 fluorescence intensity and nuclear size across the rostral-middle-caudal axis in the 14pcw sample

From these data and our analysis, we show that KIF22 protein levels change throughout the cell cycle in a positive correlation with KI67: KIF22 protein is present in G1 and increases through S and G2 phase to peak in mitosis. This is shown as a summary schematic in Figure 87.

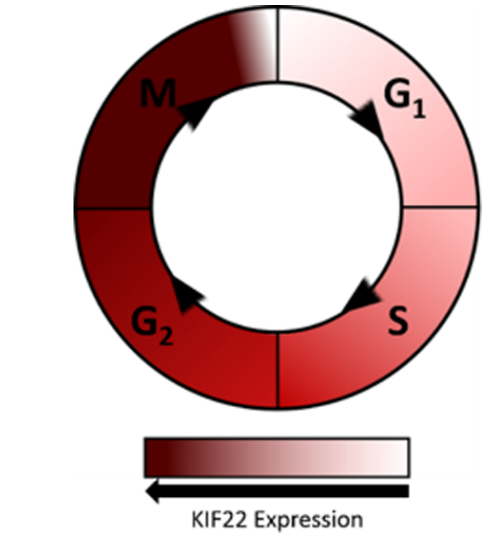


Figure 87: KIF22 model

Model based on our results of how KIF22 protein levels change throughout the cell cycle

4.3.4 Conclusion of results

Here we summarise the results of this chapter:

We used colourmetric and fluorescent immunohistochemistry to describe KIF22 protein expression in the human fetal cortex during neural development. First, we used the colourmetric IHC to show KIF22 protein as being expressed in a subset of cells in the germinative regions of the cortex; highest in the VE followed by the VZ and SVZ. We showed that this expression pattern was consistent across the ages studied and across the rostral-middle-caudal axis.

Next, we used double immunofluorescence (IF) staining for KIF22 and KI67 to determine that the subset of KIF22 expression cells in these regions are indeed proliferating, a complementary result to that of the scRNA-seq data in Figure 69. Our final, and most interesting, finding of this chapter is our observation, again complementary to the result of the scRNA-seq analysis, that KIF22 protein levels may fluctuate throughout the cell cycle, suggesting they increase as cells move from G1 to S to G2 phase, peaking in mitosis.

4.4 Discussion

The *16p11.2* CNV is a polygenic mutation heavily implicated in ASD. However, to begin to understand the many ways in which this CNV affects development it is critical to understand the roles of each of its individual genes in development; here focussing on human fetal cerebral cortex development. Based on the scRNA-seq analysis in 3.3.2, we focused on KIF22 for further studies. We have showed it to be expressed in progenitors of the cortex, and its protein levels to vary with the cell cycle. In this section of chapter four we will discuss our results obtained in this chapter and the caveats of our work.

4.4.1 KIF22 protein is expressed in progenitor cells in the VZ and SVZ at varying levels during the cell cycle

Our results show that KIF22 protein is expressed in proliferating cells of the VZ and SVZ. Our data showing the highest proportion of $KIF22^{+}$ cells in the VE, the location of mitosis for apical radial glial cells, and the correlation of KI67 and KIF22 proteins in the VZ and SVZ (where cells undergo mitotic somal translocation to divide) suggest that KIF22 protein levels peak in mitosis and its expression level is closely tied to the cell cycle. This variation of KIF22 protein levels during the cell cycle suggests its correct function in cell division is dependent on its correct dosage.

Literature tells us that KIF22 is a multifunctional protein that regulates cell proliferation through at least two distinct mechanisms. Its first role is as a kinesin-like microtubule-based motor which acts to bind microtubules and chromosomes during mitosis and regulates mitotic spindle stability and symmetric/asymmetric cell division (Tokai *et al.*, 1996; Tokai-Nishizumi *et al.*, 2005; Sun and Hevner, 2014). Its second function is summarised in Figure 73, it regulates expression of cell-cycle regulator CDC25 by transcriptionally

repressed CDC25C and inhibits mitosis. The transcriptional repression of CDC25C depends on KIF22 being phosphorylated on Thr463 (Ohsugi *et al.*, 2003; Yu *et al.*, 2014) and KIF22 depletion in tumour cell lines has been shown to accelerate the G2/M transition and slow M/G1 transition (Yu *et al.*, 2014).

4.4.2 KIF22 in the *16p11.2* CNV

Based on the literature, it appears that KIF22 protein can act at several points in the cell cycle making it difficult to predict how increased or decreased dosage of KIF22 in the *16p11.2* duplication or deletion respectively would impact cell cycle in the specific context of cerebral cortical neural progenitors. However, we can still postulate potential hypotheses in the context of our results. We observed complementary results from both our scRNA-seq analysis and *in vivo* protein studies. Our observations, while purely correlative at the protein level, suggest that *KIF22* mRNA and KIF22 protein levels may both increase during the cell cycle to achieve highest levels in G2/M phase that drop as cells enter G1 phase. These data imply that KIF22 protein does not persist for long after it is translated and is degraded at the end of M-phase suggesting that both transcriptional and post-translational mechanism regulate its levels.

In our results we can clearly see that KIF22 protein levels correlate with KI67 in neural progenitors and steadily rise as the cell progresses through G1>S>G2>M phases, culminating in the maximum level during M-phase. One possibility, shown in Figure 88, is that KIF22 is required to reach a threshold for mitosis to occur, after which its levels must decrease sufficiently to allow mitotic exit. Whether cells undertake proliferative or neurogenic divisions is a process heavily controlled by cell cycle length (described in 1.6.6.2.1) (Calegari and Huttner, 2003; Lange, Huttner and Calegari, 2009). Perturbing *KIF22* gene dosage as a consequence of the *16p11.2* CNV may affect the timing of KIF22 protein reaching this threshold in neural progenitors

and therefore affect cell-cycle kinetics and perturb neurogenesis and neuronal output.

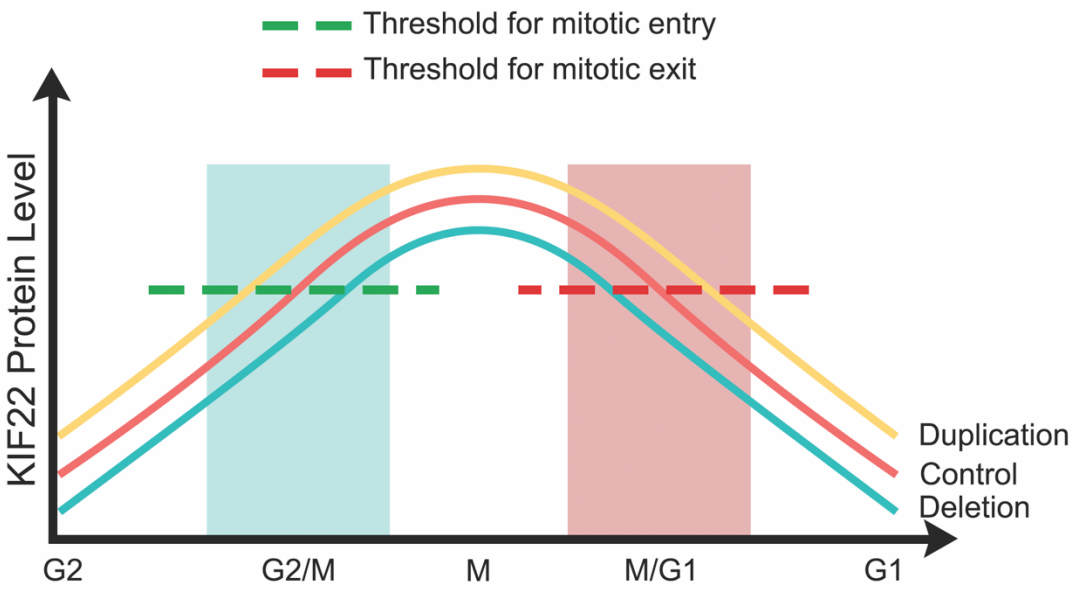


Figure 88: Hypothesis of altering KIF22 protein dosage in the *16p11.2* CNV

Here we show our hypothesis of how altering KIF22 protein dosage could alter mitosis. In the deletion (blue line), KIF22 protein levels are decreased so it takes longer for the required KIF22 protein levels required for mitotic entry to be reached, however these lower levels may also mean it is depleted quicker and the time for KIF22 levels to decrease sufficiently to allow mitotic exit is reduced.
Conversely in the duplication (yellow line), KIF22 protein levels are higher so the time for protein levels to accumulate enough to enter mitosis is reduced, but the elevated levels also mean the time needed for KIF22 protein to be depleted to allow mitotic exit is increased. In both the duplication and deletion this is likely to have substantial effects on the overall timing of the cell cycle with the potential to disrupt the balance of proliferation and differentiation.

To consider another mechanism by which altering KIF22 levels could impact proliferation and subsequent brain size is its transcriptional role in cell cycle modulation by transcriptional repression of CDC25C. CDC25 is known to primarily facilitate mitotic progression by dephosphorylating and activating the CDK1/CyclinB complex. Therefore, by altering KIF22 levels in the *16p11.2* CNV, this would alter CDC25C expression levels in a mirrored fashion (increased CDC25C expression in the deletion and decreased in the duplication), potentially impacting the time of mitotic entry and therefore cell cycle length. When KIF22 levels are depleted (as in the deletion), CDC25C expression would increase thus accelerating mitotic entry and shortening the cell cycle making NPCs more likely to undergo proliferative divisions. This matches with the phenotype of the *16p11.2* del mouse model in which increased progenitor proliferation is observed (Pucilowska *et al.*, 2015). Conversely, when KIF22 levels are heightened (as in the duplication), CDC25C expression would be repressed further slowing mitotic entry, prolonging the cell cycle and making NPCs more likely to take neurogenic divisions. This hypothesis is summarised in Figure 89. However, it is important to consider that *KIF22* is only one of the 29 genes whose dosage is altered in the CNV, therefore predicting effects of only altering *KIF22* dosage does not take into account the likely interactions or compensatory mechanisms occurring between *KIF22* and other CNV genes.

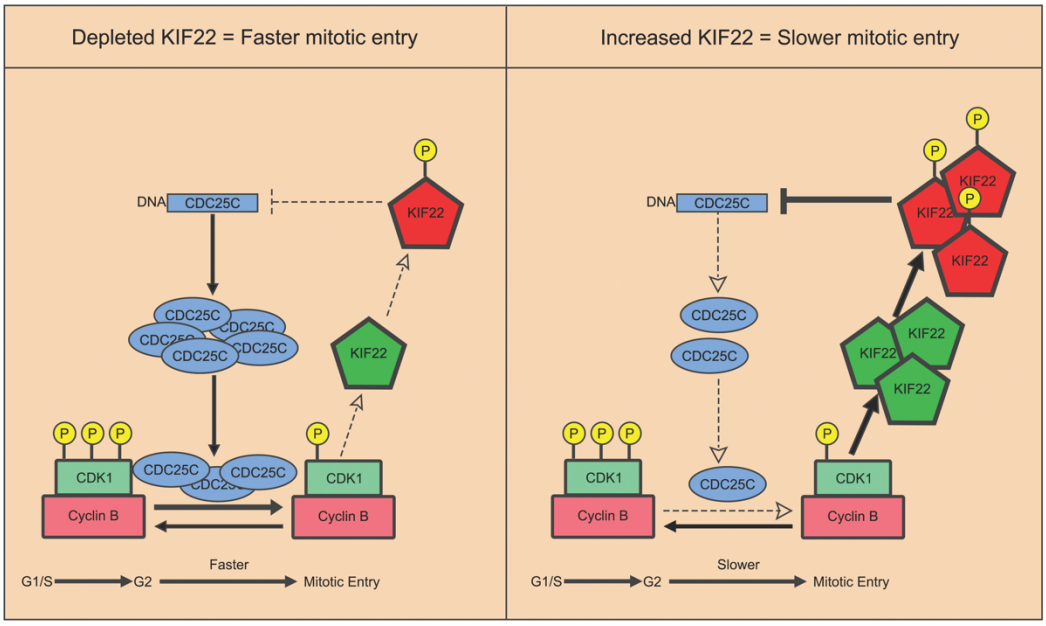


Figure 89: Potential model for the effect of altered KIF22 levels on the cell cycle

This schematic shows how altering KIF22 protein dosage could impact cell cycle length. In box one, reduced KIF22 dosage reduces the transcriptional repression of CDC25C increasing its levels which speeds up mitotic entry shortening the cell cycle length. In the second scenario the opposite is seen, more KIF22 represses CDC25C slowing mitotic exit and increasing cell cycle length.

4.4.3 Experimental limitations

When examining our results in the context of the literature, it also critically important to consider the limitations of our experiments.

In this chapter, we only predominantly focus on the germinative zones, the VZ and SVZ for our analysis. However, our IHC analysis (4.3.1) showed a small number of cells in the subplate/cortical plate region to be expressing KIF22 protein. While we did not study them in this work, this could warrant further study to identify what these might be; one hypothesis could be they are migratory interneurons.

Our result that KIF22 peaks in the M-phase of the cell cycle is dependent on correlations between KIF22 and KI67. We make the assumption here that KI67 protein levels change accurately throughout the cell cycle, however there is a solid body of literature confirming that it does (Scholzen and Gerdes, 2000; Miller *et al.*, 2018) and assume that the KI67 protein levels vary with the cell cycle in our system in the same way as the systems from the literature. We can be relatively confident of that by our result that we see many of the highest KI67 expression cells are found at the VE, the location at which aRGCs undergo mitosis.

Our result is dependent on correlations, and ideally, we would have liked to confirm by co-immunostaining for KIF22 and pH3, an M-phase marker. We did make several attempts at this, however the conditions required for each antibody which are shown in Table 41 were not compatible.

Table 41: Optimisation of KIF22 and pH3 antibodies

Fixation method	**Processing**	**KIF22 antibody**	**pH3 antibody**
Methacarne	paraffin	Worked	Failed
4% PFA	cryo	Failed	Worked
4% PFA	Paraffin	Failed	Failed

In addition, the cell counting could be improved with automation, to eliminate any chance of human bias. Automated cell counting would allow many more cells to be counted in a short space of time without the need for a human, however the very high nuclear density of the ventricular zone meant no automated cell counting programs could accurately identify individual cells.

4.4.4 Future directions

What we have presented here is a clear, descriptive study using high level histology of KIF22 protein expression in the human fetal cerebral cortex. For studying the normal expression pattern of genes in human corticogenesis, we were fortunate to be able to access human fetal tissue which is ideal for a descriptive study such as this. However, to take this study further there are several lines of experimental investigation that were beyond the scope of this thesis. One fundamental limitation of human fetal tissue is its lack of manipulability; however, this can be overcome to an extent by complementary experiments involving the use of human cerebral organoids. Before anything significant can be concluded from these artificial cortical structures, it is important to first confirm that the gene of interest is expressed in the same pattern in organoids as in the human fetal brain, which is part of why our descriptive study in human tissue is so powerful as a foundation study. It would be of great value to take this study further to manipulate *KIF22* levels in cerebral organoids and study the effects of their proliferative growth as well as growing cerebral organoids from cells obtained from *16p11.2* CNV patients.

Narrowing the picture from the context of the *16p11.2* CNV to just the *KIF22* gene, it would be interesting to understand a little more about how the many functions of the KIF22 protein contribute to normal brain development. In Figure 72 we show the many domains of KIF22 protein and their associated functions. Therefore, it could be interesting to disrupt specific KIF22 protein domains in organoids to understand more about how the individual protein

domains contribute to normal cortical proliferation. This has been previously performed in cell culture (Tokai-Nishizumi *et al.*, 2005).

Our stains (Figure 76 is a clear example), also show high KIF22 protein expression in the ganglion eminence (GE) of the brain. The GE is the birthplace of inhibitory interneurons which migrate up to the cortex as described in Figure 5, therefore studying the role of KIF22 in ventral organoids could provide interesting insight into its role in interneuron development and migration.

Overall there are many other interesting avenues for future study of *KIF22,* both as an independent gene and as part of the more complex system of the *16p11.2* CNV, to fully understand its role in neural development and how altering its dosage may contribute to the patient phenotype.

5 ALDOA protein levels do not differ between proliferating and post-mitotic cells in the developing cerebral cortex

5.1 Introduction

In chapter three we identified two candidate genes from the *16p11.2* locus whose transcripts were enriched in progenitors and significantly downregulated in post-mitotic cells. We are particularly interested in the genes whose transcripts are progenitor-enriched and then downregulated as cells become post-mitotic as they are likely to function specifically in neurogenesis. In chapter four we describe more about *KIF22* and in this chapter we will further study the second of these genes: *ALDOA*.

A summary of the analysis of the scRNA-seq dataset which identified *ALDOA* as a gene of interest is shown in Figure 90. In this violin plot we see the numbers of cells expressing different levels of *ALDOA* mRNA in the different cardinal cell classes. This shows that while the greater proportion of cells expressing the highest levels of *ALDOA* mRNA are progenitors (blue plot), there are also a substantial number of principal cells (red plot) expressing similarly high levels of *ALDOA* transcripts although very few interneurons (green plot) express high levels of *ALDOA* transcripts.

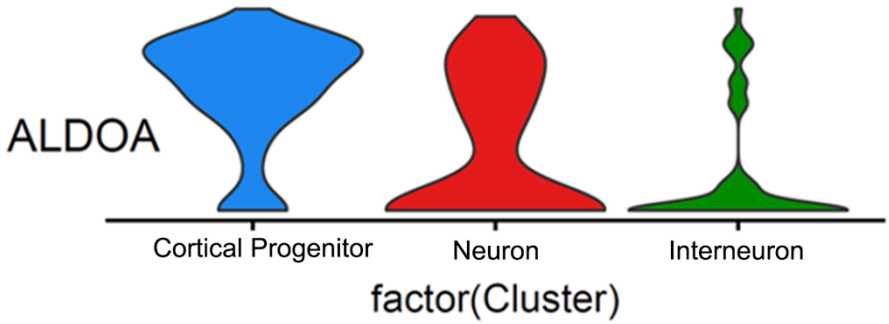


Figure 90: *ALDOA* mRNA expression

Violin plot showing the distribution of *ALDOA* in different cell types. Cortical progenitors = blue, post-mitotic neurons = red and interneurons = green.

We wanted to study the ALDOA protein expression during human cerebral cortex development to understand more about its role in neurogenesis. In this introduction we will briefly summarise what is known about ALDOA protein from the literature and outline the aims of this chapter.

5.1.1 The glycolytic pathway

The process of brain development requires a vast and consistent supply of energy. Glucose is the predominant energy substrate for the fetal brain (Gustafsson, 2009), therefore efficient and controlled glycolysis is essential for normal brain development. Glycolysis is the 10-step process by which glucose is broken down into two molecules of pyruvate and a net gain of two ATP molecules and two NADH molecules, both of which are high energy molecules. The process is shown in Figure 91 and comprise two main phases: The preparatory (green) phase, where energy is consumed, and the payoff phase (pink) which has the end net gain of high-energy molecules, ATP and NADH. This glycolysis process produces pyruvate as its final step which then feeds into the citric acid cycle facilitating further reactions within the cell. Interestingly elevated pyruvate levels have been identified in ASD children as well as alterations to the mitochondrial citric acid cycle (Giulivi *et al.*, 2010) suggesting that any variation to these pathways may contribute to the ASD phenotype.

Disruption of glycolysis is seen commonly in cancer cells, which undergo rapid and uncontrolled proliferation and perform glycolysis at ten times the rate of non-cancerous cells of the same type (Alfarouk *et al.*, 2014). Maintenance of glycolysis is essential for normal function of most, if not all, of the body's functions and disruption to any part of the pathway is likely to have wide-ranging effects.

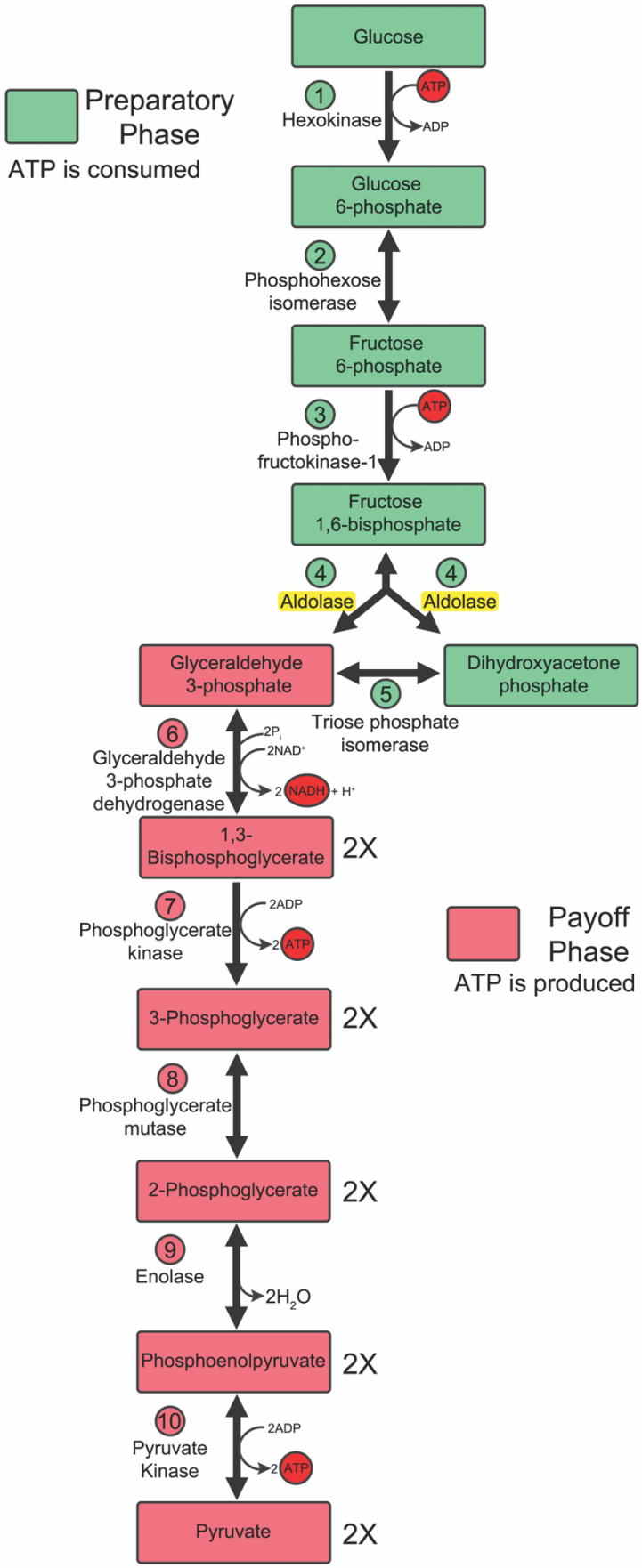

Preparatory Phase
ATP is consumed
Glucose
1
Hexokinase
ATP
ADP
Glucose 6-phosphate
2
Phosphohexose isomerase
Fructose 6-phosphate
3
Phospho-fructokinase-1
ATP
ADP
Fructose 1,6-bisphosphate
4
Aldolase
4
Aldolase
Glyceraldehyde 3-phosphate
5
Triose phosphate isomerase
Dihydroxyacetone phosphate
6
Glyceraldehyde 3-phosphate dehydrogenase
2Pi
2NAD+
2 NADH + H+
1,3-Bisphosphoglycerate
2X
7
Phosphoglycerate kinase
2ADP
2 ATP
3-Phosphoglycerate
2X
8
Phosphoglycerate mutase
2-Phosphoglycerate
2X
9
Enolase
2H2O
Phosphoenolpyruvate
2X
10
Pyruvate Kinase
2ADP
2 ATP
Pyruvate
2X
Payoff Phase
ATP is produced

Figure 91: The glycolytic pathway

This figure summarises the glycolytic pathway. This pathway comprises ten steps which can be split into two main phases: The preparatory, or investment, phase in green, and the payoff phase in pink.

Step 1) An ATP molecule is consumed from which a phosphate molecule is transferred to the Glucose producing Glucose 6-phosphate.

Step 2) Glucose 6-phosphate is rearranged into its isomer Fructose 6-phosphate.

Step 3) Another ATP molecule is consumed to turn the Fructose 6-phosphate into Fructose 1,6-bisphosphate.

Step 4) Aldolase catalyses the reversible reaction Fructose 1,6-bisphosphate to Dihydroxyacetone phosphate and Glutaraldehyde 3-phosphate.

Step 5) Dihydroxyacetone is converted into Glutaraldehyde 3-phosphate.

Step 6) Simultaneous oxidisation of Glutaraldehyde 3-phosphate and transformation of NAD^+ into NADH and H^+ providing energy to phosphorylate the oxidised molecule creating 1,3-bisphosphoglycerate.

Step 7) The first molecule of ATP is produced by 1,3-bisphosphoglycerate donating a phosphate group to ADP and transforming into 3-phosphoglycerate.

Step 8) 3-phosphoglycerate is converted to its isomer 2-phosphoglycerate.

Step 9) 2-phosphoglycerate loses a water molecule, creating the highly unstable phosphoenolpyruvate molecule.

Step 10) In the final step, the second ATP molecule is generated by the unstable phosphoenolpyruvate donating a phosphate group to ADP and becoming the end pyruvate molecule.

5.1.2 Aldolase

The fourth step of glycolysis, the reversable conversion of fructose 1,6-bisphosphate to dihydroxyacetone phosphate and glutaraldehyde 3-phosphate, is catalysed by the Aldolase family members (highlighted in yellow in Figure 91). There are three isoforms of Aldolase, each encoded by a different gene and differentially expressed during development. They are shown in Table 42 (Alfarouk *et al.*, 2014).

Table 42: Aldolase isoforms

Isoform	Gene Location	Expression Pattern of Protein
ALDOA	*16p11.2*	High in the developing embryo and in adult muscle tissue
ALDOB	*9q31.1*	Expressed in the adult liver
ALDOC	*17q11.2*	Expressed in the brain

In this chapter we are focussed on the Aldolase A (from now on this will be consistently referred to as ALDOA). *ALDOA* transcripts were identified as being significantly down-regulated as cells moved from the progenitor to post-mitotic state in our scRNA-seq analysis 3.3.1.2.

5.1.3 ALDOA

5.1.3.1 Cytoplasmic ALDOA

ALDOA is a multifunctional protein, however it was originally known for its essential cytoplasmic role in glycolysis (Horecker *et al.*, 1981). As shown in Figure 91 it catalyses the reversible reaction Fructose 1,6-bisphosphate to Dihydroxyacetone phosphate and Glutaraldehyde 3-phosphate by cleaving the C3-C4 bond. Interestingly ALDOA protein is highly expressed in many cancers, likely enhancing glycolysis allowing for the unconstrained growth

associated with cancer cells and indeed, its upregulation can correlate with overall reduced patient survival rates in several cancer types (Chang *et al.*, 2018).

However, ALDOA is a protein of many roles, and it also has other non-glycolytic “moonlighting” roles within the cytoplasm. These include regulating cytoskeleton stability and mitochondrial function by interacting with macromolecules not associated with glycolysis such as F-actin, α-tubulin and dynein light chain 8 (Pagliaro and Taylor, 1992; Kao *et al.*, 1999; Jewett and Sibley, 2003; Buscaglia *et al.*, 2006; Merkulova *et al.*, 2011). These roles of ALDOA protein, both metabolic and other moonlighting functions, are all cytoplasmic and indeed, ALDOA does typically localise to the cytoplasm. However, there is evidence for ALDOA protein to localise to the nucleus.

5.1.3.2 Nuclear ALDOA protein

Despite its stereotypical cytoplasmic localisation, ALDOA protein has been identified in the nucleus of certain cell types. It was first observed in the nucleus of pig cardiomyocytes in 2004 (Mamczur and Dzugaj, 2004) and then in porcine smooth muscle cells in 2008 (Mamczur and Dzugaj, 2008). Perhaps of most interest to this thesis was a study in 2013 that showed ALDOA to localise to the nucleus of proliferating cells (Mamczur *et al.*, 2013), and a complementary study the previous year showed silencing of ALDOA inhibited cell proliferation (Ritterson Lew and Tolan, 2012) suggesting that ALDOA acts to affect cellular proliferation through a mechanism unrelated to its glycolytic functions. However, this effect on proliferation could also be related to ALDOA’s metabolic role as cells will require energy to divide.

One proposed nuclear mechanism is that ALDOA influences the cell cycle by positively regulating cyclin D1 expression to mediate G1/S phase progression (Fu *et al.*, 2018). Therefore, it appears that the nuclear localisation of ALDOA can allow it to exert an effect directly on cellular

proliferation independent of glycolysis, an interesting consideration for our result that showed the *ALDOA* transcript to be highest in the proliferating progenitors. An interesting and unanswered question is the mechanism by which ALDOA protein is transported to the nucleus, as the protein lacks a typical nuclear localisation sequence (Mamczur *et al.*, 2013).

5.1.4 ALDOA in the human fetal cortex

The literature suggests a variety of roles for ALDOA protein, including both cytoplasmic glycolytic function and mediating cellular proliferation through nuclear localisation, making it a very interesting candidate for study. We know from our scRNA-seq analysis in 3.3.1.2 that ALDOA transcripts are highly expressed in progenitors and down-regulated as cells become post-mitotic, and our *in situ* hybridisation in 3.3.5.4 results show its mRNA to be very strongly expressed in the germinative zones. While there are several published studies, as described above, on the many functions of ALDOA protein, nothing is known about its expression or potential roles during human fetal cortex development. Indeed, the majority of ALDOA protein studies have used highly abnormal cancer tissue, or artificial cell culture systems in which glycolytic enzymes have been shown to be increased. scRNA-seq analysis showed increased *ALDOA* mRNA in human cerebral organoids compared to human fetal brain samples suggesting that cell culture systems may not fully reflect the physiological roles of ALDOA protein (Ritterson Lew and Tolan, 2012; Mamczur *et al.*, 2013; Fu *et al.*, 2018; Pollen *et al.*, 2019). Therefore, a key question is to ask how these observations of ALDOA in a variety of systems relate to its role in human cerebral cortex development.

5.1.5 Chapter aims

This chapter has the following aims:

1) To characterize the anatomical expression pattern of the ALDOA protein during human fetal cortical development using high-level immunohistochemistry.
2) To identify if, in this system, ALDOA protein levels are higher in proliferating cells and if the levels vary with the cell cycle.
3) To determine if ALDOA protein nuclear localisation is higher in proliferating cells and if nuclear localisation varies with the cell cycle.

5.2 Methods

5.2.1 Fetal tissue used

The fetal tissue used for this chapter was paraffin embedded tissue from the HDBR. Details of tissue preparation is provided in 2.1.1 and the ages used are shown in Table 43.

Table 43: Fetal ages used in this chapter

Age	Anatomical location
12pcw	Middle
14pcw	Middle
16pcw	Middle

5.2.2 Immunofluorescence of ALDOA and KI67

For our immunofluorescence study of ALDOA, the IF protocol was performed as described in 2.3.1. The antibodies used for this chapter are detailed in Table 44.

Table 44: Antibodies used for ALDOA IF

Antibody	Concentration	Vendor	Secondary Antibody
ALDOA rabbit polyclonal	1:100	Sigma-Aldrich	Alexa Fluor Goat anti-rabbit 568
KI67 mouse IgG1 monoclonal	1:750	Novus Biological	Alexa Fluor Goat anti-mouse 488 IgG1

5.2.2.1 Cell Counting for immunofluorescence

For the IF study of ALDOA at the single-cell level we performed high-magnification confocal imaging of tissue from the germinative zones: the VZ and SVZ. Using FIJI, a grid of counting boxes (20x145μm) was overlaid and cells were randomly selected on the DAPI channel ensuring their selection was blind to their ALDOA and KI67 staining. As shown in Figure 92 a, we ensured that once a cell was included for analysis, all nuclei adjacent to it were excluded from analysis to ensure their cytoplasm's would not overlap. For each cell nucleus we then moved the Z-plane to focus through the middle of the cell (the widest part) and we drew a region of interest (ROI) around the nucleus, as shown in Figure 92 b, which was identified by the DAPI stain. We then took measurements of this nuclear ROI on the green KI67 and red ALDOA channels and recorded that as the nuclear mean fluorescence value for KI67 and ALDOA.

Next we wanted to obtain the ALDOA cytoplasmic intensity, our method for this is shown in Figure 92 b. To do this we traced around the nuclear outline as shown by the DAPI stain with a four-pixel thick tool to create a "donut" which is indicated by the cytoplasm label in Figure 92 b. Switching then to the red ALDOA channel, this allowed us to obtain a reading of the mean fluorescence of ALDOA protein in just the cytoplasmic area. To produce the final cell body ALDOA value we added together the mean fluorescence values of the nucleus and the cytoplasm of each cell. This analysis was performed for ten cells in each box and the counts from individual boxes combined to give final values.

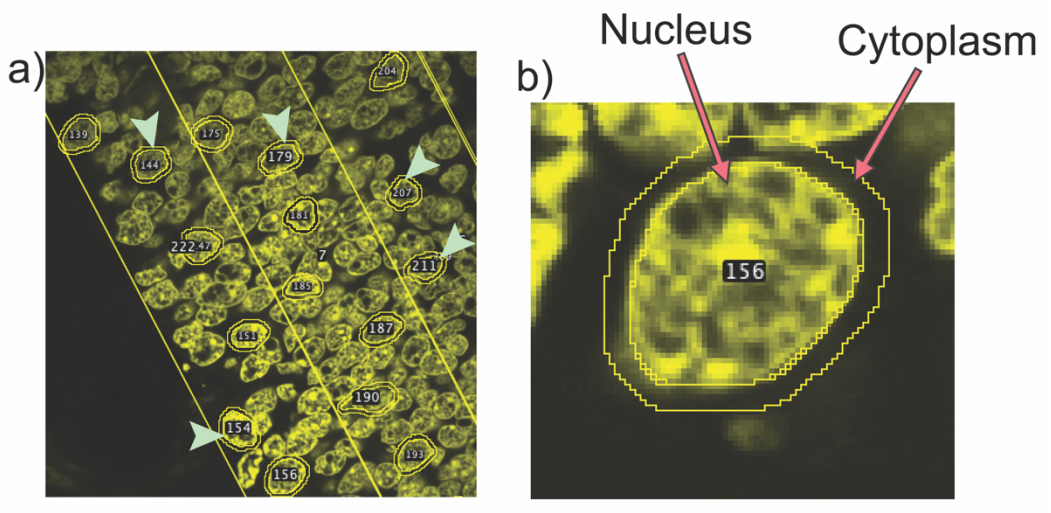


Figure 92: Method of cell selection and delineating nucleus and cytoplasm

a) Our method of selecting cells for analysis. Cell nuceli were selected blind on the DAPI channel and cells with an adjacent nucleus to an already included cell were excluded to ensure their cytoplasm's would no overlap. Ten cells were counted per box and examples of cells included are indicated by green arrows.

b) Our method of delineating the nucleus and cytoplasm. This shows the DAPI channel which stains the cell nucleus. We drew around the nucleus and took the mean fluorescence value of KI67 and ALDOA, and then drew an additional four-pixel ring around this to allow an approximation of the cytoplasm and recorded the mean fluorescence value of ALDOA.

5.2.2.2 Data analysis and statistics for immunofluorescence

We examined ALDOA and KI67 levels by recording fluorescent intensity as described above. Cells were considered $KI67^-$ if their expression level fell below that of a background reading taken from recording mean fluorescence of several clearly $KI67^-$ cells and averaging those values.

We wanted to determine if cell body ALDOA levels differed between $KI67^+$ and $KI67^-$ so for each cell we combined the mean fluorescent value for the nucleus and cytoplasm to give the cell body value. Prior to statistical testing we log transformed the data to produce normal distribution where possible. However, as some values = 0 it was necessary to do a +1(log) transformation where we added a value of one to every data point prior to the log transformation. We then created boxplots from the individual cell values in each sample, determined the distribution of the data and performed the appropriate test based on the distribution for each data. To determine distribution, data distribution was displayed as a histogram and distribution determined by eye: bimodal distribution required a Mann-Whitney test and normal distribution required an unpaired t-test with Welch's correction. Our analysis is accompanied by the clear caveat that any result is derived from pseudo-replication as it a dataset from within a single sample, however it allowed us to gauge if there was significant difference in any of the samples. As we saw no difference at any age, we then combined the data by averaging the ALDOA values for each age in $KI67^+$ and $KI67^-$ cells. We considered each age a separate n value, +1(log) transformed the data and then performed a paired t-test which overcame the above issue of pseudoreplication.

For determining if nuclear ALDOA levels were higher in KI67^{+} cells compared to KI67^{-} cells we used only nuclear values from the cells measured. We handled and transformed the data in the same way as above.

All analysis was performed, and graphs generated, in GraphPad Prism version 6.0 with the exception of the correlation graphs which were created in R studio.

5.3 Results

5.3.1 ALDOA protein is highest in the germinal zones of the cortex

In chapter three our bioinformatics analysis in 3.3.1.2, and our *in situ* hybridisation in 3.3.5.4, show that *ALDOA* mRNA levels are highest in progenitor cells and decrease as cells move towards a neuronal fate. First, we wanted to characterize ALDOA protein expression across the telencephalic wall, for which we used immunofluorescence to visualise ALDOA protein at three developmental timepoints: 12, 14 and 16pcw. Our results are shown in Figure 93 (a-c) and are consistent across the ages; ALDOA immunofluorescence appears most intense in the proliferative VZ and SVZ before decreasing in the intermediate zone and cortical plate in a pattern that matches the mRNA in 3.3.5.4. Looking at higher magnification in Figure 93 d, with double immunofluorescence for KI67 and ALDOA and all cell nuclei stained with DAPI, we can see that the ALDOA protein (red) is predominantly localized outside the blue $DAPI^{+}$ nucleus in the cytoplasm and ALDOA is expressed by both $KI67^{+}$ proliferating cells (green) and non-proliferating $KI67^{-}$ cells.

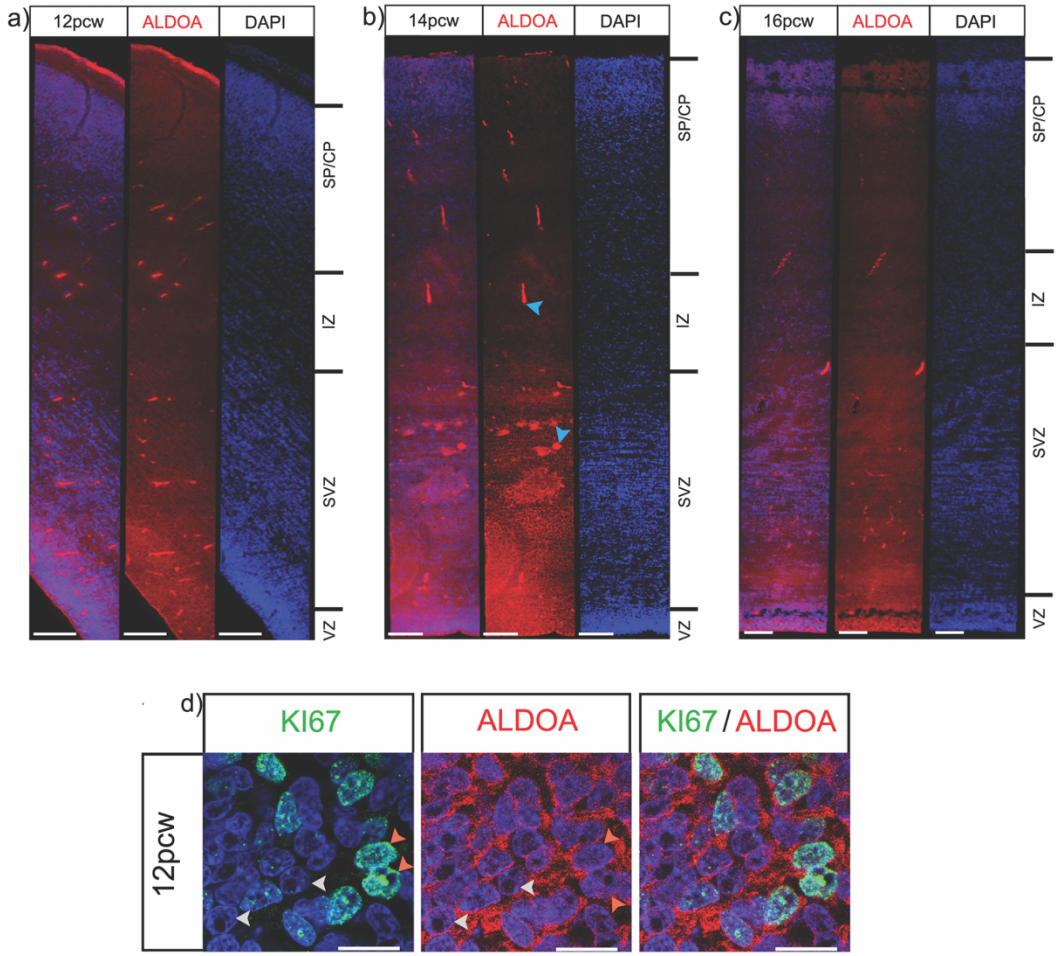


Figure 93: ALDOA expression in the cortex

ALDOA protein fluorescence intensity across the telencephalic wall at a) 12, b) 14 and c) 16pcw. Scale bars = 100μm. d) High magnification immunofluorescence of KI67 and ALDOA proteins. Scale bar = 10μm. Blue = DAPI, green = KI67, red = ALDOA
On b) blue arrows indicate non-specific antibody binding to blood vessels. On d) grey arrows indicate KI67^{-} cells and ALDOA expression in KI67^{-} cells. Orange arrows indicate KI67^{+} cells and ALDOA expression in KI67^{+} cells.

5.3.2 Cell body ALDOA protein levels are not higher in proliferating cells

Our scRNA-seq analysis in Figure 90 indicated that ALDOA mRNA was enriched in progenitor cells. To see if this was also the case for the protein, we used our analysis of immunofluorescence to compare cell body ALDOA protein levels in proliferating $KI67^{+}$ and non-proliferating $KI67^{-}$ cells. Our results are shown in Figure 94 and we saw no significant difference at 12pcw (Figure 94 a), 14pcw (Figure 94 b) nor 16pcw (Figure 94 c). Combining all ages as described in 5.2.2.2, also revealed no significant difference in the cellular levels of ALDOA between proliferating and non-proliferating cells (Figure 94 d).

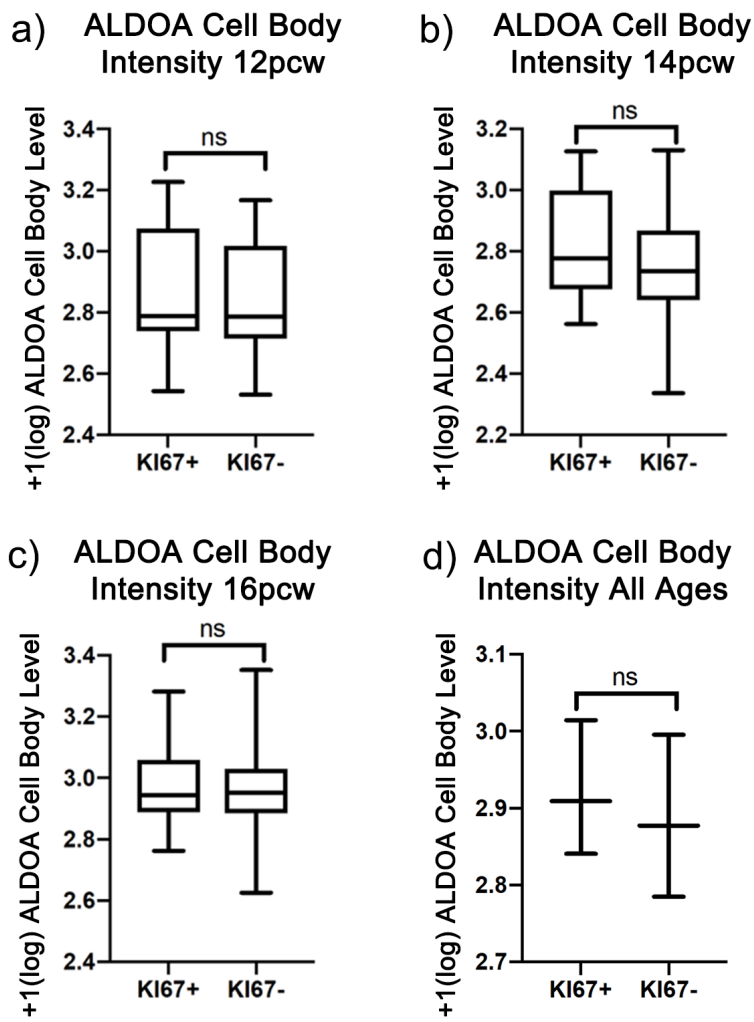


Figure 94: Cell body ALDOA quantification

Cell body ALDOA protein fluorescence intensity in KI67^{+} and KI67^{-} cells.

a) 12pcw, raw data transformation = +1(log), bimodal distribution, Mann-Whitney test, p = 0.3702.
b) 14pcw, raw data transformation = +1(log), normal distribution, unpaired t-test with Welch's correction, p = 0.2032.
c) 16pcw, raw data transformation = +1(log), normal distribution, unpaired t-test with Welch's correction, p = 0.3523.
d) 12, 14 and 16pcw individual datasets averaged, raw data transformation = +1(log), paired t-test, p = 0.0836.

Next, we wanted to see if there was any fluctuation in ALDOA protein levels with the cell cycle, so we quantified immunofluorescence for KI67 and cell body ALDOA (nucleus and adjacent cytoplasm) using the same analysis as for KIF22 as described in (4.2.3). Our analysis presented in Figure 95 showed no correlation or pattern at 12pcw (Figure 95 a) (R^2=0.018), 14pcw (Figure 95 b) (R^2=2e-4), 16pcw (Figure 95 c) (R^2=0.00992) or when we combined all the data points from the three ages onto one graph (Figure 95 d) (R^2=0.0049). These data show that in human cerebral cortex development, cellular ALDOA protein levels do not correlate with proliferation nor fluctuate with the cell cycle.

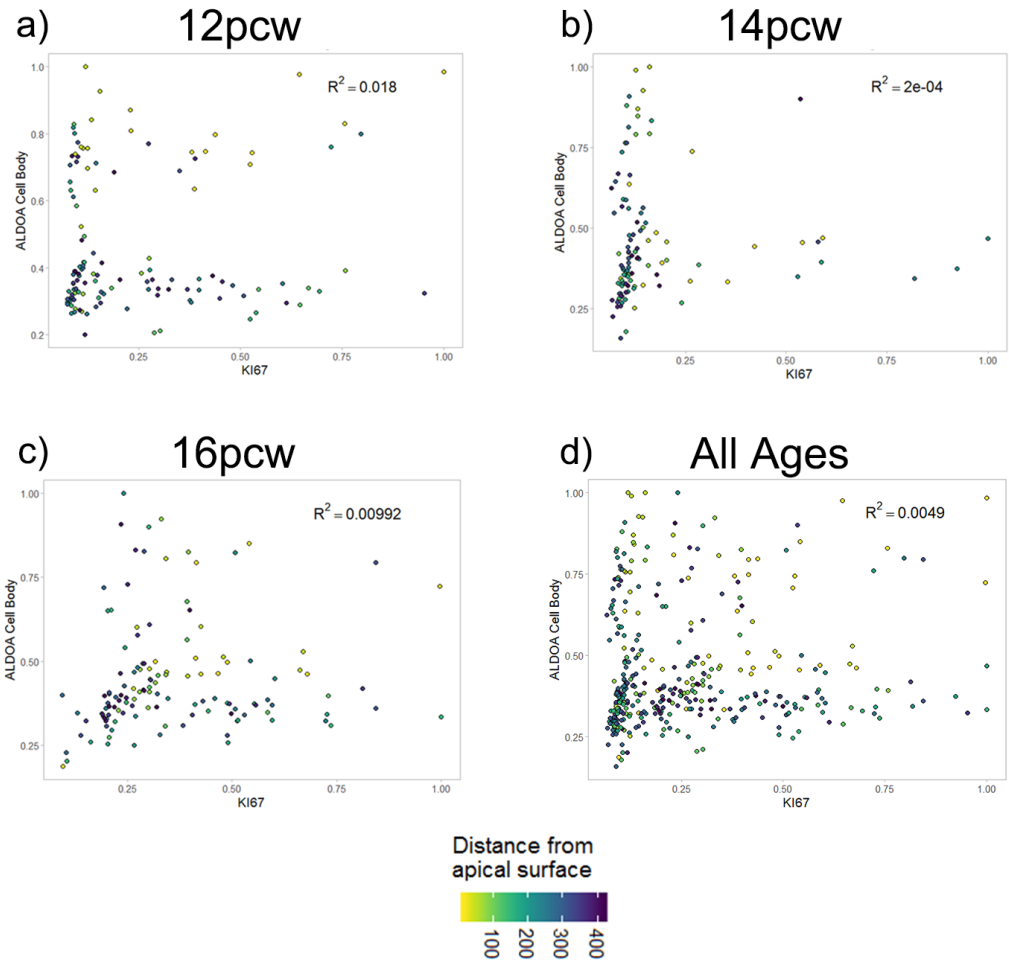


Figure 95: Quantification of cell body ALDOA and nuclear KI67

Correlation of cell body ALDOA and nuclear KI67 fluorescence intensity levels at a) 12pcw, b) 14pcw, c) 16pcw and d) data from all the ages combined onto one graph. Colour coding indicates distance of each cell from the apical surface (key at bottom).

5.3.3 Nuclear ALDOA protein levels are not higher in proliferating cells

As described in the introductory part of this chapter (5.1.3.2), previous work in different systems has shown ALDOA protein to localise to the nucleus of proliferating cells (Mamczur *et al.*, 2013). We wanted to see if this was the case in the developing human cortex, so we quantified both nuclear ALDOA and KI67 to compare ALDOA protein levels in the nucleus of proliferating $KI67^+$ cells and non-proliferating $KI67^-$ cells at three developmental ages as shown in Figure 96. We saw no significance difference in nuclear ALDOA fluorescence intensity between $KI67^+$ and $KI67^-$ cells at 12pcw (Figure 96 a), 14pcw (Figure 96 b), 16pcw (Figure 96 c), nor when we combined the average values for each of the ages as in Figure 96 d.

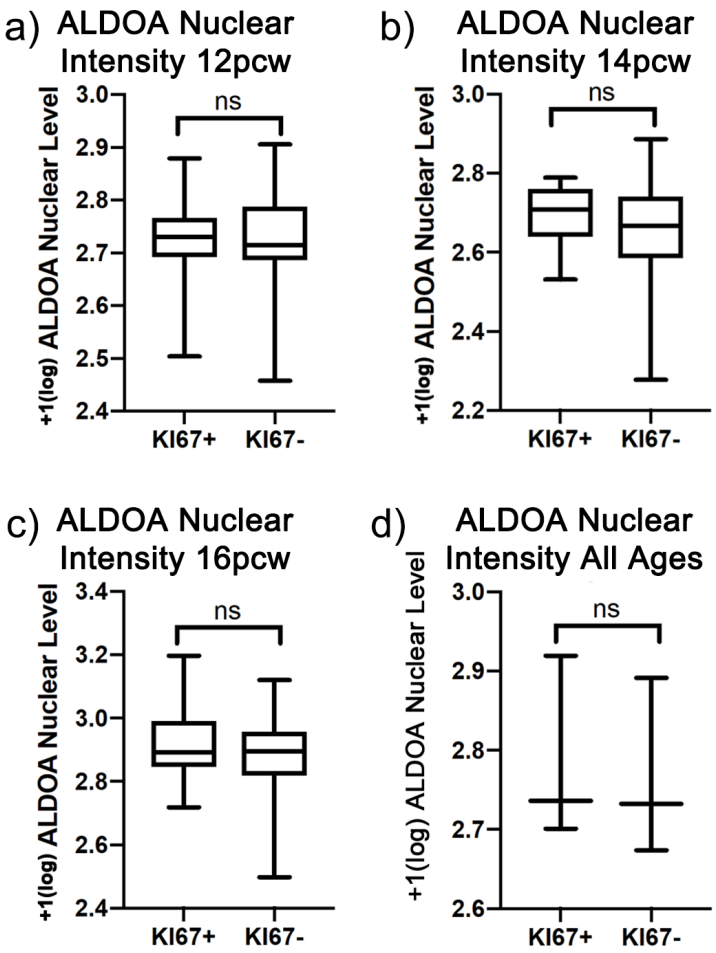


Figure 96: Nuclear ALDOA quantification

Nuclear ALDOA fluorescence intensity in $KI67^+$ and $KI67^-$ cells

a) 12pcw, raw data transformation = +1(log), normal distribution, unpaired t-test with Welch's correction, p = 0.7543.

b) 14pcw, raw data transformation = +1(log), normal distribution, unpaired t-test with Welch's correction, p =0.0694.

c) 16pcw, raw data transformation = +1(log), normal distribution, unpaired t-test with Welch's correction, p = 0.0772

d) 12, 14 and 16pcw individual datasets averaged, raw data transformation = +1(log), paired t-test, p = 0.1330.

Next, we wanted to see if the nuclear localisation of ALDOA correlated in any way with the cell cycle by analysing nuclear ALDOA and KI67 fluorescence intensity with the results in Figure 97. Our analysis established no correlation or pattern at 12pcw (Figure 97 a) (R^2 = 5e-05), 14pcw (Figure 97 b) (R^2 = 0.0365), 16pcw (Figure 97 c) (R^2 = 0.0723) nor when we combined all the data points from the three ages onto one graph (Figure 97 d) (R^2 = 6e-04).

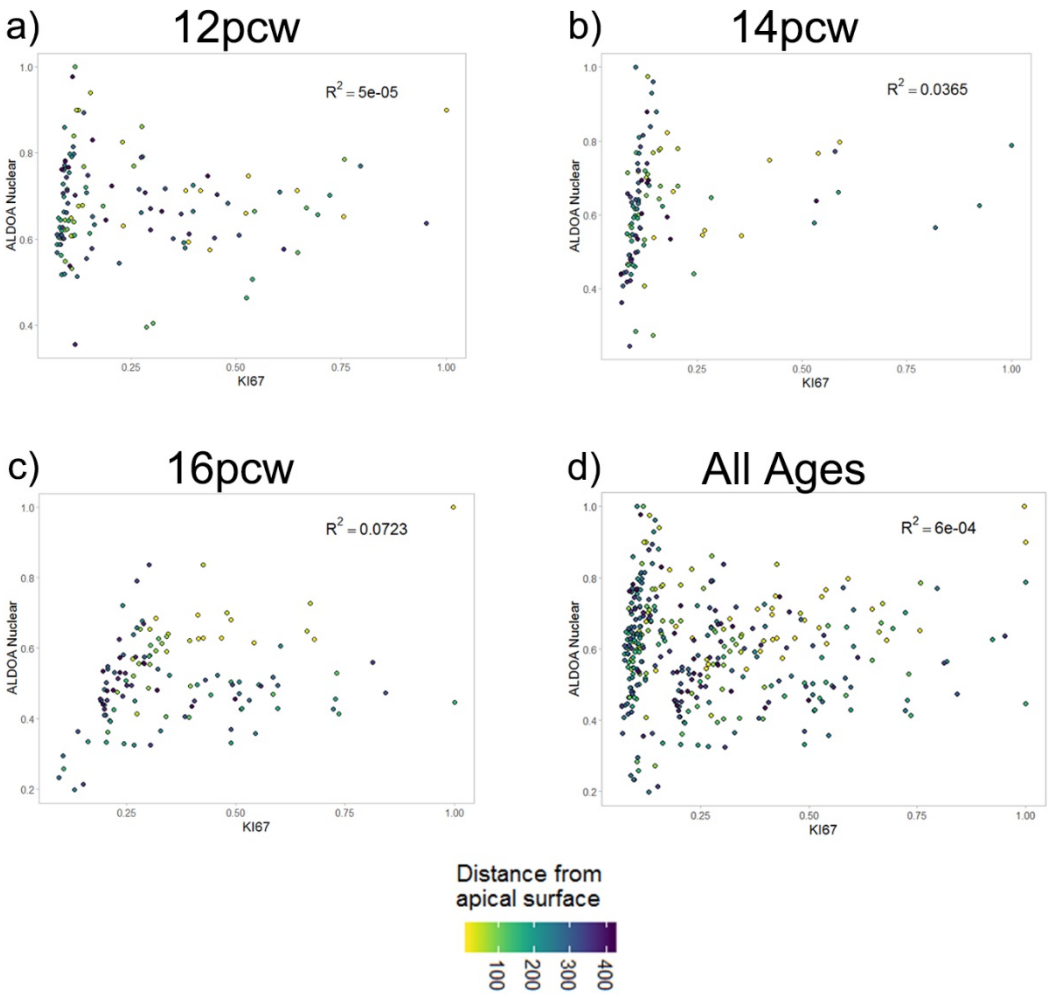


Figure 97: Quantification of nuclear ALDOA and KI67 fluorescence intensity

Correlation of nuclear ALDOA and nuclear KI67 fluorescence intensity levels at a) 12pcw, b) 14pcw, c) 16pcw and d) data from all the ages combined onto one graph. Colour coding indicates distance of each cell from the apical surface (key at bottom).

5.3.4 Summary of results

Our results have shown ALDOA protein to be present across the telencephalic wall, with stronger staining in the germinative zone. At the sub-cellular level, we have shown ALDOA protein to be predominantly localised to the cytoplasm of the cell, with a small amount in the nucleus; an expression pattern that does not differ between proliferating and non-proliferating cells as summarised in Figure 98.

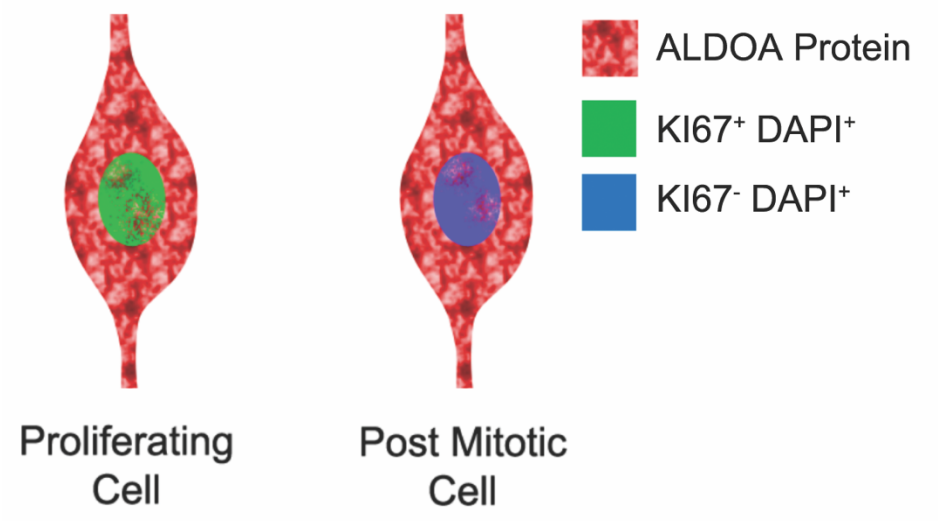


Figure 98: Schematic of ALDOA protein localisation

Summary schematic showing our findings that ALDOA protein is predominantly located in the cytoplasm and less in the nucleus in both KI67^{+} proliferating cells and KI67^{-} post-mitotic cells.

We have also shown that, unlike the KIF22 protein, ALDOA protein levels do not change in relation to the cell cycle as shown in summary Figure 99.

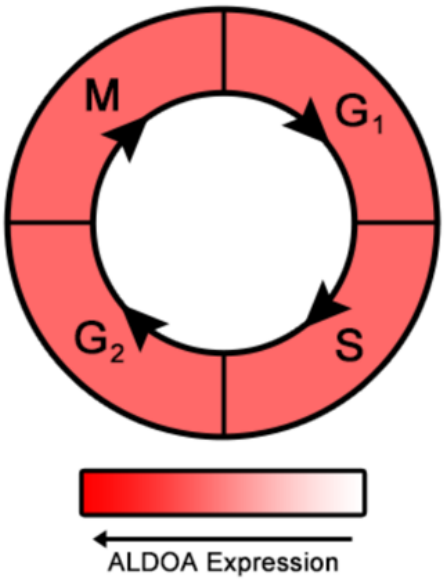


Figure 99: ALDOA cell cycle summary

Model showing that ALDOA protein levels do not change with the cell cycle.

5.4 Discussion

In beginning to understand the many ways by which the polygenic *16p11.2* CNV affects prenatal brain development, we must understand the roles of its individual genes in neural development. As altered brain size at birth is one of the most common patient phenotypes, this suggests some of the genes in the CNV may impact neural proliferation and ultimately brain size. Hence, we aimed to identify genes from this locus whose transcripts were enriched in progenitors and downregulated as cells became post-mitotic. Our scRNA-seq analysis in chapter three presented *ALDOA* as a candidate for further study. We have shown its protein to be expressed across the telencephalic wall, highest in the germinative zones and the protein to be predominantly localised to the cytoplasm in both proliferating and non-proliferating cells of the germinative zones. In this discussion section of chapter five we will discuss our results, the limitations of our study and avenues of future exploration.

5.4.1 ALDOA protein expression in the developing cerebral cortex

At the tissue level, our results show ALDOA protein to be expressed across the width of the cortex, highest in the germinative regions, the VZ and SVZ. This result matches that of the *in situ* hybridisation demonstrating that *ALDOA* mRNA and protein are expressed in a similar pattern. However, some of this result, may arise from the fact that the germinative zones, and in particular the VZ, have such a high cell density. With the fibres from germinative zone cells extending all the way to the pial surface it is also impossible to know whether cytoplasmic ALDOA observed in the upper regions belongs to the cells in those regions, is localised to the fibres of VZ and SVZ progenitor cells, or both possibilities. As the scRNA-seq analysis focusses on the VZ and SVZ, these are the regions we focused on for our single-cell protein analysis.

5.4.2 ALDOA subcellular localisation

Our results show that at all ages ALDOA predominantly localises to the cytoplasm, and its level in cells does not vary between proliferating and non-proliferating cells. From our study of the literature, we found that nuclear ALDOA localisation has been previously linked to proliferation (Mamczur *et al.*, 2013; Fu *et al.*, 2018) and hypothesised that may also be the case in the human fetal cortex. However, while we were able to detect ALDOA in the nucleus, it was predominantly localised to the cytoplasm and we found no clear relationship between ALDOA protein levels and cellular proliferation state.

This finding differs to our result from the scRNA-seq analysis which showed *ALDOA* mRNA levels to be higher in proliferating than non-proliferating cells. Our quantitative analysis of ALDOA protein did not identify any increase in ALDOA levels in the cell body of proliferating KI67$^+$ cells. One possibility is that this could be a technical problem with our analysis, however this is unlikely as we used an identical analysis to determine KIF22 protein levels are increased in KI67$^+$ cells in 4.3.2.1. However, given that neuronal cells have an elongated shape, with processes extending from the nucleus, rather than the more stereotypical “fried egg” shape and that ALDOA protein (unlike KIF22) fills the whole cell, and our analysis cannot identify protein localised to the radial fibres of neural progenitors, nor that localised to the neurites of post-mitotic cells. However, given how densely packed with cells the developing brain is, it would be impossible to use a stain like N-cadherin to identify individual cell bodies and their fibres as we would not be able to determine which process belonged to which cell.

We wanted to consider possible explanations for why *ALDOA* mRNA does not correlate well with the protein levels. One possibility is that *ALDOA* mRNA expression is higher in progenitors and subsequently downregulated

as cells become post-mitotic, but the ALDOA protein is very stable, so it persists in post-mitotic cells after *ALDOA* mRNA levels decline. A second possibility is that the total ALDOA protein levels are higher in progenitors but the subcellular distribution of the protein changes as cells become post-mitotic, e.g. ALDOA protein in the radial processes of progenitors redistributes towards the nucleus as cells become postmitotic, such that the levels of protein in and around the nucleus remain constant. These two possibilities are summarised in Figure 100.

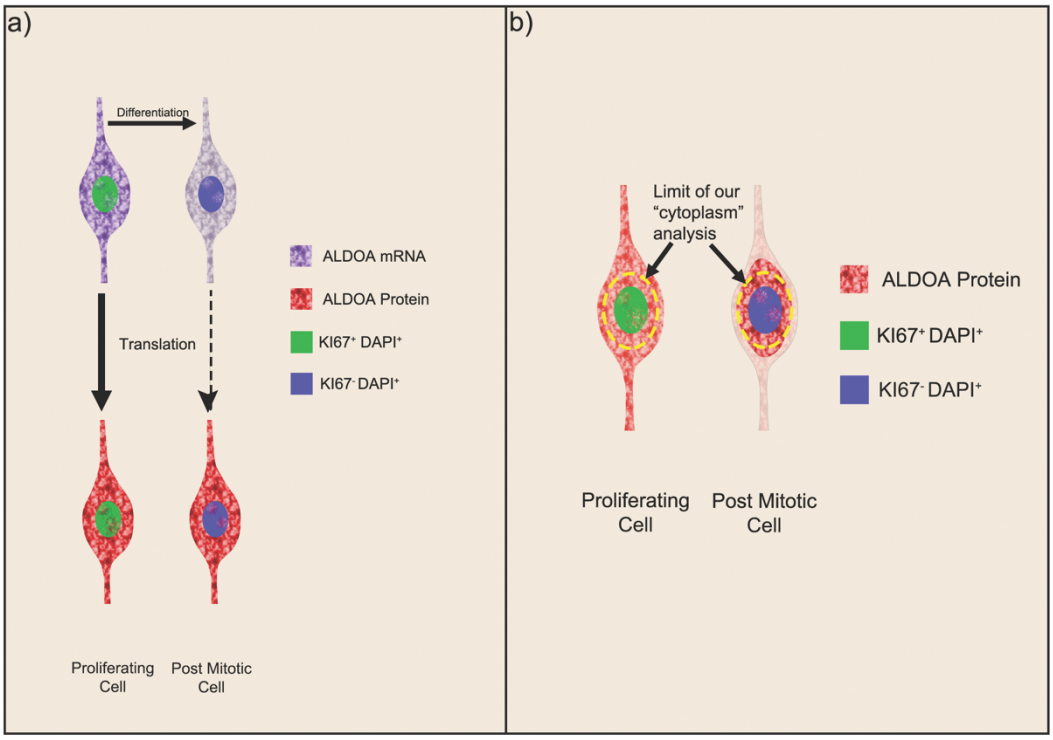


Figure 100: Possibilities for ALDOA protein levels

Schematic showing two possible explanations for why *ALDOA* mRNA does not correlate well with the protein levels.

a) In possibility one, the mRNA expression (purple) is higher in the progenitors and decreasing when the cell becomes post-mitotic but the protein (red) is stable and persists in the post-mitotic cells.

b) In the second possibility, there is a redistribution of the protein as cells become post-mitotic and while the progenitors have higher overall protein level it is widely distributed and when cells become post-mitotic the protein redistributes towards the nucleus and the area we considered to be the cytoplasm in our analysis which is indicated in yellow. Therefore, we saw no difference within our area of analysis which may not be representative of the entire cell.

With the post-mortem tissue available for this study we cannot distinguish between these two possibilities. One solution to this could be to test in dissociated cell culture where the whole cell is visible, but this was not possible in this work and in any case such an experiment would not be physiological which is a particular concern as *ALDOA* mRNA is known to be elevated in some culture contexts (Pollen *et al.*, 2019). Overall our results show that ALDOA protein is abundant in progenitor cells of the developing human cerebral cortex and therefore, may well play a role in neurogenesis phenotypes. However, the fact that we see persistent expression of ALDOA protein as cells become post-mitotic suggests potential roles for ALDOA in differentiated neurons. Some of these will likely be glycolytic as neurons are metabolically active, however given its known interactions with the cytoskeleton (Pagliaro and Taylor, 1992), it's possible ALDOA may have additional roles in neuronal migration or axon branching.

5.4.3 Altering ALDOA dosage

When ALDOA dosage is altered in the developing brain, as in the CNV, it is likely to substantially impact energy metabolism by altering the flow of metabolites through the glycolytic pathway and impacting subsequent pathways that feed on the outputs of glycolysis. There is some evidence for this in the literature: disruption to energy metabolism during development has previously been linked to ASD and ADHD (Rash *et al.*, 2018). Additionally the offspring of hyperglycaemic mice presented microcephaly, a phenocopy of the microcephaly observed in *16p11.2* microduplication patients (Rash *et al.*, 2018). Disruptions to energy metabolism, including increased glycolysis, may contribute to the microcephaly seen in the offspring of Zika infected mothers which is also a phenocopy (albeit Zika microcephaly is much more severe) of the *16p11.2* microduplication patients (Gilbert-Jaramillo *et al.*, 2019). So, although the interactions between these pathways are highly

complex, there is some evidence suggesting increased glucose of glycolysis contribute to microcephaly. Therefore, this suggests that disrupting glycolysis by altering ALDOA dosage will have significant effects on brain growth and potentially cellular composition.

Looking at medical literature revealed no homozygous null ALDOA patients, suggesting that ALDOA is essential for life and its functions cannot be compensated for by the ALDOB or ALDOC isoforms. However, there were some case studies in which patients with changes to ALDOA levels were identified. One patient presented with microcephaly, and was found to have reduced ALDOA activity (Kreuder *et al.*, 1996), and another presented reduced ALDOA activity and intellectual disability (Beutler *et al.*, 1973). Interestingly, one case study identified upregulated cortical ALDOA levels in post-mortem brains of schizophrenia patients (Beasley *et al.*, 2006) which is particularly noteworthy as the *16p11.2* microduplication is strongly associated with risk of schizophrenia. This, coupled with our results that ALDOA is expressed in all cell types of the developing human cerebral cortex, makes it clear that any changes to ALDOA dose will affect energy metabolism at many stages in the brain, impacting development.

5.4.4 Future directions

Here we have presented a clear, descriptive study to show the expression pattern and sub-cellular localisation of ALDOA protein in the developing human cerebral cortex. In order to take this study further, there are a few interesting lines of investigation that were outwith the scope of this PhD project. A limitation of our study was our inability to clearly visualise the entire cell due to the density of the post-mortem brain tissue. One option for building on this could be to dissociate cells and image the whole cell, however this would need to be considered carefully for a protein like ALDOA, whose mRNA transcript levels in artificial cell culture have been shown to be elevated compared to post-mortem tissue samples (Pollen *et al.*, 2019). A follow-on question from this could be to again utilise cell culture, using hIPSC derived NPCs where ALDOA levels have been reduced and asses any impact on cellular proliferation.

Considering ALDOA protein within the context of the *16p11.2* CNV, it would be interesting to again use cell culture where NPCs with the *16p11.2* CNV are used to determine if ALDOA protein sub-cellular localisation differs between *16p11.2* CNV and control lines, and in response to addition of different metabolic substrates.

Overall, there are many interesting future studies for *ALDOA* both as an independent gene and as part of the *16p11.2* patient phenotype and there are many unanswered questions to fully understand its role in human neurogenesis.

Summary and concluding remarks

In this thesis we have presented a descriptive study utilising bioinformatics analysis, *in situ* hybridisation of mRNA and high-level histology of protein expression in the developing human cerebral cortex. We began examining the expression level of the transcripts of each gene in the ASD-implicated *16p11.2* CNV locus to identify which genes had transcripts enriched in progenitor cells and down-regulated as cells become post-mitotic with the reasoning that these genes would be important for regulating neurogenesis. Following the scRNA-seq analysis and complementary *in situ* hybridisation we identified two significantly progenitor enriched candidate genes for further study: *ALDOA* and *KIF22*. We then proceeded to characterize their protein expression in sections of the developing human fetal cortex during the late first and early second trimester. We showed KIF22 protein to be restricted to proliferating cells and its levels to vary with the cell cycle, while ALDOA was expressed throughout the cortex although highest in the germinative zones and predominantly localised to the cytoplasm in both proliferating and non-proliferating cells. This work is publicly available in a manuscript published on preprint server BioRXiv (Morson *et al.*, 2019) (see appendix).

This is a powerful descriptive study using human fetal material, and it has indicated that further investigation into the normal function of *ALDOA* and *KIF22* and the consequences of altering their dosage on neurogenesis is likely to further our understanding of how the *16p11.2* CNV impacts brain development. We have set the foundations for describing these genes in the developing brain and proposed hypotheses about their function for future testing in more genetically manipulatable systems such as cerebral organoids.

It is important to also consider that the *16p11.2* CNV is a polygenic mutation and there are likely to a multitude of interactions between the CNV genes which contribute to the phenotype. Therefore, it is important to study these genes both independently and as part of the CNV to fully understand their

role in neural development, and the impact of altering their dosage. Only by considering both these approaches will we fully begin to understand the genetic basis of the *16p11.2* pathogenesis.

References

Abrahams, B. S. and Geschwind, D. H. (2008) ‘Advances in autism genetics: on the threshold of a new neurobiology.’, *Nature reviews. Genetics*. England, 9(5), pp. 341–355. doi: 10.1038/nrg2346.

Aldinger, K. A. (2009) ‘Copy number variation at 1q21.1 results in a spectrum of developmental abnormalities’, *Clinical Genetics*. John Wiley & Sons, Ltd (10.1111), 75(5), pp. 425–427. doi: 10.1111/j.1399-0004.2009.01186_2.x.

Alfarouk, K. O. *et al.* (2014) ‘Glycolysis, tumor metabolism, cancer growth and dissemination. A new pH-based etiopathogenic perspective and therapeutic approach to an old cancer question’, *Oncoscience*. Impact Journals LLC, 1(12), pp. 777–802. doi: 10.18632/oncoscience.109.

Allach El Khattabi, L. *et al.* (2018) ‘16p13.11 microduplication in 45 new patients: refined clinical significance and genotype-phenotype correlations.’, *Journal of medical genetics*. England. doi: 10.1136/jmedgenet-2018-105389.

Allen, R. D. *et al.* (1982) ‘Fast axonal transport in squid giant axon.’, *Science (New York, N.Y.)*. United States, 218(4577), pp. 1127–1129. doi: 10.1126/science.6183744.

Amaral, D. G., Schumann, C. M. and Nordahl, C. W. (2008) ‘Neuroanatomy of autism.’, *Trends in neurosciences*. England, 31(3), pp. 137–145. doi: 10.1016/j.tins.2007.12.005.

Association, A. P. and Force, A. P. A. D.-5 T. (2013) *Diagnostic and statistical manual of mental disorders : DSM-5.* Fifth edit, *DSM-5*. Fifth edit. Arlington, VA: Arlington, VA : American Psychiatric Association.

Ayala, R., Shu, T. and Tsai, L.-H. (2007) ‘Trekking across the brain: the journey of neuronal migration.’, *Cell*. United States, 128(1), pp. 29–43. doi: 10.1016/j.cell.2006.12.021.

Balan, S. *et al.* (2013) ‘Major vault protein (MVP) gene polymorphisms and drug resistance in mesial temporal lobe epilepsy with hippocampal sclerosis.’, *Gene*. Netherlands, 526(2), pp. 449–453. doi: 10.1016/j.gene.2013.05.067.

Barnabe-Heider, F. *et al.* (2005) ‘Evidence that embryonic neurons regulate the onset of cortical gliogenesis via cardiotrophin-1.’, *Neuron*. United States, 48(2), pp. 253–265. doi: 10.1016/j.neuron.2005.08.037.
Barres, B. A. and Barde, Y.-A. (2000) ‘Neuronal and glial cell biology’, *Current Opinion in Neurobiology*, 10(5), pp. 642–648. doi: https://doi.org/10.1016/S0959-4388(00)00134-3.
Bayatti, N. *et al.* (2008) ‘A molecular neuroanatomical study of the developing human neocortex from 8 to 17 postconceptional weeks revealing the early differentiation of the subplate and subventricular zone’, *Cerebral cortex (New York, N.Y. : 1991)*. 2007/10/26. Oxford University Press, 18(7), pp. 1536–1548. doi: 10.1093/cercor/bhm184.
Bayer, S. A. and Altman, J. (2002) *The Human Brain During the Late First Trimester*. Taylor & Francis (Atlas of Human Central Nervous System Development). Available at: https://books.google.co.uk/books?id=ck4PAQAAMAAJ.
Bayer, S. A. and Altman, J. (2005) *The Human Brain During The Second Trimester*. CRC Press Taylor & Francis Group (Atlas Of Human Central Nervous System Development). Available at: https://books.google.co.uk/books?id=-W1yPwAACAAJ.
Beasley, C. L. *et al.* (2006) ‘Proteomic analysis of the anterior cingulate cortex in the major psychiatric disorders: Evidence for disease-associated changes.’, *Proteomics*. Germany, 6(11), pp. 3414–3425. doi: 10.1002/pmic.200500069.
Bedoyan, J. K. *et al.* (2010) ‘Duplication 16p11.2 in a child with infantile seizure disorder.’, *American journal of medical genetics. Part A*. United States, 152A(6), pp. 1567–1574. doi: 10.1002/ajmg.a.33415.
Benger, M., Kinali, M. and Mazarakis, N. D. (2018) ‘Autism spectrum disorder: prospects for treatment using gene therapy’, *Molecular Autism*, 9(1), p. 39. doi: 10.1186/s13229-018-0222-8.
Bernier, R. *et al.* (2014) ‘Disruptive CHD8 mutations define a subtype of autism early in development.’, *Cell*. United States, 158(2), pp. 263–276. doi: 10.1016/j.cell.2014.06.017.

Beutler, E. *et al.* (1973) ‘Red cell aldolase deficiency and hemolytic anemia: a new syndrome.’, *Transactions of the Association of American Physicians*. United States, 86, pp. 154–166.
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*. Netherlands, 52(2–3), pp. 77–87. doi: 10.1016/j.ejmg.2009.03.006.
Blaker-Lee, A. *et al.* (2012) ‘Zebrafish homologs of genes within 16p11.2, a genomic region associated with brain disorders, are active during brain development, and include two deletion dosage sensor genes’, *Disease models & mechanisms*. 2012/05/01. The Company of Biologists Limited, 5(6), pp. 834–851. doi: 10.1242/dmm.009944.
Blumenthal, I. *et al.* (2014) ‘Transcriptional consequences of 16p11.2 deletion and duplication in mouse cortex and multiplex autism families.’, *American journal of human genetics*. United States, 94(6), pp. 870–883. doi: 10.1016/j.ajhg.2014.05.004.
Boldrini, M. *et al.* (2018) ‘Human Hippocampal Neurogenesis Persists throughout Aging.’, *Cell stem cell*. United States, 22(4), pp. 589-599.e5. doi: 10.1016/j.stem.2018.03.015.
Brady, S. T., Lasek, R. J. and Allen, R. D. (1982) ‘Fast axonal transport in extruded axoplasm from squid giant axon.’, *Science (New York, N.Y.)*. United States, 218(4577), pp. 1129–1131. doi: 10.1126/science.6183745.
Britsch, S. (2007) ‘The neuregulin-I/ErbB signaling system in development and disease.’, *Advances in anatomy, embryology, and cell biology*. Germany, 190, pp. 1–65.
Brodmann, K. (1910) ‘Vergleichende Lokalisationslehre der Grosshirnrinde’, *The Journal of Nervous and Mental Disease*, (37), pp. 783–784.
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*. United States, 40(12), pp. 1466–1471. doi: 10.1038/ng.279.
Budday, S., Steinmann, P. and Kuhl, E. (2015) ‘Physical biology of human

brain development.', *Frontiers in cellular neuroscience*. Switzerland, 9, p. 257. doi: 10.3389/fncel.2015.00257.
Bulavin, D. V *et al.* (2003) 'Dual phosphorylation controls Cdc25 phosphatases and mitotic entry.', *Nature cell biology*. England, 5(6), pp. 545–551. doi: 10.1038/ncb994.
Buscaglia, C. A. *et al.* (2006) 'Characterization of an aldolase-binding site in the Wiskott-Aldrich syndrome protein.', *The Journal of biological chemistry*. United States, 281(3), pp. 1324–1331. doi: 10.1074/jbc.M506346200.
Bystron, I., Blakemore, C. and Rakic, P. (2008) 'Development of the human cerebral cortex: Boulder Committee revisited', *Nature Reviews Neuroscience*. Nature Publishing Group, 9, p. 110. Available at: https://doi.org/10.1038/nrn2252.
Calegari, F. and Huttner, W. B. (2003) 'An inhibition of cyclin-dependent kinases that lengthens, but does not arrest, neuroepithelial cell cycle induces premature neurogenesis.', *Journal of cell science*. England, 116(Pt 24), pp. 4947–4955. doi: 10.1242/jcs.00825.
Campbell, K. R. and Yau, C. (2018) 'Uncovering pseudotemporal trajectories with covariates from single cell and bulk expression data', *Nature Communications*, 9(1), p. 2442. doi: 10.1038/s41467-018-04696-6.
Casanova, E. L. and Casanova, M. F. (2014) 'Genetics studies indicate that neural induction and early neuronal maturation are disturbed in autism ', *Frontiers in Cellular Neuroscience* , p. 397. Available at: https://www.frontiersin.org/article/10.3389/fncel.2014.00397.
Casanova, M. F. (2014) 'Autism as a sequence: from heterochronic germinal cell divisions to abnormalities of cell migration and cortical dysplasias', *Medical hypotheses*. 2014/04/13, 83(1), pp. 32–38. doi: 10.1016/j.mehy.2014.04.014.
de Castro, F. and Bribian, A. (2005) 'The molecular orchestra of the migration of oligodendrocyte precursors during development.', *Brain research. Brain research reviews*. Netherlands, 49(2), pp. 227–241. doi: 10.1016/j.brainresrev.2004.12.034.
Chang, Y.-C. *et al.* (2018) 'Roles of Aldolase Family Genes in Human

Cancers and Diseases.', *Trends in endocrinology and metabolism: TEM*. United States, 29(8), pp. 549–559. doi: 10.1016/j.tem.2018.05.003.
Chen, J.-F. *et al.* (2014) 'Microcephaly disease gene Wdr62 regulates mitotic progression of embryonic neural stem cells and brain size', *Nature communications*, 5, p. 3885. doi: 10.1038/ncomms4885.
Chen, Y. *et al.* (2015) 'Pten Mutations Alter Brain Growth Trajectory and Allocation of Cell Types through Elevated β-Catenin Signaling', *The Journal of Neuroscience*, 35(28), pp. 10252 LP – 10267. doi: 10.1523/JNEUROSCI.5272-14.2015.
Chenn, A. *et al.* (1998) 'Intrinsic polarity of mammalian neuroepithelial cells.', *Molecular and cellular neurosciences*. United States, 11(4), pp. 183–193. doi: 10.1006/mcne.1998.0680.
Chenn, A. (2009) 'A top-NOTCH way to make astrocytes.', *Developmental cell*. United States, pp. 158–159. doi: 10.1016/j.devcel.2009.01.019.
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*. Public Library of Science, 8(3), p. e1002592. Available at: https://doi.org/10.1371/journal.pgen.1002592.
Clayton-Smith, J. and Laan, L. (2003) 'Angelman syndrome: a review of the clinical and genetic aspects', *Journal of Medical Genetics*, 40(2), pp. 87 LP – 95. doi: 10.1136/jmg.40.2.87.
Cook, E. H. J. and Scherer, S. W. (2008) 'Copy-number variations associated with neuropsychiatric conditions.', *Nature*. England, 455(7215), pp. 919–923. doi: 10.1038/nature07458.
Coppinger, J. *et al.* (2009) 'Identification of familial and de novo microduplications of 22q11.21-q11.23 distal to the 22q11.21 microdeletion syndrome region', *Human molecular genetics*. 2009/02/03. Oxford University Press, 18(8), pp. 1377–1383. doi: 10.1093/hmg/ddp042.
Courchesne, E. *et al.* (2011) 'Neuron Number and Size in Prefrontal Cortex of Children With Autism', *JAMA*, 306(18), pp. 2001–2010. doi: 10.1001/jama.2011.1638.
Courchesne, E., Carper, R. and Akshoomoff, N. (2003) 'Evidence of brain

overgrowth in the first year of life in autism.', *JAMA*. United States, 290(3), pp. 337–344. doi: 10.1001/jama.290.3.337.
DeFelipe, J. and Farinas, I. (1992) 'The pyramidal neuron of the cerebral cortex: morphological and chemical characteristics of the synaptic inputs.', *Progress in neurobiology*. England, 39(6), pp. 563–607. doi: 10.1016/0301-0082(92)90015-7.
Deshpande, A. *et al.* (2017) 'Cellular Phenotypes in Human iPSC-Derived Neurons from a Genetic Model of Autism Spectrum Disorder.', *Cell reports*. United States, 21(10), pp. 2678–2687. doi: 10.1016/j.celrep.2017.11.037.
DeStefano, F. (2007) 'Vaccines and autism: evidence does not support a causal association.', *Clinical pharmacology and therapeutics*. United States, 82(6), pp. 756–759. doi: 10.1038/sj.clpt.6100407.
Englund, C. *et al.* (2005) 'Pax6, Tbr2, and Tbr1 Are Expressed Sequentially by Radial Glia, Intermediate Progenitor Cells, and Postmitotic Neurons in Developing Neocortex', *The Journal of Neuroscience*, 25(1), pp. 247 LP – 251. doi: 10.1523/JNEUROSCI.2899-04.2005.
Ey, E., Leblond, C. S. and Bourgeron, T. (2011) 'Behavioral profiles of mouse models for autism spectrum disorders.', *Autism research : official journal of the International Society for Autism Research*. United States, 4(1), pp. 5–16. doi: 10.1002/aur.175.
Feindel, W. (1962) 'Thomas Willis (1621-1675)-The Founder of Neurology', *Canadian Medical Association journal*, 87, pp. 289–296.
Ferrier, D. (1886) *The functions of the brain*. 2nd edn. G.P. Putnam's Sons.
Filges, I. *et al.* (2014) 'Brain MRI abnormalities and spectrum of neurological and clinical findings in three patients with proximal 16p11.2 microduplication.', *American journal of medical genetics. Part A*. United States, 164A(8), pp. 2003–2012. doi: 10.1002/ajmg.a.36605.
Filimonenko, M. *et al.* (2010) 'The selective macroautophagic degradation of aggregated proteins requires the PI3P-binding protein Alfy.', *Molecular cell*. United States, 38(2), pp. 265–279. doi: 10.1016/j.molcel.2010.04.007.
Fisch, G. S. (2012) 'Nosology and epidemiology in autism: Classification counts', *American Journal of Medical Genetics Part C: Seminars in Medical*

Genetics. John Wiley & Sons, Ltd, 160C(2), pp. 91–103. doi: 10.1002/ajmg.c.31325.
Florio, M. and Huttner, W. B. (2014) 'Neural progenitors, neurogenesis and the evolution of the neocortex', *Development*, 141(11), pp. 2182 LP – 2194. doi: 10.1242/dev.090571.
Folstein, S. E. and Rosen-Sheidley, B. (2001) 'Genetics of autism: complex aetiology for a heterogeneous disorder.', *Nature reviews. Genetics*. England, 2(12), pp. 943–955. doi: 10.1038/35103559.
Fox, I. J. and Kornblum, H. I. (2005) 'Developmental profile of ErbB receptors in murine central nervous system: implications for functional interactions.', *Journal of neuroscience research*. United States, 79(5), pp. 584–597. doi: 10.1002/jnr.20381.
Fritsch, G. and Hitzig, E. (1870) 'Ueber die elektrische Erregbarkeit des Grosshirns', *Archiv für Anatomie, Physiologie und wissenschaftliche Medicin*, pp. 300–332.
Fu, H. *et al.* (2018) 'Aldolase A promotes proliferation and G1/S transition via the EGFR/MAPK pathway in non-small cell lung cancer.', *Cancer communications (London, England)*. England, 38(1), p. 18. doi: 10.1186/s40880-018-0290-3.
Funabiki, H. and Murray, A. W. (2000) 'The Xenopus chromokinesin Xkid is essential for metaphase chromosome alignment and must be degraded to allow anaphase chromosome movement.', *Cell*. United States, 102(4), pp. 411–424. doi: 10.1016/s0092-8674(00)00047-7.
Gage, F. H. (2002) 'Neurogenesis in the adult brain.', *The Journal of neuroscience : the official journal of the Society for Neuroscience*. United States, pp. 612–613.
Gal, J. S. *et al.* (2006) 'Molecular and Morphological Heterogeneity of Neural Precursors in the Mouse Neocortical Proliferative Zones', *The Journal of Neuroscience*, 26(3), pp. 1045 LP – 1056. doi: 10.1523/JNEUROSCI.4499-05.2006.
Gallagher, D. *et al.* (2015) 'Ankrd11 Is a Chromatin Regulator Involved in Autism that Is Essential for Neural Development', *Developmental Cell*, 32(1),

pp. 31–42. doi: https://doi.org/10.1016/j.devcel.2014.11.031.
Ghebranious, N. *et al.* (2007) 'A novel microdeletion at 16p11.2 harbors candidate genes for aortic valve development, seizure disorder, and mild mental retardation.', *American journal of medical genetics. Part A*. United States, 143A(13), pp. 1462–1471. doi: 10.1002/ajmg.a.31837.
Giaroli, G. *et al.* (2014) 'Does rare matter? Copy number variants at 16p11.2 and the risk of psychosis: a systematic review of literature and meta-analysis.', *Schizophrenia research*. Netherlands, 159(2–3), pp. 340–346. doi: 10.1016/j.schres.2014.09.025.
Gilbert-Jaramillo, J. *et al.* (2019) 'The potential contribution of impaired brain glucose metabolism to congenital Zika syndrome', *Journal of Anatomy*. John Wiley & Sons, Ltd (10.1111), 0(0). doi: 10.1111/joa.12959.
Gilbert, J. and Man, H.-Y. (2017) 'Fundamental Elements in Autism: From Neurogenesis and Neurite Growth to Synaptic Plasticity.', *Frontiers in cellular neuroscience*. Switzerland, 11, p. 359. doi: 10.3389/fncel.2017.00359.
Girirajan, S. and Eichler, E. E. (2010) 'Phenotypic variability and genetic susceptibility to genomic disorders.', *Human molecular genetics*. England, 19(R2), pp. R176-87. doi: 10.1093/hmg/ddq366.
Giulivi, C. *et al.* (2010) 'Mitochondrial dysfunction in autism', *JAMA*, 304(21), pp. 2389–2396. doi: 10.1001/jama.2010.1706.
Golzio, C. *et al.* (2012) 'KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant', *Nature*, 485(7398), pp. 363–367. doi: 10.1038/nature11091.
Götz, M. and Huttner, W. B. (2005) 'The cell biology of neurogenesis', *Nature Reviews Molecular Cell Biology*, 6(10), pp. 777–788. doi: 10.1038/nrm1739.
Goulas, A. *et al.* (2019) 'The architecture of mammalian cortical connectomes in light of the theory of the dual origin of the cerebral cortex.', *Cortex; a journal devoted to the study of the nervous system and behavior*. Italy. doi: 10.1016/j.cortex.2019.03.002.
Gross, C. G. (2009) 'Neuroscience, Early of History', *Journal of Enclyclopedia of Neuroscience*, pp. 843–847. doi: 10.1016/B978-008045046-

9.00992-X.
Gustafsson, J. (2009) ‘Neonatal energy substrate production.’, *The Indian journal of medical research*. India, 130(5), pp. 618–623.
Hansen, D. V *et al.* (2010) ‘Neurogenic radial glia in the outer subventricular zone of human neocortex.’, *Nature*. England, 464(7288), pp. 554–561. doi: 10.1038/nature08845.
Hansen, D. V *et al.* (2013) ‘Non-epithelial stem cells and cortical interneuron production in the human ganglionic eminences’, *Nature neuroscience*. 2013/10/06, 16(11), pp. 1576–1587. doi: 10.1038/nn.3541.
Hanson, E. *et al.* (2015) ‘The cognitive and behavioral phenotype of the 16p11.2 deletion in a clinically ascertained population.’, *Biological psychiatry*. United States, 77(9), pp. 785–793. doi: 10.1016/j.biopsych.2014.04.021.
Haslinger, D. *et al.* (2018) ‘Loss of the Chr16p11.2 ASD candidate gene QPRT leads to aberrant neuronal differentiation in the SH-SY5Y neuronal cell model.’, *Molecular autism*. England, 9, p. 56. doi: 10.1186/s13229-018-0239-z.
Herculano-Houzel, S., Manger, P. R. and Kaas, J. H. (2014) ‘Brain scaling in mammalian evolution as a consequence of concerted and mosaic changes in numbers of neurons and average neuronal cell size ’, *Frontiers in Neuroanatomy* , p. 77. Available at: https://www.frontiersin.org/article/10.3389/fnana.2014.00077.
Hermanson, O., Jepsen, K. and Rosenfeld, M. G. (2002) ‘N-CoR controls differentiation of neural stem cells into astrocytes.’, *Nature*. England, 419(6910), pp. 934–939. doi: 10.1038/nature01156.
Hirabayashi, Y. *et al.* (2009) ‘Polycomb limits the neurogenic competence of neural precursor cells to promote astrogenic fate transition.’, *Neuron*. United States, 63(5), pp. 600–613. doi: 10.1016/j.neuron.2009.08.021.
Hirokawa, N. *et al.* (1989) ‘Submolecular domains of bovine brain kinesin identified by electron microscopy and monoclonal antibody decoration.’, *Cell*. United States, 56(5), pp. 867–878. doi: 10.1016/0092-8674(89)90691-0.
Hirokawa, N. *et al.* (2009) ‘Kinesin superfamily motor proteins and intracellular transport.’, *Nature reviews. Molecular cell biology*. England,

10(10), pp. 682–696. doi: 10.1038/nrm2774.
Hoerder-Suabedissen, A. and Molnár, Z. (2015) ‘Development, evolution and pathology of neocortical subplate neurons’, *Nature Reviews Neuroscience*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 16, p. 133. Available at: https://doi.org/10.1038/nrn3915.
Hogart, A. *et al.* (2010) ‘The comorbidity of autism with the genomic disorders of chromosome 15q11.2-q13’, *Neurobiology of disease*. 2008/09/18, 38(2), pp. 181–191. doi: 10.1016/j.nbd.2008.08.011.
Horecker, B. L. *et al.* (1981) ‘Aldolase and Fructose Bisphosphatase: Key Enzymes in the Control of Gluconeogenesis and Glycolysis’, in Estabrook, R. W. and Srere, P. B. T.-C. T. in C. R. (eds) *Biological Cycles*. Academic Press, pp. 181–197. doi: https://doi.org/10.1016/B978-0-12-152818-8.50016-5.
Horev, G. *et al.* (2011) ‘Dosage-dependent phenotypes in models of 16p11.2 lesions found in autism’, *Proceedings of the National Academy of Sciences of the United States of America*. 2011/10/03. National Academy of Sciences, 108(41), pp. 17076–17081. doi: 10.1073/pnas.1114042108.
Hua, W., Chung, J. and Gelles, J. (2002) ‘Distinguishing inchworm and hand-over-hand processive kinesin movement by neck rotation measurements.’, *Science (New York, N.Y.)*. United States, 295(5556), pp. 844–848. doi: 10.1126/science.1063089.
Iyer, J. *et al.* (2018) ‘Pervasive genetic interactions modulate neurodevelopmental defects of the autism-associated 16p11.2 deletion in Drosophila melanogaster.’, *Nature communications*. England, 9(1), p. 2548. doi: 10.1038/s41467-018-04882-6.
Jessen, K. R. and Mirsky, R. (1980) ‘Glial cells in the enteric nervous system contain glial fibrillary acidic protein’, *Nature*, 286(5774), pp. 736–737. doi: 10.1038/286736a0.
Jewett, T. J. and Sibley, L. D. (2003) ‘Aldolase forms a bridge between cell surface adhesins and the actin cytoskeleton in apicomplexan parasites.’, *Molecular cell*. United States, 11(4), pp. 885–894.
Kageyama, R., Ohtsuka, T. and Kobayashi, T. (2008) ‘Roles of Hes genes in

neural development.', *Development, growth & differentiation*. Japan, 50 Suppl 1, pp. S97-103. doi: 10.1111/j.1440-169X.2008.00993.x.
Kanner, L. (1943) 'Autistic disturbances of affective contact.', *Nervous Child*, 2, pp. 217–250.
Kao, A. W. *et al.* (1999) 'Aldolase mediates the association of F-actin with the insulin-responsive glucose transporter GLUT4.', *The Journal of biological chemistry*. United States, 274(25), pp. 17742–17747. doi: 10.1074/jbc.274.25.17742.
Katayama, Y. *et al.* (2016) 'CHD8 haploinsufficiency results in autistic-like phenotypes in mice', *Nature*. Macmillan Publishers Limited, part of Springer Nature. All rights reserved., 537, p. 675. Available at: https://doi.org/10.1038/nature19357.
Kempermann, G. *et al.* (2018) 'Human Adult Neurogenesis: Evidence and Remaining Questions.', *Cell stem cell*. United States, 23(1), pp. 25–30. doi: 10.1016/j.stem.2018.04.004.
Kessaris, N., Pringle, N. and Richardson, W. D. (2001) 'Ventral neurogenesis and the neuron-glial switch.', *Neuron*. United States, 31(5), pp. 677–680.
Kim, K.-Y., Hysolli, E. and Park, I.-H. (2011) 'Neuronal maturation defect in induced pluripotent stem cells from patients with Rett syndrome.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 108(34), pp. 14169–14174. doi: 10.1073/pnas.1018979108.
Kirov, G., Rees, E. and Walters, J. (2015) 'What a psychiatrist needs to know about copy number variants', *BJPsych Advances*. 2018/01/02. Cambridge University Press, 21(3), pp. 157–163. doi: DOI: 10.1192/apt.bp.113.012039.
Klein-Tasman, B. P. and Mervis, C. B. (2018) 'Autism Spectrum Symptomatology Among Children with Duplication 7q11.23 Syndrome.', *Journal of autism and developmental disorders*. United States, 48(6), pp. 1982–1994. doi: 10.1007/s10803-017-3439-z.
Kreuder, J. *et al.* (1996) 'Inherited Metabolic Myopathy and Hemolysis Due to a Mutation in Aldolase A', *New England Journal of Medicine*. Massachusetts Medical Society, 334(17), pp. 1100–1105. doi:

10.1056/NEJM199604253341705.
Kriegstein, A. and Alvarez-Buylla, A. (2009) 'The Glial Nature of Embryonic and Adult Neural Stem Cells', *Annual Review of Neuroscience*, 32(1), pp. 149–184. doi: 10.1146/annurev.neuro.051508.135600.
Kumar, R. A. *et al.* (2008) 'Recurrent 16p11.2 microdeletions in autism.', *Human molecular genetics*. England, 17(4), pp. 628–638. doi: 10.1093/hmg/ddm376.
Laje, G. *et al.* (2010) 'Autism spectrum features in Smith-Magenis syndrome', *American journal of medical genetics. Part C, Seminars in medical genetics*, 154C(4), pp. 456–462. doi: 10.1002/ajmg.c.30275.
Lalli, M. A. *et al.* (2016) 'Haploinsufficiency of BAZ1B contributes to Williams syndrome through transcriptional dysregulation of neurodevelopmental pathways', *Human Molecular Genetics*, 25(7), pp. 1294–1306. doi: 10.1093/hmg/ddw010.
LaMonica, B. E. *et al.* (2012) 'OSVZ progenitors in the human cortex: an updated perspective on neurodevelopmental disease.', *Current opinion in neurobiology*. England, 22(5), pp. 747–753. doi: 10.1016/j.conb.2012.03.006.
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*. United States, 5(3), pp. 320–331. doi: 10.1016/j.stem.2009.05.026.
Levitt, P. and Rakic, P. (1980) 'Immunoperoxidase localization of glial fibrillary acidic protein in radial glial cells and astrocytes of the developing rhesus monkey brain.', *The Journal of comparative neurology*. United States, 193(3), pp. 815–840. doi: 10.1002/cne.901930316.
Levy, D. *et al.* (2011) 'Rare de novo and transmitted copy-number variation in autistic spectrum disorders.', *Neuron*. United States, 70(5), pp. 886–897. doi: 10.1016/j.neuron.2011.05.015.
Liu, B. *et al.* (2016) 'MAZ mediates the cross-talk between CT-1 and NOTCH1 signaling during gliogenesis', *Scientific reports*. Nature Publishing Group, 6, p. 21534. doi: 10.1038/srep21534.
Liu, Y.-T. *et al.* (2016) 'PRRT2 mutations lead to neuronal dysfunction and

neurodevelopmental defects', *Oncotarget*. Impact Journals LLC, 7(26), pp. 39184–39196. doi: 10.18632/oncotarget.9258.
Llinares-Benadero, C. and Borrell, V. (2019) 'Deconstructing cortical folding: genetic, cellular and mechanical determinants', *Nature Reviews Neuroscience*, 20(3), pp. 161–176. doi: 10.1038/s41583-018-0112-2.
Lord, C. and Jones, R. M. (2012) 'Annual research review: re-thinking the classification of autism spectrum disorders', *Journal of child psychology and psychiatry, and allied disciplines*, 53(5), pp. 490–509. doi: 10.1111/j.1469-7610.2012.02547.x.
Lui, J. H., Hansen, D. V and Kriegstein, A. R. (2011) 'Development and evolution of the human neocortex', *Cell*, 146(1), pp. 18–36. doi: 10.1016/j.cell.2011.06.030.
Luskin, M. B. and Shatz, C. J. (1985) 'Studies of the earliest generated cells of the cat's visual cortex: cogeneration of subplate and marginal zones.', *The Journal of neuroscience : the official journal of the Society for Neuroscience*. United States, 5(4), pp. 1062–1075.
Ma, T. *et al.* (2013) 'Subcortical origins of human and monkey neocortical interneurons.', *Nature neuroscience*. United States, 16(11), pp. 1588–1597. doi: 10.1038/nn.3536.
Mabie, P. C., Mehler, M. F. and Kessler, J. A. (1999) 'Multiple roles of bone morphogenetic protein signaling in the regulation of cortical cell number and phenotype.', *The Journal of neuroscience : the official journal of the Society for Neuroscience*. United States, 19(16), pp. 7077–7088.
Macmillan, M. (2001) 'John Martyn Harlow: "Obscure country physician"?', *Journal of the history of the neurosciences*. England, 10(2), pp. 149–162. doi: 10.1076/jhin.10.2.149.7254.
Maillard, A. M. *et al.* (2015) 'The 16p11.2 locus modulates brain structures common to autism, schizophrenia and obesity.', *Molecular psychiatry*. England, 20(1), pp. 140–147. doi: 10.1038/mp.2014.145.
Mamczur, P. *et al.* (2013) 'Nuclear localization of aldolase A correlates with cell proliferation', *Biochimica et biophysica acta*, 1833. doi: 10.1016/j.bbamcr.2013.07.013.

Mamczur, P. and Dzugaj, A. (2004) ‘Nuclear localization of aldolase A in pig cardiomyocytes.’, *Histology and histopathology*. Spain, 19(3), pp. 753–758. doi: 10.14670/HH-19.753.
Mamczur, P. and Dzugaj, A. (2008) ‘Aldolase A is present in smooth muscle cell nuclei.’, *Acta biochimica Polonica*. Poland, 55(4), pp. 799–805.
Manuel, M. N. *et al.* (2015) ‘Regulation of cerebral cortical neurogenesis by the Pax6 transcription factor’, *Frontiers in cellular neuroscience*. Frontiers Media S.A., 9, p. 70. doi: 10.3389/fncel.2015.00070.
Mao, H. *et al.* (2015) ‘Rbm8a haploinsufficiency disrupts embryonic cortical development resulting in microcephaly.’, *The Journal of neuroscience : the official journal of the Society for Neuroscience*. United States, 35(18), pp. 7003–7018. doi: 10.1523/JNEUROSCI.0018-15.2015.
Marchetto, M. C. *et al.* (2017) ‘Altered proliferation and networks in neural cells derived from idiopathic autistic individuals’, *Molecular psychiatry*. 2016/07/05, 22(6), pp. 820–835. doi: 10.1038/mp.2016.95.
Marin-Padilla, M. (1983) ‘Structural organization of the human cerebral cortex prior to the appearance of the cortical plate.’, *Anatomy and embryology*. Germany, 168(1), pp. 21–40. doi: 10.1007/bf00305396.
Marín-Padilla, M. (2011) *The human brain prenatal development and structure*. Berlin : Berlin .
Marín, O. and Rubenstein, J. L. R. (2001) ‘A long, remarkable journey: Tangential migration in the telencephalon’, *Nature Reviews Neuroscience*, 2(11), pp. 780–790. doi: 10.1038/35097509.
Marshall, C. A. G., Novitch, B. G. and Goldman, J. E. (2005) ‘Olig2 directs astrocyte and oligodendrocyte formation in postnatal subventricular zone cells.’, *The Journal of neuroscience : the official journal of the Society for Neuroscience*. United States, 25(32), pp. 7289–7298. doi: 10.1523/JNEUROSCI.1924-05.2005.
Martinez, S. and Puelles, L. (2000) ‘Neurogenetic compartments of the mouse diencephalon and some characteristic gene expression patterns.’, *Results and problems in cell differentiation*. Germany, 30, pp. 91–106.
McCarthy, S. E. *et al.* (2009) ‘Microduplications of 16p11.2 are associated

with schizophrenia.', *Nature genetics*. United States, 41(11), pp. 1223–1227. doi: 10.1038/ng.474.
Meechan, D. W. *et al.* (2011) 'Three phases of DiGeorge/22q11 deletion syndrome pathogenesis during brain development: patterning, proliferation, and mitochondrial functions of 22q11 genes.', *International journal of developmental neuroscience : the official journal of the International Society for Developmental Neuroscience*. England, 29(3), pp. 283–294. doi: 10.1016/j.ijdevneu.2010.08.005.
Meechan, D. W. *et al.* (2015) 'Modeling a model: Mouse genetics, 22q11.2 Deletion Syndrome, and disorders of cortical circuit development', *Progress in neurobiology*. 2015/04/09, 130, pp. 1–28. doi: 10.1016/j.pneurobio.2015.03.004.
Mefford, H. C. *et al.* (2008) 'Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes.', *The New England journal of medicine*. United States, 359(16), pp. 1685–1699. doi: 10.1056/NEJMoa0805384.
Mehler, M. F. (2002) 'Mechanisms regulating lineage diversity during mammalian cerebral cortical neurogenesis and gliogenesis.', *Results and problems in cell differentiation*. Germany, 39, pp. 27–52.
Merkulova, M. *et al.* (2011) 'Aldolase directly interacts with ARNO and modulates cell morphology and acidic vesicle distribution', *American journal of physiology. Cell physiology*. 2011/02/09. American Physiological Society, 300(6), pp. C1442–C1455. doi: 10.1152/ajpcell.00076.2010.
Mervis, C. B. *et al.* (2015) 'Children with 7q11.23 duplication syndrome: psychological characteristics', *American journal of medical genetics. Part A*. 2015/04/21, 167(7), pp. 1436–1450. doi: 10.1002/ajmg.a.37071.
Miller, I. *et al.* (2018) 'Ki67 is a Graded Rather than a Binary Marker of Proliferation versus Quiescence.', *Cell reports*. United States, 24(5), pp. 1105-1112.e5. doi: 10.1016/j.celrep.2018.06.110.
Miyata, T. *et al.* (2004) 'Asymmetric production of surface-dividing and non-surface-dividing cortical progenitor cells.', *Development (Cambridge, England)*. England, 131(13), pp. 3133–3145. doi: 10.1242/dev.01173.

Molina, J. *et al.* (2008) ‘Abnormal social behaviors and altered gene expression rates in a mouse model for Potocki-Lupski syndrome.’, *Human molecular genetics*. England, 17(16), pp. 2486–2495. doi: 10.1093/hmg/ddn148.
Molnár, Z. (2004) ‘Thomas Willis (1621–1675), the founder of clinical neuroscience’, *Nature Reviews Neuroscience*, 5(4), pp. 329–335. doi: 10.1038/nrn1369.
Molne, M. *et al.* (2000) ‘Early cortical precursors do not undergo LIF-mediated astrocytic differentiation.’, *Journal of neuroscience research*. United States, 59(3), pp. 301–311.
Molofsky, A. V. and Deneen, B. (2015) ‘Astrocyte development: A Guide for the Perplexed.’, *Glia*. United States, 63(8), pp. 1320–1329. doi: 10.1002/glia.22836.
Moreno-De-Luca, D. *et al.* (2010) ‘Deletion 17q12 Is a Recurrent Copy Number Variant that Confers High Risk of Autism and Schizophrenia’, *The American Journal of Human Genetics*, 87(5), pp. 618–630. doi: https://doi.org/10.1016/j.ajhg.2010.10.004.
Morson, S. *et al.* (2019) ‘Expression of genes in the 16p11.2 locus during human fetal cortical neurogenesis’, *bioRxiv*, p. 633461. doi: 10.1101/633461.
N Filimonoff, I. (1947) ‘A rational subdivision of the cerebral cortex’, *Archives of neurology and psychiatry*, 58, pp. 296–311. doi: 10.1001/archneurpsyc.1947.02300320047002.
Nadarajah, B. *et al.* (2001) ‘Two modes of radial migration in early development of the cerebral cortex.’, *Nature neuroscience*. United States, 4(2), pp. 143–150. doi: 10.1038/83967.
Namihira, M. *et al.* (2009) ‘Committed neuronal precursors confer astrocytic potential on residual neural precursor cells.’, *Developmental cell*. United States, 16(2), pp. 245–255. doi: 10.1016/j.devcel.2008.12.014.
Natacci, F. *et al.* (2016) ‘Chromosome 17q21.31 duplication syndrome: Description of a new familiar case and further delineation of the clinical spectrum’, *European Journal of Paediatric Neurology*, 20(1), pp. 183–187. doi: https://doi.org/10.1016/j.ejpn.2015.09.010.

Nazeer, A. and Ghaziuddin, M. (2012) ‘Autism spectrum disorders: clinical features and diagnosis.’, *Pediatric clinics of North America*. United States, 59(1), pp. 19–25, ix. doi: 10.1016/j.pcl.2011.10.007.
Newschaffer, C. J. *et al.* (2007) ‘The epidemiology of autism spectrum disorders.’, *Annual review of public health*. United States, 28, pp. 235–258. doi: 10.1146/annurev.publhealth.28.021406.144007.
Niarchou, M. *et al.* (2019) ‘Psychiatric disorders in children with 16p11.2 deletion and duplication’, *Translational Psychiatry*, 9(1), p. 8. doi: 10.1038/s41398-018-0339-8.
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.
Nuttle, X. *et al.* (2016) ‘Emergence of a Homo sapiens-specific gene family and chromosome 16p11.2 CNV susceptibility.’, *Nature*. England, 536(7615), pp. 205–209. doi: 10.1038/nature19075.
Ohsugi, M. *et al.* (2003) ‘Cdc2-mediated phosphorylation of Kid controls its distribution to spindle and chromosomes.’, *The EMBO journal*. England, 22(9), pp. 2091–2103. doi: 10.1093/emboj/cdg208.
Ohsugi, M. *et al.* (2008) ‘Kid-mediated chromosome compaction ensures proper nuclear envelope formation.’, *Cell*. United States, 132(5), pp. 771–782. doi: 10.1016/j.cell.2008.01.029.
Orosco, L. A. *et al.* (2014) ‘Loss of Wdfy3 in mice alters cerebral cortical neurogenesis reflecting aspects of the autism pathology.’, *Nature communications*. England, 5, p. 4692. doi: 10.1038/ncomms5692.
Ostrem, B. E. L. *et al.* (2014) ‘Control of outer radial glial stem cell mitosis in the human brain’, *Cell reports*. 2014/07/31, 8(3), pp. 656–664. doi: 10.1016/j.celrep.2014.06.058.
Ousley, O. *et al.* (2017) ‘Examining the Overlap between Autism Spectrum Disorder and 22q11.2 Deletion Syndrome’, *International journal of molecular sciences*. MDPI, 18(5), p. 1071. doi: 10.3390/ijms18051071.
Packer, A. (2016) ‘Neocortical neurogenesis and the etiology of autism spectrum disorder.’, *Neuroscience and biobehavioral reviews*. United States,

64, pp. 185–195. doi: 10.1016/j.neubiorev.2016.03.002.
Pagliaro, L. and Taylor, D. L. (1992) '2-Deoxyglucose and cytochalasin D modulate aldolase mobility in living 3T3 cells.', *The Journal of cell biology*. United States, 118(4), pp. 859–863. doi: 10.1083/jcb.118.4.859.
Park, H. R. *et al.* (2016) 'A Short Review on the Current Understanding of Autism Spectrum Disorders', *Experimental neurobiology*. 2016/01/28. The Korean Society for Brain and Neural Science, 25(1), pp. 1–13. doi: 10.5607/en.2016.25.1.1.
Picco, N. *et al.* (2018) 'Mathematical Modeling of Cortical Neurogenesis Reveals that the Founder Population does not Necessarily Scale with Neurogenic Output', *Cerebral cortex (New York, N.Y. : 1991)*. Oxford University Press, 28(7), pp. 2540–2550. doi: 10.1093/cercor/bhy068.
Pollak, R. M. *et al.* (2018) 'Neuropsychiatric Phenotypes in 3q29 Deletion Syndrome', *bioRxiv*, p. 386243. doi: 10.1101/386243.
Pollen, A. A. *et al.* (2015) 'Molecular identity of human outer radial glia during cortical development.', *Cell*. United States, 163(1), pp. 55–67. doi: 10.1016/j.cell.2015.09.004.
Pollen, A. A. *et al.* (2019) 'Establishing Cerebral Organoids as Models of Human-Specific Brain Evolution.', *Cell*. United States, 176(4), pp. 743-756.e17. doi: 10.1016/j.cell.2019.01.017.
Porteus, M. H. *et al.* (1994) 'DLX-2, MASH-1, and MAP-2 expression and bromodeoxyuridine incorporation define molecularly distinct cell populations in the embryonic mouse forebrain', *The Journal of neuroscience : the official journal of the Society for Neuroscience*. Society for Neuroscience, 14(11 Pt 1), pp. 6370–6383. doi: 10.1523/JNEUROSCI.14-11-06370.1994.
Preibisch, S., Saalfeld, S. and Tomancak, P. (2009) 'Globally optimal stitching of tiled 3D microscopic image acquisitions', *Bioinformatics (Oxford, England)*. 2009/04/03. Oxford University Press, 25(11), pp. 1463–1465. doi: 10.1093/bioinformatics/btp184.
Price, D. *et al.* (2011) 'Building Brains: An Introduction to Neural Development', in, pp. 29–32.
Pucilowska, J. *et al.* (2012) 'Disrupted ERK signaling during cortical

development leads to abnormal progenitor proliferation, neuronal and network excitability and behavior, modeling human neuro-cardio-facial-cutaneous and related syndromes.', *The Journal of neuroscience : the official journal of the Society for Neuroscience*. United States, 32(25), pp. 8663–8677. doi: 10.1523/JNEUROSCI.1107-12.2012.

Pucilowska, J. *et al.* (2015) 'The 16p11.2 deletion mouse model of autism exhibits altered cortical progenitor proliferation and brain cytoarchitecture linked to the ERK MAPK pathway.', *The Journal of neuroscience : the official journal of the Society for Neuroscience*. United States, 35(7), pp. 3190–3200. doi: 10.1523/JNEUROSCI.4864-13.2015.

Qiu, X. *et al.* (2017) 'Single-cell mRNA quantification and differential analysis with Census.', *Nature methods*. United States, 14(3), pp. 309–315. doi: 10.1038/nmeth.4150.

Qureshi, A. Y. *et al.* (2014) 'Opposing brain differences in 16p11.2 deletion and duplication carriers.', *The Journal of neuroscience : the official journal of the Society for Neuroscience*. United States, 34(34), pp. 11199–11211. doi: 10.1523/JNEUROSCI.1366-14.2014.

Radonjić, N. V *et al.* (2014) 'Diversity of cortical interneurons in primates: the role of the dorsal proliferative niche', *Cell reports*. 2014/12/11, 9(6), pp. 2139–2151. doi: 10.1016/j.celrep.2014.11.026.

Rakic, P. (1972) 'Mode of cell migration to the superficial layers of fetal monkey neocortex.', *The Journal of comparative neurology*. United States, 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 (New York, N.Y.)*. United States, 183(4123), pp. 425–427. doi: 10.1126/science.183.4123.425.

Rakic, P. (1988) 'Specification of cerebral cortical areas.', *Science (New York, N.Y.)*. United States, 241(4862), pp. 170–176.

Rakic, P. (1990) 'Principles of neural cell migration', *Experientia*, 46(9), pp. 882–891. doi: 10.1007/BF01939380.

Rakic, P. (1995) 'A small step for the cell, a giant leap for mankind: a

hypothesis of neocortical expansion during evolution', *Trends in Neurosciences*, 18(9), pp. 383–388. doi: https://doi.org/10.1016/0166-2236(95)93934-P.
Rash, B. G. *et al.* (2018) 'Metabolic regulation and glucose sensitivity of cortical radial glial cells', *Proceedings of the National Academy of Sciences*, 115(40), pp. 10142 LP – 10147. doi: 10.1073/pnas.1808066115.
Reiner, O. *et al.* (2016) 'Regulation of neuronal migration, an emerging topic in autism spectrum disorders', *Journal of Neurochemistry*. John Wiley & Sons, Ltd (10.1111), 136(3), pp. 440–456. doi: 10.1111/jnc.13403.
Richter, M. *et al.* (2018) 'Altered TAOK2 activity causes autism-related neurodevelopmental and cognitive abnormalities through RhoA signaling', *Molecular Psychiatry*. doi: 10.1038/s41380-018-0025-5.
Ritterson Lew, C. and Tolan, D. R. (2012) 'Targeting of several glycolytic enzymes using RNA interference reveals aldolase affects cancer cell proliferation through a non-glycolytic mechanism.', *The Journal of biological chemistry*. United States, 287(51), pp. 42554–42563. doi: 10.1074/jbc.M112.405969.
Roberts, P. A. (1992) *Neuroanatomy*. New York: Springer-Verlag.
Rosenfeld, J. A. *et al.* (2010) 'Speech delays and behavioral problems are the predominant features in individuals with developmental delays and 16p11.2 microdeletions and microduplications.', *Journal of neurodevelopmental disorders*. England, 2(1), pp. 26–38. doi: 10.1007/s11689-009-9037-4.
Rosenfeld, J. A. *et al.* (2013) 'Estimates of penetrance for recurrent pathogenic copy-number variations', *Genetics in medicine : official journal of the American College of Medical Genetics*. 2012/12/20. Nature Publishing Group, 15(6), pp. 478–481. doi: 10.1038/gim.2012.164.
Rowitch, D. H. and Kriegstein, A. R. (2010) 'Developmental genetics of vertebrate glial-cell specification.', *Nature*. England, 468(7321), pp. 214–222. doi: 10.1038/nature09611.
Rutter, M. (1968) 'Concepts of autism: A review of research.', *Child Psychology & Psychiatry & Allied Disciplines*. United Kingdom: Pergamon

Press, pp. 1–25. doi: 10.1111/j.1469-7610.1968.tb02204.x.
Sacco, R., Gabriele, S. and Persico, A. M. (2015) 'Head circumference and brain size in autism spectrum disorder: A systematic review and meta-analysis.', *Psychiatry research*. Ireland, 234(2), pp. 239–251. doi: 10.1016/j.pscychresns.2015.08.016.
Saladin, K. S. (2011) *Human anatomy*. New York: McGraw-Hill.
Saldarriaga, W. *et al.* (2014) 'Fragile X syndrome', *Colombia medica (Cali, Colombia)*. Universidad del Valle, 45(4), pp. 190–198. Available at: https://www.ncbi.nlm.nih.gov/pubmed/25767309.
Sanosaka, T. *et al.* (2008) 'Identification of genes that restrict astrocyte differentiation of midgestational neural precursor cells.', *Neuroscience*. United States, 155(3), pp. 780–788. doi: 10.1016/j.neuroscience.2008.06.039.
Sardi, S. P. *et al.* (2006) 'Presenilin-dependent ErbB4 nuclear signaling regulates the timing of astrogenesis in the developing brain.', *Cell*. United States, 127(1), pp. 185–197. doi: 10.1016/j.cell.2006.07.037.
Sauer, M. E. and Walker, B. E. (1959) 'Radioautographic Study of Interkinetic Nuclear Migration in the Neural Tube.', *Proceedings of the Society for Experimental Biology and Medicine*. SAGE Publications, 101(3), pp. 557–560. doi: 10.3181/00379727-101-25014.
Scholzen, T. and Gerdes, J. (2000) 'The Ki-67 protein: from the known and the unknown.', *Journal of cellular physiology*. United States, 182(3), pp. 311–322. doi: 10.1002/(SICI)1097-4652(200003)182:3<311::AID-JCP1>3.0.CO;2-9.
Sebat, J. *et al.* (2007) 'Strong association of de novo copy number mutations with autism', *Science (New York, N.Y.)*. 2007/03/15, 316(5823), pp. 445–449. doi: 10.1126/science.1138659.
Shinawi, M. *et al.* (2010) 'Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size', *Journal of medical genetics*. 2009/11/12, 47(5), pp. 332–341. doi: 10.1136/jmg.2009.073015.
Shiroguchi, K. *et al.* (2003) 'The second microtubule-binding site of

monomeric kid enhances the microtubule affinity.', *The Journal of biological chemistry*. United States, 278(25), pp. 22460–22465. doi: 10.1074/jbc.M212274200.
Smart, I. H. (1973) 'Proliferative characteristics of the ependymal layer during the early development of the mouse neocortex: a pilot study based on recording the number, location and plane of cleavage of mitotic figures', *Journal of anatomy*, 116(Pt 1), pp. 67–91. Available at: https://www.ncbi.nlm.nih.gov/pubmed/4777782.
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 (New York, N.Y. : 1991)*. Oxford University Press, 12(1), pp. 37–53. doi: 10.1093/cercor/12.1.37.
Soeda, S., Yamada-Nomoto, K. and Ohsugi, M. (2016) 'The microtubule-binding and coiled-coil domains of Kid are required to turn off the polar ejection force at anaphase.', *Journal of cell science*. England, 129(19), pp. 3609–3619. doi: 10.1242/jcs.189969.
Sofroniew, M. V and Vinters, H. V (2010) 'Astrocytes: biology and pathology', *Acta neuropathologica*. 2009/12/10. Springer-Verlag, 119(1), pp. 7–35. doi: 10.1007/s00401-009-0619-8.
Sorrells, S. F. *et al.* (2018) 'Human hippocampal neurogenesis drops sharply in children to undetectable levels in adults.', *Nature*. England, 555(7696), pp. 377–381. doi: 10.1038/nature25975.
Stoner, R. *et al.* (2014) 'Patches of Disorganization in the Neocortex of Children with Autism', *New England Journal of Medicine*. Massachusetts Medical Society, 370(13), pp. 1209–1219. doi: 10.1056/NEJMoa1307491.
Stoykova, A. and Gruss, P. (1994) 'Roles of Pax-genes in developing and adult brain as suggested by expression patterns.', *The Journal of neuroscience : the official journal of the Society for Neuroscience*. United States, 14(3 Pt 2), pp. 1395–1412.
Sun, T. and Hevner, R. F. (2014) 'Growth and folding of the mammalian cerebral cortex: from molecules to malformations', *Nature reviews.*

Neuroscience, 15(4), pp. 217–232. doi: 10.1038/nrn3707.
Sur, M. and Rubenstein, J. L. R. (2005) 'Patterning and Plasticity of the Cerebral Cortex', *Science*. American Association for the Advancement of Science, 310(5749), pp. 805–810. doi: 10.1126/science.1112070.
Sztainberg, Y. and Zoghbi, H. Y. (2016) 'Lessons learned from studying syndromic autism spectrum disorders', *Nature Neuroscience*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 19, p. 1408. Available at: https://doi.org/10.1038/nn.4420.
Takahashi, T., Nowakowski, R. S. and Caviness, V. S. (1995) 'The cell cycle of the pseudostratified ventricular epithelium of the embryonic murine cerebral wall', *The Journal of Neuroscience*, 15(9), pp. 6046 LP – 6057. doi: 10.1523/JNEUROSCI.15-09-06046.1995.
Takiguchi-Hayashi, K. *et al.* (2004) 'Generation of reelin-positive marginal zone cells from the caudomedial wall of telencephalic vesicles.', *The Journal of neuroscience : the official journal of the Society for Neuroscience*. United States, 24(9), pp. 2286–2295. doi: 10.1523/JNEUROSCI.4671-03.2004.
Takizawa, T. *et al.* (2001) 'DNA methylation is a critical cell-intrinsic determinant of astrocyte differentiation in the fetal brain.', *Developmental cell*. United States, 1(6), pp. 749–758.
Taylor, L. E., Swerdfeger, A. L. and Eslick, G. D. (2014) 'Vaccines are not associated with autism: an evidence-based meta-analysis of case-control and cohort studies.', *Vaccine*. Netherlands, 32(29), pp. 3623–3629. doi: 10.1016/j.vaccine.2014.04.085.
Temple, S. (2001) 'The development of neural stem cells.', *Nature*. England, 414(6859), pp. 112–117. doi: 10.1038/35102174.
Tokai-Nishizumi, N. *et al.* (2005) 'The chromokinesin Kid is required for maintenance of proper metaphase spindle size.', *Molecular biology of the cell*. United States, 16(11), pp. 5455–5463. doi: 10.1091/mbc.e05-03-0244.
Tokai, N. *et al.* (1996) 'Kid, a novel kinesin-like DNA binding protein, is localized to chromosomes and the mitotic spindle.', *The EMBO journal*. England, 15(3), pp. 457–467.
Trapnell, C. *et al.* (2014) 'The dynamics and regulators of cell fate decisions

are revealed by pseudotemporal ordering of single cells', *Nature biotechnology*. 2014/03/23, 32(4), pp. 381–386. doi: 10.1038/nbt.2859.
Tyler, W. A. and Haydar, T. F. (2013) 'Multiplex Genetic Fate Mapping Reveals a Novel Route of Neocortical Neurogenesis, Which Is Altered in the Ts65Dn Mouse Model of Down Syndrome', *The Journal of Neuroscience*, 33(12), pp. 5106 LP – 5119. doi: 10.1523/JNEUROSCI.5380-12.2013.
Vaage, S. (1969) 'The segmentation of the primitive neural tube in chick embryos (Gallus domesticus). A morphological, histochemical and autoradiographical investigation.', *Ergebnisse der Anatomie und Entwicklungsgeschichte*. Germany, 41(3), pp. 3–87.
Vale, R. D., Reese, T. S. and Sheetz, M. P. (1985) 'Identification of a novel force-generating protein, kinesin, involved in microtubule-based motility', *Cell*, 42(1), pp. 39–50. doi: 10.1016/s0092-8674(85)80099-4.
Vicari, S. *et al.* (2019) 'Copy number variants in autism spectrum disorders.', *Progress in neuro-psychopharmacology & biological psychiatry*. England, 92, pp. 421–427. doi: 10.1016/j.pnpbp.2019.02.012.
Walters, R. G. *et al.* (2010) 'A new highly penetrant form of obesity due to deletions on chromosome 16p11.2', *Nature*, 463(7281), pp. 671–675. doi: 10.1038/nature08727.
Walther, C. and Gruss, P. (1991) 'Pax-6, a murine paired box gene, is expressed in the developing CNS.', *Development (Cambridge, England)*. England, 113(4), pp. 1435–1449.
Wang, X. *et al.* (2011) 'A new subtype of progenitor cell in the mouse embryonic neocortex', *Nature neuroscience*. 2011/04/10, 14(5), pp. 555–561. doi: 10.1038/nn.2807.
Wenger, T. L. *et al.* (2016) '22q11.2 duplication syndrome: elevated rate of autism spectrum disorder and need for medical screening.', *Molecular autism*. England, 7, p. 27. doi: 10.1186/s13229-016-0090-z.
WHO (2019) *ICD-11 for Mortality and Morbidity Statistics*.
Wilkinson, G., Dennis, D. and Schuurmans, C. (2013) 'Proneural genes in neocortical development.', *Neuroscience*. United States, 253, pp. 256–273. doi: 10.1016/j.neuroscience.2013.08.029.

Wisniowiecka-Kowalnik, B. and Nowakowska, B. A. (2019) ‘Genetics and epigenetics of autism spectrum disorder-current evidence in the field.’, *Journal of applied genetics*. England, 60(1), pp. 37–47. doi: 10.1007/s13353-018-00480-w.

Xie, Y. *et al.* (2013) ‘The phosphatase PP4c controls spindle orientation to maintain proliferative symmetric divisions in the developing neocortex.’, *Neuron*. United States, 79(2), pp. 254–265. doi: 10.1016/j.neuron.2013.05.027.

Yajima, J. *et al.* (2003) ‘The human chromokinesin Kid is a plus end-directed microtubule-based motor.’, *The EMBO journal*. England, 22(5), pp. 1067–1074. doi: 10.1093/emboj/cdg102.

Yildiz, A. *et al.* (2004) ‘Kinesin walks hand-over-hand.’, *Science (New York, N.Y.)*. United States, 303(5658), pp. 676–678. doi: 10.1126/science.1093753.

Yoo, H. (2015) ‘Genetics of Autism Spectrum Disorder: Current Status and Possible Clinical Applications’, *Experimental neurobiology*. 2015/12/16. The Korean Society for Brain and Neural Science, 24(4), pp. 257–272. doi: 10.5607/en.2015.24.4.257.

Yu, Y. *et al.* (2014) ‘Inhibition of KIF22 suppresses cancer cell proliferation by delaying mitotic exit through upregulating CDC25C expression.’, *Carcinogenesis*. England, 35(6), pp. 1416–1425. doi: 10.1093/carcin/bgu065.

Zecevic, N., Hu, F. and Jakovcevski, I. (2011) ‘Interneurons in the developing human neocortex’, *Developmental neurobiology*, 71(1), pp. 18–33. doi: 10.1002/dneu.20812.

Zoghbi, H. Y. and Bear, M. F. (2012) ‘Synaptic dysfunction in neurodevelopmental disorders associated with autism and intellectual disabilities.’, *Cold Spring Harbor perspectives in biology*. United States, 4(3). doi: 10.1101/cshperspect.a009886.

Appendix

The following manuscript includes results from this thesis and is publicly available on the preprint server BioRXiv. DOI: https://doi.org/10.1101/633461

Expression of genes in the *16p11.2* locus during human fetal cortical neurogenesis.

Sarah Morson[1,2], Yifei Yang[1,2], David J. Price[1,2], Thomas Pratt[1,2,3]

[1]Simons Initiative for the Developing Brain,

[2]Centre for Discovery Brain Sciences,

Hugh Robson Building, Edinburgh Medical School Biomedical Sciences, The University of Edinburgh, Edinburgh, United Kingdom, EH8 9XD, United Kingdom.

[3]Correspondence:

Thomas Pratt

t.pratt@ed.ac.uk

Tel: +44(0)1316503732

Running Title: 16p11.2 gene expression during human corticogenesis

Abstract

The 593 kbp *16p11.2* copy number variation (CNV) affects the gene dosage of 29 protein coding genes, with heterozygous *16p11.2* microduplication or microdeletion implicated in about 1% of autism spectrum disorder (ASD) cases. The *16p11.2* CNV is frequently associated with macrocephaly or microcephaly indicating early defects of neurogenesis may contribute to subsequent ASD symptoms, but it is unknown which *16p11.2* transcripts are expressed in progenitors and whose levels are likely, therefore, to influence neurogenesis. Analysis of human fetal gene expression data revealed that of all the *16p11.2* transcripts only two, *ALDOA* and *KIF22*, are significantly enriched in progenitors. To investigate the role of *ALDOA* and *KIF22* in human cerebral cortex development we used immunohistochemical staining to

describe their expression in late first and early second trimester human cerebral cortex. KIF22 protein is restricted to proliferating cells with its levels increasing during the cell cycle and peaking at mitosis. ALDOA protein is expressed in all cell types and does not vary with cell-cycle phase. Our expression analysis suggests the hypothesis that the simultaneous changes in KIF22 and ALDOA dosage in cortical progenitors causes defects in neurogenesis that may contribute to ASD in *16p11.2* CNV patients.

Keywords:
ALDOA, Autism, Cerebral Cortex, CNV, KIF22

Large, recurrent Copy Number Variations (CNVs) are implicated in many neuropsychiatric disorders including autism spectrum disorders (ASD), epilepsy, intellectual disability (ID) and schizophrenia (McCarthy et al. 2009; Girirajan and Eichler 2010; Levy et al. 2011; Sanders et al. 2011; Malhotra and Sebat 2012). The *16p11.2* CNV (OMIM 611913) encompasses a 593 kb DNA sequence in the p11.2 region of human chromosome 16 (BP4-BP5). This region harbors 29 protein coding genes and is strongly linked to neurodevelopmental disorders (NDDs) including ASD (Kumar et al. 2008; Bijlsma et al. 2009; Rosenfeld et al. 2010; Shinawi et al. 2010; Zufferey et al. 2012). This *16p11.2* region is flanked by two homologous 147kbp sequences that arose after the evolutionary divergence of humans from other primates, generating a hot-spot for mis-aligned recombination that explains the high frequency of the *16p11.2* CNV in the human population and also the high frequency of *de novo 16p11.2* CNV (Nuttle et al. 2016). In humans, the *16p11.2* microdeletion is associated with transient infant brain overgrowth and focal thickening of the cerebral cortex, while the *16p11.2* microduplication is associated with reduced brain size (microcephaly) (Qureshi et al. 2014; Blackmon et al. 2018). The early manifestation of anatomical phenotype in newborns, along with the onset of ASD symptoms in infancy, suggests crucial roles for *16p11.2* genes during neural development. *16p11.2* is the most prevalent CNV associated with ASD, ~1% incidence, making this CNV particularly intriguing and providing motivation for investigating the role played by

16p11.2 genes in brain development and function (Weiss et al. 2008). Available lines of evidence from *16p11.2* rodent models, *16p11.2* patient derived lymphoblastoid cell lines, and human induced pluripotent stem cells genetically engineered to harbor the *16p11.2* CNV indicate that all *16p11.2* mRNAs' levels reflect the altered gene dosage of 16p11.2 genes (50% in microdeletion and 150% in microduplication heterozygotes) (Horev et al. 2011; Blumenthal et al. 2014; Pucilowska et al. 2015; Tai et al. 2016). This indicates that multiple *16p11.2* transcript levels are affected by the *16p11.2* CNV and that the pathology of the *16p11.2* CNV could stem from altered dosage of one or more them.

While none of the individual *16p11.2* genes have been identified as sole causative genes for the *16p11.2* phenotype, the use of different models has implicated individual CNV genes in various phenotypes. *MAPK3*, *QPRT, KCTD13*, *ALDO*A, and *KIF22* have been individually associated with a variety of neural phenotypes including cell proliferation, neuronal morphology, axonal projection and spine morphogenesis, altered head size, and behavioral phenotypes. A study in which *16p11.2* genes were systematically knocked down in zebrafish embryos reported neuroanatomical phenotypes for the vast majority of *16p11.2* genes tested, suggesting that the 16p11.2 phenotype is likely to be polygenic (Blaker-Lee et al. 2012; de Anda et al. 2012; Golzio et al. 2012; Pucilowska et al. 2015, 2018).

The cellular mechanisms by which the CNV causes the patient phenotype are poorly understood. One plausible hypothesis is that disrupted neurogenesis causes changes in neuronal output that produces a brain with abnormal cell number or composition and that this contributes to the *16p11.2* pathology. Consistent with this hypothesis, the *16p11.2* deletion mouse model exhibits proliferation defects in cortical progenitors during pre-natal brain development and subsequently develops ASD-like symptoms (Horev et al. 2011; Pucilowska et al. 2015). However, it is unknown which of the proteins produced by *16p11.2* CNV genes are expressed by progenitor cells in the developing human cerebral cortex and are therefore candidates for regulating neurogenesis.

We focused on the potential for the *16p11.2* CNV to affect neurogenesis in developing human cerebral cortex by identifying *16p11.2* genes that are highly expressed in cerebral cortex progenitors and down-regulated as cells become post-mitotic. We analysed previously published human fetal cortex single cell RNA sequencing (scRNA-seq) data (Pollen et al. 2015) to identify candidate genes and characterize their expression in sections of developing human fetal cerebral cortex from the late first and early second trimester.

Material and Methods

Human Tissue

Human embryos ranging in age from 12-16 post-conceptual weeks were obtained from the MRC/Wellcome-Trust funded Human Developmental Biology Resource at Newcastle Univeristy (HDBR, http://www.hdbr.org/) with appropriate maternal written consent and approval from the Newcastle and North Tyneside NHS Health Authority Joint Ethics Committee. HDBR is regulated by the UK Human Tissue Authority (HTA; www.hta.gov.uk) and operates in accordance with the relevant HTA Codes of Practice.

For cryosections 12 PCW week brains were fixed in 4%PFA/PBS for 1 week then cryoprotected with 30% sucrose/PBS and then embedded in 50:50 30%sucrose:OCT, flash frozen and sectioned at 12µm using a Leica Cryostat.
The following ages were used for this study: 12 PCW (2 brains), 14 PCW and 16 PCW.

scRNA-seq Analysis

The publicly available scRNA-seq data set (Pollen et al. 2015) was used to identify candidate genes. Prior to dataset publication the reads were aligned, we normalized

the RPKMs as log(x+1). Analysis was performed using R studio. To determine genes with significant changes a Wilcox test by *FindAllMarkers* in Seurat package was used. Monocle2 R package was used to order cells in pseudotime. To identify cell-cycle phase specific transcripts we used function CellCycleScoring from Seurat R package.

Immunohistochemistry

Immunohistochemistry was carried out on paraffin sections obtained from HDBR. Antigen retrieval consisting of boiling sections in 10mM sodium citrate pH6 for 10 mins was used for all stains. Primary antibodies were diluted in 20% blocking serum in pH7.6 Tris buffered saline (TBS) and sections incubated overnight at 4°C. Primary antibodies used: KID 1/5000 DAB, 1/2000 fluorescent (Invitrogen PA5-29490), KI67 1/800 (Novus Biologicals NBP2-22112), ALDOA 1/100 (Sigma HPA004177).

For colourmetric stains, sections were incubated 1hr at room temperature with biotinylated secondary antibody (1/200) followed by incubation for 1 hour with ABC (Vector Labs) and developed with diaminobenzidine solution (Vector Labs), washed, counterstained with nuclear fast red, dehydrated and then mounted using DPX.

For immunofluorescence sections were incubated with secondary antibodies 1/200 1 hour room temperature, counterstained with 4',6-diamidino-2-phenylindole dihydrochloride (DAPI;ThermoFisher) and mounted with Vectashield H1400 Hardset Mounting Medium (Vector Labs). Extensive TBS washes were carried out between each step.

***In Situ* Hybridisation**

Cryosections of 12 PCW brain were used for ALDOA and KIF22 *in situ* hybridisation.

PCR primers used to clone *in situ* probes from human cDNA into pGEMTeasy for

preparation of DIG labelled RNA were as follows: ALDOA Forward: CTG TCA CTG GGA TCA CCT T, ALDOA Reverse: GTG ATG GAC TTA GCA TTC AC. KIF22 Forward: CGA GAG CGG ATG GTG CTA AT, KIF22 Reverse: GAG ACC CAG GAT GTT TGC CT. *In situ* hybridisation was performed as described previously (Radonjić et al. 2014). Briefly, 12µm cryosections were dried at 37°C for 3 hours then incubated overnight at 70°C in hybridization mix containing 1x salts (200mM NaCl, 10mM Tris HCl (pH 7.5), 1mM Tris Base, 5mM $NaH_2PO_42H_2O$, 5mM Na_2HPO_4, 0.5M EDTA: Sigma-Aldrich), 50% deionized formamide, 10% dextran sulfate, 1mg/ml rRNA, 1x Denhardt's, and DIG-labelled RNA probe. Next day sections were washed 3 times at 70°C in wash buffer comprising 1x SSC, 50% formamide, 0.1% Tween-20 and then 3 times at RT in 1x MABT (20mM Maleic acid, 30mM NaCl, 0.5% Tween-20 and pH adjusted to 7.5 with 10mM NaOH). Sections were incubated 1hr RT in 1x MABT blocking solution (20% sheep serum, 2% blocking reagent) and then incubated overnight with anti-DIG antibody 1:1500 in blocking solution at 4°C followed by colour reaction overnight at RT.

Microscopy and Imaging

DAB and *in situ* hybridisation images were taken using a Leica DMNB microscope with an attached Leica DFC480 Camera. Fluorescence images were obtained with a Leica DM5500B epifluorescence microscope with a DFC360FX camera. Confocal images were obtained using Nikon A1R FILM microscope and analysed in ImageJ.

Image Analysis and Quantification

For DAB stains and *in situ* hybridisation the images were stitched in ImageJ using the stitching plugin (Preibisch et al. 2009).

For KIF22 analysis of DAB stains rectangular counting boxes (34x88µm) were overlaid across the section. Using ImageJ cell counting plugin cells in each box were counted and denoted KIF22+ (brown) or KIF22- (red). The distinction between the regions (VZ, SVZ, IZ/CP) was determined anatomically by cell density. The count for each box was averaged with other boxes in the region to provide the final value.

For analysis of KIF22/KI67 double staining counting boxes (20x145µm) were

overlaid over the VZ and SVZ (determined based on cell density). For determining intensity cells were randomly selected on the DAPI channel, the nucleus outlined and intensity of KIF22 and KI67 recorded. 20 cells were selected per box and the counts from individual boxes combined to give final values.

For subcellular ALDOA analysis counting boxes (20x145µm) were overlaid over the VZ and SVZ. Cells were randomly selected on the DAPI channel, far enough apart to ensure their cytoplasm would not overlap, the Z plane through the center of the cell was used and the nucleus outlined. The KI67 and ALDOA intensity was measured constituting the nuclear value. To obtain ALDOA cytoplasmic intensity the nuclear outline was duplicated and extended 4 pixels allowing a reading of just the cytoplasmic area to be obtained (see Fig.5e). This was performed for 10 cells in each box and the counts from individual boxes combined to give final values.

Data Analysis and Statistics

Where error bars are shown they are expressed as mean ± SEM. Boxplots show median and upper and lower quartiles. Statistical comparison between two groups was performed with a *t* test. Statistical comparison between more than two groups was performed with ANOVA followed by *post hoc* test. $P<0.05$ was considered statistically significant. Analysis was performed using GraphPad Prism.

Results

Analysis of scRNA-seq data identifies *KIF22* and *ALDOA* as progenitor-enriched *16p11.2* transcripts in the developing human fetal cerebral cortex.

The *16p11.2* CNV involves microduplication or microdeletion of a 593 kb locus on human chromosome 16 containing 29 protein coding genes (Fig.1a). The aim of the current study is to identify *16p11.2* genes that are potential candidates for being involved in neurogenesis in the developing human cerebral cortex (Fig.1b) and whose altered dosage in *16p11.2* microdeletion or microduplication patients may disrupt neurogenesis and contribute to the CNV phenotype. We reasoned that *16p11.2* genes important for neurogenesis would be highly expressed in proliferating progenitor cells and down-regulated as cells became post-mitotic.

We took advantage of a published single cell RNA-sequencing (scRNA-seq) data-set acquired from 393 cells of the ventricular zone (VZ) and subventricular zone (SVZ) of gestational week (GW) 16-18 human fetal cerebral cortex (equivalent to post conception week (PCW) 14-16) to perform an unbiased screen to identify *16p11.2* transcripts that matched this expression profile (Pollen et al. 2015). Dimensional reduction of the scRNA-seq data separated the cells into three clusters based on transcriptome similarity (Fig.1c – each dot on the tSNE plot represents an individual cell) that were subsequently identified as the three cardinal cell classes of progenitors (blue), post-mitotic neurons/ principal cells (red) and interneurons (green), by expression of cell-type specific transcripts. We next used the monocle2 R package to order the cells in pseudotime using the normalized expression levels of selected differentially expressed genes (DEGs) as input to order the cells (Trapnell et al. 2014; Qiu et al. 2017) (Fig.1d) moving from the progenitor state (left) to post mitotic state (right) along the X-axis. We plotted the average expression of each *16p11.2* transcript at each pseudotime-point on the Y-axis. We found that two genes, *KIF22* (blue line) and *ALDOA* (brown line), were notable for having high expression in progenitors that declined as cells became post-mitotic. A Wilcox test identified *ALDOA* and *KIF22* as the only *16p11.2* transcripts that were significantly higher in progenitor than neuronal populations ($p<<0.05$). Although not significantly enriched in progenitors *HIRIP3* (orange line), *MAZ* (red line), *PAGR1* (green line), and *SPN* (turquoise line) transcripts were expressed in progenitors at higher levels than the remaining *16p11.2* transcripts (shown as grey lines), many of which were barely expressed at all.

We focused on *ALDOA* and *KIF22* transcripts and used a combination of *in situ* hybridisation on 12 PCW human cerebral cortex and analysis of the scRNA-seq data to investigate their expression in more detail. *In situ* hybridisation shows that *KIF22* mRNA expression is most prominent in the VZ and SVZ with a few expressing cells in the IZ and SP/CP (Fig.1e). Violin plots of the numbers of cells expressing different levels of *KIF22* mRNA in the different cardinal cell classes show that *KIF22* is expressed predominantly in progenitors (Fig.1f) and mapping the expression level of *KIF22* onto the tSNE plot (Fig.1g) revealed that *KIF22* expression is highest in a

subset of the progenitor cluster (arrow in Fig.1g) with a substantial proportion of progenitor cells expressing relatively low levels of *KIF22.* Very few post-mitotic neurons, both interneurons and principal cells, express appreciable levels of *KIF22* (Fig.1f,g). The expression of *KIF22* in a subset of progenitors prompted us to ask whether its expression was related to the cell-cycle phase. We used the expression of cell-cycle phase specific transcripts using function CellCycleScoring from Seurat R package to divide the cells into three classes (Macosko et al. 2015; Tirosh et al. 2016), G1/S, G2/M, and post-mitotic neurons, and compared *KIF22* transcript levels between these three groups using a violin plot (Fig.1h). We found that the majority of cells in G2/M phase expressed higher levels of *KIF22* (red plot), cells in G1/S expressed lower levels (blue plot) while the vast majority of post-mitotic cells expressed low levels of *KIF22* (green plot).

We performed the same analysis for *ALDOA* transcripts. *In situ* hybridisation showed that while *ALDOA* mRNA expression is most prominent in the proliferative VZ and SVZ there are substantial numbers of *ALDOA* expressing cells in the SP/CP (Fig.1i). Violin plots (Fig.1j) show that while a greater proportion of cells expressing the highest levels of *ALDOA* are progenitors (blue plot) there are also a substantial number of principal cells (red plot) expressing similarly high levels of *ALDOA* transcripts although very few interneurons (green plot). Mapping *ALDOA* expression level onto the tSNE plot (Fig.1k) shows cells expressing high levels of ALDOA are evenly distributed throughout the progenitor cluster with appreciable numbers of principal cells expressing high levels of *ALDOA* and a much lower proportion of interneurons. In contrast to *KIF22*, there is no clear difference in the partitioning of *ALDOA* expression level between different phases of the cell-cycle (Fig.1l).

To conclude, of all the 29 *16p11.2* transcripts, only two, *KIF22* and *ALDOA*, are significantly enriched in progenitors compared to post-mitotic cells making them candidates for having specific roles in neurogenesis in the developing human fetal cerebral cortex. Although both are enriched in progenitors, *KIF22* and *ALDOA* transcript expression shows notable differences: *KIF22* transcripts are more restricted

to progenitors and their levels vary as the cell-cycle progresses. We next describe the expression of KIF22 and ALDOA protein over a range of developmental stages.

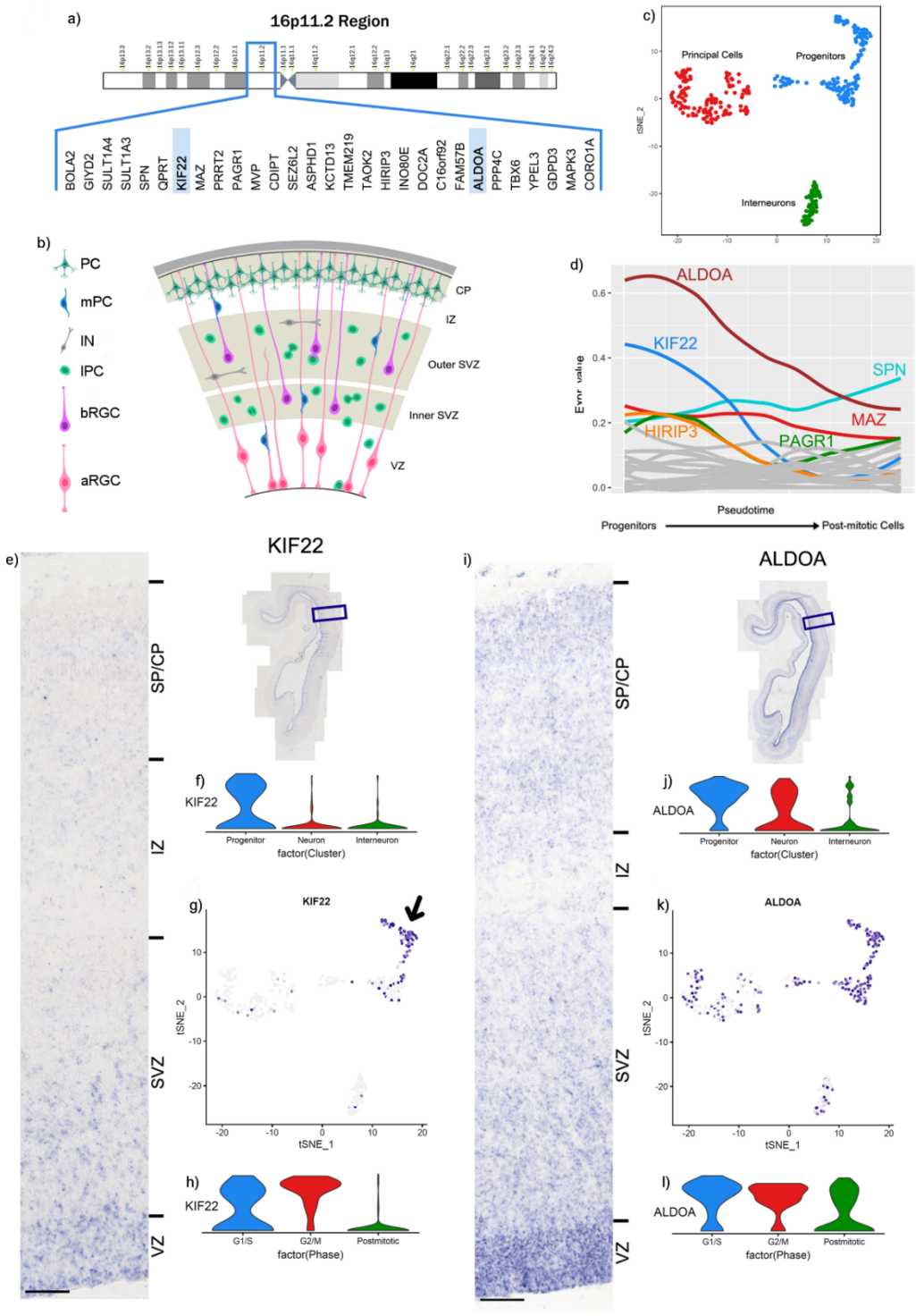

a)
16p11.2 Region
BOLA2
GIYD2
SULT1A4
SULT1A3
SPN
QPRT
KIF22
MAZ
PRRT2
PAGR1
MVP
CDIPT
SEZ6L2
ASPHD1
KCTD13
TMEM219
TAOK2
HIRIP3
INO80E
DOC2A
C16orf92
FAM57B
ALDOA
PPP4C
TBX6
YPEL3
GDPD3
MAPK3
CORO1A
b)
PC
mPC
IN
IPC
bRGC
aRGC
CP
IZ
Outer SVZ
Inner SVZ
VZ
c)
Principal Cells
Progenitors
Interneurons
tSNE_2
d)
ALDOA
KIF22
SPN
MAZ
HIRIP3
PAGR1
Expr value
Pseudotime
Progenitors
Post-mitotic Cells
e)
KIF22
SP/CP
IZ
SVZ
VZ
f)
KIF22
Progenitor
Neuron
Interneuron
factor(Cluster)
g)
KIF22
tSNE_1
tSNE_2
h)
KIF22
G1/S
G2/M
Postmitotic
factor(Phase)
i)
ALDOA
SP/CP
IZ
SVZ
VZ
j)
ALDOA
Progenitor
Neuron
Interneuron
factor(Cluster)
k)
ALDOA
tSNE_1
tSNE_2
l)
ALDOA
G1/S
G2/M
Postmitotic
factor(Phase)

Fig 1: a) 16p11.2 region and genes. b) schematic of human cortical structure during development. c) tSNE clustering of cell types. d) changing mRNA expression levels of 16p11.2 genes as cells move from progenitors to neurons with KIF22 and ALDOA identified as changing significantly. e) *in situ* hybridisation images of KIF22 at 12 PCW. f) Violin plots showing distribution of KIF22 in different cell types. g) KIF22 gradient plot. h) Violin plots showing distribution of KIF22 at different cell cycle stages i) *in situ* hybridisation images of ALDOA at 12 PCW. j) Violin plots showing distribution of ALDOA mRNA levels in different cell types. k) ALDOA gradient plot. l) Violin plots showing distribution of ALDOA mRNA at different cell cycle stages. Scale bars = 100μm.

KIF22 protein is expressed in germinal zones of 12, 14 and 16 PCW cortex

Here we characterize KIF22 protein expression during human corticogenesis. Coronal cortex sections spaced along the rostral-caudal axis were immunostained for KIF22 protein and counterstained with Nuclear Fast Red (NFR) to show cytoarchitecture. $KIF22^{+}$ (brown) and $KIF22^{-}$ (red) cells were counted for each region in the telencephalic wall (VE, VZ, SVZ, IZ and CP) (see methods for details of sampling) and lamination was identified by cell density (Bayer and Altman 2002, 2005). These data are shown for three developmental stages, 12 PCW (Fig 2 a-d), 14 PCW (Fig 2 a'-d'), and 16 PCW (Fig 2 a*-d*). At all stages and rostro-caudal positions KIF22 expressing cells appear most abundant in the VE followed by the VZ and SVZ with the IZ and CP presenting a very low to complete absence of KIF22 (Fig.2 c, c', and c* with higher magnification of boxed regions from each zone shown in d, d', and d* respectively, green arrows indicate examples of individual $KIF22^{+}$ cells).

We next pooled $KIF22^{+}$ cell count data in two ways to compare between all ages (Fig 2e) and anatomical regions (Fig 2f) and found that the percentage of $KIF22^{+}$ cells in the VE (40-50%) was consistently higher than other regions, followed by the VZ (20-

30%) and SVZ (10%), with even fewer cells (<10%) in the IZ, and CP (Fig.2e,f). This result describes KIF22 protein as predominantly restricted to a subset of cells in the germinal zones of the developing cortex at all stages studied.

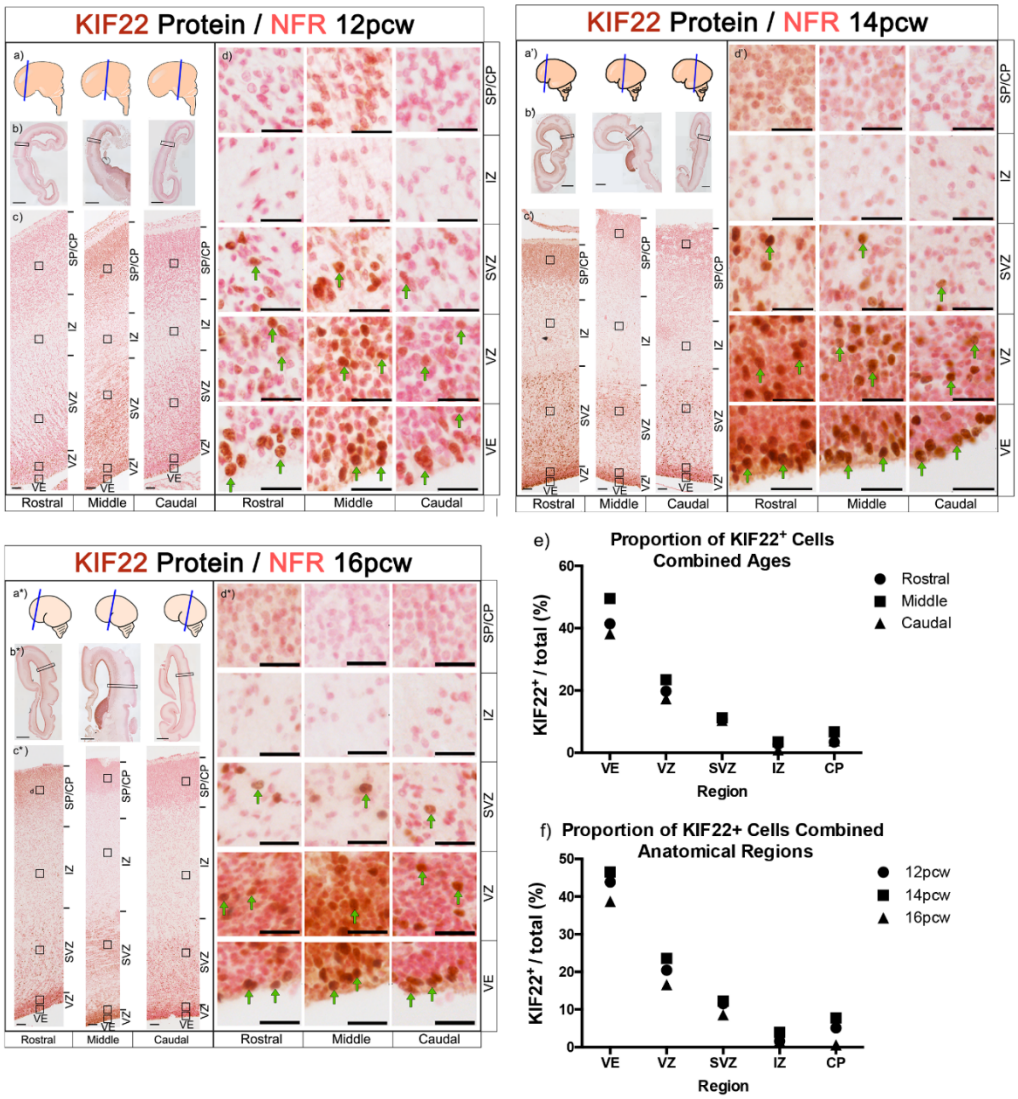


Fig 2. KIF22 protein expression levels in the cerebral cortex at 12, 14 and 16 PCW: a, a', a*) schematic showing brain regions sectioned. b, b', b*) images of whole brain section, scale bars =2mm. c, c', c*) sections spanning the rostral-caudal axis showing KIF22 expression in the telencephalic wall, scale bars = 100µm. d, d', d*) high magnification images of different cortical zones rostral-caudal. $KIF22^+$ cells in brown and examples indicated by green arrows, $KIF22^-$ cells in pink, scale bars =25 µm. e) Quantification of KIF22 expressing cells with all three ages combined. f) Quantification of KIF22 expressing cells with rostral, middle, caudal values combined.

KIF22 protein expression is restricted to proliferating cells

KIF22 protein expression is almost exclusively restricted to a subset of cells in the proliferative regions. From the scRNA-seq data, we expect these to be progenitor cells (Fig.1). To identify these KIF22 positive cells we performed double immunofluorescence for KIF22 and KI67 (a protein expressed in all proliferating cells (Scholzen and Gerdes 2000; Miller et al. 2018)). $KIF22^+$ cells were predominantly located in the VE, VZ and SVZ (Fig.2), therefore these regions were examined for analysis. Low magnification of KI67/KIF22 staining are shown in 12 PCW (Fig.3a) and 14 PCW (Fig.3b) with higher magnification showing individual cells in Fig.3c. Cell counts for $KIF22^+/KI67^+$ labelled cells show that the majority (80-90%) of $KI67^+$ cells also express KIF22 both at 12 (Fig.3d) and 14 PCW (Fig.3e) or across the rostral-caudal axis. Combining the data for anatomical locations and ages revealed significantly more $KIF22^+/KI67^+$ cells than $KIF22^+/KI67^-$ and $KIF22^-/KI67^+$ cells (Fig.3f).

Quantification of nuclear KIF22 fluorescence level at the single-cell level shows KIF22 protein is variable, but also significantly higher in $KI67^+$ (proliferating) cells than in $KI67^-$, across the rostral caudal axis at 12 (Fig.4a-c) and 14 PCW (Fig.4g-i). Combining data from the two ages and locations showed significantly higher KIF22 fluorescence in $KI67^+$ cells (Fig.4m). Thus, we confirm $KIF22^+$ cells are proliferating and that $KI67^+$ cells have higher KIF22 protein levels.

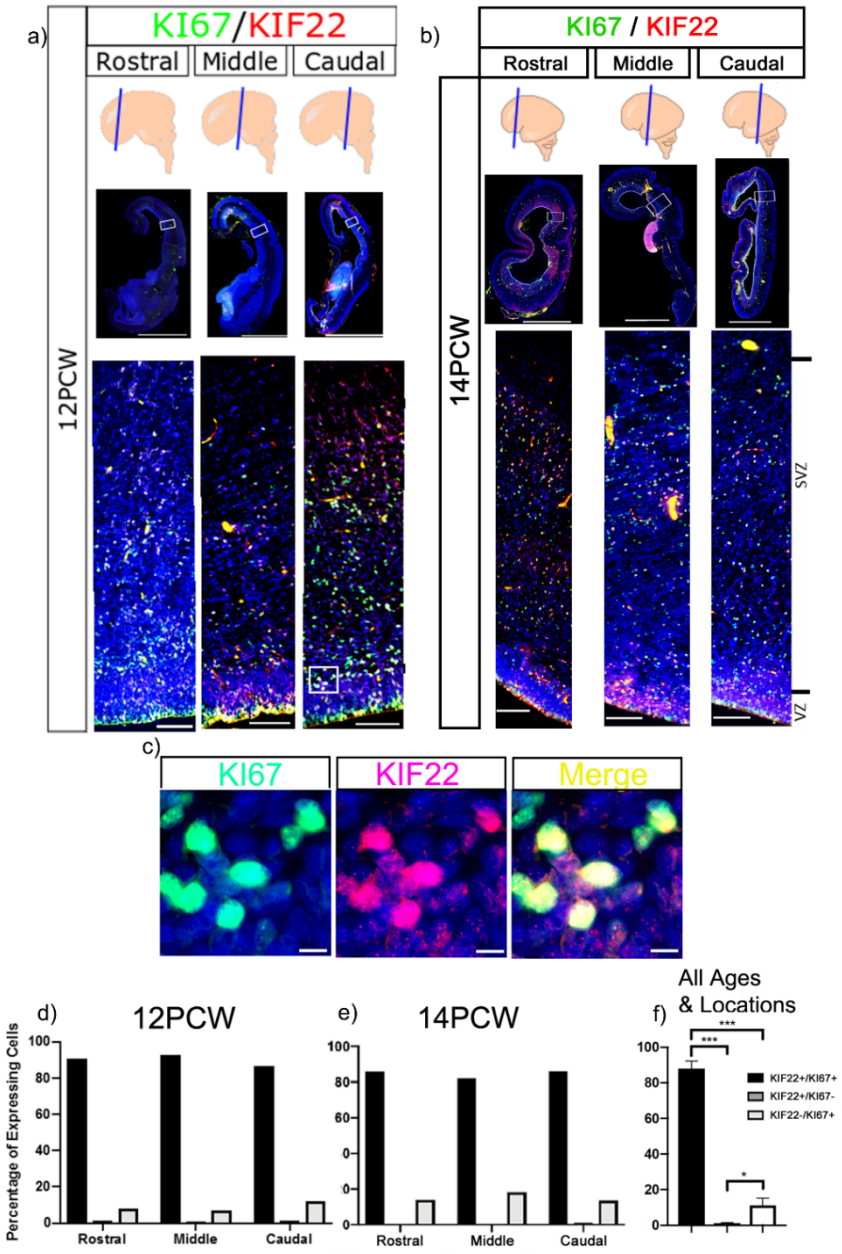

a)
KI67/KIF22
Rostral
Middle
Caudal
12PCW
b)
KI67 / KIF22
Rostral
Middle
Caudal
14PCW
SVZ
VZ
c)
KI67
KIF22
Merge
d)
12PCW
Percentage of Expressing Cells
100
80
60
40
20
0
Rostral
Middle
Caudal
e)
14PCW
100
80
60
0
0
Rostral
Middle
Caudal
f)
All Ages
& Locations
***
***
*
100
80
60
40
20
0
KIF22+/KI67+
KIF22+/KI67-
KIF22-/KI67+

Fig 3. Immunofluorescence of KIF22 and KI67 in the cortex: a) KIF22 and KI67 at 12 PCW, low magnification scale bars = 4mm, high magnification scale bars = 100 μm. b) KIF22 and KI67 at 14 PCW, low magnification scale bars = 4mm, high magnification scale bars = 100 μm. c) high magnification of KI67/KIF22 expressing cells. Scale bars = 10 μm. d) Percentage of cells expressing KIF22, KI67 or both at 12 PCW. e) Percentage of cells expressing KIF22, KI67 or both at 14 PCW. f) Combined data of percentage of cells expressing KIF22, KI67 or both.

KIF22 levels vary with cell-cycle phase

From the scRNA-seq analysis, and the variable KIF22 protein levels in $KI67^+$ cells, we hypothesised that KIF22 protein levels change throughout the cell cycle. To test this, we quantified nuclear immunofluorescence intensity of KIF22 and KI67 in two 12 PCW brains (see methods for details of sampling procedure). KI67 protein levels vary during the cell cycle: lowest in G1 phase, increasing through S and G2 to peak in mitosis (Fig.4o) (Scholzen and Gerdes 2000; Miller et al. 2018). We found a strong correlation between KIF22 and KI67 intensity (Brain 1 rostral R^2=0.8095, middle R^2=0.8139, caudal R^2= 0.6691. Brain 2 rostral R^2 = 0.7489, middle R^2=0.7447, caudal R^2=0.7763) (Fig.4d-f). To ensure the correlation observed was not a result of nucleus size changing with cell cycle, we confirmed that KIF22 protein levels did not correlate with nuclear size by DAPI staining (Brain 1 rostral KIF22 R^2=0.104, middle KIF22 R^2=0.0874, caudal KIF22 R^2=0.2969. Brain 2 rostral KIF22 R^2=0.1183, middle KIF22 R^2=0.0512, caudal KIF22 R^2=0.2287). A strong positive correlation was also observed at 14 PCW (Brain 1 rostral R^2=0.7465, middle R^2=0.6668, caudal R^2= 0.634, Fig.4j-l). Again, we confirmed that KIF22 protein levels did not correlate with nuclear size (rostral KIF22 R^2=0.1239, middle KIF22 R^2=0.0599, caudal KIF22 R^2=0.0229). This demonstrates that the correlation between KI67 and KIF22 is consistent between ages and rostral-caudal location. Combining all values of KI67/KIF22 nuclear intensity values showed that KIF22 was expressed at significantly higher levels in $KI67^+$ cells (Fig.4m) with a strong correlation (R^2=0.7236) between KIF22 and KI67 levels

(Fig.4n). Although KIF22 expressing cells were scattered throughout the VE, VZ and SVZ there was a general trend for the cells expressing the highest levels of KIF22 to be closest to the apical surface (yellow coloured dots on scatterplots Fig.4d-g, g-i,n) with lower expressing cells tending to be further from the apical surface (blue coloured dots on scatterplots Fig.4 d-g, g-I,). During interkinetic nuclear movement radial glial cell nuclei move to the apical surface to perform mitosis so this spatial distribution suggests KIF22 is expressed at high levels by radial glial cells undergoing mitosis at the apical surface of the VZ.

From these data we show that KIF22 protein levels change throughout the cell cycle in positive correlation with KI67: KIF22 is present in G1 and increases through S and G2 phase to peak in mitosis (Fig.4p).

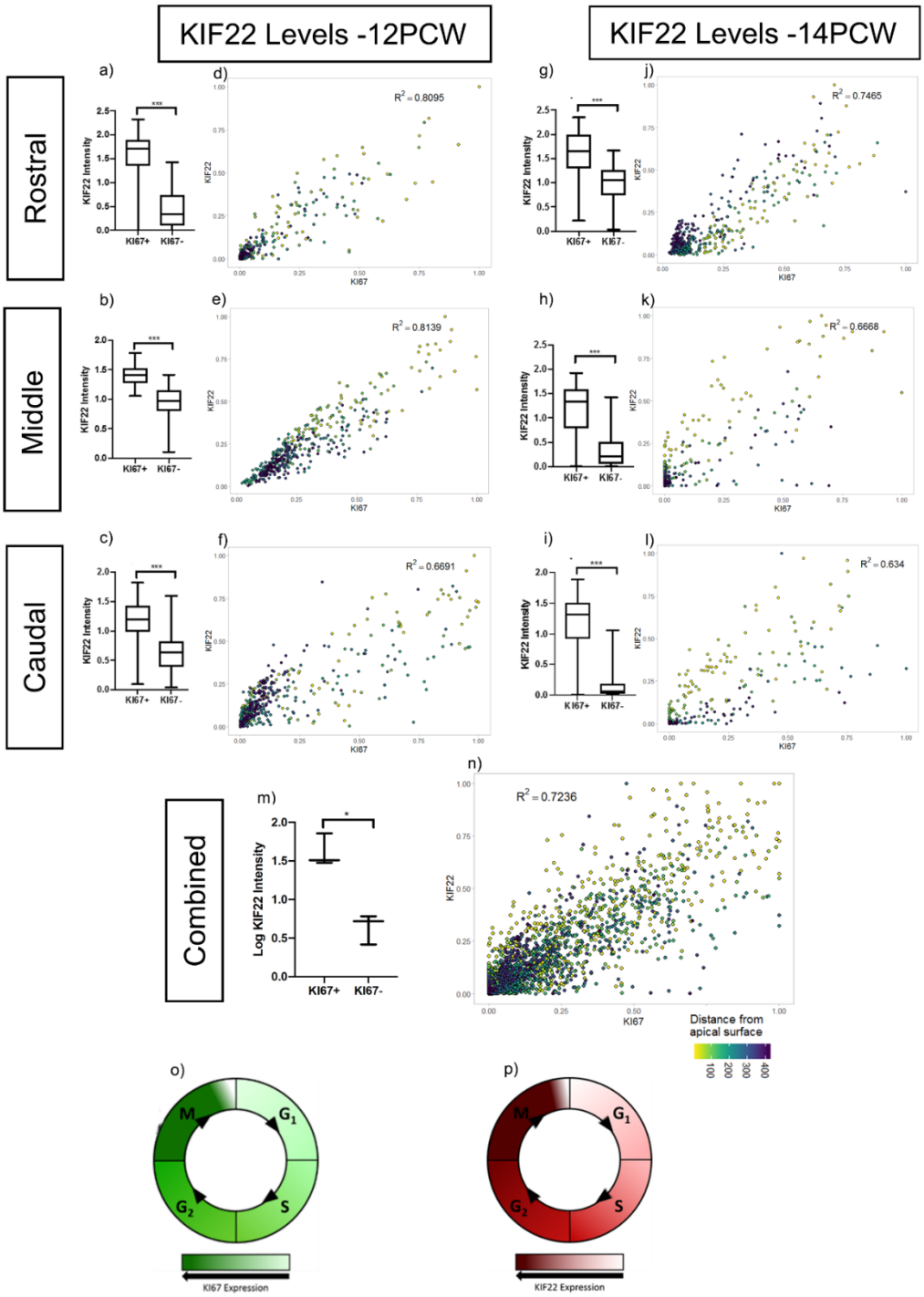

KIF22 Levels -12PCW
KIF22 Levels -14PCW
Rostral
Middle
Caudal
Combined
a)
b)
c)
d)
e)
f)
g)
h)
i)
j)
k)
l)
m)
n)
o)
p)
KIF22 Intensity
Log KIF22 Intensity
KI67+
KI67-
KIF22
KI67
R² = 0.8095
R² = 0.8139
R² = 0.6691
R² = 0.7465
R² = 0.6668
R² = 0.634
R² = 0.7236
Distance from apical surface
M
G1
S
G2
KI67 Expression
KIF22 Expression

Fig 4. Quantification of KIF22 Levels: a,b,c) 12 PCW quantification of KIF22 fluorescence intensity in $KI67^{+}/KI67^{-}$ cells (raw data transformation =+1(log), unpaired t-test with Welch's correction, p=<0.001). d,e,f) intensity correlations of KIF22 and KI67 nuclear fluorescence intensity at 12 PCW. g, h, i) 14 PCW quantification of KIF22 fluorescence intensity in $KI67^{+}$/KI67- cells (raw data transformation =+1(log), unpaired t-test with Welch's correction, p=<0.001). j, k, l) intensity correlations of KIF22 and KI67 nuclear fluorescence intensity at 14 PCW. m) quantification of KIF22 fluorescence intensity in $KI67^{+}/KI67^{-}$ cells 12 and 14 weeks combined (raw data transformation =(log), paired t-test, p=0.0122). n) intensity correlations of KIF22 and KI67 nuclear fluorescence intensity for rostral-caudal points at 12 and 14 PCW with distance from apical surface indicated by dot colour. o) diagram of KI67 protein levels throughout the cell cycle. p) model based on our results of KIF22 protein levels throughout the cell cycle.

ALDOA protein is highest in the germinal zones of the cortex

Bioinformatics analysis and *in situ* hybridisation show *ALDOA* mRNA levels decrease as progenitor cells move towards a neuronal fate (Fig.1). Here we used immunofluorescence to characterize ALDOA protein expression across the telencephalic wall at 3 developmental time points; at 12, 14 and 16 PCW, ALDOA immunofluorescence is most intense in the VZ and SVZ before decreasing in the cortical plate (Fig.5 a-c). Double immunofluorescence for KI67 and ALDOA viewed at high magnification shows that ALDOA protein is primarily localized outside $DAPI^{+}$ nuclei in the cytoplasm and that ALDOA is expressed by $KI67^{+}$ progenitor cells and also by cells that do not express KI67 (Fig.5d). The schematic (Fig.5e) illustrates the

areas used for quantification of nuclear and whole cell ALDOA fluorescence presented in Fig.6.

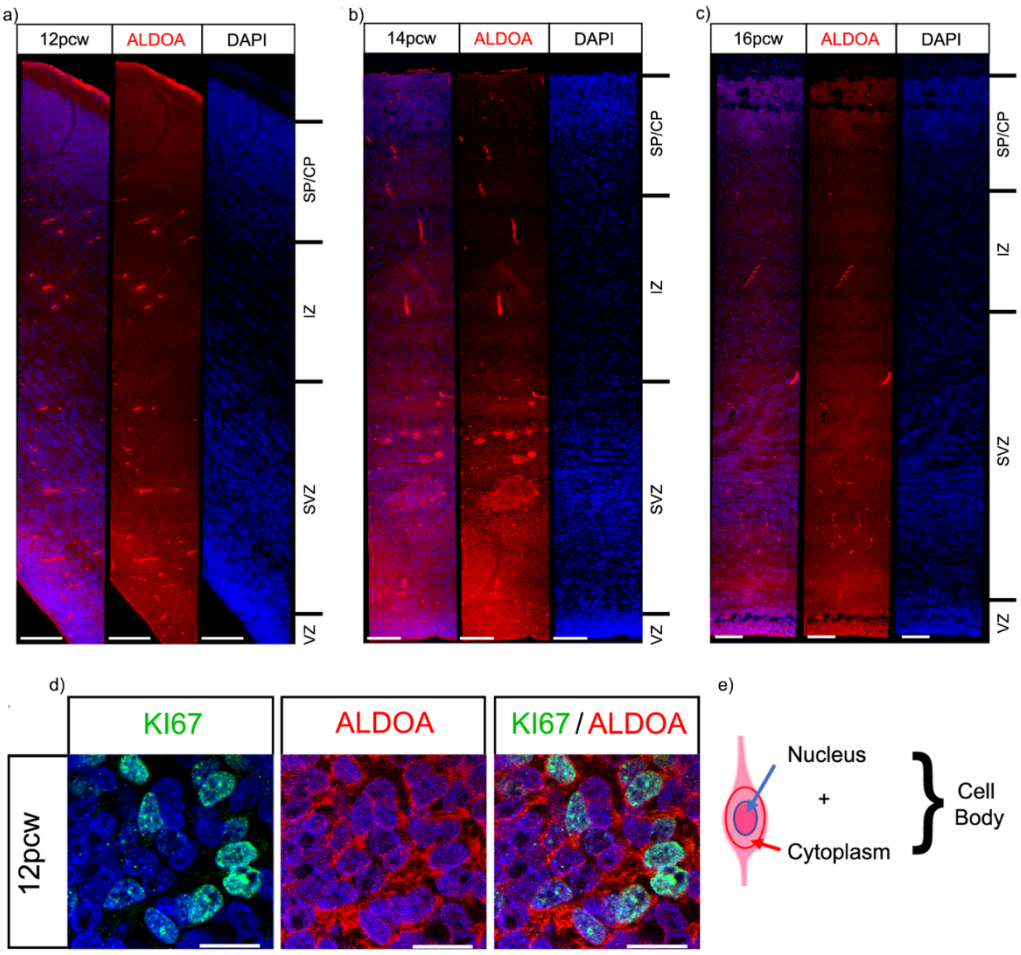


Fig 5. ALDOA Expression in the Cortex: ALDOA protein fluorescence intensity across the telencephalic wall at a) 12, b) 14 and c) 16 PCW. Scale bars = 100 µm. d) High magnification immunofluorescence of KI67 and ALDOA proteins scale bar = 10µm. e) Schematic of quantification method.

ALDOA protein levels do not correlate with proliferation

Although examination of ALDOA mRNA expression indicated it was enriched in progenitors we were unable to find a significant difference in ALDOA protein levels between $KI67^+$ and $KI67^-$ cells at 12 (Fig.6a), 14 (Fig.6b) and 16 PCW (Fig.6c) in the human cortex. To look for any fluctuation in ALDOA levels with the cell cycle we quantified immunofluorescence for KI67 and cell body ALDOA (nucleus and adjacent cell body) using the same analysis as that described above for KIF22, and found no correlation or discernible pattern at 12 (Fig.6e) (R^2= 0.018), 14 (Fig.6f) (R^2= 2e-4) or 16 PCW (Fig.6g) (R^2= 0.00992). These data show that in human cortex development, cellular ALDOA protein levels do not correlate with proliferation or fluctuate with cell cycle.

Previous work in different models demonstrated nuclear ALDOA level is greater in proliferating cells (Mamczur et al. 2010, 2013). To see if this was the case in human cortex development, we quantified nuclear ALDOA and KI67 (Fig.5e) but found no significant difference in nuclear ALDOA fluorescence between $KI67^+$ and $KI67^-$ cells at 12 (Fig.6i), 14 (Fig.6j) or 16 PCW (Fig.6k). We next tested if nuclear ALDOA levels in proliferating cells varied with cell cycle. Analysis of ALDOA and KI67 nuclear intensity established no correlation or pattern at 12 (Fig.6m) (R^2= 5e-05), 14 (Fig.6n) (R^2= 0.0365) or 16 PCW (Fig.6o) (R^2= 0.0723). This shows nuclear ALDOA levels do not increase with proliferation, nor fluctuate with cell cycle. We combined results across the 12, 14 and 16 PCW. There was no significant difference between $KI67^+$ and $KI67^-$ cells when examining ALDOA protein intensity in the whole cell (Fig.6d) or the nucleus (Fig.6l). Using the pooled data, there was no correlation or discernible pattern when nuclear ALDOA intensity was graphed against nuclear KI67 level for the whole cell (Fig.6h) (R^2= 0.0049) or the nucleus (Fig.6p) (R^2= 6e-04).

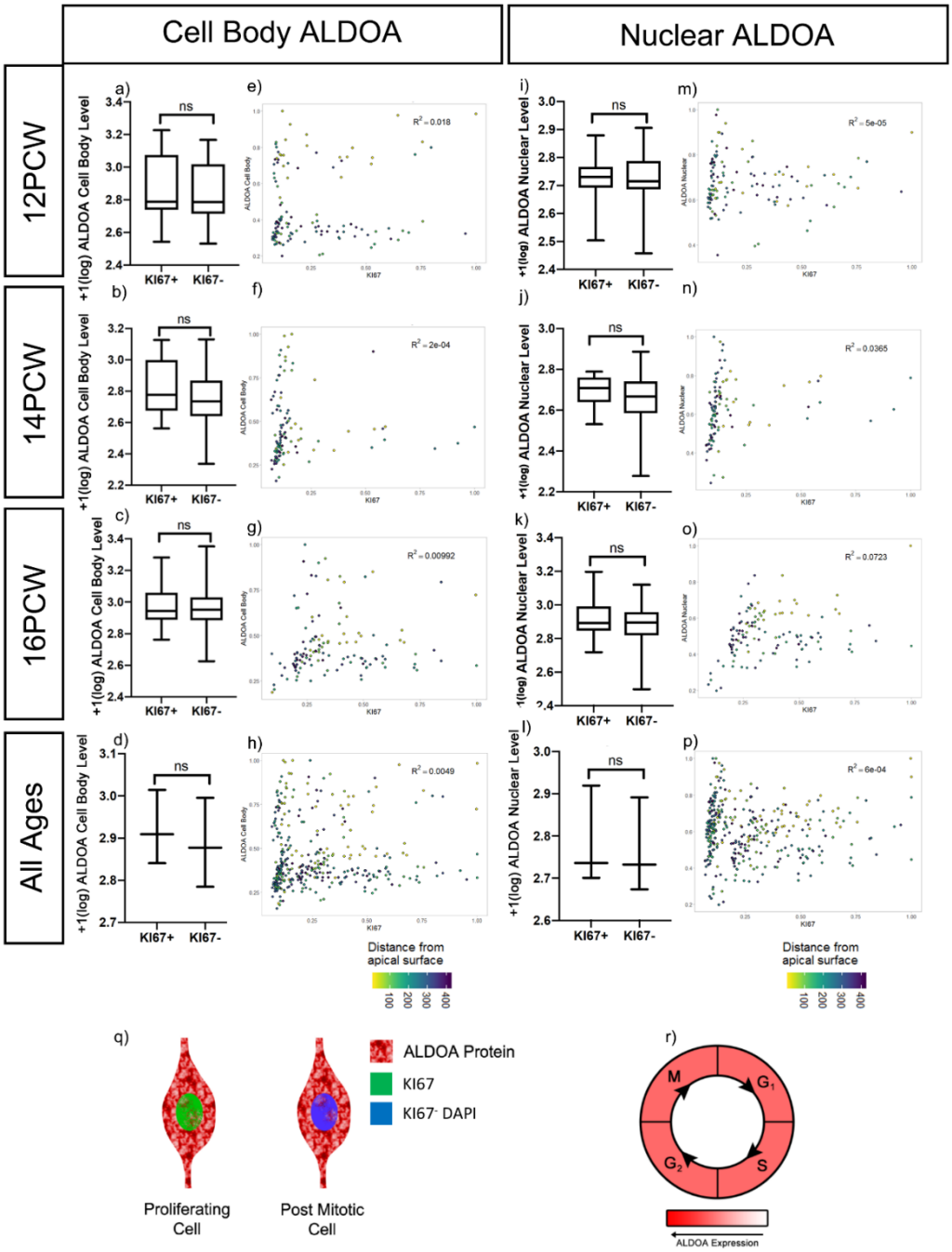

Cell Body ALDOA
Nuclear ALDOA
12PCW
14PCW
16PCW
All Ages
+1(log) ALDOA Cell Body Level
+1(log) ALDOA Nuclear Level
KI67+
KI67-
ns
Distance from apical surface
100
200
300
400
ALDOA Protein
KI67
KI67- DAPI
Proliferating Cell
Post Mitotic Cell
M
G1
S
G2
ALDOA Expression

Fig 6. ALDOA Quantification: a-d) Cell body ALDOA protein fluorescent intensity in $KI67^+$ and $KI67^-$ cells at a) 12 PCW (raw data transformation =+1(log), bimodal distribution, Mann-Whitney test, p = 0.3702), b) 14 PCW (raw data transformation =+1(log), normal distribution, unpaired t-test with Welch's correction, p = 0.2032), c) 16 PCW (raw data transformation =+1(log), normal distribution, unpaired t-test with Welch's correction, p = 0.3523). d) ALDOA cell body protein fluorescent intensity in $KI67^+$ and $KI67^-$ cells, 12, 14 and 16 PCW individual datasets averaged, (raw data transformation =+1(log)), paired t-test, p = 0.0836. e-h) ALDOA cellular protein intensity levels ls correlated to nuclear KI67 protein intensity at e)12, f)14 and g)16 PCW with distance from ventricular edge indicated. h) ALDOA whole cell protein intensity levels correlated to nuclear KI67 protein intensity pooled 12, 14, 16 PCW. i-l) Nuclear ALDOA protein fluorescent intensity in $KI67^+$ and $KI67^-$ cells at i) 12 PCW (raw data transformation =+1(log), normal distribution, unpaired t-test with Welch's correction, p = 0.7543), j) 14 PCW (raw data transformation =+1(log), normal distribution, unpaired t-test with Welch's correction, p = 0.0694), k) 16 PCW (raw data transformation =+1(log), normal distribution, unpaired t-test with Welch's correction, p = 0.0772). l) ALDOA nuclear protein fluorescent intensity in $KI67^+$ and $KI67^-$ cells, 12, 14 and 16 PCW individual datasets averaged, (raw data transformation =+1(log)), paired t-test, p = 0.1330. m-p) ALDOA nuclear protein intensity levels is correlated to nuclear KI67 protein intensity at m)12, n)14 and o)16 PCW with distance from ventricular edge indicated. p) ALDOA nuclear protein intensity levels correlated to nuclear KI67 protein intensity pooled 12, 14, 16 PCW. q) schematic demonstrating ALDOA protein is predominantly in the cytoplasm and lower in the nucleus in both $KI67^+$ proliferating cells and $KI67^-$ post mitotic cells. r) model showing ALDOA levels do not change with the cell cycle.

Discussion

16p11.2 transcript expression during human neurogenesis.

The *16p11.2* CNV is a polygenic mutation that causes NDDs and the current study identified a number of the 29 *16p11.2* transcripts expressed in progenitor cells of the cerebral cortex. In addition to *ALDOA* and *KIF22* that are significantly enriched in progenitors, several other transcripts (e.g. *HIRIP3*, *PAGR1*, *MAZ*, and *SPN*) are also expressed in progenitors albeit at lower levels and are not significantly down-regulated as cells become post-mitotic. The simultaneous expression of multiple *16p11.2* genes in cells undergoing neurogenesis suggests that these cells may be particularly vulnerable to simultaneous alteration in their dosage as a consequence of the *16p11.2* microdeletion or microduplication. This lends support to the hypothesis that neurogenesis is disrupted in *16p11.2* CNV patients and that this contributes to subsequent development of NDDs.

KIF22

KIF22 is a multifunctional protein that can regulate cell proliferation through at least two distinct mechanisms. First, KIF22 is a kinesin-like microtubule-based motor that binds microtubules and chromosomes during mitosis and regulates mitotic spindle microtubule stability and symmetric/asymmetric cell division (Tokai et al. 1996; Tokai-Nishizumi et al. 2005; Sun and Hevner 2014). Second, KIF22 regulates the expression of the cell-cycle regulator CDC25C. During cell division, CDC25C dephosphorylates CDK1, thus activating the CDK1-cyclinB complex while the CDK1-cyclin B complex phosphorylates CDC25C, causing an amplification loop to drive cells to mitosis (Nilsson and Hoffmann 2000). KIF22 directly transcriptionally

represses CDC25C and inhibits mitosis; this transcriptional repression of CDC25C is dependent on KIF22 being phosphorylated on Thr463 (Ohsugi et al. 2003; Yu et al. 2014). KIF22 depletion in a tumor cell line accelerates the G2/M transition and slows M/G1 transition (Yu et al. 2014).

Overall it therefore appears that KIF22 can act at several different points in the cell cycle making it difficult to predict how increased or decreased dosage of KIF22 in the *16p11.2* microduplication or microdeletion respectively would impact cell cycle in the specific context of cerebral cortex neural progenitors. Our observation that *KIF22* mRNA and KIF22 protein levels both increase during the cell cycle to achieve highest levels in G2/M phase that drop as cells enter G1 phase implies that KIF22 protein does not persist for long after it is translated and is degraded at the end of M-phase suggesting that both transcriptional and post-transcriptional mechanisms regulate its levels. A clear outcome of our study is that KIF22 levels positively correlate with KI67 in neural progenitors and steadily rise as the cell progresses through G1>S>G2>M phases culminating in the maximum level during M-phase. One possibility is that KIF22 is required to reach a threshold for mitosis to occur, after which its levels must decrease sufficiently to allow mitotic exit. Whether cells undertake proliferative or neurogenic divisions is a process heavily controlled by cell cycle length (Borrell and Calegari 2014). Perturbing *KIF22* gene dosage as a consequence of the *16p11.2* CNV might affect the timing of KIF22 protein reaching this threshold in neural progenitors and therefore affect cell-cycle kinetics and perturb neurogenesis and neuronal output. Our results suggest the hypothesis that KIF22 regulates neurogenesis in the human developing cortex through cell-cycle regulation.

ALDOA

The process of brain development requires a vast and consistent supply of energy. Glucose is the predominant energy substrate for the fetal brain (Gustafsson 2009), therefore efficient and controlled glycolysis is essential for normal brain development. ALDOA is required for the fourth step of glycolysis, conversion of fructose 1,6-biphospate to dihyroxyacetone phosphate and gluteraldehyde 3-phosphate. The

metabolic role of cytoplasmic ALDOA is well established, and ALDOA also has other non-glycolytic "moonlighting" roles such as regulating mitochondrial function and cyctoskeleton stability (Orosz et al. 1988; Pagliaro and Taylor 1992; Kao et al. 1999; Jewett and Sibley 2003; Buscaglia et al. 2006). In addition to its cytoplasmic roles, ALDOA has been observed in the nucleus (Mamczur and Dzugaj 2008; Mamczur et al. 2010, 2013) where it has been suggested to impact cell cycle by positively regulating cyclin D1 expression to mediate G1/S progression (Ritterson Lew and Tolan 2012; Fu et al. 2018). Cell-culture studies show ALDOA sub-cellular localisation depends on the availability of energetic substrates, with addition of glucose driving ALDOA protein to the cytoplasm (Mamczur et al. 2013). Therefore, it is likely the primary role for ALDOA is metabolic when cells require, and have available to them, large amounts of energy. The majority of ALDOA studies have used highly abnormal cancer tissue, or artificial cell culture systems in which glycolytic enzymes have been shown to be increased (Ritterson Lew and Tolan 2012; Mamczur et al. 2013; Fu et al. 2018; Pollen et al. 2019). How these observations of ALDOA in a variety of systems relate to its role in human cerebral cortex development is unclear.

Altering ALDOA dosage in the developing brain will likely impact energy metabolism by altering the flow of metabolites through the glycolytic pathway and impacting subsequent pathways which feed on outputs of glycolysis. Disruption to energy metabolism during development has previously been linked to ASD and ADHD (Rash et al. 2018). The offspring of hyperglycaemic mice presented microcephaly, a phenocopy of the microcephaly observed in *16p11.2* microduplication patients (Rash et al. 2018) and disruptions to energy metabolism may contribute to the microcephaly seen in the offspring of Zika infected mothers (Gilbert-Jaramillo et al. 2019). No homozygous null ALDOA patients have been identified suggesting it is essential for life, but patients with changes to ALDOA levels have been identified; one patient with reduced ALDOA activity presented microcephaly (Kreuder et al. 1996) and another presented intellectual disability (Beutler et al. 1973). Of particular interest is the finding of schizophrenia patients with upregulated cortical ALDOA levels (Beasley et al. 2006) and *16p11.2* microduplication is strongly associated with risk of schizophrenia. This information, coupled with our results that ALDOA is expressed in

all cell types, make it clear that any changes to ALDOA dose will perturb energy metabolism at many stages in the brain, impacting its development.

ALDOA is much more abundant in the cytoplasm and we also found no clear relationship between cell body ALDOA levels and cell proliferation status. Nuclear ALDOA has been linked to cell proliferation (Mamczur et al. 2010, 2013; Fu et al. 2018). While we were able to detect nuclear ALDOA we found no clear relationship between ALDOA protein levels and cell proliferation status. While *ALDOA* mRNA levels are higher in proliferating cells compared to non-proliferating cells, quantitative analysis of ALDOA protein failed to identify increased levels in proliferating $KI67^{+}$ cells. This seems unlikely to reflect a technical problem with our quantification method because an identical analysis of KIF22 protein found increased levels of KIF22 protein in $KI67^{+}$ cells (see above). However, our quantification method was restricted to measuring signal in the nucleus and adjacent cytoplasm so may be more appropriate for a nuclear protein (like KIF22) than a protein that fills the whole cell (like ALDOA) and will specifically fail to identify protein localized to the radial fibers of neural progenitors or neurites of post-mitotic cells. Our finding that *ALDOA* mRNA levels do not appear to correlate well with protein levels could therefore have several explanations. One possibility is that *ALDOA* mRNA expression is higher in progenitors and subsequently down-regulated in cells that become post-mitotic but ALDOA protein translated when the cells are progenitors is very stable so persists in post-mitotic cells after *ALDOA* mRNA levels decline. Another possibility is that total ALDOA protein levels are higher in progenitors but the subcellular distribution changes as cells become post-mitotic, for example ALDOA protein in the radial processes of progenitors redistributes towards the nucleus as cells become post mitotic, such that the levels of protein in and around the nucleus remain constant. We cannot distinguish between these possibilities using the fixed post-mortem material available for this study, but it would be possible to test in dissociated cell culture where the whole cell is visible. In any case, it is clear that while ALDOA protein is abundant in progenitor cells of the developing human cerebral cortex and so may play a role in neurogenesis phenotypes, the persistent expression of ALDOA protein as cells become

post-mitotic raises the additional possibility that ALDOA also plays roles in differentiated neurons.

Conclusion

Our study of gene expression in developing human fetal cerebral cortex indicates that further investigation into the normal function of ALDOA and KIF22 and the consequences of altering their dosage on neurogenesis is likely to be a fruitful line of enquiry for understanding the *16p11.2* phenotypes. Further studies are required to unpick the mechanisms involved, but given the nature of the tissue, the scope for studying this *in vivo* is currently limited. However, growth of new model systems in which *16p11.2* gene expression can be manipulated such as human cerebral organoids will provide the opportunity address these questions.

Acknowledgements

This work was supported by a BBSRC grant (BB/M00693X/1) to TP and an EASTBIO BBSRC funded PhD studentship to SM. The human embryonic and fetal material was provided by the Joint MRC / Wellcome (MR/R006237/1) Human Developmental Biology Resource (www.hdbr.org).

We thank Katherine Howe for her assistance with designing *in situ* probes and undergraduate student Emma Fowler (EF funded by a WR Henderson Scholarship) for her assistance with quantification

References

Bayer SA, Altman J. 2002. The Human Brain During the Late First Trimester. Taylor & Francis.

Bayer SA, Altman J. 2005. The Human Brain During The Second Trimester. CRC Press Taylor & Francis Group.

Beasley CL, Pennington K, Behan A, Wait R, Dunn MJ, Cotter D. 2006. Proteomic analysis of the anterior cingulate cortex in the major psychiatric disorders: Evidence for disease-associated changes. Proteomics. 6:3414–3425.

Beutler E, Scott S, Bishop A, Margolis N, Matsumoto F, Kuhl W. 1973. Red cell aldolase deficiency and hemolytic anemia: a new syndrome. Trans Assoc Am Physicians. 86:154–166.

Bijlsma EK, Gijsbers AC, Schuurs-Hoeijmakers JH, van Haeringen A, Fransen van de Putte DE, Anderlid BM, Lundin J, Lapunzina P, Perez Jurado LA, Delle Chiaie B, Loeys B, Menten B, Oostra A, Verhelst H, Amor DJ, Bruno DL, van Essen AJ, Hordijk R, Sikkema-Raddatz B, Verbruggen KT, Jongmans MC, Pfundt R, Reeser HM, Breuning MH, Ruivenkamp CA. 2009. Extending the phenotype of recurrent rearrangements of 16p11.2: deletions in mentally retarded patients without autism and in normal individuals. Eur J Med Genet. 52:77–87.

Blackmon K, Thesen T, Green S, Ben-Avi E, Wang X, Fuchs B, Kuzniecky R, Devinsky O. 2018. Focal Cortical Anomalies and Language Impairment in 16p11.2 Deletion and Duplication Syndrome. Cereb Cortex. 28:2422–2430.

Blaker-Lee A, Gupta S, McCammon JM, De Rienzo G, Sive H. 2012. Zebrafish homologs of genes within 16p11.2, a genomic region associated with brain disorders, are active during brain development, and include two deletion dosage sensor genes. Dis Model Mech. 5:834–851.

Blumenthal I, Ragavendran A, Erdin S, Klei L, Sugathan A, Guide JR, Manavalan P, Zhou JQ, Wheeler VC, Levin JZ, Ernst C, Roeder K, Devlin B, Gusella JF, Talkowski ME. 2014. Transcriptional consequences of 16p11.2 deletion and duplication in mouse cortex and multiplex autism families. Am J Hum Genet.

94:870–883.

Borrell V, Calegari F. 2014. Mechanisms of brain evolution: regulation of neural progenitor cell diversity and cell cycle length. Neurosci Res. 86:14–24.

Buscaglia CA, Penesetti D, Tao M, Nussenzweig V. 2006. Characterization of an aldolase-binding site in the Wiskott-Aldrich syndrome protein. J Biol Chem. 281:1324–1331.

de Anda FC, Rosario AL, Durak O, Tran T, Graff J, Meletis K, Rei D, Soda T, Madabhushi R, Ginty DD, Kolodkin AL, Tsai LH. 2012. Autism spectrum disorder susceptibility gene TAOK2 affects basal dendrite formation in the neocortex. Nat Neurosci. 15:1022–1031.

Fu H, Gao H, Qi X, Zhao L, Wu D, Bai Y, Li H, Liu X, Hu J, Shao S. 2018. Aldolase A promotes proliferation and G1/S transition via the EGFR/MAPK pathway in non-small cell lung cancer. Cancer Commun. 38:18.

Gilbert-Jaramillo J, Garcez P, James W, Molnár Z, Clarke K. 2019. The potential contribution of impaired brain glucose metabolism to congenital Zika syndrome. J Anat.

Girirajan S, Eichler EE. 2010. Phenotypic variability and genetic susceptibility to genomic disorders. Hum Mol Genet. 19:R176–R187.

Golzio C, Willer J, Talkowski ME, Oh EC, Taniguchi Y, Jacquemont S, Reymond A, Sun M, Sawa A, Gusella JF, Kamiya A, Beckmann JS, Katsanis N. 2012. KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. Nature. 485:363–367.

Gustafsson J. 2009. Neonatal energy substrate production. Indian J Med Res. 130:618–623.

Horev G, Ellegood J, Lerch JP, Son YE, Muthuswamy L, Vogel H, Krieger AM, Buja A, Henkelman RM, Wigler M, Mills AA. 2011. Dosage-dependent phenotypes in models of 16p11.2 lesions found in autism. Proc Natl Acad Sci U S A. 108:17076–17081.

Jewett TJ, Sibley LD. 2003. Aldolase forms a bridge between cell surface adhesins and the actin cytoskeleton in apicomplexan parasites. Mol Cell. 11:885–894.

Kao AW, Noda Y, Johnson JH, Pessin JE, Saltiel AR. 1999. Aldolase mediates the association of F-actin with the insulin-responsive glucose transporter GLUT4. J

Biol Chem. 274:17742–17747.

Kreuder J, Borkhardt A, Repp R, Pekrun A, Göttsche B, Gottschalk U, Reichmann H, Schachenmayr W, Schlegel K, Lampert F. 1996. Brief report: inherited metabolic myopathy and hemolysis due to a mutation in aldolase A. N Engl J Med. 334:1100–1104.

Kumar RA, KaraMohamed S, Sudi J, Conrad DF, Brune C, Badner JA, Gilliam TC, Nowak NJ, Cook EH Jr, Dobyns WB, Christian SL. 2008. Recurrent 16p11.2 microdeletions in autism. Hum Mol Genet. 17:628–638.

Levy D, Ronemus M, Yamrom B, Lee YH, Leotta A, Kendall J, Marks S, Lakshmi B, Pai D, Ye K, Buja A, Krieger A, Yoon S, Troge J, Rodgers L, Iossifov I, Wigler M. 2011. Rare de novo and transmitted copy-number variation in autistic spectrum disorders. Neuron. 70:886–897.

Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA. 2015. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell. 161:1202–1214.

Malhotra D, Sebat J. 2012. CNVs: harbingers of a rare variant revolution in psychiatric genetics. Cell. 148:1223–1241.

Mamczur P, Dzugaj A. 2008. Aldolase A is present in smooth muscle cell nuclei. Acta Biochim Pol. 55:799–805.

Mamczur P, Gamian A, Kolodziej J, Dziegiel P, Rakus D. 2013. Nuclear localization of aldolase A correlates with cell proliferation. Biochim Biophys Acta. 1833:2812–2822.

Mamczur P, Mazurek J, Rakus D. 2010. Ubiquitous presence of gluconeogenic regulatory enzyme, fructose-1,6-bisphosphatase, within layers of rat retina. Cell Tissue Res. 341:213–221.

McCarthy SE, Makarov V, Kirov G, Addington AM, McClellan J, Yoon S, Perkins DO, Dickel DE, Kusenda M, Krastoshevsky O, Krause V, Kumar RA, Grozeva D, Malhotra D, Walsh T, Zackai EH, Kaplan P, Ganesh J, Krantz ID, Spinner NB, Roccanova P, Bhandari A, Pavon K, Lakshmi B, Leotta A, Kendall J, Lee Y-H, Vacic V, Gary S, Iakoucheva LM, Crow TJ, Christian SL, Lieberman JA, Stroup TS, Lehtimäki T, Puura K, Haldeman-Englert C, Pearl J, Goodell M,

Willour VL, Derosse P, Steele J, Kassem L, Wolff J, Chitkara N, McMahon FJ, Malhotra AK, Potash JB, Schulze TG, Nöthen MM, Cichon S, Rietschel M, Leibenluft E, Kustanovich V, Lajonchere CM, Sutcliffe JS, Skuse D, Gill M, Gallagher L, Mendell NR, Wellcome Trust Case Control Consortium, Craddock N, Owen MJ, O'Donovan MC, Shaikh TH, Susser E, Delisi LE, Sullivan PF, Deutsch CK, Rapoport J, Levy DL, King M-C, Sebat J. 2009. Microduplications of 16p11.2 are associated with schizophrenia. Nat Genet. 41:1223–1227.

Miller I, Min M, Yang C, Tian C, Gookin S, Carter D, Spencer SL. 2018. Ki67 is a Graded Rather than a Binary Marker of Proliferation versus Quiescence. Cell Rep. 24:1105–1112.e5.

Nilsson I, Hoffmann I. 2000. Cell cycle regulation by the Cdc25 phosphatase family. Prog Cell Cycle Res. 4:107–114.

Nuttle X, Giannuzzi G, Duyzend MH, Schraiber JG, Narvaiza I, Sudmant PH, Penn O, Chiatante G, Malig M, Huddleston J, Benner C, Camponeschi F, Ciofi-Baffoni S, Stessman HAF, Marchetto MCN, Denman L, Harshman L, Baker C, Raja A, Penewit K, Janke N, Tang WJ, Ventura M, Banci L, Antonacci F, Akey JM, Amemiya CT, Gage FH, Reymond A, Eichler EE. 2016. Emergence of a Homo sapiens-specific gene family and chromosome 16p11.2 CNV susceptibility. Nature. 536:205–209.

Ohsugi M, Tokai-Nishizumi N, Shiroguchi K, Toyoshima YY, Inoue J, Yamamoto T. 2003. Cdc2-mediated phosphorylation of Kid controls its distribution to spindle and chromosomes. EMBO J. 22:2091–2103.

Orosz F, Christova TY, Ovádi J. 1988. Modulation of phosphofructokinase action by macromolecular interactions. Quantitative analysis of the phosphofructokinase-aldolase-calmodulin system. Biochim Biophys Acta. 957:293–300.

Pagliaro L, Taylor DL. 1992. 2-Deoxyglucose and cytochalasin D modulate aldolase mobility in living 3T3 cells. J Cell Biol. 118:859–863.

Pollen AA, Bhaduri A, Andrews MG, Nowakowski TJ, Meyerson OS, Mostajo-Radji MA, Di Lullo E, Alvarado B, Bedolli M, Dougherty ML, Fiddes IT, Kronenberg ZN, Shuga J, Leyrat AA, West JA, Bershteyn M, Lowe CB, Pavlovic BJ, Salama SR, Haussler D, Eichler EE, Kriegstein AR. 2019.

Establishing Cerebral Organoids as Models of Human-Specific Brain Evolution. Cell. 176:743–756.e17.

Pollen AA, Nowakowski TJ, Chen J, Retallack H, Sandoval-Espinosa C, Nicholas CR, Shuga J, Liu SJ, Oldham MC, Diaz A, Lim DA, Leyrat AA, West JA, Kriegstein AR. 2015. Molecular identity of human outer radial glia during cortical development. Cell. 163:55–67.

Preibisch S, Saalfeld S, Tomancak P. 2009. Globally optimal stitching of tiled 3D microscopic image acquisitions. Bioinformatics. 25:1463–1465.

Pucilowska J, Vithayathil J, Pagani M, Kelly C, Karlo JC, Robol C, Morella I, Gozzi A, Brambilla R, Landreth GE. 2018. Pharmacological Inhibition of ERK Signaling Rescues Pathophysiology and Behavioral Phenotype Associated with 16p11.2 Chromosomal Deletion in Mice. J Neurosci. 38:6640–6652.

Pucilowska J, Vithayathil J, Tavares EJ, Kelly C, Karlo JC, Landreth GE. 2015. The 16p11.2 deletion mouse model of autism exhibits altered cortical progenitor proliferation and brain cytoarchitecture linked to the ERK MAPK pathway. J Neurosci. 35:3190–3200.

Qiu X, Hill A, Packer J, Lin D, Ma Y-A, Trapnell C. 2017. Single-cell mRNA quantification and differential analysis with Census. Nat Methods. 14:309–315.

Qureshi AY, Mueller S, Snyder AZ, Mukherjee P, Berman JI, Roberts TP, Nagarajan SS, Spiro JE, Chung WK, Sherr EH, Buckner RL. 2014. Opposing brain differences in 16p11.2 deletion and duplication carriers. J Neurosci. 34:11199–11211.

Radonjić NV, Ortega JA, Memi F, Dionne K, Jakovcevski I, Zecevic N. 2014. The complexity of the calretinin-expressing progenitors in the human cerebral cortex. Front Neuroanat. 8:82.

Rash BG, Micali N, Huttner AJ, Morozov YM, Horvath TL, Rakic P. 2018. Metabolic regulation and glucose sensitivity of cortical radial glial cells. Proc Natl Acad Sci U S A. 115:10142–10147.

Ritterson Lew C, Tolan DR. 2012. Targeting of several glycolytic enzymes using RNA interference reveals aldolase affects cancer cell proliferation through a non-glycolytic mechanism. J Biol Chem. 287:42554–42563.

Rosenfeld JA, Coppinger J, Bejjani BA, Girirajan S, Eichler EE, Shaffer LG, Ballif

BC. 2010. Speech delays and behavioral problems are the predominant features in individuals with developmental delays and 16p11.2 microdeletions and microduplications. J Neurodev Disord. 2:26–38.

Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, Chu SH, Moreau MP, Gupta AR, Thomson SA, Mason CE, Bilguvar K, Celestino-Soper PB, Choi M, Crawford EL, Davis L, Wright NR, Dhodapkar RM, DiCola M, DiLullo NM, Fernandez TV, Fielding-Singh V, Fishman DO, Frahm S, Garagaloyan R, Goh GS, Kammela S, Klei L, Lowe JK, Lund SC, McGrew AD, Meyer KA, Moffat WJ, Murdoch JD, O'Roak BJ, Ober GT, Pottenger RS, Raubeson MJ, Song Y, Wang Q, Yaspan BL, Yu TW, Yurkiewicz IR, Beaudet AL, Cantor RM, Curland M, Grice DE, Gunel M, Lifton RP, Mane SM, Martin DM, Shaw CA, Sheldon M, Tischfield JA, Walsh CA, Morrow EM, Ledbetter DH, Fombonne E, Lord C, Martin CL, Brooks AI, Sutcliffe JS, Cook EH Jr, Geschwind D, Roeder K, Devlin B, State, M. W. 2011. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron. 70:863–885.

Scholzen T, Gerdes J. 2000. The Ki-67 protein: from the known and the unknown. J Cell Physiol. 182:311–322.

Shinawi M, Liu P, Kang SH, Shen J, Belmont JW, Scott DA, Probst FJ, Craigen WJ, Graham BH, Pursley A, Clark G, Lee J, Proud M, Stocco A, Rodriguez DL, Kozel BA, Sparagana S, Roeder ER, McGrew SG, Kurczynski TW, Allison LJ, Amato S, Savage S, Patel A, Stankiewicz P, Beaudet AL, Cheung SW, Lupski JR. 2010. Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size. J Med Genet. 47:332–341.

Sun T, Hevner RF. 2014. Growth and folding of the mammalian cerebral cortex: from molecules to malformations. Nat Rev Neurosci. 15:217–232.

Tai DJC, Ragavendran A, Manavalan P, Stortchevoi A, Seabra CM, Erdin S, Collins RL, Blumenthal I, Chen X, Shen Y, Sahin M, Zhang C, Lee C, Gusella JF, Talkowski ME. 2016. Engineering microdeletions and microduplications by targeting segmental duplications with CRISPR. Nat Neurosci. 19:517–522.

Tirosh I, Izar B, Prakadan SM, Wadsworth MH 2nd, Treacy D, Trombetta JJ, Rotem A, Rodman C, Lian C, Murphy G, Fallahi-Sichani M, Dutton-Regester K, Lin J-R, Cohen O, Shah P, Lu D, Genshaft AS, Hughes TK, Ziegler CGK, Kazer SW, Gaillard A, Kolb KE, Villani A-C, Johannessen CM, Andreev AY, Van Allen EM, Bertagnolli M, Sorger PK, Sullivan RJ, Flaherty KT, Frederick DT, Jané-Valbuena J, Yoon CH, Rozenblatt-Rosen O, Shalek AK, Regev A, Garraway LA. 2016. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science. 352:189–196.

Tokai N, Fujimoto-Nishiyama A, Toyoshima Y, Yonemura S, Tsukita S, Inoue J, Yamamota T. 1996. Kid, a novel kinesin-like DNA binding protein, is localized to chromosomes and the mitotic spindle. EMBO J. 15:457–467.

Tokai-Nishizumi N, Ohsugi M, Suzuki E, Yamamoto T. 2005. The chromokinesin Kid is required for maintenance of proper metaphase spindle size. Mol Biol Cell. 16:5455–5463.

Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL. 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol. 32:381–386.

Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, Saemundsen E, Stefansson H, Ferreira MA, Green T, Platt OS, Ruderfer DM, Walsh CA, Altshuler D, Chakravarti A, Tanzi RE, Stefansson K, Santangelo SL, Gusella JF, Sklar P, Wu BL, Daly MJ. 2008. Association between microdeletion and microduplication at 16p11.2 and autism. N Engl J Med. 358:667–675.

Yu Y, Wang XY, Sun L, Wang YL, Wan YF, Li XQ, Feng YM. 2014. Inhibition of KIF22 suppresses cancer cell proliferation by delaying mitotic exit through upregulating CDC25C expression. Carcinogenesis. 35:1416–1425.

Zufferey F, Sherr EH, Beckmann ND, Hanson E, Maillard AM, Hippolyte L, Mace A, Ferrari C, Kutalik Z, Andrieux J, Aylward E, Barker M, Bernier R, Bouquillon S, Conus P, Delobel B, Faucett WA, Goin-Kochel RP, Grant E, Harewood L, Hunter JV, Lebon S, Ledbetter DH, Martin CL, Mannik K, Martinet D, Mukherjee P, Ramocki MB, Spence SJ, Steinman KJ, Tjernagel J, Spiro JE, Reymond A, Beckmann JS, Chung WK, Jacquemont S. 2012. A 600

kb deletion syndrome at 16p11.2 leads to energy imbalance and neuropsychiatric disorders. J Med Genet. 49:660–668.

.