Molecular mechanisms of PCDH19 regulation of neurogenesis

A thesis submitted to Cardiff University in accordance with the requirements for the degree of Doctor of Philosophy in the discipline of Neuroscience.

Ian William John Fox

Supervisor: Dr. Isabel Martinez-Garay

Co-supervisor: Dr. Kerrie Thomas

February 2023

The works of the LORD are great, sought out of all them that have pleasure therein.

Psalm 111:2 KJV

Acknowledgements

I am extremely fortunate to have been helped and guided by a multitude of kind and generous people throughout my PhD experience. This thesis is the result of their support, wisdom, and encouragement.

First and foremost, I would like to thank my supervisor, Dr. Isabel Martinez-Garay. A scholar, mentor, and friend – your knowledge and experience has been instrumental in supporting me throughout this project. I am incredibly grateful for your insight and encouragement, and I consider it a privilege to have worked under your supervision.

I would also like to thank Professor Yves-Alain Barde and Dr. David Petrik, to whom I am grateful for all of the advice and valuable input you gave me during our common lab meetings. By extension, I would also like to thank Dr. Francesco Bedogni and Dr. Erik Mire for their comments and thoughts during our cortical development lab meetings.

This project was funded by the Wellcome Trust, without which this research would not have been possible. Thank you as well to the BBSRC and PCDH19 Alliance for the funding they have provided to the lab.

There have been many people that have helped me complete the research that led to this thesis, and it is my pleasure to thank all of them. Thank you to Dr. Xinsheng Nan, your knowledge and expertise in the lab is tremendously appreciated. Thank you to Angela Marchbank and Dr. Sumukh Deshpande for your help in preparing and studying the RNA-seq data. A special thank you to our collaborators Dr. Sergi Roig-Puiggros and Dr. Natalia Baumann for your marvellous help in analysing the single cell RNA-seq data; the detail that we were able to achieve was not possible without you. I would also like to thank everyone from the Pheese lab – Dr. Toby Pheese, Dr. Valerie Meniel, and Dr. Giusy Tornillo – for your help in troubleshooting the luciferase experiments. Thank you to all of the current, past, and adopted members of the BIOSCI Floor 2 labs: Ms. Eleanor Lewis, Ms. Sara Jörgensen, Dr. Tetyana Tsugorka, Dr. Andrew Want, Dr. Gloria Cimaglia, Dr. Sarah Ateaque, and Dr. Erin Wosnitzka for your advice, support, and friendship. A special thank you as well to the past members of the IMG lab: Dr. Cristina Llinares-Benadero, Dr. Sylvia Newbold, Dr. Jessica Griffiths, Dr. Natalia Galindo-Riera, and Dr. Monika Sledziowska. Cris and Sylvia – thank you for the many adventures, both in the lab and abroad. A special thank you to you, Cris, for your friendship, patience, and mentorship throughout this PhD; you've been the greatest lab partner. And to my brilliant students Ruhani Makkar and John McLarnon, thank you for the laughs, for listening, and for everything you did in the lab. I am proud of both of you.

Thank you to my supporting family for everything you have done for me. To my younger sister, Alice, for your endless encouragement and for being the best of friends. To my father, Keith, for supporting my ambitions and for always being there for me. To my cousins, Matthew, Gwenan, Bethan, Rob, Joe, Frances, Julian, Henry, Capucine, Cameron, and Rachel – you're the best family a man could ask for. To Grandad and Nana, for everything you did for me and Alice from our first moments on the Earth. You will always be in my prayers. To my mother, Ann, for always believing in me. To Ellie, without you I think none of this would have been possible. Your patience and love keep me upright and disciplined; I love you very much.

Finally, I want to thank God for giving me the ability to make a difference.

Statement of contributions

I would like to acknowledge and express my gratitude to the following individuals who have provided significant contributions and support throughout the completion of my thesis.

1. Dr. Isabel Martinez-Garay

Dr. Isabel Martinez-Garay served as the primary supervisor and provided guidance throughout the entire project. Dr. Martinez-Garay had a crucial role in shaping the research objectives and formulating the research methodology. She also provided valuable insights in data analysis, interpretation, and the formulation of conclusions. Dr. Martinez-Garay also offered valuable feedback and suggestions for improving the thesis content, structure, and overall quality. I am grateful for the guidance and support she has provided me over the years.

2. Dr. Cristina Llinares-Benadero

Dr. Cristina Llinares-Benadero was the post-doctorate researcher of the lab and worked on understanding the cellular mechanisms of PCDH19 regulation of neurogenesis. Therefore, there was a close collaboration between Dr. Llinares-Benadero and myself, as both projects were strongly related. Dr. Llinares-Benadero provided valuable insights in data analysis and discussion of results. She also provided much of my initial training in the lab, and especially in the teaching of *in utero* electroporation, immunohistochemistry, perfusion, confocal microscopy, and brain dissection. She also assisted in developing the experimental protocol for the dissection technique of the embryonic samples, for both the bulk RNA-seq and single cell (sc)RNA-seq experiments. Dr. Llinares-Benadero also had a crucial role in the scRNA-seq experiment by performing the genotyping of samples while I conducted the sample dissociation process. This collaboration was of utmost importance due to time constraints, as it was necessary to determine the genotypes before completing the dissociation. Cristina's expertise and dedication significantly contributed to the successful completion of this time-sensitive task. I am grateful for her valuable assistance, which played a pivotal role in the timely progress of my research.

3. Dr. Angela Marchbank

Dr. Angela Marchbank is a senior researcher at the Genomics Hub facility at the School of Biosciences, Cardiff University. Dr. Marchbank’s expertise on genomic and transcriptomic research was invaluable in the shaping of the bulk and scRNA-seq experiments. Dr. Marchbank offered in-depth advice on the structural approach of both experiments. She also contributed significantly to the initial handling steps of the RNA samples by testing the quantity and quality of

the RNA. Dr. Marchbank also performed the cDNA conversion, fragmentation, adaptor ligation, PCR amplification, and sequencing of the samples for the bulk RNA-seq experiment. Dr. Marchbank and myself also worked together on the initial phases of the scRNA-seq protocol, including the cell partitioning phase and cDNA conversion. From there, Dr. Marchbank performed the remaining steps, including sample cleaning, fragmentation, adaptor ligation, PCR amplification, sequencing, and provided me with the raw data files. These services were offered by the Genomics Hub at Cardiff University, and I am very grateful to all members of the facility and Dr. Marchbank for their assistance.

4. Dr. Sumukh Deshpande

Dr. Sumukh Deshpande is a senior bioinformatician at Cardiff University. Dr. Deshpande was hired to help with the processing and analysis of the scRNA-seq data – as outlined in the BBSRC grant proposal – which included collaborating with an experienced bioinformatician for this part of the project. Dr. Deshpande's expertise and dedication played a pivotal role in the successful processing of the scRNA-seq data, following the Cell Ranger pipeline and all associated steps. He executed the initial data analysis workflow, ensuring the accuracy and reliability of the results. His meticulous approach in preprocessing, quality control, and alignment of the scRNA-seq data was commendable. After processing the data, Dr. Deshpande and I worked together for a short time to analyse the results in R. I am grateful for Dr. Deshpande's commitment to the project and for his expertise in the Cell Ranger pipeline. His contribution significantly enhanced our ability to explore and analyse the data within the time constraints of the project.

5. Prof. Denis Jabaudon, Dr. Sergi Roig-Puiggros, and Dr. Natalia Baumann

Prof. Denis Jabaudon is a professor at the Neurocenter in Geneva University. Dr. Sergi Roig-Puiggros and Dr. Natalia Baumann are post doctorate researchers in the Jabaudon lab. Our collaboration with the Jabaudon lab began after we reached out to them and explained that we had no prior experience in scRNA-seq analysis and would greatly benefit from their expertise. Their combined experience in scRNA-seq analysis was invaluable to us during the initial exploration of the scRNA-seq data. Dr. Roig-Puiggros' expertise and guidance were instrumental in navigating the complex data landscape. Dr. Roig-Puiggros and I analysed the data together and he provided extensive training for me in scRNA-seq analysis. His insightful suggestions and recommendations paved the way for meaningful data exploration and analysis. His wealth of knowledge and experience significantly enriched the early stages of my research, allowing for a comprehensive understanding of the dataset. Dr. Baumann's proactive involvement and support

were invaluable during the initial data exploration phases. Her keen attention to detail and analytical skills contributed to uncovering key patterns and insights when Dr. Roig-Puiggros and I would reach a stumbling block. Her collaboration and enthusiasm for the project enhanced the depth and rigor of the data exploration process. The training I received from Dr. Roig-Puiggros and Dr. Baumann was instrumental in the completion of the scRNA-seq project. Additionally, both researchers shared their machine learning models for us to assess whether suitable differences in neuron differentiation rates could be detected in the dataset. This was an algorithm they had been working closely on for years and I am thankful they were willing to share it with me and provide me with experience on more advanced scRNA-seq analysis techniques. This collaboration was no small feat for them, and I am very grateful for their assistance, dedication, and expertise.

6. Dr. Toby Pheese, Dr. Valerie Meniel, and Dr. Giusy Tornillo

Dr. Toby Pheese is a research associate at Cardiff University. Dr. Valerie Meniel and Dr. Giusy Tornillo are post doctorate researchers working in Dr. Pheese's lab. Their expertise in the Wnt signalling pathway was instrumental for us to design and structure our experiments involving Wnt signalling. Dr. Meniel and Dr. Tornillo also provided training for conducting luciferase experiments, offered advice on optimising these experiments, and the best approaches for analysing the results. I am very grateful for their contribution to this thesis and for helping to establish the luciferase assay technique in our lab.

7. Members of the Martinez-Garay, Barde, Mire, Bedogni, and Petrik labs

I am also grateful for the contributions made by the past and present members of the Martinez-Garay, Barde, Mire, Bedogni, and Petrik labs. I am grateful for the insightful comments I received during lab meetings which had a meaningful impact on the direction of the project as well as the overall quality of the research. I am also grateful for each interaction, discussion, and suggestion, which overall contributed to the success of this project.

Abstract

The neurons that compose the mammalian cerebral cortex are born from a seemingly uniform population of progenitor cells through a process called “cortical neurogenesis”. This process is strictly regulated to ensure progenitors produce neurons at the right time and in correct numbers, and defects in neurogenesis are linked to neurological disorders. Mutations in the X-linked gene protocadherin-19 (*PCDH19*) lead to PCDH19-epilepsy, a disorder that causes early onset epilepsy and cognitive impairment. Although the disease mechanism is unknown, PCDH19 tissue mosaicism is thought to be a crucial underlying factor. Recent work at the Isabel Martinez-Garay (IMG) lab has suggested that the presence of PCDH19 in neural progenitor cells may be important for regulating cortical neurogenesis. Progenitors that expressed either the wildtype (WT) or knockout (KO) PCDH19 allele showed delayed and premature neurogenesis, respectively. Interestingly, this only occurred in the developing brain of heterozygous (HET) mice, since homozygous WT and KO animals showed normal neurogenesis rates. This thesis aimed to examine the molecular mechanisms by which PCDH19 may regulate cortical neurogenesis.

To identify molecular hallmarks by which PCDH19 may regulate neurogenesis, bulk RNA-seq and single cell (sc)RNA-seq were performed on cortical tissue from WT, HET, and KO embryonic day (E)11 animals. The bulk RNA-seq initially found few differentially expressed (DE) genes between mutant and WT same sex counterparts. However, enrichment analyses suggested several pathways were upregulated in HET and KO animals, including pathways related to ribosomes, BMP signalling, and Wnt signalling. The results from the scRNA-seq analysis revealed numerous DE genes between mutants and WTs when examined within cell clusters of interest (progenitors and excitatory neurons), including genes related to neurogenesis, Wnt signalling, and Shh signalling. Interestingly, HETs also displayed a marked reduction in ribosomal gene expression compared to WTs and KOs. However, one of the most striking findings from this analysis was that despite WT and KO cells within the HET animals displaying altered neurogenesis rates, no DE genes were uncovered between these cells when compared to each other, suggesting they are transcriptionally similar. Data was also generated that revealed a role of PCDH19 in negatively modulating Wnt signalling, potentially through an interaction with β-catenin and the noncanonical Wnt receptor, RYK. Moreover, immunostaining using pLRP6 antibodies revealed no differences in the proportion of pLRP6 cells in HETs, however KO animals showed a slightly increased proportion. Taken together, this thesis provides the first examination of the molecular mechanisms of PCDH19 in cortical development. The transcriptional data generated will help direct future work to unravel the role of PCDH19 as a regulator of neurogenesis.

List of abbreviations

-	Minus
%	Percentage
+	Plus
<	Less than
>	Greater than
°C	Degrees Celsius
ADAM10	A Disintegrin and metalloproteinase domain-containing protein 10
AJ	Adherens junctions
AKR1	Aldo-Keto Reductase Family 1
ANOVA	Analysis of variance
APCDD1	adenomatosis polyposis down-regulated 1
ASC1	Asc-type amino acid transporter 1
ASD	Autism spectrum disorder
AU	Arbitrary units
bHLH	Basic helix-loop-helix
BMP	Bone morphogenetic protein
BMPR	Bone morphogenetic protein receptor
bp	Base pair
BrdU	Bromodeoxyuridine
CA	Cornu Ammonis
CaCl2	Calcium chloride
CDH	Cadherin
CDK	Cyclin-dependent kinase
cDNA	complementary DNA
CK	Cyclin kinase
CM	Cytoplasmic motif
CNS	Central nervous system
CO2	Carbon dioxide
CoIP	Co-immunoprecipitation
CP	Cortical plate
CR	Cajal-Retzius

CRISPR	Clustered regularly interspaced short palindromic repeats
CTF	Cytoplasmic fragment
CTIP2	COUP-TF-interacting protein 2
CTNNB1	Catenin Beta 1
CYFIP2	Cytoplasmic FMR1-interacting protein 2
DAPI	4′,6-diamidino-2-phenylindole
ddH20	Double distilled water
DDX3X	DEAD-box helicase 3 X-linked
DE	Differentially expressed
DEPC	Diethyl pyrocarbonate
DG	Dentate gyrus
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DTT	Dithiothreitol
DVL	Dishevelled
E	Embryonic
EC	Extracellular repeat
EDTA	Ethylenediaminetetraacetic acid
EdU	5-ethynyl-2'-deoxyuridine
EIEE9	Early Infantile Epileptic Encephalopathy 9
EtOH	Ethanol
FBS	Fetal bovine serum
FGF	Fibroblast growth factor
FGFR	Fibroblast growth factor receptor
FI	Fluorescence intensity
FL	Full length
FMR1	Fragile X messenger ribonucleoprotein 1
FZD	Frizzled
GABA	γ-aminobutyric acid
GC	Guanine Cytosine
gDNA	Genomic DNA
GFP	Green fluorescent protein
GLAST	Glutamate aspartate transporter

GLI	Glioma-associated oncogene
GO	Gene ontology
GSEA	Gene set enrichment analysis
GSK3	Glycogen synthase kinase-3
HA	Human influenza hemagglutinin
HEK293	Human embryonic kidney 293
HES	Hairy and enhancer of split
HET	Heterozygous
HMGA2	High mobility group protein 2
HRP	Horseradish peroxidase
HVG	High variable genes
ICC	Immunocytochemistry
ICD	Intracellular domain
IHC	Immunohistochemistry
IMG	Isabel Martinez-Garay
INM	Interkinetic nuclear migration
IP	Intermediate progenitor
iPSC	Induced pluripotent stem cell
ISH	In situ hybridisation
IUE	In utero electroporation
IZ	Intermediate zone
JAG1	Jagged canonical notch ligand 1
kb	Kilobase
KCL	Potassium chloride
KD	Knockdown
KDM5D	Lysine demethylase 5D
KEGG	Kyoto Encyclopedia of Genes and Genomes
KH2PO4	Potassium dihydrogen phosphate
KI67	Marker Of Proliferation Ki-67
KO	Knockout
KPNB1	Importin subunit beta-1
LB	Luria-Bertani
LEF1	Lymphoid enhancer binding factor 1
LFC	Log fold change

LRP	LDL receptor related protein 6
LSD1	Lysine-specific demethylase 1
M	Molar
MADM	Mosaic analysis with double markers
MAPK	Mitogen-activated protein kinases
MAZ	Multipolar-accumulation zone
mESC	Mouse embryonic stem cells
MgCl2	Magnesium chloride
miRNA	Micro RNA
mM	Millimolar
mRNA	Messenger RNA
MZ	Marginal zone
NA2HPO4	Disodium phosphate
NaCl	Sodium chloride
NAP1	Nucleosome assembly protein 1
NDE1	NudE neurodevelopment protein 1
NE	Neuroepithelial
NEDD1	Neural precursor cell expressed, developmentally down-regulated 1
NES	Nestin
NEUROG2	Neurogenin-2
ng	Nanogram
NGN1	Neurogenin 1
NGS	Next generation sequencing
NICD	Notch inctracellular domain
NKX2	NK2 homeobox 1
nM	Nanomolar
NMDA	N-methyl-D-aspartate receptor
NONO	Non-POU domain containing octamer binding
OCT	Optimal cutting temperature
OTX2	Orthodenticle homeobox 2
PAR3	Partitioning defective protein 3
PAX6	Paired box protein 6
PBS	Phosphate buffered saline

PC	Principal component
PCA	Principal component analysis
PCDH	Protocadherin
PCDH19-GCE	Protocadherin 19 girls clustered epilepsy
PCR	Polymerse chain reaction
PFA	Paraformaldehyde
pHH3	Phosphor-histone H3
PP	Preplate
PTCH1	Patched 1
PTP1	Protein tyrosine phosphatase PTP1
QC	Quality control
qPCR	Quantitative PCR
RGC	Radial glial cell
RIN	RNA integrity number
RLU	Relative luminescence units
RNA	Ribonucleic acid
RNA-seq	RNA-sequencing
RP	Ribosomal protein
rpm	Rotations per minute
rRNA	Ribosomal RNA
RSPO	R-spondin
RT	Room temperature
RYK	Receptor like tyrosine kinase
SALL1	Spalt like transcription factor 1
SATB2	SATB Homeobox 2
scRNA-seq	Single cell RNA sequencing
SDS PAGE	Sodium dodecyl-sulfate polyacrylamide gel electrophoresis
SHH	Sonic hedgehog
SHISA2	Shisa family member 2
SMAD	Mothers against decapentaplegic homolog 1
SMEK	Suppressor of MEK
SNP	Single nucleotide polymorphism
SOX	SRY-related HMG-box
SVZ	Subventricular zone

TAF1	TATA-box binding protein associated factor 1
TBR	T-box brain transcription factor
TCF	T-cell factor
TF	Transcription factor
TOPO	Topoisomerase based cloning
TPPP3	Tubulin polymerization promoting protein family member 3
TSIX	X-inactive specific transcript (antisense)
UMAP	Uniform manifold approximation and projection
UTR	Untranslated region
UTY	Ubiquitously transcribed tetratricopeptide repeat containing, Y-linked
vs.	Versus
VZ	Ventricular zone
WAVE	WASP family verprolin-homologous protein
WB	Western blot
WIRS	WAVE interacting regulatory sequence
WNT	Wingless-related integration site
WNT/STOP	WNT stablisation of proteins
WRC	WAVE regulatory sequence
WT	Wildtype
XCI	X chromosome inactivation
XIST	X-inactive specific transcript (sense)
α	Alpha
β	Beta
β-GAL	Beta-galactosidase
γ	Gamma
δ	Delta
ε	Epsilon
µg	Microgram
µl	Microliter
µM	Micromolar
µm	Micrometer

Table of contents

List of figures

List of tables

List of Appendices

Chapter 1: Introduction

1.1. Cell adhesion and the cadherin superfamily

Cell-to-cell contact and communication are vital processes that underlie the correct development and function of all tissues in multicellular organisms. Cell adhesion is mediated by numerous families of cell adhesion molecules, including the cadherin superfamily. The cadherins comprise of over 110 transmembrane glycoproteins that were originally described by Takeichi (1988) as cell-surface molecules that mediate calcium-dependent cell-cell adhesion, coining the term “cadherin” as a portmanteau of “calcium-adherens”.

The molecular structure of cadherins typically involves the calcium-binding extracellular domain, a transmembrane domain, and an intracellular domain (Hirano and Takeichi, 2012). The extracellular domain of cadherins is characterised by five or more repeated motifs of ~110 amino acids, called extracellular cadherin (EC) repeats that are linked via Ca^{2+} binding motifs. The extracellular domain is remarkably conserved between cadherin members and mediates primarily homophilic cell-cell interactions, although heterophilic interactions also occur (Hulpiau and van Roy, 2009). The intracellular domain has a primary role in anchoring the cadherins to the underlying cytoskeleton and modulating cytoskeletal dynamics in response to extracellular contacts. The intracellular domains of many cadherins have also been shown to have an active role in various cytoplasmic signal transduction events, the extent of which is slowly being unravelled (Shapiro and Weis, 2009). The cadherin superfamily is divided into several sub-categories, however a consensus on the exact categorisation has yet to be completed due in part to the functional diversity of some cadherin members, which are thus categorised differently between publications (Hirano and Takeichi, 2012; Hulpiau and van Roy, 2009). Nevertheless, members of the cadherin superfamily can be divided into three broad phylogenetic groups: classical cadherins, protocadherins, and desomosomal cadherins (Hulpiau and van Roy, 2009) (see Figure 1.1).

Many cadherins and especially protocadherins are widely expressed throughout the developing and mature nervous system (Kim et al., 2007; Redies and Takeichi, 1993; Zou et al., 2007). These molecules have been shown to play a vital role in organising various events during almost all stages in brain development, including progenitor proliferation, neurogenesis, migration, axon elongation, synaptogenesis, and synaptic plasticity (Hirano and Takeichi, 2012; Punovuori et al., 2021). Therefore, cadherin and protocadherin function likely underlie much of the molecular and cellular complexity of the central nervous system.

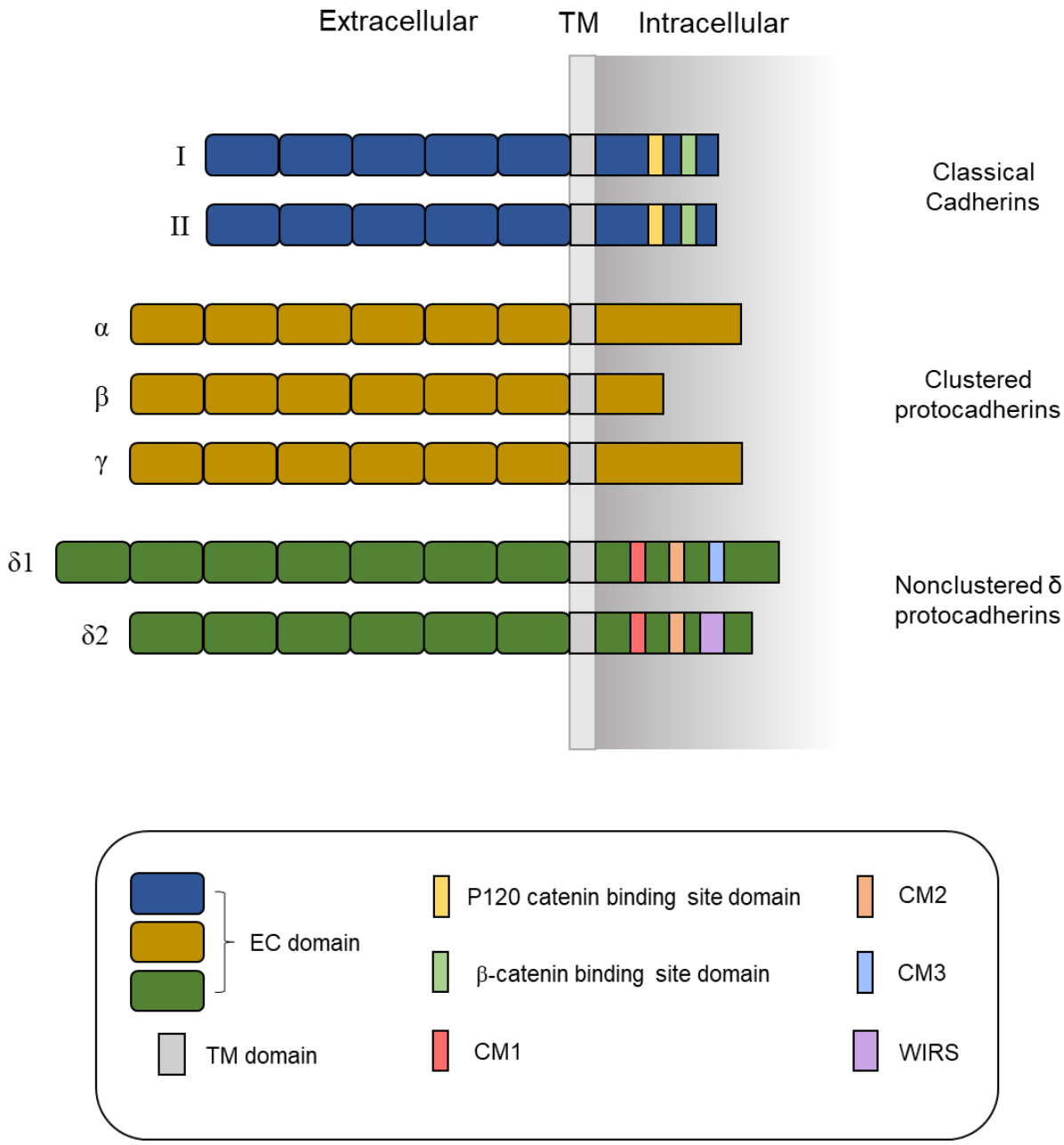


Figure 1.1. Structural representation of the main cadherin subfamilies (adapted from Hayashi and Takeichi, 2015). A representation of the structural features of the main cadherins of the cadherin superfamily, namely the classical cadherins, clustered protocadherins, and nonclustered protocadherins. The conserved cytoplasmic domains of each cadherin group are highlighted in the black box below.

1.1.1. Classical cadherins

The classical cadherin (CDH) subfamily is further divided into two subcategories, namely type I (CDH1 (E-cadherin), CDH2 (N-cadherin), CDH3 (P-cadherin), CDH4 (R-cadherin), CDH15 (M-cadherin)) and type II (CDH5 (VE-cadherin), CDH5 (K-cadherin), CDH6, CDH7, CDH8, CDH9 (T1-cadherin), CDH10 (T2-cadherin), CDH11 (OB-cadherin), CDH12 (N2-cadherin), CDH18, CDH1, CDH20, CDH22, CDH24) (see Figure 1.1). The division is based on slight differences in sequence homology, especially in the intracellular domain (Sotomayor et al., 2014). The strand-swap interface on the EC1 domain of type I and type II classical cadherins is also different, wherein type II classical cadherins have two conserved tryptophan side chain whereas type I cadherins only have one (Patel et al., 2006). Both type I and type II cadherins harbour five EC domains and mediate cell-cell adhesion via a strand-swap mechanism wherein two cadherins create a *trans*-interaction by forming an X-shaped *trans*-dimer near the EC1-EC2 domains (Harrison et al., 2010; Zhang et al., 2009). Although slight variations in the sequence of the intracellular domain define type I and type II groups, the intracellular domain of all classical cadherins show a relatively conserved function. For example, classic cadherins are able to interact with various cytoplasmic catenins, including α-catenin, β-catenin, and p120-catenin (Nelson, 2008). β-catenin and p120-catenin can interact directly with cadherins, whereas α-catenin interacts indirectly via β-catenin (Pokutta et al., 2008). The cadherin-catenin complex is important for many functions, especially to link cadherins to the actin cytoskeleton and modulate cytoskeletal dynamics in response to extracellular cues.

Cadherins have been found to play essential roles in the development of the nervous system. For example, during early development, the neural tube switches expression of E-cadherin to N-cadherin which is necessary for neurulation (Hatta and Takeichi, 1986). N-cadherin and R-cadherin, along with the aforementioned catenins, also form the adherens junctions (AJs) between neuroepithelial cells which are integral for maintaining the structural integrity and morphogenesis of the neural tube and later embryonic brain (Gänzler-Odenthal and Redies, 1998). In addition to their roles in maintaining tissue architecture, many cadherins – especially N-cadherin – have been associated with regulating various signalling pathways that are important for brain development (Tanriover et al., 2004; Yamagata et al., 2020). The details of this are discussed later in this chapter.

1.1.2. Protocadherins

Protocadherins (PCDH) form largest subfamily within the cadherin superfamily and are further divided into two subgroups: clustered and nonclustered protocadherins (Hayashi and Takeichi, 2015). Collectively, clustered and nonclustered protocadherins are structurally very similar to cadherins, however they contain either six or seven EC repeated instead of five (Sano et al., 1993) (see Figure 1.1). The extracellular domain of protocadherins has overall different adhesive properties compared to cadherins (Hayashi and Takeichi, 2015). It is generally considered that protocadherins form much weaker homophilic bonds compared to classical cadherins, and it is theorised that protocadherins operate mostly by forming *cis*-interactions with other cadherin/protocadherins and thus regulate adhesion by forming multi-cadherin complexes (Harrison et al., 2020; Kim et al., 2011; Rubinstein et al., 2017). Moreover, the intracellular domain of protocadherins typically lacks the binding sites for β-catenin and p120-catenin (Hayashi and Takeichi, 2015). In addition to clustered and nonclustered protocadherins, there are also the seven-pass transmembrane cadherins which include the flamingo/CELRS protocadherins, Fat-like protocadherins, and others. However, for simplicity the clustered and nonclustered protocadherins will only be focused on in the sections below.

1.1.3. Clustered protocadherins

The clustered protocadherins are organised into three subgroups: *Pcdhα, Pcdhβ*, and *Pcdhγ.* These three groups are arranged in tandem i.e., a “cluster” on a single chromosome, namely chromosome 5q31 in humans and chromosome 18 in mice (Wu and Maniatis, 1999). Interestingly, most clustered protocadherins exhibit variable exons that encode for the six EC extracellular domain, the transmembrane domain, and a portion of the intracellular domain. *Pcdhα* harbours 14 variable exons, whereas *Pcdhβ* and *Pcdhγ* harbour 22. Each variable exon contains a specific promoter that is regulated by various epigenetic mechanisms which allow for stochastic expression and the production of ~60 different isoforms (Canzio and Maniatis, 2019; Chen and Maniatis, 2013). *Pcdhα* and *Pcdhγ* also exhibit small constant regions that encode for a portion of the intracellular domain, however *Pcdhβ* does not have any constant regions (Hayashi and Takeichi, 2015; Wu et al., 2001).

The last two and last three variable exons of the *Pcdhα* and *Pcdhγ* gene clusters (*αC1, αC2, γC3, γC4, γC5*) vary slightly in sequence compared to the other clustered protocadherins and are referred to as C-type clustered protocadherins (Wu et al., 2001). These C-type protocadherins are expressed constitutively throughout the central nervous system, however all the other

isoforms show sparse and stochastic expression. For example, Purkinje cells express ~10 alternate clustered protocadherins in a stochastic fashion, but also express all five C-type protocadherins constitutively (Esumi et al., 2005). Serotonergic neurons express just *αC*-type protocadherins constitutively, in addition to alternative expression of the other clustered protocadherins (Canzio and Maniatis, 2019; Chen et al., 2017). The combination of random and ubiquitous expression of a diverse range of protocadherin isoforms has been theorised to provide neurons with a unique molecular "barcode" (Rubinstein et al., 2015, 2017). The clustered protocadherins form complex *cis*-dimeric units on the cell membrane that can be used to make *trans*-interactions with other units and form zipper-like assemblies. The *cis*-dimeric complexes are assembled via EC6 whereas the *trans*-interactions are mediated by EC1-4 in a "head-to-tail" conformation (Goodman et al., 2016, 2017; Rubinstein et al., 2015). It is believed that these zipper-like assemblies are necessary for neuronal self-recognition, triggering downstream signalling pathways that leads to self-avoidance (Fan et al., 2018). Recent X-ray crystallography and electron microscopy studies have shown full length clustered protocadherin extracellular domains forming linear zipper structures (Brasch et al., 2019). Moreover, the zipper assemblies have been shown to form remarkably precise *trans*-interactions with no evidence of heterophilic binding occurring between different isoforms (Goodman et al., 2022). Therefore, clustered protocadherins are an incredibly diverse array of proteins that provide neurons with a molecular signature that likely underlies much of the complexity involving neuron connectivity and positioning.

1.1.4. Nonclustered protocadherins

Contrary to clustered protocadherins, the nonclustered protocadherins are scattered randomly throughout the genome and are organised into three subgroups: δ1, δ2, and ε (Kim et al., 2011). δ1-protocadherins (protocadherin-1, 7, 9, and 11) contain seven EC repeats and three conserved motifs in the cytoplasmic domain, whereas the δ2-protocadherins (protocadherin-8, 10, 17, 18, and 19) contain six EC repeats and only two of the conserved cytoplasmic motifs (Hulpiau and van Roy, 2009; Redies et al., 2005; Vanhalst et al., 2005). The intracellular domain of δ2-protocadherins also contains a WAVE interacting regulatory sequence (WIRS) that binds to the WAVE complex, which regulates actin and cytoskeletal dynamics to promote cell membrane motility (Chen et al., 2014) (see Figure 1.1). ε-protocadherins (protocadherin-15, 16, and 21) contain variable numbers of EC repeats. Finally, protocadherin-12 and 20 are considered as atypical δ-protocadherins that exhibit structurally different intracellular domains compared to δ1 and δ2-protocadherins (Hulpiau and van Roy, 2009).

The biophysical properties of δ-protocadherins have recently been elucidated. For example, it was found that δ1 and δ2-protocadherins form weak yet highly specific *trans*-interactions, but only with members of the same δ-subgroup (Harrison et al., 2020). The homophilic *trans*-interactions formed between nonclustered protocadherins appear to employ a similar "head-to-tail" mechanism as the one observed in the clustered protocadherins, wherein *trans*-dimers form through antiparallel EC1-EC4 interfaces (Harrison et al., 2020). However, unlike clustered protocadherins, no evidence suggests that nonclustered protocadherins can form *cis*-dimers although that does not imply that they cannot interact via their transmembrane or intracellular domains, which may be possible.

The δ-protocadherins are highly expressed in the nervous system in a regulated spatiotemporal pattern that is similar to classical cadherins (Kim et al., 2007; Krishna-K et al., 2011). The expression of δ-protocadherins is especially prominent during brain development and these proteins have been shown to play multiple roles in various developmental events (Kim et al., 2007). For example, Pcdh7 has been repeatedly shown to be vital for the closure of the neural tube in *Xenopus* (Rashid et al., 2006). Likewise, knockdown (KD) of Pcdh19 in zebrafish causes impaired convergence during neurulation (Emond et al., 2009). Almost all δ-protocadherins are also localised at close proximity to or within axons, dendrites, and synapses (Yamagata et al., 2020). With the exception of Pcdh18, δ-protocadherins have been shown to be vital for dendritic initiation, axon growth, arborisation, and synaptic refinement (Kim et al., 2011). The δ-protocadherins also show specific combinatorial expression patterns within individual neurons that cells adjust to regulate their adhesivity (Bisogni et al., 2018; Krishna-K et al., 2011). Based on this evidence, it is widely believed that δ-protocadherins contribute to a molecular recognition code important for precise cell-cell connectivity such as during neural circuitry formation (Bisogni et al., 2018). In addition to neuronal function, δ-protocadherins also have a role in neural progenitor function. For example, several δ2-protocadherins were recently shown to be important in regulating neural progenitor proliferation in zebrafish via interactions with the non-canonical Wnt receptor, receptor tyrosine kinase (Ryk) (Biswas et al., 2021).

1.1.5. Protocadherins in disease

Because protocadherins exhibit expansive and diverse roles during the development of the brain and other tissues, mutations in these genes have been associated with a variety of human disorders. For example, dysregulation of several protocadherins has been linked to cancer pathogenesis (Van Roy, 2014). Clustered protocadherins are downregulated in many cancers; because the clustered protocadherins are organised in tandem on chromosome 5q31 the genes

are susceptible to long-range hypermethylation and epigenetic silencing (Dallosso et al., 2012). This process affects the expression of almost all of the genes in tandem and has been linked to aggressive tumour proliferation (El Hajj et al., 2017). Additionally, single nucleotide polymorphisms (SNPs) found within multiple clustered protocadherins have been associated with various neuropsychiatric disorders, including autism and schizophrenia (Pardiñas et al., 2018; Smoller et al., 2013). It is theorised that SNPs cause structural alterations within the clustered protocadherins which may affect the binding affinities of the zipper-like complexes, thus disrupting normal neurite repulsion and connections (Flaherty and Maniatis, 2020). Furthermore, SNPs in several nonclustered protocadherins, including *PCDH9 and PCDH17* have been associated with major mood disorders (Chang et al., 2017a; Xiao et al., 2017). SNPs within *PCDH8* and *PCDH10* have also been linked to autism (Butler et al., 2015; Morrow et al., 2008). Finally, mutations in *PCDH19* cause PCDH19-epilepsy, a rare form of monogenic early-onset epilepsy with autism features and developmental delays (Depienne et al., 2009; Dibbens et al., 2008). As this thesis is focused on PCDH19, PCDH19-epilepsy will be discussed in further detail later in this chapter (see section 1.3).

1.2. Protocadherin-19

1.2.1. PCDH19/Pcdh19 structure and function

Protocadherin-19 (PCDH19) is a member of the δ2-protocadherin subfamily and contains six EC motifs, a transmembrane domain, and an intracellular domain which itself contains two conserved motifs and a WIRS domain (see Figure 1.2). In humans, the *PCDH19* gene is located on the X chromosome (Xq22.1) and is encoded by six exons which produce an 1,148 amino acid sequence. Exon 1 codes for the entire extracellular domain, transmembrane domain, and some of the intracellular domain – the rest of the gene is encoded by exons 2-6 (Dibbens et al., 2008; Redies et al., 2005) (see Figure 1.2). Like other protocadherins, PCDH19 can form weak homophilic *trans*-interactions by forming an antiparallel contact described as a “forearm-handshake” (Cooper et al., 2016). PCDH19 can also form *cis* complexes with other cadherins, most notably N-cadherin (Biswas et al., 2010) (see Figure 1.2). The PCDH19-N-cadherin complex has been shown to increase the adhesive capabilities of PCDH19, allowing it to make far more robust *trans*-interactions than PCDH19 alone (Emond et al., 2011). It is also suggested that PCDH19 is the dominant partner in the complex and acts as a switch to engage more specific and durable adhesive interactions (Emond et al., 2011). As previously stated, N-cadherin has multiple important roles during brain development, including regulating progenitor cell proliferation and neurogenesis (Sakane and Miyamoto, 2013), regulating radial migration of neurons

(Martinez-Garay et al., 2016), and synapse formation (Cen et al., 2018). Since PCDH19 is an important functional partner of N-cadherin, it has been theorised that PCDH19 would also play an important role in these events alongside N-cadherin. The exact functions of PCDH19 in brain development will be discussed in further detail later in this chapter.

As previously discussed, the intracellular domain of PCDH19 also interacts with the WAVE complex via its WIRS domain (Chen et al., 2014). Members of the WAVE regulatory complex (WRC), including cytoplasmic FMR1-interacting protein 2 (CYFIP2), Nck-associated protein 1 (NAP1), hematopoietic stem/progenitor cell protein 300 (HSPC300), and Abelson interactor 1 (Abi-1) can interact with PCDH19 directly (Chen et al., 2014; Tai et al., 2010) (see Figure 1.2). The WRC, in turn, interacts with the actin cytoskeleton – thus it is postulated that PCDH19 is important for regulating cytoskeletal dynamics (Hayashi et al., 2017).

A plethora of interacting partners of PCDH19 have recently been identified. Mass spectrometry analysis performed *in vitro* has also shown that PCDH19 interacts with neural precursor cell expressed developmentally downregulated protein 1 (NEDD1) and other microtubule associated proteins (Emond et al., 2021). An *in vivo* mass spectrometry analysis of PCDH19 binding partners at different developmental ages was also performed recently at the Isabel Martinez-Garay (IMG) lab, revealing a number of known and novel interactors involved in cell adhesion, synapse function, and nuclear transport (unpublished data, 2023). Moreover, the C-terminus of PCDH19 has been shown to interact with a conserved subunit of the GABA-A α1 receptor subunit and can regulate the levels of GABA receptors on the cell surface (Bassani et al., 2018) (see Figure 1.2). PCDH19 can also interact with non-POU-domain-containing octamer binding protein (NONO) which is a paraspeckle DNA/RNA-binding protein (Pham et al., 2017). The interaction of NONO with the intracellular domain of PCDH19 was unusual since, at the time, PCDH19 was considered to have no nuclear function. However, recent publications as well as work performed by Dr. Sylvia Newbold at the IMG lab have shown that PCDH19 is capable of undergoing proteolytic cleavage wherein the intracellular domain is processed by A Disintegrin and metalloproteinase domain-containing protein 10 (ADAM10) and possibly γ-secretase that is dependent on neuronal activity (Gerosa et al., 2022; unpublished data, 2023). Moreover, the cytoplasmic fragment (CTF) of PCDH19 interacts with karyopherin subunit β 1 (KPNB1) which transports the CTF to the nucleus (unpublished data, 2023). There, the CTF forms a transcription complex with the chromatin remodeler lysine-specific demethylase 1 (LSD1) and represses transcription of early-intermediate genes (Gerosa et al., 2022). Moreover, as previously discussed the zebrafish Pcdh19 homolog has also been shown to interact with the Wnt receptor Ryk (Biswas et al., 2021) Therefore,

emerging evidence surrounding the interactors of PCDH19 is furthering the understanding of the role of PCDH19 as a cell adhesion and signalling molecule.

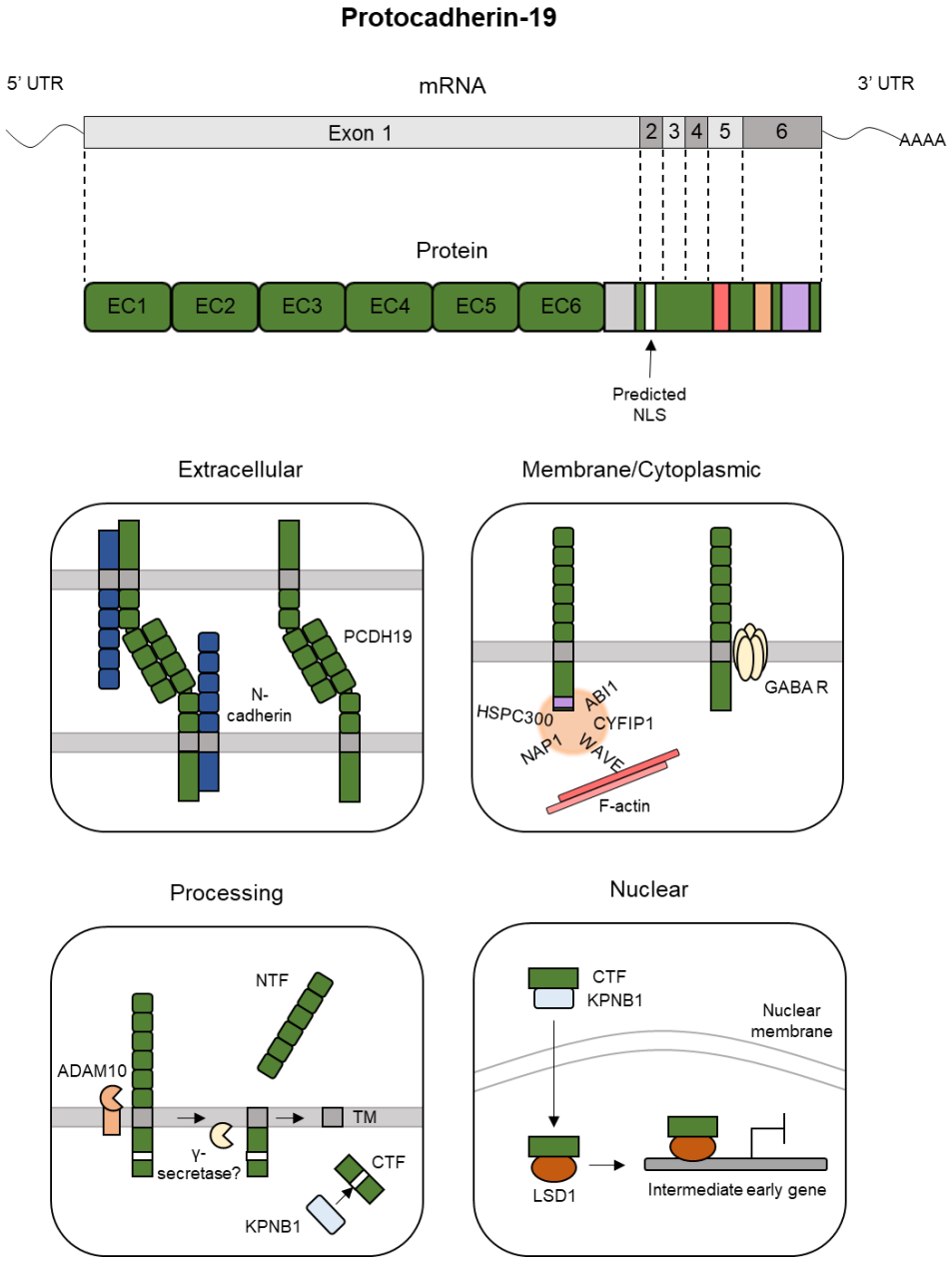


Figure 1.2. PCDH19 structure and function (adapted from Gerosa et al. 2019). Schematic diagram illustrating the mRNA and exons that encode for each portion of the PCDH19 protein. Each box represents one of the main functions of PCDH19. Cell adhesion and its interaction with N-cadherin (Extracellular), association with the WAVE complex and GABA receptor (Membrane/Cytoplasmic), proteolytic processing by ADAM10 and possibly γ-secretase (Processing), and the translocation of the CTF to the nucleus by KPNB1 and its formation with LSD1 to repress early intermediate gene expression (Nuclear).

1.2.2. PCDH19 expression in the mammalian nervous system

PCDH19/Pcdh19 is expressed in a tissue-specific manner throughout the developing and mature nervous system of mammals (Gaitan and Bouchard, 2006). At embryonic day (E)9 in mice, *Pcdh19* expression was detected in the neuroepithelium of the forebrain and some expression was observed in the hindbrain (Gaitan and Bouchard, 2006). At E12, expression of *Pcdh19* was decreased in the telencephalon but was observed in the neural retina, nasal epithelium, and spinal cord (Gaitan and Bouchard, 2006). In the postnatal (P) mouse brain, *Pcdh19* expression has been detected in the cortex, hippocampus, and regions within the limbic system such as the amygdala, the ventral hypothalamus, and the nucleus accumbens (Schaarschuch and Hertel, 2018). Within the cortex, *Pcdh19* expression is mostly limited to the upper regions of layer V and layers II/III (Hertel and Redies, 2011; Pederick et al., 2016). Likewise, at P0, expression within the hippocampus is limited to *cornus ammonis* (CA)1 and CA3 regions, however expression is diminished in the dentate gyrus (DG) (Schaarschuch and Hertel, 2018). Intriguingly, this expression pattern is reversed in the adult cortex, suggesting PCDH19 may play a role in the DG later in life, potentially as a contributor of adult neurogenesis. Expression of *Pcdh19* was also shown to be limited to specific nuclei of the amygdala at P0, such as the caudate putamen, central amygdaloid nucleus, and the dorsal and ventral regions of the lateral amygdaloid nucleus. Interestingly, expression of *Pcdh19* within the lateral amygdaloid nucleus decreased between P0-P7, however expression within the caudate putamen and central amygdaloid nucleus was consistent across postnatal development (Schaarschuch and Hertel, 2018). In regards to non-neuronal tissues, expression has been observed in endothelial tissues, the blood-brain barrier, and within digestive tissues (Gaitan and Bouchard, 2006; Higurashi et al., 2015; Schaarschuch and Hertel, 2018).

A detailed analysis of the expression of *Pcdh19* in specific cortical neuronal subtypes has recently been conducted by employing *in situ* hybridization (ISH) against *Pcdh19* combined with immunohistochemistry (IHC) for neurons and interneurons at P10-20 using specific markers (Galindo-Riera et al., 2021). At P10, expression of *Pcdh19* was found in RORB-positive (+), $SATB2^+$, $CTIP2^+$, and $TBR1^+$ cells, which correspond to layer IV neurons, callosal projection neurons, corticospinal neurons, and corticothalamic neurons, respectively – although expression levels varied between these cell populations (Galindo-Riera et al., 2021). *Pcdh19* expression was also detected in cells positive for parvalbumin and calbindin cells, which are markers for interneurons. Expression analysis was also confirmed using publicly available single cell (sc)RNA-seq databases for mouse and human, which demonstrated that *PCDH19/Pcdh19*

expression shows layer specificity in glutamatergic and GABAergic cortical cells (Galindo-Riera et al., 2021). With regards to the subcellular localisation of PCDH19, western blot (WB) analysis of synaptosome fractions of primary cultured hippocampal neurons has shown strong levels of PCDH19 (Pederick et al., 2016). Further analysis has also revealed that PCDH19 co-localises with synaptic markers Synapsin-1/2 on excitatory neurons (Hayashi et al., 2017) and does not localise within the synapses of GABAergic neurons (Bassani et al., 2018). Recently, the presence of PCDH19 at the presynaptic mossy fibers was also reported (Hoshina et al., 2021). However, it is important to note that to date, few of the commercially available PCDH19 antibodies has been shown to work for well immunohistochemistry without the need of a harsh antigen-retrieval protocol, and thus cannot be used in combination with other antibodies.

Taken together, the spatiotemporal expression of *PCDH19/Pcdh19* suggests that PCDH19 has a multifaceted role in the developing and mature nervous system. In adult brains, PCDH19 exerts a primary role of mediating cell-cell contact between neurons and can orchestrate changes in the cytoskeleton (Gerosa et al., 2019). However, recent publications, as well as research from the IMG lab, have illustrated that PCDH19 also acts as an important signalling molecule in response to neuronal activity (Gerosa et al., 2022; unpublished data, 2023). In the developing nervous system, the role of PCDH19 is only beginning to be unraveled, and recent reports have suggested that PCDH19 may be a key modulator of cell sorting and cortical neurogenesis (Biswas et al., 2021; Pederick et al., 2018; unpublished data, 2019). A detailed analysis of the expression pattern of *Pcdh19* in the developing cortex, as well as the role of PCDH19 during neurogenesis, was recently conducted within the IMG lab (unpublished data, 2019) and will be discussed in greater detail later in this chapter.

1.3. Protocadherin-19 epilepsy

As previously mentioned, mutations in *PCDH19* are linked to a rare form of monogenic epilepsy called PCDH19-epilepsy. PCDH19-epilepsy was first described in 1971 as Epilepsy and Mental Retardation Limited to Females (EMRF) (Juberg and Hellman, 1971). However, the disorder was later referred to as Early Infantile Epileptic Encephalopathy 9 (EIEE9) or Juberg-Hellman syndrome until the gene responsible for the disorder was identified (Depienne et al., 2009; Dibbens et al., 2008; Scheffer et al., 2008). Thereafter, the disorder was also known as PCDH19-epilepsy or PCDH19 girls clustering epilepsy (PCDH19-GCE) (Pederick et al., 2018). For simplicity, throughout this thesis the disorder will be referred to as PCDH19-epilepsy.

PCDH19-epilepsy is a disorder characterised by epileptic and non-epileptic symptoms. The hallmark feature of PCDH19-epilepsy is the early onset of clustered seizures that begin at around 6-36 months of age (Depienne et al., 2011) and that the seizures can be triggered by prolonged fever (Depienne et al., 2011; Dibbens et al., 2008). Seizure types include focal, tonic, tonic-clonic, atonic, absences, and myoclonic jerks and their severity ranges from moderate to severe, being relatively drug-resistant (Dell'Isola et al., 2022). In some patients, the seizures become less severe or even disappear in adolescence, however the non-epileptic symptoms are typically the most disabling for adult patients. PCDH19-epilpesy is comorbid with other disturbances such as intellectual disability and behavioural disorders that occur in ~75% and ~55% of patients, respectively (Camacho et al., 2012). Comorbidity with autism spectrum disorder and obsessive-compulsive disorders has also been widely reported (Kolc et al., 2019). Cognitive delay can occur early in development at ~2 years of age, however it has been reported that ~15% patients display cognitive delay before seizure onset (Samanta, 2020). Therefore, the cognitive impairments are only partially related to the epileptic symptoms which suggests that other genetic, biological, or environmental factors may underlie the phenotypic spectrum (Dell'Isola et al., 2022).

There have been ~150 mutations identified within *PCDH19* that have been linked to PCDH19-epilepsy, including familial and *de novo* mutations (Kolc et al., 2019). Many of them (~86%) have been shown to aggregate within the extracellular domain of PCDH19, especially within EC3-EC4, which is important for PCDH19 to form *trans*-interactions. Most mutations have been identified as missense or frameshift mutations, with a fraction of nonsense changes as well (Kolc et al., 2019). It is generally theorised that PCDH19-epilepsy is caused by loss-of-function mutations that reduce the adhesive properties of PCDH19. However, the underlying cause of PCDH19-epilepsy is still widely unknown.

As previously mentioned, *PCDH19* is located on the X chromosome and, therefore, PCDH19-epilepsy is an X-linked disorder. Typically, in most X-linked disorders the hemizygous males (XY) exhibit the disease phenotype whereas the heterozygous females (XX) are spared or display weaker symptoms. This is due to hemizygous males only possessing one copy of the mutant gene, while heterozygous females carry the mutant and normal copy of the gene. Thus, due to random X chromosome inactivation (see section 1.3.1) in females, the mutant allele is expressed in approximately 50% of the cells, whereas in males the mutant allele will be expressed in all cells, which means that any deleterious effects caused by the mutation are typically more severe in males. However, unlike most X-linked disorders, PCDH19-epilepsy is characterised by mostly affecting heterozygous females, whereas hemizygous males are spared (Dibbens et al., 2008).

This unusual inheritance pattern was initially explained by the presence of *PCDH11Y* on the Y chromosome which was theorised to exert a compensatory effect in males (Dibbens et al., 2008). However, this was disproven when male patients were found to display symptoms of PCDH19-epilepsy (Depienne et al., 2009; Kolc et al., 2020; de Lange et al., 2017). Further analysis revealed that deleterious postzygotic mutations were the likely cause of these males developing PCDH19-epilepsy. Interestingly, an individual with Klinefelter-syndrome, a condition in which males exhibit an extra X chromosome (XXY) was also described to display symptoms of PCDH19-epilepsy (Romasko et al., 2018). Taken together, this encouraged the development of an expanded theory of the underlying cause of PCDH19-epilepsy, known as the cellular interference hypothesis.

1.3.1. X chromosome inactivation and the cellular interference hypothesis

As most X-linked genes, *PCDH19* is subjected to random X chromosome inactivation (XCI). XCI occurs when one of the two X chromosomes in females is inactivated early in embryonic development (Lyon, 1961). XCI arises mostly in mammals and is an evolutionary trait that compensates for genetic imbalances between homogametic and heterogametic sexes, thus achieving dosage equivalency for X-linked genes between males and females. XCI is normally random, with equal probability that either X chromosome will be inactivated in a given cell (Lyon, 1961). This also means that in females, the alleles of X-linked genes are expressed in a mosaic fashion depending on which X chromosome is inactivated. Likewise, heterozygous female patients with PCDH19-epilepsy harbour two cell populations due to the mosaic expression of normal ($PCDH19^+$) and mutant ($PCDH19^-$) cells that reside in the same tissue.

According to the cellular interference hypothesis, this co-existence of $PCDH19^+$ and $PCDH19^-$ cells in the brain of PCDH19-epilepsy females would underpin the disorder (see Figure 1.3). The basis of this hypothesis is that the presence of $PCDH19^+$ and $PCDH19^-$ cells disrupts normal cell-cell communication in the brain (Depienne et al., 2009, 2011). Moreover, mosaic expression of PCDH19 is thought to impair neuronal recognition between cells which leads to scrambled defects in neural wiring and circuitry formation, thus leading to improper control of neural activity which may underlie the epileptic and behavioural symptoms of PCDH19-epilepsy. The cellular interference hypothesis also assumes that hemizygous males do not display a disease phenotype since it is the presence of both $PCDH19^+$ and $PCDH19^-$ cells that drives the disorder. Therefore, it is possible that a compensation mechanism exists in hemizygous males that counteracts the loss-of-function of PCDH19.

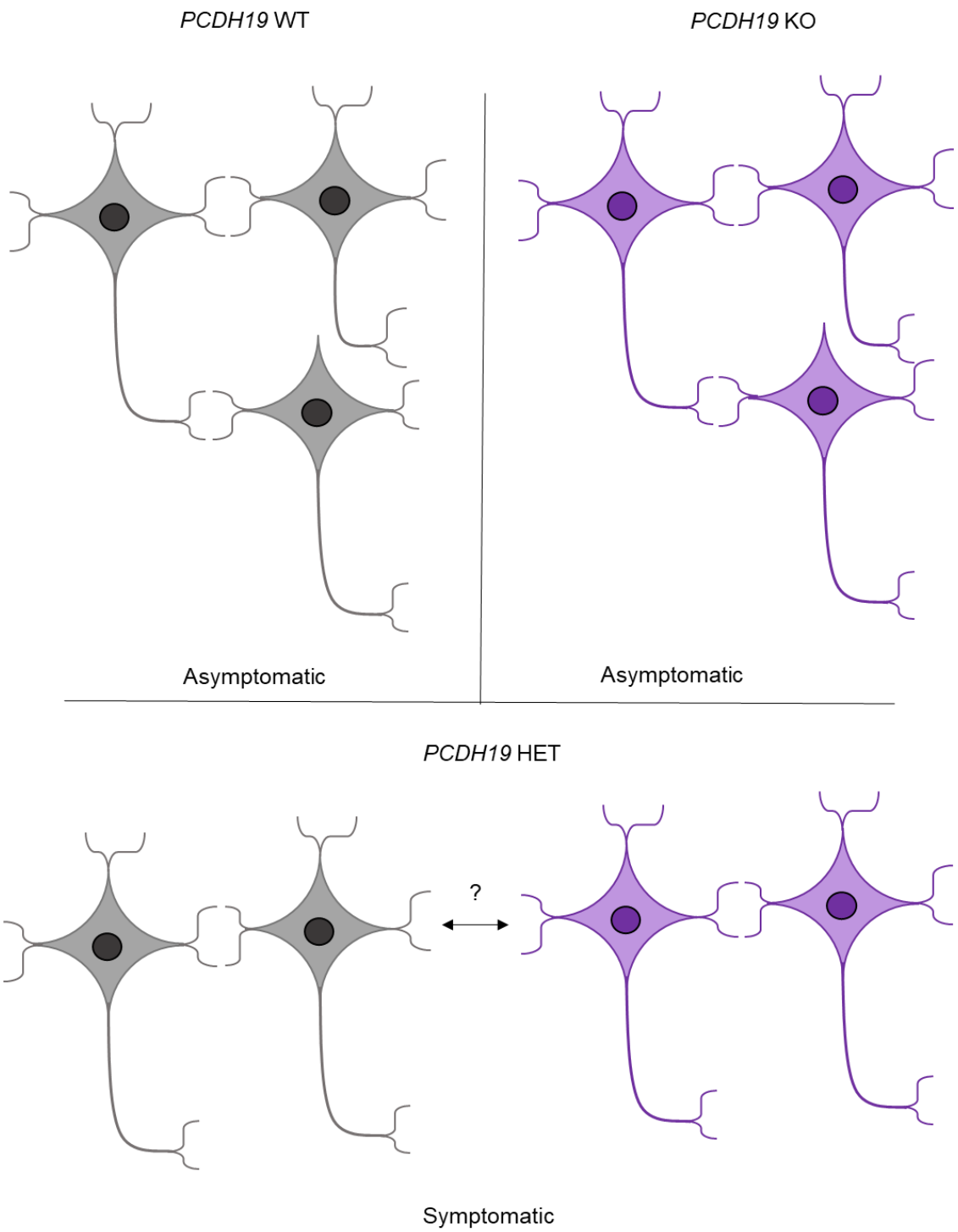


Figure 1.3. The cellular interference hypothesis of PCDH19-epilepsy (adapted from Depienne et al. 2009). Representation of PCDH19 WT cells in healthy individuals, PCDH19 KO cells in hemizygous individuals, and PCDH19 WT and KO cells co-existing in the same brain in heterozygous individuals.

1.3.2. Neurosteroid hypothesis

Although the cellular interference hypothesis is the most widely accepted theory of the underlying cause of PCDH19-epilepsy, other non-mutually excluding theories have also been formulated. In particular, it has been shown that PCDH19-epilepsy patients display altered levels of neurosteroids, such as allopregnanolone deficiency and deficiency in the allopregnanolone synthesizing enzyme, aldo-keto reductase 1C (AKR1C) (Tan et al., 2015). Allopregnanolone has been shown to be a positive regulator of GABA-A receptors and thus stimulates neural inhibitory activity. Moreover, other neurosteroids are also decreased in PCDH19-epilepsy patients, including pregnenolone sulphate, 17OH-progesterone, progesterone, and cortisol (Trivisano et al., 2017). A synthetic analogue of allopregnanolone called ganaxalone has recently been used to treat seizures in PCDH19-epilepsy patients (Violet study). Ganaxalone functions similarly to allopregnanolone by acting as a positive allosteric modulator of GABA receptors. Interestingly, treatment with this compound has initially shown to reduce seizures in ~60% PCDH19-epilepsy patients, however these results are still considered as preliminary (Moncayo et al., 2022). Nevertheless, there appears to be a strong link between neurosteroid imbalances and PCDH19-epilepsy symptoms and although a mechanism has yet to be described, research in this area may offer a viable treatment for this disorder.

1.4. Modelling PCDH19 heterogeneity

Mosaic presence of $PCDH19^{+}$ and $PCDH19^{-}$ cells in the brain in addition to alterations in the levels of neurosteroids are likely both key drivers behind PCDH19-epilepsy. Several *in vitro* and *in vivo* models have been employed to study the effects of PCDH19 heterogeneity at the cellular and tissue levels, respectively.

1.4.1. In vivo models of PCDH19 heterogeneity

Taconic Biosciences generated a *Pcdh19* knockout (KO) mouse model by replacing the first three exons of the gene (Exon 1-3) with a β-galactosidase (*β-gal*) neomycin cassette reporter (see Figure 1.4), which is under the control of the endogenous *Pcdh19* promoter. The initial characterisation of this model reported no gross brain abnormalities between wildtype (WT), heterozygous (HET) and KO mice, as measured by cortical thickness (Pederick et al., 2016). Further characterisation of these mice revealed no gross abnormalities in cortical lamination, however subtle differences in neuronal positioning within the cortex were detected in the HETs (Galindo-Riera et al., 2021). Another *Pcdh19* KO mouse model has also been generated by insertion of *LacZ* into the *Pcdh19* locus (Hayashi et al., 2017). Here, researchers were able to

show that PCDH19 localised within Homer^{+} and Synapsin^{+} synapses, leading to the possibility that heterogenous expression of *Pcdh19* may affect synaptic architecture. However, no differences were found in the number of spines between WT, HET, and KO animals. Moreover, no gross abnormalities were also observed in the overall morphology of neurons between groups (Hayashi et al., 2017). Interestingly, another mouse model of *Pcdh19* heterogeneity has been recently generated via CRISPR/Cas9 induced KO (Hoshina et al., 2021). Here, the researchers were able to demonstrate that PCDH19 is localised predominantly within the presynaptic mossy fibers. Moreover, *Pcdh19* HET animals showed disrupted presynaptic development due in part to the lack of a PCDH19/N-cadherin complex at the presynapse which causes synaptic mismatch (Hoshina et al., 2021).

Finally, another mouse model of PCDH19 heterogeneity has been developed wherein endogenous *Pcdh19* has been tagged with a hemagglutinin (HA)-FLAG epitope sequence that attaches to the C-terminus of the PCDH19 protein (Pederick et al., 2018). These animals were then crossed with the Taconic Bioscience *Pcdh19* KO line to generate *Pcdh19*$^{HA\text{-}FLAG/\beta\text{-}gal}$ double transgenics. Interestingly, a striking cell sorting phenotype was observed in the developing HET brain, wherein PCDH19-WT and PCDH19-KO neural progenitor cells segregated away from each other and formed distinctive columnar structures in the cortex (Pederick et al., 2018). It was further confirmed that this separation was not due to random positioning after XCI, but because PCDH19-WT and PCDH19-KO cells segregate from each other. This finding was not observed in hemizygous KO animals, but has been replicated in the IMG lab wherein IHC using β-GAL antibodies has demonstrated a striking cell sorting phenotype which was observed only in the cortex of HETs and as early as E11-E12 (see Figure 1.4) (unpublished data, 2019).

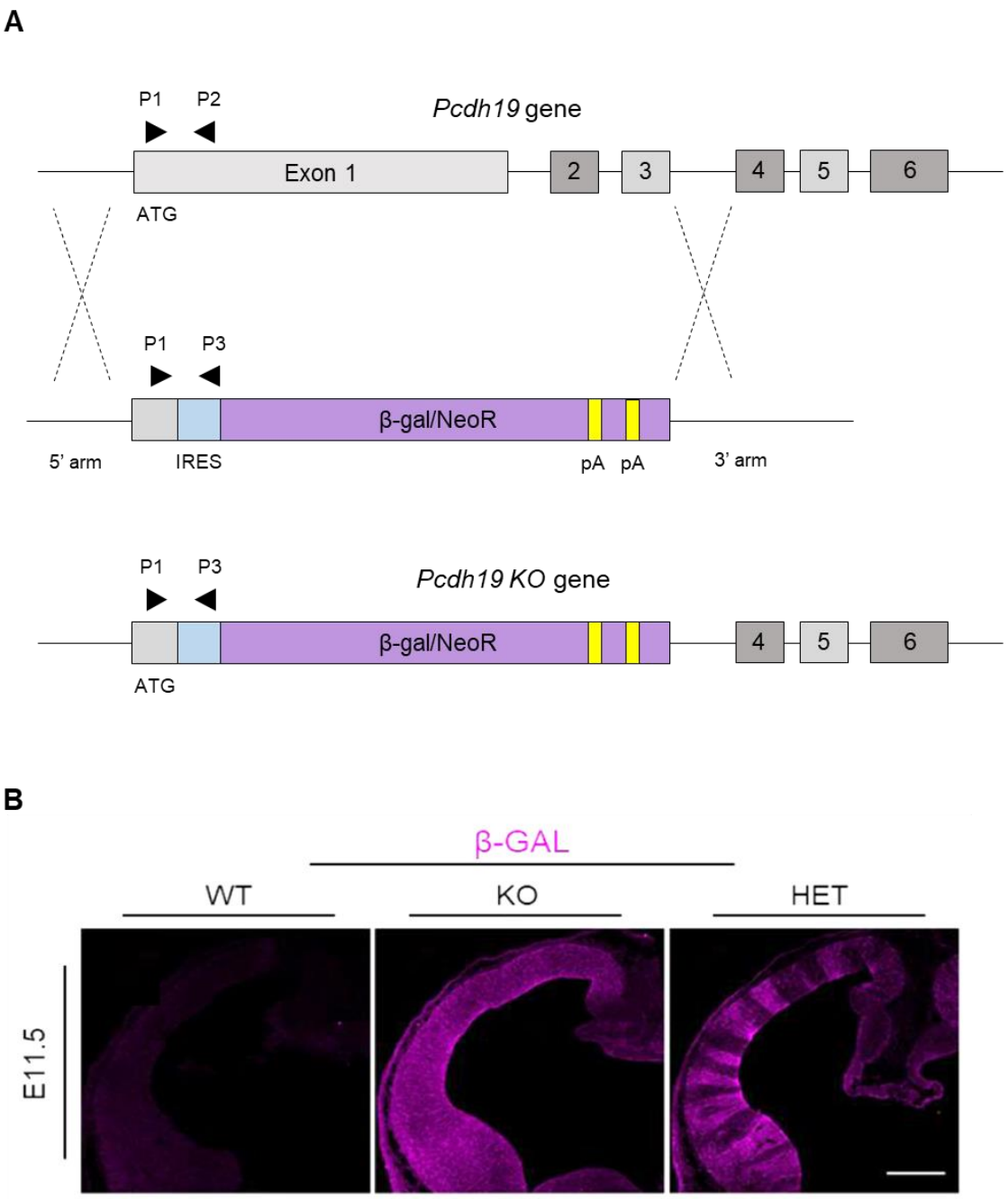


Figure 1.4. *Pcdh19* KO mouse model (adapted from Pederick et al. 2016). (A) Schematic representing the Taconic Biosciences *Pcdh19* KO mouse line, demonstrating the replacement of exons 1-3 of *Pcdh19* with β-galactosidase/neomycin cassette. P1-P2, primers for detecting WT allele. P1-P3, primers for detecting KO allele. IRES, internal ribosome entry site; pA, polyA. (B) Immunostaining of β-GAL on E11 *Pcdh19* WT, KO, and HET cortices, demonstrating the abnormal cell-sorting behaviour of $PCDH19^{+}$ and $PCDH19^{-}$ cells. Taken from the work of Dr. Jessica Griffiths (unpublished data, 2019).

1.4.2. In vitro models of PCDH19 heterogeneity

To study the effects of heterogeneous and KO expression of *PCDH19/Pcdh19* at the cellular level, several *in vitro* models have been generated. For instance, the cellular mosaicism present in the *Pcdh19* HET brain has been modelled *in vitro* by co-culturing the neural progenitor cells from E14 WT and KO embryos in a 1:1 ratio to generate neurospheres (Homan et al., 2018). Interestingly, PCDH19 KO progenitors showed increased neurogenesis when cultured separately from WT cells and when co-cultured with WT cells (Homan et al., 2018). Induced pluripotent stem cell (iPSC) lines have also been generated from the fibroblasts of PCDH19-epliepsy patients and these cells also show increased neurogenic behaviours *in vitro* (Borghi et al., 2021; Homan et al., 2018). Interestingly, the increase in neurogenesis was recently attributed to perturbations in positioning of the mitotic spindle configuration (Borghi et al., 2021). However, it is worth mentioning that, in that study, many of the patient-derived iPSCs displayed more than two centrosomes during mitosis, which would be expected to lead to aneuploidy and apoptosis, rather than increased neurogenesis. In the same study, cortical organoids derived from iPSCs of PCDH19-epilepsy patients were shown to have reduced size and altered overall morphology compared to non-patient-derived organoids (Borghi et al., 2021). However, it was not explored whether the reduced size was due to perturbations in neurogenesis events, as this was only illustrated using iPSCs. Organoid models share a striking resemblance to the developing brain, including presenting proliferative layers such as a ventricular zone and subventricular zone, in addition to presenting a cavity that is similar to a ventricle (Lancaster et al., 2013). Therefore, it would be interesting to observe whether mosaic expression of *PCDH19* in organoids causes a similar cell sorting phenotype as described *in vivo*. Moreover, it would also be interesting to examine whether perturbed expression of *PCDH19* alters neurogenic events in organoids.

1.4.3. Conclusions from the models of PCDH19 heterogeneity

Taken together, modelling the mosaic expression of *PCDH19/Pcdh19* has illustrated various effects both *in vivo* and *in vitro*. The most striking phenotype observed *in vivo* was the unusual cell sorting of neural progenitors which occurs only in HET animals (Pederick et al., 2018). This is in line with the cellular interference hypothesis in that the presence of $PCDH19^{+}$ and $PCDH19^{-}$ cells causes changes in cell positioning and connectivity which may underlie PCDH19-epilepsy. This is further supported by the fact that many symptoms of PCDH19-epilepsy are early onset, which suggests that the cause of the disorder may be linked to perturbations that occur during early brain development events. The role of PCDH19 in early brain development is only beginning to be unraveled (Gerosa et al., 2019; unpublished data, 2019). However, experiments performed

in vitro strengthens the idea that PCDH19 may play a role in neurogenesis in neural progenitors (Borghi et al., 2021; Homan et al., 2018). Because PCDH19 is expressed in cortical neural progenitors during early brain development, PCDH19 may also have a role in orchestrating neurogenesis *in vivo,* and mosaic or KO expression of *PCDH19* may disrupt this process. To date, the role of PCDH19 in neurogenesis events *in vivo* has yet to be reported, however data generated by Dr. Jessica Griffiths from the IMG lab suggest that mosaic expression of *Pcdh19* does disrupt normal neurogenesis rates during early cortical development (unpublished data, 2019). The mechanisms of cortical neurogenesis and the involvement of PCDH19 are discussed below.

1.5. Cortical development

The mammalian cerebral cortex is a highly organised brain structure that is the seat of higher cognitive and sophisticated motor and sensory functions. The extraordinary capabilities of the cortex are due in part to the remarkable volume of the six interconnected layers which contain billions of neuronal cell types and other cell classes (Herculano-Houzel, 2009). The development of the cortex occurs in an “inside-out” manner, wherein the first-born neurons populate the deep layers of the cortex and the later born neurons migrate past them and form the superficial layers. The increase in cortical size during evolution is widely considered to be a major component of the evolution of mammalian species (Espinós et al., 2022). The most important element of cortical expansion is the number of neurons and glial cells that are produced. The mechanisms of corticogenesis are highly conserved between vertebrates, however they become more elaborate in higher mammalian species, underlying the expansion of the cortex in these organisms (Molnár, 2011; Molnár and Clowry, 2012). In this section, the different steps of corticogenesis will be described in detail, with a particular emphasis on the development of excitatory neurons during mouse cortical development. Because neurogenesis is an important component of this thesis, the cellular and molecular mechanisms of neurogenesis will be discussed in greater detail later in this chapter (see section 1.6).

1.5.1. Pre-neurogenesis: progenitor expansion

After neurulation, the walls of the newly-formed neural tube are composed of neuroepithelial (NE) cells which are considered the first multipotent neural progenitor cell (Lawson et al., 2001; McShane et al., 2015). NE cells form the pseudostratified neuroepithelium and, similar to other epithelial cells, are highly polarised along the apical-basal axis wherein the apical processes connect to the neural canal and the basal processes span the width of the neural tube (Götz and

Huttner, 2005). As NEs divide, they undergo a phenomenon called interkinetic nuclear migration (INM) where the nucleus of the cell moves in the apico-basal axis at certain stages of the cell cycle: in a basal direction during G_1 phase, then in an apical direction during S-phase, until finally undergoing M-phase and dividing at the apical surface (Sauer, 1935) (see Figure 1.5). NEs are also characterised by occludin$^+$ tight junctions at the apical endfeet which enable cell-cell communication through various adhesion and small molecules (Aaku-Saraste et al., 1996; Elias and Kriegstein, 2008). They also possess a prominin$^+$ primary cilium which protrudes into the neural canal (and later the lateral ventricles) to sense signalling molecules from the cerebrospinal fluid (CSF) (Lehtinen et al., 2011). Initially, NEs undergo extensive symmetric proliferative divisions to increase the progenitor cell pool (Rakic, 1995), which is sometimes referred to as the "expansion" phase of cortical development (see Figure 1.5). This continues until the onset of neurogenesis (~E10 in mice) when NEs begin to transition into a more fate-restricted progenitor cell type, called radial glial cells (RGCs) (Aaku-Saraste et al., 1996; Götz and Huttner, 2005) (see Figure 1.5). This transition is characterised by the replacement of the tight junctions by an adherens junction (AJ) belt that exhibits specific markers such as N-cadherin and zona occludens 1 (ZO-1) (Aaku-Saraste et al., 1996). Another feature of the NE-RGC transition is that RGCs express astroglial markers such as glutamate-aspartate transporter (GLAST) and other key neurogenic transcription factors (TF), such as PAX6 (Englund et al., 2005; Götz et al., 1998, 2015). RGCs also retain several characteristics of NEs, including the apical-basal polarity and they also undergo INM (see Figure 1.5). However, unlike NEs which undergo INM across the entire width of the cortical primordium, RGC IMN is restricted closer to the lateral ventricle (Götz and Huttner, 2005). Therefore, the boundary in which RGCs undergo INM is called the ventricular zone (VZ). Similar to NEs, RGCs also undergo symmetric proliferative divisions to further expand the progenitor cell pool, however their proliferative potential is more restricted than NEs (Takahashi et al., 1996). Unlike NEs, in which asymmetric divisions are very are, RGCs can also divide asymmetrically to give rise to daughter cells that will differentiate into post-mitotic neurons. The transition of RGCs from symmetric proliferative to asymmetric neurogenic divisions is called the neurogenic switch, and occurs ~E10 through to E12 in mice (Noctor et al., 2001, 2004).

1.5.2. Direct and indirect neurogenesis

Upon the onset of the neurogenic switch, RGCs are the predominant neural progenitor cell type and will give rise to most of the neurons in the cortex (Noctor et al., 2001). RGCs produce neurons by either direct or indirect neurogenesis (see Figure 1.5). Direct neurogenesis occurs when RGCs undergo asymmetric division and produce a neuronal daughter cell directly. Indirect neurogenesis

refers to the process when RGCs undergo asymmetric division but instead give rise to an intermediary progenitor cell type that subsequently produces neurons. In mice, the most common intermediary progenitors cell types are called intermediate progenitors (IPs) (Götz and Huttner, 2005). Unlike RGCs, IPs are multipolar, detach from the apical surface and migrate basally, positioning above the VZ and forming a new germinal area called the subventricular zone (SVZ). The transcriptional landscape of IPs also changes as they reduce expression of PAX6 and start to express TBR2 and neuronal markers (Englund et al., 2005). In higher mammals such as primates, IPs are known to undergo symmetric proliferative divisions, however this occurs rarely in mice, where IPs typically undergo symmetric self-consuming divisions to give rise to two neurons (Miyata et al., 2004; Noctor et al., 2004) (see Figure 1.5). Nevertheless, from one RGC indirect neurogenesis division, the subsequent IP would produce two neurons, thus doubling the neuronal output compared to RGC direct neurogenesis divisions. It is widely considered that IPs and other transit-amplifying cells are responsible for the expansion and gyrification of the cortex across evolution (Fernández et al., 2016; Kowalczyk et al., 2009; Llinares-Benadero and Borrell, 2019; Martínez-Cerdeño et al., 2006; Noctor et al., 2004; Tarabykin et al., 2001). This is demonstrated by the fact that direct neurogenesis occurs for the majority of RGC divisions in vertebrates with smaller brains, including snakes and birds. However, in mammals, RGCs mostly undergo indirect neurogenesis which subsequently increases neuronal output (Cárdenas et al., 2018). This is also evident in higher mammalian species such as primates, as other transit-amplifying cells exist including basal RGCs (bRGCs) which further divide the SVZ into the outer SVZ (oSVZ) and inner SVZ (iSVZ) (Reillo et al., 2011). However, these cells rarely exist in murine cortical development, where it is IPs that form the majority of transit-amplifying progenitors.

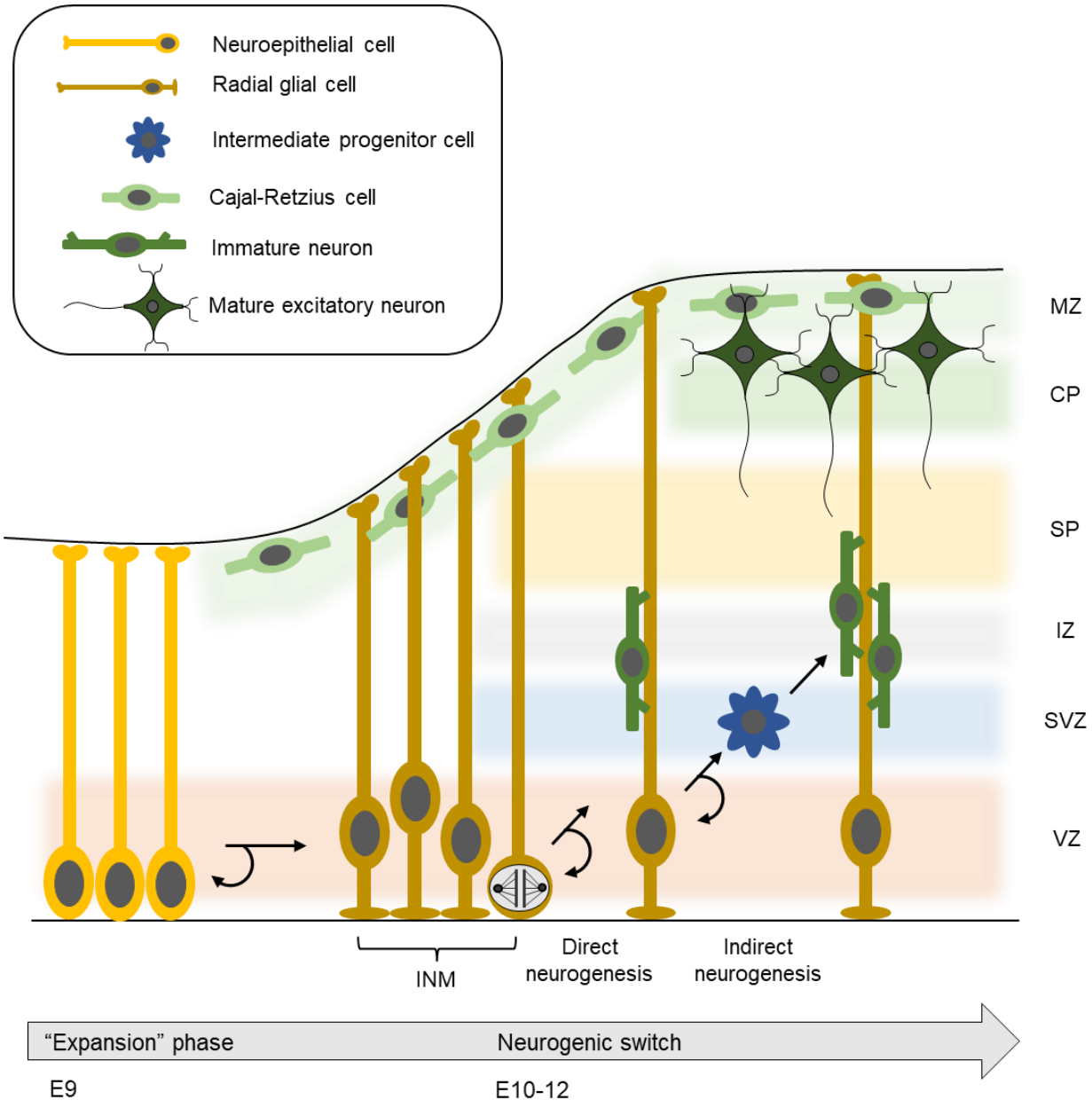


Figure 1.5. Stages of neurogenesis during mouse cortical development. Representation of the main stages of mouse cortical neurogenesis. The box above illustrates the names of the cell types in the diagram. MZ, marginal zone; CP, cortical plate; SP, subplate; IZ, intermediate zone; SVZ, subventricular zone; VZ ventricular zone; INM, interkinetic nuclear migration. Adapted from the work of Dr. Sylvia Newbold (unpublished data, 2023).

1.5.3. Neuronal migration

The migration of neurons during development is termed “radial migration” and is central to the inside-out development of the cortex (Rakic, 1972). Once born, cortical neurons become multipolar and detach from the apical surface and then migrate basally until positioning themselves close to the SVZ, in a region called the multipolar-accumulation zone (MAZ). Once positioned in the MAZ, early born neurons convert from multipolar to bipolar cells and extend leading processes that attach to the basal surface. The leading process is then shortened, and this subsequently drags the soma of the neurons towards the pial surface. This type of migration is independent of radial fibers and is referred to as somal translocation (Nadarajah et al., 2003). As neurogenesis proceeds, somal translocation occurs less frequently since the cortical area expands and the basal surface moves further away from the VZ and SVZ. Therefore, newborn neurons are more reliant on locomotion along the RGC basal fibers which act as a scaffold (Kawauchi et al., 2010). Once the leading process of the migrating neuron reaches the marginal zone (MZ), they detach from the RGC fiber and anchor to the basal surface, finishing with a translocation step similar to somal translocation, which is sometimes referred to as terminal translocation (Kawauchi et al., 2010; Sekine et al., 2011). This continues until the development of the six layered cortex is complete, when the remaining RGCs in the VZ start giving rise to various glia, including astrocytes and oligodendrocytes, beginning around E17 in mice.

It is important to note that a specific type of excitatory neuron, called Cajal-Retzius (CR) cells, migrates tangentially into the cortex. These neurons typically arise from the cortical hem and caudomedial telencephalon and thalamic eminence, and migrate laterally into the developing cortex at around E10 in mice (Sekine et al., 2014). Early in development, CR cells, early born neurons, and other cells, form the preplate (PP) at the basal surface of the cortex. As neurogenesis proceeds, migrating neurons split the PP into the subplate (SP) and MZ. CR cells end up exclusively in the MZ and secret the extracellular matrix protein, reelin (RELN) which aids in neuron positioning (Sekine et al., 2014). As more neurons settle, the space between the SP and the MZ (called the cortical plate (CP)) increases, forming layers II to VI of the cortex (see Figure 1.5). Once neurons are positioned, they begin to extend axons and form highly specific cell-cell connections. It is generally considered that neurons within the deep layers (layers V and VI) form subcortical connections with areas in the thalamus (layer VI) and spinal cord (layer V) whereas superficial layers (layers II, III, IV) form intracortical connections, although some neurons in the deep layers also connect within the cortex.

1.5.4. Deterministic and multipotency of neural progenitors

As detailed previously, neurogenesis occurs in an inside-out manner, wherein new-born neurons establish the deep layers of the cortex and later-born neurons establish the superficial layers. Although the cortex is composed of an incredible array of diverse neurons, these cells are generated from a seemingly uniform population of progenitor cells. Understanding the capability of RGCs an IPs to produce such a complicated range of neurons has been a leading focus in the field of cortical development. It was widely considered that at the beginning of neurogenesis, RGCs are a homogenous multipotent cell population that can generate all types of cortical neurons but follow a fine-tuned temporal sequence to produce neurons at the right time (Klingler and Jabaudon, 2020; Leone et al., 2008). This was demonstrated *in vitro* when mouse embryonic stem cells were cultured in morphogen-free medium and yet were able to sequentially generate a diverse range of neurons that broadly resembled the different layers of the cortex (Gaspard et al., 2008). Moreover, several genetic fate mapping analyses performed *in vivo* using mosaic analysis with double markers (MADM) or *in utero* electroporation (IUE) have found that RGCs that express *Sox9* and *Fezef2* are multipotent and can produce neurons for any of the cortical layers (Guo et al., 2013; Kaplan et al., 2017). Clonal analysis of IP lineage has also shown that IPs are able to generate neurons for any of the cortical layers (Vasistha et al., 2015). Taken together, it has been theorised that RGCs and IPs are inherently multipotent and can produce neurons in an orderly and deterministic manner (Gao et al., 2014).

On the other hand, there is evidence that some neural progenitors are inherently more fate restricted which has become a significant controversy in the field (Eckler et al., 2015; Gil-Sanz et al., 2015). For example, genetic fate mapping *in vivo* and *in vitro* have identified a sub-lineage of RGCs that produce neurons mainly for the superficial layers (Franco et al., 2012). Moreover, a subset of RGCs in the mouse cortex has also been shown to exhibit delayed neurogenesis at the initial neurogenic switch, and instead undergo neurogenesis later in development to produce upper layer callosal neurons (García-Moreno and Molnár, 2015). Taken together, these experiments suggests that a subset of RGCs may specialise in producing specific types of cortical neurons. Llorca et al. (2019) recently concluded that the lack of an established model on how progenitors produce neurons is partly due to the absence of studies that have combined multiple approaches to examine progenitor fates. To address this, the researchers used retroviral labelling, MADM, and genetic fate mapping to investigate the clonal production of neurons by utilising the advantages of each technique (Llorca et al., 2019). They demonstrated from all three approaches that most progenitors produced neurons for both deep and superficial layers of the cortex,

however a sizable fraction (1 in 6) of progenitors were more restricted to produce neurons only for specific layers. Moreover, using Bayesian mathematical modelling, Llorca et al. (2019) were able to demonstrate that a stochastic developmental approach in which progenitors undergo a series of probability-based decisions was the best model to explain the wide diversity of neuron outputs capable by most RGCs. Nevertheless, the debate over whether RGCs are restricted to a pre-determined lineage or are multipotent is still on going, and more fate-restricted progenitors are being identified using single cell transcriptomics (Eze et al., 2021; Ratz et al., 2022; Telley et al., 2019).

1.6. Regulatory mechanisms of neurogenesis

As detailed above, RGCs undergo an initial phase of symmetric proliferative divisions before changing to asymmetric neurogenic divisions, ~E10 in mice. The switch from symmetric proliferative to asymmetric neurogenic divisions is termed the “neurogenic switch” (Götz and Huttner, 2005). In mice, the neurogenic switch is permanent i.e., once RGCs commit to neurogenic divisions they cannot return to a pure proliferative state. Therefore, the regulation of the neurogenic switch is vital for proper brain development; neurogenesis must start at an exact time to ensure that the pool of progenitors is sufficient to produce the correct number of neurons. For example, microcephaly and macrocephaly have been linked to premature or delayed onset of the neurogenic switch, respectively (Pang et al., 2008). Even subtle changes in neurogenesis have been linked to various neurological disorders, including drug-resistant epilepsies and intellectual disability (Guarnieri et al., 2022). Therefore, understanding the many extrinsic and intrinsic mechanisms that regulate neurogenesis and the neurogenic switch is essential for understanding cortical development in both health and disease contexts.

1.6.1. Cleavage plane

The geometric orientation of the cleavage plane during mitosis has long been theorised to play a direct role in controlling symmetric vs. asymmetric RGC divisions (Pietro et al., 2016). As RGCs divide during mitosis, the orientation of the cleavage plane determines whether the daughter cells equally or unequally inherit specific cell fate determinants that are distributed throughout the mother cell (Chenn and McConnell, 1995). For example, progenitors that exhibit a cleavage furrow that is perpendicular to the ventricular surface (vertical) are more likely to evenly distribute key cell fate determinants to each daughter cell. One such determinant is the apical domain, which during a vertical division is bisected and inherited by both daughter cells. Consequently, the daughter cells are more likely to retain a progenitor identity. On the other hand, if progenitors

exhibit a cleavage furrow that is slightly tilted (oblique) or parallel to the ventricular surface (horizontal), the daughter cells will unevenly inherit cell fate determinants. For example, one daughter cell may inherit the whole apical domain, whereas for the other daughter cell it is bypassed. This would lead to an asymmetric cell division as the daughter cells undertake different cell identities (Chenn and McConnell, 1995; Konno et al., 2007; Kosodo et al., 2004). Although it was traditionally thought that inheritance of the apical domain was critical for daughter cells to retain a progenitor identity, this type of inheritance was only observed in a relatively low fraction of cells (Konno et al., 2007). Therefore, the inheritance of cell fate determinants other than the apical domain is important for controlling asymmetric divisions. For instance, the inheritance of the basal process has also been shown to be pro-proliferative (Tsunekawa et al., 2012). Here, it was described that the mRNA of cyclinD2, an important molecule in cell cycle progression, aggregates within the basal process of RGCs. Thus, daughter cells that inherit the basal process have increased likelihood to remain within the cell cycle. Likewise, it has also been demonstrated that daughter cells that inherit the basal process are able to regrow the apical endfoot (Alexandre et al., 2010). Therefore, inheritance of RGC cellular features is likely an important factor for determining daughter cell identity.

1.6.2. Mitotic spindle orientation

The mitotic spindle assembly largely governs cleave plane orientation and is also involved in modulating symmetric vs. asymmetric RGC divisions. The spindle assembly is generated by the translocation of the centrosomes from the apical process to the center-side of the soma where they duplicate and form the spindle poles (Hu et al., 2013). The spindle poles then associate with astral microtubules, which connect the centrosome to the cell cortex (Pietro et al., 2016). During mitosis, the mitotic spindle assembly is important for positioning organelles and other cell fate determinants to ensure viable inheritance for each daughter cell. Disruption of the mitotic spindle assembly has been associated with cortical malformations. For example, mutations in the LIS1 regulator nuclear distribution protein homolog 1 (NDE1) causes defects in spindle orientation which perturbs neurogenesis (Feng and Walsh, 2004). Knockdown (KD) of *Aspm* a key spindle regulator, decreases the number of vertical divisions in NEs and causes precocious cell delamination (Fish et al., 2006). Interestingly, ASPM is also associated with human primary microencephaly. Likewise, overexpression of citron kinase (*Citk*), another gene associated with microencephaly, rescues mitotic spindle defects caused by *Aspm* KD and increases the number of NE vertical divisions (Gai et al., 2017). KD of *Lgn*, a key protein that anchors the mitotic spindle poles to astral microtubules and the cell cortex, randomises the position of the spindle assembly

(Konno et al., 2007). Interestingly, despite altering spindle positioning within the progenitors, *Lgn* KD did not affect neurogenesis output. Taken together, spindle and cleavage plane orientation impacts symmetric vs. asymmetric divisions, however their exact importance is still a matter of controversy (Uzquiano et al., 2018).

1.6.3. Cell cycle length

Early studies of the neurogenic switch have linked the cell cycle length to symmetric and asymmetric divisions. In principle, progenitor cell cycle length increases over the course of neurogenesis i.e., asymmetrically dividing progenitors exhibit a longer cell cycle length than symmetrically dividing progenitors (Takahashi et al., 1993). Later studies demonstrated that the lengthening of G_1 phase is the main change in cell cycle length between symmetric and asymmetric dividing progenitors (Arai et al., 2011; Calegari et al., 2005; Pilaz et al., 2009). In fact, artificial lengthening of G_1 phase was shown to induce premature neurogenic divisions in RGCs, as demonstrated by an increase in the number of new-born IPs. Likewise, the shortening of G_1 phase delayed neurogenesis and caused RGCs to over proliferate (Pilaz et al., 2009). Manipulating the cell cycle in these experiments was achieved by inhibiting the function of G_1 specific cycling proteins and kinases such as cyclinD1/2 and the cyclin-dependent kinases (CDK) (Calegari and Huttner, 2003; Lange and Calegari, 2010). Interestingly, these proteins are activated in response to various extrinsic signalling pathways, such as Wnt, FGF, and Notch. The role of signalling pathways in regulating neurogenesis will be discussed in greater detail later in this section (see section 1.6.6.).

1.6.4. Transcription factors

Various homeodomain and basic helix-loop-helix (bHLH) transcription factors (TF) play a crucial role in regulating neurogenesis (Paridaen and Huttner, 2014). For instance, the homeodomain TF PAX6 is considered one of the master regulators of neurogenesis and is widely expressed in RGCs, often used as a marker of the VZ (Englund et al., 2005; Osumi et al., 2008). PAX6 is necessary for controlling the balance between symmetric and asymmetric divisions of RGCs by forming transcription complexes with other TFs that are either pro-proliferative (such as HES1) or pro-neurogenic (such as ASC1 and NEUROG2) (Asami et al., 2011; Osumi et al., 2008; Sansom et al., 2009). In addition to regulating gene expression, PAX6 is also involved in spindle orientation and can regulate cell cycle reentry after mitosis (Asami et al., 2011). The best-known transcriptional targets of PAX6 are the neurogenins (NGN1/2) which are a group of bHLH TFs. *Ngn1/2* are transiently expressed in RGCs that divide asymmetrically to give rise to IPs. Thus, a

common model of TF expression that underlies the RGC-IP-neuron transition is that RGCs express *Pax6* which in turn drives expression of *Ngn1/2* and subsequently *Tbr2*. TBR2 then drives expression of neuronal TFs such as *Tbr1*, which is expressed by newborn neurons (Englund et al., 2005; Manuel et al., 2015). This model suggests that TF expression is sequential and thus possibly may underlie the temporal order in which RGCs give rise to the incredible array of neurons in the cortex. Nevertheless, other evidence has shown that proliferative and neurogenic TFs are expressed in the same cell at the same time but in oscillatory manner, until one of the TFs becomes dominant and thus instigates the fate choice of that progenitor (Imayoshi et al., 2013). Therefore, it is possible that the oscillatory expression of several proliferative and neurogenic TFs is important for progenitor multipotency, however the decision for one type of TF to dominate is likely driven by other internal or external influences that work synergistically to tip the balance in favour of a particular cell fate.

1.6.5. Epigenetic and post-transcriptional regulatory mechanisms

Numerous forms of epigenetic and post-transcriptional regulatory mechanisms, including DNA methylation, histone modification, and microRNAs have also been shown to control gene expression during neurogenesis. During early stages of neurogenesis, RGCs display high expression of chromatin regulator proteins such as the HMG family that decreases over the course of neurogenesis (Hirabayashi et al., 2009; Kishi et al., 2012). It is typically considered that RGCs in early neurogenesis phases possess more open chromatin however it becomes more condensed overtime, potentially reflecting the change of RGCs from multipotency to strict cell fates (Kishi et al., 2012). PAX6 is also able to interact with BAF155 and BAF170 which form the mSWI/SNF nucleosome remodeling complex that can repress expression of *Tbr2, Cux2,* and *Tle2* thus delaying neurogenesis (Tuoc et al., 2013). Several microRNAs have also been shown to bind to the 3' untranslated region (UTR) of mRNA targets to regulate neuron production. For example, miR-92 can exert a pro-proliferative effect on RGCs by suppressing expression of *Tbr2* (Bian et al., 2013; Nowakowski et al., 2013). The let-7 miRNAs can also suppress the expression of cell cycling regulators such as *Ccnd1, Cdc25a*, and proneuronal genes such as *Ngn1* and *Ascl1* (Cimadamore et al., 2013).

1.6.6. Signalling pathways

As previously mentioned, there are many signalling pathways that are known to be crucial for neurogenesis and patterning. Arguably the most important extrinsic modulators of neurogenesis, the signalling pathways of Notch, sonic hedgehog (Shh), fibroblast growth factor (Fgf), and

wingless-related integration (Wnt) have been studied extensively (Espinós et al., 2022; Götz and Huttner, 2005; Paridaen and Huttner, 2014). Moreover, many of these pathways undergo considerable crosstalk and employ some of the same signalling molecules. Importantly, many of the intrinsic mechanisms that regulate neurogenesis that were discussed previously, such as mitotic spindle orientation, cell cycle length, and activation of transcription factors, act downstream of these signalling pathways. The Wnt signalling pathway was a particular interest of this thesis and will be described in further detail below, as well as the other major signalling pathways.

1.6.6.1. Notch

One of the major signalling pathways involved in regulating neurogenesis during cortical development is the Delta/Notch pathway. In addition to its function during the NE-RGC transition (Gaiano et al., 2000), Notch signalling plays a well-documented role in binary fate decisions of neurogenic progenitors (Chitnis et al., 1995; Henrique et al., 1997). Cells expressing the Notch ligand, Delta (or Jagged) at the cell membrane can activate Notch receptors on adjacent cells, causing Notch to be cleaved by ADAM10 and γ-secretase. The cleavage of Notch releases the Notch intracellular domain (NICD) that is translocated to the nucleus and promotes the transcription of the *Hes* bHLH TF family (*Hes1-5*), which in turn represses expression of pro-neural genes, such as *Ascl* and *Ngn1/2* (Hatakeyama et al., 2004). Therefore, the traditional view is that cells with high or low Notch activity commit to a proliferative or neurogenic cell fate, respectively.

Recent publications have illustrated that the asymmetric inheritance of specific Notch modulators and signalling molecules plays a significant role in cell fate decisions. During asymmetric divisions, daughter cells that inherit Delta and Mindbomb1 (a Delta protein modulator) are more likely to differentiate while their sister cell – which inherits the NICD – remains proliferative (Dong et al., 2012). Interestingly, it was also shown that daughter cells that inherit the apical or basal processes are more likely to commit to a neurogenic or proliferative fate, respectively. Inheritance of the apical process means that the daughter cell inherits apical polarity proteins, including PAR3, αPKC, Pard3, and Numb (Alexandre et al., 2010; Dong et al., 2012). These proteins have been shown to negatively regulate Notch signalling in the apical daughter cell, likely by sequestering Mindbomb1 unequally to the apical daughter during asymmetric division (Bultje et al., 2009; Liu et al., 2015; Tozer et al., 2017). Subsequently, Mindbomb1 supports Notch *trans*-activation by regulating levels of Delta in the apical daughter cell and triggering Notch cleavage in the basal daughter cell. Asymmetric activation of Notch signalling can also occur from the asymmetric inheritance of Sara-expressing endosomes (Coumailleau et al., 2009). Sara endosomes are

short-lived organelles that sequester components of the Notch pathway and are asymmetrically inherited by the basal daughter cell (Coumailleau et al., 2009; Kressmann et al., 2015). After mitosis, Notch signalling is activated in this cell which thus remains proliferative. Interestingly, Par3 was recently shown to traffic Sara endosomes to basal areas of dividing progenitors via the dynein motor complex (Zhao et al., 2021). Taken together, the asymmetrical inheritance of Notch components underpins a molecular signalling structure that is vital for regulating neurogenesis.

1.6.6.2. FGF

The fibroblast growth factors (FGFs) and respective receptors (FGFRs) also play a significant role in modulating cell fate decisions during neurogenesis (Vaccarino et al., 1999). Signalling occurs when FGFs and heparan sulfate proteoglycan binds to FGFRs, causing the dimerisation and autophosphorylation of the FGFR intracellular tyrosine kinase domain (Yayon et al., 1991). This is followed by phosphorylation of various downstream signalling proteins involved in the MPK, PI3K-AKT, PLCγ, and STAT pathways which can induce downstream gene expression (Goetz and Mohammadi, 2013; Ornitz and Itoh, 2015). Activated FGFRs are then downregulated by ubiquitination, endocytosis, and degraded via lysosomal degradation (Katzmann et al., 2002). The 23 FGF morphogens have shown to individually possess different affinities for the four FGFRs and can bind separately or synergistically to evoke a complex array of intracellular cascades. For instance, FGF2 and FGF8 binding to FGFR1-2 has shown to promote proliferation in RGCs by activating ERK and AKT (Vaccarino et al., 1999). In contrast, FGF15 promotes the expression of the proneural gene CoupTF in progenitors which represses proliferation and drives neurogenesis (Borello et al., 2008). Therefore, different FGF ligands can inhibit or promote neurogenesis by modulating internal signalling responses.

1.6.6.3. Shh

Sonic Hedgehog (Shh) signalling is another pathway that is heavily involved in cortical development events. Shh signalling was first examined in *Drosophila* where it was found to be integral for dorsoventral patterning of the CNS (Echelard et al., 1993). This was further shown to be the case in mammalian development as well, however vertebrates also possess additional sets of Shh genes such as desert hedgehog (Dhh) and Indian hedgehog (Ihh) (De Luca et al., 2016). Signalling occurs when Shh peptides are cleaved via autoproteolytic processing, producing an N-terminus fragment with a signalling domain (Shh-N) and a C-terminus fragment devoid of signalling potential (Shh-C) (Dessaud et al., 2008). Shh-N is secreted and binds to the Patched1 receptor (Ptch1) at the primary cilium, which attracts and accumulates intracellular Smoothened

(Smo) to the cilium (Kang et al., 2007; Tenzen et al., 2006). Smo typically phosphorylates the Gli1-3 TFs, however its accumulation at the cilium stablises Gli1-3, which then translocate to the nucleus to activate (Gli1-2) or repress (Gli3) Shh gene targets (De Luca et al., 2016). Shh regulates RGC division behaviours primarily by promoting proliferation (Cayuso et al., 2006; Wang et al., 2011). Likewise, the gradual decrease in Shh signalling in the cortex coincides with progenitors undergoing neurogenesis. However, it is known that Shh signalling is also important for promoting RGC-IP divisions as well (Shikata et al., 2011) and has a function later in development during gliogenesis. Shh signalling exhibits considerable crosstalk with the other signalling pathways. For example, Shh activity is known to promote proliferation in progenitors by driving transcription of *Hes1*, a downstream target of Notch signalling (Dave et al., 2011). Moreover, Shh and Wnt undergo considerable crosstalk during the patterning of the cortex. Wnt morphogens are secreted from the dorsal telencephalon which patterns adjacent RGCs to give rise to the hippocampus (Backman et al., 2005). Similarly, Shh-N is secreted from the ventral telencephalon and the combined interaction between Wnt and Shh along the cortex is necessary for promoting expression of TFs that mediate neurogenesis (Mallamaci and Stoykova, 2006).

1.6.6.4. Wnt

The Wnt signalling pathway is a highly conserved and is known to play a vital role in cortical development and neurogenesis. As aforementioned, Wnt morphogens are secreted very early in development and are vital for the patterning of the dorso-ventral and anterior-posterior axes (Harrison-Uy and Pleasure, 2012). However, later in development, Wnt plays a significant role in orchestrating proliferation and neurogenesis in RGCs (Chenn and Walsh, 2002). Wnt signalling encompasses multiple pathways that are broadly categorised as canonical (β-catenin dependent) and non-canonical (β-catenin independent). Canonical Wnt signalling occurs through the binding of Wnt ligands to the receptors Frizzled (Fzd) and LDL receptor related protein (LRP)5/6 on the cell membrane (Bhanot et al., 1996; Tamai et al., 2000). This interaction phosphorylates the intracellular domain of LRP5/6 which recruits Disheveled (Dvl) and in turn inhibits the activity of the intracellular β-catenin phosphodestruction complex (composed of Axin, GSK3-β, CK1, APC, Dvl, and other proteins) (Chen et al., 2003; Davidson et al., 2005; Mao et al., 2001; Zeng et al., 2005). The inhibition of the destruction complex enables β-catenin to accumulate and translocate to the nucleus where it forms a transcription complex with TCF/LEF and promotes expression of target genes (Barker, 2008; Clevers and Nusse, 2012).

On the other hand, non-canonical Wnt signalling is far more diverse and involves a range of different mechanisms including Wnt-planar cell polarity (Wnt-PCP) and Wnt-Ca^2 pathways

(Gómez-Orte et al., 2013). Typically, a non-canonical pathway is categorised as such if Wnt ligands trigger an intracellular cascade without involving β-catenin. For example, Wnt ligands have been shown to interact with the membrane protein receptor-like tyrosine kinase (RYK) via its Wnt inhibitory factor (WIF) domain. This interaction instigates the proteolytic cleavage of RYK, and its intracellular domain is translocated to the nucleus to drive gene expression (Inoue et al., 2004). Likewise, another Wnt signalling pathway was recently described that involves the stablisation of other targets of the GSK3-β phosphodestruction complex aside from β-catenin (Acebron et al., 2014). The Wnt stablisation of proteins (WNT/STOP) pathway has been shown to orchestrate various Wnt cascades that occur at the protein level instead of altering transcription (Da Silva et al., 2021). Therefore, Wnt signalling is incredibly diverse, and these pathways have been shown to converge and coordinate cell fate decisions, especially during cortical development.

The overall role of canonical Wnt signalling in cortical development is complex. For example, canonical Wnt signalling has been found to be necessary for promoting proliferation of NEs and RGCs (Chenn and Walsh, 2002). This occurs through the transcription of Wnt target genes such as *Ccnd1/2,* which are cyclin genes that promote cell cycle re-entry (Machon et al., 2003; Mutch et al., 2010; Woodhead et al., 2006a). However, canonical Wnt signalling has also been documented to promote asymmetric divisions in progenitors by driving expression of pro-neural genes, such as *N-myc* and *Ngn1/2* (Hirabayashi et al., 2004; Munji et al., 2011). Based on this data, it was hypothesised that Wnt signalling is necessary for promoting self-renewal of RGCs during early stages of neurogenesis, however at later stages it switches to become pro-neuronal (Paridaen and Huttner, 2014). Likewise, it was also theorised that Wnt signalling may promote proliferation in RGCs but also promote neurogenesis in IPs (Munji et al., 2011). Nevertheless, it is important to note that the pro-proliferative effects of canonical Wnt signalling have mostly been observed *in vivo*, whereas its pro-neuronal effects have been observed *in vitro*. Furthermore, the pro-neuronal effects of Wnt signalling have yet to be demonstrated *in vivo*. One suggested explanation to these inconsistent results is that the manipulation of the canonical pathway can exert different effects depending on the epistatic level at which the pathway is manipulated. For example, overexpression of *β-catenin* causes over-proliferation of RGCs and delays neurogenesis, whereas *β-catenin* KD causes premature neurogenesis (Machon et al., 2003; Woodhead et al., 2006b). Interestingly, ablation of LRP6 caused RGCs to exhibit normal proliferation rates but decreased neurogenesis onset (Zhou et al., 2006). Further complications arise when considering that multiple molecules upstream of β-catenin in the Wnt signalling pathway are employed in noncanonical pathways. For example, the inhibition of GSK3-β by LRP6

has been shown to stablise other proteins aside from β-catenin, such as SOX4 and SOX11, which are necessary for promoting RGC-IP divisions and neurogenesis (Da Silva et al., 2021). Importantly, these signalling effects occur without affecting β-catenin activity. Therefore, it is likely that canonical Wnt signalling regulates neurogenesis in progenitors by working in conjunction with noncanonical pathways.

Another interesting noncanonical Wnt signalling pathway shown to be involved in cortical neurogenesis is the Wnt/RYK pathway. RYK is an atypical member of the tyrosine kinase family, exhibiting an extracellular WIF domain, transmembrane, and intracellular domain (Fradkin et al., 2010; Patthy, 2000). The intracellular domain resembles an abnormal kinase domain that lacks residues that would enable phosphorylation activity. Thus, the RYK intracellular domain exhibits no kinase activity and is sometimes referred to as a pseudo-kinase domain (Inoue et al., 2004; Yoshikawa et al., 2001). Instead, the RYK intracellular domain is used as a signalling molecule after Wnt-induced proteolytic cleavage by ADAM10 and possibly γ-secretase (Green et al., 2014). Wnt/RYK signalling has previously been associated with promoting cell polarity, migration, axon guidance, and neurite outgrowth (Keeble and Cooper, 2006; Lu et al., 2004; Macheda et al., 2012). However, a report by Lyu et al. (2008) first demonstrated that at E11, RYK is expressed widely throughout the VZ. Interestingly, RYK was localised primarily at the membrane of RGCs but showed a more nuclear distribution in neurons of the CP (Lyu et al., 2008). Further experiments revealed that RYK is proteolytically processed and that the RYK intracellular domain is transported to the nucleus to drive neurogenesis. Later studies revealed that the RYK intracellular domain is transported to the nucleus by SMEK1/2 and promotes the expression of various pro-neural genes such as *Dlx1/2* and *NeuroD1* (Chang et al., 2017b). It was also recently shown in the neural progenitors of zebrafish that Ryk can form a protein complex with various δ-protocadherins, including Pcdh19 (Biswas et al., 2021). Interestingly, KD of the δ-protocadherins increased Wnt/β-catenin signalling and caused hyper-proliferation of neural progenitors and delayed neurogenesis. This suggested that δ-protocadherins may inhibit Wnt/RYK signalling to modulate neurogenic divisions of progenitors. The role of cadherins and protocadherins in regulating neurogenesis will be discussed in its own section below.

1.6.7. Cadherins and protocadherins in cortical neurogenesis

Multiple cadherins/protocadherins have been identified as potential regulators of neurogenesis (Yamagata et al., 2020). The most studied cadherin in this context is N-cadherin. N-cadherin is known to directly interact with β-catenin to keep it sequestered at the cell membrane, therefore limiting the amount of β-catenin in the cytoplasm available for Wnt signalling (Linask et al., 1997;

Sakane and Miyamoto, 2013). Likewise, it has also been argued that N-cadherin can regulate β-catenin levels by forming a protein complex with LRP5 and coupling β-catenin to the GSK3-β destruction complex (Haÿ et al., 2009; Maher et al., 2009). N-cadherin KD has also been shown to cause cytosolic levels of β-catenin to increase, which causes RGCs within the ventral midbrain to over-proliferate (Sakane and Miyamoto, 2013). Therefore, it is generally considered that N-cadherin negatively regulates Wnt signalling to modulate RGC proliferative and neurogenic divisions. However, other evidence has demonstrated that N-cadherin may also stablise β-catenin by interacting with LRP6 and activating AKT, which subsequently phosphorylates and stimulates β-catenin signaling (Zhang et al., 2013).

N-cadherin has also been shown to regulate neurogenesis through other signalling pathways. For example, N-cadherin plays an intricate role in controlling the gradual detachment of neurogenic daughter cells from the apical surface (Hatakeyama et al., 2014). Likewise, N-cadherin performs a cooperative feedback mechanism with Notch wherein the detaching daughter cells would activate Notch signaling in neighboring progenitors before delaminating and differentiating (Baek et al., 2018; Hatakeyama et al., 2014). N-cadherin can also interact with FGFR1, stabilising the receptor and preventing its ubiquitination, and thus prolonging FGF signalling, which has been shown to be important for promoting neurogenesis in mouse embryonic stem cells (mESCs) (Punovuori et al., 2019; Takehara et al., 2015). Lastly, some evidence has shown that N-cadherin may have a small regulatory role in Shh signalling within telencephalic and spinal progenitors, however the exact mechanism is not well understood (Chalasani and Brewster, 2011).

Several protocadherins have also been linked to neurogenesis. For example, PCDH11x is widely expressed in the VZ and SVZ, and has been shown to exert a pro-proliferative function in progenitors (Zhang et al., 2014). As previously discussed, it was recently reported that KO of various δ-protocadherins in zebrafish, including *pcdh1a, pcdh7a, pcdh9, pcdh17, pcdh18b*, and *pcdh19*, caused an increase in neural progenitor proliferation and delayed neurogenesis onset (Biswas et al., 2021). Likewise, the hyper-proliferative phenotype was linked to an increase in Wnt signalling, as demonstrated by a significant upregulation of Wnt signalling molecules, including *axin2* and *lef1*. Interestingly, it was revealed that the δ-protocadherins could directly interact with Ryk and form a protein complex on the cell membrane. KD of *ryk* in *pcdh* mutants rescued the over-proliferative phenotype of the progenitors, suggesting that the protocadherins negatively regulate the Wnt/Ryk signalling (Biswas et al., 2021).

Taken together, cadherins and protocadherins are important regulators of cortical neurogenesis and likely work synergistically with various signalling pathways and internal regulatory

mechanisms to accomplish this. However, the exact mechanisms on how protocadherins regulate neurogenesis are unknown. Interestingly, there is increasing attention towards PCDH19 as a regulator of neurogenesis (Biswas et al., 2021; Fujitani et al., 2017; Homan et al., 2018) which will be discussed separately.

1.7. PCDH19 and neurogenesis

1.7.1. Previous publications

As previously discussed, PCDH19 is widely expressed in the progenitors of the developing cortex between E10-E12 but starts to decrease by E13 (Dibbens et al., 2008; Gaitan and Bouchard, 2006; Hertel and Redies, 2011). Interestingly, the onset of cortical neurogenesis also occurs around this time in mice, highlighting that the limited time window of *Pcdh19* may be linked to neurogenesis onset (Götz and Huttner, 2005). Indeed, various studies have demonstrated a link between PCDH19 and neurogenesis. For example, it has been shown that *Pcdh19* KO neural progenitors display increased neurogenesis output at the expense of proliferation *in vitro* (Homan et al., 2018). iPSC-derived progenitors from PCDH19-epilepsy patients also displayed increased neurogenesis (Homan et al., 2018). On the other hand, short-harpin (sh)RNA-induced KD of *Pcdh19* was found to increase RGC proliferation to delay neurogenesis (Fujitani et al., 2017). Here, it was also found that miR-484 can delay IP production by interfering with *Pcdh19* expression by binding to its 3' UTR region (Fujitani et al., 2017). Furthermore, *pcdh19* KO was shown to increase neural progenitor proliferation *in vivo* in a zebrafish model (Biswas et al., 2021). Taken together, increasing evidence suggests that PCDH19 is a regulator of neurogenesis.

Nevertheless, there are several limitations of the previous publications that must be addressed. For instance, much of the research that has linked PCDH19 to neurogenesis has been performed *in vitro* (Homan et al., 2018). Culturing cortical progenitors means that the complex structural architecture of the brain is lost. This is especially relevant since PCDH19 is an adhesion molecule and thus requires a complex tissue environment to exert many of its functions. This may also explain why studies have reported contradicting results, either suggesting that KO/KD of *Pcdh19* decreases or increases neurogenesis in progenitors (Fujitani et al., 2017; Homan et al., 2018). Recently, researchers crossed the *Emx1-CreER*^{T2} mouse line with the *Tbr2*$^{fl/fl}$ line to generate conditional KO (cKO) of *Tbr2* in cortical progenitors (Lv et al., 2019). After further crossing these mice into the MADM line, the researchers were able to show that *in vivo* removal of *Tbr2* in RGCs significantly decreased their individual neuronal outputs and caused individual neurons to become more laterally dispersed within the cortex. Interestingly, analysis involving CHIP-seq and CHIP-

seq-qPCR suggested that TBR2 binds upstream and downstream of the endogenous *Pcdh19* genomic locus, and positively regulates its expression. KD of *Pcdh19* in the developing cortex was shown to decrease RGC individual neuronal outputs and cause a similar lateral dispersion phenotype as observed in the *Tbr2* cKO mice, suggesting *Pcdh19* is important for regulating RGC neuronal production and neuronal positioning. Finally, retroviral-induced upregulation of *Pcdh19* in the *Tbr2* cKO mice was shown to significantly increase neurogenesis rates and normalise neuronal positioning in the cortex (Lv et al., 2019). Taken together, these data suggest a mechanism wherein TBR2 mediates neurogenesis in RGCs by regulating the expression of *Pcdh19*, however the mechanism in which PCDH19 controls neurogenesis in RGCs was not explored. Moreover, concentrated efforts to explore whether PCDH19 regulates neurogenesis *in vivo* have yet to be published. Work at the IMG lab has begun to unravel the role of PCDH19 involvement in cortical neurogenesis which will be discussed below.

1.7.2. Research from the IMG lab

To investigate this further, research studying the role of PCDH19 during mouse cortical development was performed by Dr. Jessica Griffiths in the IMG lab (unpublished data, 2019). Using WT embryos, the spatiotemporal expression of *Pcdh19* was examined using ISH between E11-E14. Likewise, ISH was also combined with IHC using antibodies against PAX6 (RGCs), TBR2 (IPs), and TBR1 (neurons) (Englund et al., 2005). ISH analysis found that *Pcdh19* is widely expressed in the lateral cortex at E11 and E12, however expression recedes into medial regions from E13 until expression in the cortex is diminished by E14. Interestingly, ISH and IHC analysis showed that expression of *Pcdh19* was primarily limited to the VZ. Taken together, these results confirmed that *Pcdh19* was expressed in the cortex in a temporally regulated fashion, which supported the theory that PCDH19 has a specific function in the RGCs at the time of the neurogenic switch.

To investigate whether mosaic expression affected neurogenesis, the *Pcdh19* mutant mouse line from Taconic (Pederick et al., 2016) was crossed with WT mice to generate embryos that were WT, HET, or KO for *Pcdh19* (unpublished data, 2019). By performing IHC using β-GAL antibodies, the cell segregation phenotype was observed in HET samples at E11 and E12, confirming the previously reported phenotype (Pederick et al., 2018) (see Figure 1.4B). Importantly, this also allowed for the quantification of specific cell markers between PCDH19-WT and PCDH19-KO cells within the HETs to determine whether mosaic expression affected neurogenesis. To examine this, IHC analysis of the mitotic fraction (phosphohistone H3 staining), S-phase fraction (2hr EdU labeling + KI67 staining), and cell cycle quitting fraction (24hr EdU labelling - KI67

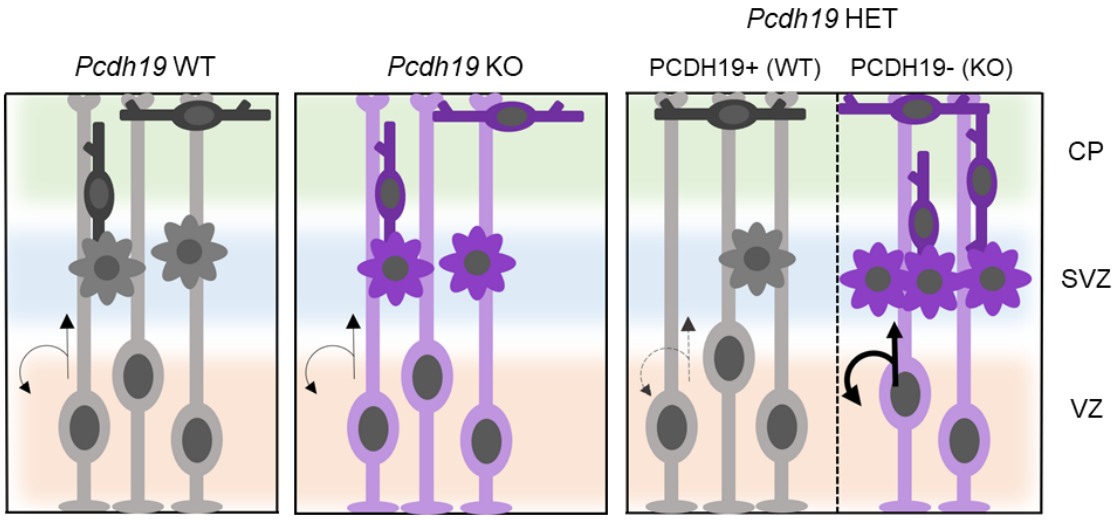
staining) was performed (see Figure 1.6B). Interestingly, homozygous WTs and KOs showed no differences in the fraction of cells in mitosis, S-phase, or exiting the cell cycle. Collectively, HETs also showed no differences compared to WTs and KO – however, when examining differences **within** the HETs, PCDH19-WT cells showed reduced fraction of cells in mitosis, S-phase, and exiting the cell cycle, whereas PCDH19-KO cells had increased fractions (see Figure 1.6B). Moreover, when examining the proportion of RGCs, IPs, and neurons, no differences were found between WT, KO, and HETs collectively. However, within the HETs, there was an increase in the proportion of PCDH19-WT RGCs and a decrease in the proportion of IPs and neurons, whereas there was a decrease in the proportion of PCDH19-KO RGCs and an increase in the proportion of IPs and neurons (see Figure 1.6C). Therefore, these results provided firsthand evidence that PCDH19 is involved in cortical neurogenesis *in vivo*, and that perturbed expression of *Pcdh19* alters neurogenesis rates but **only** in the developing HET brain.

1.7.3. Conclusions and further study

Taken together, previous publications and research from the IMG lab have demonstrated that PCDH19 is a novel regulator of neurogenesis, both *in vitro* and *in vivo* (Fujitani et al., 2017; Homan et al., 2018; unpublished data, 2019). Interestingly, whereas *in vitro* it was demonstrated that *Pcdh19* KO causes increased neurogenesis in progenitors, this was not replicated *in vivo*, wherein only the HETs showed altered neurogenesis (Homan et al., 2018; unpublished data, 2019). This illustrates that the structural integrity of the brain may be for PCDH19 to coordinate neurogenesis and the mechanism by which it does this is likely disturbed when culturing cells *in vitro*. Moreover, this result may also suggest that complete loss of PCDH19 instigates a compensatory mechanism to ensure correct progression of neurogenesis. This may also explain why hemizygous carriers of the *PCDH19* mutant allele are spared from displaying PCDH19-epilepsy symptoms. Nevertheless, a crucial question that has yet to be addressed is how PCDH19 influences neurogenesis. As discussed previously, Pcdh19 has been recently shown to regulate progenitor neurogenesis behaviours during zebrafish development by operating through the Wnt/Ryk signalling pathway (Biswas et al., 2021). However, this has not been explored in mammals and a detailed mechanism was not examined. A transcriptome analysis between *Pcdh19* WT and KO cultured progenitors was recently performed using microarrays and illustrated various dysregulated genes and pathways related to neurogenesis (Homan et al., 2018). However, as detailed above, perturbed expression of *Pcdh19* in progenitors produces different neurogenic effects if examined *in vitro* or *in vivo* and sometimes produces different consequences between separate *in vitro* systems, likely because of losing the complex architecture of the brain (Fujitani

et al., 2017; Homan et al., 2018; unpublished data, 2019). Therefore, examining the mechanisms of how PCDH19 regulates neurogenesis *in vivo* is vital to elucidate the role of PCDH19 during brain development, both in health and disease contexts.

A



B

Experiment	Genotype			
%	WT	KO	HET (WT)	HET (KO)
Mitosis	2.9	2.7	1.6	4.3
S-phase	46.3	46.8	46.8	56.1
Quitting	11.7	12.2	5.1	16.5
RGCs	87.3	86.4	89.6	80.5
IPs	21	22.4	10.9	26
Neurons	9.7	9.6	7.4	11.5

Figure 1.6. Differences in cortical neurogenesis between *Pcdh19* WT, HET, and KO animals. (A) Illustration of cortical neurogenesis events in WT, KO, and HET cortices. Figure also represents the subtle differences in neurogenesis between WT and KO cells within the HETs. Arrows indicate the rate of the progenitor cell cycle; grey-dashed arrows indicate a slower cell cycling, thick black arrows indicate an increase in cell cycling. (B) Table summarising the main findings of Dr. Jessica Griffith's thesis involving the role of PCDH19 in neurogenesis. Coloured numbers represent statistically significant results compared to all animals (red = significant decrease, green = significant increase). The original data can be found in Appendix 1.

1.8. Hypothesis and aims

The work conducted by Dr. Jessica Griffiths from the IMG lab has presented novel evidence suggesting that PCDH19 plays a role in regulating murine cortical neurogenesis *in vivo*, with mosaic expression of *Pcdh19* disrupting this process (unpublished data, 2019). However, the precise mechanism by which PCDH19 influences neurogenesis remains unknown. Moreover, the mechanism that drives the disrupted neurogenesis in the HET animals and the normal neurogenesis in the KO animals is also unknown. The plethora of molecular mechanisms that regulate cortical neurogenesis means there could be many avenues in which PCDH19 may operate. Therefore, a focused investigation is necessary to comprehend the involvement of PCDH19 in cortical neurogenesis by examining the molecular architecture of the cortex of these animals in addition to assessing the role of PCDH19 in candidate molecular pathways.

The fundamental hypothesis of this thesis is that PCDH19 plays an important role in murine cortical neurogenesis by functioning through an undetermined molecular mechanism. Moreover, this mechanism is disrupted in HET animals due to mosaic expression, and potentially compensated in KO animals. It is hypothesised that identifying key molecular differences between WT, HET, and KO embryos will elucidate specific pathways that can highlight the molecular role of PCDH19 in cortical neurogenesis and will shed light on the cause of the neurogenesis phenotypes in the HET and KO animals. Moreover, identifying differences between the WT and the KO cells **within** the HET animals will also aid in elucidating this mechanism. Finally, based on previous links between protocadherins and the Wnt signalling pathway, it is also hypothesised that PCDH19 may play a role in this candidate signalling pathway which has an important impact on cortical neurogenesis. Taken together, the following research aims were generated:

- Examine and compare the transcriptional landscape of E11 WT, HET, and KO cortical tissue from mice using bulk RNA-seq as a proxy for assessing molecular differences between these animals.
- Assess the transcriptional landscape of E11 WT, HET, and KO cortical tissue from mice at the single cell level using scRNA-seq and analyse the profiles of WT and KO cells **within** the HET animals.
- Analyse the role of PCDH19 in the Wnt signalling pathway and determine whether the role of PCDH19 in the Wnt pathway is important for Wnt-mediated cortical neurogenesis in mice.

Chapter 2: Material and methods

2.1. Animals

2.1.1. Husbandry

Experiments were performed in accordance with the local ethical and Home Office approval under the UK Animals (Scientific Procedures) Act 1986. All animals were housed in diurnal light conditions (12 hours light, 12 hours dark) and received food and water *ad libitum*. Nesting, material, carboard and plastic tubes, wooden chew-sticks, and sunflower seeds were provided in all home-cages. Mice were weaned four weeks post birth and kept in cages of no more than five animals per cage. For timed pregnancies, female mice were housed with a male mouse overnight. The next morning, the female was checked for a coagulated seminal plug and was considered embryonic day (E) 0.5 at noon of the day the plug was found.

2.1.2. Mouse lines

Pcdh19 knock-out (KO) mice were purchased from Taconic Biosciences (TF2108). The *Pcdh19* KO mice were characterised by the replacement of exons 1-3 of the *Pcdh19* allele by a β-galactosidase (β-gal)/neomycin (neo) reporter cassette under the control of the endogenous *Pcdh19* promoter. This mouse line has been rigorously characterised both anatomically (Pederick et al., 2016) and behaviourally (Galindo-Riera et al., 2021).

D4/XEGFP (XGFP) mice were kindly provided by Prof. Ros John's research group at Cardiff University. These mice carry an X-linked enhanced green florescent protein (*EGFP*) transgene under the control of a *chicken β-actin* promoter and have been widely used to study early sexual dimorphism effects in embryos (Hadjantonakis et al., 1998).

TCF/Lef1-HIST1H2BB/EGFP (TCF/Lef:H2B-GFP) mice were purchased from the Jackson laboratory (RRID: IMSR_JAX:013752). These mice carry a modified GFP cassette that harbours the histone H2B sequence (H2B-GFP). The H2B-GFP fusion localises GFP to the nucleus by binding to chromatin. Expression of the cassette is under the control of six *T cell-specific transcription factor/lymphoid enhancer binding factor 1* (*TCF/Lef*) response elements and the *Hsp68* minimal promoter. As such, the mouse model has been widely used to study Wnt signalling during embryogenesis (Ferrer-Vaquer et al., 2010).

All C57BL/6J mice were purchased from Charles River Laboratories. Moreover, all mouse lines were maintained on a B6 background.

2.1.3. Genotyping

Tail samples from embryonic mice and ear-notches from adult mice were obtained and stored at -20°C for genotyping. Crude genomic DNA was extracted from the samples using the Mouse Direct PCR kit (Biotool, B4001). Briefly, pre-made lysis buffer (Buffer L) and Protease Plus was added to the samples. The samples were incubated at 55°C for 15 minutes for effective digestion. The samples were then incubated at 95°C for 5 minutes to inactivate the protease. The digested solution was used directly as the DNA template for the polymerase chain reaction (PCR).

For the PCR reaction, a solution containing 2x M-PCR OPTI™ mix (obtained from the Mouse Direct PCR kit, containing dNTPs, $MgCl_2$, and Taq DNA polymerase) distilled water (ddH20), specific primers (10 µM), and the DNA template, were combined and loaded into a T100 Thermal Cycler (BIORAD). Primer pairs, annealing temperatures, and amplicon sizes are shown in **Table 2.1**. PCR products were then stored at 4°C until separated on a 1-2% agarose gel via electrophoresis and visualised using ethidium bromide.

To genotype for *Pcdh19* wild type (WT) and knockout (KO) alleles, two PCR reactions were used as previously described (Pederick et al., 2016). Briefly, one PCR reaction was WT specific and absent in the KO allele. Moreover, another PCR reaction was mutant specific and absent in the WT allele. Therefore, by using both PCR reactions, WT (WT present, KO absent), KO (KO present, WT absent), and HET (WT present, KO present) genotypes could be determined. Additionally, to genotype XGFP animals and TCF/Lef1:H2B-GFP animals, PCR reactions were used to target their specific transgenes, as recommended by the Jackson laboratory (Protocol 29499 and Protocol 29498, respectively).

Sex determination of embryos was also conducted using genotyping PCR, as described previously (Clapcote and Roder, 2005). The PCR detected the X-chromosome-specific gene *Jarid1c* and the Y-chromosome-specific *Jarid1d*. Both genes have homologous exons however intron 9-10 is 29 base pairs (bp) longer in *Jarid1c* than *Jarid1d*. Therefore, the reaction is designed to amplify DNA fragments of 331 bp from X-specific *Jarid1c* and 302 bp from Y-specific *Jarid1d*. When visualised on an 2% agarose gel, females and males would be identified as having one or two band(s), respectively.

Table 2.1. Genotyping primers and annealing temperatures

Primer Name	Fwd/Rev	Annealing Temperature (°C)	Product Size (bp)	Primer Sequence (5' – 3')
***Pcdh19* WT Primers**				
TF2108-F2	Fwd	56.5°C	123 bp	TAGAGGTTCTTGCTGAAGACTTCC
TF21-8-R2	Rev			TCAACTGTTTCGATGAGACACTGC
***Pcdh19* KO Primers**				
TF2108-10	Fwd	57.2°C	437 bp	GTGCGTACCAGGCGGGAGC
GT-IRES	Rev			CCCTAGGAATGCTCGTCAAGA
XGFP Primers				
oIMR0872	Fwd	63.8°C	324 bp	CCTACGGCAAGCTGACCCTGAAG
oIMR1416	Rev			GTCGCCGATGGGGGTGTTCTG
TCF/Lef1-GFP Primers				
11775	Fwd	60°C	530 bp	ACAACAAGCGCTCGACCATCAC
11776	Rev			AGTCGATGCCCTTCAGCTCGAT
Sex Determination Primers				
Jardid1c	Fwd	54°C	331 bp	GTCAAGCTTTTGGCTTTGAG
Jardid1d	Rev		302 bp	CCACTGCCAAATTCTTTGG

2.2. Tissue processing

2.2.1. Fixing of embryonic samples for sectioning

To acquire E11.5 embryos, male and female mice were mated and plug-checked the following morning, as described previously (see Section 2.1.1.). At E11.5, pregnant mice were terminated via schedule 1 and the embryos were extracted from the uterine horns. The heads of the embryos were dissected and fixed in 4% paraformaldehyde (PFA) made in 1X phosphate-buffer saline (10X PBS; NaCl 1.37 M, KCL 27 mM, Na_2HPO_4 100 mM, KH_2PO_4 18 mM). The heads were fixed for 4-6 hours at 4°C after which the PFA was removed and replaced with 1X PBS.

2.2.2. Sectioning

Fixed E11.5 heads were cryoprotected in 1X PBS with 30% sucrose and incubated overnight at 4°C. The heads were then frozen in Optimal Cutting Medium (OCT; Thermofisher, 23730571) and stored at -80°C. Tissue was then sectioned at 50 µM using a cryostat (Leica Biosystems, CM3050) and mounted onto poly-lysine microslides (VWR, 6311560). Slides were then kept at -80°C until needed.

2.3. Immunohistochemistry (IHC)

Slides were removed from -80°C and washed three times with 1X PBS to rehydrate the sections. The sections were then washed again with 1X PBS (2x 5-minutes) and then with 1X PBS with 0.25% Triton X100 (PBS-T; 3x 5-minutes) to permeabilise the cell membranes. After this, the sections were incubated in blocking solution (1X PBS, 4% bovine serum albumin (BSA), 3% donkey serum, 1% Triton X100) for 1 hour at room temperature (RT). The blocking solution was then removed, and the sections were incubated in primary antibody made in blocking solution and kept overnight at 4°C. Details of the primary antibodies are provided below (see Table 2.2) The next day, sections were washed with PBS-T (3x 5-minutes) and then incubated in secondary antibody in blocking solution for 2 hours at RT. Details of the secondary antibodies are provided below (see Table 2.3). After this, the sections were washed again in PBS-T (3x 5-minutes) and then counterstained with DAPI (1:3000 in 1X PBS) for 10 minutes at RT. The sections were finally washed with PBS (2x 5-minutes) and left to dry for ~15 minutes before adding coverslips using DAKO mounting medium.

2.4. Imaging

2.4.1. Acquisition

E11.5 brain sections were imaged using a confocal microscope (Carl Zeiss, LSM 780) with Zen Black software (Carl Zeiss, version 2.0). For IHC experiments, whole brain images were taken using a 20x objective.

2.4.2. Image analysis

Images were processed and analysed using imaging software FIJI. To analyse fluorescence intensity, individual nuclei were highlighted using the free draw function and the intensity of fluorescence was measured using the multi-measure function. Fluorescent intensity was also taken from an area devoid of tissue and the value was subtracted from nuclear measurements to account for background signal.

Table 2.2. Primary antibodies used for IHC

Antibody	Host	Type	Supplier	Cat. No.	Dilution
PAX6	Rabbit	pAb	Biolegend	PRB-278P	1:500
TBR1	Rabbit	pAb	Abcam	Ab31940	1:500
GFP	Chicken	mAb	Fisher Scientific	10524234	1:500
β-galactosidase (β-GAL)	Chicken	pAb	Abcam	Ab9361	1:500
pLRP6 (Thr1479)	Rabbit	pAb	Niehrs lab (da Silva et al. (2021))	In-house	1:200
pLRP6 (Ser1490)	Rabbit	pAb	Cell signalling technology	2568	1:200
pHH3	Rat	mAb	Abcam	Ab10543	1:300

Table 2.3. Secondary antibodies used for IHC

Antibody	Use	Dilution
Chicken Alexa-488	GFP	1:2000
Chicken Alexa-647	β-GAL	1:1000
Rabbit Alexa-555	PAX6, TBR1	1:2000
Rabbit Alexa-488	pLRP6 (Thr1479), pLRP6 (Ser1490)	1:1000
Rat Alexa-488	pHH3	1:1000

2.5. Cell culture

2.5.1. Mycoplasma testing

The Lookout Mycoplasma PCR detection kit (Sigma, MP0035) was used to routinely check for potential mycoplasma infection in all cell lines.

2.5.2. HEK293 cells

Human embryonic kidney 293 (HEK293) cells were cultured with Dulbecco's modified eagle medium (DMEM) supplemented with 10% heat-inactivated foetal bovine serum (FBS; Biosera, FB-1001/500-122551), 1% Glutamax (Gibco, 35050061), and 1% penicillin streptomycin (Gibco, 15140122). The cells were cultured in T75 Nunc flasks (Thermofisher, F7552) at a starting concentration of $1x10^6$ cells per 10 ml media and stored in a sealed HeraCell150i CO_2 incubator (Thermofisher). The incubator maintained a constant temperature of 37°C with 7% CO_2 and 100% relative humidity which provided the optimal environment for the cells to grow.

HEK293 cells were split every 3-4 days or when they had reached ~80% confluency. Briefly, the cells were placed in a Maxisafe 2020 Class II Biological Safely Cabinet which had previously been cleaned with 70% ethanol. The confluent media was removed and the cells were washed once with sterile 1X PBS. The cells were then incubated in TrypLE™ Express Enzyme (Gibco, 12604013) and kept at 37°C for 2 minutes. To quench the trypsinisation, the cells were washed in 10 ml media, collected into a 15 ml falcon, and centrifuged for 2 minutes at 1500 rotations per minute (rpm). The supernatant was then aspirated and the cells were resuspended in 1 ml fresh media. To calculate the number of cells for accurate cell seeding, a 1:10 diluent of cells was collected and quantified using the NucleoCounter® NC-200™ (Chemometec) automated cell counter, following the manufacturer's instructions. A working dilution of cells was then calculated with the concentration of cells being dependent on the type of experiment.

2.5.3. Cell transfection

HEK293 cells were transfected 24 hours after seeding using the Lipofectamine™ 2000 Transfection Reagent kit (Thermofisher Scientific), following the manufacturer's instructions. The lipofectamine transfection system involves using cationic lipids which form liposomes in aqueous solution. These positively charged molecules are capable of taking up negatively-charged plasmid DNA; they then bind onto the cell membrane and enter the cell via endocytosis. Once inside the cell, the lipid-based vesicles harbouring the plasmid DNA would burst open, due to the lowering

of the osmotic potential. The DNA is then unpacked by endogenous cell enzymes which allow the plasmid to freely diffuse into the nucleus, where it is expressed.

Plasmid DNA was mixed with lipofectamine reagent (1:2 ratio) and incubated in reduced serum media (OptiMEM®; Gibco, 31985062). The amount of DNA, lipofectamine, and OptiMEM® for each experiment is provided below (see Table 2.4). The mixture was prepared and each sample was vortexed and left to incubate at room temperature for 10 minutes to allow the DNA-liposome complexes to form. The transfection solution was then added to the corresponding plate/well in a dropwise manner. All plates were returned to the CO_2 incubator and left at 37°C in 7% CO_2 for 24 hours.

Table 2.4. Cell culture transfections with lipofectamine

Plate	Amount of DNA	Amount of OptiMEM®
6 well plate (Co-immunoprecipitation)	1 µg per plasmid (2 plasmids)	100 µl
6 well plate (Western blot)	1 µg	100 µl
24 well plate (Luciferase assay)	500 ng	10 µl

2.6. Plasmids and cloning

A detailed overview of the plasmids is provided below (see Table 2.5). Most plasmids used for transfection experiments were already available in the Martinez-Garay lab (Cardiff University) or were purchased. However, plasmids related to *Related to receptor tyrosine kinase* (*Ryk*) had to be developed for this project.

2.6.1. TOPO cloning

Ryk FL was first amplified from E11.5 cortical template cDNA via PCR (see Table 2.6) using Phusion® High-Fidelity DNA Polymerase (NEB, M0530S). The amplicons were then inserted into pCR™4-TOPO™ TA vector (Invitrogen, K4575J10) and transformed, following the manufacturer's instructions. Several *TOPO-Ryk* plasmid colonies were selected for miniprep purification (see Section 2.6.5) and sequenced via Sanger sequencing (see Section 2.6.6). Minipreps that were shown to have the correct *Ryk* sequence were subsequently used as template DNA for further cloning.

2.6.2. Amplification

Plasmids containing the FL or the intracellular domain (ICD) sequence of *Ryk* were amplified for cloning into the pcDNA3.1-HA vector (Addgene, 128034) using a standard Phusion® reaction for genes with high GC content. Briefly, ~50 ng *TOPO-Ryk* plasmid was used as the DNA template and mixed into Phusion® reaction solution (0.5 µM forward primer, 0.5 µM reverse primer, 200 µM dNTPs, 3% DMSO, 1X Phusion GC Buffer, 1 unit Phusion DNA polymerase). The PCR was then performed, and the modified amplicons were detected via agarose gel electrophoresis. An overview of the primers and annealing temperatures for all PCR reactions can be seen below (see Table 2.5). The amplicons were then extracted and purified using the QIAquick Gel Extraction Kit (QIAGEN, 28704), following the manufacturer's instructions.

2.6.3. Digestion and ligation

The amplicons and 2 µg pcDNA-HA vector were digested using appropriate restriction enzymes (see Table 2.6) for 4 hours at 37 °C. Both products were then separated via gel electrophoresis and purified, as mentioned previously (see Section 2.6.2). The amplicons and the vector were then subjected to ligation using a T4 DNA Ligase (NEB, M0202) for 4 hours at RT. The ligation product was then immediately transformed.

2.6.4. Bacterial transformation

Plasmids were transformed into DH5α *Escherichia coli* (*E coli*) chemically competent cells by mixing 50 µl DH5α cells with ~100-200 ng plasmid DNA or ligation product. The cells and DNA were then incubated on ice for 30 minutes to allow the DNA to localise close to the cell membrane. The cells were then placed in a water bath and heat shocked for 30 seconds at 42°C to initiate the opening of the cell membranes and allowing the DNA to enter the cells. The cells were placed back on ice for 2 minutes and then mixed with 950 µl Luria Bertani (LB; Sigma, L3522) media and incubated at 37°C for 1 hour to give the cells time to begin expressing the antibiotic resistance gene. Afterwards, the cells were centrifuged for 1 minute at 8000 rpm and most of the supernatant was removed. The cell pellet was then resuspended in the remaining media and plated on a LB Agar plate (Sigma, L3147) that was supplied with the appropriate antibiotic. The plates were then left to grow overnight at 37°C.

2.6.5. Miniprep

Viable bacteria colonies were selected from LB Agar plates by individually isolating the colonies using a sterile pipette tip. The colonies were then inoculated in 3 ml LB media supplemented with the appropriate antibiotic and left to incubate overnight in motion to stimulate bacterial aeration growth. The next day, bacterial samples were centrifuged at 8000 rpm for 3 minutes at RT and the plasmid DNA was extracted and purified using the QIAprep® Spin Miniprep Kit (Qiagen, 27106), following the manufacturer's instructions. To examine the sequence fidelity, 2 µl of each miniprep was digested using appropriate restriction enzymes for 2 hours at 37°C and subjected to agarose gel electrophoresis. Minipreps that showed the correct predicted band sizes were then sent for sanger sequencing to confirm sequence fidelity.

2.6.6. Sanger sequencing

Sangar sequencing of miniprep samples was performed by Eurofins Genomics UK. Approximately 100 ng/µl DNA was sent to Eurofins along with 1 pmol/µl sequencing primers if those primers were not already supplied by Eurofins in house. The chromatogram sequences were the analysed using FinchTV® software (version 1.4.0) and CodonCode Aligner® (version 9.0.2). Plasmid maps were also generated using SnapGene Viewer® (version 6.1).

2.6.7. Maxiprep

Plasmids that were confirmed to contain the correct sequences were further amplified using the EndoFree® Plasmid Maxi kit (Qiagen, 12362), following the manufacturer's instructions. Miniprep plasmids that were sequenced were transformed and plated on an Agar plate, as mentioned previously (see Section 2.6.4). A single colony was then selected and inoculated in 200 ml LB media supplemented with the appropriate antibiotic and left to incubate overnight at 37°C. The bacterial culture was then centrifuged, and the plasmid DNA was extracted using the aforementioned kit. Approximately 5-10 µg/µl plasmid DNA was extracted from each maxiprep.

Table 2.5. Plasmids used for cell culture experiments

Plasmid	Description	Source
pCBA-Pcdh19-FL-EGFP	Chicken-beta-actin promoter which drives expression of the full length *Pcdh19* gene with a C-terminal EGFP tag.	IMG
pcDNA-Ryk-FL-HA	CMV promoter which drives the expression of the full length *Ryk* gene, followed by a C-terminal HA tag.	IWJF
pcDNA-β-catenin-HA	CMV promoter which drives the expression of the full length *β-catenin* gene, followed by a C-terminus HA tag.	IMG
pCBA-N-cadherin-FL-EGFP	Chicken-beta-actin promoter which drives expression of the full-length *N-cadherin* gene with a C-terminal EGFP tag.	IMG
M50 Super 8x TOPFlash	Reporter plasmid containing a minimal *fos* promoter and six TCF/LEF binding sites upstream of a firefly luciferase gene.	Addgene (12456)
M50 Super 8x FOPFlash	Reporter plasmid containing a minimal *fos* promoter and six mutated non-functional TCF/LEF binding sites upstream of a luciferase gene.	Addgene (12457)

2.7. Co-immunoprecipitation

Tissue or cells were lysed in IP lysis buffer (20 mM Tris-HCl, 150 mM NaCl, 1% Triton X100, 1% protease inhibitor cocktail (Sigma), 1% phosphatase inhibitor cocktail (Sigma)). Samples were then incubated on ice for 10 minutes to properly lyse. The lysates were then centrifuged at 14,000 rpm for 10 minutes at 4°C, the cell pellet was discarded, and the supernatant was moved to fresh eppendorfs. In parallel, 10 µl Protein G Sepharose® beads (Abcam, ab193259) were washed twice with ice-cold 1X PBS by centrifugation at 2000 rpm for 2 minutes at 4°C. After the final wash, the protein supernatant of each sample was added to the beads for pre-cleaning and incubated for 30 minutes at 4°C in constant rotation. The samples were then centrifuged (2000 rpm for 2 minutes at 4°C) and 10% of the cleaned lysate was put aside to be used as the input sample. The remaining 90% of the lysate was then used for the immunoprecipitation; lysates were mixed with 20 µl of pre-washed Protein G Sepharose® beads and 2 µl antibody of interest (GFP-Trap (Chromotek). The samples were then incubated for 2 hours at 4°C in constant rotation. After incubation, the bead-antibody-protein complexes were precipitated by centrifugation and washed three times with lysis buffer. To elute the protein complexes from the beads, the samples were then mixed with LDS buffer (for 500 µl: 250 µl 4X LDS, 50 µl DTT, and 150 µl ddH_2O) in a 1:1 ratio and incubated at 70°C for 10 minutes. The beads were then removed by centrifugation (2000 rpm for 5 minutes at RT). IP and input samples were stored at -80°C and analysed via western blot (see Section 2.5).

2.8. Western blotting

2.8.1. SDS PAGE

Protein lysates were prepared by the addition of LDS buffer and 10% 0.5 M DTT in a 1:1 ratio. The samples were then heated at 70°C for 10 minutes to denature the proteins, followed by centrifugation at 14,000 rpm for 1 minute at RT. Samples were then loaded onto a NuPAGE Novex 4-12% Bis-Tris acrylamide gel (Novex Life Technologies, WC1020) along with a Colour Prestained Protein Standard ladder (New England Biolabs, P7719S) and subjected to gel electrophoresis for 90 minutes at 120 V. Once separated, the proteins were transferred to a nitrocellulose membrane (GE Healthcare Life Science, 10600001) using the wet transfer Mini-Trans Blot Cell (Bio-Rad), which then ran for 120 minutes at 100 V. After the transfer was complete, the membranes were stained using Ponceau red to ensure the proteins had successfully transferred.

2.8.2. Immunoblotting

Membranes were incubated in blocking solution (5% milk powder (Bio-Rad) in 1X TBS-T) for 60 minutes at RT on a rocker-shaker. The blocking solution was then removed, and the membranes were then placed in primary antibody made in blocking solution and incubated overnight at 4°C on a rocker-shaker. Details of the primary antibodies are provided below (see Table 2.6). The following day, membranes were washed three times with 1X TBS-T and then incubated in secondary antibody made in blocking solution for 2 hours at RT. Details of the secondary antibodies are listed below (see Table 2.7). The membranes were then washed three times with TBS-T and developed with 1 ml WesternBright ECL substrate (Advansta) and immediately imaged using a ChemiDoc XR+ (Bio-Rad) with ImageLab software (Bio-Rad, version 6.0.1).

Table 2.6. Primary antibodies used for western blot

Antibody	Host	Type	Supplier	Cat. No.	Dilution
Anti-HA	Rat	mAb	Roche	ROAHAHA	1:2000
Anti-GFP	Rabbit	pAb	Abcam	Ab290	1:2000

Table 2.7. Secondary antibodies used for western blot

Antibody	Host	Type	Supplier	Cat. No	Dilution
Anti-Rat-HRP	Goat	pAb	Promega	W4011	1:2000
Anti-Rabbit-HRP	Goat	pAb	Promega	HAF005	1:2000

2.9. Luciferase assay

Approximately 50,000 HEK293 cells were seeded into 24 well plate and transfected the following day with 5 ng pRL Renilla vector (Promega), 150 ng TOP-FLASH or FOP-FLASH (Addgene) vectors, and 150 ng plasmid of interest. The cells were lysed using the Dual-Luciferase® Reporter Assay System (Promega, E1910) following the manufacturer's instructions. Firefly luminescence activity was measured first using an FLOUstar® Omega microplate reader (BMG Labtech). After measuring the firefly activity, Renilla luminescence was then measured using the same plate reader protocol. The relative light units (RLU) of firefly activity were then normalised by dividing by the RLU of Renilla to calculate the relative response ratios for each condition. Moreover, to stimulate Wnt signalling, cells were treated with 200 ng human WNT3A 24 hours before lysis.

2.10. Bulk RNA sequencing

2.10.1. Generation of embryonic samples for RNA seq

To generate E11.5 samples for RNA-seq, *Pcdh19* HET females were mated with WT males and plug-checked the following morning, as described previously (see Section 2.1.1The litters produced from this mating strategy would include WT females ($X^{19\text{-}WT}X^{19\text{-}WT}$), *Pcdh19* HET females ($X^{19\text{-}WT}X^{19\text{-}KO}$), WT males (Y $X^{19\text{-}WT}$), and hemizygous KO males (Y $X^{19\text{-}KO}$).

2.10.2. Tissue dissection

Timed pregnant dams were terminated via schedule 1 at E11.5 and the uterus were extracted and placed in a 10 cm petri dish with ice-cold 1X dimethyl pyrocarbonate (DEPC)-treated PBS. The embryos were then extracted from their amniotic sacs and dissected one at a time. For the dissection, the embryos were placed under an illuminated M-series stereo microscope (Lecia, 10450167) in 1X DEPC-PBS and decapitated. The skin and meninges were then removed, and the brain was carefully dissected to extract tissue only from the developing lateral cortex whilst avoiding tissue from the ganglionic eminences and medial cortex (i.e., cortical hem and medial hippocampal primordium). The collected tissue was stored in RNA*later*™ stabilisation solution (Invitrogen, AM7020) and kept at -80°C. Samples were collected from four different litters across different days to obtain a sufficient number of samples for each condition (WT male, WT female, KO male, HET female). The tail of each embryo was also removed and kept at -20°C for genotyping, as mentioned previously (see Section 2.1.3).

2.10.3. RNA extraction and sequencing

Samples were thawed on ice and lysed using RLT lysis buffer (from RNeasy Mini kit; Qiagen, cat no. 74104) containing 1% β-mercaptoethanol. Total RNA was then extracted and purified using the RNeasy Mini kit following the manufacturer's instructions with DNase treatment. All steps were carried out in an RNase-free workspace. Quality control of the samples was done via Tapestation (Agilent Technologies) and the RNA integrity number (RIN) was determined for all samples. QUBIT was used to determine the concentration of RNA for each sample. Library preparation and sequencing was performed at the Cardiff University Genomics Hub by Angela Marchbank. Because the RNA input was low for all samples, libraries were prepared using the NEBNext® Single Cell/Low Input RNA Library Prep Kit for Illumina® (New England Biolabs, E6420S). Briefly, mRNA was purified from total RNA using poly-T oligos, fragmented, and converted to cDNA. The cDNA was then amplified via PCR. The sequencing was carried out using the Illumina Nextseq 500 platform using cartridges for single-end (SE) (1x75bp) sequencing. Each individual sample was sequenced twice to obtain at least 30-40 M reads per sample. The two sequencing files per sample were therefore merged before quality control began.

2.10.4. Quality control, alignment, and read counting

Quality control was performed using FastQC. Adaptor sequences were trimmed using Trimmomatic default parameters. The majority of reads were around 60-75bp long, therefore in order to improve processing speed in later steps, reads that were smaller than 50bp were also trimmed. Reads were mapped to the mouse GRCm39 reference genome using STAR (Dobin et al., 2013) and counted using featureCounts (Liao et al., 2014), utilising the GRCm39 Ensembl gene build GTF. The reference genome and the GTF files were downloaded from the Ensembl FTP site (http://www.ensembl.org/info/data/ftp/index.html/).

2.10.5. Differential gene analysis in R

Differential gene expression analysis was conducted in R (v.4.02) via RStudio (v.1.2.1335) using the DESeq2 package and associated workflow (Love et al., 2014). The Benjamini-Hochberg correction was used for multiple testing. Plotting was done using the R package "ggplot2" (v.3.3.5). Volcano plots were generated using the "EnhancedVolcano" package (v.1.10.0). Gene Set Enrichment and overrepresentation analysis was performed using the "clusterProfiler" (v.4.0.2) package. Dotplots and Cnet plots were also generated using the "DOSE" (v.3.18.1) and "enrichplot" (v.1.12.2) packages, respectively. Finally, heatmaps were generated using the

"pheatmap" (v.1.0.12) package. Gene symbols were added to the DESeq2 dataset via the "AnnotationDbi" (v.1.54.1) and "org.Mm.eg.db" (v.3.13.0) packages.

2.11. Single cell RNA sequencing

2.11.1. Generation of embryonic samples for single cell RNA seq

E11.5 samples were generated for single cell RNA-seq (sc-RNA seq) using a similar strategy as mentioned previously (see Section 2.11.1) with slight variation. *Pcdh19* HET females were mated with *XGFP* males to generate litters that would include *XGFP* WT females, *XGFP Pcdh19* HET females, WT males, and hemizygous KO males.

2.11.2. Tissue extraction and dissociation

Timed pregnant dams were terminated via schedule 1 at E11.5 and the uterus were extracted and placed in a 10 cm petri dish with ice-cold 1X DEPC-PBS with 3% FBS (F-PBS). The embryos were then dissected using a similar strategy as mentioned previously (see Section 2.10.2) with slight variation for sc-RNA seq. Single cell suspensions were generated using a modified protocol for single cell dissociation of mouse embryonic tissue (Ibarra-Soria et al., 2018). Lateral cortical tissue was collected from each embryo and placed in individual eppendorfs on ice in F-PBS. Samples were then centrifuged at 500 g for 5 minutes and washed twice with 1X DEPC-PBS. The samples were then re-suspended in TrypLE™ Express Enzyme (Gibco, 12604013) and incubated in a 37°C water bath for 10 minutes. The trypsinisation was then quenched by washing the samples with F-PBS. Finally, the cells were resuspended in PBS with 0.04% bovine serum albumin (BSA) supplemented with DNase I (Sigma, 10104159001) and filtered using a 0.40 µm nylon cell strainer. The number of cells was then calculated using an 0.0025 cm^2 Haemocytometer (Neubauer). Tail samples were collected and each embryo was genotyped alongside the dissociation process by Dr. Cristina Llinares-Benadero.

2.11.3. Cell partitioning

Cell suspensions were partitioned into individual microfluidic emulsions using the Chromium Single Cell 3' v3 kit (10x Genomics, 1000269) following the manufacturer's instructions. Briefly, cell samples were diluted with nuclease-free water to reach the optimal range of cell concentration that would maximise the likelihood of recovering approximately 10,000 cells per sample. The cells were then mixed with a master mix solution (containing template switch oligos, reducing agents, and enzymes) and loaded onto a Chromium Chip-B cassette. Gel beads, which contain surface small nucleotide sequences necessary for single cell mRNA capture, were then loaded into the

wells above the cells. Partitioning oil was then added into the wells below the cell suspensions and the cassette was loaded into a 10x Chromium Controller. The Controller then performed parallel sample partitioning and molecular barcoding for each individual cell by stimulating a stable flow rate that enabled each gel bead to bind to a single cell. Once bound, the gel bead/cell complex was immersed in the partitioning oil which served as a reaction vesicle. The gel beads in emulsion (GEMs) reaction lysed the cells and caused the bead to dissolve and relinquish the mRNA. The captured mRNA was then immediately converted to cDNA via reverse transcription using the recommended cycling protocol. Single cell cDNA samples were then kept at -20°C until ready for sequencing.

2.11.4. Sequencing and data pre-processing

Library preparation and sequencing was performed at the Cardiff University Genomics Hub by Angela Marchbank. cDNA quality of each library was checked via Tapestation before sequencing. The samples were sequenced on the Illumina Novaseq™ 6000 platform using the S2 Novaseq™ cartridges to achieve a read depth of ~50,000 reads per cell. The initial processing of the single cell data, including alignment of the reads, filtering, barcode counting, and UMI counting, were performed using the Cell Ranger pipeline (v.7.0) by Dr. Sumukh Deshpande from the College of Biomedical Science, Cardiff University.

2.11.5. Single cell transcriptional analysis in R

The analysis of sc RNA-seq data was conducted in R (v.4.02) via RStudio (v.1.2.1335) using the Seurat package (v.4.1.1) and associated workflow (Hao et al., 2021). The Benjamini-Hochberg correction was used for multiple testing. Plotting was done using the R package “ggplot2” (v.3.3.5). The initial cell clustering and downstream analyses were performed in collaboration with Dr. Sumukh Deshpande from the College of Biomedical Science, Cardiff University, and Dr. Sergi Roig-Puiggros from the Jabaudon lab, University of Geneva.

2.12. Quantitative PCR (qPCR)

For RNA extraction, samples were collected and protected with RNAlater (ThermoFisher) at -80°C for RNA extraction, which was performed using RNeasy Mini Kit (Qiagen) followed by RNase-Free DNase set (Qiagen). Maxima First Strand cDNA Synthesis Kit was used to generate the cDNA template for quantitative real-time PCR (ThermoFisher). RT-PCR were developed using Applied Biosystems StepOne Plus and analysed using the corresponding software StepOne Software Version 2.0 (Applied Biosystems). Quantification was performed using a standard curve.

The primers used included Pcdh19-F (5'–GGAGGAGACAGACAAGATGAATG–3') and Pcdh19-R (5'–CTGCTGGTGGTAGTCGAAATAG–3').

2.13. Statistical analysis

Statistical analysis was conducted using R (v.4.02) via RStudio (v.1.2.1335). Homogeneity of variance was examined using the Levene's test of equality of error variances. Normality was examined using the Shapiro-Wilk test. Data was considered to have equal variances and considered normally distributed if both aforementioned tests produced a p-value of > 0.05. For analyses that involved comparing more than two groups, a one-way ANOVA was used. For analyses that involved comparing two or more groups across more than one factor, a repeated-measured ANOVA was used. For analyses that compared only two groups, a two-tailed unpaired t-test was used. Post-hoc Bonferroni corrections were used for multiple comparisons.

Where data was shown to violate parametric assumptions, including normality or homogeneity of variance, the data was either logarithmically transformed or non-parametric tests were employed. For analyses comparing more than two groups, a Kruskal-Wallis test was used. For post-hoc tests, a pairwise Wilcox test was used. All significant results are shown using asterisks (* = 0.05, ** = 0.01, *** < 0.001).

Chapter 3: Investigating the transcriptional landscape of WT, HET, and KO Protocadherin 19 embryos using bulk RNA-seq

3.1. Introduction

As previously described, heterozygous mutations in the X chromosome gene protocadherin-19 (*PCDH19*) lead to PCDH19-epilepsy in predominantly female patients (Dibbens et al., 2008; Specchio et al., 2011). The cellular and molecular mechanisms underlying PCDH19-epilepsy are poorly understood, however early onset of PCDH19-epilepsy suggests that *PCDH19* may play a role in brain development. Consistent with this, *Pcdh19* has been shown to be expressed during mouse embryogenesis, especially in neural tissues (Gaitan and Bouchard, 2006). Expression has been shown to begin in the neuroepithelium as early as embryonic day 9 (E9) and continues through the onset of neurogenesis at E10-12. Therefore, PCDH19 likely has a role in brain development that is beginning to be explored.

Previous work in the Isabel Martinez-Garay IMG lab has shown that at E11, predominant expression of *Pcdh19* mRNA was found in the cortical primordium and this receded overtime in a lateral-to-medial direction, until at E14 when expression was negligible (unpublished data, 2019). Expression of *Pcdh19* was also primarily localised to radial glial cells (RGCs), which at E11 are the major subtype of neural progenitor cells that populate the ventricular zone (VZ) of the developing cortex (Pinto and Götz, 2007). *Pcdh19* was also shown to be downregulated in other subtypes of progenitors, especially intermediate progenitors (IPs), which populate the subventricular zone (SVZ) (unpublished data, 2019). During cortical development, RGCs undergo extensive symmetric proliferative divisions to expand the progenitor pool but switch to asymmetric neurogenic divisions at ~E10-E11 in mice (Takahashi et al., 1993, 1996). Taken together, the short window of expression of *Pcdh19* in the RGCs of the cortical primordium suggests that *Pcdh19* is playing an important role in RGCs around the time when the cells switch to divide asymmetrically to produce neurons.

Increasing evidence is emerging that identifies PCDH19 as a regulator of neurogenesis. Knockdown (KD) of *Pcdh19* has been shown to cause hyperproliferation of neural progenitors in the developing hindbrain of zebrafish (Biswas et al., 2021). Interestingly, the increased proliferation of progenitors was linked to an overactivation of the Wnt/β-catenin signalling pathway via the Wnt co-receptor Ryk. Further evidence has shown that neural progenitors extracted from knockout (KO) mouse telencephalic vesicles display altered neurogenesis *in vitro*, wherein KO

progenitors produced more neurons and less non-neuronal cells than WTs (Homan et al., 2018). Moreover, it has also been shown that shRNA KD of *Pcdh19 in vivo* at E11 causes a decrease in the production of layer 2-4 and 5-6 neurons, as demonstrated using mosaic analysis with double markers (MADM) (Lv et al., 2019).

Previous reports exploring the role of PCDH19 in neurogenesis are often contradicting. For example, it is not clear whether KD of *Pcdh19* causes neural progenitor cells to over-proliferate or overcommit to neurogenic divisions (Biswas et al., 2021; Fujitani et al., 2017; Homan et al., 2018). Additionally, few studies have investigated whether mosaic expression of *Pcdh19* affects neurogenesis, especially *in vivo*. Homan et al. (2018) reported that when co-culturing cortical progenitor cells extracted from E14 WT and KO embryos, only KO progenitors showed altered neurogenic behaviour *in vitro*. However, as aforementioned, work in the IMG lab has shown that at E14 there is no expression of *Pcdh19* in cortical RGCs, suggesting that culturing KO progenitors causes unusual neurogenic behaviour that is not strictly related to loss of *Pcdh19*. Moreover, it has been shown *in vivo* that mosaic expression of *Pcdh19* causes WT and KO progenitors to segregate away from each other and form striking column structures in the cortex (Pederick et al., 2018). To investigate whether mosaic expression affected neurogenesis, work at the IMG lab has shown that in heterozygous (HET) embryos at E11 and E12, WT and KO progenitors produce less and more neurons, respectively (unpublished data, 2019). Interestingly, no neurogenesis defects were observed between homozygous WT and KO embryos, demonstrating that alterations in neurogenesis caused by the loss of *Pcdh19* are different between *in vitro* and *in vivo* systems (Homan et al., 2018).

The results described above suggest that PCDH19 is likely involved in regulating the timely transition of RGCs to undergo neurogenesis and that mosaic expression of *Pcdh19* causes RGCs prolonged proliferation and premature neurogenesis in WT and KO RGCs, respectively. The altered neurogenic phenotype observed in HETs could also be a potential underpinning reason of PCDH19-epilepsy in humans. However, despite these findings, the mechanism by which PCDH19 regulates neurogenesis in RGCs is unknown. One approach to identify pathways related to the role of PCDH19 in neurogenesis in an ‘unbiased’ fashion is to employ next generation sequencing (NGS). To date, only one transcriptional analysis on WT and KO progenitors has been reported (Homan et al., 2018). In this experiment, *Pcdh19* KO progenitors were extracted from E14 embryos and differentiated into neurons *in vitro*. Microarray analysis on the KO cells revealed several upregulated pathways related to differentiation and neurogenesis and downregulation of pro-proliferative genes (Homan et al., 2018). However, as aforementioned the

expression of *Pcdh19* is diminished in RGCs by E14, suggesting that the experiment may not be an accurate representation. Moreover, it appears that KO progenitors show altered neurogenic behaviour *in vitro* but not *in vivo* (Homan et al., 2018; unpublished data, 2019). Therefore, the transcriptional differences that have been previously reported *in vitro* cannot confidently demonstrate the legitimate molecular mechanisms of PCDH19 involvement in neurogenesis *in vivo*. To investigate this further, transcriptional analysis of *Pcdh19* WT, HET, and KO tissue samples taken *in vivo* may reveal more reliable evidence.

3.1.1. Aims

The aim of this chapter was to provide a detailed examination of the expression landscape of *Pcdh19* WT, HET, and KO embryonic cortical tissue at E11.5 by conducting whole transcriptome RNA-seq analysis. Additionally, this chapter also aimed to uncover potential transcriptional changes that arise from the total loss of *Pcdh19* or, importantly, whether heterogeneous expression of *Pcdh19* causes transcriptional alterations that can be detected using bulk RNA-seq methods. This was the first step in understanding the molecular underpinnings of PCDH19 involvement in neurogenesis.

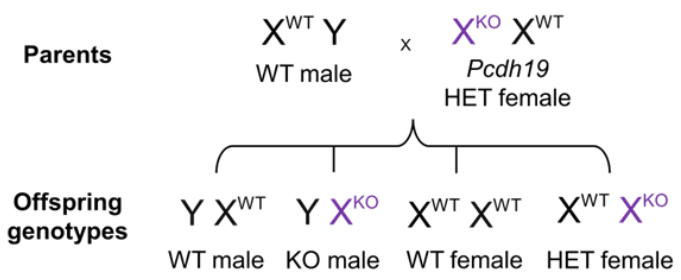
3.2. Results

3.2.1. RNA quality and sample processing

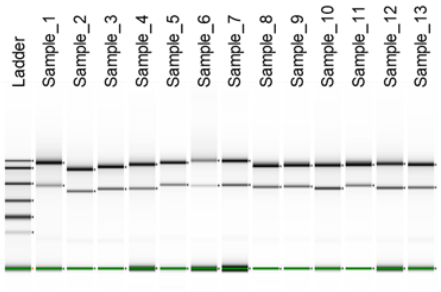
The E11 cortical samples from *Pcdh19* WT male, WT female, HET, and KO embryos were collected, and the RNA was extracted and sequenced, as detailed previously (see Section 2.11). Samples were collected from four different litters across several days in order to achieve at least three biological replicates per group (final group count: WT male n = 3, WT female n = 3, HET female n = 4, KO male n = 3). Ideally, all biological replicates for all groups would come from the same litter to ensure that all embryos were at the exact same developmental time point. However, due to varying litter sizes and the unlikelihood that the N for all genotypes would be present a single litter, it was decided to collect samples across multiple litters. Moreover, samples were collected at the same time of day (noon) which would help to slightly reduce the overall effect of development-based biases. Because litter could prove to be a major confounding factor in the experiment, its effect on the data was further explored below.

All samples were checked for RNA integrity and were shown to possess a RIN score of 9.0 or above (see Figure 3.1). Sample concentrations were assessed using the QUIBIT and were shown to range between 0.4 ng/µl – 100 ng/µl. After sequencing, FASTQ files were quality checked using FASTQC and all samples were shown to possess a Mean Quality Score (Phred Scores) of 30 or above (i.e., more than 99.9% base call accuracy). The GC content for all samples was relatively high (~60%), however this is expected from RNA-seq data (Benjamini and Speed, 2012). Duplication levels were also normal (~50%). All samples passed quality control metrics and were therefore trimmed, aligned, and counted as described previously (see Section 2.11).

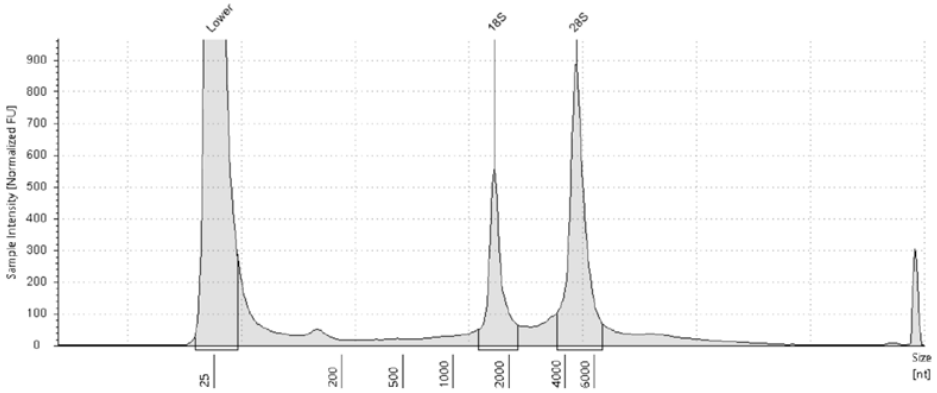
A



B



C



Well	RINe	28S/18S (Area)	Conc. [ng/µl]	Sample Description	Alert	Observations
B1	9.9	2.0	13.3		⚠	Caution! Expired ScreenTape device (used after two weeks of first use); Sample concentration outside functional range for RINe and the assay

Peak Table

Size [nt]	Calibrated Conc. [ng/µl]	Assigned Conc. [ng/µl]	Peak Molarity [nmol/l]	% Integrated Area	Peak Comment	Observations
25	40.0	40.0	4710	-		Lower Marker
1677	2.56	-	4.49	33.62		18S
4678	5.06	-	3.18	66.38		28S

Figure 3.1. RNA samples passed RNA Integrity Number (RIN) quality control. (A) Visual aid showing the offspring genotypes after mating a WT male with a *Pcdh19* HET female. The offspring produced from this type of crossing were used as the main groups for the bulk RNA-seq experiment (WT male n = 3, KO male n = 3, WT female n = 3, HET female n = 4). (B) Tapestation gel showing the RINs of all RNA samples (C) Example of RNA scores calculated from the 28S/18S RNA peaks.

3.2.2. Hierarchical clustering

The RNA-seq dataset was first subjected to hierarchical clustering to assess overall similarity between samples. This was achieved by calculating the Euclidian distance between the samples and plotting the values in a heatmap (see Figure 3.2 and Figure 3.3). Hierarchical clustering illustrated a noticeable separation of samples by sex (see Figure 3.2). Regarding the males, samples 10, 7, 2, 3, and 4 displayed similar distance values and thus appear to strongly correlate, whereas sample 13 was the only major outlier. In the females, samples 6, 11, and 9 show noticeable clustering, however this is not as strong as the male clustering effect. The distance values of samples 8 and 1 are also moderately related to the other female samples, but the clustering effect is markedly weaker. Interestingly, samples 5 and 12 correlate more strongly with the male samples rather than the female samples.

Regarding genotype, it was observed that WT male and KO samples appeared to show a significant clustering effect, suggesting a strong similarity between these groups. The clustering of WT female samples is less evident although still noticeable, as demonstrated by samples 6 and 9. Nevertheless, sample 12 was relatively divergent compared to the other WT female samples. Strikingly, HET samples showed the weakest within-group correlations; samples 1, 8, and 11 show a significantly weak correlation when compared to each other. Sample 5 shows a slightly stronger sample-sample distance comparison to the other HET samples; however, it is most strongly similar to multiple WT male and KO samples. This indicates that the HET samples are likely the most variable of the genotype groups and hold the most diverse transcriptional profiles. Overall, hierarchical clustering analysis illustrates that samples mainly cluster by sex.

It is also worth noting that the dissimilarity observed between these samples may be due to the confounding variability introduced by including samples from different litters. To address this issue, litter annotations were added to the hierarchical clustering map to examine whether samples would cluster by litter (see Figure 3.3). Interestingly, litters 2, 3, and 4 appeared to show a clustering effect however, because the sample numbers within each litter were not equal, it was difficult to determine whether this effect was driven by litter or by another factor such as sex. For example, samples 2 and 3 (litter 2) were male, whereas samples 8, 6, 11, and 9 (litter 3 and 4) were female.

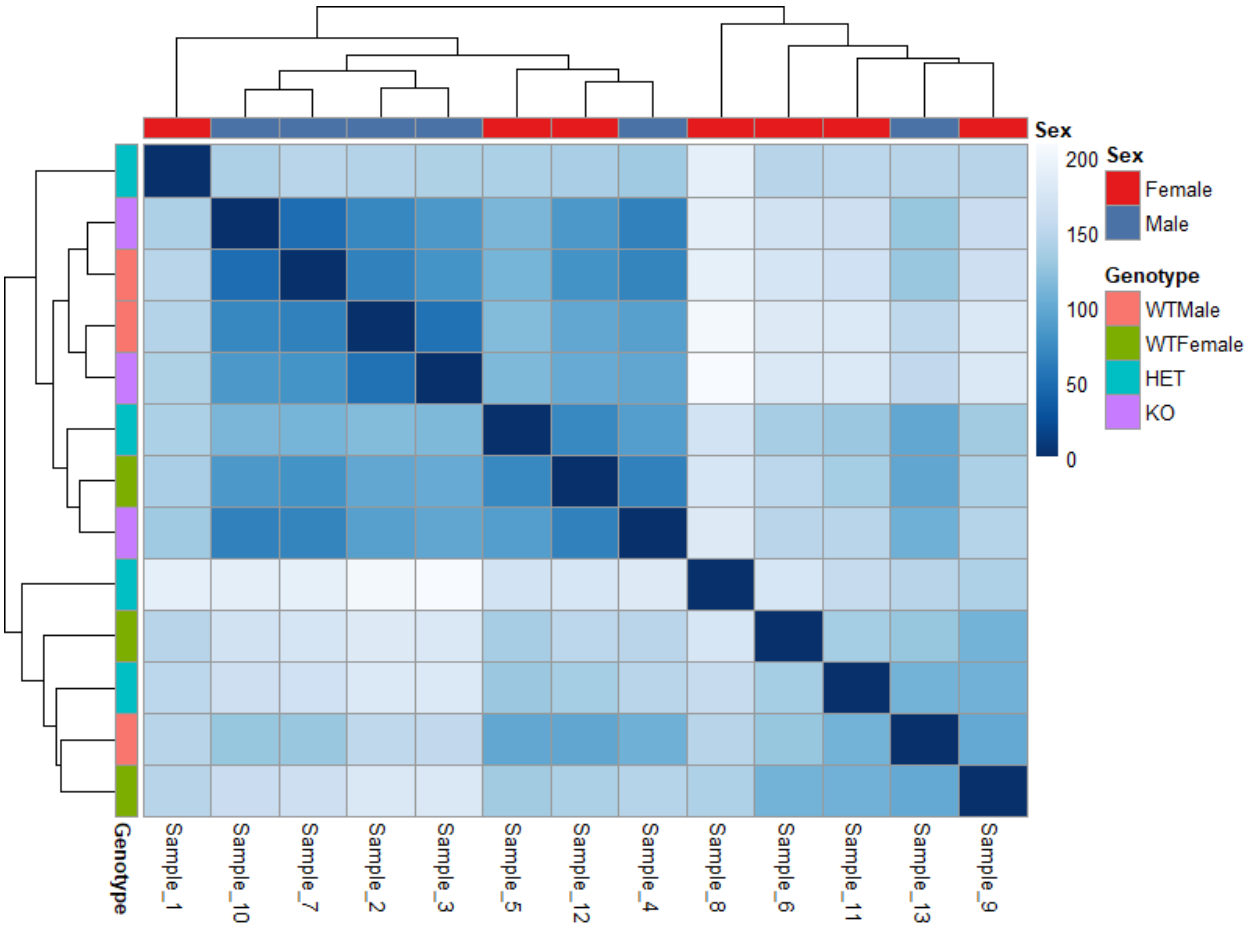


Figure 3.2. Correlation matrix showing the sample-sample distances via hierarchical clustering based on total gene expression for sex and genotype. Blue scale denotes the distance between samples, with darker blue corresponding to smaller distances (i.e., greater transcriptional similarity between samples).

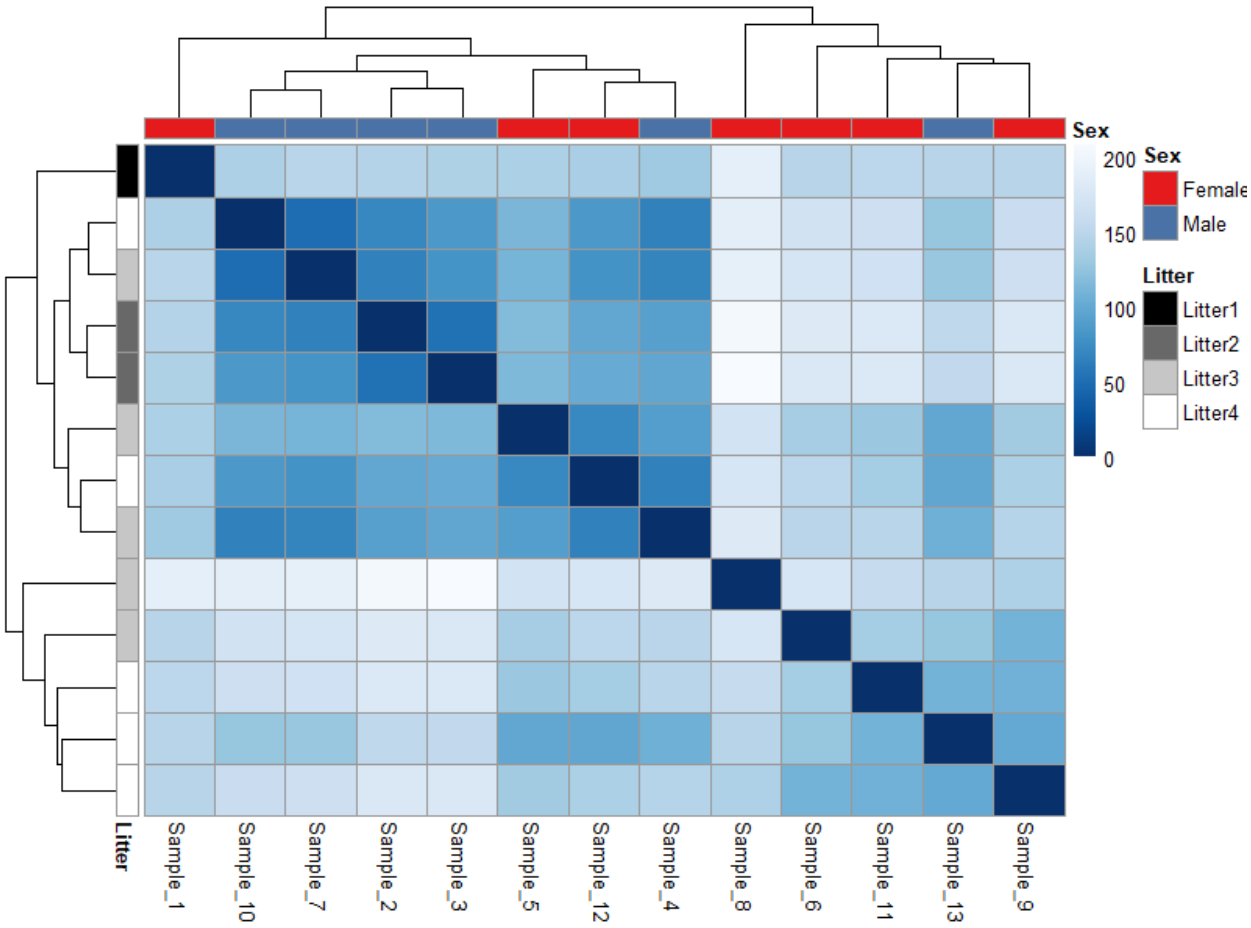


Figure 3.3. Correlation matrix showing the sample-sample distance values via hierarchical clustering based on total gene expression for sex and litter. Blue scale denotes the distance between samples, with darker blue corresponding to smaller distances (i.e., greater transcriptional similarity between samples).

3.2.3. Principal component analysis (PCA)

To examine the distances between the samples further, the data was subjected to a principal component analysis (PCA). PCA is a technique for transforming multidimensional data into a coordinate system wherein the variation within the dataset is described by two principal components (PC). The PCs of each data observation are plotted in two-dimensional space which helps interpret the relationships between each observation by examining their proximity to each other. Observations that cluster strongly to each other are considered closely related. PCA on all samples revealed that PC1 accounted for 63% of the variance in the dataset, whereas PC2 accounted for 12% (see Figure 3.4). Similarly to the hierarchical clustering, in the PCA analysis the samples were mostly separated by sex (seeFigure 3.4A). Male samples showed the strongest clustering in the dataset with the exception of sample 13 which did not co-cluster with the rest. In contrast, female samples displayed a noticeably weaker clustering effect compared to males. Samples 6, 9, and 11 were the most co-clustered female samples, whereas samples 1, 5, 12, and 8 were mostly divergent. When examining the effect of genotype on the PCA, WT and KO male samples appeared to co-cluster together regardless of genotype. Moreover, except for sample 12, WT female samples also showed a moderate clustering effect. Interestingly, the HET samples did not cluster together, as previously demonstrated with the hierarchical clustering analysis. Overall, this illustrates that sex is likely the primary factor driving the transcriptional variance in the dataset, especially for the male samples.

To further examine the effect of litter on the variance of the dataset, the PCA was re-plotted to include litter annotations (seeFigure 3.4B). This analysis revealed that litter 2 displayed the strongest clustering effect, and that some samples from litter 3 and 4 also clustered together. As previously stated, it is difficult to separate the effects of sex and litter. For example, the two samples from litter 2 clustered closely, but those two samples of the litter groups are also both male. This is also observed for samples 7 and 4 (litter 3), and samples 11 and 9 (litter 4). There are also multiple samples that do not co-cluster with samples from the same litter, such as samples 5, 6, and 8, and samples 13, 12, and 10. Therefore, it is probable that litter does have a marginal effect on the variance of the dataset, however sex is likely the most important variable. This observation demonstrates that variability between litters is not a main driver of variance between samples and that combining embryos from the different litters for the RNA-seq analysis is unlikely to have a significant impact on the results.

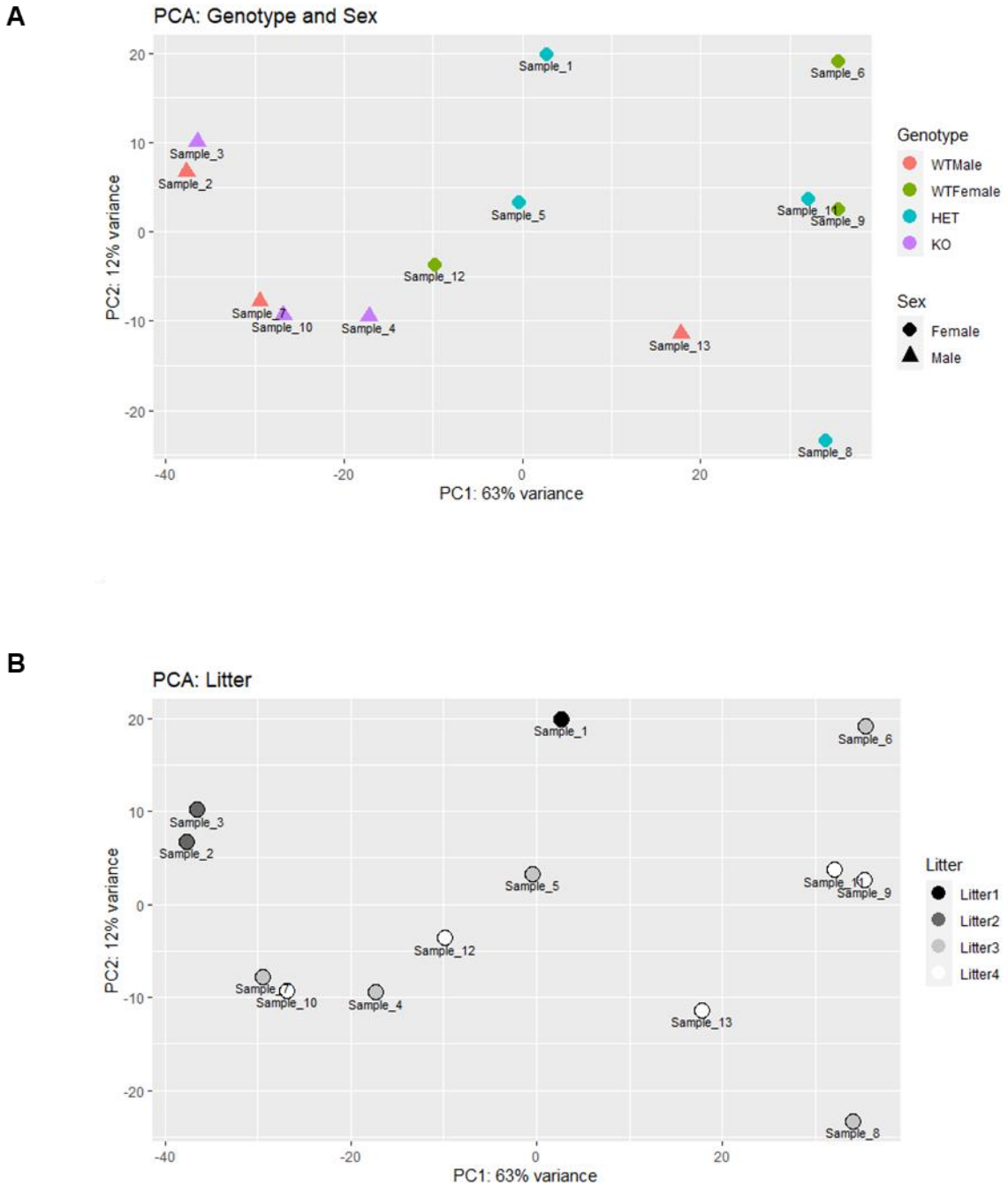


Figure 3.4. Principal component analysis (PCA). (A) PCA of all samples illustrating the effect of genotype (colour) and sex (shape) on sample-sample distances (B) PCA of all samples illustrating the effect of litter (colour) on sample-sample distances.

3.2.4. Protocadherin-19 and β-galactosidase expression

To confirm the genotype of the samples, which were initially assessed by PCR, expression of *Pcdh19* and *β-gal* were examined for each sample. Expression was analysed by extracting the normalised count values of both genes from the RNA-seq data frame and the expression from each genotype was compared (see Figure 3.5). *Pcdh19* expression levels were relatively comparable between WTs and HETs, except for two WT female samples that displayed significantly higher expression of *Pcdh19*. As expected, KO samples showed the lowest expression levels of *Pcdh19*. This was expected since exons 1-3 of the *Pcdh19* allele have been replaced by the *β-gal* neo reporter cassette, as mentioned previously. KO samples also showed the highest expression of *β-gal* (see Figure 3.5B). Interestingly, the expression of *β-gal* in the HETs was significantly more variable than in the KOs.

To investigate the higher expression of *Pcdh19* in WT females relative to WT males, quantitative PCR (qPCR) was performed against *Pcdh19* using cDNA generated from cortical tissue of E11.5 WT male (n = 3) and WT female (n = 3) (see Figure 3.5C). Comparing the normalised expression levels of *Pcdh19* between both sex groups using an independent t-test revealed no differences between WT females and WT males ($T(2) = -0.85, p = 0.47$), however the variance within the WT female group was much higher than in the WT male group.

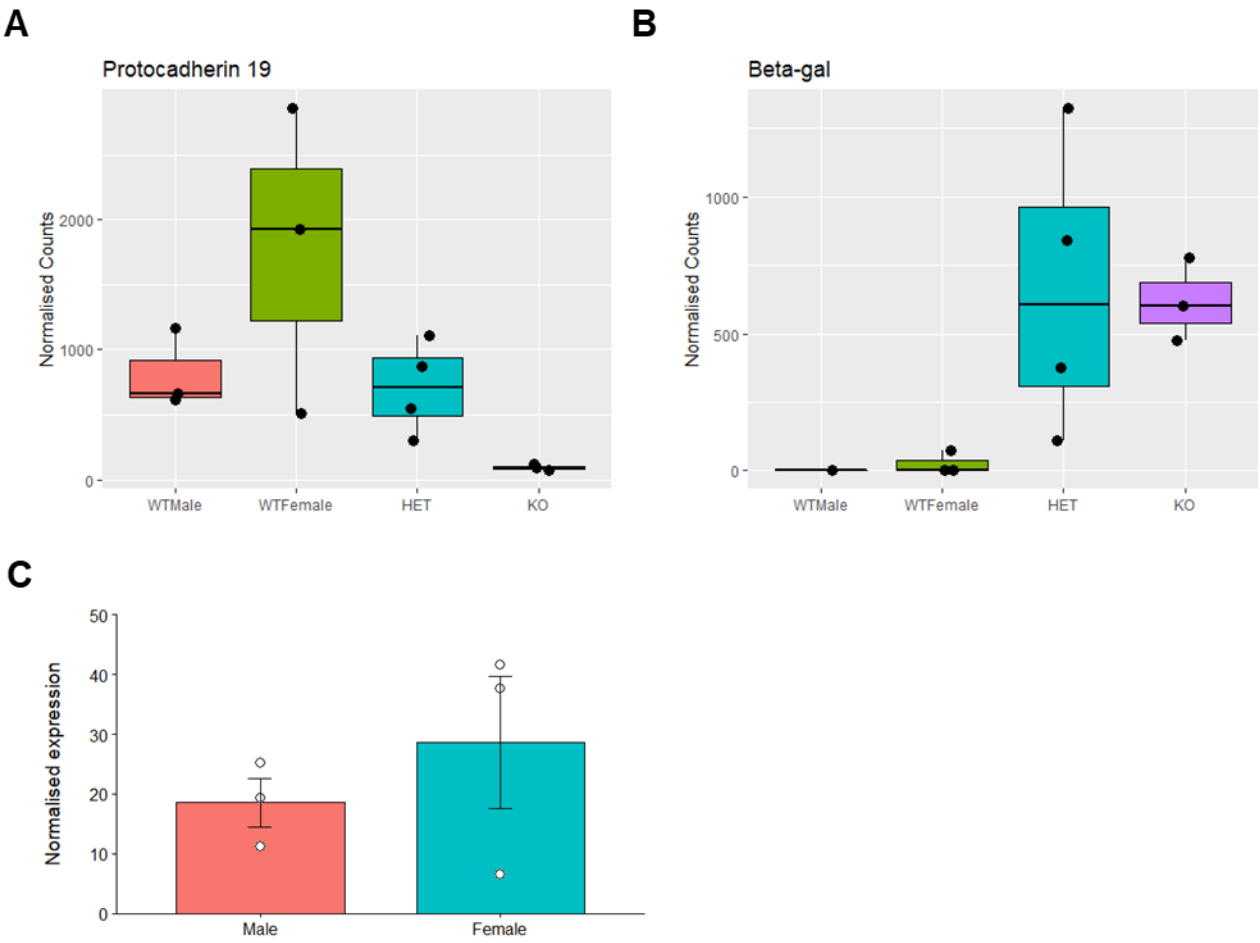


Figure 3.5. Normalised expression counts for *Pcdh19* and *β-gal*. (A) Normalised counts of *Pcdh19* between genotype groups. (B) Normalised counts of *β-gal* between genotype groups. (C) qPCR results illustrating the averaged quantitative expression of *Pcdh19* between E11.5 WT male (n = 3) and WT female (n = 3) samples. Expression levels were normalised using a standard curve generated from a series of cDNA dilutions. No difference in expression level was found between WT male and WT female samples (p = 0.47). Data is represented as mean + standard error. White dots indicate data points.

3.2.5. Expression of selective markers

Analysing the expression of key marker genes is a further way to verify the validity of the transcriptional profiles of the samples. As aforementioned, the predominant cell type at E11.5 in the developing cortex are RGCs. Therefore, it was predicted that RGC makers, such as *Pax6, Nestin,* and *Emx2* would be upregulated, whereas IP markers such as *Eomes* and neuronal makers such as *Tbr1, NeuroD1,* and *NeuroD2*, would show lower expression levels. To investigate this, transformed normalised counts of those genes were extracted from the RNA-seq dataset and plotted via a heatmap. As predicted, there was a noticeable separation by cell marker, wherein RGC genes were markedly upregulated compared to neuronal and IP genes (see Figure 3.6A). Interestingly, there was also a noticeable clustering effect by sex, whereby female samples appeared to show slightly increased expression levels of neuronal and IP markers including *NeuroD1, Eomes,* and *Tbr1*. *NeuroD2* was also slightly upregulated in all samples, especially in males. These differences in expression of neuronal markers may indicate that the samples were at marginally different stages of development and that therefore the levels of neurogenesis between the samples are different. To explore this possibility, the data was replotted to include litter annotations. Interestingly, there was no meaningful clustering effect by litter (see Figure 3.6B).

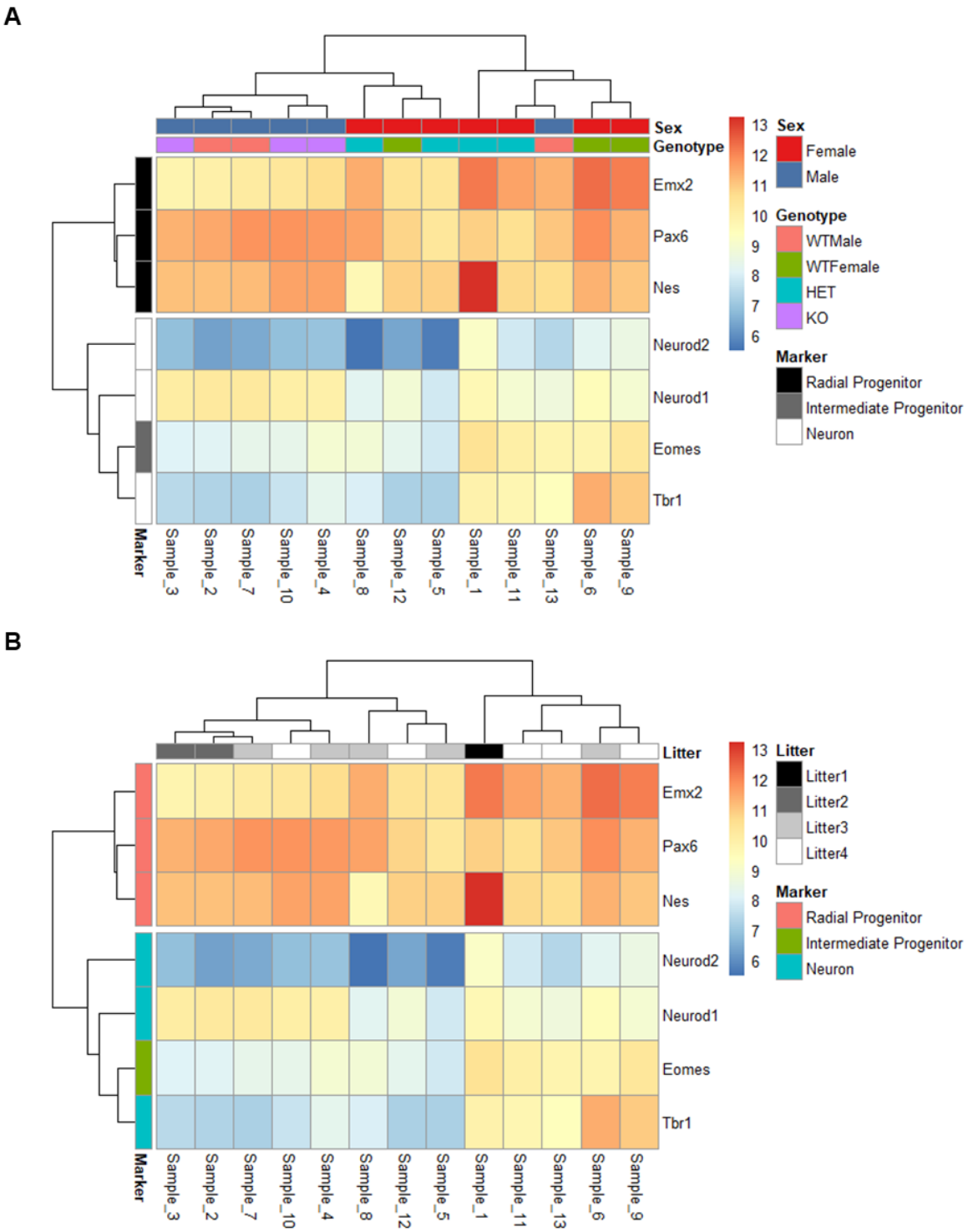


Figure 3.6. Expression of selected progenitor and neuronal genes. (A) Normalised expression of progenitor and neuronal marker genes annotated by sex, marker, and genotype. (B) Normalised expression of progenitor and neuronal marker genes annotated by litter and marker.

3.2.6. Differential gene expression analysis

To identify up-and-downregulated genes between samples, differential expression (DE) analysis was performed. The threshold for significance was set to $p < 0.05$ and a $\log_2$ fold change cut off at > 0.58 or < -0.58 which corresponds to positive and negative fold changes of 1.5, respectively. All p-values reported are p-adjusted values.

3.2.6.1. WT female vs. WT male

As previously discussed, sex appears to account for a significant proportion of the variance in the RNA-seq dataset. Therefore, an initial DE comparison was performed between WT female and WT male groups in order to determine the most significant sex-related DE genes (see Figure 3.7). This analysis revealed four highly differentially expressed genes: *Kdm5d* (p = 3.945013e-25) *Xist* (p = 5.791020e-2), *Uty* (p = 1.702442e-18), and *Tsix* (p = 1.702442e-18). *Kdm5d* and *Uty* are demethylases located on the Y-chromosome and are thus upregulated in male samples. Conversely, *Xist* and *Tsix* are non-coding RNAs located on the X-chromosome that have an important role in X-inactivation during female embryonic development. These four genes were also shown to be highly expressed within KO and HET samples following the same sex-specific expression pattern, illustrating that the expression of these genes is independent of genotype and solely sex-linked (see Figure 3.8). Additionally, there was a small number of other DE genes that were shown to just reach the significance cut-off threshold and were upregulated in female samples, including *Fgf8* (p = 5.627565e-08), *Grb14* (p = 4.135953e-02), *Shc4* (p = 2.191449e-02), *Nenf* (p = 2.191449e-02), *Pkp2* (p = 2.261791e-02), and *Dipk2a* (p = 4.135953e-02) (see Figure 3.8). Taken together, it was concluded that the differences in expression of those genes was exclusively due to sex; therefore, it was decided to remove all of them from subsequent analyses to increase the potential of identifying DE genes that were related to genotype.

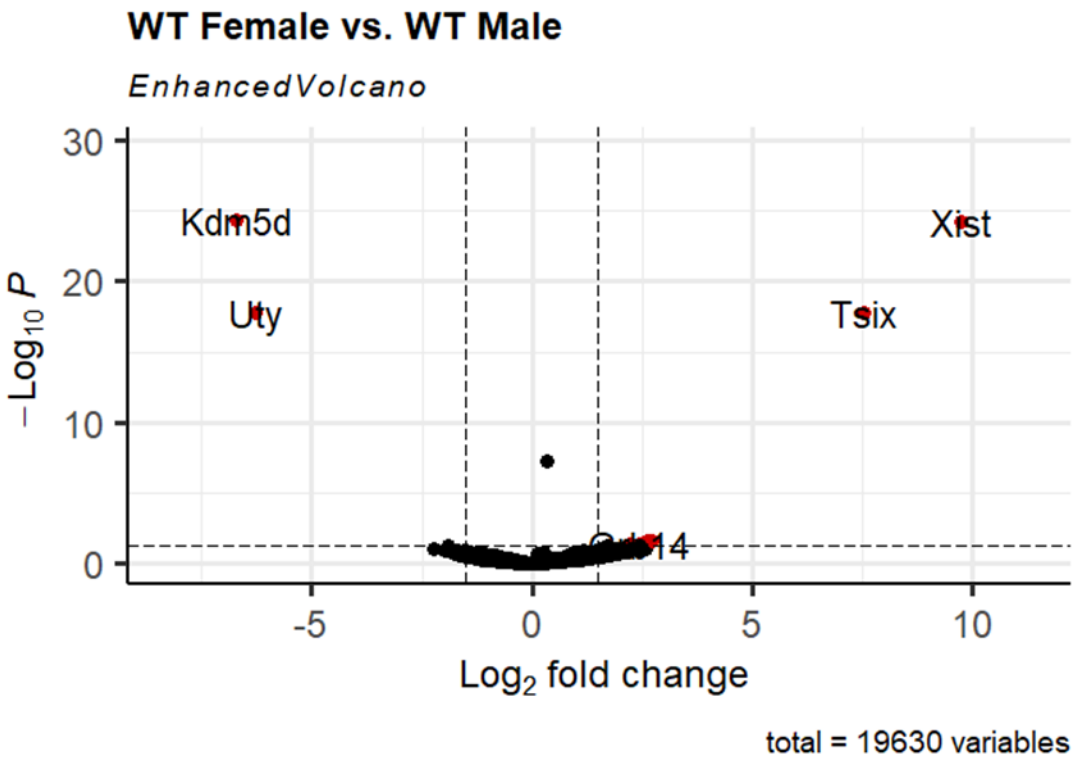


Figure 3.7. WT female vs. WT male differential gene expression analysis. Volcano plot of the WT female vs. WT male differential expression analysis. Significantly expressed genes are shown in red. Nonsignificant genes are in black. Threshold for significance = p.adjusted < 0.05, cut-off > 1.5.

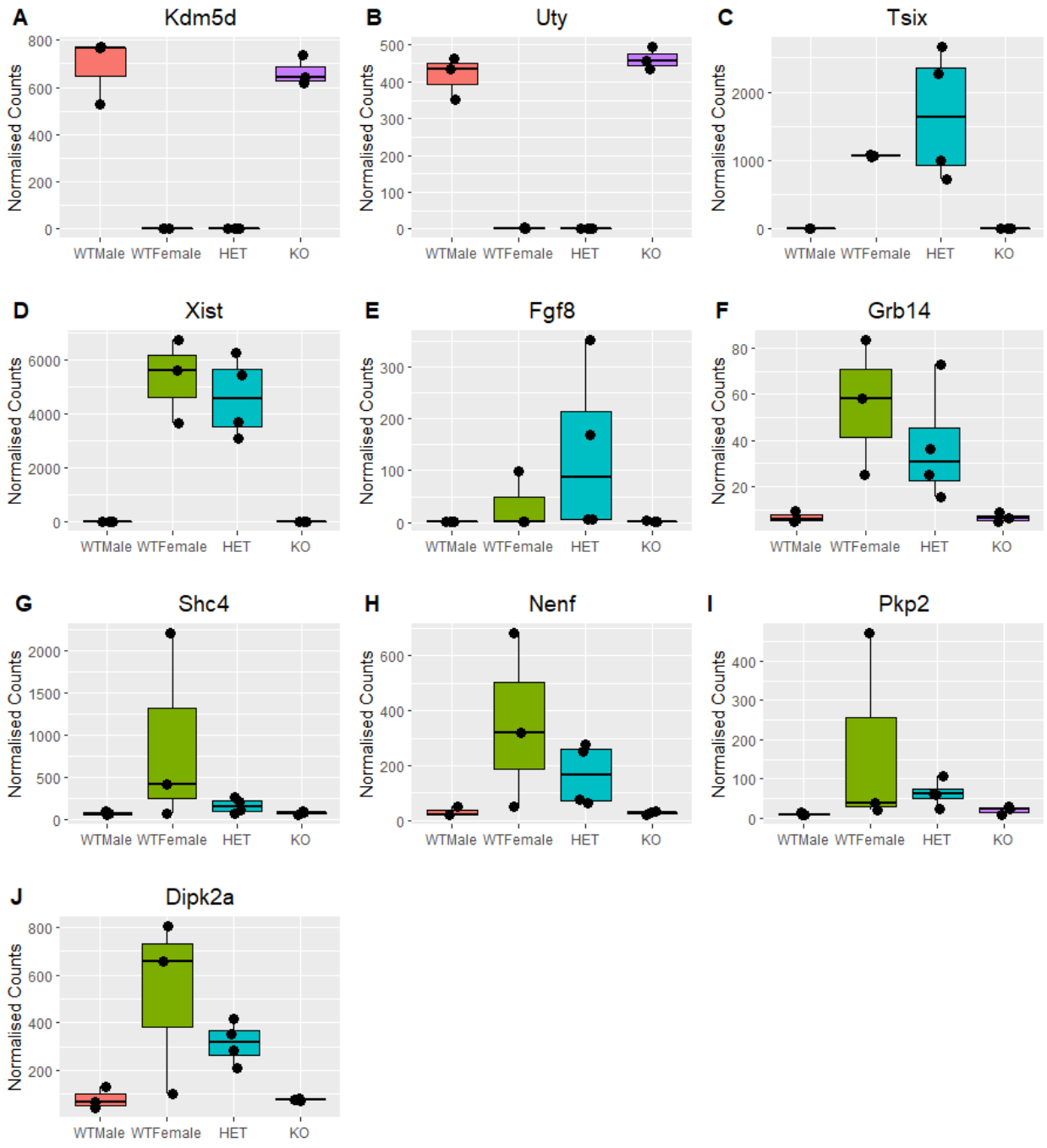


Figure 3.8. Normalised expression counts of DE genes from WT female vs. WT male comparison. Expression of differentially expressed genes found between WT female vs. WT male DE analysis and shown for all genotype groups, including *Kdm5d* (A), *Uty* (B), *Tsix* (C), *Xist* (D), *Fgf8* (E), *Grb14* (F), *Shc4* (G), *Nenf* (H), *Pkp2* (I), and *Dipk2a* (J).

3.2.6.2. KO males vs. WT males

As previously mentioned, KO males are hemizygous for the KO allele, and show complete reduction of *Pcdh19* expression. Interestingly, KO mice display normal rates of neurogenesis at E11 and E12, potentially suggesting that loss of *Pcdh19* is compensated in order to maintain normal brain development. To investigate whether loss of *Pcdh19* causes upregulation of compensatory genes, KO males were compared to WT males (see Figure 3.9A). Interestingly, only two genes were shown to be dysregulated: *Pcdh19* was shown to be significantly downregulated ($p = 0.005$) in the KOs, and *β-gal* was shown to be significantly upregulated ($p = 0.0009$).

3.2.6.3. HET females vs. WT females

Similarly, in order to identity potentially differentially expressed genes within the HET animals and elucidate genetic candidates for *Pcdh19* involvement in neurogenesis, HET samples were then compared to WT female samples (see Figure 3.9B). Interestingly, no differentially expressed genes were identified.

3.2.6.4. KO males vs. WT females and HET females vs. WT males

The results from the previous analysis suggested that genotype accounted for a minor proportion of the variance and that sex was the primary source of variation. To determine whether genotype had no effect on the variance within the dataset, it was decided to compare the mutant groups (KOs and HETs) against their opposite WT sex counterparts (see Figure 3.10). If genotype had a minimal influence on the variance, then these two comparisons would show similar results to the WT female vs. WT male comparison, as previously shown and produce an equal list of DE genes. Remarkably, both comparisons revealed numerous DE genes. In the KO male vs. WT female comparison, 39 genes were shown to be upregulated in the KO and 641 were shown to be downregulated (see Figure 3.10A). In the HET female vs. WT male comparison, 22 genes were shown to be upregulated in the HETs, and 17 were shown to be downregulated (see Figure 3.10B) The top 20 DE genes for the KO male vs. WT female and HET female vs. WT male comparison are depicted below (see Figure 3.11). All the DE genes can be found in the appendices at the back of this thesis.

Based on the findings above, it was possible that a synergistic effect existed between sex and genotype, i.e., DE genes were only detectable when utilising both variables in the analysis. To investigate this further, the top 20 DE genes from both analyses were screened by comparing the

normalised expression values of each gene between all genotype groups. It was decided to only focus on the top 20 DE genes because of the high volume of DE genes in the KO vs. WT female comparison, and because only 22 DE genes were uncovered between the HET vs. WT male comparison. It was theorised that if the DE genes from each comparison were dysregulated because of a synergistic effect between sex and genotype, then the expression values would only be different between the respective mutant group (HET or KO) and the opposite WT sex counterpart. However, if genes were dysregulated between the mutant group and the groups of the opposite sex – regardless of genotype – then it could be suggested that these genes are dysregulated due to sex. Likewise, if genes were found to be dysregulated in only one of the mutant groups, then these genes could be considered differentially expressed due to genotype.

By examining the DE genes from the KO vs. WT female comparison, it was found that the top three upregulated genes, *Gad2*, *Dlx6os1*, and *Dlx1*, were noticeably upregulated only in the KO animals (see Figure 3.12). Many of the downregulated genes in the KOs (*Faf1, Akirin2, Vegfa, Sla2*, and *Mrps24*) were also found to be highly expressed in WT females and some HET samples, however expression was relatively low in the WT males. This was the case for all of the other top DE genes. This may potentially suggest that these genes are dysregulated mainly due to sex differences in expression, whereas a small synergistic effect between sex and genotype may also be present.

By examining the DE genes from the HET vs. WT male comparison, it was found that four DE genes, *Bmp6*, *Rpso1, Rpso3*, and *Apcdd1* were noticeably upregulated only in the HET animals (see Figure 3.13). Interestingly, *Cenpw* was also upregulated in the HETs and WT females when compared to males, yet HETs showed slightly increased expression levels compared to WT females. The overall expression of *Six3* and *Six3os1* were also upregulated in the HETs, however this was mainly driven by one HET sample which showed greater expression of both genes; most HET samples showed similar expression to WT females. Lastly, *Pcdhga11* was downregulated in the HETs and WT females, whereas expression was upregulated in the male samples. All other DE genes were shown to be dysregulated in HETs and WT females i.e., due to sex.

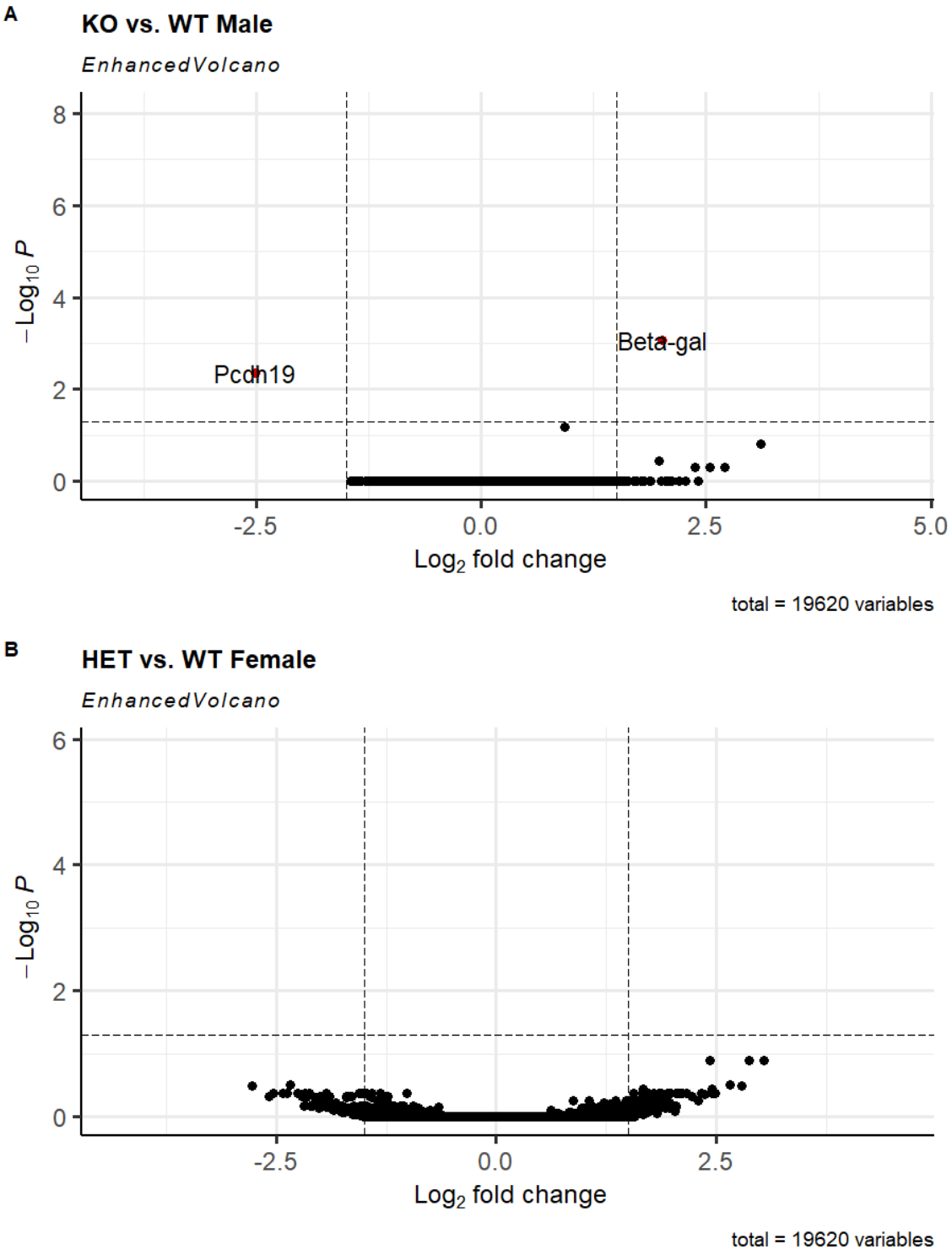


Figure 3.9. Mutant vs. WT same sex counterparts. (A) Volcano plot of the KO vs. WT male comparison. (B) Volcano plot of the HET vs. WT female comparison. Significantly expressed genes are shown in red. Nonsignificant genes are in black. Threshold for significance = p. adjusted < 0.05, cut-off > 1.5.

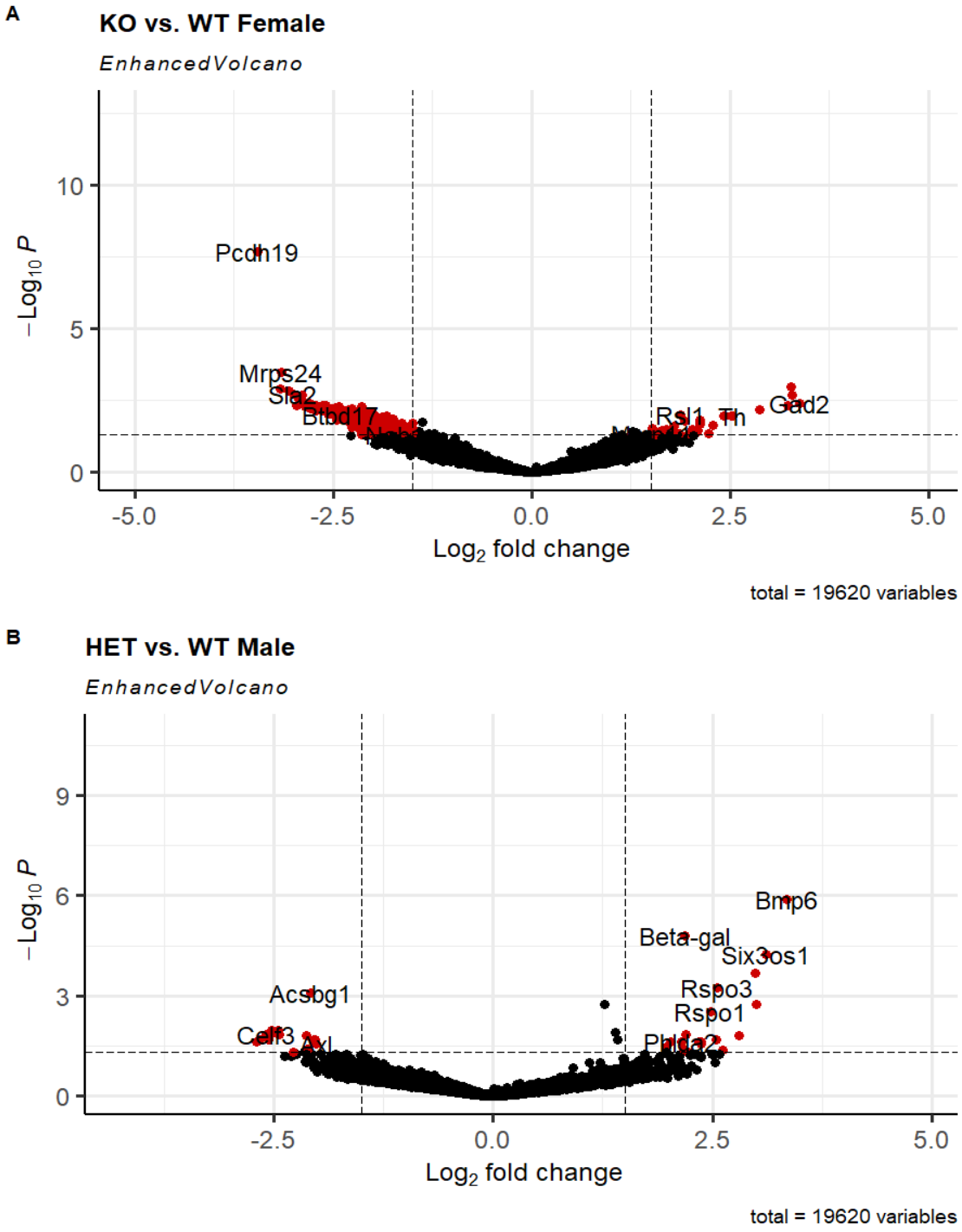


Figure 3.10. Mutant vs. WT opposite sex counterparts. (A) Volcano plot of the KO vs. WT female comparison. (B) Volcano plot of the HET vs. WT male comparison. Significantly expressed genes are shown in red. Nonsignificant genes are in black. Threshold for significance = p. adjusted < 0.05, cut-off > 1.5.

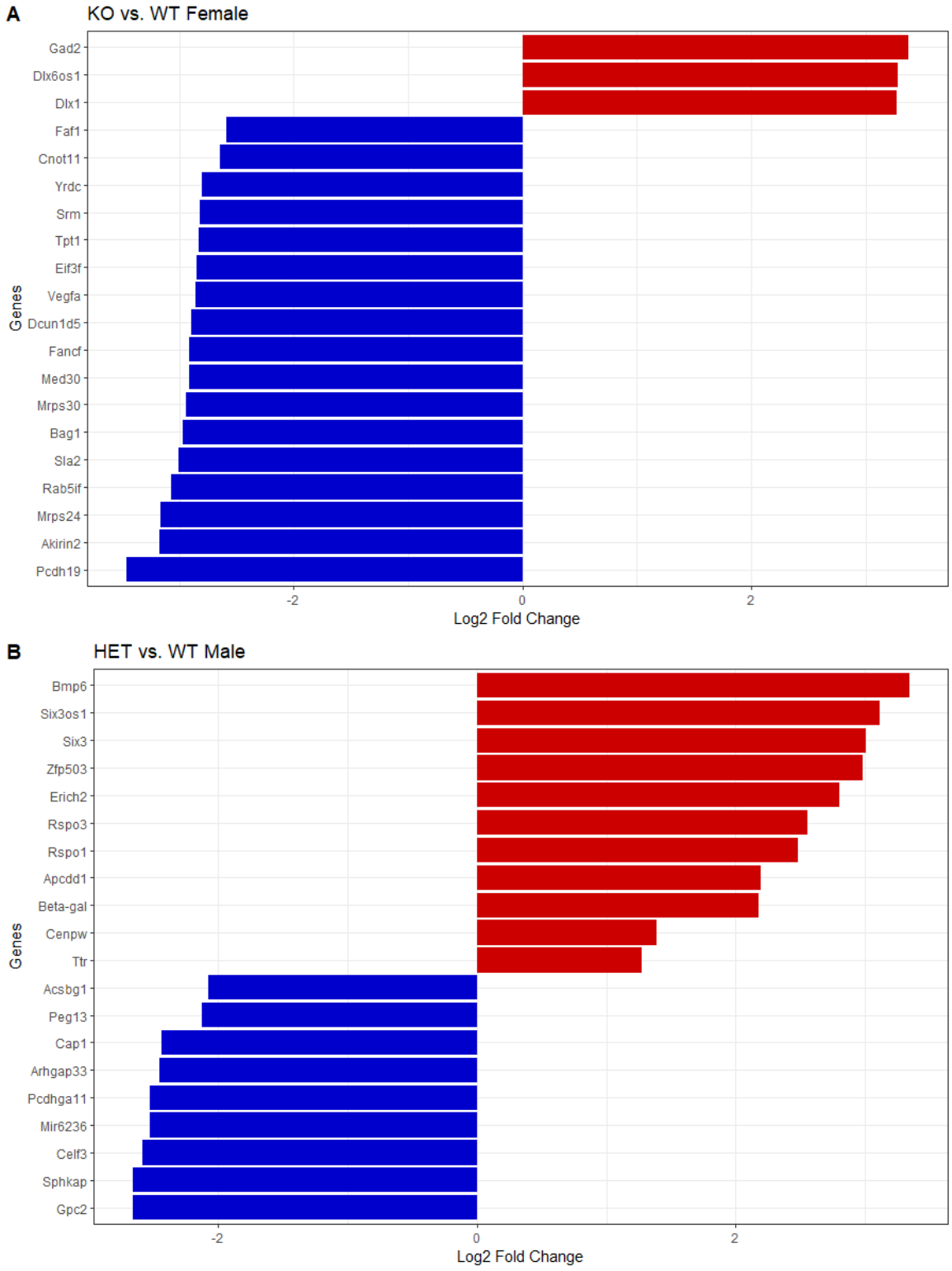


Figure 3.11. Top 20 differentially expressed genes from Mutant vs. WT opposite sex counterpart comparison. (A) KO vs. WT female. (B) HET vs. WT male. Upregulated genes are shown in red. Downregulated genes are shown in blue.

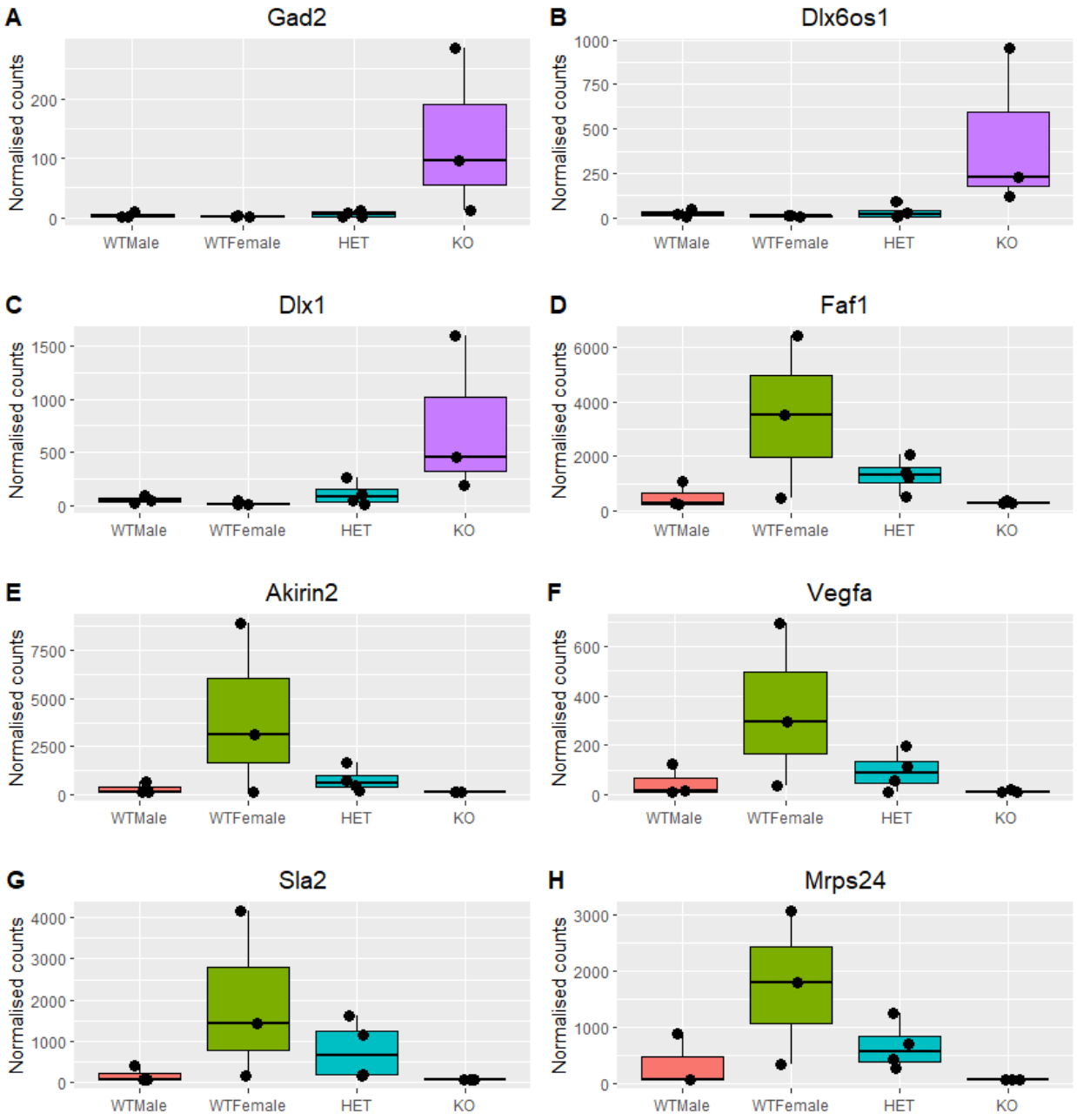


Figure 3.12. Normalised expression counts of DE genes found from KO vs. WT female comparison. Expression of differentially expressed genes found between KO vs. WT female DE analysis and shown for all genotype groups, including *Gad2* (A), *Dlx6os1* (B), *Dlx1* (C), *Faf1* (D), *Akirin2* (E), *Vegfa* (F), *Sla2* (G), and *Mrps24* (H).

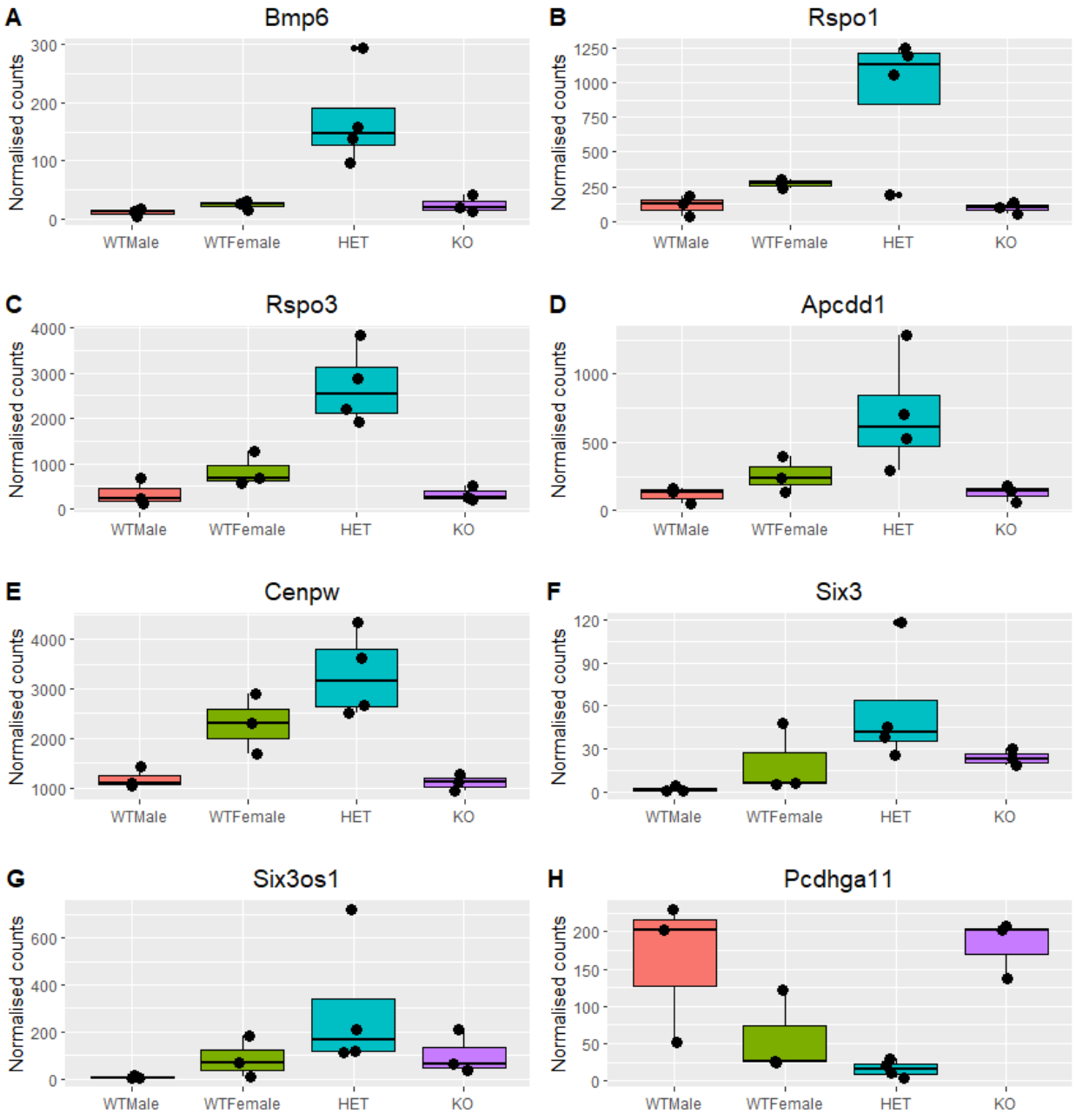


Figure 3.13. Normalised expression counts of DE genes found from the HET vs. WT male comparison. Expression of differentially expressed genes found between HET vs. WT male DE analysis and shown for all genotype groups, including *Bmp6* (A), *Rspo1* (B), *Rspo3* (C), *Apcdd1* (D), *Cenpw* (E), *Six3* (F), *Six3os1* (G), and *Pcdhga11* (H).

3.2.7. Gene set enrichment analysis (GSEA)

As previously discussed, DE analysis of KO male vs. WT males, and HET vs. WT females revealed few or no DE genes. Despite these analyses revealing no large gene expression changes, it is possible that there could be small but coordinated changes in sets of functionally related genes between these groups. To gain deeper insight into the underlying biological differences between KO vs. WT male and HET vs. WT female groups, gene set enrichment analysis (GSEA) was performed using the "clusterProfiler" package (Yu et al., 2012).

GSEA is a functional class scoring (FCS) tool which hypothesises that small changes in expression of genes within pre-defined functionally related gene sets could reflect meaningful biological effects (Subramanian et al., 2005). Therefore, all genes in the dataset are considered in this analysis regardless of their individual scores. GSEA works by ranking the genes of a gene set based on the correlation of their differential expression values (i.e., $\log_2$ fold changes) with the phenotype (i.e., the two groups being compared). The goal of GSEA is to determine whether the genes of a gene set (*S*) are randomly distributed throughout the ranked gene list (*L*) or if they aggregate towards the top of bottom ends (which indicate whether the gene set is "activated" or "suppressed", respectively). An enrichment score (ES) is generated which represents the degree that *S* is over-represented at the extreme ends of *L*. The p-values are then calculated by permutating the gene list *L* and recomputing the ES for the permutated data. This is repeated numerous times to generate a null distribution of enrichment scores; the p-value is then calculated by measuring the original ES value against the null distribution.

It is important to note that GSEA predicts the activation or suppression of gene sets by positive or negative $\log_2$ fold changes only and does not consider the biological function of the genes. For example, if several known repressor genes are upregulated, GSEA will consider the gene set to be "activated" even though the upregulation of these genes likely means that the pathway is biologically supressed. For this analysis, gene sets were chosen using gene ontology (GO) terms. In order to identify statistically and biologically meaningful pathways, the p-value cut off threshold was set to $p < 0.05$.

3.2.8. KO males vs. WT males (GSEA)

GSEA for the KO vs. WT male comparison revealed numerous enriched pathways for “Biological Processes” (GO: BP, 643 terms). For simplicity, these results were filtered by eliminating terms related to non-relevant organ systems (such as heart, liver, skin etc.) in addition to “offspring” terms. Several terms related to cell fate commitment were shown to be “activated” in the KO, including *negative regulation of neurogenesis, cell fate commitment,* and *regulation of neuron differentiation* (see Figure 3.14). Only four of the selected pathways were shown to be “suppressed”, including *regulation of the cell cycle G2/M phase transition*, *DNA repair, cell cycle checkpoint,* and *ncRNA processing*. Interestingly, many terms related to cell adhesion were also shown to be “activated”. These include terms such as *cell-cell junction organisation, extracellular matrix organisation,* and *cell junction assembly.* Moreover, terms related to Wnt signalling were also shown to be “activated” in the KO, such as *cell-cell signalling by Wnt*. GSEA plots were generated to illustrate the running enrichment score of *cell-cell signalling by Wnt*, *cell junction assembly*, and *regulation of neuron differentiation* (see Figure 3.15A). Additionally, category netplots were also generated to show the fold changes of the genes associated with the aforementioned pathways (see Figure 3.15B).

3.2.9. HET vs. WT female (GSEA)

GSEA analysis for the HET vs. WT female comparison revealed fewer enriched pathways (45 terms) compared to the KO vs. WT male comparison. Nevertheless, the terms from the HET vs. WT female analysis were subsequently filtered to remove those related to non-relevant organ systems and to remove “offspring” terms (see Figure 3.16). Several pathways related to ribosomal function were shown to be “activated” in the HETs, including *ribosomal small subunit assembly, ribosomal assembly,* and *rRNA processing*. Only one pathway was shown to be “supressed”, which included *vesicle localisation*. GSEA plots were generated to show the enrichment score of the top dysregulated pathways in the HETs (see Figure 3.17A). Because the top dysregulated pathways in the HETs were related to ribosomal function, only one category netplot was generated, since the shared list of genes related to these pathways were almost identical (see Figure 3.17B).

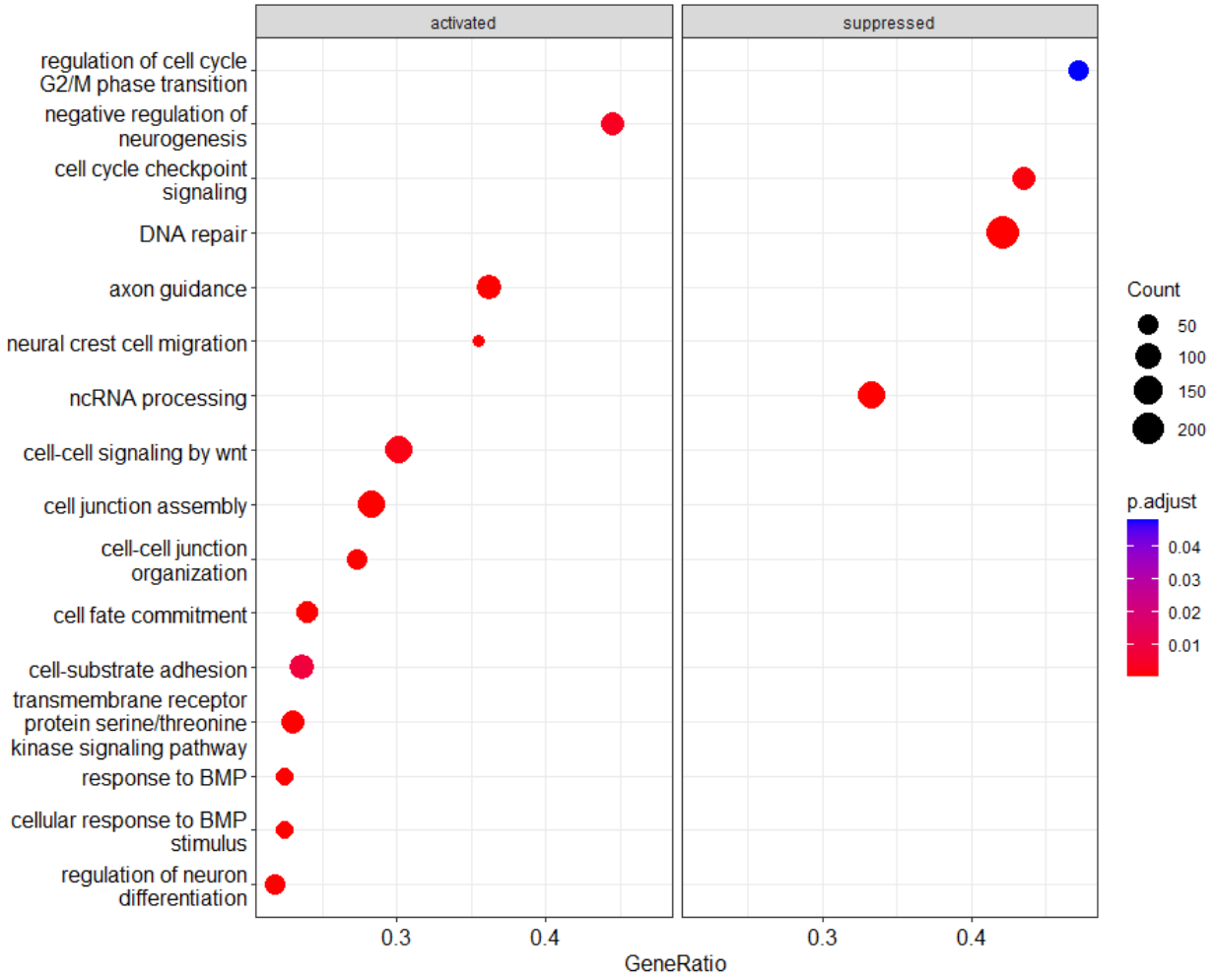


Figure 3.14. Dotplot showing the top 20 enriched GO terms from the KO vs. WT male comparison. Circle size indicates gene count for each GO term. Colour scheme represents p-adjusted values.

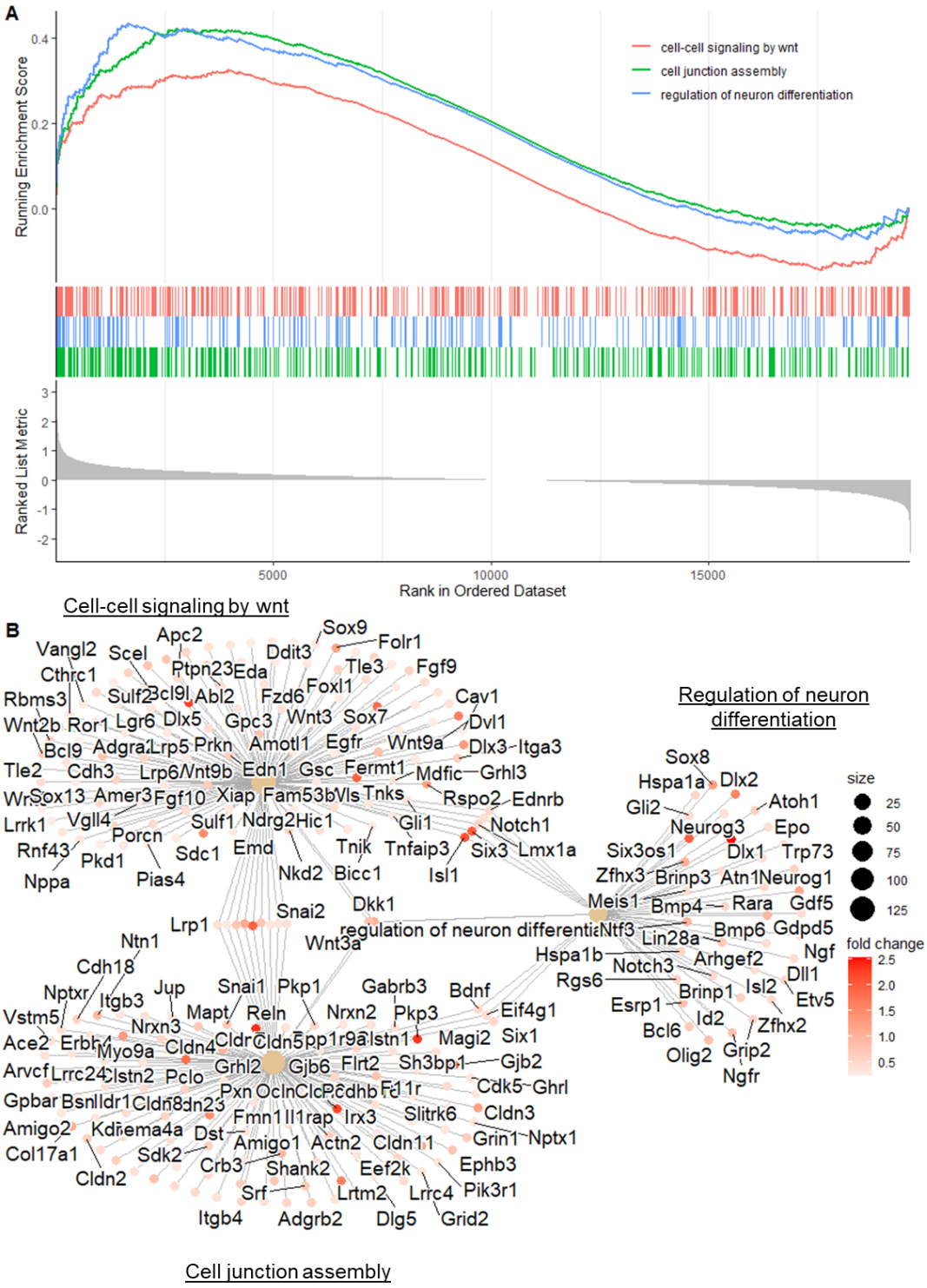


Figure 3.15. GSEA plot and category netplots: KO vs. WT males. (A) GSEA plots of genes associated with the three main enriched pathways. Genes are ranked by $\log_2$ fold changes. (B) Category netplots of the three main enriched pathways. Colour scheme represents the $\log_2$ fold changes of the individual genes.

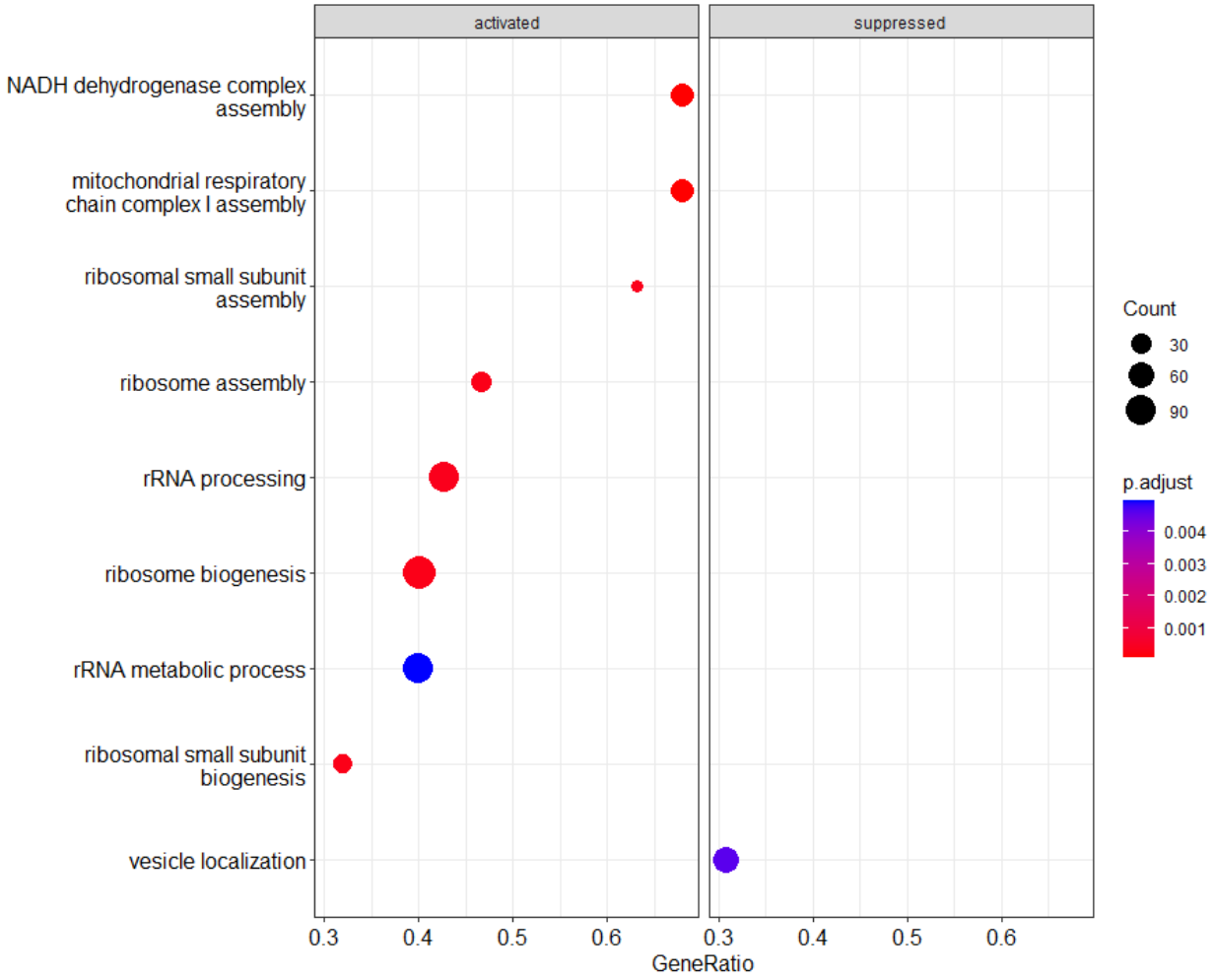


Figure 3.16. Dotplot showing the top enriched GO terms from the HET vs. WT female comparison. Circle size indicates gene count for each GO term. Colour scheme represents p-adjusted values.

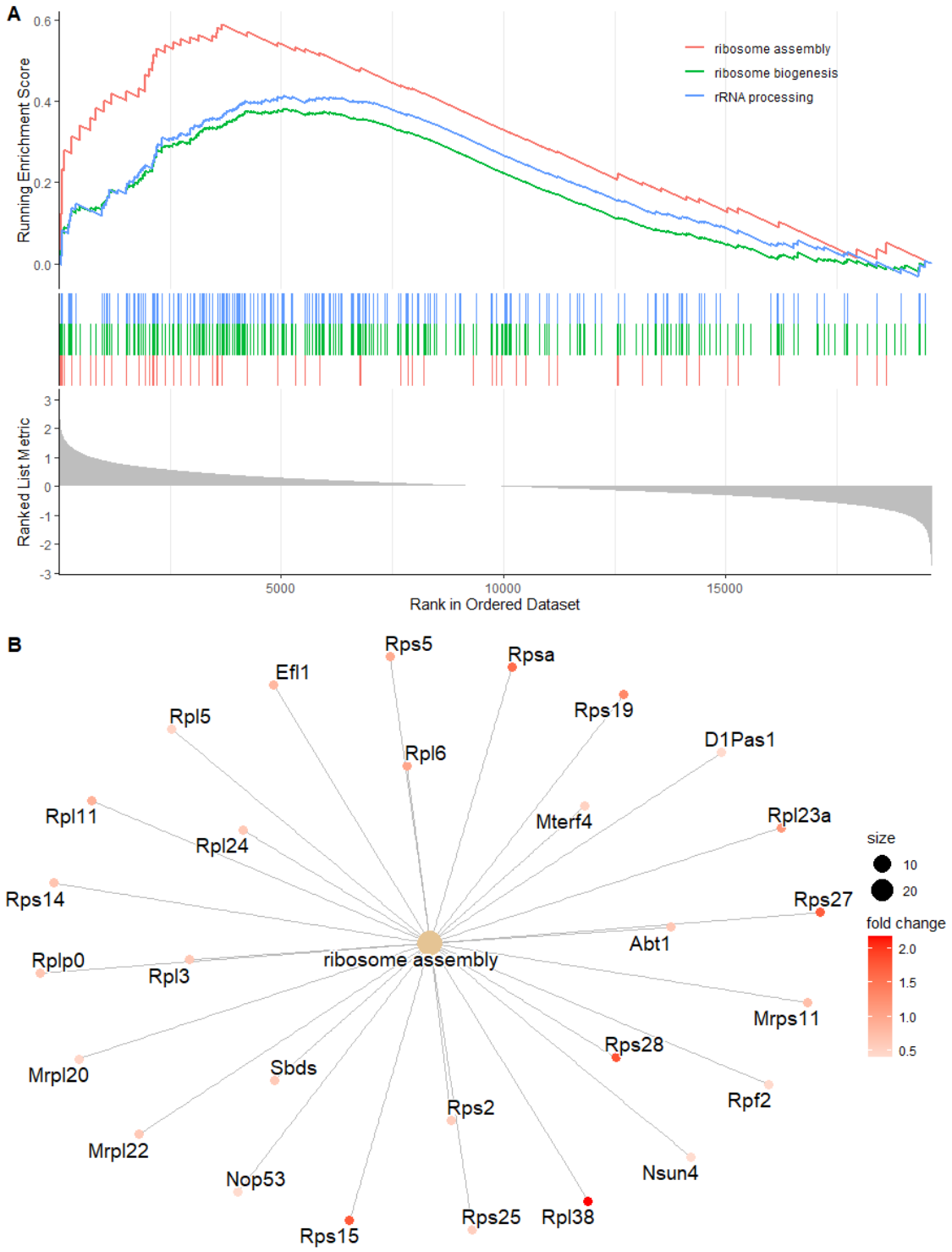


Figure 3.17. GSEA plot and category netplots: HET vs. WT female. (A) GSEA plots of genes associated with the three main enriched pathways. Genes are ranked by $\log_2$ fold changes. (B) Category netplots of the ribosome assembly pathway. Colour scheme represents the $\log_2$ fold changes of the individual genes.

3.3. Discussion

In this chapter, a detailed examination of the transcriptional profiles of cortical tissue from *Pcdh19* WT, HET, and KO E11 embryos was conducted using bulk RNA-seq. To date, there have been few studies that have investigated transcriptional alterations in KO or mosaic *Pcdh19*-expressing neural progenitors (Homan et al., 2018). Moreover, there have been no previous reports analysing gene expression changes from KO or mosaic cells *in vivo*. Therefore, this chapter provides a first insight into the transcriptional landscape of *Pcdh19* WT, HET, and KO embryos which presents the first step in understanding the molecular underpinnings of PCDH19 involvement in neurogenesis.

3.3.1. WT, HET, and KO embryos cluster mainly by sex rather than by genotype

One of the most striking findings from this analysis was that the samples clustered mainly based on sex rather than on any other variable. This was most pronounced for the male samples, which displayed the strongest clustering effect regardless of genotype. The effect of sex was further illustrated by the differential expression (DE) analysis performed on WT male and WT female samples, which revealed *Kdm5d, Xist, Uty,* and *Tsix* as the highest differentially expressed genes found in the entire experiment. Sex was a difficult variable to control for due to the X-linked inheritance of *Pcdh19*. For instance, to control for sex in the analysis, it would have been better to include male and female samples for each genotype group, i.e., male and female WTs, HETs, and KOs. Although it would be possible to generate KO female samples by mating hemizygous KO males (Y X^{KO}) with a homozygous KO female (X^{KO} X^{KO}), because *Pcdh19* is X-linked it is impossible to generate HET males from our model animals. This would require introducing mosaic mutations, which would be much more challenging. It would have been interesting to include KO females in this analysis, however including samples generated from different mating strategies was hypothesised to potentially introduce further extraneous variability so it was decided not to include KO females in the analysis. However, since sex appears to explain a significant proportion of the transcriptional variance at E11, it would be interesting to see whether KO females would also cluster by sex. Taken together, sex was a significant variable in this analysis, which suggests that disrupted expression of *Pcdh19* has a lesser impact on the overall transcriptional signature of the samples than was previously hypothesised.

3.3.2. WT females show unusual Pcdh19 normalised expression levels

The *Pcdh19* gene is considered to be subjected to X-chromosome inactivation (XCI) (Cotton et al., 2015; Dibbens et al., 2008). Moreover, when XCI occurs in humans, the random inactivation

of the normal and mutant *PCDH19* alleles has been thought to be a leading mechanism behind PCDH19-epilepsy known as the cellular interference hypothesis (Depienne et al., 2009). Therefore, in mice, if *Pcdh19* does not escape XCI, the expression levels should be similar between WT males, WT females, and HETs, whereas KOs would show no expression. As predicted, KO samples showed the lowest normalised expression values of *Pcdh19*, and WT males and HETs showed similar expression levels. Interestingly, WT females showed unusual *Pcdh19* expression, wherein two samples showed almost twice or three times the amount of expression than any other sample. During XCI, most genes on the inactivated chromosome are silenced, however it has been found that ~3% genes can escape XCI in mice (Berletch et al., 2011). To investigate whether *Pcdh19* could escape X-inactivation in WT females, qPCR was performed on cDNA extracted from WT male and WT female samples to verify this finding. Although no statistical difference was observed between both groups, it was noted that two WT female samples exhibited almost twice the amount of *Pcdh19* expression levels compared to the WT male average, however one WT female sample showed expression levels that was ~6 times lower than other WT females. The increased variance would therefore explain why no statistical differences were observed between WT males and WT females. However, this result does show that expression of *Pcdh19* between WT females is more variable than between WT males, suggesting that *Pcdh19* may escape XCI. It is possible that *Pcdh19* may not completely escape XCI. For instance, ~10% of XCI escapee genes show variable levels of bi-allelic expression i.e., although the supposed inactivated allele is expressed, its expression levels are markedly lower compared to the activated allele (Fang et al., 2019). Therefore, the qPCR result must be confirmed using more embryo samples and, likewise, it would also be interesting to determine whether *Pcdh19* WT allele showed evidence of XCI escape in the KO cells of HET samples.

3.3.3. Differential expression analysis between mutant and WT groups

One of the most interesting findings from this experiment was that DE genes from the mutant groups were only detected after comparing to the opposite WT sex counterpart. Initially, comparing HETs and KOs to the same sex WT counterpart group in order to control for sex revealed few or no DE genes, which suggested that genotype had a minimal effect on the variance in the dataset. However, this hypothesis was challenged after comparing the mutant groups to the opposite WT sex counterpart, which revealed numerous DE genes. If genotype had a minimal impact on the variance and sex was the most significant variable, then it would be expected that comparing mutants to the opposite WT counterpart would produce a similar result to that of the WT male and WT female comparison. However, since many DE genes were uncovered, this

suggested that transcriptional differences were present in the mutant groups but were only detectable after including genotype and sex into the analysis. This suggests that a partial synergistic effect between sex and genotype was also present in the data.

To determine whether these genes were dysregulated due to a synergistic effect between sex and genotype, the expression of the top 20 DE genes was checked between each group. It was theorised that if these genes were differentially expressed due to a synergistic effect, then it would be expected that the expression values would be mostly different between the respective mutant group (HET or KO) and the opposite WT sex counterpart. However, if the gene expression values were different between the mutant group and the opposite sex groups – regardless of genotype – then it could be concluded that these genes are likely dysregulated due to sex. Likewise, if genes were shown to be dysregulated only in the mutant group, then these could be considered dysregulated due to genotype. For example, in the KO vs. WT female comparison, the three most upregulated genes, *Gad2, Dlx6os1*, and *Dlx1*, were found to be exclusively upregulated in the KOs. Yet since these genes are widely known as interneuron markers, their upregulation suggests that some of the tissue from the of medial ganglionic eminences (MGE) was likely harvested during the dissection. Moreover, most of the downregulated genes in the KO were shown to be widely upregulated in the WT female and the HETs, however the difference was more apparent in the WT females. It can be concluded that these genes were mostly dysregulated due to sex than genotype, but since there were many downregulated genes detected from this comparison, it is possible that some of the genes that were not studied may be exclusively downregulated in the KOs compared to the other groups, or only downregulated when compared to WT females.

Several of the DE genes in the HET vs. WT male comparison were shown to be almost exclusively increased in the HETs, including *Bmp6, Rspo1, Rspo3*, *Apcdd1,* and *Cenpw*. For some of these genes, the HET samples showed noticeable within group variability which might explain why these genes were not considered DE expressed when compared to WT females and after adjusting for multiple comparisons with the other genes. Interestingly, the bone morphogenic protein (BMP) pathway has been known to play an important role in patterning the dorsal telencephalic midline via secretion of BMP morphogens from the cortical hem and choroid plexus (Le Dréau, 2022; Furuta et al., 1997). The BMP pathway has also been shown to promote RGC symmetric and asymmetric divisions during early neurogenesis events (Panchision et al., 2001), and the upregulation of *Bmp6* may hint at a role of the BMP pathway in the developing HET brain. The genes *Rspo1* and *Rspo3* encode proteins that are also secreted morphogens and act by amplifying Wnt signalling (Mulligan and Cheyette, 2012). Interestingly, the gene *Apcdd1* encodes

a transmembrane protein that has been shown to negatively regulate Wnt signalling (Shimomura et al., 2010) and the BMP pathway (Vonica et al., 2020) by interacting with LRP5 and the BMP receptor BMPR1A, which attenuates the signal transductions of both pathways. Thus, APCDD1 has been considered a major regulator of both pathways by coordinating Wnt/BMP signalling during gastrulation, brain patterning, neurogenesis, and axon pathfinding (Vonica et al., 2020). Finally, *Cenpw* encodes for a component of the centromere and is important for kinetochore functioning during cell division. Although few studies have investigated the role of this gene in cortical development, it has been widely studied in the context of cancer and has been shown to underlie invasive proliferation of hepatocellular carcinoma cells (Zhou et al., 2021). Many of the other DE genes in the HET vs. WT male comparison were also found to be upregulated in the WT females, suggesting that these genes are dysregulated due to sex and a partial synergistic effect between sex and genotype, as previously discussed. Taken together, the upregulation of these genes in the HETs hints at a role of both the BMP and Wnt pathways in HET brain development and exploring these further may help understand the unusual neurogenesis phenotype in these animals.

HETs are a mosaic of WT-expressing and KO-expressing RGCs which display altered neurogenic behaviour (unpublished data, 2019). The altered neurogenesis of these RGCs is theorised to be driven by changes in the expression of neurogenic-related genes in both cell populations. Importantly, the transcriptional environment of the WT and KO RGCs are likely different. Therefore, because bulk RNA-seq offers no possibility of separating WT and KO cell populations, it is possible that by merging the cellular profiles, any transcriptional differences underlying the altered neurogenesis phenotype are being masked. This demonstrates a critical caveat of performing bulk RNA-seq on mosaic tissue and therefore, further specialised techniques will be needed to uncover dysregulated genes between WT and KO cells within the HETs (see chapter 4 and 5).

Previous transcriptional analyses between homogenous WT and KO cells *in vitro* revealed numerous DE genes related to neurogenesis (Homan et al., 2018). Therefore, it was surprising in this chapter that the comparison between KOs and WT males – which is equivalent to comparing tissue that is homozygous for *Pcdh19* or KO of *Pcdh19* – revealed no DE genes except for *Pcdh19* and *β-gal*. An upregulation of pro-neurogenic genes in the KO neural progenitors likely underlies the reported increase in neurogenesis rates from these cells (Homan et al., 2018). However, since no differences were observed between WTs and KOs *in vivo*, it was theorised that the transcriptional differences reported *in vitro* could not legitimately represent the molecular

mechanisms by which PCDH19 regulates neurogenesis. Likewise, the fact that no DE genes were uncovered in the KOs when compared to WT males supports the notion that KO of *Pcdh19 in vivo* does not cause major changes in transcription to compensate for the loss of the gene. Based on the results from the gene set enrichment analysis (GSEA), KO of *Pcdh19* rather induces small and coordinated changes in expression of genes related to cell cycle, adhesion, and neurogenesis.

3.3.4. Gene set enrichment analysis reveals activation of pathways in KOs and HETs

Despite finding no differentially expressed genes between the mutant groups and their same sex WT counterparts, valuable information could still be extracted via a pathway analysis such as GSEA. GSEA on KO samples revealed several dysregulated pathways related to neurogenesis and cell junction assembly. Another interesting finding was that "*cell-cell signalling by Wnt"* was significantly enriched in the KOs. Previous reports have illustrated that protocadherins can module Wnt signalling. For example, PcdhγC3 has been shown to negatively regulate canonical Wnt signalling by interacting with Axin1 and hindering its involvement in LRP6 phosphorylation, thus inhibiting the intracellular cascade caused by Wnt binding (Mah et al., 2016). Moreover, several δ-Pcdhs, including Pcdh19, have been shown to interact with the non-canonical Wnt receptor Ryk in zebrafish and control the response of neural progenitors to Wnt activity (Biswas et al., 2021). Taken together, these data might suggest a role of PCDH19 in Wnt signalling and loss of *Pcdh19* may cause small adjustments in signalling activity that would sustain normal neurogenesis.

GSEA revealed that several pathways related to ribosomal function were enriched in the HETs. This unusual result could suggest that protein synthesis rates or translation events are dysregulated in the HETs. Interestingly, it has been reported that during forebrain development in mice, neuroepithelial cells at E8 showed a significant upregulation of genes related to protein biosynthesis (Chau et al., 2018). Moreover, in the same study GSEA revealed that several pathways related to ribosomes, including *Ribosomes/Translation* and *Ribosome Biogenesis* were upregulated at E8 but noticeably decreased by E10 (Chau et al., 2018). Therefore, enrichment of pathways related to ribosomal function may serve as a biomarker signature that defines the developmental age of progenitor cells. Enrichment of these pathways in the HETs could therefore indicate that these progenitors are behaving at a different 'age' of development. However, because it has been shown that WT and KO progenitors within the HET animals produce neurons at different rates (unpublished data, 2019), it is possible that the ribosomal effects are different within the WT and KO cells. However, as previously mentioned, since WT and KO profiles cannot

be separated in this analysis, it is not possible to examine this hypothesis using conventional bulk RNA-seq methods.

3.3.5. Conclusion

This chapter provided the first look at the transcriptional landscape of cortical tissue taken from *Pcdh19* WT, HET, and KO embryos at E11. Overall, although samples obtained from animals of the same genotype were predicted to share similar transcriptional profiles, most samples clustered by sex regardless of genotype. Interestingly, it was found that KO of *Pcdh19* does not induce major transcriptional changes *in vivo*. Instead, small but coordinated changes in the expression of genes related to neurogenesis and Wnt signalling were discovered, which may underpin the normal neurogenesis rates that have been observed from KO embryos (unpublished data, 2019). Ribosomal function and protein biogenesis may also be disrupted in HETs; however, WT and KO progenitors could not be separated in this analysis due to the limitations of bulk RNA-seq. The following chapter aims to resolve this issue by employing single cell RNA-seq (scRNA-seq) to perform DE analysis on RGCs from WT, HET, and KOs at the single cell level.

Chapter 4: Single cell RNA-seq analysis of Protocadherin 19 WT, HET, and KO embryos Part I: Quality control and cell clustering analysis

4.1. Introduction

Recent publications, as well as research from the IMG lab, have begun to unravel the role of PCDH19 as a key regulator of neurogenesis in neural progenitor cells, (Biswas et al., 2021; Fujitani et al., 2017; Homan et al., 2018, unpublished data, 2019). Transcriptional analysis has been performed – via microarray – on *Pcdh19* knockout (KO) progenitors *in vitro*, revealing numerous differentially expressed (DE) genes related to neurogenesis and differentiation (Homan et al., 2018). Moreover, mosaic expression of *Pcdh19 in vivo* has also shown to cause wild type (WT) and KO progenitors to undergo neurogenesis at different rates (unpublished data, 2019). Interestingly, hemizygous KOs did not display altered neurogenesis *in vivo*, suggesting that compensatory mechanisms can occur when *Pcdh19* is ablated in a tissue environment. Although transcriptional analysis has been performed on *Pcdh19* KO progenitors *in vitro*, a detailed transcriptional analysis of perturbed *Pcdh19* expression *in vivo* has yet to be completed.

In chapter 3, a first glimpse into the transcriptional landscape of the developing cortex of WT, heterozygous (HET), and KO embryos was accomplished by applying bulk RNA-seq to whole tissue samples. A significant limitation in this experiment is that because RNA is sequenced from the whole lateral cortex, reads from particular cell types cannot be isolated. For example, it was not possible to separate the reads of radial glial cells (RGCs) from intermediate progenitors (IPs) or neurons, or potential non-neuronal contaminants carried over from the dissection such as meningeal cells etc. This also meant that reads from WT and KO cells within the HET tissue could not be compared. Nevertheless, the greater sequencing depth offered by bulk RNA-seq meant that expression data could be collected at levels not possible using alternative transcriptional analysis techniques (Ding et al., 2020). Moreover, because the E11 cortex is mostly abundant in RGCs (Götz and Huttner, 2005), it was theorised that the dataset would be enriched in reads from those cells. However, DE analysis between HET, KO and their WT same sex counterparts revealed no DE genes, and there was significant co-variability between genotype and sex. Therefore, despite the greater count depth, it is possible that the averaging effect that occurs by merging reads from all cells likely masked the transcriptional features that underlie the unusual neurogenesis phenotype in HET animals and the compensatory mechanism within the KOs.

To further investigate the transcriptional alterations within the HET and KOs, single cell (sc)RNA-seq was employed. Unlike bulk RNA-seq, scRNA-seq permits the comparison of expression profiles at the single cell level. This means that scRNA-seq is ideal to assess transcriptional similarities and differences within a population of cells, and reveals much of the biological variability that is missed by pooling cells for RNA-seq (Haque et al., 2017). Moreover, scRNA-seq has been widely used in the context of brain development to examine the temporal dynamics and population diversity of RGCs (Ruan et al., 2021; Telley et al., 2019). Therefore, because scRNA-seq can be used to perform analysis on heterogenous cell samples at an unprecedented resolution, this makes it an ideal technique to use to study the transcriptional layout of WT, HET, and KO embryos.

scRNA-seq is a rapidly growing field and thus, multiple protocols have emerged that offer different advantages to single cell analysis (Ding et al., 2020; Eze et al., 2021; Ruan et al., 2021; Telley et al., 2019). Despite many differences, scRNA-seq protocols can be broadly categorised into two groups: low-throughput and high-throughput protocols. Low-throughput protocols such as SMART-seq2 typically use plate-based sorting methods to separate single cells into individual wells of a 96 or 384 well plate. The cells are then lysed, and the mRNA is extracted for sequencing. Low-throughput techniques therefore only collect data from a few hundred cells however they offer the highest sequencing sensitivity (Haque et al., 2017). High-throughput techniques such as 10X chromium use droplet-based cell partitioning combined with molecular tagging which permits the capture and sequencing of thousands of cells, but at the cost of reduced count depth. The type of protocol used plays into the nature of the research question; high-throughput protocols are typically used for identifying novel cell types (Ruan et al., 2021) whereas low-throughput protocols are used when the maximum amount of information from each cell is needed (Telley et al., 2019). However, it must be emphasised that all scRNA-seq protocols are good at capturing moderate to the most abundant RNAs in each cell (Haque et al., 2017). SMART-seq2 was initially considered for this experiment since one of the main goals of the project was to examine the transcriptional differences between WT and KO cells within the HET animals. However, because SMART-seq2 only captures several hundred cells, it was decided to use 10X chromium to capture as many cells of interest as possible for downstream analysis.

4.1.1. Aim

The first aim of this chapter is to perform the initial characterisation of the scRNA-seq dataset by subjecting it to a series of quality control (QC) metrics. The dimensionality of the dataset and the key variables that drove the variance were also explored. As previously noted, dissection of E11 cortical tissue provides samples abundant in RGCs, however a small proportion of other cell types are also present, including IPs, neurons, and various non-neuronal cells such as meninges, vascular, and endothelial cells. Therefore, another aim of this chapter was to identify the cell types in the dataset using cell cluster analysis.

4.2. Results

4.2.1. RNA quality and sample processing

Cortical samples were dissected from E11.5 *Pcdh19* WT, HET, and KO embryos and the RNA was extracted and sequenced, as previously described (see Section 2.12). Samples were collected from two different litters across two separate days in order to achieve at least two biological replicates per group (final group count: WT male = 2, WT female = 2, HET female = 2, KO male = 2). It is important to note that cells from WT male and WT female samples were mixed and then ran on the same well on the Chromium Chip-B cassette (see Figure 4.1). It was predicted that male and female cells could be separated using the sex markers identified in the previous chapter (see Section 3.2.7). Additionally, because uncovering DE genes within the HETs was a primary focus of this experiment, it was decided to run the HET samples twice i.e., on two wells each in order to maximise the number of captured cells for this group (see Figure 4.1).

All samples were checked for RNA integrity by Angela Marchbank at the Genomics Hub, Cardiff University. Moreover, all were shown to possess a RIN score of 8.0 or above (see Figure 4.1). Sample concentrations were assessed using the QUIBIT and were shown to range between 22.1 ng/µl – 68 ng/µl. After sequencing, FASTQ files were quality checked using MISEQC. All samples passed quality control metrics and were therefore aligned, filtered, and counted using the Cell Ranger pipeline. The individual sample files were then merged and revealed a total number of 46,606 captured cells.

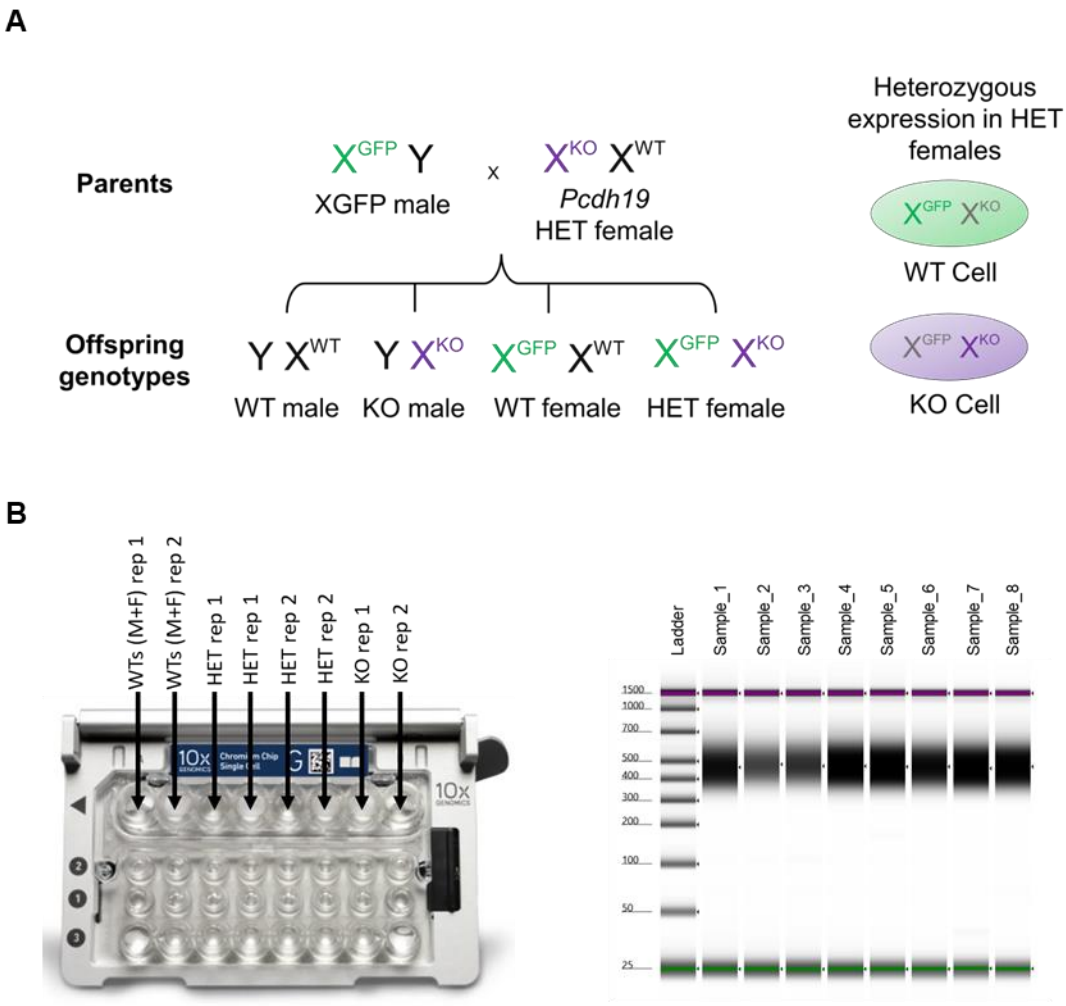


Figure 4.1. RNA samples for single cell RNA-seq passed RIN quality control. (A) Visual aid showing the offspring genotypes after mating an *XGFP* mate with an *Pcdh19* HET female. The diagram also depicts an example of the mosaic expression of the *XGFP* and *Pcdh19 KO* alleles in the HET females. The offspring genotypes highlighted in this diagram were used as the main groups for the sc-RNA-seq experiment (WT male n = 2, WT female n = 2, HET female n = 2 (sorted twice), KO male n = 2). (B) Diagram representing the loading of the sample replicates (rep) into the Chromium Chip B cassette (left) and Tapestation gel showing the density of RNA for all samples (right).

4.2.2. Quality control and cell filtering

The data was first subjected to quality control (QC) checks to identify and remove cells of poor-quality data. Currently, there is no consensus on the exact QC filtering strategies, however three QC covariates can be employed to identify poor-quality cells (Luecken and Theis, 2019). These include the number of counts per cell (count depth), the number of genes per cell, and the percentage of counts corresponding to mitochondrial genes per cell (Ilicic et al., 2016; Stegle et al., 2015). Small library sizes, low number of detected genes, and high fraction of mitochondrial gene content, are markers of cells with degrading RNA that are possibly undergoing apoptosis. Likewise, cell profiles that exhibit unusually large count depths may be indicative of doublets. It is also worth examining the proportion of counts corresponding to ribosomal genes which often make up a large fraction of reads from neuronal-based datasets (~15-20%).

To ensure that all single cell gene expression data corresponded to viable cells, the distribution of these QC covariates was examined from the whole dataset (see Figure 4.2). Using thresholds set by Seurat, it was revealed that most captured cells had a high read depth (cells with > 2000 RNA counts = 60%). Moreover, the frequency of doublets was low (cells with > 20,000 RNA counts = 15%). Interestingly, approximately half of the cells showed a high number of gene features (cells with > 2000 genes = 48%). To determine the quantity of potential apoptotic cells, it was shown that most cells exhibited low expression of mitochondrial genes (cells < 10% mitochondrial gene content = 97%). Moreover, most cells also exhibited normal levels of ribosomal gene content (average ribosomal gene content = 17% and cells < 30% ribosomal gene content = 72%).

It was found that there was a strong positive correlation between count depth and the number of gene features (Pearson's r = 0.79) (see Figure 4.3A), indicating that cells that possessed a high number of genes also exhibited greater count depth, as expected. This correlational analysis also showed two covariate peaks, wherein a subset of cells exhibited moderately low gene features (< 5000 genes) but relatively high count depth (see Figure 4.3A, arrow). It is common that datasets containing heterogenous mixtures of cell types display multiple covariate peaks (Luecken and Theis, 2019). Therefore, to determine the biological relevance of these cells, they were kept for the subsequent identification analyses. Further correlational analysis revealed no correlation between count depth and mitochondrial content (Pearson's r = -0.07), illustrating that those cells exhibiting high count depth had low fraction of mitochondrial genes (see Figure 4.3B). Moreover, a small negative correlation was observed between count depth and ribosomal content (Pearson's r = -0.25) (see Figure 4.3C). Visual examination illustrated a broad covariate peak, wherein the

cells with high count depth had moderate ribosomal content (~0-35%) and cells exhibiting high fraction of ribosomal content had low count depth. Furthermore, there was a subtle negative correlation between mitochondrial content and ribosomal content (Pearson's r = -0.19) (see Figure 4.3D).

QC covariate analysis demonstrated that cells exhibiting high mitochondrial content showed low count depth which is therefore a reasonable indicator of poor-quality cells (see Figure 4.3B and Figure 4.4D). Interestingly, cells with high ribosomal content also displayed low count depth, however ribosomal content is a less clear indicator of cell quality since the cells that exhibited high ribosomal gene content had low mitochondrial gene content (see Figure 4.3D). Furthermore, the analysis revealed the presence of cells exhibiting low gene features but high expression activity which required further exploration. Taken together, by comparing the QC covariates, several univariate thresholding decisions were made (see Figure 4.4). To ensure the subsequent analyses were performed on good quality cells, a conservative filtering approach was used. It was decided to remove cells with a count depth of less than 2000 reads and cells with less than 1000 gene features. Cells with a count depth of greater than 75,000 were considered doublets and were also removed. Moreover, cells with greater than 10% mitochondrial gene content were also eliminated. Collectively, the final cell count was 26,944 cells.

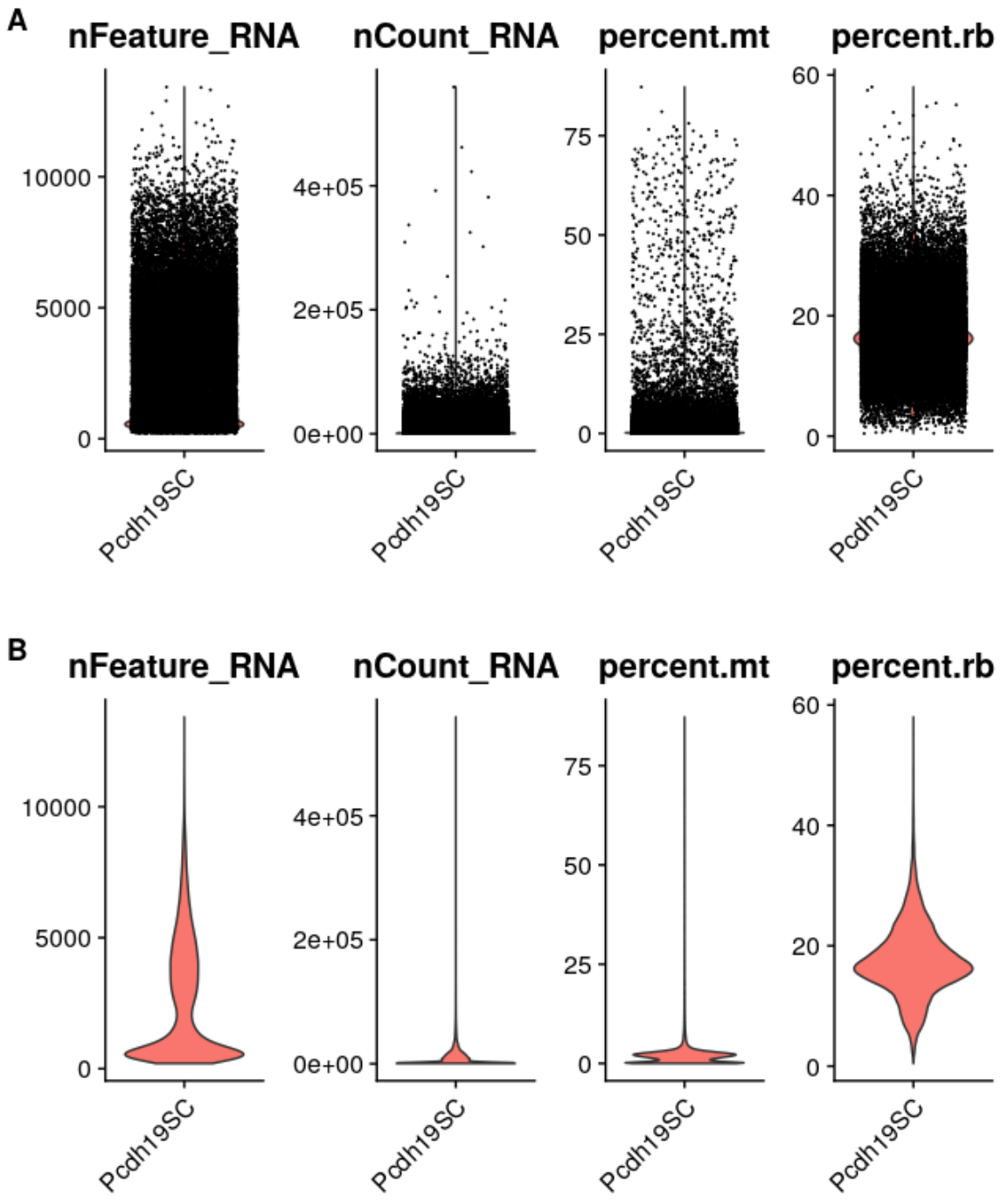


Figure 4.2. Distribution of quality control covariates. (A) QC covariates commonly used for examining the data quality of captured cells. (nFeature_RNA = number of gene features, nCount_RNA = count depth, percent.mt = fraction of mitochondrial genes, percent.rb = fraction of ribosomal genes). (B) The same QC covariates plotted without datapoints to better illustrate the violin distribution.

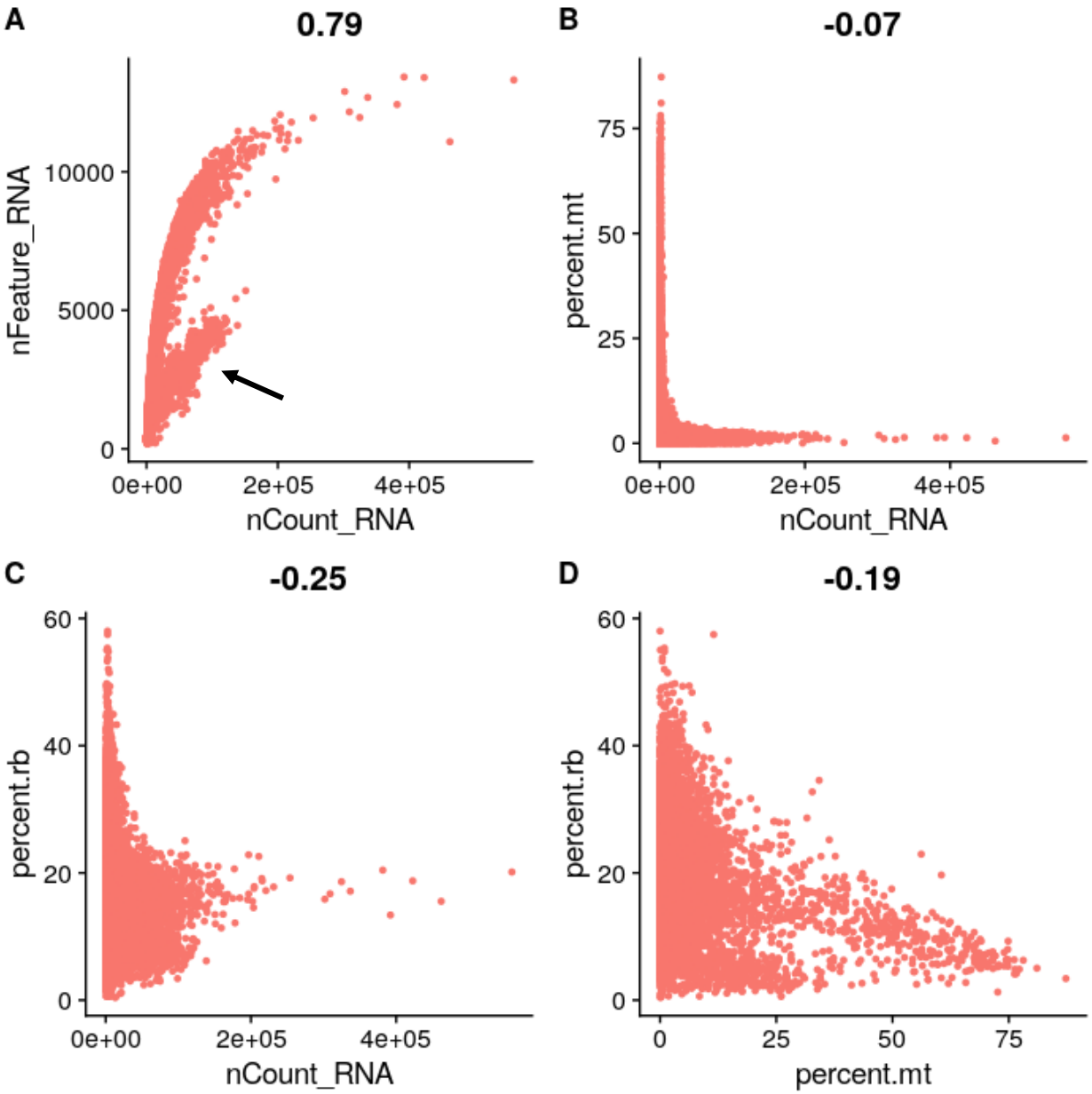


Figure 4.3. Correlational analysis between QC covariates. (A) Correlation between number of detected genes (nFeature_RNA) and count depth (nCount_RNA). The arrow indicates a second covariate peak of cells with moderately low gene features but increased count depth. (B) Correlation between fraction of mitochondrial genes (percent.mt) and count depth. (C) Correlation between fraction of ribosomal genes (percent.rb) and count depth. (D) Correlation between fraction of ribosomal genes and fraction of mitochondrial genes. Correlation coefficients are shown above in each graph.

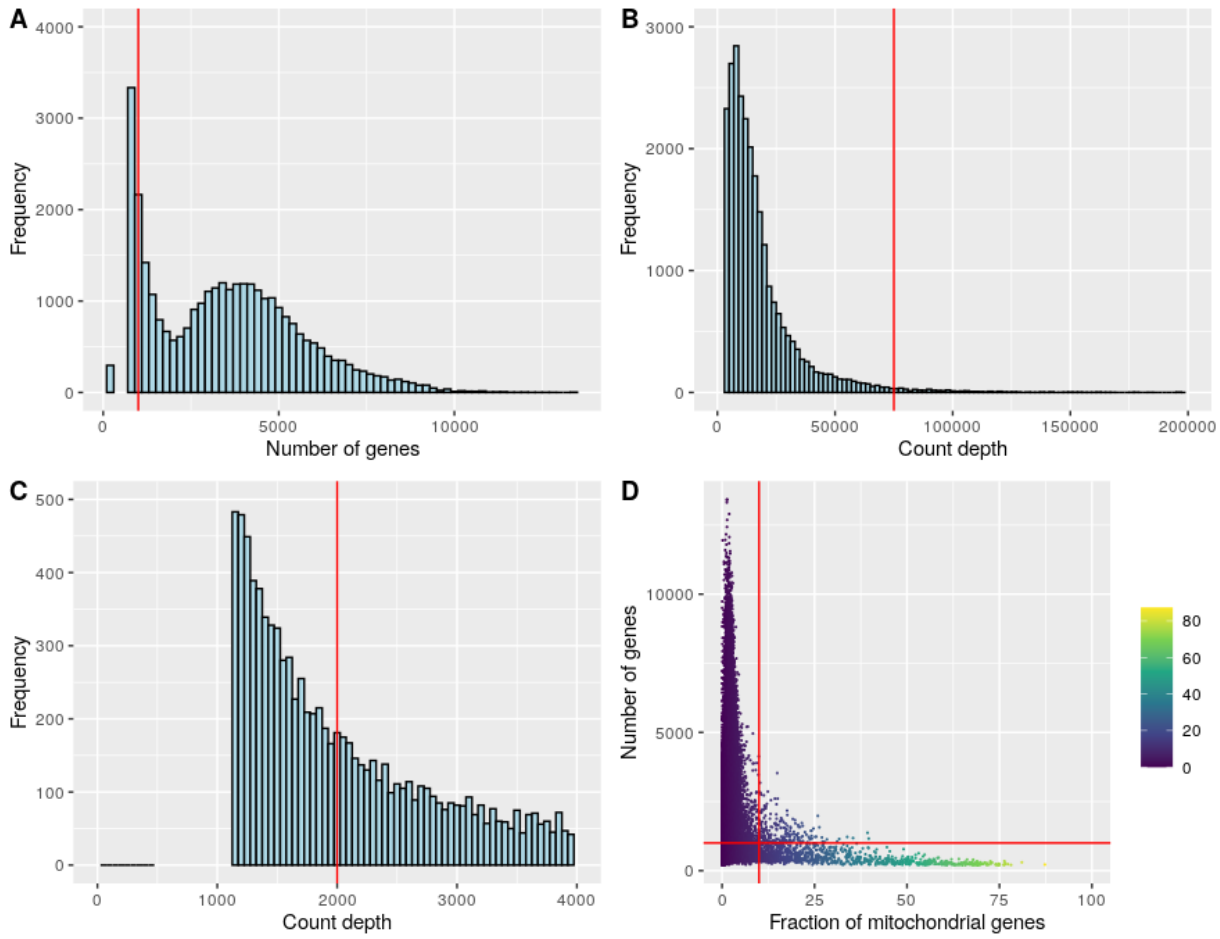


Figure 4.4. Quality control covariates with filtering decisions. (A) Histogram of number of genes detected per cell. Red line indicates the threshold applied to remove cells with less than 1000 genes. (B) Histogram of count depth per cell. Red line indicates the threshold applied to remove cells with greater than 75,000 reads i.e., doublets. (C) Histogram where the count depth has been zoomed-in to view cells with depths below 4000. Red line indicates the threshold applied to remove cells with less than 2000 reads. (D) Scatterplot showing the number of genes vs. fraction of mitochondrial genes and coloured by the fraction of mitochondrial genes. Red lines display the gene count threshold and mitochondrial threshold as noted above.

4.2.3. Feature selection, principal component analysis (PCA), and dimensionality reduction

Another critical step in scRNA-seq analysis is dimensionality reduction. A typical scRNA-seq dataset will contain up to 15,000 variance dimensions (Luecken and Theis, 2019) and much of this information may not be useful for downstream analysis. Therefore, several dimensionality reduction techniques were employed to identify the primary drivers of the variance in the dataset and to reduce computational burden and noise for future downstream analysis.

A common first step in dimensionality reduction is feature selection, wherein the dataset is filtered to include only the genes that are informative on the variability of the dataset, known as high variable genes (HVGs) (Brennecke et al., 2013). For this experiment, HVGs were chosen by selecting the top 2000 genes with the highest variance-to-mean ratio (see Figure 4.5A). Interestingly, many of the most highly expressed HVGs were related to blood cell function (including *Hbb-bs, Hbb-bt, Hba-2a*, and *Hbb-bh1*) and the immune system (including *C1qc, C1qb*, and *Tyrobp*). These HVGs indicated the presence of blood cells in the dataset which possessed highly conserved expression profiles when averaged to other cells.

To further investigate the underlying factors on the variance in the dataset, principal component analysis (PCA) was employed only on the cells expressing the 2000 HVGs identified previously. To identify the PCs that capture the most residual variance, the top 50 PCs were ranked based on the percentage of variance explained by each PC (see Figure 4.5B). It was found that PC1-5 covered a high proportion of the variance in the dataset, however PC-6-20 also accounted for a significant proportion. To examine this further, the top genes associated with PC1-6 were visualised (see Figure 4.5C). PCA indicated the presence of numerous non-neuronal-related genes that drive a major proportion of the variance in the dataset, including several genes related to meninges (such as *Col4a2*, *Col4a1*, and *Col3a1*) and endothelial cells (such as *Cdh5*, and *Cdh11*). Nevertheless, several genes related to progenitors and neurons were also present, including *Eomes, Tubb3, Tbr1,* and *Dcx.* This indicated that cell type, including neuronal and non-neuronal cell types, was a major driver of the variance of the dataset. Taken together, the dataset was filtered to include cells that expressed the top 2000 HVGs and the dimensionality was reduced to include PC1-20.

Another important step in dimensionality reduction is correcting for unwanted variability caused by technical and biological covariates (Büttner et al., 2018; Luecken and Theis, 2019). The most common technical effects typically include count depth and batch effects. Batch effects occur

when the cells are harvested at different times which may influence the measurement of the transcriptome (Luecken and Theis, 2019). Likewise, count depth effects are typically corrected for during normalisation however residue artefact may still be present due to poor cell sampling. To investigate whether count depth and batch affected the variance in the dataset, both variables were visualised using PCA and uniform approximation and projection method (UMAP) plots (see Figure 4.3). It was found that most cells co-clustered regardless of count depth, indicating that depth was properly corrected for during normalisation (see Figure 4.6A). Interestingly, a sub-cluster of cells did display noticeably reduced count depth, suggesting these cells may be poor quality. Therefore, it was decided to remove these cells from further downstream analyses. Moreover, it was also shown that most cells appeared to overlap regardless of batch except for a small proportion of cells from batch 1, indicating that batch had a minimal impact on the variance of the dataset (see Figure 4.6B). Therefore, batch was initially not corrected for to minimise the effect of over-correction (Büttner et al., 2018).

The most common biological covariate that is corrected for is the effect of cell cycle (Luecken and Theis, 2019). To examine cell cycling effects on the data, cell cycle scores were generated and annotated onto the first two PCs (see Figure 4.6C). Interestingly, cells also appeared to overlap regardless of Phase, however there was a noticeable separation of some G1 cells along PC2. Likewise, UMAP analysis revealed that G1 cells formed distinctive clusters away from S and G2M cells. Moreover, G2M and S cells were also separated into subclusters but showed stronger transcriptional similarity than cells in G1. Although cell cycle phase could be corrected for in the dataset, biological data correction is largely contextual and should only be performed if the covariates show evidence of interfering with the data (Büttner et al., 2018). Likewise, biological covariates such as cell cycle may provide useful information on the underlying biology which may be masked if corrected for. Therefore, cell cycle phase was not corrected for in the dataset.

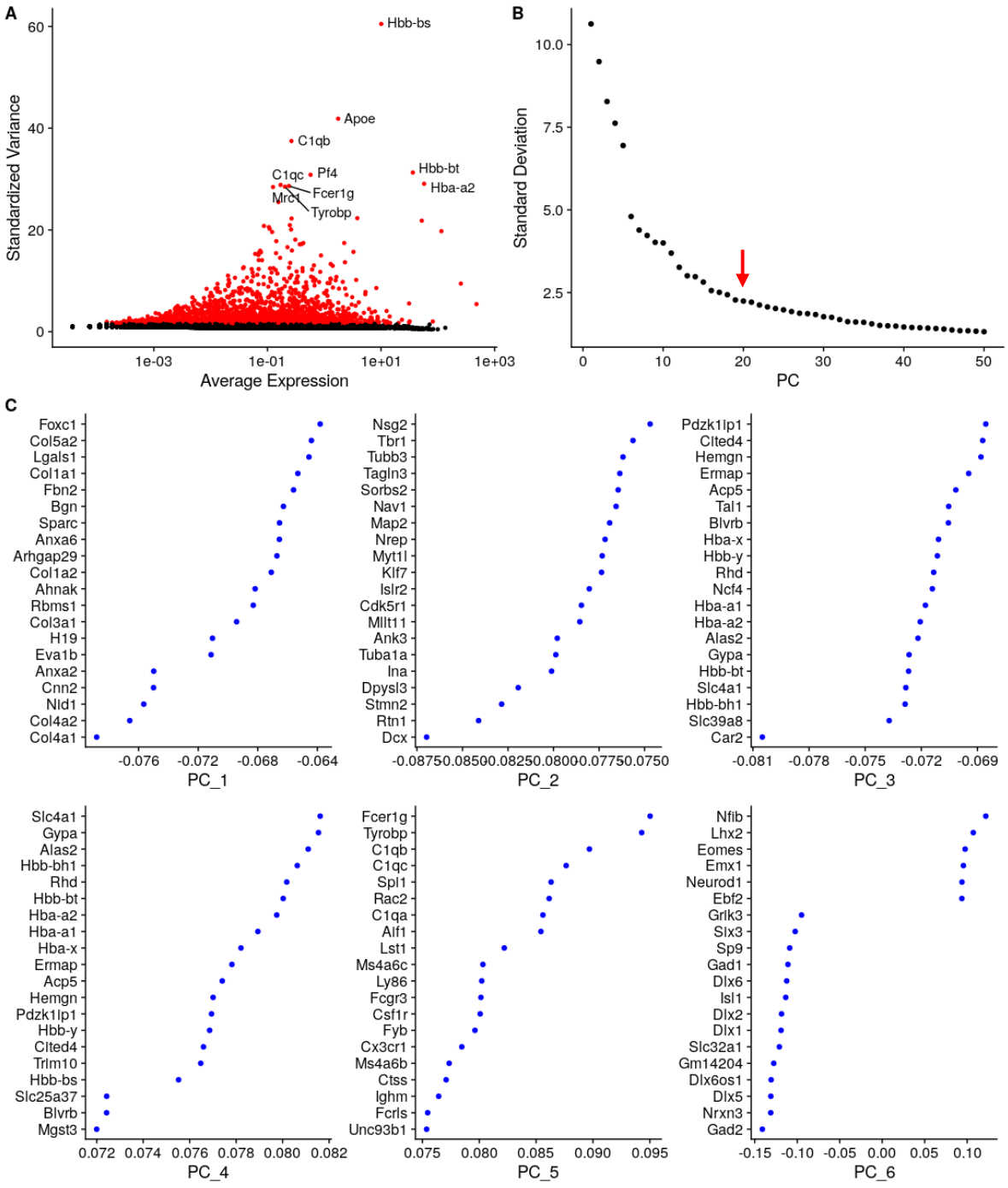


Figure 4.5. Feature selection and principal component analysis. (A) Scatterplot showing the top 2000 HVGs in the scRNA-seq dataset. HVGs are labelled in red. (B) Elbow plot illustrating the proportion of variance explained by PC1-50. Red arrow indicates the PC range used for dimensionality reduction for further downstream analysis. (C) Dot plot illustrating the top 20 genes associated with the first six PCs.

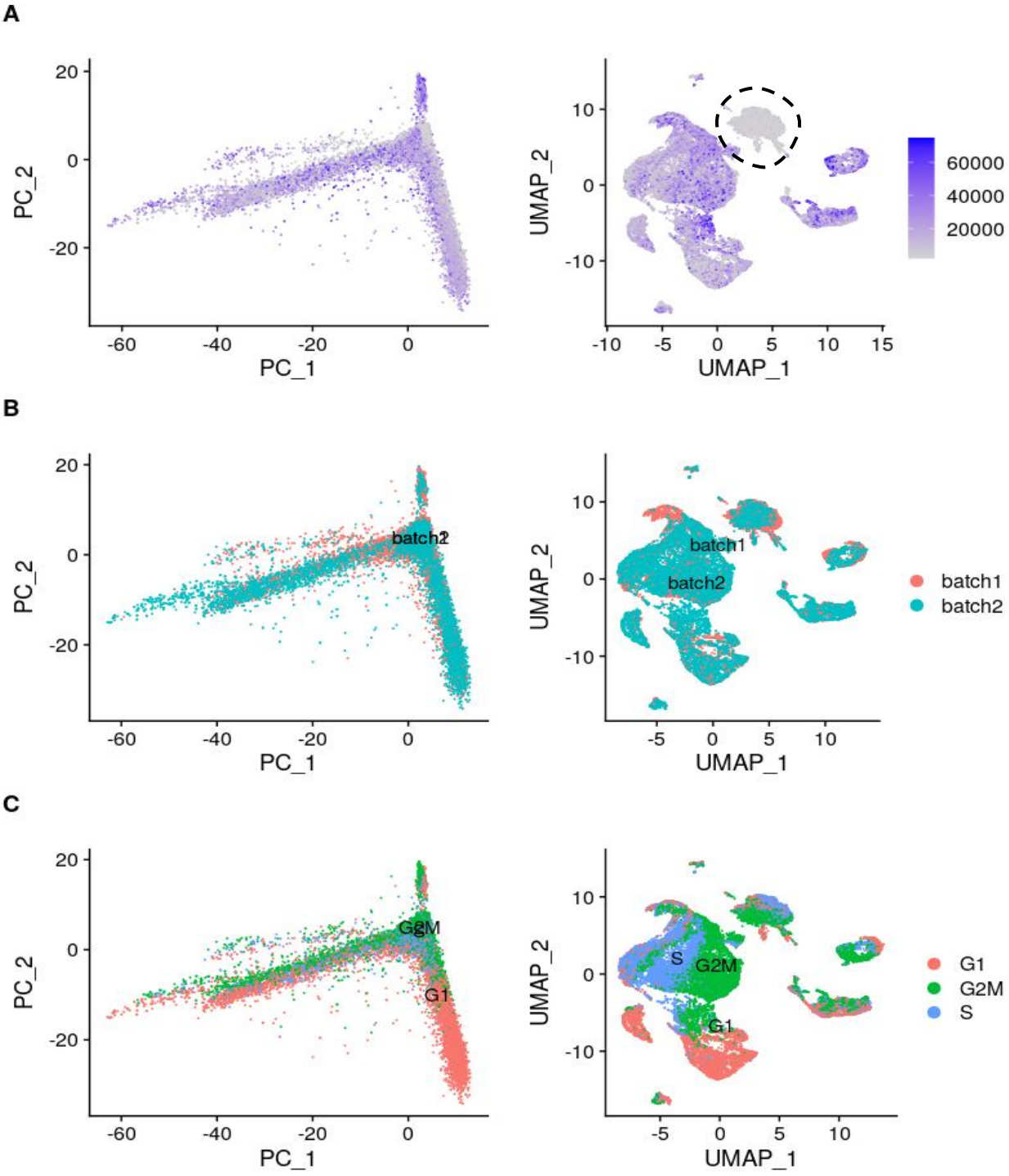


Figure 4.6. PCA and UMAP of technical and biological covariates. (A) PCA and UMAP analysis illustrating the effect of count depth on the variance in the dataset. The encircled cluster showed low RNA levels and was removed. (B) PCA and UMAP analysis illustrating the effect of batch on the variance in the dataset. (C) PCA and UMAP analysis illustrating the effect of cell cycle phase on the variance in the dataset.

4.2.4. Cell cluster analysis

After dimensionality reduction, the dataset was subjected to cell clustering analysis to begin inferring on the identity of the cells. Cell clusters are obtained by analysing the expression profiles of the cells and grouping them based on similarity scoring. Similarity scores were obtained by calculating the Euclidian distances using the community detection method *K*-nearest neighbour (KNN) approach. The cells were then plotted in reduced PC-expression space based on their Euclidian distance values. Cell clusters were generated by detecting dense regions of cells in the reduced PC-expression space and calculating the *k*-means centroid of these regions. Cells were assigned into clusters based on their position to the nearest cluster centroid.

The initial clustering analysis revealed 19 distinct cell clusters (see Figure 4.7). In order to infer on the cell identities of these clusters, differential expression (DE) analysis was performed by comparing the overall transcriptional profiles of each cluster against all other clusters. This produced a list of DE gene signatures i.e., marker genes for each cluster which were used to characterise the clusters. The expression of the most meaningful gene markers was examined visually using UMAPs (see Figure 4.8 and Figure 4.9). Expression of the broad progenitor markers *Sox2* and *Nestin* was most abundant in cell clusters 0, 1, 2, 5, 6, 8, 11, and 14 (see Figure 4.8A). To determine whether these cells were RGCs, the expression of *Pax6* was visualised. Interestingly, expression of *Pax6* was mostly limited to clusters 0, 1, 2, and 8, with noticeably reduced expression in 5, 6, 11, and 14. To determine the identity of the other progenitor cell clusters, expression of *Eomes,* and *Rspo3* was visualised, which are markers for IPs and cortical hem progenitors, respectively (see Figure 4.8B) (Englund et al., 2005). Moreover, the expression of *Otx2* was also examined, a marker for progenitors from the medial ganglionic eminence (MGE) and medial/hippocampal progenitors (Hoch et al., 2015). Expression of *Eomes* was abundant in clusters 8, 10, and some of cluster 4. Moreover, expression of *Rspo3* and *Otx2* was observed in clusters 11 and 12, and 5, 6 and 16, respectively.

To determine the identity of the other cell clusters, the broad neuronal marker *Tubb3* was examined (see Figure 4.8C). Expression of *Tubb3* was most abundant in clusters 4, 8, 10, 12, and 14. To determine whether these were excitatory neurons, the expression of *Tbr1* and *Slc17ac6* was analysed. It was shown that expression of both genes was mainly localised to cluster 4, with some expression in clusters 10 and 12. To infer more on the identity of these cells, expression of *Gad2, Reln,* and *Dcx* was also examined, which are markers for interneurons, Cajal-Retzius cells, and immature neurons, respectively (see Figure 4.8D). Interestingly, expression of

Gad2 was mainly localised to clusters 9 and 14, whereas *Reln* was most abundant in cluster 12 with some expression in cluster 4. *Dcx* expression was also observed in most neuronal clusters.

Cell cluster analysis also revealed the presence of distinct cell clusters which were noticeably separated from the main cluster matrix, such as clusters 3, 7, 13, 15, 17, and 18. The separation of these clusters indicated that the transcriptional profiles of these cells are vastly different from the main matrix, suggesting these were non-neuronal cells. To identify these cells, expression of several markers was examined, including *Col1a1, Cdh5, Sfn, Hba-a1, Tyrobp,* and *Rgs5*, which are markers for meninges, endothelial cells, epidermal cells, blood cells, platelets, and vascular cells, respectively (see Figure 4.9). Expression of *Col1a1* was localised to clusters 3 and 15. Moreover, expression of *Cdh5, Sfn, Hba-a1, Tyrobp*, and *Rgs5*, was also exclusively localised to clusters 13, 18, 7, 17, and 15 respectively.

After analysing the gene signatures of each cluster using marker genes, the clusters were re-annotated with an appropriate biological label (see Figure 4.10A). It was found that most of the scRNA-seq dataset were progenitor cells, with also a large plurality of diverse neuronal cells (see Figure 4.10B). The most abundant non-neuronal cell type in the dataset were meningeal cells, which accounted for ~8% of the total cells. As aforementioned, the initial QC covariate analysis between number of genes and count depth revealed the presence of two covariate peaks, wherein a subset of cells exhibited relatively low gene features but high count depth (see Figure 4.3A). To infer on the identity of these cells, the correlational analysis was re-plotted to include cell identity (see Figure 4.10C). Interestingly, all cells within the smaller covariate peak were blood cells.

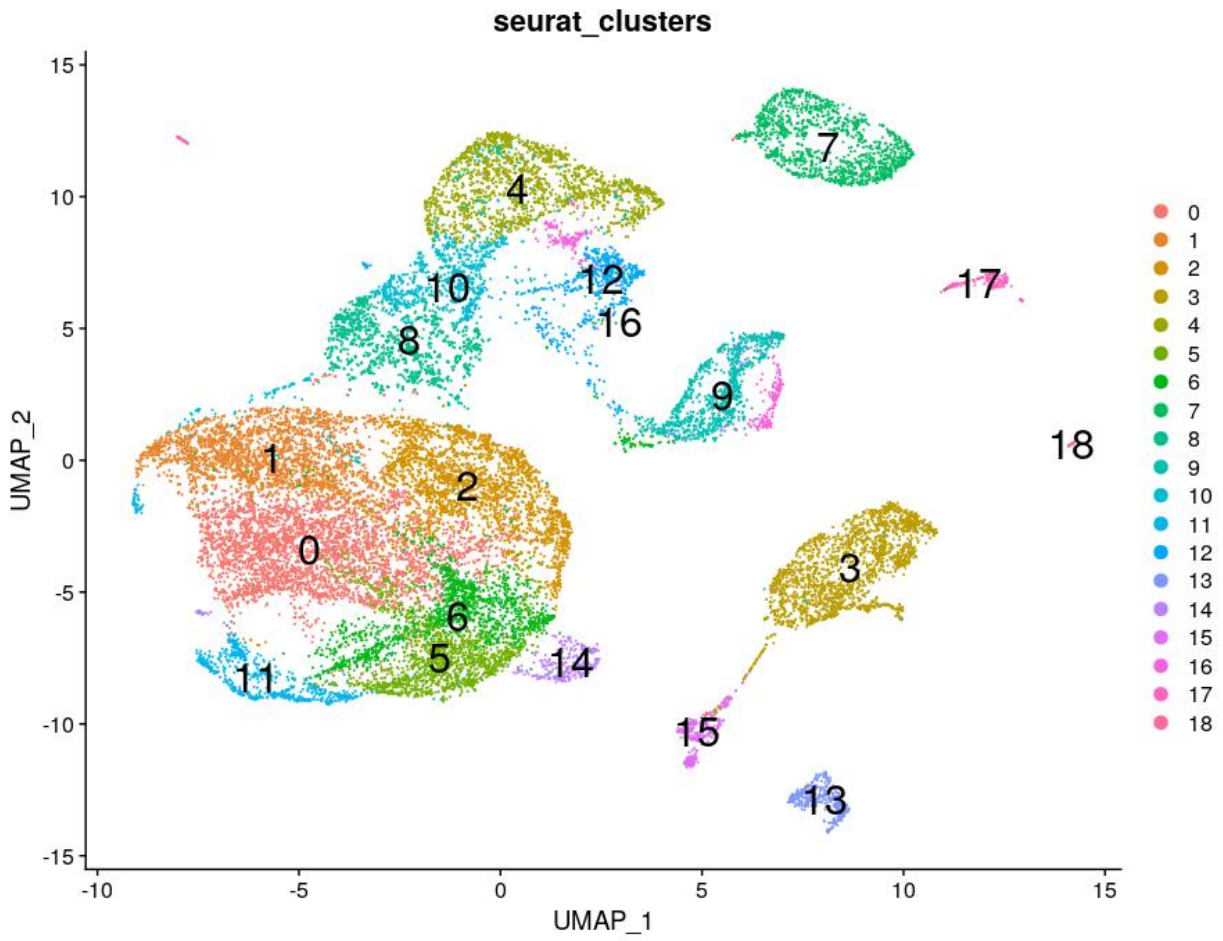


Figure 4.7. UMAP showing results from cell cluster analysis. 19 distinct cell clusters identified in the scRNA-seq dataset. Clusters are generated using the KNN community distance detection method, which groups cells into clusters based on the similarity of their expression profiles and the proximity of the cells to the *k*-means centroid. Colour code indicates cell cluster identity.

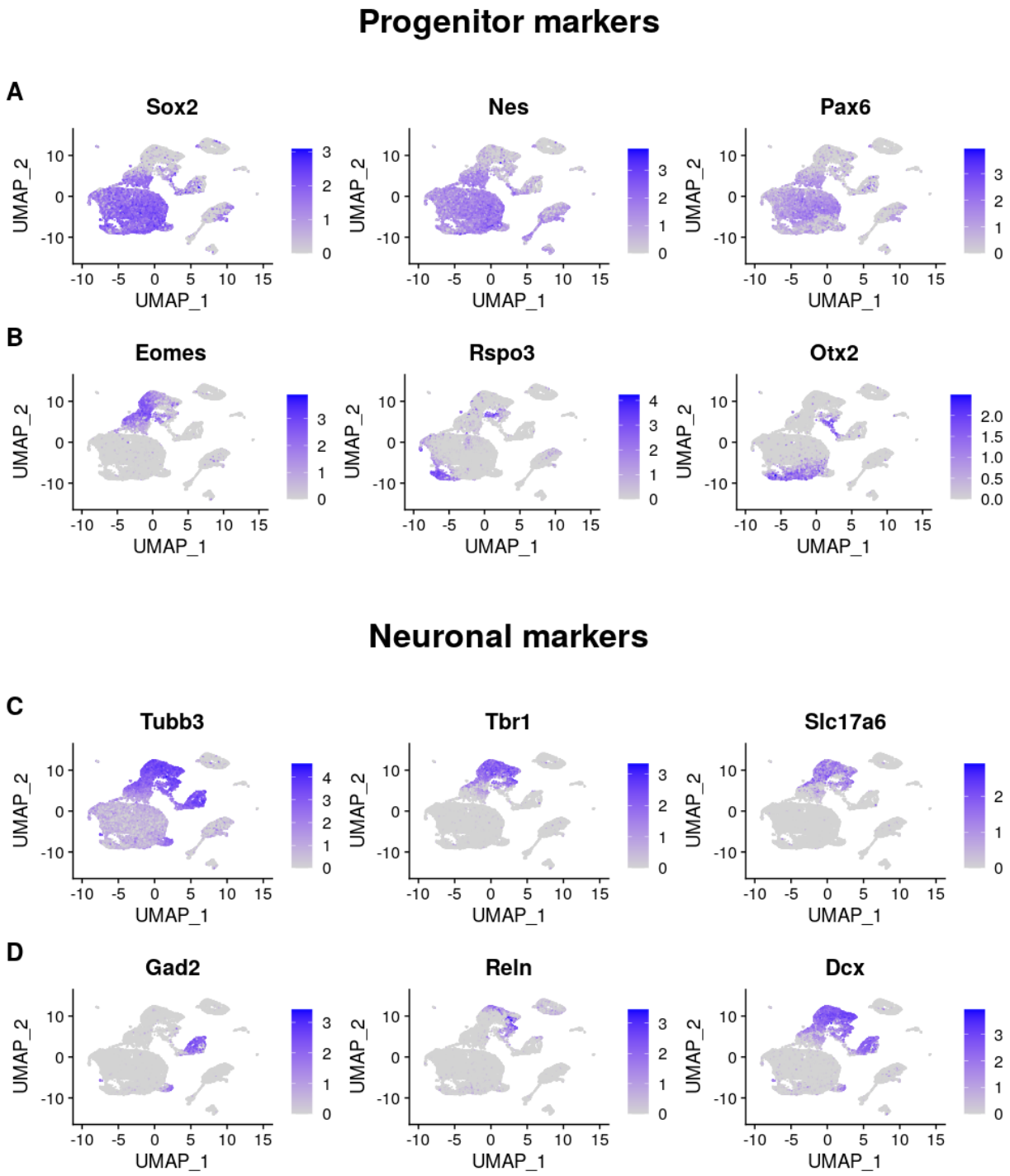


Figure 4.8. Expression analysis using progenitor and neuronal markers. (A) Clusters showing expression of *Sox2, Nestin,* and *Pax6.* (B) Clusters showing expression of *Eomes, Rspo3,* and *Oxt2.* (C) Clusters showing expression of *Tubb3, Tbr1*, and *Slc17a6.* (D) Clusters showing expression of *Gad2, Reln,* and *Dcx.*

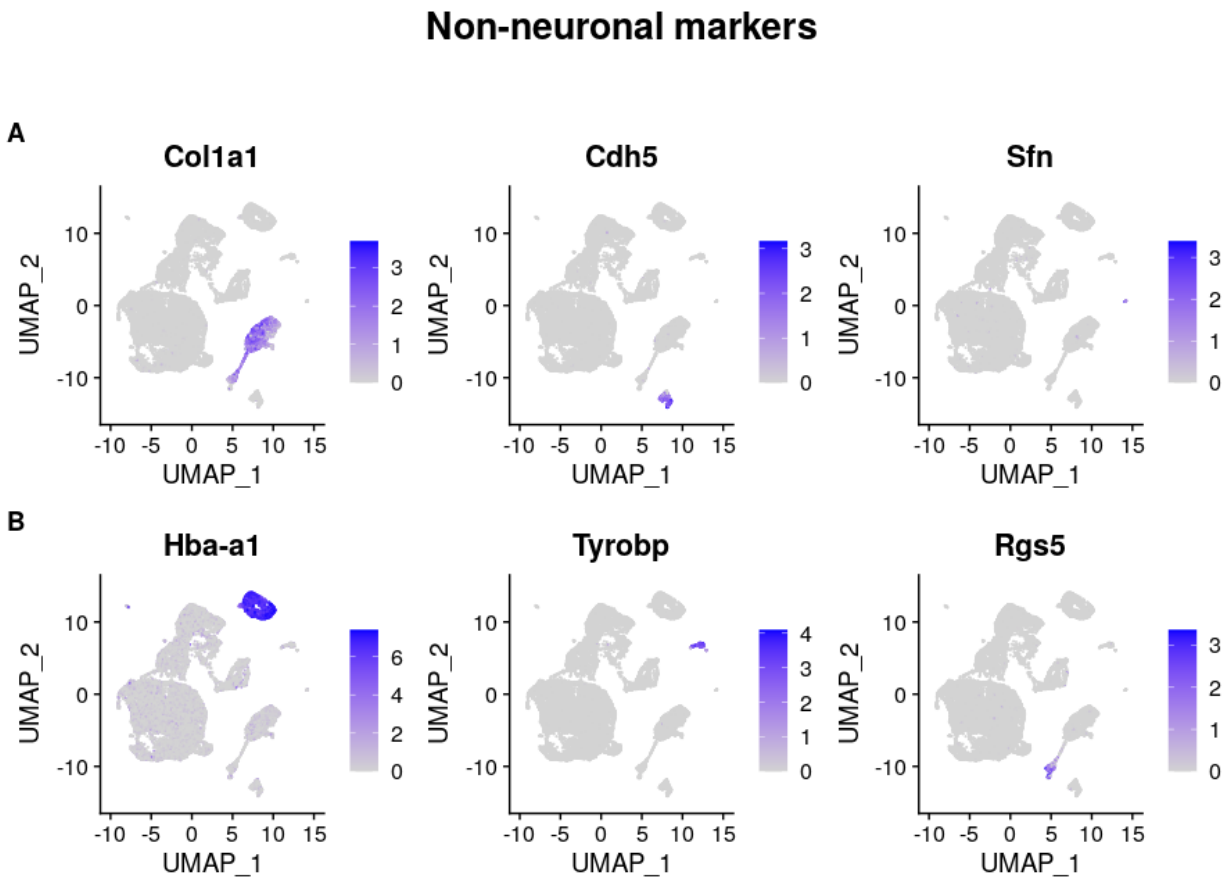


Figure 4.9. Expression analysis using non-neuronal markers. (A) Clusters showing expression of *Col1a1, Cdh5,* and *Sfn.* (B) Clusters showing expression of *Hba-a1, Tyrobp,* and *Rgs5.*

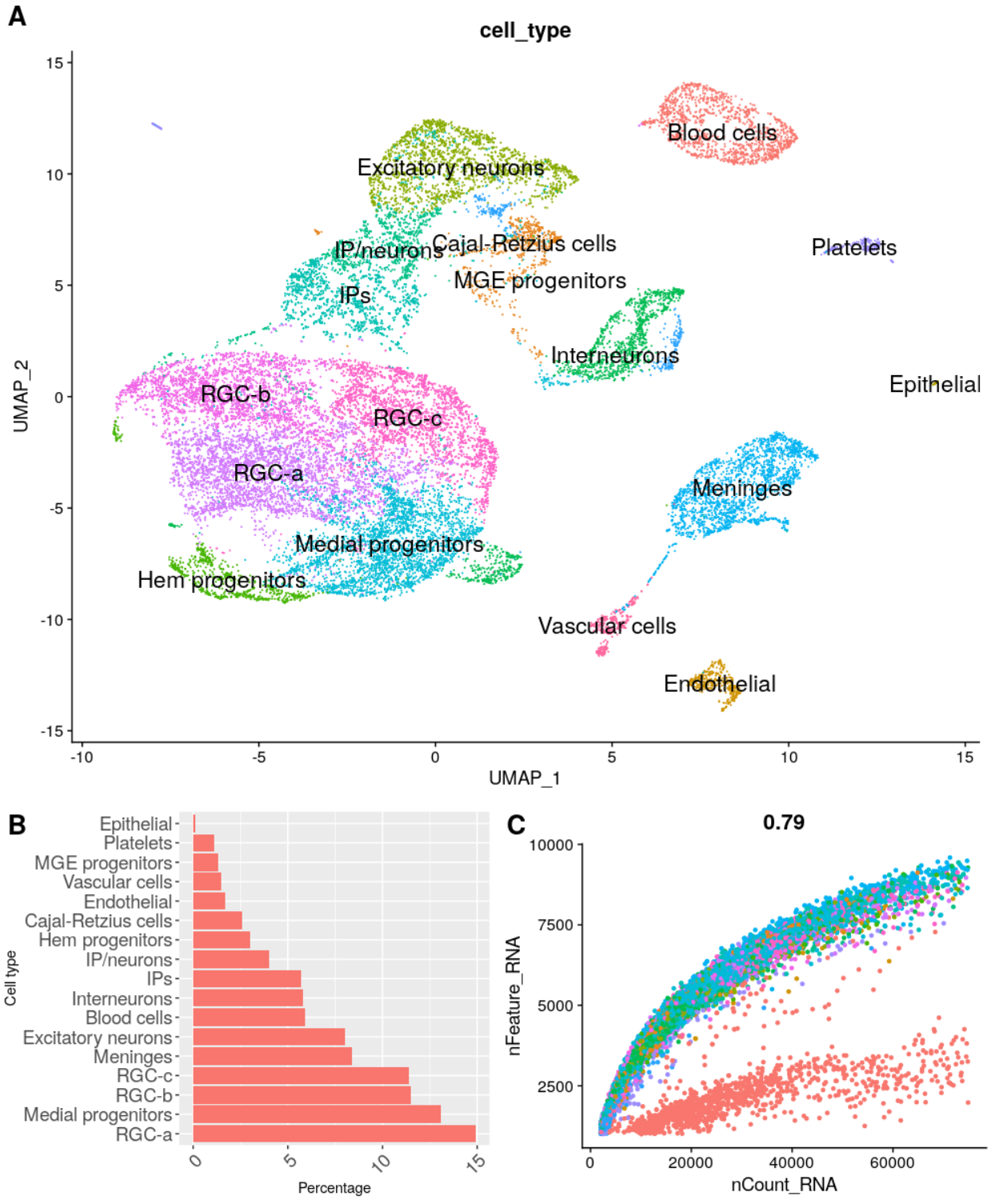


Figure 4.10. Cluster analysis results showing cell cluster identities. (A) UMAP displaying the cell cluster identities. (B) Bar plot illustrating the proportion of each cell type in the dataset. (C) Correlational analysis between number of genes and count depth, labelled with cell type. The smaller correlational peak is completely composed of blood cells.

4.2.5. Expression of Pcdh19, β-gal, and EGFP at the single cell level

To further examine the validity of the cell clusters, the expression of the key markers *Pcdh19*, *β-gal,* and *EGFP* were analysed (see Figure 4.11). Based on existing PCDH19 expression data, it was predicted that expression of *Pcdh19* would mainly be localised to RGC clusters (Pederick et al., 2018; unpublished data, 2019). Moreover, since *β-gal* expression is under the control of the endogenous *Pcdh19* promoter, it was predicted that *β-gal* would show the same expression pattern as *Pcdh19*. Finally, because *EGFP* was under the control of a CAG promoter, it was predicted that *EGFP* expression would be more dispersed across all the cell clusters. As expected, *Pcdh19* was most abundant in the RGC cell clusters (see Figure 4.11A). Interestingly, *Pcdh19* expression was also observed in IP, hem, and medial progenitor clusters, with minor expression in excitatory neuron, interneuron, and meningeal clusters. Moreover, it was found that *β-gal* expression was also most abundant in the RGC cell clusters, in addition to the IP, hem, and medial progenitor clusters. Expression was also observed in the excitatory neuron, interneuron, and meningeal clusters, similar to *Pcdh19*. Expression of *EGFP* was also observed in RGC, IP, medial progenitor, and neuronal clusters, however the expression level was significantly reduced compared to the other markers (see Figure 4.11B).

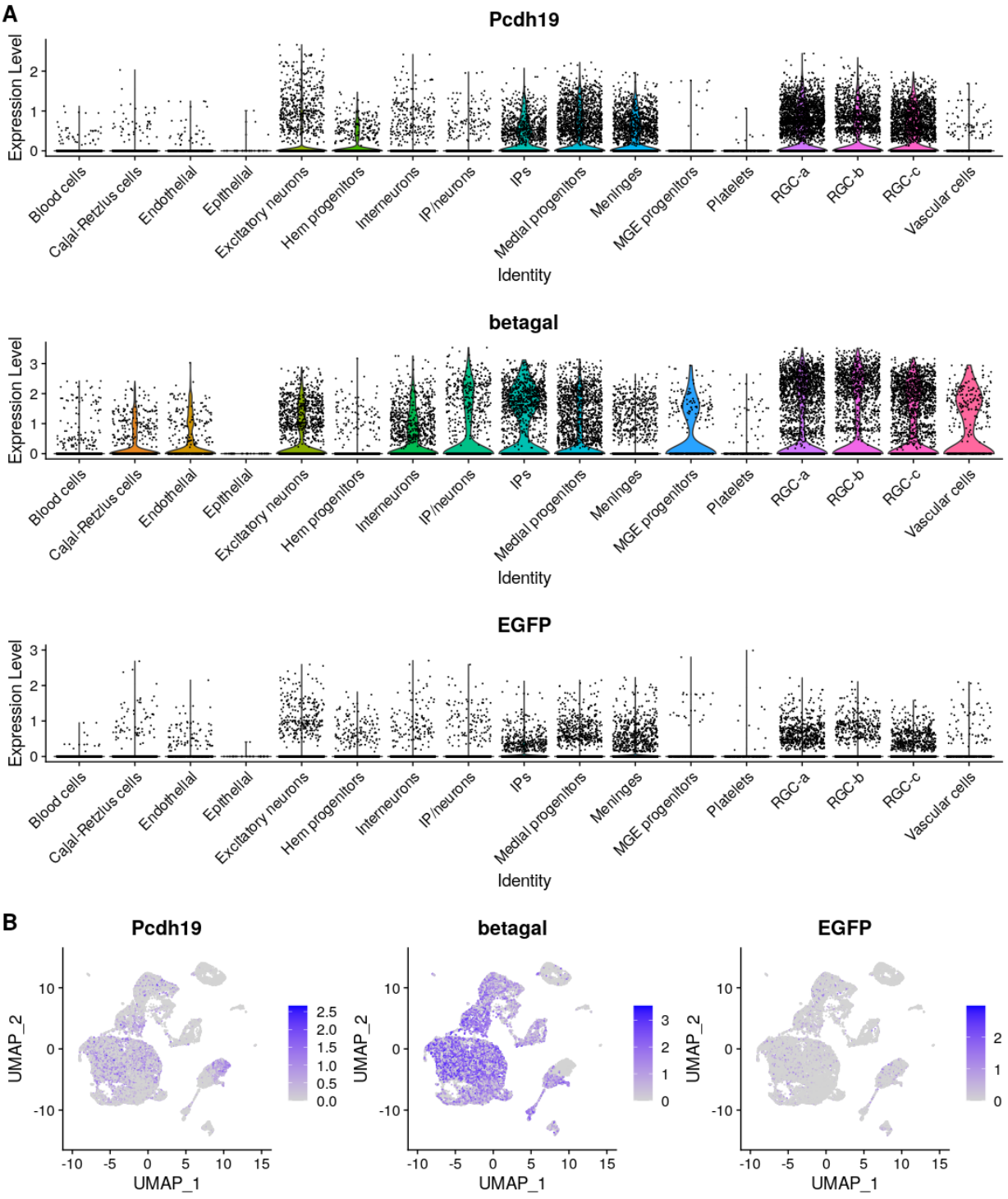


Figure 4.11. Expression of *Pcdh19*, *β-gal,* and *EGFP*, in the cell clusters. (A) Violin plots showing expression of *Pcdh19, β-gal* and *EGFP* in the cell clusters. Expression of *Pcdh19* and *β-gal* was mainly localised to progenitor cell clusters. *EGFP* expression was relatively low in most cell cluster groups. (B) Expression of the genes displayed using UMAPs.

4.2.6. Expression of genotype and sex markers

To confirm the validity of the genotypes, the expression of *Pcdh19, β-gal,* and *EGFP* was examined between WT, HET, and KO samples (see Figure 4.12A). As expected, expression of *Pcdh19* was most abundant in WT and HET samples. Interestingly, low levels of *Pcdh19* were also detected in the KO samples. *β-gal* was also only detected in HET and KO samples, whereas no expression was detected in WT samples. Likewise, *EGFP* was only detected in WT and HET samples, and no expression was observed in KO samples, as expected. As aforementioned, cells from WT male and female samples were mixed and ran on the same well in the Chromium Chip-B cassette (see Section 4.2.1) since it was theorised that the cells could be separated using the sex-specific markers identified in the previous chapter (see Section 3.2.7). To examine whether sex markers can be used to distinguish male and female cells, the expression of female markers (*Tsix* and *Xist*) and male markers (*Uty* and *Kdm5d*) were analysed across WT, HET, and KO groups (see Figure 4.12B and Figure 4.12C). As expected, the female markers were only observed within WT and HET samples, whereas the male markers was only observed within WT and KO samples. To determine whether WT male and WT female cells could be separated, the aforementioned sex markers were used to label the cells as female (cells > 0 *Xist* expression) or male (cells > 0 *Uty* expression) cells and the expression of *EGFP* was examined, which should only be present in WT female samples (see Figure 4.12D). As predicted, *EGFP* was mainly detected in female cells, with only background levels detected in male cells. Therefore, male and female cells can be accurately filtered using the previously identified sex markers.

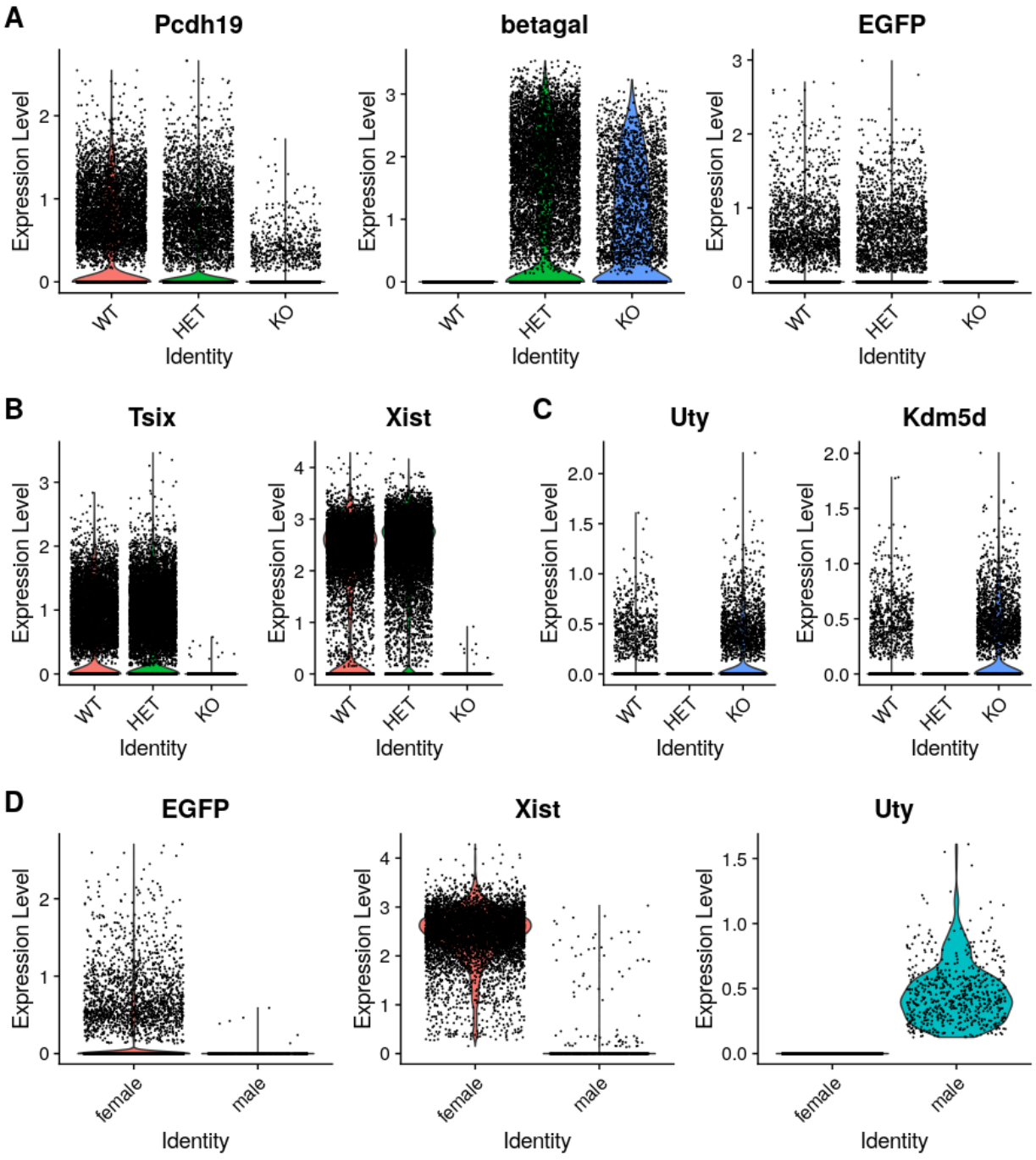


Figure 4.12. Expression of genotype and sex markers at the single cell level. (A) Violin plot showing expression of *Pcdh19, β-gal,* and *EGFP* within WT, HET, and KO samples. (B) Violin plot showing expression of female sex markers *Tsix* and *Xist* in WT, HET, and KO samples. (C) expression of male sex markers *Uty* an *Kdm5d* in WT, HET, and KO samples. (D) Violin plots showing expression of *EGFP*, *Xist*, and *Uty* in male and female cells that were filtered using the aforementioned sex markers.

4.2.7. Proportion of WT, HET, and KO cells in RGC, IP, and neuron clusters

The results from the cell cluster analysis revealed successful capture and sequencing of the key cell types for downstream analysis, namely RGCs, IPs, and neurons. Therefore, the total number and proportion of WT, HET, and KO cells in RGC, IP, and neuron cell clusters was also examined to confirm that each genotype had a satisfactory level of cells in each cluster for future downstream DE analysis (see Figure 4.13A). HET cells were the most abundant in the RGC, IP, and neuron clusters, capturing 5651, 1482, and 1080 cells, respectively. This was expected since the HET replicates were run twice to increase the cellular output for this group (Figure 4.1). Interestingly, KOs cells were the least abundant in all cell clusters, capturing 426, 189, and 225 RGCs, IPs, and neurons, respectively. WT cells were the second most abundant in each cell cluster, capturing 2738, 600, and 556 RGCs, IPs, and neurons, respectively. When examining the proportion of RGCs, IPs, and neurons captured from each genotype group, it was found that RGCs were the largest cell group captured, followed by IPs and neurons (see Figure 4.13B). Interestingly, although KOs showed the smallest number of captured cells, they showed the highest proportion of IPs and neurons.

To investigate whether WT, HET, and KO cells showed similar transcriptional profiles within the RGC, IP, and neuronal clusters, the cells were plotted using UMAPs and annotated by genotype (see Figure 4.13C and Figure 4.13D). It was shown that WT and KO cells co-clustered at the RGC, IP, and neuronal level – however HET RGCs were shown to segregate away almost completely from WT and KO RGCs. Interestingly, this segregation effect was not observed within the IP and neuron cluster, suggesting HET RGCs exhibited strikingly different transcriptional profiles compared to WT and KO cells which disappeared at the IP and neuron level.

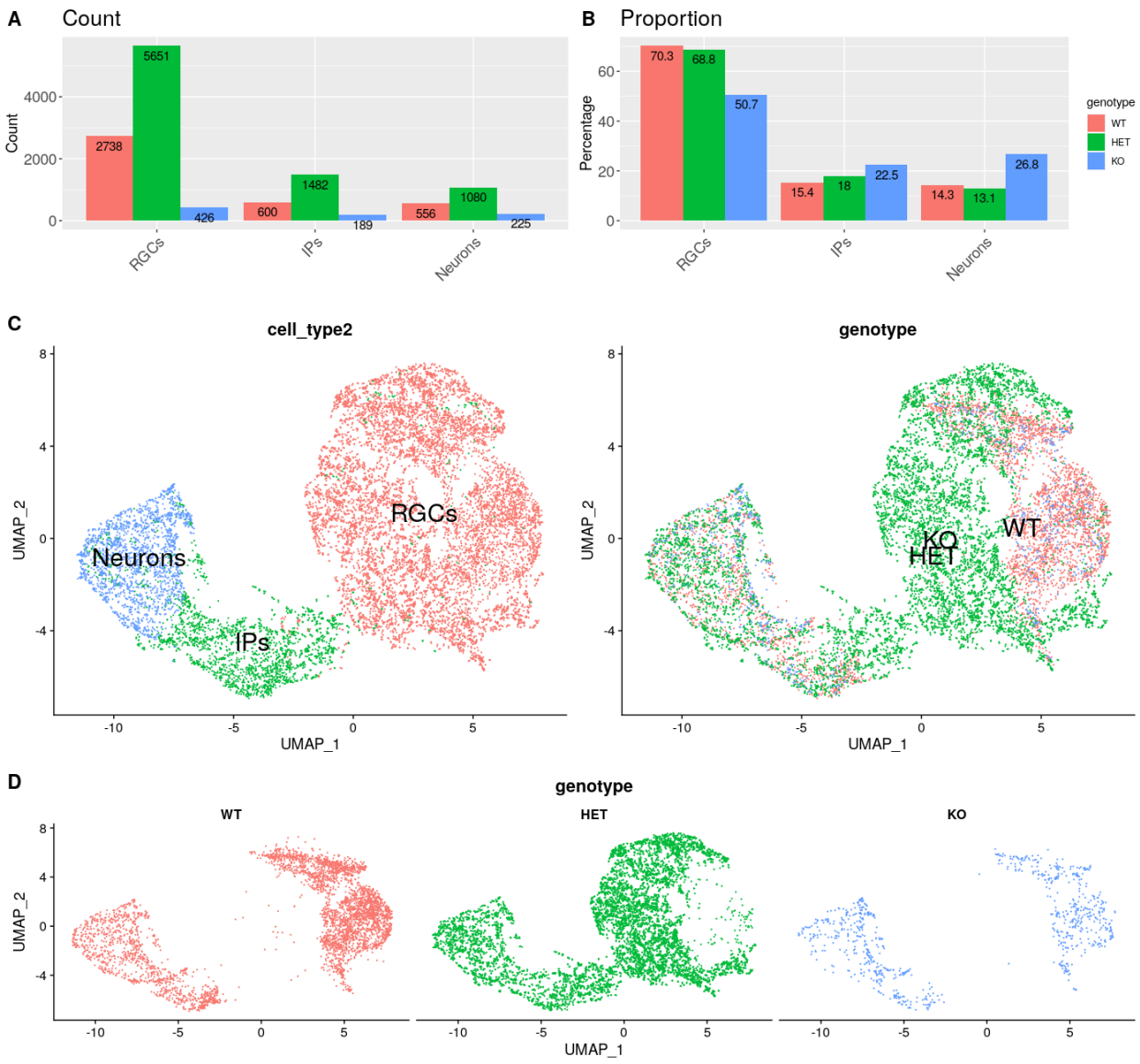


Figure 4.13. Proportion of WT, HET, and KO cells within the RGC, IP and neuron clusters. (A) Bar plot illustrating the total number of RGCs, IP, and neurons belonging to each genotype group. (B) Bar plot illustrating the proportion of RGCs, IPs, and neurons belonging to each genotype group. (C) UMAP showing the position of total RGCs, IPs, and neurons labelled as cell type (left) and genotype (right). HET cells were mostly segregated from WT and KO cells, but only in the RGC cluster. (D) UMAP showing the position of WT, HET, and KO cells separately within the RGCs, IPs, and neuronal clusters.

4.3. Discussion

In this chapter, the first analyses were performed on the scRNA-seq dataset. Recent publications have attempted to identify transcriptional alterations that potentially underlie the mechanism in which PCDH19 can regulate neurogenesis (Homan et al., 2018). In addition, the previous chapter attempted to examine these differences further by analysing the transcriptional profile of WT, HET, and KO E11 tissue using bulk RNA-seq. The bulk RNA-seq experiment demonstrated that total loss of *Pcdh19* does not cause major transcriptional alterations in KO samples *in vivo*. Furthermore, no transcriptional differences were also observed in the HETs, however the underlying profiles of the mosaic WT and KO cells within the HET could not be separated and examined using bulk RNA-seq methods. Moreover, the RNA-seq experiment was performed on whole cortical tissue, which meant that theoretically most reads would originate from RGCs, however reads from IPs, neurons, and non-neuronal cells would also be present and could not be removed. Therefore, the averaging that occurs by pooling reads from many different cell types, in addition to pooling reads from WT and KO cells within HET samples, did not provide enough resolution to assess the transcriptional layout of perturbed expression of *Pcdh19* and its effect on cortical neurogenesis. To further elucidate the potential transcriptional alterations that occur *in vivo* from perturbed *Pcdh19* expression, scRNA-seq was performed. scRNA-seq offers a key advantage compared to bulk RNA-seq, namely it permits the comparison of transcriptomes from individual cells (Haque et al., 2017). Therefore, scRNA-seq enables the ability to sort cells of interest from a wider population of cells and perform transcriptional analysis at a resolution not possible with conventional bulk RNA-seq. However, before downstream analysis could be performed on WT, HET, and KO samples, the scRNA-seq dataset had to first be subjected to quality control (Luecken and Theis, 2019). In this chapter, the scRNA-seq dataset was explored using common QC metrics and the identities of the captured cells were also verified via cell cluster analysis.

4.3.1. QC metrics revealed the majority of captured cells were suitable for analysis

scRNA-seq datasets can contain many technical artefacts, for example cell bursting causes RNA leakage, inadequate cell partitioning can lead to the capture of multiple cells (doublets or multiplets), and dying cells exhibit RNA degradation (Nayak and Hasija, 2021). Assessing the overall quality of data within the scRNA-seq dataset can be performed through multiple avenues, however a common approach is to examine the relationship between QC covariates (Ilicic et al., 2016; Stegle et al., 2015). For example, examining the number of genes per cell and the number of RNA molecules per cell (count depth) can reveal a subset of cells exhibiting low scores for both

covariates. Cells displaying low genes and RNA are usually indicative of damaged cells and are not suitable for downstream analysis (Brennecke et al., 2013). Comparing these covariates within the scRNA-seq dataset revealed that most captured cells exhibited decent RNA and gene numbers (see section 4.2.2). Moreover, the doublet rate was also relatively low, overall suggesting that the cell dissociation and partitioning techniques worked well. However, a significant plurality of cells did display low RNA counts and gene numbers, indicating that many cells in the dataset were poor quality compared to typical QC thresholds. Another way of assessing cell quality is analysing the fraction of mitochondrial genes per cell; a greater than 10% mitochondrial gene content is an indicator the cell membrane was compromised, and cytoplasmic RNA has leaked (Nayak and Hasija, 2021). Interestingly, the rate of cells exhibiting high mitochondrial gene fraction was very low, therefore the aforementioned cells that displayed low RNA counts did not automatically display high mitochondrial gene expression. This result demonstrates that considering the QC covariates in isolation can lead to misinterpretation of the data – for example it is possible that the cells with low genes/RNA are not necessarily damaged but possibly quiescent (Luecken and Theis, 2019). This also demonstrates that the covariables used for QC are an *indicator* of cell quality, rather than specific measurements of cell quality. One way of directly measuring cell quality is to use RNA spike-ins, which are synthetically-generated short RNA polymers added to the samples before amplification (Brennecke et al., 2013). The RNA spike-ins are not amplified during library preparation, rather they are sequenced and measured against the amount of endogenous amplified RNA. Cells that exhibit a high ratio between spike-ins and endogenous RNA are considered poor quality and removed.

Taken together, ~42% of cells were removed from the scRNA-seq dataset – including most of the cells with low count depth – to ensure that the subsequent remaining cells were of good quality for downstream analysis. If RNA spike-ins had been used for this experiment it is possible that more of the cells could have been used for analysis, however this is something that should be considered for future experiments.

4.3.2. Cell cluster and identity analysis showed that the highest proportion of captured cells were progenitors

The first intermediate result of scRNA-seq analysis is organising the cells into cluster groups based on overall similarity of their expression profiles (Luecken and Theis, 2019). Cluster annotation is then followed, wherein the cell clusters are subjected to DE analysis to identify the gene signatures that distinguish the clusters, which are used to annotate the clusters with a

meaningful biological label. As mentioned, the most abundant cell type in the developing cortex at E11.5 are RGCs, with a small proportion of IPs as well as preplate (PP) and some newborn excitatory neurons, and interneurons (Götz and Huttner, 2005). Therefore, it was expected that the most of captured cells within the filtered scRNA-seq dataset would be RGCs with a noticeable presence of IP and neurons (see section 4.2.4). As predicted, many of the captured cells expressed key progenitor markers such as *Sox2* and *Nestin* and most of these cells also expressed the RGC marker *Pax6* (Englund et al., 2005). However not all progenitor cells expressed *Pax6* – indeed, further marker analysis revealed the presence of other types of progenitor cells, including a fraction of cells from the cortical hem, MGE, and primordial hippocampus/ medial cortical region. Likewise, a sizable fraction of IP and excitatory neuronal clusters were also identified, in addition to numerous non-neuronal clusters, including meninges, blood cells, endothelial cells, and others. Nevertheless, proportional analysis revealed that the most abundant cell type in the dataset were the RGC clusters in addition to the medial progenitor population. This highlights that despite employing careful dissection and dissociation techniques of the lateral cortex at E11.5, it is very difficult to completely dispense with nonrelevant cell types - not only to avoid other areas of the brain (such as contaminants from the primordial hippocampus and MGE) but also non-neuronal cell types. Nevertheless, this result does demonstrate that dissection of the E11.5 cortex for RNA-seq does provide an enrichment of RGCs for downstream analysis. This further highlights the advantage of scRNA-seq over bulk RNA-seq, wherein the latter offered no ability to remove reads from nonrelevant cell contaminants from the analysis.

4.3.3. Expression of Pcdh19 was observed in RGCs, as well as IPs, and neurons

Performing scRNA-seq at E11.5 also offered the opportunity to explore the expression levels of *Pcdh19* in cells of interest, namely RGCs, IPs, and neurons (see section 4.2.5). As previously discussed, *Pcdh19* is expressed in the mouse cortex as early as E9 (Gaitan and Bouchard, 2006; Pederick et al., 2018) but because antibodies for PCDH19 have shown to be nonspecific or non-functioning in immunostaining experiments, a detailed expression analysis of PCDH19 has not been published. Nevertheless, an analysis involving *in situ* hybridisation (ISH) for *Pcdh19* combined with immunohistochemistry (IHC) for progenitor markers has been performed in the IMG lab and revealed that expression of *Pcdh19* is greatest at E11 and E12 and diminishes by E13. Moreover, expression is most abundant in RGCs of the lateral cortex which recedes in a lateral-medial pattern across development (unpublished data, 2019). Interestingly, *Pcdh19* colocalised stronger with $PAX6^+$ RGCs, in addition to partial localisation with $Calretinin^+$ interneurons/CZ cells, but reduced *Pcdh19* signal was detected in $TBR2^+$ IPs (unpublished data,

2019). Similarly, scRNA-seq expression analysis revealed that *Pcdh19* was most widely expressed in the RGC clusters, however contrary to the previous analysis, noticeable expression was observed in IPs and minor expression in neurons. One possible explanation to this finding is that *Pcdh19* is expressed in IPs however the mRNA levels are more tightly regulated. As previously discussed, the expression window of *Pcdh19* in RGCs of the lateral cortex is short, suggesting that the expression of *Pcdh19* is under strict control by selective regulatory mechanisms. Likewise, these mechanisms may also control the expression of *Pcdh19* in IPs albeit in a more severe manner than in RGCs. Another explanation could be that the mRNA of *Pcdh19* was only detected in newly-generated IPs which retained a small level of the mRNA from their mother cell. Indeed, when examining the IPs that express *Pcdh19* in two-dimensional space, IPs that expressed *Pcdh19* clustered closer to the RGC clusters (see Figure 4.11). Likewise, IPs that clustered closer to neurons showed noticeably less *Pcdh19* expression. One way to examine this further would be to assess the protein levels of PCDH19 in RGCs and IPs i.e., determine whether *Pcdh19* mRNA is translated in both cell types and to confirm whether protein levels of PCDH19 are different between both cell types. However, the protein pattern of PCDH19 in the cortex cannot be addressed until a specific PCDH19 antibody is generated and has become commercially available.

An interesting result was also found when comparing the expression of *Pcdh19* between WT, HET, and KO groups. As expected, *Pcdh19* was mostly expressed within the WT and HET groups, however a small degree of expression was also observed in the KOs. Low levels of *Pcdh19* mRNA have also been detected in a previous RNA-seq experiment performed in the IMG lab (unpublished data, 2021). It has also been demonstrated that the remaining exons in the KO model (exons 4-6) are likely being transiently expressed (unpublished data, 2022). Moreover, recent work at the IMG lab has demonstrated that the truncated mRNA product is also possibly translated, however a full analysis concerning the stability of the protein and its function in cells has not been assessed. Therefore, further work is required to assess the validity of the KO model.

4.3.4. Sex markers could accurately separate male and female samples from the WT cluster

As mentioned before, one of the aims of the scRNA-seq experiment was to perform DE analysis between WT and KO cells within the HET samples to determine the transcriptional underpinnings behind the unusual neurogenesis phenotype displayed in these animals. Therefore, to maximise the number of cells for that analysis, it was decided to use two wells each for both HET biological

replicates to increase the cell output for that group. Because the 10X chromium controller only allows the partition of 8 single cell dissociation samples, it was decided to merge WT male and female samples and run them in the same well under the assumption that these cells could be separated for the downstream analysis using the highly expressed sex-specific markers identified in the previous chapter (see section 3.2.6.1). To explore this, the expression of the markers was analysed. Expression of *Tsix* and *Xist* was shown to be female-specific, as demonstrated by high expression of both markers from the WT and HET groups, but not from the KOs. Likewise, expression of *Uty* and *Kdm5d* was also shown to be male-specific, as demonstrated by expression in WT and KO groups, but not HETs. These markers were used to label WT cells with female and male annotations, respectively, and the accuracy of these labels was tested by analysing the expression of *EGFP* between both groups, which should only be expressed in WT females. As expected, *EGFP* expression was almost completely limited to the female cells, demonstrating that using the previously identified sex markers can be used to sort male and female samples at the single cell level.

4.3.5. HET cells showed distinct separation from WT and KO cells, but only at the RGC level and in some IPs

After cell filtering, cluster analysis, and verifying the cell identities, the next step was to determine whether enough cells from each genotype group were captured for downstream analysis (see section 4.3.5). As expected, HET cells were the most numerous in the dataset, followed by WTs and KOs. Interestingly, KOs showed the lowest cell count in the RGC, IP and neuronal clusters, which was likely a technical artefact since the amount of KO cells harvested for cell partitioning was lower than HETs and WTs in both batches. Nevertheless, several hundred KO RGCs were collected for analysis, which is suitable for downstream analysis (Ding et al., 2020). After examining the distribution of WT, HET, and KO cells in the RGCs, IPs, and neuron clusters, a striking cell separation pattern was found, wherein the HET cells segregated away from the WT and KO cells. Surprisingly, this effect was mostly observed in the RGC cluster, whereas some IP cells also clustered away from WT and KO IPs. Interestingly, WT, HET, and KO neurons appeared to co-cluster indiscriminately. This suggests that the transcriptional profiles of the HET RGCs are significantly different to the WT and KO RGCs, but this difference disappears at the IP and neuronal level. One possible explanation is that mosaic expression of *Pcdh19* alters the transcriptional landscape of these cells. Because *Pcdh19* expression diminishes as RGCs commit to IPs and neurons, these transcriptional changes disappear over time. This may also explain why some HET IPs also separate from WT and KO IPs, since it is possible that these are new-born

IPs that still exhibit expression of *Pcdh19*. However, as the cells commit closer to neurogenesis, *Pcdh19* expression reduced, and the transcriptional profiles of the cells become more similar to WTs and KOs. It is also possible that because it was decided to load HET samples twice to increase the cell output for this group, the small proportion of HET IP cells that are separated from the WTs and KOs could be an earlier lineage of IPs that were only captured from the HETs and not from the WTs and KOs. Identifying the transcriptional changes in RGCs, IPs and neurons was the focus of the next chapter.

Chapter 5: Single cell RNA-seq analysis of Protocadherin-19 WT, HET, and KO embryos Part II: Differential expression analysis

5.1. Introduction

One of the main goals of this project was to identify transcriptional differences between *Pcdh19* wild type (WT), heterozygous (HET), and knockout (KO) E11.5 embryos, to decipher the molecular mechanisms underlying the role of *Pcdh19* in cortical neurogenesis. Likewise, another goal was to identify differences between WT and KO cells within HETs to provide insight into the unusual neurogenesis phenotype that occurs in these animals. To this end, single cell (sc)RNA-seq was employed to examine the differences of gene expression of WT, HET, and KO cells and WT and KO cells within the HET at the single cell level.

As aforementioned, the main advantage of scRNA-seq over bulk RNA-seq is that cell heterogeneity can be accounted for by performing differential expression (DE) analysis within cell identity clusters (Kang et al., 2017). Therefore, unlike with bulk RNA-seq, expression data can be examined specifically from cells of interest without confusing the data by averaging reads from other cell types or other confounding factors. Nevertheless, although scRNA-seq provides superior cellular resolution, it also exhibits unique challenges such as a high degree of technical artefact, high cell-to-cell variability, and lower read depth (Hicks et al., 2018). In chapter 4, technical and biological variability was rigorously explored. Using metrics designed for scRNA-seq, the dataset was filtered to include only data-viable cells and the dimensionality was reduced to remove as much technical noise as possible (Brennecke et al., 2013; Ilicic et al., 2016; Stegle et al., 2015). Moreover, most DE methods for scRNA-seq also take these artefacts into account, although reducing their influence on the data is an important step to maximise the signal (Kharchenko et al., 2014).

Another important step is deciding which method to use for DE analysis. The Seurat package that is widely used for scRNA-seq analysis has an inbuilt DE analysis method which employs the Wilcoxon ranked sum test to distinguish DE genes between two identity classes (Hao et al., 2021). Although useful, the Seurat method can often produce inflated p-values since it treats every cell as a sample. Moreover, since single cells from a sample are not independent of each other (i.e., they come from the same sample), it is difficult to separate the variation of the whole cell group from the variation within the cell group, i.e., the variation of the population from the variation of the individual sample. Indeed, methods that ignore within sample variation are prone to false

discoveries (Squair et al., 2021). One way around this is to adapt the dataset to employ bulk RNA-seq methods of DE analysis, called a “pseudo-bulk” analysis (Luecken and Theis, 2019). To do this, expression data from cells of interest are gathered and the counts/metadata are aggregated to the samples by averaging the reads. That way, the expression dataset resembles a dataset compatible with bulk RNA-seq analysis however the count data has been selected from cells of interest. Indeed, pseudo-bulk analysis has been shown to reduce statistical sensitivity and bias (Soneson and Robinson, 2018). Nevertheless, averaging the reads across cell populations can completely diminish much of the information regarding cell heterogeneity, which can also be necessary for identifying DE genes between conditions. Therefore, to determine the appropriate DE method to use in this chapter, both Seurat and pseudo-bulk methods were employed to examine DE genes between WT, HET, and KO samples.

It is also important to decide which cell identities to focus on for DE analysis. Because scRNA-seq provides rich expression information on a large collection of cells, DE analyses can become complex. Therefore, deciding which cells to analyse between conditions *a priori* is important for focusing hypotheses. In mice at E11, the most abundant cell type of the developing lateral cortex are the radial glial cells (RGCs) of the ventricular zone (VZ). These cells initially undergo symmetric proliferative divisions but around E10-E12 in mice the cells undergo a neurogenic switch and begin producing intermediate progenitors (IPs) and neurons (Götz and Huttner, 2005; Noctor et al., 2001). The IPs migrate basally and form the subventricular zone (SVZ) whereas new-born excitatory neurons migrate along the radial fibres of the RGCs and form the cortical plate (CP) (Noctor et al., 2004). Histological analysis at E11.5 and E12.5 demonstrated that *Pcdh19* expression is mainly limited to RGCs of the VZ (unpublished data, 2019) although in chapter 4 residual expression of *Pcdh19* was also observed in what were likely new-born IPs and to a lesser extent, neurons (see section 4.2.5). Moreover, further histological analysis demonstrated that mosaic expression of *Pcdh19* caused altered neurogenic behaviours in WT and KO RGCs within the HETs which was not observed in homozygous WTs and KOs. Strikingly, when examining the transcriptional similarities of RGCs, IPs, and neurons from WT, HET, and KO samples, HET cells showed a clear separation from WT and KO cells. Likewise, WT and KO cells co-clustered regardless of genotype. Interestingly, the separation of the HET cells was only observed from the RGCs and some IPs, whereas WT, HET, and KO neurons were shown to co-cluster. This result suggested that mosaic expression of *Pcdh19* causes transcriptional alterations in RGCs, but these differences are diminished when progenitors commit to neurogenesis and *Pcdh19* expression is reduced.

Taken together, it would be useful to examine the transcriptional landscape of RGCs, IPs, and neurons between WT, HET, and KO groups. The RGC population would be the primary cell type of interest to focus on since these cells widely express *Pcdh19* and mosaic expression of the gene causes the cells to display altered neurogenic and transcriptional activity that must be investigated (unpublished data, 2019; see section 4.2.7). Moreover, although IPs show reduced expression of *Pcdh19* the aforementioned results demonstrate that new-born IPs likely exhibit residual *Pcdh19* expression. Therefore, it would be worth exploring the transcriptional profiles of the IPs to determine whether DE genes also exist between the genotype groups. Finally, it was decided to also examine the transcriptional landscape of neurons between the genotype groups. It was theorised that since WT, HET, and KO neurons co-cluster, the transcriptional differences between these groups would be minimal compared to the potential differences found within RGC and IP clusters. If proven correct, then it would support the hypothesis that the transcriptional differences between the genotype groups would be mostly found within the RGCs.

5.1.1. Aim

The primary aim of this chapter was to identify the transcriptional changes between WT, HET, and KO groups by performing DE analysis at the single cell level. This would be achieved by subsetting cells of interest, namely RGCs in addition to IPs and neurons, and performing DE analysis within these clusters. This was to examine the transcriptional differences that drive the separation effect observed from the HETs and whether these differences remain, diminish, or change, in IPs and neurons. Moreover, DE analysis was to be initially performed using both the Seurat and pseudo-bulk methods which would provide expression information at the individual cell and population levels, respectively. Additionally, another key aim of this chapter was to perform DE analysis between WT and KO cells within the HETs to identify DE genes that may underlie the unusual neurogenesis phenotype of these animals. Overall, the goal of this chapter was to identify DE genes which may highlight molecular avenues that underscore the role of *Pcdh19* in neurogenesis.

5.2. Results

5.2.1. Single cell DE analysis: WT vs. HET vs. KO

As a first approach to determine the most differentially expressed genes between WT, HET, and KO groups, DE analysis was performed by comparing each group within the RGC, IP, and neuron clusters using the Seurat DE analysis technique. To perform DE analysis between the genotype groups within each cell population, each cell type was individually subset from the main cellular matrix. Moreover, the RGC, IP, and neuron matrices were filtered further to remove any cell that did not express *Pax6*, *Eomes,* or *Tbr1*, respectively. Each dataset was then subjected to re-normalisation, scaling, and cluster analysis before DE was performed. For the analysis, the threshold for significance was set to p adjusted value > 0.05 and a log fold change (LFC) of 0.25. These threshold parameters were chosen based on following the standard parameters in the Seurat FindMarkers function. The list of DE genes for all comparisons in this chapter can be found in the appendices at the back of this thesis.

DE analysis of WT, HET, and KO groups within the RGC cluster revealed numerous DE genes (73 genes; 48 upregulated and 25 downregulated) (see Figure 5.1). Surprisingly, analysis within the IP cluster revealed many more DE genes (184 DE genes; 110 upregulated, 74 downregulated) and analysis within the neuron cluster revealed 85 DE genes (49 upregulated, 36 downregulated) (see Figure 5.1). It was noticed that many of the top DE genes between the genotype groups were related to ribosomes and histones. Surprisingly, most of these genes were shown to be downregulated in the HETs, whereas levels were typically equal between WTs and KOs (see Figure 5.2). To investigate this further, the overall levels of ribosome mRNA were examined between each genotype group. Interestingly, HETs were shown to possess noticeably reduced fraction of ribosomal genes (see Figure 5.1 and Figure 5.2). Although this could mean that in further DE analysis on HET samples ribosomal genes would almost always be the most differentially expressed. It was calculated that ~37% of DE genes were explicitly related to ribosomes and histones, meaning ~63% DE genes were not. To better visualise those unrelated genes, genes explicitly related to ribosomes and histones were removed and the remaining DE genes were re-visualised (see Figure 5.3). Interestingly, in the RGC comparison, expression of several genes linked to progenitor proliferation and neurogenesis was shown to be increased in the HETs, including *Hmg2a* (LFC = 0.3*,* p = 1.007867e-101), *Gli3* (LFC = 0.32*,* p = 1.162256e-84)*, Slc24a5* (LFC = 0.4*,* p = 1.017848e-78), and *Ddx3x* (LFC = 0.32*,* p = 4.751749e-8). Several of these genes were also dysregulated in the IP cluster. Moreover, the expression of several neurogenesis genes such as *Bcl11a* (LFC = 0.27*,* p = 7.305902e-10)*, Bcl11b* (LFC = 0.29*,* p =

4.502827e-15), and *Zeb2* (LFC = 0.33, p = 1.000000e+00), was also increased in the HETs within the neuron comparison.

Taken together, DE analysis revealed numerous biologically significant genes that were dysregulated in the HETs; however, these genes were overshadowed by a tremendous decrease in expression of many ribosomal and histone genes. In the following analyses, ribosome and histone genes were also found to be some of the most dysregulated genes. Although the dysregulation of these genes are interesting, it was decided to remove the genes from the subsequent DE analyses to focus on uncovering more distinctive DE genes.

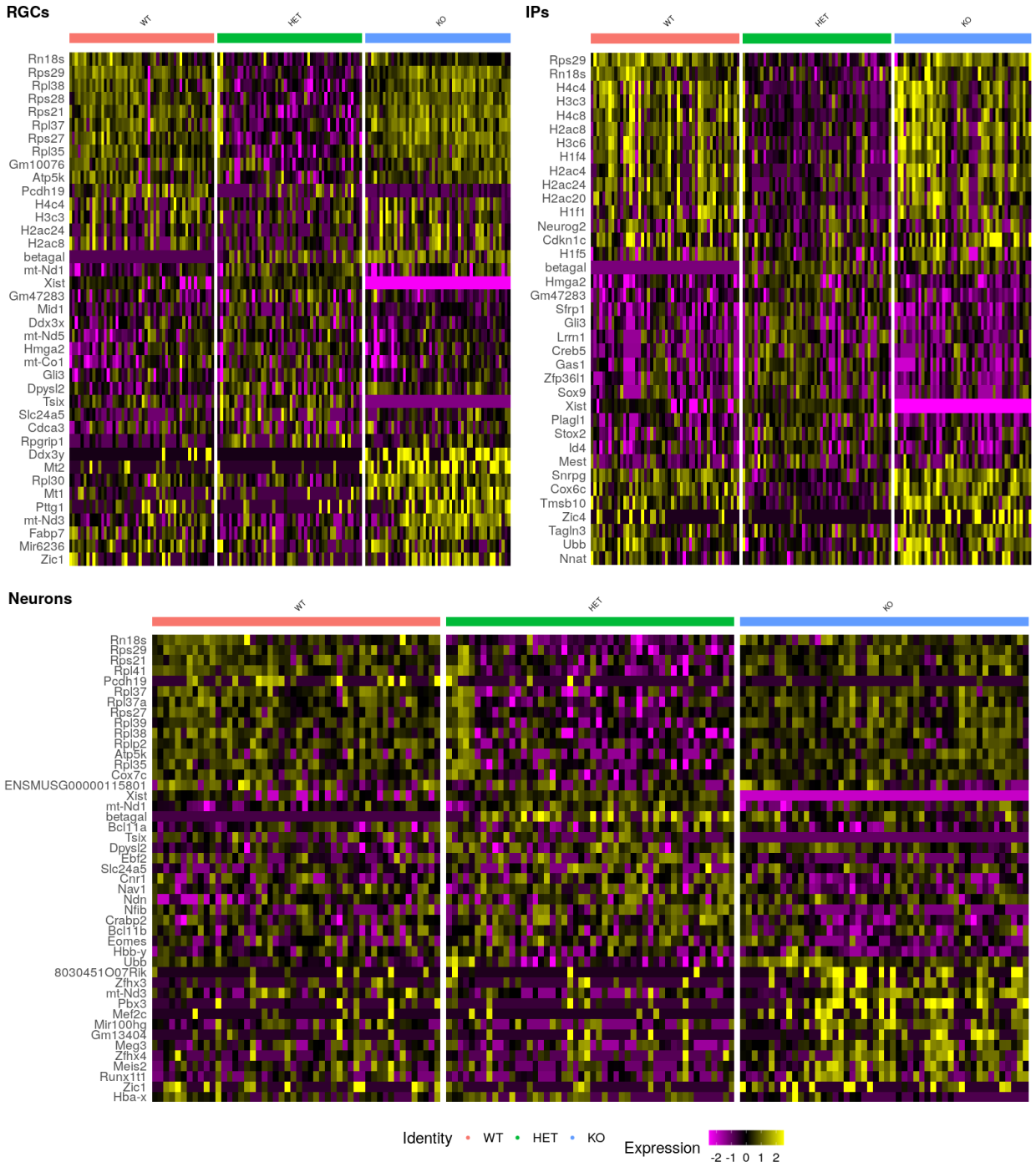


Figure 5.1. DE results between WT, HET, and KO within RGC, IP, and neuron clusters. Heatmap displaying the top 30 DE genes between WT, HET, and KO genotypes within the RGC cluster (top left), IP cluster (top right), and neuron cluster (bottom). Each point on the heatmap corresponds to a cell. The colour represents the expression level difference of the associated gene (yellow = upregulation, purple = downregulation).

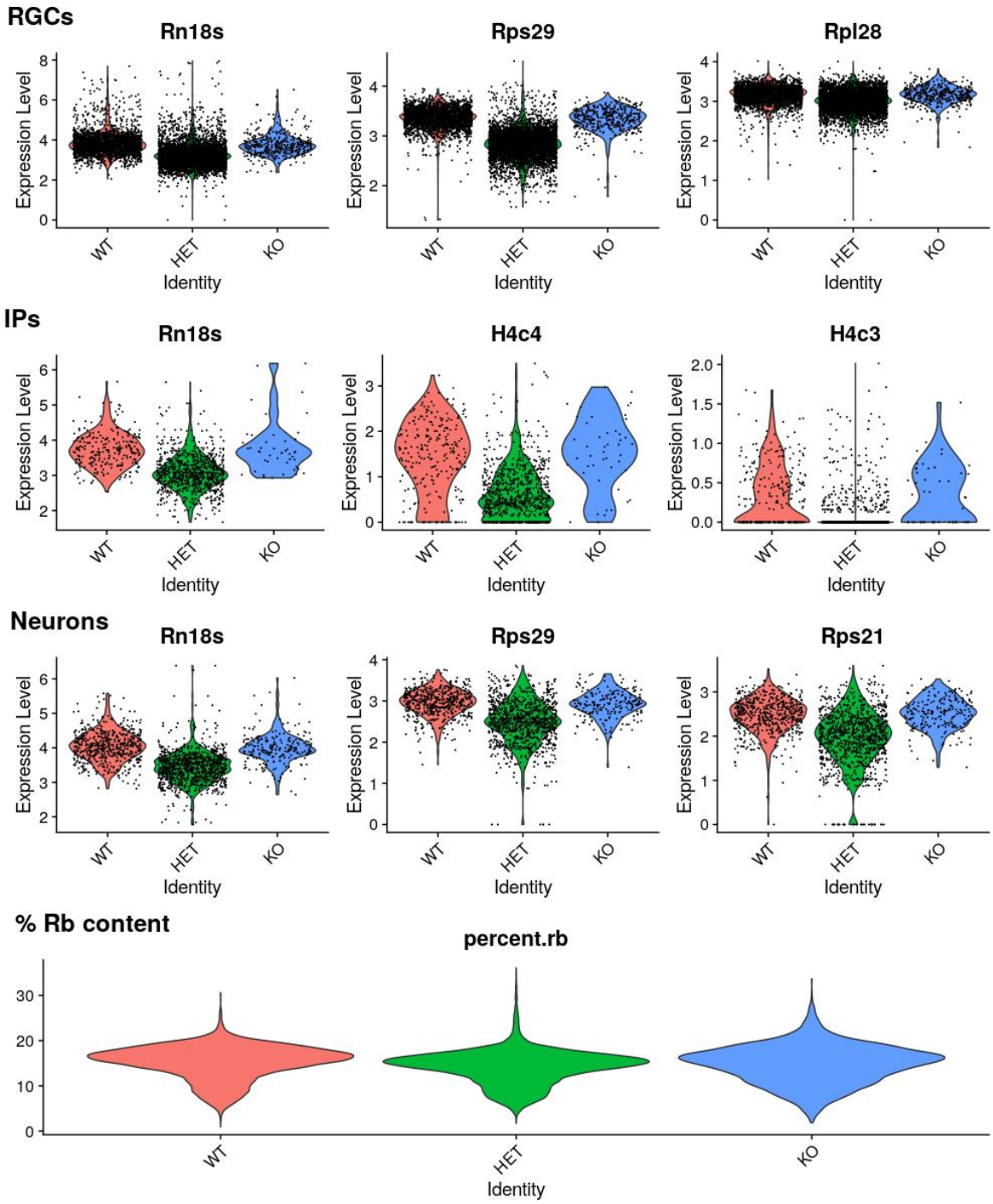


Figure 5.2. Expression of DE ribosomal and histone genes between WT, HET, and KO groups. Violin plots displaying the top three DE genes between WT, HET, and KO groups within the RGC cluster (top), IP cluster (top-middle), and neuron cluster (bottom-middle). The proportion of all ribosomal gene content between WT, HET, and KOs is also displayed (bottom). Expression of ribosomal genes were noticeably downregulated in the HET group.

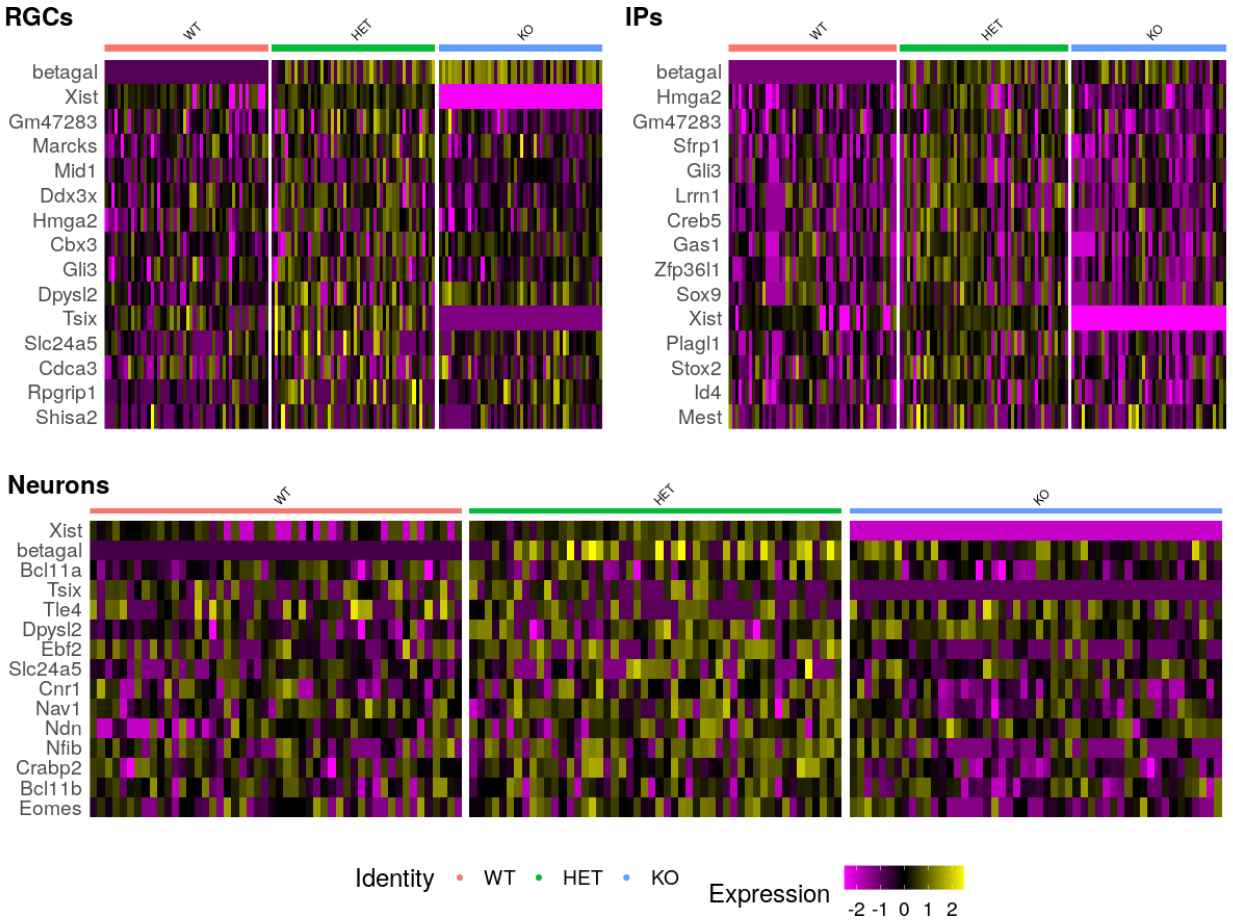


Figure 5.3. DE results between WT, HET, and KO groups within RGC, IP, and neuron clusters after filtering ribosomal and histone genes. Heatmap displaying the top 15 DE genes between the genotype groups in the RGC cluster (top left), IP cluster (top right), and neuron cluster (bottom). Each point on the heatmap corresponds to a cell. The colour represents the expression level difference of the associated gene (yellow = upregulation, purple = downregulation).

5.2.2. DE analysis using DESeq2: WT vs. HET and WT vs. KO

As aforementioned, single cell DE analysis can often produce exaggerated lists of DE genes due to high statistical sensitivity. To work around this, a pseudo-bulk DE analysis was also performed within each cell type to determine whether DE genes can be detected at the population level, rather than the cell level. This was achieved by subsetting the RGCs, IPs, and neurons into separate cell matrices and extracting the raw counts for each gene. The metadata and counts were then averaged and aggregated for each genotype sample, which provided a dataset suitable to perform pseudo-bulk analysis using DESeq2 (Love et al., 2014). The threshold for significance was set to p adjusted value < 0.05 and a log fold change of 0.5, as recommended by the guidelines of the statistical package. Moreover, because DESeq2 can perform DE analysis only between two groups, each mutant genotype group was separately compared to WTs within each cell cluster (i.e., WT vs. HET, and WT vs. KO) (see Figure 5.4).

DE analysis between WTs and HETs within the RGC cluster revealed only three DE genes, including *Pcdh19* (LFC = -0.8, p = 1.314660e-07), and *β-gal* (LFC = 4.43, p = 4.184170e-08) which were downregulated and upregulated in the HETs, respectively. Additionally, *Pianp* (LFC = -0.66, p = 7.650085e-29) was also downregulated in the HETs. Likewise, analysis within the IP cluster revealed *β-gal* (LFC = 4.13, p = 1.477877e-17) as the only DE gene which was upregulated in the HETs. Analysis within the neuron cluster also revealed three DE genes. As expected, *Pcdh19* (LFC = -1.04, p = 1.385405e-07) and *β-gal* (LFC = 2.8, p = 2.083954e-15) were downregulated and upregulated, respectively, however *Ccnd2* (LFC = 0.84, p = 3.873490e-06) was also shown to be upregulated in the HETs.

Furthermore, DE analysis between WTs and KOs within the RGC cluster revealed four DE genes, including *Pcdh19* (LFC = -1.73, p = 1.741206e-33) and *β-gal* (LFC = 4.72, p = 1.151672e-29) which were downregulated and upregulated in the KOs, respectively. Moreover, the female markers *Xist* (LFC = -0.82, p = 1.670376e-23) and *Tsix* (LFC = -2.1, p = 5.239346e-2) were also shown to be significantly downregulated in the KOs. Analysis in the IP cluster revealed several other DE genes, including *Sulf1* (LFC = 1.01, p =1.116576e-05)*, Nkx2-1* (LFC = 1.6, p = 9.227850e-07)*,* and *Tppp3* (LFC = 1.58, p = 6.100731e-09). Finally, analysis within the neuron cluster revealed four DE genes, including *Pcdh19* (LFC = -1.22, p = 2.637939e-06), *β-gal* (LFC = 4.22, p = 5.845163e-18), *Xist* (LFC = -3.29, p = 4.747670e-16) and *Tsix* (LFC = -1.96, p = 1.441014e-09).

It was decided that DE genes detected using pseudo-bulk should be re-examined at the single cell level, except for the genotype and sex markers *Pcdh19*, *β-gal, Xist,* and *Tsix* (see Figure 5.5). Interestingly, *Ccnd2* was shown to be upregulated in HETs both in neurons and in RGCs, however levels of the *Ccnd2* seemed relatively normal in IPs compared to WTs (see Figure 5.5). Moreover, *Ccnd2* was also downregulated in the KO RGC cluster (see Figure 5.5). Expression differences of *Pianp*, *Sulf1*, and *Tppp3* were also shown, however very few cells actually expressed these genes. It is possible that the aggregation effect of pseudo-bulk means that genes that are expressed in few cells can become statistically inflated at the population level. Therefore, it was decided not to use pseudo-bulk analysis for subsequent comparisons.

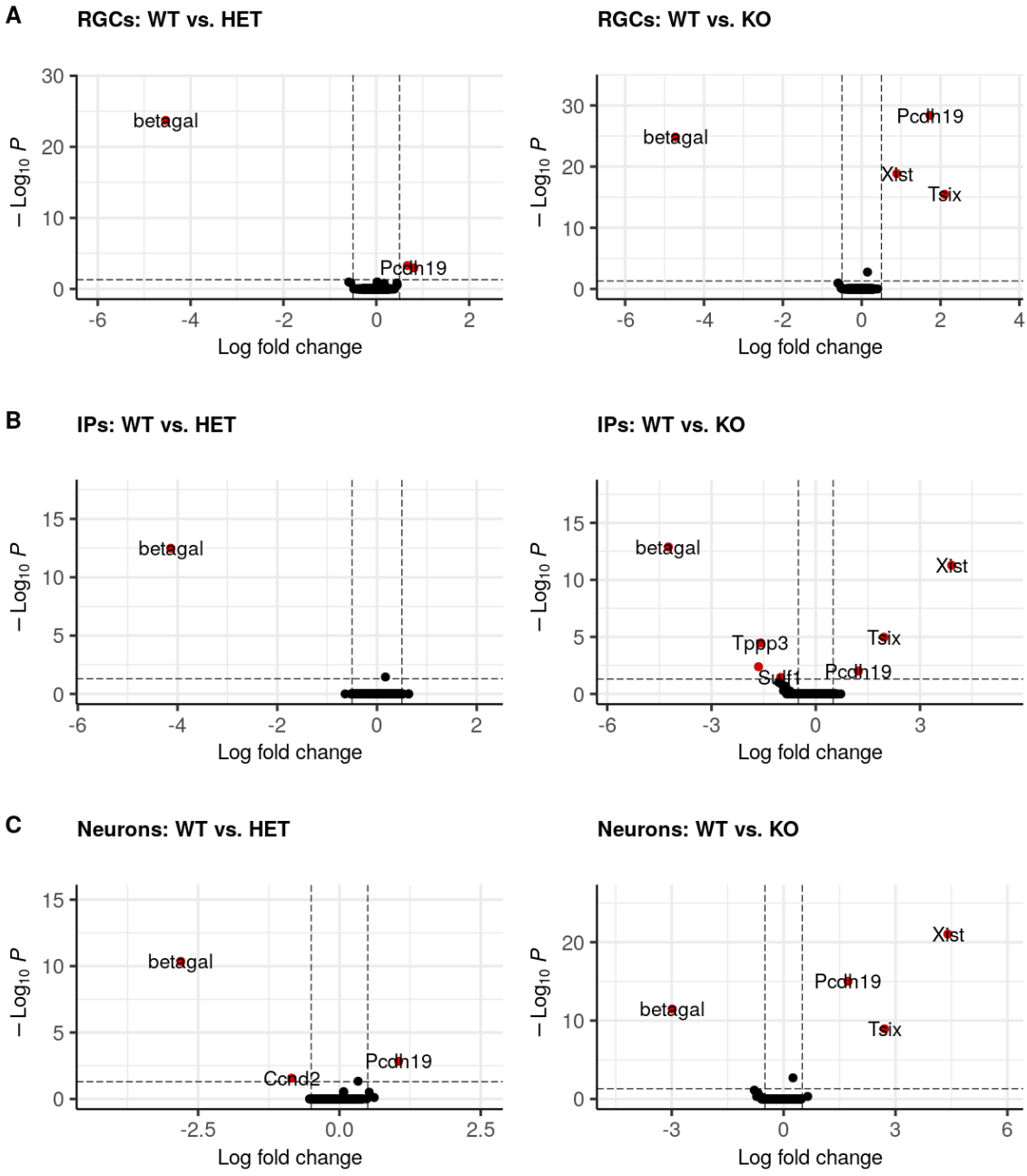


Figure 5.4. DE results of WT vs. HET and WT vs. KO in RGC, IP, and neuron clusters using DESeq2. (A) Volcano plot displaying DE genes between WT and HET (left) and WT and KO (right) within the RGC cluster. (B) Volcano plot displaying DE genes between WT and HET (left) and WT and KO (right) within the IP cluster. (C) Volcano plot displaying DE genes between WT and HET (left) and WT and KO (right) within the neuron cluster. DE genes are represented in red. non-DE genes are represented in black. Plots were produced after performing the pseudo-bulk analysis on the sc-RNA-seq data.

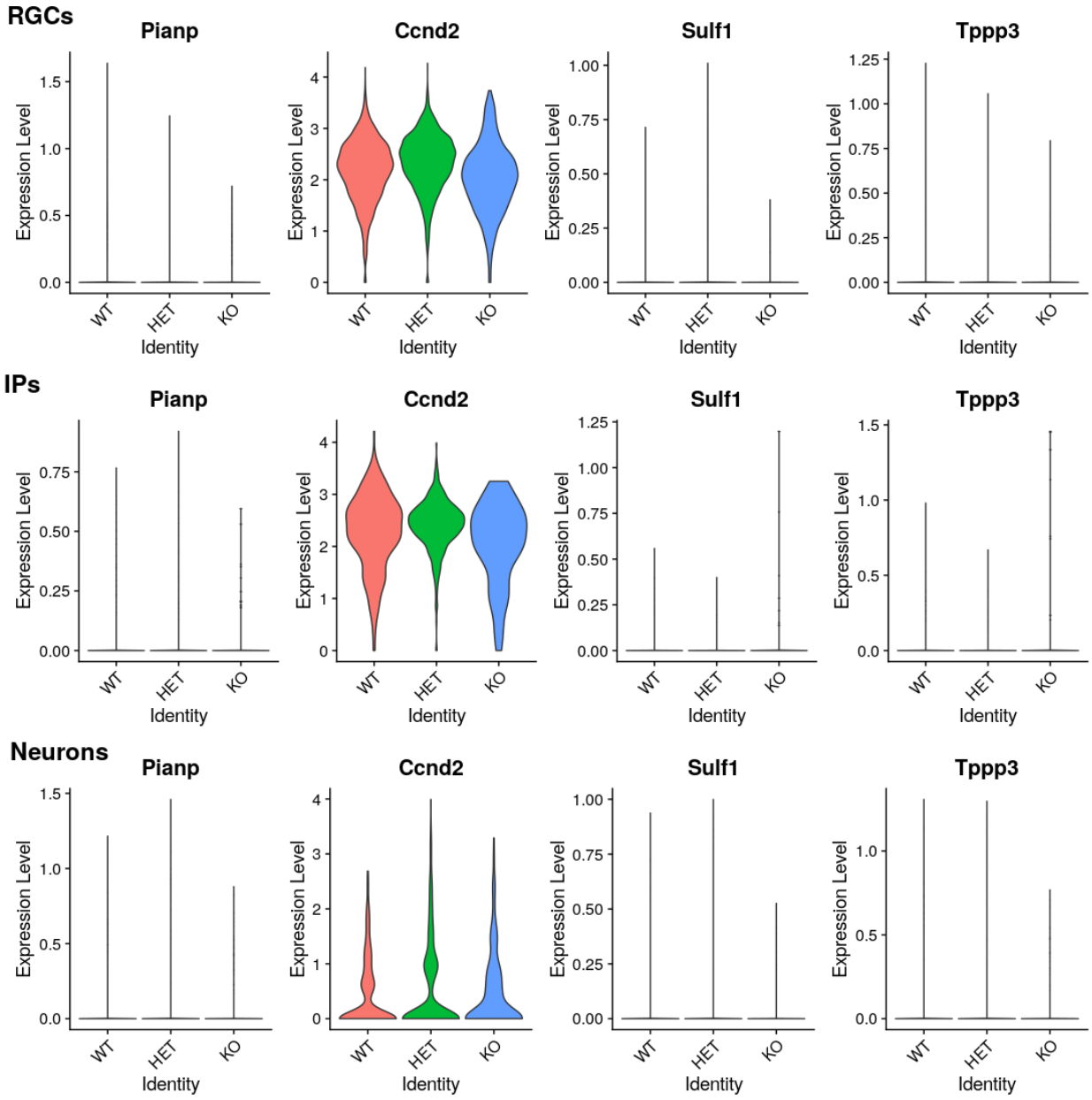


Figure 5.5. Expression of DE genes from pseudo-bulk using single cell DE analysis. Violin plots displaying the expression of the genes *Pianp, Ccnd2, Sulf1,* and *Tppp3* in RGCs (top), Ips (middle), and neurons (bottom) between WT, HET, ad KO groups. Genes were discovered as differentially expressed from the pseudo-bulk analysis and replotted in Seurat to examine their expression at the single cell level.

5.2.3. Single cell DE analysis: Mutants vs. same sex WT counterparts

In the previous section, it was noticed that several of the most differentially expressed genes between WT, HET, and KO were related to sex, especially *Xist, Tsix,* and *Ddx3y*. Because sex has been shown to be a major confounding variable (see chapter 3), this raises the possibility that sex may play a role in the transcriptional alterations observed in the previous sections. To further focus the DE analysis and reveal genes differentially expressed due to genotype, WT male and female cells were subset from the dataset using the parameters established previously (see section 4.2.6). Mutant groups (HETs and KOs) were then compared to their WT same sex counterparts within the RGC, IP, and neuron clusters (see Figure 5.6 and Figure 5.7).

Analysis between WT female and HETs revealed numerous DE genes between both groups within the RGC (97 genes; 68 upregulated, 31 downregulated), IP (288 genes; 132 upregulated, 156 downregulated) and neuron (70 genes; 52 upregulated, 18 downregulated) comparisons. Likewise, DE analysis between WT male and KOs revealed numerous DE genes within the RGC (40 genes; 28 upregulated, 20 downregulated) and neuron (13 genes; 6 upregulated, 7 downregulated) clusters. Interestingly, no DE genes were found within the IP cluster.

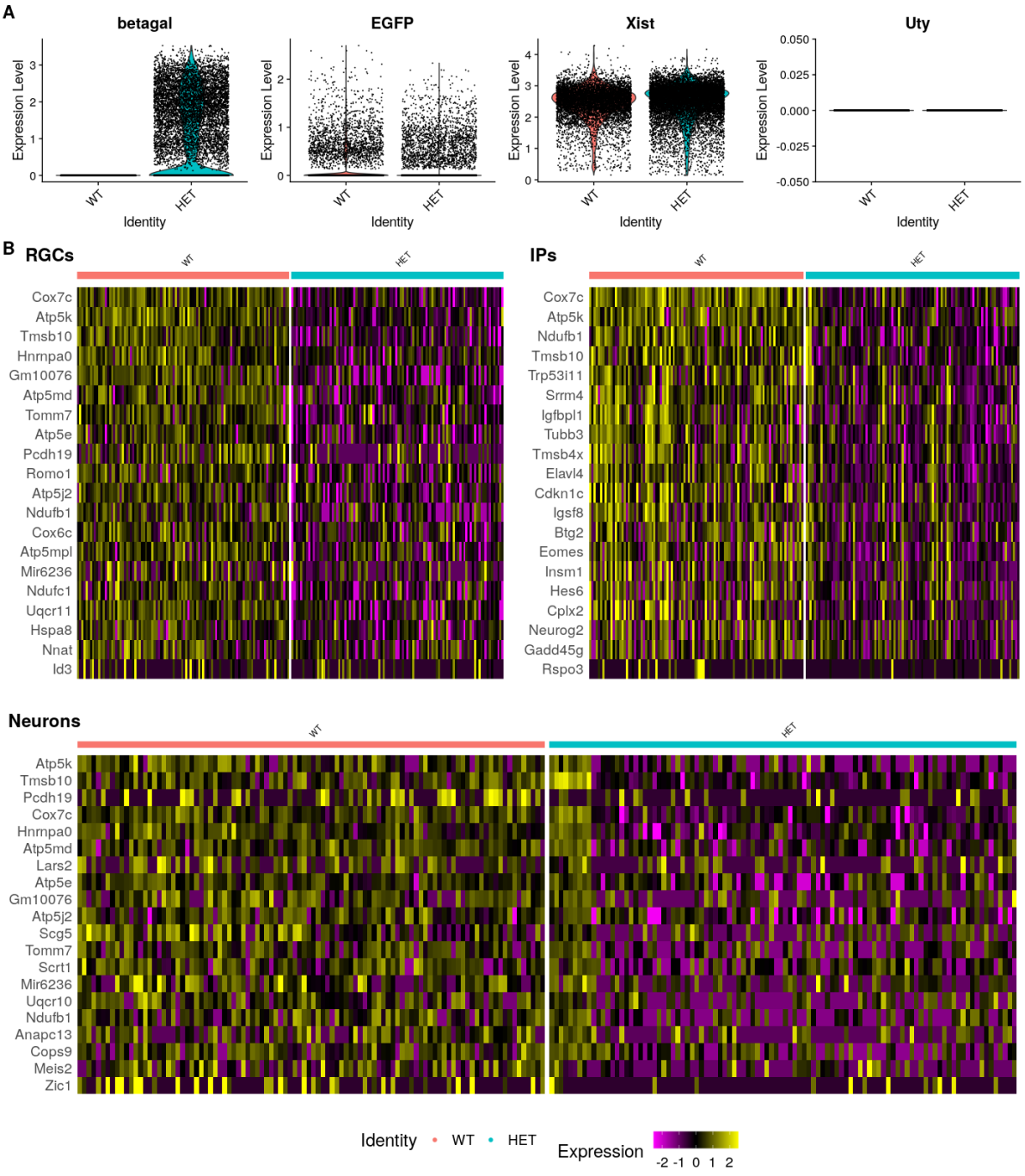


Figure 5.6. DE results of WT female vs. HETs in RGC, IP, and neuron clusters. (A) Violin plots illustrating he expression of *β-gal, EGFP, Xist,* and *Uty* to demonstrate proper filtering of WT female cells. (B) Heatmaps displaying the top 20 DE genes within the RGC (top left), IP (top right), and neuron (bottom) clusters between WT female and HET groups. Each point on the heatmap corresponds to a cell. The colour represents the expression level difference of the associated gene (yellow = upregulation, purple = downregulation).

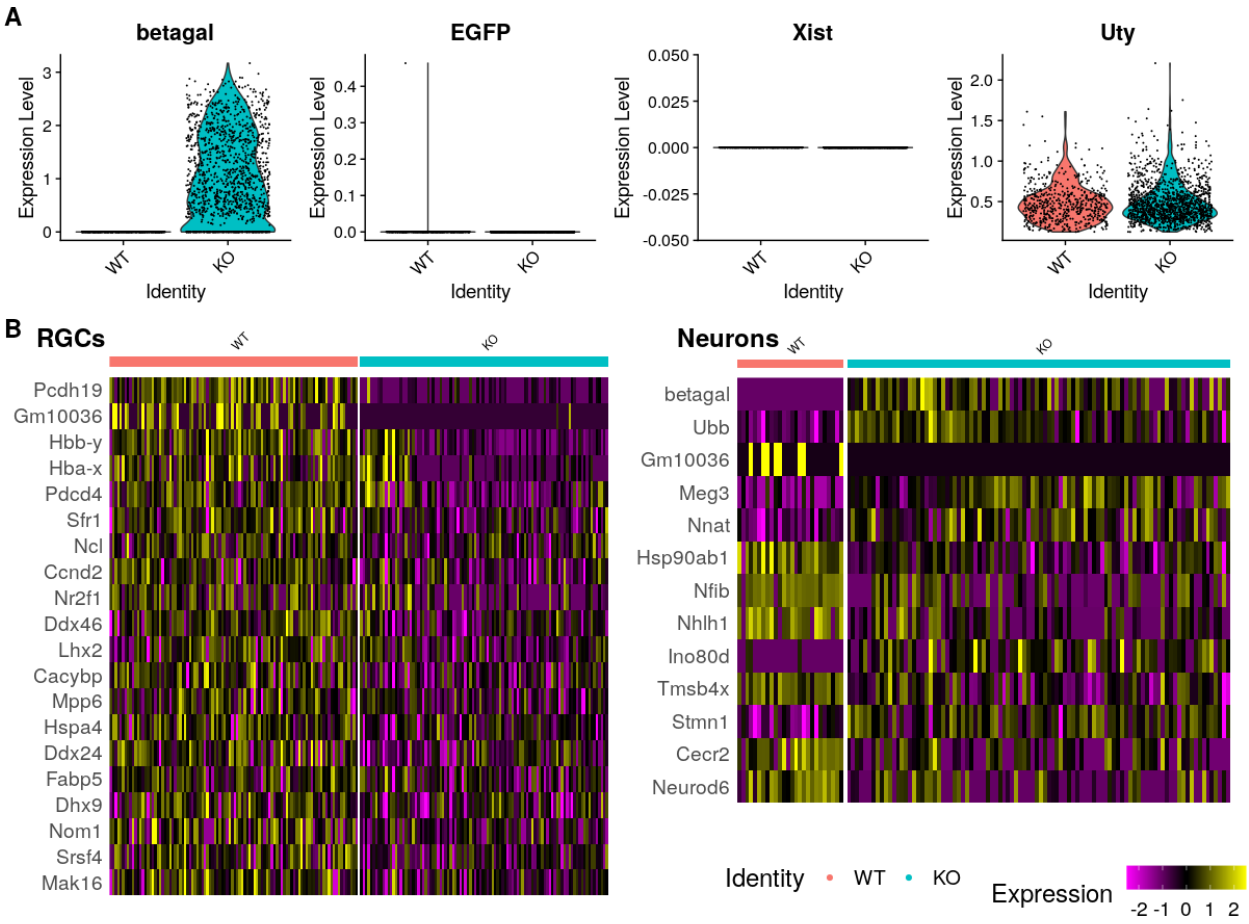


Figure 5.7. DE results of WT male vs. KOs in RGC, IP, and neuron clusters. (A) Violin plots illustrating he expression of *β-gal, EGFP, Xist,* and *Uty* to demonstrate proper filtering of WT male cells. (B) Heatmaps displaying the top 20 DE genes within the RGC (left), and neuron (right) clusters between WT male and KO groups. Each point on the heatmap corresponds to a cell. The colour represents the expression level difference of the associated gene (yellow = upregulation, purple = downregulation).

5.2.4. Enrichment analysis: WT female vs. HET and WT male vs. KO

To better interpret the biological significance of the DE genes from the RGC clusters, gene ontology (GO) analysis was performed using the online resource ShinyGo (version 0.7.3). DE genes were grouped using GO terms (Biological Processes (BP)) and the threshold for enrichment was set as FDR < 0.05 (see Figure 5.8 and Figure 5.10). All overrepresented pathways can be found in the appendices at the back of this thesis.

From the WT female vs. HET comparison, many DE genes were found to be overrepresented in pathways related to development and neurogenesis in the RGC comparisons, including *central nervous system neuron differentiation* (p = 2.9E-06)*, neurogenesis* (p = 3.1E-06), and *brain development* (p = 4.2E-03) (see Figure 5.8). Interestingly, *cell-cell signalling by Wnt* (p = 1.0E-05) was also shown to be significantly overrepresented. To investigate this further, the expression of the DE genes that were related to *cell-cell signalling by Wnt* were examined between WT and HETs within the RGC cluster (see Figure 5.9). Interestingly, all of the DE genes considered in this category, namely *Dd2x, Rspo3, Wnt8b, Gli3, Sall1, Shisa2,* and *Hmga2* were found to be upregulated in the HETs.

Similarly, GO analysis was also performed using the DE genes from the WT male vs. KO comparison (see Figure 5.10). DE genes were overrepresented in several pathways associated with development and neurogenesis, however other pathways involved in neuron morphogenesis and phosphorylation were also found, including *neuron projection morphogenesis* (p = 0.006) and *axon development* (p = 0.006). Interestingly, when examining the most overrepresented KEGG pathways, the Wnt signalling was also found to be overrepresented. The DE genes categorised within the Wnt signalling KEGG were *Ccnd2* and *Lhx2* and were all significantly downregulated in the KOs (see Figure 5.11).

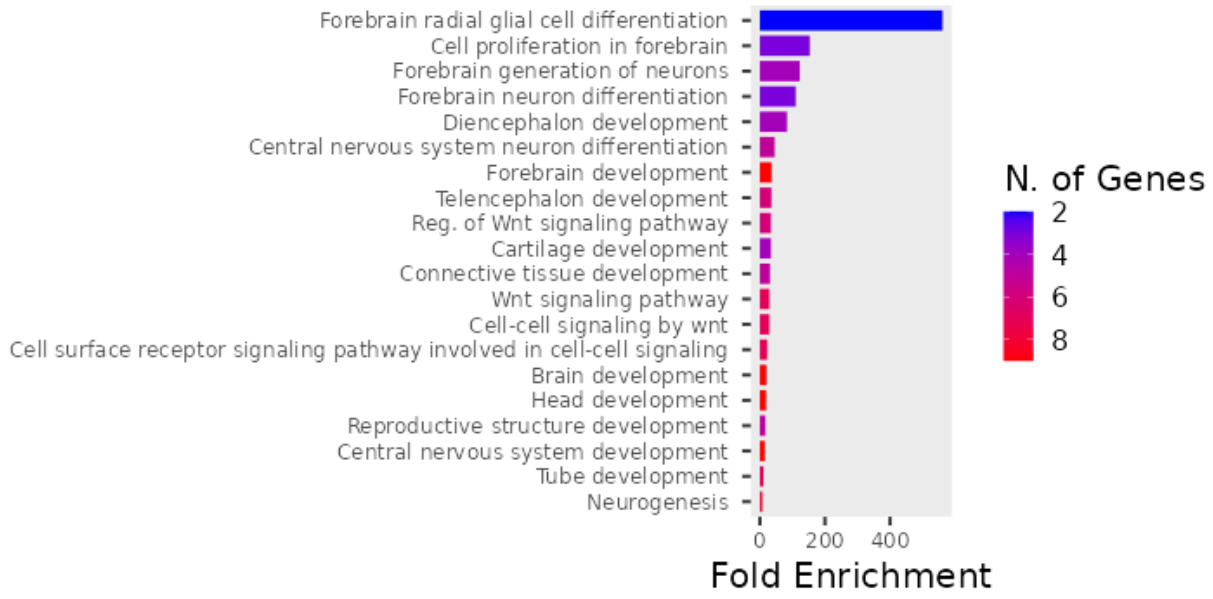


Figure 5.8. Enrichment analysis results of DE genes between WT female and HET groups. Bar plots depicting the top 20 enriched GO terms from DE genes within the RGC cluster. Interestingly pathways shown to be significantly overrepresented include *central nervous system neuron differentiation* (p = 2.9E-06)*, neurogenesis* (p = 3.1E-06), *brain development* (p = 4.2E-03), and *cell-cell signalling by Wnt* (p = 1.0E-05).

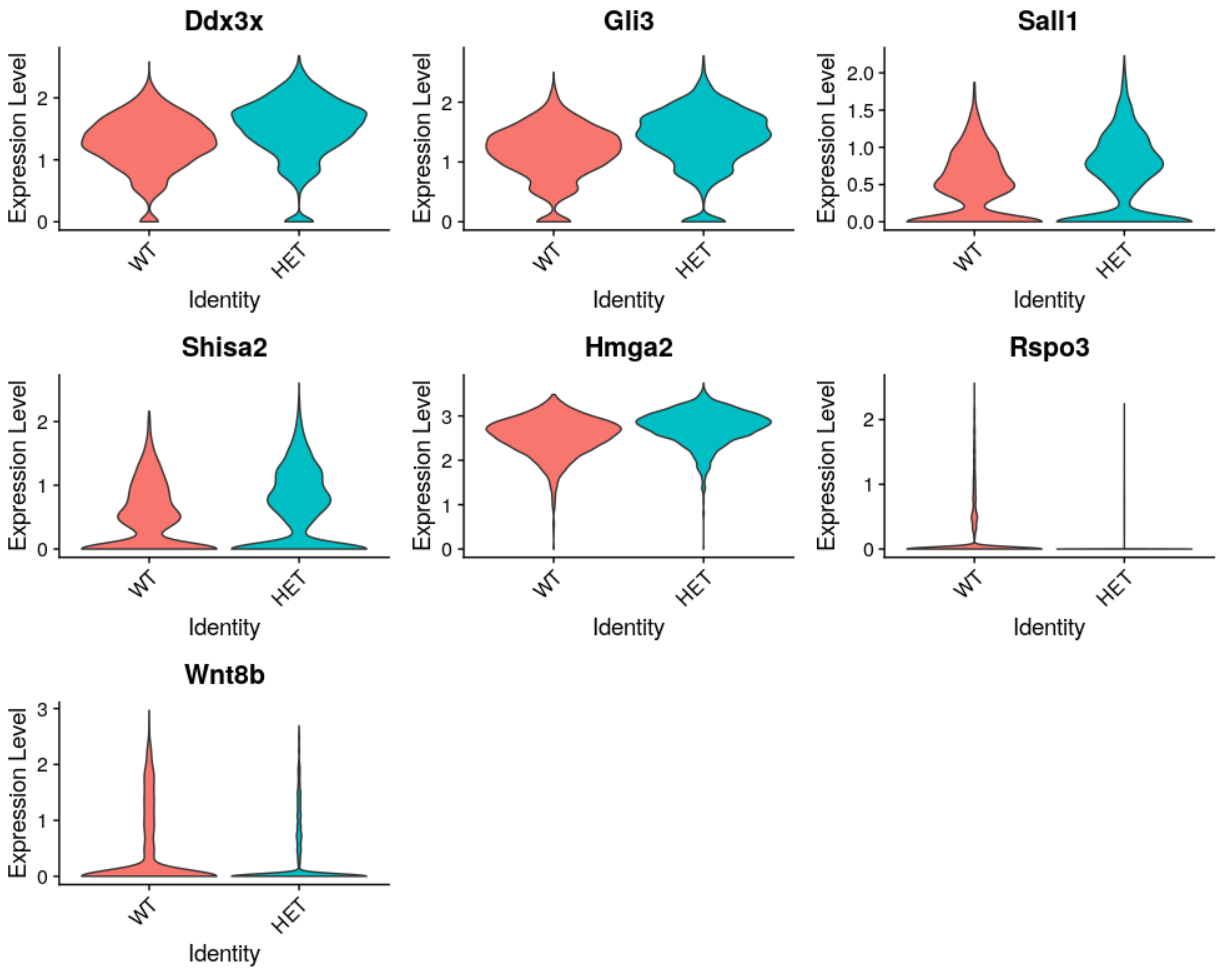


Figure 5.9. Expression of DE genes related to Wnt signalling between WT female and HET groups. Violin plots illustrating the expression of *Ddx3x, Gli3, Sall1, Shisa2,* and *Hmg2a* between WT female and HETs in RGCs. The DE genes were selected after performing overrepresentation analysis using ShinyGO, and determining which genes were categorised in the Wnt signalling pathway GO term.

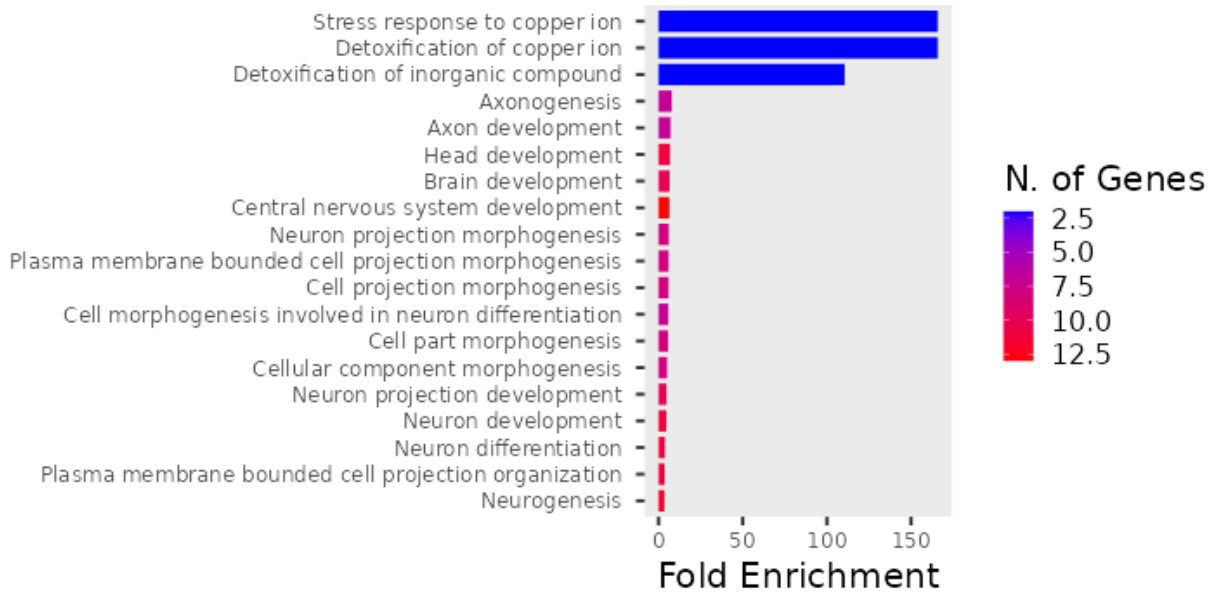


Figure 5.10. Enrichment analysis results of DE genes between WT male and KO groups. Bar plots depicting the top 20 enriched GO terms from DE genes within the RGC cluster. Interesting pathways included *neuron projection morphogenesis* (p = 0.006) and *axon development* (p = 0.006).

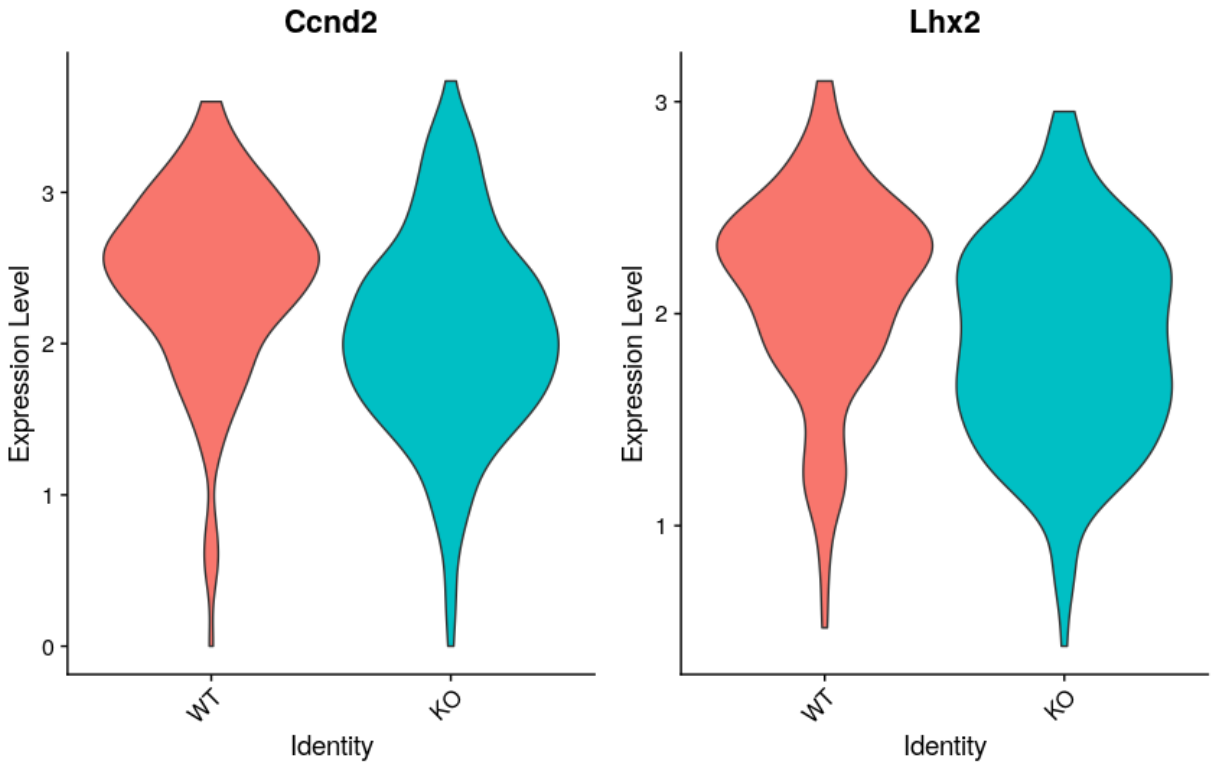


Figure 5.11. Expression of DE genes related to Wnt signaling between WT male and KO groups. Violin plots illustrating the expression of *Ccnd2* and *Lhx2,* between WT male and KOs in RGCs. The DE genes were selected after performing overrepresentation analysis using ShinyGO, and determining which genes were categorised in the Wnt signalling pathway GO term.

5.2.5. Single cell DE analysis: WT and KO cells within the HET

To perform analysis on WT and KO cells within the HET, the cells were first annotated as “HET (WT)”, or “HET (KO)” based on their expression levels of *β-gal, EGFP,* and *Pcdh19*. Briefly, cells that expressed > 0 *Pcdh19* and no *β-gal* were considered as HET (WT), and cells that expressed > 0 *β-gal* but no *EGFP* were considered as HET (KO) (see Figure 5.12A). It is important to note that *Pcdh19* expression was not used to define the HET (KO) cells due to the aforementioned residual expression of Pcdh19 in KO animals (see section 4.3.3). To examine the overall transcriptional similarity of the cells, the cells were visualised using UMAPs (see Figure 5.12B). Interestingly, within the RGC cluster, HET WT and HET KO cells seemed to co-cluster regardless of cell genotype. Moreover, this effect was also observed in the IP and neuron clusters, however it was also noticed that the number of HET WT cells was strikingly lower than HET KO cells (see Figure 5.12B).

Several DE genes were found within the IP (8 genes; 2 upregulated, 6 downregulated) and neuron (11 genes; 9 upregulated, 2 downregulated) comparisons. Astonishingly, the DE analysis between HET WT and HET KO cells within the RGC cluster revealed only three DE genes – *Pcdh19* (LFC = 1.02, p = 0.000000e+00) and *EGFP* (LFC = 0.58, p = 1.637703e-273) were shown to be upregulated in the HET WT cells, whereas *β-gal* (LFC = -3.29, p = 0.000000e+0) was found to be downregulated.

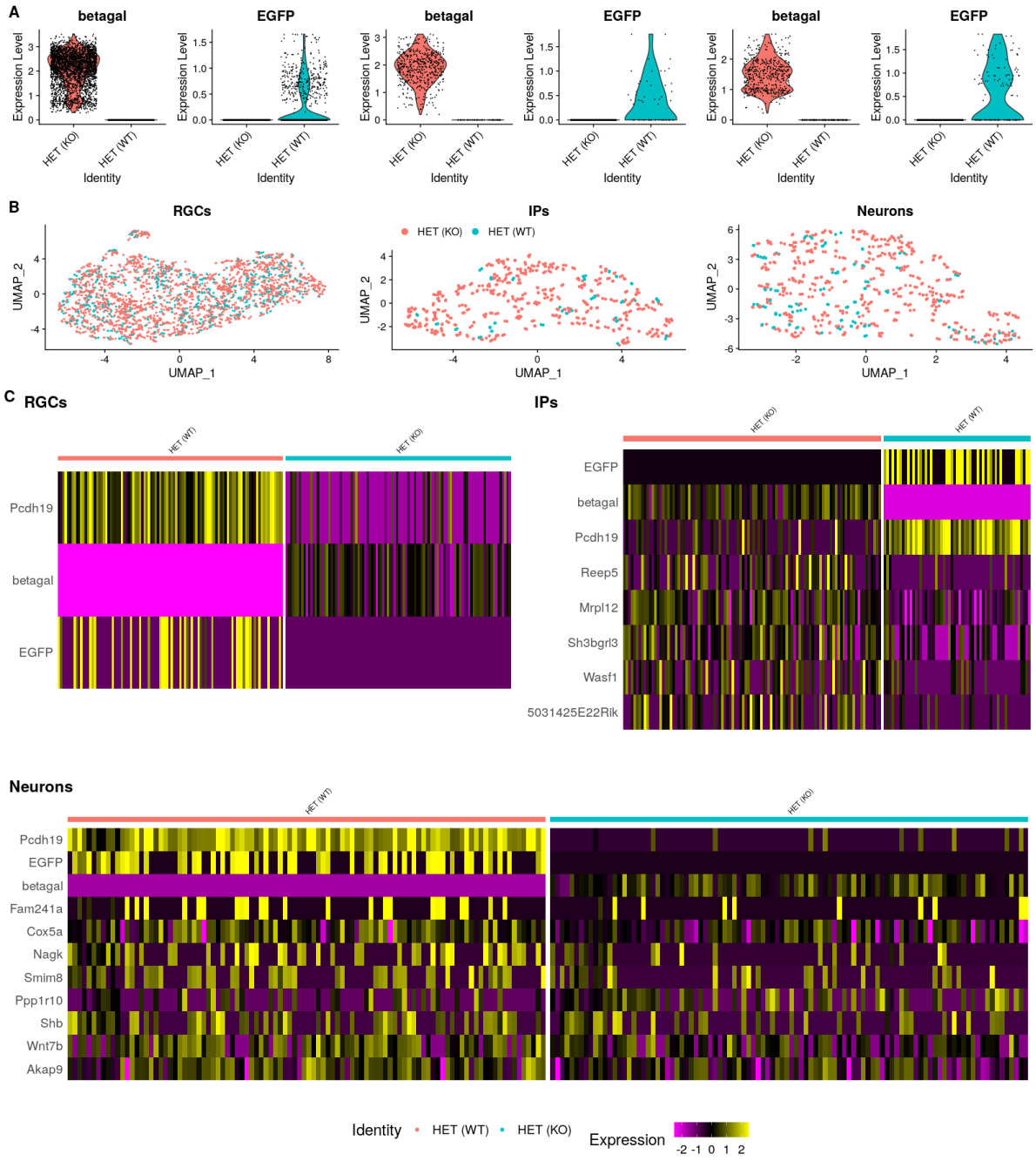


Figure 5.12. DE results comparing HET WT and HET KO cells from RGC, IP, and neuron clusters. (A) Violin plots showing the expression levels of *β-gal* and *EGFP* in RGCs (left), IPs (middle) and neurons (right). (B) UMAPs displaying the position of the HET WT and HET KO RGCs (left), IPs (middle), and neurons (right). (C) Heatmap showing DE results from the RGC (top left), IP (top right), and neuron (bottom) comparisons between HET WT and HET KO cells. Each point on the heatmap corresponds to a cell. The colour represents the expression level difference of the associated gene (yellow = upregulation, purple = downregulation).

5.2.6. Single cell DE analysis: WT female vs. WT/KO HET cells

Although few DE genes were uncovered between HET WT and HET KO RGCs, it was hypothesised that because these cells exhibit altered neurogenic behaviour, the potential underlying transcriptional alterations may instead be identified by comparing each cell genotype from the HETs to WT female cells i.e., “normal” cells. To achieve this, HET WT and HET KO RGCs were annotated using the same approach described previously (see section 5.2.5) and then separately compared to WT female RGCs. Interestingly, analysis between WT female and HET WT RGCs revealed numerous DE genes (99 genes; 64 upregulated, 45 downregulated). Likewise, analysis between WT female and HET KO RGCs also revealed numerous DE genes (99 genes; 68 upregulated, 31 downregulated).

In a previous analysis, it was shown that when comparing WT female RGCs to the RGCs of HETs as a whole, multiple DE genes were found (see section 5.2.3). Therefore, it was theorised that many of these DE genes would also be present after comparing HET WT or HET KO RGCs to WT females, meaning that it was probable that these DE genes are likely differentially expressed in the HET regardless of cell genotype. However, it was also possible that a subset of DE genes may also be distinctly dysregulated in each cell genotype group. To examine this, the list of DE genes from the WT female vs. HET RGC comparison was compared to the lists of DE genes from the WT female vs. HET WT and HET KO comparisons. Interestingly, ~86% DE genes from the WT female vs. HET WT comparison were also found in the WT female vs. whole HET comparison. Moreover, ~96% DE genes from the WT female vs. HET KO comparison were also identified in the WT female vs. whole HET comparison. Therefore, each list was filtered to only include the distinct DE genes (see Figure 5.13). Interestingly, 16 genes were shown to be distinctly dysregulated in the HET WT RGCs (8 upregulated, 8 downregulated) and 6 genes were shown to be distinctly dysregulated in the HET KO RGCs (4 upregulated, 2 downregulated). Examining the list of DE genes revealed that several genes were related to progenitor proliferation and neurogenesis. Genes related to proliferation, including *Foxg1* (LFC = 0.28, p = 1.379092e-16), *Id4* (LFC = 0.25, p = 5.449485e-15), and *Cdon* (LFC = 0.31, p = 6.041457e-15), were all upregulated in the HET WT cells. Moreover, genes related to neurogenesis, including *Neurog2* (LFC = -0.27, p = 1.452112e-04), and *Cdkn1c* (LFC = -0.28, p = 2.720537e-3) were mostly downregulated in the HET WTs except for *Foxp1* (LFC = 0.26, p = 7.054160e-05). Examining the list of DE genes from the HET KO cells also revealed that *Foxg1* (LFC = 0.29, p = 1.578255e-23), expression was increased in the HET KO cells compared to WTs.

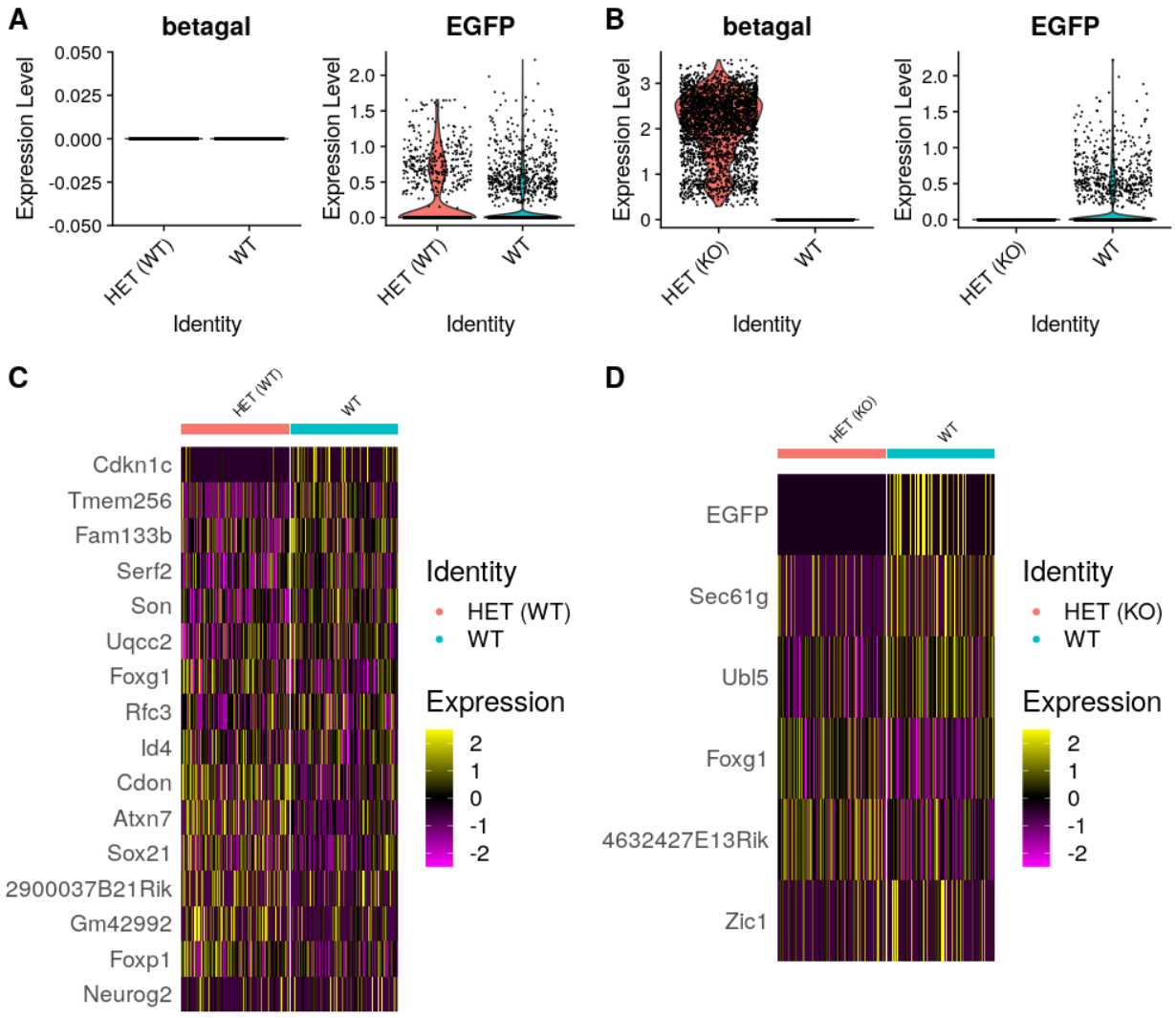


Figure 5.13. DE results comparing HET WT and HET KO RGCs to WT female RGCs. (A) Violin plots illustrating the expression of cell genotype markers *β-gal* and *EGFP* in HET WT and WT female RGCs. (B) Violin plots illustrating the expression of cell genotype markers *β-gal* and *EGFP* in HET KO and WT female RGCs. (C) Heatmap displaying the DE results comparing HET WT RGCs to WT female RGCs. (D) Heatmap displaying the DE results comparing HET KO RGCs to WT female RGCs. Each point on the heatmap corresponds to a cell. The colour represents the expression level difference of the associated gene (yellow = upregulation, purple = downregulation).

5.3. Discussion

The primary aim of this chapter was to perform DE analysis between WT, HET, and KO groups at the single cell level and identify DE genes that may underpin the role of *Pcdh19* in neurogenesis. Analysis was performed between these groups within the RGC, IP, and neuronal clusters. Moreover, two methods of DE analysis were also employed to examine DE genes at the single cell and population levels using Seurat and pseudo-bulk, respectively. Another goal was to perform DE analysis between WT and KO cells within the HET to identify DE genes that may allude to the unusual neurogenesis phenotype of these animals.

5.3.1. DE analysis from IPs produced more DE genes than RGCs and neurons

One of the most consistent findings in this chapter was that DE analysis between WT, HET, and KO animals within the IP cluster sometimes produced the largest lists of DE genes compared to RGCs and most neuron comparisons. As aforementioned, the primary cell type of interest to analyse in this chapter was the RGCs, since RGCs express the highest level of *Pcdh19* and mosaic expression of the gene within the HETs causes from WT and KO RGCs to separate. Therefore, it was believed that the RGCs would show the most significant DE genes. Furthermore, new-born HET IPs also separate from WT and KO IPs, but this diminishes as the cells become more neuronal-like. Likewise, WT, HET, and KO neurons co-cluster indiscriminately, which overall suggested that mosaic expression of *Pcdh19* in the RGCs and new-born IPs of HETs causes significant transcriptional alterations compared to WT and KO cells, yet these differences are diminished as *Pcdh19* expression is silenced and the cells commit to neurogenesis. However, since significantly more DE genes were found from the IP and neuron comparisons, this result initially suggests that these cells exhibit greater transcriptional variation than the RGCs.

Nevertheless, one possible explanation of this finding is that IPs may exhibit higher cell-cell gene expression variability than RGCs. Cell-cell variability is when genes that are detected as expressed at moderate or high levels in some cells are also detected lowly expressed in other cells from the same sample (Kharchenko et al., 2014). Therefore, these genes are often detected as differentially expressed; however, this may be due to a number of confounding reasons. For example, levels of cell-cell variability in gene expression can be greater within a cell cluster wherein the cells are undergoing significant co-ordinated changes in gene transcription at different times. This is best exemplified by IPs which at E11 are more transcriptionally heterogenous than RGCs because their transcriptional profiles become more fate-restricted towards neurogenesis (Munsky et al., 2012; Ruan et al., 2021). In other words, the IP cluster can be considered as a

continuum wherein at the transcriptional level, new-born IPs are more progenitor-like and become more neurogenic overtime. This can be seen via the UMAP in the previous chapter whereby the IPs form a “bridge”-like cluster that connects the RGC and neuron clusters, which visually represents the IPs transcriptional landscape changing from progenitor-like to neuronal-like in a continuous process (see section 4.2.7). Therefore, compared to RGCs and neurons, IPs exhibit considerable within-cluster gene variability which may explain why DE analyses within this cluster produced the largest list of DE genes. One way to improve this analysis would be to perform DE analyses along specific points within the IP continuum. For example, DE analysis could be performed between genotype groups on IP cells subset from the main IP cluster that were more progenitor-like (closer to RGCs), or more neuronal-like (closer to neurons), or in the middle. This would in theory correct some of the transcriptional variability and focus analyses by uncovering DE genes due to genotype at different time points within the IP cluster. Taken together, although interesting, the DE analyses from IP and some neuron comparisons should be considered with caution since it is difficult to deduce whether the list of DE genes produced from these comparisons are related to changes driven by genotype or other extraneous variables.

5.3.2. Single cell and pseudo-bulk methods produced varying lists of DE genes

As previously discussed, identifying DE genes between conditions from scRNA-seq can be challenging. Single cell DE methods such as the inbuilt technique employed by Seurat treats every cell as a sample in the DE analysis. This is often useful since it considers within-sample variation in gene expression (Hao et al., 2021). However, because the analyses can involve thousands of cells, the p-values can be significantly inflated. Therefore, in this chapter single cell DE analysis tools were initially used alongside a “pseudo-bulk” analysis method which works around the high statistical sensitivity of single cell analysis methods by averaging the reads across cells and considers the transcriptional profile of the overall population (Luecken and Theis, 2019; Squair et al., 2021).

DE analysis between WTs, HETs, and KOs using both methods showed that pseudo-bulk produced far fewer DE genes than the single cell Seurat method. Indeed, it was found that the most DE genes from the pseudo-bulk between WT vs. HET and WT vs KO were the genotype markers *Pcdh19* and *β-gal* as well as the sex markers *Xist* and *Tsix*. Nevertheless, other DE genes were identified; in the RGC comparison, the gene *Pianp* was shown to be downregulated in HET samples, and in the neuron comparison, *Ccnd2* was downregulated in the HETs. Interestingly, KOs displayed fewer DE genes except for the aforementioned genotype and sex

markers, however in the IP comparison, it was found that *Pianp*, *Tppp3,* and *Sulf1* were upregulated in the KOs.

The dysregulated genes have all been associated with neural progenitor behaviour and neurogenesis. For example, *Pianp* has been shown to be involved in regulating neurogenesis in human RGCs as a gene imprinting regulatory element (Liang et al., 2022). Moreover, *Sulf1* has been mostly studied as a modulator of neural progenitor fate in the spinal cord by controlling sonic hedgehog (Shh) signalling (Touahri et al., 2012). *Tppp3* is less well studied in the context of bran development, however recent evidence has linked the gene as a regulator of the tubulin cytoskeleton during epithelial-mesenchymal transition in glioblastoma (Xu et al., 2022). Therefore, it is possible it may provide a similar function during RGC-IP transitions. Lastly, *Ccnd2* encodes for the Cyclin D2 protein which is a key component of the cell cycle machinery and integral in controlling the transition between G1 and S phases (Lange and Calegari, 2010). Along with the other cyclin proteins (Cyclin D1 and D3) and associated cyclin-dependent kinases (CDKs) these molecules are critical in maintaining cell cycle progression in RGCs (Tsunekawa et al., 2012), however they are known to promote neurogenic divisions as well (Pauklin and Vallier, 2013). Interestingly, *Ccnd1/2* are also major downstream targets for multiple morphogen pathways that modulate neurogenesis in the cortex, including, Notch, STAT5, Shh, and the Wnt signalling pathway (Cohen et al., 2010; Kalita et al., 2013; Shahi et al., 2010; Shtutman et al., 1999).

Nevertheless, although the pseudo-bulk approach to DE analysis is considered less statistically biased than conventional single cell DE techniques, statistical bias can still ensue. It was found that when re-examining the expression of the DE genes from the pseudo-bulk at the single cell level, *Pianp, Tppp3,* and *Sulf1* were only expressed in very few cells within the RGC, IP, and neuron populations. Therefore, it is likely that these genes were considered differentially expressed at the population level only because of high cell-cell variability which became exacerbated after aggregating the averaged gene counts to perform pseudo-bulk analysis. Interestingly, the only DE gene found in the pseudo-bulk that was widely expressed at the single cell level was *Ccnd2*. Moreover, this gene was also noticeably decreased in the HETs within the RGC and neuron clusters, however levels were relatively normal in the IPs. Taken together, the statistical limitations of single cell DE methods were considered in this chapter and pseudo-bulk was initially used to explore the data via an alternative avenue. As it has been demonstrated from this chapter, each method presents its own variabilities when identifying DE genes, and thus any findings from these analyses must be confirmed via another technique such as quantitative PCR.

5.3.3. Ribosomal genes were significantly downregulated in HETs in all cell types

One of the most striking findings of the single cell DE analysis between WTs, HETs, and KOs was that many of the most DE genes from the RGC, IP, and neuron comparisons were related to ribosomes and histones. Interestingly, it was consistently found that these genes were mostly downregulated in the HETs, whereas expression levels were relatively normal between WTs and KOs. Indeed, when examining the total fraction of ribosomal RNAs (rRNAs) between the groups, HETs showed a noticeable decrease in ribosomal gene expression (~10%), and ~37% of all DE genes from the HETs were rRNAs or related to ribosomes. Overall, these initial results suggested that ribosomal biogenesis, function, and translation maybe significantly reduced in HETs.

Ribosomes are supramolecular ribonucleoprotein complexes that are integral in the translation of mRNAs into protein. Eukaryotic ribosomes are typically formed from two subunits: the small 40S subunit and the large 60S subunit. The small 40S subunit is composed of the 18S rRNA and 33 ribosomal proteins (RPs) whereas the 60S is composed of 5S, 5.8S, and 28S rRNAs and 47 RPs (Chaillou et al., 2014). The rRNAs 18S, 28S, and 5.8S are transcribed in the nucleolus by RNA polymerase I, however 5S rRNA is transcribed in the nucleoplasm as well as most of the RPs. RPs are then translated in the cytoplasm but are translocated back into the nucleolus to assemble the 40S and 60S subunits (Teng et al., 2013). Ribosomal biogenesis has been shown to be critical for regulating the synthesis of proteins during cell cycling and cell proliferation. For example, deletion of *Rps6* and *Rpl26* inhibits cells from re-entering the cell cycle due to failure of translating genes that would promote further proliferation (Kirn-Safran et al., 2007; Stewart and Denell, 1993). Loss of function mutations in ribosomal genes have also been associated with various “ribosomopathies” such as Diamond Blackfan Anemia (DBA) which is characterised by congenital macrocytic anemia and malformations in brain development and the development of other organs (Teng et al., 2013). Likewise, upregulation of ribosomal biogenesis has been associated with proliferation of cancer cells (Montanaro et al., 2012). Therefore, it is possible mosaic expression of *Pcdh19* may cause dysregulation of ribosomal pathways which alters the proliferative and neurogenic behaviour of the progenitors in HETs. However, the possible mechanism underlying this is unknown.

5.3.4. Various Wnt-related genes were dysregulated in HETs and KOs

DE analysis was also performed by comparing mutant groups to their WT same sex counterpart in order to control for sex and uncover DE genes driven by genotype. To infer whether the DE genes were overrepresented in specific biological pathways, GO term analysis was performed

using the online resource ShinyGO. Interestingly, DE genes related to Wnt signalling were overrepresented in HET and KO comparisons within the RGC cluster. Further examination of the Wnt genes illustrated that many genes, such as *Ddx3x, Rspo3, Wnt8b, Gli3, Sall1, Shisa2,* and *Hmga2* were all upregulated in the HETs. Conversely, *Cncd2* and *Lhx2* were downregulated in the KOs.

Many of these genes have been shown to modulate Wnt signalling in various ways. For example, *Sall1* has been demonstrated to promote proliferation in RGCs and neurogenesis in IPs, potentially by interacting with β-catenin and regulating transcription of Wnt target genes (Harrison et al., 2012; Sato et al., 2004). Moreover, SHISA2 has been found to inhibit the glycosylation of FZD3 and to reduce its presence on the membrane, thus lowering the responsiveness of cells to secreted Wnt proteins (Pascual-Vargas and Salinas, 2021). *Hmg2a* is also a well characterised transcription factor (TF) that is necessary for the neuroepithelial (NE)-to-RGC transition (Kuwayama et al., 2021) and has been shown to be a Wnt attenuator by binding to *Gata6* and modulating expression levels of *Fzd2* (Singh et al., 2014). Interestingly, *Gli3* is a known TF that acts downstream of Shh signalling and is necessary for maintaining progenitor proliferation (Wang et al., 2011). However, it has also been shown to exhibit a convergent role between Shh and Wnt signalling in the cortex (Ulloa et al., 2007). Therefore, the upregulation of these genes in the HET RGCs may suggest that Wnt signalling and potentially Shh signalling are dysregulated in these animals. Interestingly, *Rspo3* was also shown to be upregulated in the HETs from the bulk RNA-seq. Moreover, the downregulation of key Wnt components in the KO, including the Wnt TF *Lhx2* and the downstream target *Ccnd2*, suggests that canonical Wnt signalling may be slightly altered in the KOs. However, it is unclear how downregulation of canonical Wnt signalling does not affect progenitor behaviour in these animals (unpublished data, 2019). Likewise, if Wnt signalling is upregulated in the HETs as a whole, it is not clear how HET WT and HET KO progenitors undergo neurogenesis at different rates.

Another DE gene shown to be upregulated in the HETs was *Ddx3x*, which is an X-linked RNA helicase that has a significant role in transcription, splicing, RNA transport, and translation (Snijders Blok et al., 2015). *Ddx3x* also has a role in modulating Wnt β-catenin signalling, whereby it binds to CK1 and stimulates its kinase activity to phosphorylate dishevelled, and is necessary for progenitor proliferation (Cruciat et al., 2013). Interestingly, recent publications have demonstrated that DDX3X plays an important role in ribosome function in neural progenitors and is integral to the proper translation of genes that control proliferation which may be Wnt-dependent (Hoye et al., 2022). Strikingly, it was also found that loss of *Ddx3x* increases progenitor cell cycling

which reduces proliferation, but only in females – since the Y-chromosome paralog, *Ddx3y*, compensates for DDX3X perturbations in the developing male cortex. Therefore, *Ddx3x* may play an important role in the perturbed neurogenesis phenotype observed only in HET females and may partially explain why KO males do not exhibit altered neurogenesis.

5.3.5. HET WT and HET KO RGCs show no transcriptional differences

The most surprising result from this chapter was that only the cell genotype markers *EGFP, Pcdh19,* and *β-gal* were shown to be dysregulated when comparing the transcriptional profiles of HET WT and HET KO RGCs. This was especially unexpected since it is these cells that exhibit perturbed neurogenesis – because WT and KO RGCs display reduced and increased neurogenesis, respectively, it was theorised that the transcriptional profiles of these cells would be significantly different. However, this analysis illustrated that, at least at the transcriptional level, HET WT and HET KO cells are very similar.

One possible explanation for this finding is that the potential transcriptional changes could not be identified using 10X chromium. As aforementioned, different scRNA-seq approaches offer several advantages and disadvantages (Ding et al., 2020; Haque et al., 2017). The main advantage offered by 10X Chromium is that by using high-throughput emulsion-based cell partitioning, thousands of cells can be captured and sequenced, however at the cost of sequencing depth. Likewise, low-throughput approaches can capture only a few hundred cells but provide significantly greater sequencing depth. Therefore, it is possible that the transcriptional differences between HET WT and HET KO RGCs are more subtle than previously hypothesised and could only be detected using a technique that provides greater sequencing depth, such as SMART-seq2 (Ding et al., 2020). DE analysis was also performed by isolating HET WT and HET KO RGCs and comparing to WT female cells which we considered as a “normal” transcriptional background. Only several DE genes were shown to be distinctly expressed in the HET cells. The proliferative genes *Foxp1, Id4,* and *Foxg1* were shown to be upregulated in the HET WT cells (Hanashima et al., 2002; Pearson et al., 2020; Yun et al., 2004), whereas the pro-neurogenic gene *Neurog2* was downregulated (Lacomme et al., 2012). Transcriptionally, this fits the notion that the HET WT RGCs are remaining in a proliferative state longer than HET KO RGCs. Interestingly, *Foxg1* was also upregulated in HET KO cells however to a lesser extent than in HET WT cells. Taken together, the HET WT and HET KO RGCs possibly do exhibit transcriptional differences that may underlie the different neurogenic behaviours exhibited by these cells, however more sensitive techniques may be needed to unravel this further.

Nevertheless, it is also possible that the perturbed neurogenesis phenotype observed between the WT and KO progenitors within the HETs may not be driven by changes of transcription. For example, cell fate decisions can be made not just at the transcriptional level but through other biological processes. It was recently shown that the regulation of post-translational modifications of cyclin proteins during mitosis is critical for modulating RGC and IP cell fate (Da Silva et al., 2021). In this work, Da Silva et al. (2021) demonstrated that inhibition of GSK3-β – which is traditionally associated with the stablisation of β-catenin – also stablises additional downstream proteins such as SOX4 and SOX11, which promotes proliferative and neurogenic divisions in RGCs and IPs. Importantly, because this cascade occurs during mitosis – when gene transcription is markedly reduced – the cell fate decisions are made by changes in protein signalling that are independent of gene transcription. Although inferences cannot be made on possible changes of post-translational modifications between WT, HET, and KO animals in this experiment, it is worth noting that ribosomal genes were the most consistently downregulated genes within the HETs, suggesting that protein translation and modification may be a critical pathway underlying the neurogenesis phenotype in these animals.

5.3.6. Conclusion

In this chapter, DE analysis at the single cell level illustrated numerous DE genes between WT, HET, and KO animals within RGC, IP, and neuronal clusters. Although alternative statistical methods were explored to account for the inherent statistical bias of single cell DE methods, the alternative method also showed problems with genes that exhibited high cell-cell variability. Nevertheless, within the RGC comparisons, many DE genes were uncovered, and it was especially shown that ribosome biogenesis and translation may be compromised in the HETs which may cause cell cycle deficiencies, leading to altered neurogenesis. Moreover, many genes related to Wnt signalling were also shown to be upregulated in the HETs and downregulated in the KOs, suggesting that Wnt signalling may be a key player in understanding the role of *Pcdh19* in neurogenesis. This finding also suggests that Wnt may act as the driver of the perturbed neurogenesis phenotype in the HETs and maybe a compensator in the KOs. Interestingly, HET WT and HET KO RGCs showed few DE genes when compared to each other, and DE genes were only found when comparing the cells separately to WT female cells. Therefore, the potential transcriptional differences would need to be examined using more sensitive scRNA-seq techniques. Alternatively, evidence also indicates that it is possible that other biological pathways independent of transcription, such as protein modification may be dysregulated in the HETs and KOs, which may have a greater role in the underlying neurogenic phenotypes of these animals.

Chapter 6: Exploring the role of Procadherin-19 in Wnt signalling and examining differences in Wnt signalling in the developing cortex of WT, HET, and KO embryos

6.1. Introduction

As previously stated, the main goal of this thesis was to decipher the role of PCDH19 during early neurogenesis events and to uncover potential molecular avenues that may underpin this role. To this end, two RNA-seq experiments were conducted, wherein the expression landscape of the cortex of *Pcdh19* wild type (WT), heterozygous (HET), and knockout (KO) E11.5 embryos were examined using bulk and single cell (sc)RNA-seq. As detailed in the previous chapters, the Wnt signalling pathway was a re-occurring hit in both experiments. Gene set enrichment analysis (GSEA) from the bulk RNA-seq found Wnt signalling as a significantly dysregulated pathway. Moreover, multiple Wnt-related genes were differentially expressed in HETs and KOs when compared to WTs. Therefore, these results suggest that the Wnt signalling pathway is a viable candidate to explore the molecular mechanisms underlying PCDH19 involvement in neurogenesis.

Wnt signalling is a highly conserved pathway that is integral to the proper development of the cortex. As mentioned in chapter 1, the overall role of Wnt signalling in the cortex is complex. During early development, Wnt signalling is necessary for the establishment of the anterior/posterior axis of the telencephalon (Kiecker and Niehrs, 2001). Then, after neural tube closure, Wnt morphogens are secreted from the developing telencephalon which pattern adjacent radial glial cells (RGCs) to give rise to the hippocampus (Backman et al., 2005). Likewise, sonic hedgehog (Shh) morphogens are secreted from the ventral part of the telencephalon, and it is believed that the interaction of Wnt and Shh is necessary for the expression of transcription factors that mediate the development of the cortical primordium (Mallamaci and Stoykova, 2006). At later developmental stages, Wnt signalling has also been shown to regulate the switch of RGCs from proliferative to neurogenic divisions. During the expansion phase, canonical Wnt signalling (β-catenin dependent) is believed to exert a pro-proliferative effect on RGCs by driving expression of cyclin genes (Chenn and Walsh, 2002). However, canonical Wnt signalling has also been shown to promote neurogenesis in RGC-derived intermediate progenitors (IPs) (Munji et al., 2011) and promote neurogenesis in neural progenitor cells *in vitro* by driving transcription of *N-myc* (Hirabayashi et al., 2004). Moreover, various non-canonical (β-catenin independent) Wnt

pathways have been found to be important for neurogenic events in RGCs. For example, the Wnt/Ryk pathway has been shown to drive neurogenesis in RGCs via Wnt morphogens binding to the WIF domains of RYK (Lyu et al., 2008). Moreover, the Wnt-mediated phosphorylation of LRP6 and the inhibition of GSK3-β – which is the key mechanism underling β-catenin stablisation – has also been shown to stabilise other phosphor-targets of GSK3-β, such as *Ccny* and *Ccnyl1*, and is necessary for driving neurogenic divisions during mitosis (Da Silva et al., 2021). Taken together, Wnt signalling has a diverse role in cortical development and can exert both self-renewal and neurogenic effects on progenitor cells.

Due to the complex role of Wnt signalling in cortical development, there arise several challenges in accurately measuring this pathway. For instance, although the Wnt pathways are loosely categorised as β-catenin-dependent or β-catenin-independent, many of them, regardless of β-catenin, share other molecules (Harrison-Uy and Pleasure, 2012). For example, cyclin kinase 1 (CK1) kinase is important for phosphorylating LRP6 and recruiting the GSK3-β destruction complex to the membrane, thus stabilising β-catenin (Niehrs and Shen, 2010). However, phosphorylation of LRP6 and the inhibition of GSK3-β are also important for engaging the Wnt stablisation of proteins (WNT/STOP) pathway which is independent of β-catenin (Da Silva et al., 2021). Moreover, the recruitment of Dishevelled (DVL) to the Frizzled (Fzd)-LRP complex after Wnt receptor binding is involved in the canonical, Wnt/PCP, and Wnt/Ca^2 pathways (Delaunay et al., 2014; Shu et al., 2018). Therefore, many of the Wnt pathways operate using overlapping signalling molecules and thus, the epistatic level at which Wnt signalling is investigated is important to know to examine the specific pathway of interest.

One way that has been used to measure broad Wnt signalling is by using phosphor-antibodies for pLRP6 (Niehrs and Shen, 2010). The intracellular domain of LRP6 contains five PP[S/T]P (PPSP) and juxtaposed CK1 motifs (Davidson et al., 2005; Zeng et al., 2005). Antibodies have been generated that monitor the phosphorylation activity across the multiple PPSP sites of LRP6, including Thr1479, Ser1490, and Thr1493. Interestingly, research with these antibodies has illustrated that some PPSP sites are phosphorylated in a Wnt-dependent manner, but others can be phosphorylated constitutively i.e., without Wnt (MacDonald et al., 2008; Wolf et al., 2008). For example, Thr1479 and Thr1493 are phosphorylated by CK1γ and are strictly Wnt-dependent (Davidson et al., 2005; Zeng et al., 2005). However, phosphorylation at Ser1490 can occur dependently or independently of Wnt, depending on the kinase (Davidson et al., 2005; Wan et al., 2008). Importantly, phosphorylation at the Ser1490 residue can occur without Thr1479 phosphorylation and in this case does not engage internal Wnt signalling, but this state has been

considered as a “priming” of LRP6 for incoming signalling (Niehrs and Shen, 2010). The Thr1479 and Ser1490 antibodies have been used to measure cortical Wnt signalling *in vivo* in the form of fluorescence intensity (FI) measurements (Da Silva et al., 2021). Although Thr1479 is technically more specific for measuring active Wnt signalling, both antibodies have been used to detect changes in Wnt signalling activity. Overall, examining pLRP6 activity can be considered as a measurement of Wnt signalling that is “upstream” of the pathway.

Another way that has been employed to measure Wnt signalling *in vivo* has been to use reporter animals. The Wnt reporter *TCF/Lef:H2B-GFP* line is characterised by possessing six copies of a *TCF/Lef1* transcription factor (TF) binding site upstream of a minimal promoter which drives expression of a *H2B-GFP* transgene, wherein GFP is fused with a partial Histone-1 sequence that translocates the GFP to the nucleus (Ferrer-Vaquer et al., 2010). Therefore, Wnt-active cells exhibit a clear nuclear GFP signal. Importantly, because the *H2B-GFP* expression is downstream of TCF/LEF activation – which depends on nuclear β-catenin – the FI of the H2B-GFP^{+} cells could be considered as a proxy measurement of downstream transcriptional activity of canonical Wnt signalling. Although this reporter line has been used to study the effects of Wnt signalling during broader embryonic development (Ferrer-Vaquer et al., 2010) as well as to investigate the role of Wnt signalling in interneuron development (McKenzie et al., 2019), it has yet to be used as a method to measure Wnt signalling in the cortex.

Evidence also suggests that δ-protocadherins, including Pcdh19, may regulate Wnt signalling via the noncanonical Wnt/Ryk pathway (Biswas et al., 2021). In this publication, Biswas et al. (2021) demonstrated that the zebrafish homolog Ryk, a Wnt receptor that is proteolytically cleaved upon Wnt binding, can form a complex with Pcdh19 and other δ-protocadherins. Moreover, this interaction was shown to be important to regulating self-renewal/neurogenic divisions of neural progenitors in the developing hindbrain of zebrafish, and knockdown (KD) of the δ-protocadherins caused perturbed Wnt/β-catenin signalling and dysregulated expression of Wnt gene targets (Biswas et al., 2021). However, it has not been shown whether mammalian RYK and PCDH19 are able to interact. Additionally, it could also be suggested that PCDH19 could modulate Wnt signalling by regulating β-catenin. For example, N-cadherin – a known binding partner of PCDH19 – can interact directly with β-catenin and modulate its levels in the cytoplasm (Biswas et al., 2010; Linask et al., 1997; Sakane and Miyamoto, 2013). Although PCDH19 does not contain the binding site to interact with β-catenin directly, it is possible that PCDH19 may indirectly interact with β-catenin through N-cadherin. Nevertheless, whether this is important for N-cadherin-mediated regulation of Wnt signalling is unknown.

6.1.1. Aim

This chapter aims to investigate the role of PCDH19 in Wnt signalling during cortical development. Biochemical assays were performed *in vitro* to investigate the role of PCDH19 in Wnt signalling at the molecular level. Co-immunoprecipitation (CoIP) assays were used to determine whether PCDH19 interacts with β-catenin as well as the mammalian RYK. Moreover, TOP-FLASH luciferase assays were also employed to determine whether the presence of PCDH19 enhances or attenuates Wnt signalling *in vitro*. To examine canonical Wnt signalling *in vivo*, *Pcdh19* mutant mice were crossed with the Wnt reporter mice *TCF/LEF:H2B-GFP*. Immunohistochemical (IHC) analysis was performed by co-staining these animals using PAX6 antibodies and taking FI measurements of the H2B-GFP from $PAX6^{+}$ RGCs. A preliminary characterisation of the H2B-GFP signal in the cortex was also carried out. IHC was also performed on WT, HET, and KO E11.5 embryos using phosphor-antibodies for pLRP6 (Thr1479) and pLRP6 (Ser1490). FI measurements were then quantified to determine whether upstream phosphor activity of LRP6 was different between WT, HET, and KO cells.

6.2. Results

6.2.1. Co-immunoprecipitation of PCDH19 and Wnt-related signalling proteins

It is possible that PCDH19 may modulate Wnt signalling by forming protein complexes with known Wnt regulators/signalling molecules, for example RYK or β-catenin. Although Pcdh19 has been shown to interact with Ryk in zebrafish, this has not been shown in mammals. Likewise, it is possible that PCDH19 may interact with β-catenin via a common binding partner – N-cadherin. Therefore, to determine whether PCDH19 interacts with these proteins, CoIP was performed *in vitro*. It was first investigated whether the mammalian proteins PCDH19 and RYK would show an interaction (see Figure 6.1A). HEK293 cells were transfected with either *Pcdh19-GFP*, *Ryk-HA,* or both plasmids and the precipitation was achieved using the commercially available GFP-Trap. INPUT and IP samples were then blotted using GFP antibody to detect the targeted PCDH19-GFP protein and HA antibody to detect RYK-HA. On the GFP IP blot, it was observed that PCDH19-GFP was successfully captured in the single transfection and co-transfection condition. Moreover, no protein band was observed in the RYK-HA single transfection condition, illustrating that the GFP-Trap would only pull down GFP-tagged protein products. On the HA IP blot, no protein bands were observed from the PCDH19-GFP and RYK-HA single transfection conditions, however a strong RYK-HA band was observed from the co-transfection condition. This indicates that PCDH19-GFP and RYK-HA likely interact *in vitro*.

Next, it was decided to investigate whether PCDH19 interacts with β-catenin. N-cadherin was used as a positive control for the CoIP (Emond et al., 2011; Linask et al., 1997). HEK293 cells were transfected with either *Pcdh19-GFP*, *Ctnnb1-HA*, *N-cadherin-GFP*, *Pcdh19-GFP* + *Ctnnb1-HA*, or *N-cadherin-GFP* + *Ctnnb1-HA* (see Figure 6.1B). On the GFP IP blot, PCDH19-GFP and N-cadherin-GFP were successfully captured using the GFP-Trap. Moreover, no levels of β-catenin-HA were detected in the single transfection condition. Likewise, on the HA IP blot, no levels of β-catenin-HA were detected in the single transfection condition, further confirming the validity of the pulldown using GFP-Trap. Interestingly, β-catenin-HA was detected when co-transfected with PCDH19-GFP or N-cadherin-GFP. The β-catenin band was much stronger in the N-cadherin-GFP co-transfection condition than in the PCDH19-GFP co-transfection condition, which indicates the interactions may be direct and indirect, respectively.

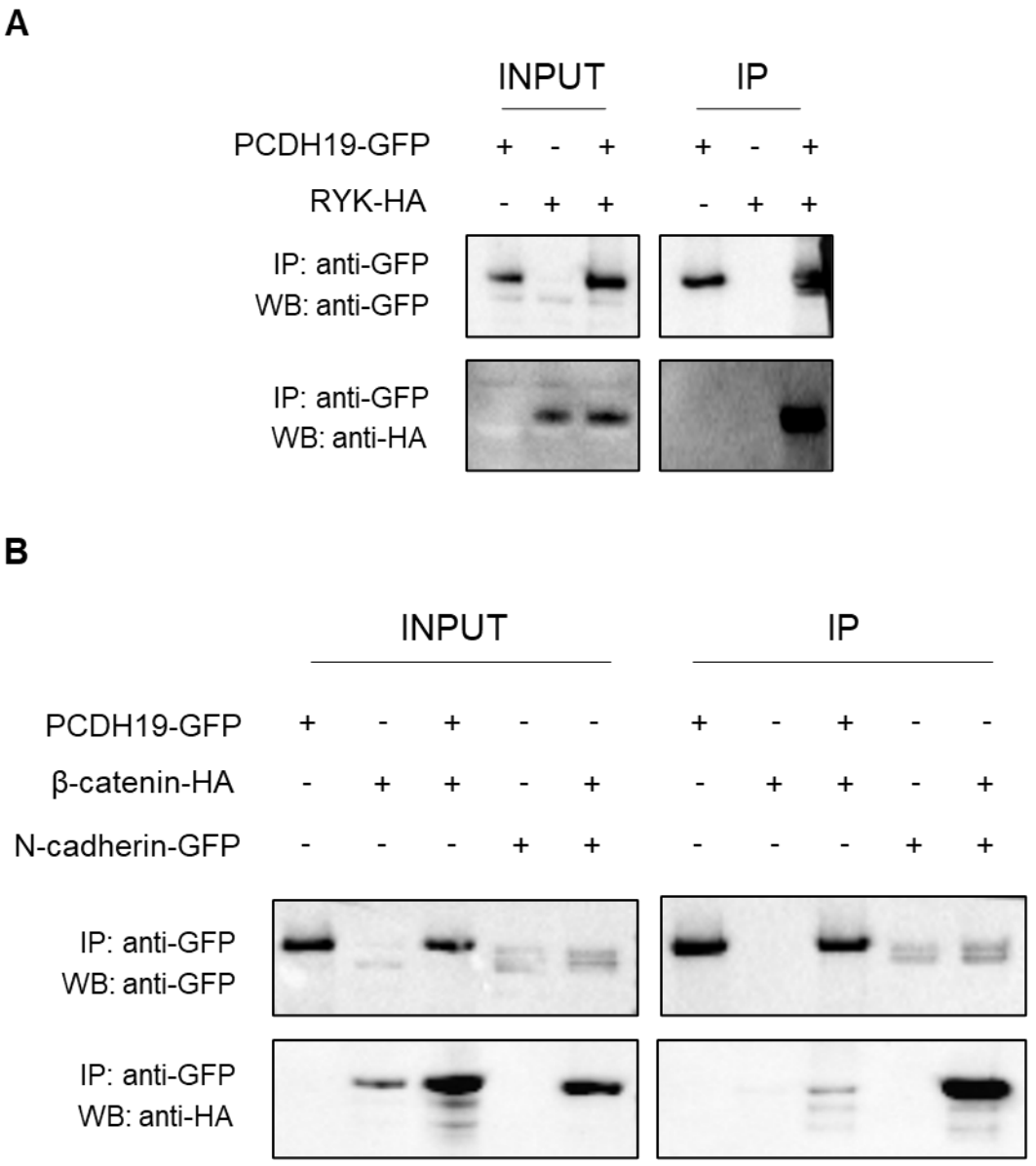


Figure 6.1. Co-immunoprecipitation results showing interactions of PCDH19 with Wnt signalling proteins. (A) CoIP of PCDH19-GFP and RYK-HA. HEK293 cells were transfected with either *Pcdh19-GFP* or *Ryk-HA* or co-transfected. (B) CoIP of PCDH19-GFP and β-catenin-HA in addition to N-cadherin-GFP as a positive control. HEK293 cells were transfected with *Pcdh19-GFP, Ctnnb1-HA, N-cadherin-GFP, Pcdh19-GFP* + *Ctnnb1-HA*, or *N-cadherin-GFP* + *Ctnnb1-HA*. Plus (+) and minus (-) symbols represent presence or absence of protein product, respectively. Overexpression of *PCDH19* and *Ctnnb1* together hints at a possible protein-protein interaction. INPUT shows relative levels of the protein product in the lysates before immunoprecipitation. IP shows protein levels after capture.

6.2.2. Analysis of PCDH19 and its involvement in Wnt signalling using luciferase assays

The results of the previous section indicated that PCDH19 may interact with β-catenin. Since N-cadherin can directly interact with PCDH19 and β-catenin, it is possible that N-cadherin may mediate an indirect interaction between these two proteins (Biswas et al., 2010; Linask et al., 1997). However, the consequences of this interaction and its effect on Wnt signalling requires further exploration. To that end, TOP-FLASH luciferase assays were performed *in vitro* to examine the effects of PCDH19 on downstream Wnt signalling (see Figure 6.2 and Figure 6.3).

Firstly, the luciferase assay was optimised before conducting experiments with PCDH19. To determine whether the luciferase assay was responsive to Wnt activity, HEK293 cells were transfected with *pCBA* empty vector and the *TOP-FLASH* luciferase reporter plasmid. Cells were then treated with either 10 ng, 25 ng, or 50 ng human recombinant WNT3A for 24 hours (see Figure 6.2A). A one-way ANOVA analysis on the relative luciferase units (RLU) showed a significant main effect ($F(3, 8) = 13.076$, $p = 0.002$). Bonferroni post-hoc analysis further revealed that cells treated with 50 ng WNT3A showed significantly higher RLUs than cells treated with 25 ng ($p = 0.03$), cells treated with 10 ng ($p = 0.005$), and cells that were not treated ($p = 0.001$). No other comparisons were significant. This demonstrated that the luciferase assay was responsive to WNT3A treatment. Moreover, this also showed that treatment of transfected HEK293 cells with 50 ng WNT3A was sufficient to engage a measurable upregulation of the *TOP-FLASH* reporter expression. Therefore, for subsequent experiments that included WNT treatment, it was decided to treat the cells with 50 ng WNT3A.

Next, it was decided to examine the validity of the *TOP-FLASH* reporter by performing the luciferase assay with the *FOP-FLASH* negative control reporter – which harbours mutated sequences upstream of the minimal promoter – thus preventing transcription factors from binding to it. Cells were transfected with the *pCBA* empty vector and either the *TOP-FLASH* or *FOP-FLASH* reporter. As a positive control for Wnt signalling, *Ctnnb1* was also transfected with either *TOP-FLASH* or *FOP-FLASH*. It was theorised that β-catenin would stimulate *TOP-FLASH* expression, however the mutated sequences within the *FOP-FLASH* reporter would block β-catenin activity (see Figure 6.2B). A two-way ANOVA analysis revealed a main effect of group ($F(3,24) = 36.594$, $p = 0.05$) wherein TOP β-catenin showed a greater luciferase response compared to TOP CBA ($p < 0.001$), FOP CBA ($p < 0.001$), and FOP β-catenin ($p < 0.001$). No other post hoc comparisons were significant. Moreover, no main effect of treatment or an interaction between group and treatment was found. Overall, these results suggested that the *TOP-FLASH* reporter is an effective measure of *in vitro* Wnt/β-catenin activity.

To further investigate whether PCDH19 affected Wnt signalling, it was decided to use *Pcdh19, N-cadherin,* and *Ctnnb1* plasmids in the TOP-FLASH assay. N-cadherin is a known negative regulator of Wnt signalling and therefore is a good candidate to use as a negative control. Likewise, β-catenin would be used as a positive control for the internal activation of the pathway. It was predicted that N-cadherin overexpression would reduce TOP-FLASH activity, whereas β-catenin overexpression would enhance the activity. Moreover, to determine whether presence of PCDH19 or N-cadherin would affect β-catenin signalling activity, these plasmids were also co-transfected separately with the *Ctnnb1* plasmid. Finally, cells were treated with and without WNT3A to determine whether the effect of PCDH19 on the Wnt pathway may be dependent on the presence of WNT protein (see Figure 6.3A)

As predicted, it was noticed that β-catenin caused a substantial increase in luciferase activity (see Figure 6.3A). Moreover, the presence of N-cadherin alone did not stimulate luciferase activity, however N-cadherin was shown to decrease the luciferase response caused by β-catenin when *N-cadherin* was co-transfected with *Ctnnb1*. Interestingly, PCDH19 alone did not change luciferase levels, however in the presence of β-catenin, PCDH19 also seemed to also reduce β-catenin luciferase activity. Statistical analysis was not performed on these conditions since it was found that conditions involving N-cadherin seriously violated the ANOVA assumptions. To increase the validity of the statistical test, conditions involving N-cadherin were dropped from the analysis (see Figure 6.3B). A two-way repeated ANOVA was performed on the remaining groups. A significant main effect of group was found ($F(3,32) = 61.295, p = 0.05$) and post hoc analysis further revealed β-catenin caused a greater luciferase response than CBA ($p < 0.001$) and PCDH19 ($p < 0.001$). Moreover, PCDH19 + β-catenin had significantly lower luciferase activity compared to β-catenin alone ($p < 0.001$). No other comparisons were significant.

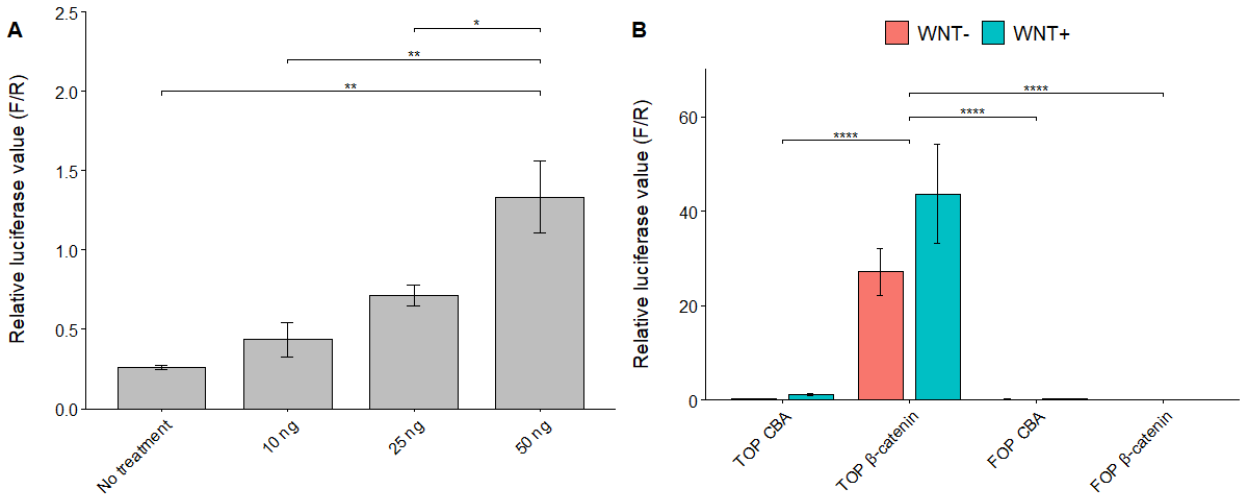


Figure 6.2. Optimization results of luciferase assays. (A) Bar graph depicting the RLU values of cells transfected with *pCBA* and the *TOP-FLASH* luciferase reporter. Cells were also treated with 10 ng, 25 ng, or 50 ng recombinant WNT3A in order to measure the responsiveness of the *TOP-FLASH* reporter to Wnt stimulation. Cells treated with 50 ng WNT3A showed a significant increase in luciferase activity compared to treatments with 25 ng ($p = 0.03$), 10 ng ($p = 0.005$), and cells with no treatment ($p = 0.001$). Each condition n = 3 (B) Bar graph depicting the RLU values of cells transfected with either *TOP-FLASH* or *FOP-FLASH* and *pCBA* or *Ctnnb1*. *FOP-FLASH* is the negative control to *TOP-FLASH*. Cells transfected with *Ctnnb1* and *TOP-FLASH* showed greater luciferase activity than cells transfected with *TOP-FLASH CBA* ($p < 0.001$), *FOP-FLASH CBA* ($p < 0.001$), and *FOP-FLASH Ctnnb1* ($p < 0.001$). Each condition n = 4. Data is represented as mean + standard error. * = $p < 0.05$. ** = $p < 0.01$. *** = $p < 0.001$. F = Firefly. R = Renilla.

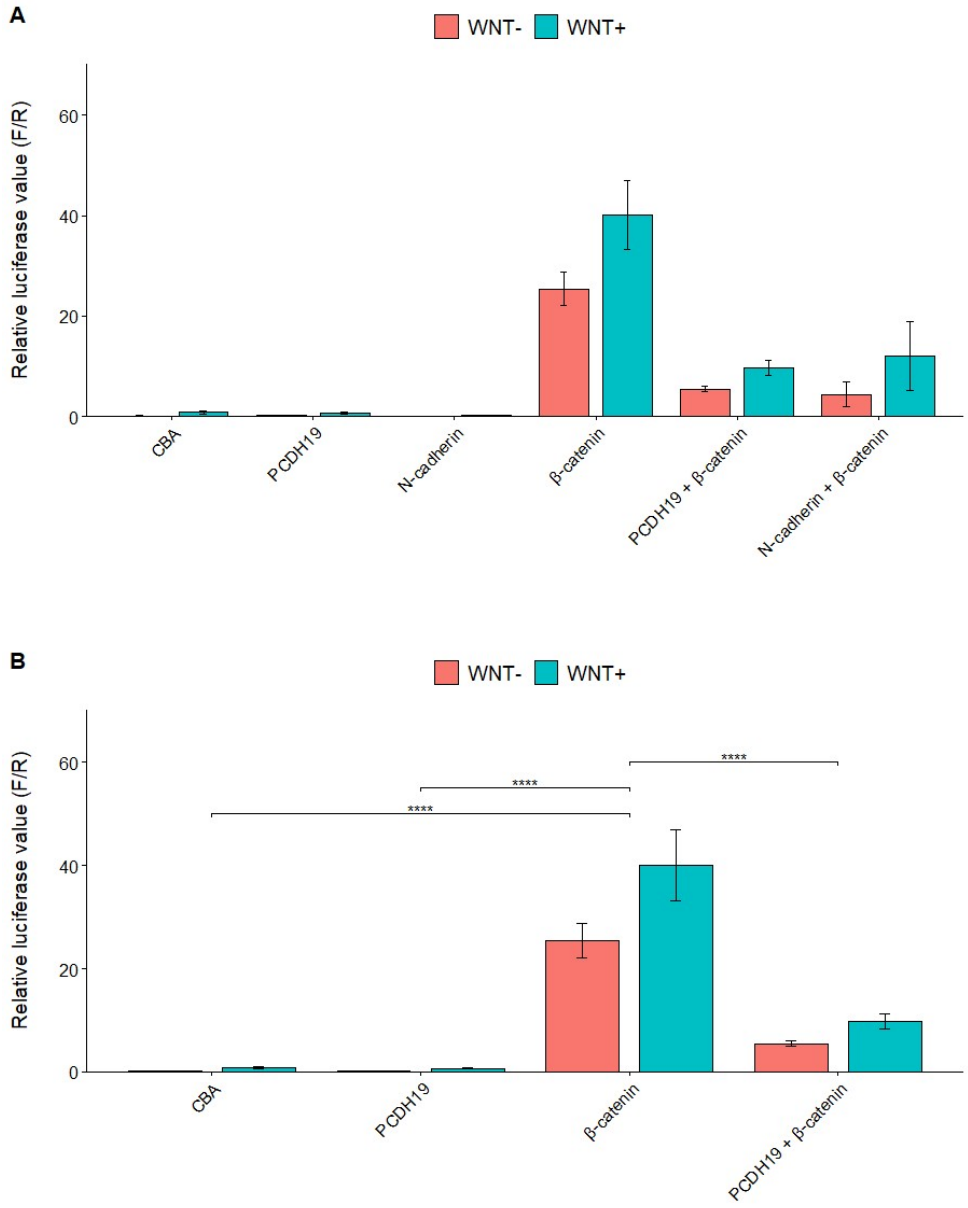


Figure 6.3. Luciferase assay results illustrating effect of PCDH19, N-cadherin, and β-catenin on downstream Wnt signalling. (A) Bar graph illustrating the RLU values of TOP-FLASH luciferase assay after overexpressing *Pcdh19*, *N-cadherin*, *Ctnnb1*, *Pcdh19* + *Ctnnb1*, and *N-cadherin* + *Ctnnb1*, with or without WNT3A treatment. (B) The same data as before however with N-cadherin conditions removed to increase statistical power. Cells transfected with *Ctnnb1* showed greater luciferase activity compared to cells transfected with *CBA* ($p < 0.001$), *PCDH19* ($p < 0.001$), and *PCDH19* + *Ctnnb1* ($p < 0.001$). Data is represented as mean + standard error. * = $p < 0.05$. ** = $p < 0.01$. F = Firefly. R = Renilla. Each condition n = 5.

6.2.3. Expression of the TCF/Lef:H2B-GFP reporter gene in RGCs and neurons

Although recent publications have shown that the *TCF/Lef:H2B-GFP* reporter gene is expressed in the medial ganglionic eminences (MGEs) (McKenzie et al., 2019), to date no publications have used these mice for studies in the developing cortex. Therefore, to demonstrate that the reporter gene is expressed in RGCs at E11.5, *Pcdh19* WT *TCF/Lef:H2B-GFP* embryos were stained with PAX6 antibody and counterstained with DAPI to label the nuclei (see Figure 6.4A). Sections were also stained with GFP antibody to enhance the endogenous signal from the H2B-GFP. Examination of the ventricular zone (VZ) showed that many $PAX6^+$ cells were also $H2B\text{-}GFP^+$, demonstrating that the Wnt reporter gene is expressed in the cortex and that Wnt signalling occurs at a sufficient level in the cortex to activate the reporter. Interestingly, it was observed that other cells above the VZ were also $H2B\text{-}GFP^+$, indicating that IPs and neurons may also express the transgene and may be subjected to Wnt signalling. To investigate this further, sections were stained with a specific IP marker, TBR2 antibody, and a specific neuronal marker, TBR1 antibody, and counterstained with DAPI (see Figure 6.4B). Interestingly, this showed that many $TBR2^+$ and $TBR1^+$ cells were also $H2B\text{-}GFP^+$, further demonstrating that the *TCF/Lef:H2B-GFP* transgene is active in RGCs, IPs, and neurons.

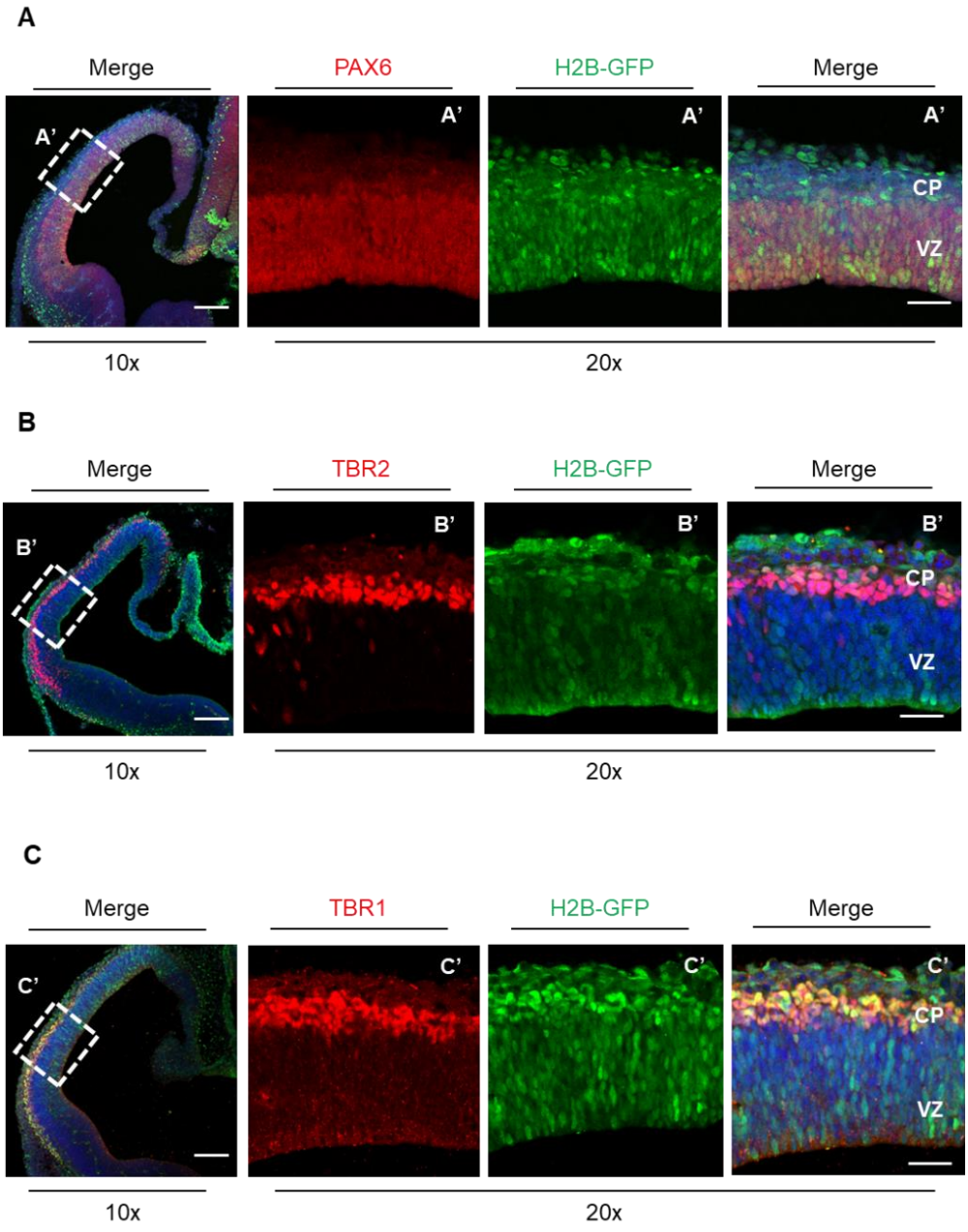


Figure 6.4. IHC stainings showing H2B-GFP expression localised to RGCs (PAX6$^+$), IPs (TBR2$^+$), and new-born neurons (TBR1$^+$). (A) Representative image of E11 *TCF/Lef:H2B-GFP* WT embryo section stained for PAX6 (red), also showing the H2B-GFP signal (green) and counterstained with DAPI (blue). (B) Representative image of E11 *TCF/Lef:H2B-GFP* WT embryo section stained for TBR2 (red). (C) Representative image of E11 *TCF/Lef:H2B-GFP* WT embryo section stained for TBR1 (red). Whole hemisphere images were taken at 10x, and high magnification images were taken at 20x (A', B', C'). CP = Cortical plate. VZ = ventricular zone. Scale bar = 200 µm (10x) and 50 µm (20x).

6.2.4. Spatial expression of the TCF/Lef:H2B-GFP reporter gene: a system for rostro-caudal and medio-lateral analysis

In the previous experiment, it was noticed that the *TCF/Lef:H2B-GFP* expression within the VZ was most intense at the medial region and gradually decreased laterally across the cortex. Moreover, expression was also relatively weak in rostral regions of the brain with intensity increasing more caudally. Therefore, a system was designed to accurately study H2B-GFP levels in the VZ of the cortex in a rostro-caudal, medio-lateral manner (see Figure 6.5A). First, it was decided to analyse three sections – a rostral, medial, and caudal section – for each embryo in each genotype group. Secondly, three areas of pre-defined size were selected spanning the radial thickness of the cortex (see Figure 6.5B). This means measurements would be taken from medial (1), middle (2), and lateral (3) areas of the cortex for each section. Finally, the nucleus of most of the $PAX6^{+}$ cells in each section was drawn around using the free-hand tool in ImageJ and FI measurements of the H2B-GFP signal were extracted. To ensure that FI measurements from the H2B-GFP signal accurately represented active Wnt signalling in RGCs as much as possible, it was decided not to use a GFP antibody to enhance the GFP signal for the analysis. This was because it was theorised that the GFP antibody may produce an oversaturated FI signal and thus any potential differences in Wnt signalling may be masked. Instead, FI measurements were taken using the endogenous H2B-GFP signal (see Figure 6.6A).

To determine whether there was a significant difference in H2B-GFP intensity across the cortex (medial, middle, and lateral) and along the anterior-posterior sections of cortex (rostral, medial, caudal), FI measurements were taken from *Pcdh19* WT *TCF/Lef:H2B-GFP* animals (males = 1, females = 3, total n = 4) following the previously mentioned experimental plan (see Figure 6.6A). A repeated measures ANOVA was then used to compare FI values between section areas and within cortical regions (see Figure 6.6B). Interestingly, there was no statistical difference in FI between sections or between regions (Brain area: $F(2,27) = 0.43$, $p = 0.62$. Region: $F(2,27) = 0.03$, $p = 0.97$. Interaction: $F(4,27) = 0.09$, $p = 0.98$).

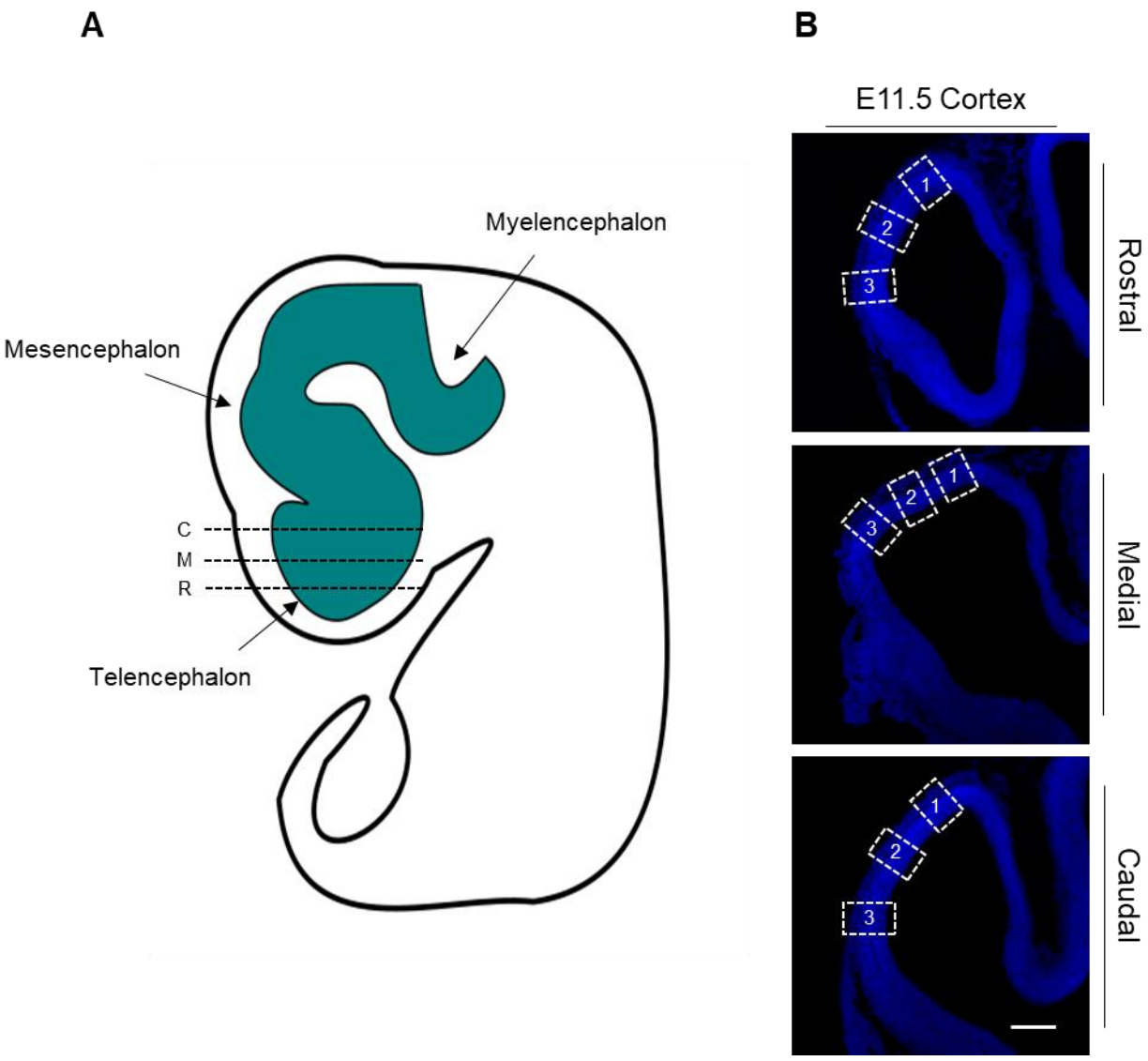


Figure 6.5. Experimental plan for IHC analysis of *TCF/Lef:H2B-GFP* animals. (A) Schematic side view diagram of an E11.5 embryo. The dashed lines indicate the rough position of the rostral (R), medial (M), and caudal (C) sections. Main regions of the developing brain are labelled (B) Representative images of rostral, medial, and caudal sections counterstained with DAPI. The dashed boxes illustrate the regions of analysis at the medial (1), middle (2), and lateral (3) regions of the cortex. Scale bar = 200 µm.

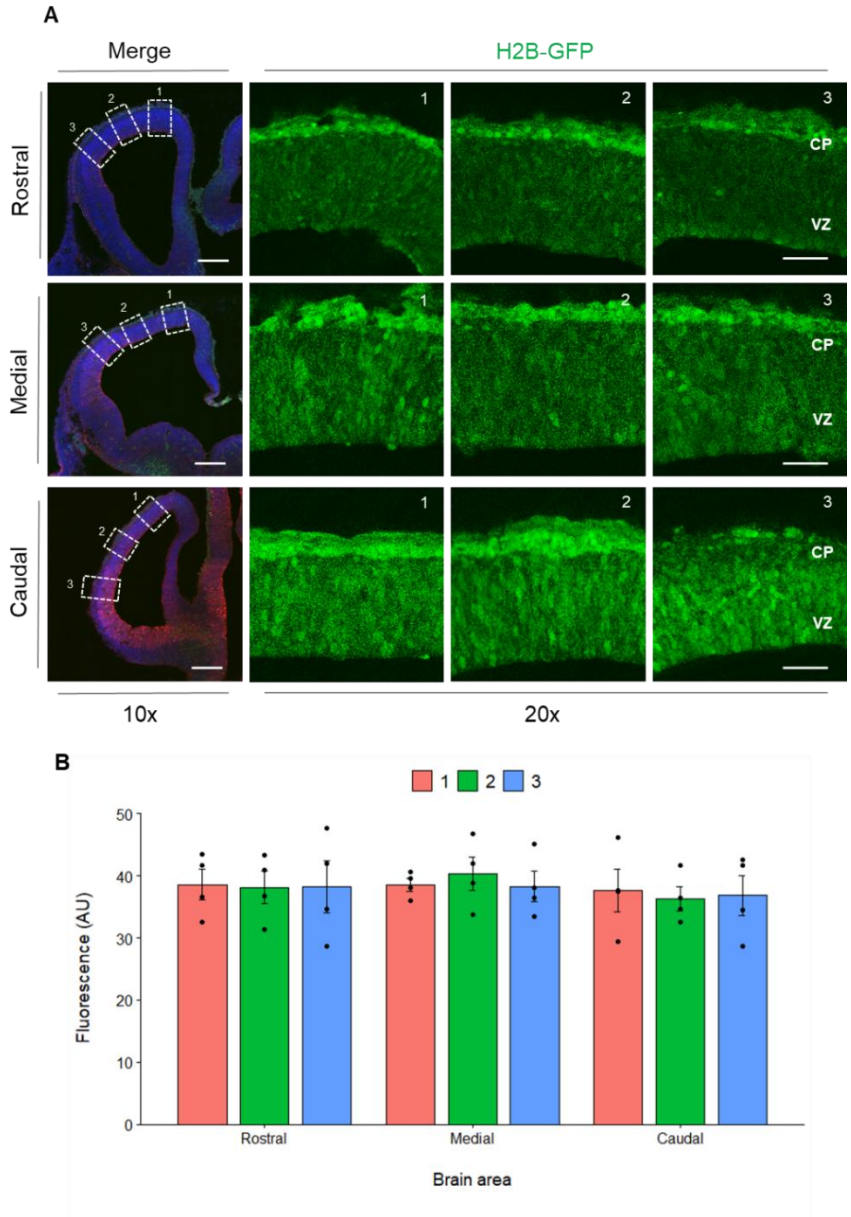


Figure 6.6. IHC analysis of fluorescence intensity of H2B-GFP from rostral, medial, and caudal sections and from medial to lateral areas of the cortex. (A) Representative images of E11 WT *TCF/Lef-H2B-GFP* rostral, medial, and caudal sections showing the endogenous H2B-GFP signal. Whole hemisphere images were taken at 10x and show PAX6 staining (red) and the DAPI counterstain (blue). Higher magnification images of H2B-GFP were taken at 20x. (B) Bar graph illustrating the quantification of the fluorescence intensity measurements in arbitrary units (AU). Two-way ANOVA analysis revealed no statistical difference in FI between brain area or cortical region. Data are represented as mean + standard error. 1 = medial, 2 = middle, 3 = lateral. VZ = ventricular zone. CP = cortical plate. Scale bar 200 µm (10x) and 50 µm (20x).

6.2.5. Differences in H2B-GFP fluorescence intensity between genotype groups

The FI of H2B-GFP from $PAX6^+$ RGCs was then compared between WT, HET, and KO cortices. To examine potential differences between PCDH19-WT and PCDH19-KO cells within the HET, sections were also stained using the β-GAL antibody to highlight PCDH19-KO cells. β-GAL- cells were considered as PCDH19-WT for the analysis. Taken together, the analyses below compare the FI of H2B-GFP from RGCs between WT and KO genotypes, as well as HET (WT) and HET (KO) epigenotypes.

Firstly, it was decided to determine whether there was a significant difference in FI of H2B-GFP between WT, KO, HET (WT), and HET (KO) groups across the whole brain (see Figure 6.7A). To do this, FI measurements were averaged from each quantification region of the cortex and averaged within each section. A one-way ANOVA was employed to assess whether there was a significant difference of FI between genotype groups (see Figure 6.7A and Figure 6.7B). Interestingly, analysis revealed no significant differences in FI between the groups ($F(3,9) = 0.08$, $p = 0.96$). To determine whether differences in FI may occur between genotype groups at specific areas of the brain, a repeated measures ANOVA was used to compare FI measurements between genotypes and within section areas (see Figure 6.7C). Interestingly, no significant main effects were found (Brain area: $F(2, 27) = 0.73$, $p = 0.48$. Genotype: $F(3, 27) = 0.28$, $p = 0.83$. Interaction: $F(6, 27) = 0.22$, $p = 0.96$).

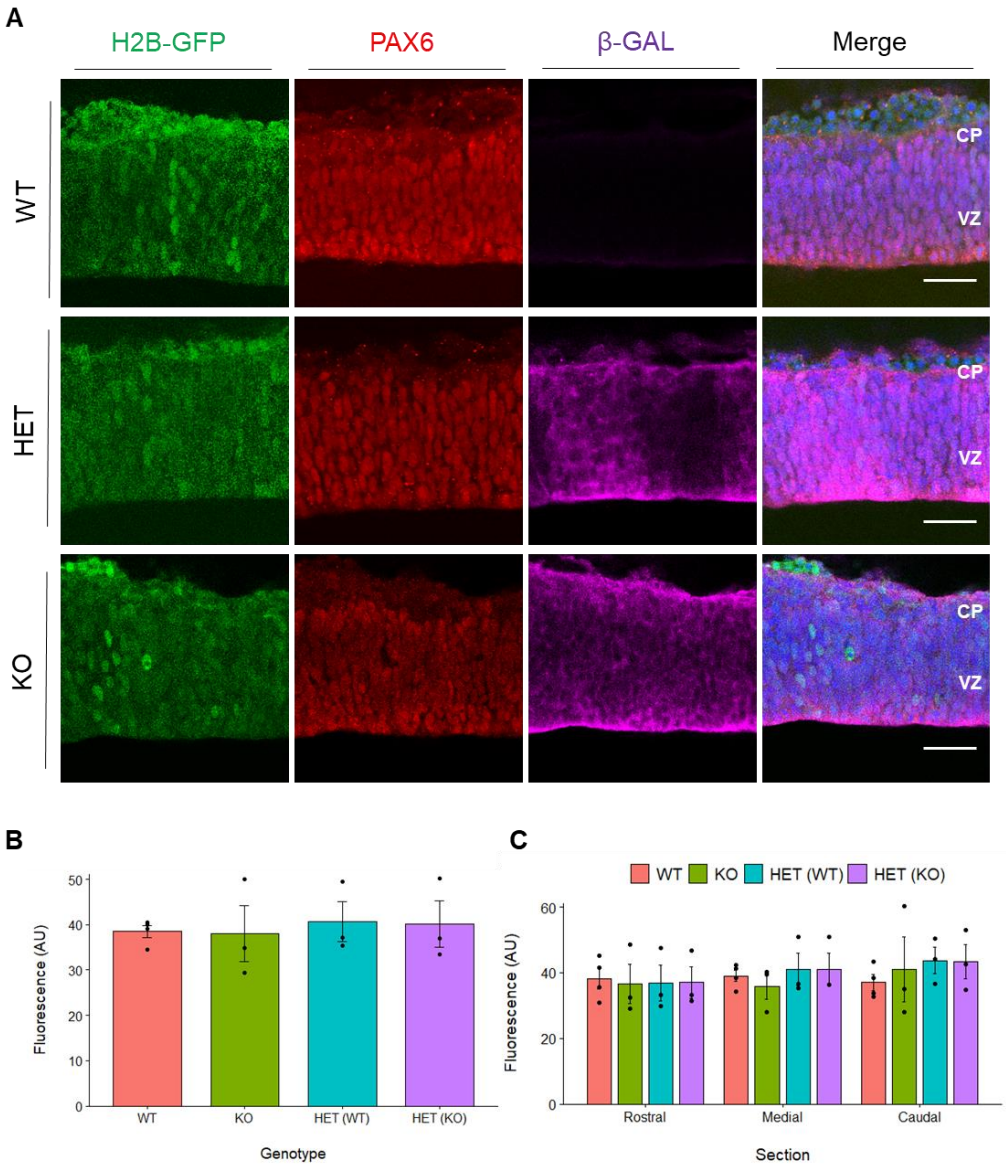


Figure 6.7. IHC analysis of fluorescence intensity of H2B-GFP between WT, KO, HET (WT) and HET (KO) RGCs. (A) Representative images of E11.5 WT, HET, and KO sections showing endogenous H2B-GFP signal (green) and stainings for PAX6 (red), β-GAL (magenta), and DAPI (blue). (B) Bar graph illustrating the quantification of fluorescence intensity measurements of H2B-GFP from the whole brain in arbitrary units (AU). One-way ANOVA revealed no difference in FI between groups. (C) Bar graph illustrating fluorescence intensity measurements of H2B-GFP from rostral, medial, and caudal regions and compared between genotypes. Two-way ANOVA revealed no difference in FI between group or brain region. Data is represented as mean + standard error. CP = cortical plate. VZ = ventricular zone. Scale bar = 50 µm. WT n = 4. HET n = 3. KO n = 3.

6.2.6. Fluorescence intensity and proportional analysis of pLRP6 (Thr1479)

In the previous section, no differences in H2B-GFP FI were detected between genotype groups. H2B-GFP signal represents the downstream transcriptional response of cells with active canonical Wnt signalling, i.e., the GFP expression is regulated by Wnt TFs TCF and LEF1 which are activated by the nuclear localisation of β-catenin. Therefore, although no differences were found in the previous section, it is possible that differences in Wnt signalling may be detectable further upstream in the pathway. To that end, IHC analysis was performed using pLRP6 (Thr1479) and pLRP6 (Ser1490) antibodies on WT, HET, and KO animals and FI measurements were taken using the same experimental design plan as previously discussed (see Figure 6.5). Likewise, sections were also co-stained with β-GAL antibody to quantify FI measurements from WT, KO, HET (WT) and HET (KO) genotype and epigenotype groups, as previously explained (see section 6.2.5). Moreover, because phosphorylation of LRP6 peaks during mitosis (Da Silva et al., 2021) it was also decided to co-stain sections using pHH3 antibody to aid in identifying mitotic cells for analysis. Interestingly, it was noticed that not all $pHH3^+$ cells were $pLRP6^+$. Therefore, it was decided to quantify the proportion of $pLRP6^+$ cells out of all cells in mitosis ($pHH3^+$) as another way of measuring Wnt signalling in the cortex.

Immunostaining using the pLRP6 (Thr1479) in WT, HET, and KO animals revealed a strong signal from mitotic cells, as expected (see Figure 6.8A). Moreover, one-way ANOVA analysis between WT, KO, HET (WT), and HET (KO) groups revealed no significant differences in FI of pLRP6 (Thr1479) ($F(3,12) = 0.31, p = 0.81$) (see Figure 6.8B). Although the FI of pLRP6 was not significantly different between genotype groups, the proportion of $pLRP6^+$ cells out of all $pHH3^+$ was still quantified. Interestingly, one-way ANOVA analysis revealed a significant main effect ($F(2,13) = 5.97, p = 0.01$). Bonferroni post-hoc analysis further revealed that KOs exhibited a higher proportion of $pLRP6^+$ cells in mitosis than WTs ($p = 0.008$) (see Figure 6.8C). Moreover, KOs also displayed a higher proportion of pLRP6+ cells in mitosis than HET (KO) cells ($p = 0.03$). No other comparisons were significant.

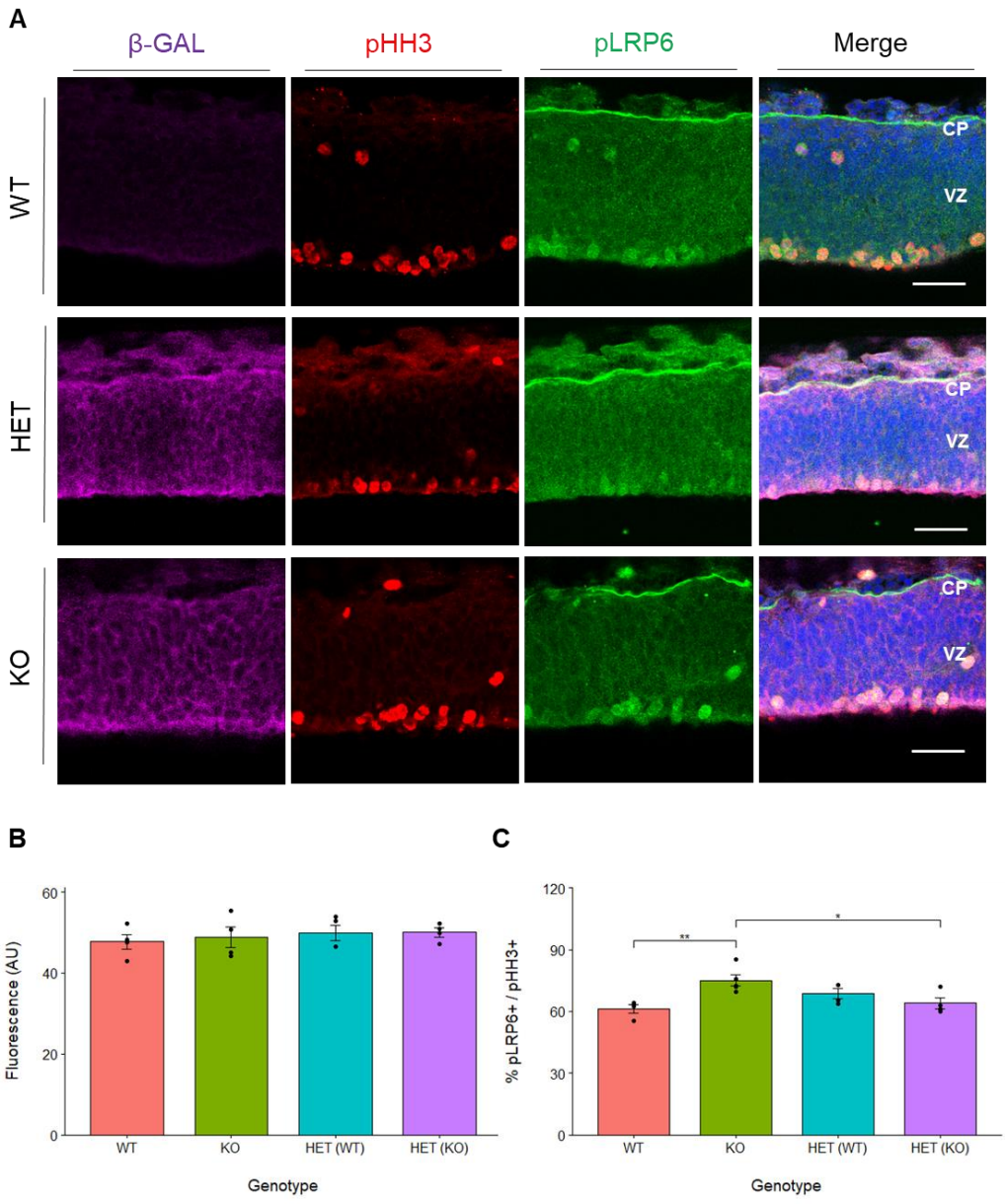


Figure 6.8. Fluorescence intensity and proportion analysis of pLRP6 (Thr1479) levels between WT, KO, HET (WT), and HET (KO) groups. (A) Representative images of E11 WT, HET, and KO sections showing immunostainings of β-GAL (magenta), pHH3 (red), pLRP6 (Thr1479) (green), and DAPI (blue). (B) Bar graph illustrating the quantification of fluorescence intensity of pLRP6 signal in arbitrary units (AU). One-way ANOVA revealed no difference between groups. (C) Bar graph illustrating the proportion of $pLRP6^+$ cells out of all $pHH3^+$ cells. One-way ANOVA revealed greater proportion of $pLRP6^+/pHH3^+$ in KOs compared to WTs ($p = 0.008$) and HET (KO) ($p = 0.03$) groups. Data is represented as mean + standard error. * = $p < 0.05$. ** = $p < 0.01$. VZ = ventricular zone. CP =cortical plate. Scale bar = 50 μm. Each group n = 4.

6.2.7. Fluorescence intensity and proportional analysis of pLRP6 (Ser1490)

Immunostaining using the pLRP6 (Ser1490) antibody revealed a strong signal from mitotic cells, as expected (see Figure 6.9A). Again, one-way ANOVA analysis between WT, KO, HET (WT), and HET (KO) revealed no significant differences in FI (*F(3,12) = 0.20 p = 0.88*) (see Figure 6.9B). Interestingly, when examining the proportion of $pLRP6^{+}$ cells from all pHH3 cells, a one-way ANOVA analysis revealed no significant differences between genotype groups (*F(3,12) = 0.87, p = 0.48*) (see Figure 6.9C).

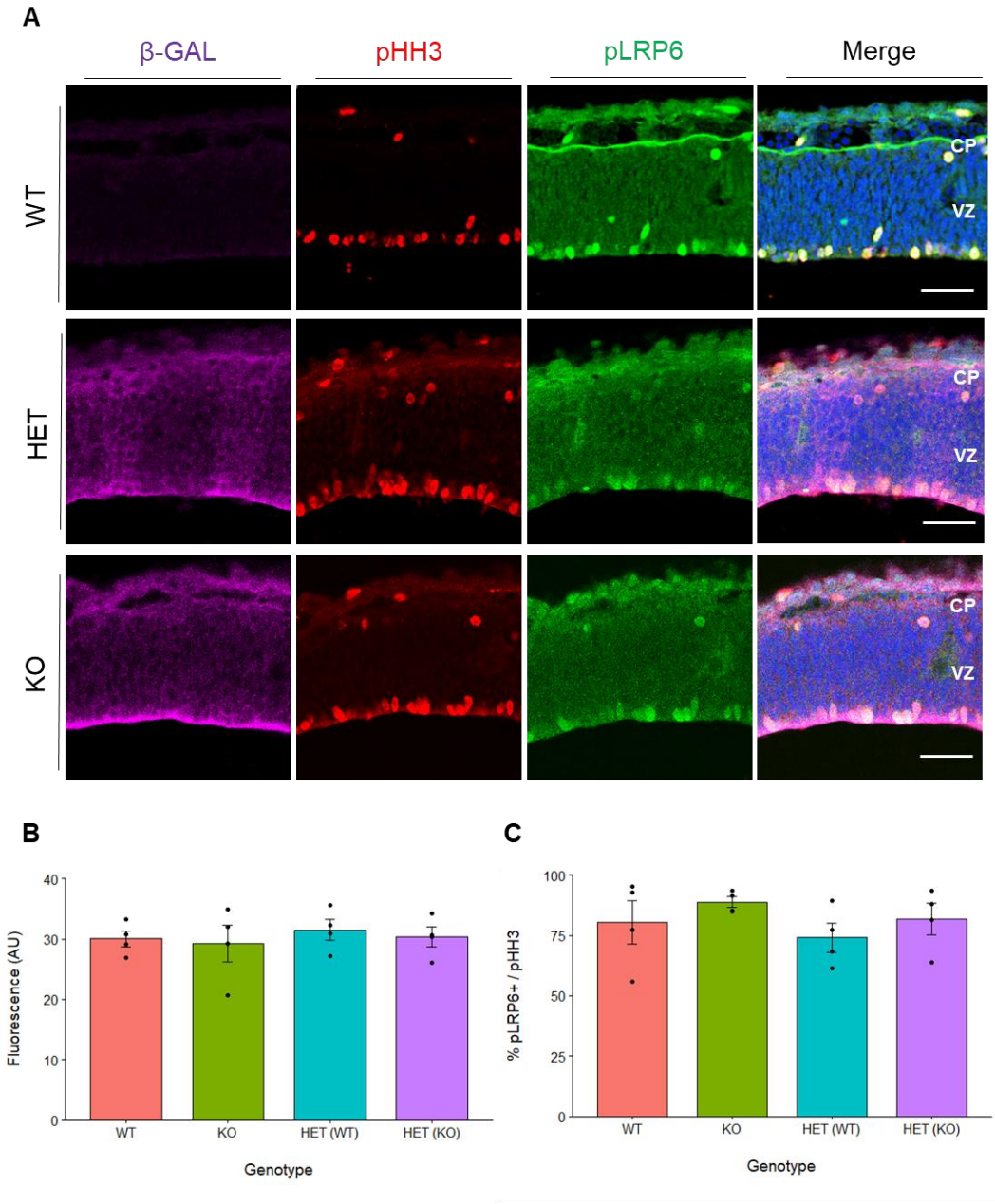


Figure 6.9. Fluorescence intensity and proportion analysis of pLRP6 (Ser1490) levels between WT, HET, and KO groups. (A) Representative images of E11 WT, HET, and KO sections showing immunostainings of β-GAL (magenta), pHH3 (red), pLRP6 (Ser1490) (green), and DAPI (blue). (B) Bar graph illustrating the quantification of fluorescence intensity of pLRP6 signal in arbitrary units (AU). One-way ANOVA revealed no differences in FI between groups. (C) Bar graph illustrating the proportion of $pLRP6^{+}$ cells out of all $pHH3^{+}$ cells. One-way ANOVA revealed no proportional differences between groups. Data is represented as mean + standard error. VZ = ventricular zone. CP = cortical plate. Scale bar = 50 μm. Each group n = 4.

6.3. Discussion

In this chapter, a first look into the role of PCDH19 in Wnt signalling was performed both *in vivo* and *in vitro*. The *Pcdh19* mutant mice were crossed with a Wnt reporter line *TCF/Lef:H2B-GFP* to examine potential differences in the levels of canonical Wnt activity between the genotype groups. Moreover, to provide a measurement of upstream Wnt signalling, IHC analysis was also performed using two phosphor-antibodies for LRP6. *In vitro* assays were also employed to examine whether PCDH19 could interact with known Wnt signalling molecules and whether overexpressing *Pcdh19* attenuated or enhanced canonical Wnt transcriptional activities.

6.3.1. WT, HET, and KO E11.5 embryos show no differences in downstream canonical Wnt signalling in vivo

As aforementioned, FI measurements of the H2B-GFP signal were taken from *Pcdh19* WT, HET, and KO E11 cortices and used as a proxy measure for downstream canonical Wnt signalling activity. Interestingly, no differences in FI were detected between the genotype groups, or between WT and KO cells within the HET. At a glance, this may suggest that the transcriptional signalling activity of canonical Wnt signalling is similar between the genotype groups. However, there are several shortcomings in this experiment that must be considered. For example, it has been reported that *Lef1* expression in the developing cortex is most intense in caudal regions (Vendrell et al., 2009). Moreover, because both transcription factors are important in the patterning of the hippocampus, the expression is also most intense within the medial regions of the cortex and gradually decreases laterally (Chodelkova et al., 2018; Galceran et al., 2000). However, although the *H2B-GFP* expression also followed this pattern, when analysing differences in FI between rostral-caudal regions and laterally across the cortex, no differences were found. Therefore, it is possible that the method of quantifying FI was not sensitive enough to translate the differences in *H2B-GFP* expression that were noticeable at the histological level. In other words, if the differences in expression levels of *H2B-GFP* between sections could not be accurately measured, then measuring differences in intensity between genotype groups would also be difficult. This is especially relevant as it was predicted that HETs and KOs would display some changes in canonical Wnt signalling, since in the previous chapter several canonical genes such as *Ccnd2* and *Lhx2,* were found to be dysregulated in these animals. This may also explain why FI differences were also not detected using the pLRP6 antibodies; FI may not be an accurate way of measuring signalling activity *in vivo*. Moreover, it was theorised that by taking FI measurements from the endogenous H2B-GFP signal rather than using antibodies to enhance

the signal, this would provide a more accurate measure of Wnt signalling activity. However, it is also possible that without an antibody, the potential differences are too subtle to be detected. Therefore, although these results suggest that downstream canonical Wnt signaling activity is conserved between WT, HET, and KOs, further optimisation using the *TCF/Lef1:H2B-GFP* line would need to be performed to confirm this result.

6.3.2. KO animals show a proportional increase in pLRP6^{+} cells despite no change in fluorescence intensity

One of the most surprising findings from the IHC analysis of pLRP6 was that KO animals showed a slight increase (~10%) in the proportion of pLRP6^{+} cells compared to WTs and HETs, but only when using the Thr1479 antibody. This was unexpected since HET animals showed no differences in proportion or in FI compared to WTs, yet these are the animals exhibiting the unusual neurogenesis phenotype (unpublished data, 2019). However, although KO animals were reported to show normal neurogenesis rates (unpublished data, 2019), it has been theorised that this may be driven by a compensation mechanism that is engaged after total loss of PCDH19. Therefore, these results provide some evidence that a slight increase in upstream Wnt signalling may underlie this compensation mechanism.

As aforementioned, Thr1479 phosphorylation is strictly Wnt-dependent whereas Ser1490 phosphorylation can occur both dependently and independently of Wnt signalling (Davidson et al., 2005; Wan et al., 2008). Therefore, the Thr1479 antibody technically provides a more accurate measure of LRP6-mediated Wnt signalling in the cortex, however Ser1490 phosphorylation is still necessary for engaging the internal signalling cascade (Niehrs and Shen, 2010). In fact, evidence suggests that Thr1479 phosphorylation may engage Wnt signalling by promoting the phosphorylation of Ser1490 (Yum et al., 2009). Likewise, Ser1490 promotes the phosphorylation of Thr1493 (Davidson et al., 2005; Zeng et al., 2005) in a sequence that has been described as a “ping-pong” mechanism (Niehrs and Shen, 2010). Taken together, these results may suggest that Wnt activity via phosphorylation of LRP6 is increased in RGCs of KO animals, but this is only detectable using a phosphor-antibody that targets a residue more upstream of the LRP6 phosphorylation cascade. Moreover, this may also suggest that loss of PCDH19 increases the sensitivity of RGCs to Wnt signalling. However, it is important to consider that further experiments would be needed to validate these claims. For example, it would be interesting to examine whether cultured WT, HET, and KO cells respond differently to Wnt3a treatment *in vitro*. Culturing these cells would allow precise control in treating the cells with Wnt3a and allow more control

compared to measuring the sensitivity of cells to Wnt signalling *in vivo*. Likewise, Wnt3a treatment would also be interesting to examine the kinetics of the phosphorylation hierarchy between WT, HET, and KO cells using phosphor-antibodies against the relevant sites of LRP6. Moreover, combining this experiment with knockdown (KD) experiments of the different kinases may yield evidence in which loss of PCDH19 increases kinase activity at the Thr1479 site but not the other residues.

6.3.3. PCDH19 interacts with RYK and β-catenin

Other ways of investigating whether PCDH19 is involved in the Wnt signalling pathway included experiments at the biochemical level. To that end, CoIP assays were performed to determine whether PCDH19 could interact with known Wnt signalling molecules, namely RYK and β-catenin. Interestingly, the IP reactions indicated that PCDH19 could interact with both molecules in an *in vitro* setting. When examining the intensity of the IP bands for both reactions, it was discovered that RYK showed a much stronger band with PCDH19 than β-catenin with PCDH19. This may suggest that RYK and PCDH19 may have a higher binding affinity or that they interact directly. Moreover, the weaker interaction observed between PCDH19 and β-catenin suggest that these molecules probably interact indirectly.

Although PCDH19 was shown to likely interact directly with RYK, the consequences of this interaction on Wnt signalling are unknown. Previous publications have shown that Pcdh19 and Ryk can interact *in vivo* within neural progenitors in zebrafish (Biswas et al., 2021). This interaction was important for regulating neurogenesis, since KD of Pcdh19 caused progenitors to over-proliferate, but a mechanism was not reported. It is possible that both proteins may orchestrate neurogenesis in progenitors via a processing mechanism. For example, publications in mice have demonstrated that RYK instigates neurogenesis in RGCs by being cleaved by γ-secretase whereby its intracellular domain is transported to the nucleus by chaperones SMEK1/2 to drive expression of neurogenic genes (Chang et al., 2017b; Lyu et al., 2008). Interestingly, it was recently shown that in neurons, PCDH19 is also proteolytically processed by ADAM10 and possibly γ-secretase in an NMDAR-dependent manner. The intracellular domain of PCDH19 is also transported to the nucleus where it drives expression of early-intermediate genes (Gerosa et al., 2022). Although processing of PCDH19 has not been reported in RGCs, data from the IMG lab has shown that a cytoplasmic fragment of PCDH19 is also detectable in E11 cortical lysates, however it is significantly weaker compared to the same fragment found in P10 cortical lysates (unpublished data, 2023). Therefore, it is likely that PCDH19 is mostly processed in neurons and in an activity-dependent manner, however a small degree of processing may also occur in

progenitors. Taken together, the fact that PCDH19 and RYK are both processed may indicate an intersecting mechanism; it is possible that PCDH19 and RYK may interact to inhibit or enhance the processing of each other in a Wnt-dependent manner, which may be important for regulating RGCs to commit to neurogenesis at the right time. Therefore, further work would be required to uncover the relationship between PCDH19 and RYK in the context of neurogenesis.

The suggested indirect interaction between PCDH19 and β-catenin is further supported by the fact that IP bands from the CoIP reaction involving N-cadherin and β-catenin are much stronger, since it is known that β-catenin interacts directly with N-cadherin at the plasma membrane (Linask et al., 1997; Sakane and Miyamoto, 2013; Zhang et al., 2013). Additionally, this could suggest that the weaker interaction observed between PCDH19 and β-catenin may be facilitated by endogenous N-cadherin. One way to confirm this would be to repeat the CoIP between PCDH19 and β-catenin with simultaneous KD of N-cadherin to investigate whether the interaction diminishes. Likewise, it would be interesting to see whether the interaction between PCDH19 and β-catenin increases after overexpressing *Pcdh19, Ctnnb1*, and *N-cadherin*.

6.3.4. PCDH19 may act as a negative regulator of canonical Wnt signalling

Although the CoIP results provided evidence that PCDH19, N-cadherin, and β-catenin may form a protein complex, the consequences of this interaction on Wnt signalling required further investigation. To that end, the TOP-FLASH luciferase assay was employed to investigate whether the single transfection or combination of PCDH19, N-cadherin, and β-catenin would affect the downstream transcriptional activities of TCF/LEF. As expected, single transfection of β-catenin caused a tremendous increase in luciferase activity, since β-catenin acts as a positive regulator for the internal Wnt signalling cascade. Likewise, the β-catenin luciferase activity was markedly increased after WNT3A treatment, however this was not statistically significant. Moreover, when combining N-cadherin and β-catenin, there was a significant decrease in luciferase activity compared to single β-catenin transfection, as expected (Sakane and Miyamoto, 2013). Interestingly, this effect was also observed after combining PCDH19 and β-catenin, suggesting that PCDH19 may also act as a negative regulator of β-catenin signalling, potentially by forming a protein complex with N-cadherin at the membrane. Overexpression of *Pcdh19* and *Ctnnb1* with simultaneous KD of *N-cadherin* would be useful to further elucidate the regulatory role of PCDH19 in Wnt/β-catenin signalling.

Nevertheless, further research would be needed to elaborate these findings. For example, it is possible that PCDH19 may regulate β-catenin signalling via a similar mechanism employed by N-

cadherin, e.g., sequestering β-catenin at the plasma membrane and thus controlling its levels in the cytoplasm (Sakane and Miyamoto, 2013). Because it was shown in this chapter that PCDH19 may need N-cadherin to interact with β-catenin, it is possible that the combination of both PCDH19 and N-cadherin would be needed for PCDH19 to regulate β-catenin levels. It is possible that PCDH19 may increase the binding affinity of N-cadherin to β-catenin and thus may act as a switch to instigate N-cadherin regulation of Wnt signalling. Although this has never been described before in the context of intracellular signalling, it is known that PCDH19 is likely the dominant partner in the PCDH19-N-cadherin complex and PCDH19 can increase the specificity and binding affinities of N-cadherin to make more durable extracellular interactions (Emond et al., 2011). Therefore, it is possible that PCDH19 may also act as a similar switch for N-cadherin interactions to intracellular proteins. However, it is also possible that PCDH19 may not be involved in the N-cadherin regulation of β-catenin and may regulate β-catenin through another mechanism. To demonstrate whether N-cadherin is needed for the regulation of β-catenin signalling by PCDH19, the same luciferase assay could be repeated with the addition of KD of endogenous N-cadherin. Examining the localisation of β-catenin after overexpression of N-cadherin and PCDH19 would also be required to assess whether presence of these proteins affects β-catenin localisation within the cell.

6.3.5. Conclusion

This chapter provided a first look into the role of PCDH19 in Wnt signalling. Differences in Wnt signalling were examined between WT, HET, and KO animals by crossing *Pcdh19* mutant mice with the Wnt reporter line *TCF/LEF:H2B-GFP* in addition to employing IHC analysis using antibodies for different phosphorylation sites of LRP6. Interestingly, WT, HET, and KO animals did not show any differences in H2B-GFP FI intensity, however KO animals showed a surprising increase in the proportion of pLRP6 $(Thr1479)^{+}$ cells. Moreover, analysis using CoIP and luciferase assays suggested that PCDH19 may indirectly interact with β-catenin and regulate its signalling capabilities, however further work is needed to expand on these findings.

Chapter 7: General discussion

This thesis aimed to elaborate on the function of PCDH19 in mouse cortical neurogenesis by investigating the molecular mechanisms that underpin this role *in vivo*. To do this, a combination of bulk and single cell (sc)RNA-seq techniques were employed to examine the transcriptional landscape of cortical tissue samples from *Pcdh19* wild type (WT), heterozygous (HET), and knockout (KO) E11 embryos. At the tissue level, few differentially expressed (DE) genes were found between genotypes, however at the single cell level multiple DE genes were uncovered that were involved in neurogenesis, ribosomal function, and Wnt signalling. An initial investigation into the role of PCDH19 in the Wnt signalling pathway was also conducted *in vitro* and *in vivo*. Analysis *in vivo* revealed potential differences in upstream Wnt signalling in KOs, whereas *in vitro* experiments showed that PCDH19 may negatively modulate Wnt signalling through canonical and possibly non-canonical pathways.

7.1. Molecular insights into the HETs

The previous work performed by Dr. Jessica Griffiths demonstrated that within HETs, the WT and KO RGCs displayed a decrease and increase in neurogenesis, respectively (unpublished data, 2019). Although these findings were statistically significant, it is important to note that the proportional difference in neurogenesis was not that profound (~15%). Major disruptions in the timing of the neurogenic switch in progenitors, such as premature neurogenesis or delayed neurogenesis, typically lead to severe brain malformations that can come in the form of macrencephaly or microencephaly (Uzquiano et al., 2018; Winden et al., 2015). However, even subtle changes in neurogenesis can lead to neurological problems (Guarnieri et al., 2022). The mosaic expression of *Pcdh19* is associated with PCDH19-epilpesy, a rare form of monogenic epilepsy characterised by early onset seizures and intellectual disability (Depienne et al., 2009; Dibbens et al., 2008). Although the underlying mechanism of the disorder is unknown, the early onset of the disease phenotypes suggested a developmental mechanism. Moreover, the work performed by Dr. Jessica Griffiths demonstrated a new avenue to explore the disease by understanding the role of PCDH19 in neurogenesis and how mosaic expression of *Pcdh19* interferes with normal neurogenesis rates. A molecular mechanism underlying the neurogenesis phenotype could therefore shed light on the potential disease mechanism that could be investigated further. In this thesis, examination of the transcriptional landscape of HET animals produced several intriguing results. Firstly, genes related to ribosomal biogenesis and function were found to be consistently dysregulated within the HETs across the bulk and single cell RNA-

seq experiments. Additionally, one of the most surprising findings in this thesis was that despite WT and KO RGCs within the HETs exhibiting different rates of neurogenesis (unpublished data, 2019), when comparing the transcriptional profiles of these cells to each other, no DE genes were uncovered except for *Pcdh19, EGFP*, and *β-gal*. Nevertheless, some DE genes from these cell populations were found by comparing to WT female cells. The major transcriptional results found in the HETs will be discussed in further detail below.

7.1.1. Defects in ribosome gene expression

Ribosomes are the molecular machinery integral for mediating translation of mRNAs into protein. The ribonucleoprotein complexes are typically formed by two subunits: the small 40S and the large 60S subunits, and these units are composed of a variety of ribosomal RNAs (rRNAs) and ribosome proteins (RPs). Disrupted expression of ribosomal genes, including ribosomal RNAs (rRNAs) and ribosomal proteins (RPs) have been associated with a variety of "ribosomopathies" such as Diamond Blackfan Anemia (DBA), which is associated with abnormal brain development, and in addition to various cancers (Teng et al., 2013). Therefore, ribosome biogenesis is an important component of normal brain development and has been shown to play a major role in neurogenesis. For instance, several RPs, including *Rps6* and *Rpl26,* have been shown to regulate cell cycle re-entry by translating mRNAs that promote cell proliferation (Kirn-Safran et al., 2007; Stewart and Denell, 1993). The transcription factor (TF) c-MYC, which is important for maintaining RGC proliferation, has also been shown to drive transcription of rRNAs during G_0/G_1 transition, which regulates cell cycling (Grandori et al., 2005). Likewise, N-MYC, another regulator of RGC neurogenic divisions, also controls the expression of RPs and rRNAs (Boon et al., 2001; Knoepfler et al., 2002). Compared to post-mitotic neurons, RGCs also exhibit higher levels of ribosome biogenesis which decreases over the course of neurogenesis (Chau et al., 2018). Therefore, ribosomal function and translation has been considered to play an important role in regulating cell cycle dynamics in cycling progenitors, and the decrease in ribosomal biogenesis is linked to neurogenic cell fate commitment (Chau et al., 2018). Taken together, ribosomal genes play an important role in neurogenesis, and the perturbed expression of ribosomal genes in the HETs may be involved in the neurogenesis phenotype observed in these animals.

However, it is important to note that the findings regarding the dysregulated expression of ribosomal genes in the HETs were conflicting between the bulk RNA-seq and scRNA-seq experiments. For instance, in the bulk RNA-seq experiment, no ribosome genes were found to be differentially expressed in the HETs, however GSEA revealed that several RP genes showed an increased fold change in the HETs compared to WT females. On the other hand, findings from

the single cell RNA-seq revealed that many RPs and rRNAs were decreased in the HETs compared to WTs and KOs. Based on these findings, it is difficult to conclude whether ribosomal genes are therefore upregulated or downregulated in the HETs. It is possible that the dysregulation of ribosomal genes in the HETs may be an artefact rather than a genuine biological signal. For example, rRNA is sometimes removed from RNA-seq datasets before data normalisation. This is because most 3' assay-based RNA-seq techniques, including 10X Chromium and NextSeq, capture polyadenylated (polyA) transcripts by using polydT sequence-based libraries. The intention is to only capture and sequence polyA-tail mRNA transcripts, however high abundance RNAs such as rRNAs can still be captured due to their sheer volume in the cell (Loi et al., 2021). This can sometimes be problematic since rRNAs can occupy much of the sequencing capacity which can crowd out low abundance mRNA transcripts, thus reducing sequencing sensitivity. In fact, kits and techniques have been designed to reduce capture of rRNAs before library preparation in order to generate libraries from enriched samples of mRNA (Herbert et al., 2018; Loi et al., 2021). Therefore, the fact that ribosome genes were upregulated in the HETs from the bulk RNA-seq, yet downregulated from the scRNA-seq, may be due to the fact that different levels of rRNAs were captured between each sequencing experiment, and thus may be due to technical variability and not be a genuine signal.

Nevertheless, although rRNAs are considered as "noise" in RNA-seq datasets, there is no consensus regarding whether only rRNA should be removed, or rRNA and mRNAs encoding for RPs. Likewise, rRNA and RPs are not always routinely removed from datasets since they can provide useful information in regards to the ribosome transcriptome (Luecken and Theis, 2019). This is especially the case if sample libraries were generated on the same day, since it can be more confidently considered that each sample underwent the same procedure and thus, differences in ribosome gene content between samples may not be rooted in technical variability. This is especially relevant considering that in this thesis, WT and KO samples showed similar levels of ribosome gene content in the scRNA-seq experiment, however HETs were the only group to show reduced ribosome content. However, since the WT and KO cells within the HETs showed a similar decrease in ribosome gene expression, it is difficult to explain how this pathway may underlie the neurogenic phenotype in these animals. Taken together, the findings regarding the abnormal expression of ribosomal genes in the HETs, as demonstrated in the bulk RNA-seq and scRNA-seq experiments, should be considered with caution since it is not clear whether these findings represent biological signal or noise. However, it would be interesting to explore potential translational defects further, which will be discussed later in this chapter.

7.1.2. Evidence of the Wnt signalling pathway in HET cortical development

Many other genes were found to be dysregulated in the HETs implicating several pathways in HET cortical development. Interestingly, several genes related to the Wnt signalling pathway were shown to be dysregulated in the HETs, as highlighted in the bulk RNA-seq (see Figure 3.13) and sc-RNA-seq (see Figure 5.8 and Figure 5.9) experiments. As previously discussed, the canonical and noncanonical Wnt pathways play a large role in early brain development events, such as cortical patterning (Backman et al., 2005; Machon et al., 2003). However, Wnt signalling is also crucial for regulating symmetric and asymmetric divisions of RGCs during neurogenesis. For example, the canonical Wnt signalling pathway (β-catenin-dependent) regulates neuroepithelial (NE) cell and RGC proliferation during early neurogenesis stages by promoting expression of pro-proliferative genes such as the aforementioned TF *c-myc* and *CyclinD1/2* (Chenn and Walsh, 2002; Niehrs and Acebron, 2012). Later in neurogenesis, canonical Wnt signalling transitions to promote neurogenic divisions by targeting expression of genes such as *n-myc, neurogenins,* and *neuroD1/2* (Hirabayashi et al., 2004; Munji et al., 2011). Although it is not clear when the switch from being pro-proliferative to pro-neurogenic occurs, and whether it affects all RGCs at the same time, the switch is generally considered to start around E14 (Munji et al., 2011). Thus, the canonical Wnt signalling pathway can generally be considered to still be pro-proliferative at E11. Therefore, since several genes known to play a role in the canonical pathway were dysregulated in the HETs, this may suggest that the pro-proliferative effect of Wnt is altered in these animals.

Interestingly, many of the DE Wnt genes found in the HETs from both RNA-seq experiments are not main components of the canonical pathway, such as *Wnt3a, Ctnnb1* etc. Instead, the identified genes function by supporting the canonical pathway; for example, RSPO1 and RSPO3 synergise with Wnt3a to enhance β-catenin stablisation by binding to LRP6 and preventing its internalisation (Binnerts et al., 2007; Wei et al., 2007). Likewise, SHISA2 also prevents the internalisation of FZD3, thus prolonging Wnt stimulation (Onishi and Zou, 2017). Upon Wnt3a stimulation, SALL1 translocates to the nucleus where it works independently but in conjunction with β-catenin to support Wnt-mediated gene expression (Sato et al., 2004). HMGA2 has also been shown to modify expression of *Fzd2* (Singh et al., 2014). Finally, APCDD1 forms a complex with LRP5 to inhibit Wnt signalling output (Shimomura et al., 2010). The upregulation of support molecules in the canonical pathway may explain why HETs do not display a profound proliferation phenotype, as previously discussed (see section 7.3.1). If the major components of the pathway were upregulated in the HETs, such as *Ctnnb1*, then the neurogenic phenotype would be expected to be much more severe (Chenn and Walsh, 2002). However, the combinatorial upregulation of Wnt

enhancing and attenuating molecules may instead subtly alter the pro-proliferative effects of Wnt signalling during HET cortical development.

A noticeable caveat of this hypothesis is that both WT and KO RGC populations displayed the increased expression of the aforementioned Wnt genes, thus inferring that both WT and KO cells should experience the same alterations in Wnt signalling. Therefore, the upregulation of Wnt support molecules on its own does not explain why WT and KO RGCs undergo neurogenesis at different rates. One possible explanation is that the upregulation of Wnt signalling works in conjunction with small, individual changes in the internal molecular architecture of the WT and KO RGCs which when taken together, underpins their neurogenic phenotypes. For instance, when compared to WT females, HET WT RGCs displayed increased expression of proliferative genes *Foxg1, Id4*, and *Cdon* (Hanashima et al., 2002; Pearson et al., 2020; Yun et al., 2004) and a downregulation of the neurogenic gene *Neurog2* (Lacomme et al., 2012). Likewise, HET KO RGCs showed decreased expression of the proliferative gene *Zic1* (Inoue et al., 2007). The upregulation of a handful of proliferative genes in the WT RGCs, and a downregulation of *Zic1* in the KO RGCs, may demonstrate that these cells are subtly primed to commit to proliferation or neurogenesis, respectively. Thus, the increased expression of Wnt enhancer and attenuator genes may be the trigger that pushes these cells to undergo neurogenesis at different rates.

7.1.3. Other pathways in HET cortical development

It is also worth noting that other pathways and mechanisms may be involved in HET brain development. For instance, the upregulation of *Bmp6* in the HETs implicates a role of the BMP pathway. BMPs are secreted morphogens that bind to heterotetrametric receptors (BMPR1a, BMPR1b) resulting in the phosphorylation of cytoplasmic SMAD proteins that translocate to the nucleus to initiate transcriptional activity (Bond et al., 2012). BMPs are secreted from the cortical hem and are known to cooperate with Wnt morphogens to promote dorsomedial telencephalic patterning (Furuta et al., 1997). Moreover, BMPs have also been shown to exert a pro-neurogenic effect on RGCs (Li et al., 1998). Interestingly, *Foxg1*, which was shown to be upregulated in WT and KO cells within the HETs, has been considered to be a pro-neurogenic downstream molecule of BMP signalling (Martynoga et al., 2005). Therefore, it is possible that the upregulation of *Bmp6* increases BMP signalling in WT and KO cells within the HETs, however both cell populations respond differently to this. Since KO RGCs may be considered more transcriptionally prone to neurogenesis (see section 7.3.2.1), the increase in BMP signalling and *Foxg1* expression may underlie why these cells undergo premature neurogenesis. Likewise, WT RGCs may be less sensitive to the pro-neurogenic effects of BMP signalling since these cells exhibit increased

expression of more pro-proliferative genes. Moreover, *Apcdd1* has been linked to attenuating Wnt and BMP signalling by binding to BMPR1a and preventing SMAD1 localisation to the nucleus (Vonica et al., 2020). The functional overlap between Wnt signalling and BMP signalling indicates that *Apcdd1* may be an important molecule in understanding how both pathways are involved in HET brain development. Moreover, although *Shisa2* and *Gli3* were considered as Wnt signalling components in this thesis, both proteins also have a widely documented role in the Shh pathway (Hasenpusch-Theil et al., 2012; Onishi and Zou, 2017; Wang et al., 2011). Taken together, many genes and pathways were identified as dysregulated in the HETs. Although further work will need to be performed to confirm the validity of these results, these findings should be useful in directing future efforts to explore the molecular mechanism underlying the unusual neurogenesis phenotype in the HET animals.

7.2. Molecular insights into the KOs

Although *PCDH19* is X-linked, mutations in the gene only affect heterozygous females or individuals harbouring somatic mutations (Depienne et al., 2011; Dibbens et al., 2008). Therefore, hemizygous males that possess only the mutant allele are developmentally normal. Although the underlying mechanism of PCDH19-epilepsy is unknown, the previous work performed by Dr. Jessica Griffiths suggests that mosaic expression of *Pcdh19* causes subtle alterations in neurogenesis output which may be an underlying factor of the disorder (unpublished data, 2019). Interestingly, this analysis also demonstrated that KO animals displayed no differences in neurogenesis compared to WT animals. Thus, loss of function of PCDH19 during neurogenesis does not alter neurogenic outputs, which may be linked to the fact that homogenous individuals with the mutant allele do not display neurological problems. One suggestion is that a compensation mechanism is tiggered upon complete loss of *Pcdh19* which stablises neurogenesis. However, since HETs display mosaic expression of *Pcdh19*, this mechanism may not engage in these animals, thus leading to impaired neurogenesis. However, whether there is a compensation mechanism in the KOs is currently unknown. Previous work has demonstrated that KO neural progenitors exhibit increased expression of genes related to neurogenesis and display increased neurogenesis *in vitro* (Homan et al., 2018). As previously discussed, whether this data represents the genuine mechanism by which PCDH19 regulates neurogenesis is questionable since KOs do not display altered neurogenesis *in vivo* (unpublished data, 2019). Likewise, because PCDH19 is an adhesion molecule, it likely requires cell-cell contact to exhibit its function in neurogenesis which is severed if the cells are dissociated and kept in culture.

Therefore, this thesis aimed to examine the molecular landscape of homozygous KOs *in vivo*, to understand why loss of *Pcdh19* does not change the neurogenesis rates in these animals.

7.2.1. Changes in transcription after loss of Pcdh19 in vivo

One of the most interesting results from this thesis was that the transcriptional profiles of KO animals were remarkably similar to those of WT animals. This was demonstrated in the PCA and UMAP analysis from the bulk and single cell RNA-seq experiments, respectively. Moreover, transcriptional comparison between KO and WT males revealed only *Pcdh19* and *β-gal* as differentially expressed, whereas small but coordinated changes in gene expression were observed in genes related to neurogenesis, cell adhesion, and, interestingly, Wnt signalling. Based on these findings, it was initially hypothesised that loss of *Pcd19* does not cause major transcriptional changes during cortical development, but rather small changes in specific genes. The small upregulation of genes related to cell adhesion, neurogenesis, and Wnt signalling may also underlie the aforementioned mechanism wherein loss of *Pcdh19* triggers a coordinated transcriptional response to compensate for the loss of the gene and normalise neurogenesis.

However, the hypothesis that loss of *Pcdh19* does not cause major transcriptional changes was challenged by several other results in this thesis. For instance, it is important to note that DE genes were found in the KO animals in the bulk RNA-seq when comparing to WT females. Interestingly, the KOs actually showed the largest list of DE genes when compared to WT females than any other comparison in chapter 3. Moreover, many of these genes were downregulated. It was discussed in that chapter that a synergistic effect between sex and genotype may explain why DE genes were more prevalent between mutants and opposite WT sex counterparts. However, examining the top 20 DE genes revealed that some genes were dysregulated only in one group. Therefore, it was possible that DE genes did exist in the KOs however were not detectable when compared to WT males due to statistical overcorrection. The many hundreds of genes that were uncovered between KOs and WT females were not fully explored in this thesis due to time constraints, however the most upregulated genes from this comparison were likely dysregulated due to accidental inclusion of non-specific tissue (medial ganglionic eminence). Moreover, the top downregulated genes were shown to be widely upregulated in the WT females in addition to the HETs, but to a lesser extent. Therefore, it is difficult to deduce whether these genes were dysregulated due to sex, a synergistic effect between sex and genotype, or were exclusively dysregulated in the KOs. It will be necessary to explore these genes further in order to determine whether loss of *Pcdh19* causes significant or small compensatory changes in transcription.

Another noteworthy factor against the initial hypothesis was that in the scRNA-seq, DE genes were uncovered between KOs and WT males. Interestingly, many of these genes were also related to neurogenesis, cell morphology, and Wnt signalling. It is also possible that these genes were considered differentially expressed due statistical inflation that can occur with single cell DE analysis methods (Hao et al., 2021; Squair et al., 2021). Moreover, it is important to mention that KOs showed the lowest numbers of captured cells for scRNA-seq analysis, which can also contribute to p-value inflation (Squair et al., 2021). On the other hand, the fact that these genes showed small changes in expression further supports the theory that loss of *Pcdh19* does not cause major transcriptional changes, but small coordinated changes in genes and pathways related to neurogenesis and Wnt signalling. Moreover, the fact that similar pathways were shown to be overrepresented in both bulk and the scRNA-seq experiments also provides support to this theory.

7.2.2. Dysregulated pathways in KOs

The dysregulation of components of the Wnt signalling pathway was a consistent finding in the KOs, suggesting that *Pcdh19* plays a role in this pathway during early neurogenesis. Interestingly, some of the dysregulated genes found from the scRNA-seq were significant components of the canonical pathway. For instance, *Ccnd2* is one of the gene target molecules in the canonical pathway (Chenn and Walsh, 2002; Machon et al., 2003). The gene *Lhx2* is also a TF that is needed for β-catenin-induced gene transcription, likely by forming a complex with the protein (Hsu et al., 2015). Interestingly, the *Lhx2* transcriptional function has also been shown to be a downstream responder to Shh signalling as well (Li et al., 2022). Most of the Wnt DE genes were downregulated in the KOs, and thus it was initially theorised that an attenuation of the canonical Wnt signalling pathway may play a role in normalising neurogenesis in these animals. However, previous publications have demonstrated that KO of *Pcdh19* in the neural progenitors of zebrafish induces hyperproliferation of these cells, due in part to an upregulation of the canonical pathway (Biswas et al., 2021). Therefore, those findings are the opposite of what was found in mice, i.e., KO of *Pcdh19* causes no difference in neurogenesis and a slight downregulation of genes in the canonical pathway. Interestingly, although KOs display slightly decreased expression of canonical genes, IHC analysis revealed that KOs were the only group to display an increase in the proportion of $pLRP6^{+}$ mitotic cells. Therefore, this conflicts with the notion that Wnt signalling may be clearly upregulated or downregulated in these animals, and how this may affect RGC divisions during KO cortical development.

One possible explanation is that certain noncanonical Wnt pathways are upregulated in the KO, whereas the canonical pathway is downregulated. For instance, the phosphorylation of LRP6 occurs upstream of many Wnt signalling pathways, including the WNT/STOP pathway, which is a post-translational regulatory pathway involved in regulating RGC divisions (Da Silva et al., 2021). Decreased LRP6 phosphorylation is associated with decreased WNT/STOP signalling, and thus the loss of *Pcdh19* might alter the balance of the different pathways downstream of Wnt. It could be that in the RGCs of the KO animal, the upregulation of non-canonical Wnt pathways ensures RGCs undergo neurogenesis at the right time despite canonical Wnt signalling being possibly decreased. Moreover, other pathways such as the Shh pathway may also be a contributing factor in normalising neurogenesis in these animals. It is also worth mentioning that data from the bulk RNA-seq experiment hinted at an upregulation of the BMP signalling pathway in the KOs, which was not explored further in this thesis due to time constraints. The fact that Wnt, Shh, and BMP were also dysregulated in the HETs highlights an avenue by which PCDH19 may be operating to regulate neurogenesis at the molecular level. This may also suggest that the mechanism behind the neurogenesis phenotype in the HETs and the potential compensation mechanism in the KOs may be similar, but the KO or mosaic expression of *Pcdh19* slightly adjusts this mechanism which leads to the differences in brain development observed in these animals (unpublished data, 2019). Taken together, slight alterations in Wnt, Shh, and BMP pathways may be a crucial underlying factor in normalising neurogenesis in the KO animals, however future work will be required to fully elucidate this further.

7.3. Molecular mechanisms of how PCDH19 may regulate cortical neurogenesis

The fact that components of the Wnt signalling pathway were found to be dysregulated in both HET and KO animals suggests that the mechanism by which PCDH19 regulates neurogenesis may be Wnt-related. The biochemical experiments performed in this thesis also strongly hint at this, in that PCDH19 may form a complex with β-catenin and potentially may be involved in the β-catenin sequestration mechanism employed by N-cadherin (Linask et al., 1997; Miyamoto et al., 2015; Sakane and Miyamoto, 2013). Whether this mechanism exists, and if it is involved in the cortical development of HET and KO animals, will require further study.

It is also possible that PCDH19 may operate through a different mechanism within the Wnt pathway, namely through the noncanonical Wnt receptor RYK. RYK has many roles during brain development, including regulating cell polarity (Andre et al., 2012), axon pathfinding (Keeble and Cooper, 2006), generation of interneurons (McKenzie et al., 2019; Zhong et al., 2011), and neurogenesis (Biswas et al., 2021; Chang et al., 2017b; Lyu et al., 2008). During neurogenesis,

RYK is primarily localised on the membrane of RGCs and upon interaction with Wnt ligands, the RYK ICD is cleaved and translocated to the nucleus to activate transcription of neurogenic genes (Chang et al., 2017b; Lyu et al., 2008). Interestingly, *Ryk* is expressed in mouse RGCs throughout the neurogenic period, starting at the onset of neurogenesis (around E11) and remains until as late as E18 (Lyu et al., 2008). However, the cleavage of RYK is minimal at the onset of neurogenesis, but this gradually increases as neurogenesis proceeds (Lyu et al., 2008). How the cleavage of RYK is regulated is unknown, however it is possible that since PCDH19 and RYK may show a direct interaction, the presence of PCDH19 in cortical RGCs at E11 and E12 may stablise RYK and prevent its proteolytic processing. Thus, when *Pcdh19* expression recedes from the cortex, RYK loses this stabilising influence and therefore is more susceptible to cleavage and instigating neurogenesis. Another possibility is that PCDH19 triggers RYK cleavage after forming a complex which increases even after *Pcdh19* expression is reduced in the cortex. Both theories may also explain why PCDH19 is active in RGCs under a strict temporal window; the activation of PCDH19 in RGCs may be necessary to begin or delay the RYK cleavage cascade, but once it begins then PCDH19 is no longer needed. This may be an underlying mechanism of the neurogenic switch and thus, it would be interesting to further explore the relationship between PCDH19 and RYK *in vivo*.

A noticeable caveat of the second theory is that it does not fully explain the neurogenesis phenotypes in the HET and KO animals. For instance, if PCDH19 is necessary for starting the RYK cleavage cascade, then it could be theorised that KO of *Pcdh19* would delay neurogenesis. Although this was observed in zebrafish (Biswas et al., 2021) this is not the case in mice (unpublished data, 2019). The first theory fits the data better: if PCDH19 delays RYK cleavage, then by extension the removal of *Pcdh19* should instigate neurogenesis prematurely. Since WT and KO RGCs in the HETs undergo neurogenesis at different rates, it could be that the presence of PCDH19 and RYK in the WT population delays the neurogenic switch in these cells, whereas the absence of PCDH19 within the KO population induces the neurogenic switch prematurely. However, as no neurogenic defects are apparent in KO animals, the question of the compensatory mechanism in the KOs remains. In any case, how the interaction between PCDH19 and RYK relates to the transcriptional data collected from HET and KO animals is unknown. For instance, RYK was not shown to be differentially expressed in either WTs, HETs, or KOs (data not shown), suggesting that perturbed expression of *Pcdh19* does not inherently affect *Ryk* expression. *Pcdh19* KO in zebrafish also induced a significant upregulation of canonical Wnt target genes, yet as previously discussed, the RGCs from KO mice showed a subtle decrease in canonical gene expression (Biswas et al., 2021). Taken together, the relationship between PCDH19 and

RYK in the context of cortical neurogenesis is still vastly unknown, however deciphering this relationship will likely be an important component to understanding the molecular mechanism by which PCDH19 regulates neurogenesis.

It is also worth mentioning that PCDH19 may operate through pathways other than the Wnt signalling pathway. For instance, several genes related to Shh signalling were dysregulated in the HETs, such as *Gli3* and *Shisa2* (Onishi and Zou, 2017; Wang et al., 2011). Shh signalling primarily promotes RGC proliferation (Cayuso et al., 2006; Wang et al., 2011) however other studies have demonstrated it is also important for the RGC-IP transition (Shikata et al., 2011) and for gliogenesis. Shh, Wnt, and Notch signalling all undergo considerable crosstalk during neurogenesis, for example by sharing the same gene targets or signalling molecules (Dave et al., 2011; Onishi and Zou, 2017; Ulloa et al., 2007). Therefore, although evidence suggests that PCDH19 may operate though the Wnt pathway, it is possible that PCDH19 may also be involved in other signalling pathways that regulate neurogenesis by either exerting a direct role in the pathway or affecting the pathway indirectly via crosstalk interference. Moreover, it is important to note that PCDH19 may regulate neurogenesis via its cell adhesive properties. For example, physical forces and the inheritance of internal cellular components during mitosis play a key role in symmetric and asymmetric divisions of RGCs (Kosodo et al., 2004). It is possible that PCDH19 functions by organising cellular components during mitosis via its interaction with the actin cytoskeleton (Chen et al., 2014). Moreover, the cell-cell interactions mediated by PCDH19 may also be important for ensuring the correct orientation of the cell as it divides. In the HET animals, the presence and absence of PCDH19 within WT and KO cells, respectively, may subtly push the cells in different orientations, which may affect the inheritance of internal cellular components and thus dysregulate symmetric and asymmetric divisions (Kosodo et al., 2004). Taken together, PCDH19 may regulate neurogenesis through many molecular avenues, not just through Wnt signalling.

7.4. Limitations of experiments

7.4.1. Advantages and disadvantages of RNA-seq

The employment of RNA-seq and especially scRNA-seq have provided invaluable insight into the molecular mechanisms that underpin many physiological functions ranging from cancer development to cortical neurogenesis (Hong et al., 2020; Ibarra-Soria et al., 2018; Telley et al., 2019). Thus, RNA-seq has become a prominent tool for most scientists to quantify whole gene expression. Both bulk RNA-seq and scRNA-seq roughly follow the same procedure: RNA

molecules are extracted from biological samples of interest and converted to cDNA, which is fragmented, amplified, and sequenced (Hrdlickova et al., 2017). Moreover, the computational analysis tools, including quality control, alignment, and gene counts are also similar. However, both techniques provide different insights into the transcriptional profiles of samples that bring advantages and disadvantages. For instance, bulk RNA-seq is typically less labourious than scRNA-seq and provides greater count depth since the RNA from the sample is sequenced as one without separating individual cells first. This means that it is more cost-effective to detect genes that exhibit relatively low expression but still may be biologically meaningful. Bulk RNA-seq is also typically the go-to method to explore transcriptional differences between conditions such as healthy and disease samples or WTs and mutants (Han et al., 2015; Hrdlickova et al., 2017) and is a very powerful technique when comparing homogenous samples.

However, one of the main caveats of performing bulk RNA-seq on tissue samples is that it averages expression data across all the cells within the tissue. A cell bulk often contains different types of cells that exhibit vastly different expression profiles and thus it is impossible to dissect which transcripts originated in each cell type (Ding et al., 2020; Hrdlickova et al., 2017). This was best exemplified in chapter 3 of this thesis since *Dlx1, Dlx6os1,* and *Gad2* were upregulated in the KO animals, however it is known that these genes are markers for interneurons (Lim et al., 2018). It is likely that a portion of the MGE was dissected from these animals and thus transcripts from interneurons were more enriched in the KO samples. However, this cannot be directly tested within the bulk RNA-seq matrix since it is not possible to assign transcripts to certain cells. This inherently is the main advantage of scRNA-seq, wherein gene expression profiles of samples can be examined at the single cell level (Islam et al., 2011; Tang et al., 2009). scRNA-seq has significantly advanced our understanding of progenitor cell biology during cortical development, such as providing further insight into progenitor heterogeneity (Eze et al., 2021; Zhong et al., 2018) and how the transcriptional landscape of progenitors changes over the course of neurogenesis (Telley et al., 2019). Moreover, the advent of faster and cheaper sequencing machines and more sensitive library preparation protocols has meant that scRNA-seq has become more affordable and viable for new labs to use (Ding et al., 2020; Haque et al., 2017; Zheng et al., 2017).

However, although scRNA-seq is a powerful technology, there are significant drawbacks to this technique (Whitley et al., 2016; Wu et al., 2018). For instance, scRNA-seq only uses a limited amount of RNA for sequencing compared to bulk RNA-seq. This means that scRNA-seq is typically noisier than bulk RNA-seq and thus accurate live/dead cell filtering and annotation can

sometimes be challenging (Ding et al., 2020; Wu et al., 2018). It also means that the transcriptional depth of scRNA-seq is markedly lower than what can be achieved with bulk RNA-seq, and therefore much of the expression data can be lost when examined at the single cell level. This may be one of the reasons why no DE genes were uncovered between WT and KO cells within the HETs, despite these cells displaying altered neurogenesis rates *in vivo* (unpublished data, 2019). A scRNA-seq technique that provides greater sequencing depth, such as SMART-seq, may be useful to examine these cells further and to elucidate more subtly dysregulated genes (Picelli et al., 2013; Wang et al., 2021). However, SMART-seq provides greater count depth at the cost of sequencing fewer cells (Wang et al., 2021). As it was demonstrated in chapter 4, 10X chromium scRNA-seq performed on E11 tissue captured thousands of RGCs, however many other cell types were also present in the matrix, including IPs, neurons, interneurons, meningeal cells etc. (see section 4.2.4). SMART-seq on the other hand, can only capture a few hundred cells. Thus, if SMART-seq were employed, it could be accurately assumed that some cell types would be captured that are not relevant to understanding the neurogenesis phenotype in the HETs. Therefore, these cells would have to be removed, and DE analysis would be performed on a much more limited population of RGCs.

Moreover, comparing biological conditions at the single cell level also comes with statistical challenges. For example, the Seurat single cell DE analysis technique considers each cell as a sample in the analysis, which means that much of the cellular variability in gene expression from one sample is not properly corrected for when compared to another other samples (Squair et al., 2021). Therefore, it is difficult to separate the variation of the whole cell group from the variation within the cell group, i.e., the variation of the population from the variation of the individual sample. This also contributes to p-value inflation and false discovery of DE genes (Squair et al., 2021). A notable example of this was the DE analysis between WT, HET, and KO animals within the IP cluster (see section 5.3.1). Compared to RGCs, there were fewer IPs in the single cell matrix, however DE analysis involving the IPs mostly produced the largest list of DE genes. This is likely because the IPs exhibited significant within-cell variation since new-born IPs and neuronally transitioning IPs displayed markedly different transcriptional profiles. There were attempts to overcome this challenge by employing “pseudo-bulk” analysis, wherein bulk RNA-seq DE methods were used on a transformed dataset that displayed the averaged transcriptional expression of genes from selected cells of interest (see section 5.3.2). Interestingly, this technique provided vastly different results compared to the single cell DE method. However, the pseudo-bulk results also demonstrated that genes, although maybe statistically differentially expressed, may not be biologically meaningful. Taken together, bulk RNA-seq and scRNA-seq exhibit many

advantages and disadvantages when examining gene expression profiles between biological groups of interest. Understanding these advantages and disadvantages are key to correctly utilising these techniques and to identify molecular hallmarks that can underpin many physiological functions. However, both techniques should be considered as exploratory and any interesting results must be confirmed in the lab via another method, such as qPCR.

7.4.2. Limitations of PCDH19 KO mouse model

Expression analysis from this thesis, as well as other work from the IMG lab (unpublished data, 2023), has demonstrated that there is residual expression of *Pcdh19* in the Taconic KO mouse line. This was first noticed when RNA-seq analysis was performed on WT and KO progenitors and neurons *in vitro* and examining the expression in KO cells revealed a moderate level of *Pcdh19* reads (unpublished data, 2023). Likewise, low-to-moderate levels of *Pcdh19* expression were also observed at the single cell level in RGCs (see section 4.2.6). Based on these findings, it was proposed that the remaining exons 4-6 of *Pcdh19* were still being expressed. To examine this further, cDNA from KO E11 cortical tissue was generated and PCR was performed using primers for *Pcdh19* exons 3-6, exons 4-6, and exons 5-6 (data not shown). It was theorised that if *Pcdh19* exons 4-6 were still being expressed in the KOs, PCR performed on cDNA obtained from KO tissue would produce amplicons with primers for exons 4-6 and exons 5-6, but not for exons 3-6. These predictions were validated, as PCR products of the expected sizes were detected using primers for exons 4-6 and exons 5-6, but no amplicons were detected when using primers for exons 3-6, suggesting that exons 4-6 of *Pcdh19* were still being expressed in the KOs. Moreover, to examine whether the residual *Pcdh19* mRNA was translated, western blot analysis using a C-terminus PCDH19 antibody was performed on E11 KO and P10 KO cortical tissue. Although full length PCDH19 was not detected, a small protein band of ~37kDa was found, which matched the predicted size of the *Pcdh19* truncated product (data not shown). Interestingly, the smaller band was barely detectable in the E11 extracts compared to the P10 extracts, despite running the same amount of protein from both samples. Moreover, mass spectrometry analysis performed after coimmunoprecipitation with PCDH19 antibodies from E11 WT and KO embryos revealed that within the KOs, no PCDH19 peptides were captured (unpublished data, 2023). Taken together, although residual *Pcdh19* mRNA was detected in the E11 KOs, protein analysis suggests that the truncated mRNA may not be stably translated at E11. Nevertheless, these results are concerning since the Taconic KO line is commercially available, but the *Pcdh19* KO allele may not be functioning as intended, especially in postnatal animals. Further work would be

required to understand why the remaining exons of *Pcdh19* are expressed and uncover whether the truncated protein product exhibits some biological function.

7.5. Future directions

The findings from this thesis have provided a strong foundation from which to further explore the molecular underpinnings of PCDH19 regulation of neurogenesis in health and disease contexts. This section will primarily address future experiments that will build upon the results found in this thesis.

7.5.1. Follow up on single cell RNA-seq results

The scRNA-seq analysis between WT, HET, and KO RGCs revealed numerous DE genes that will be useful for directing future investigations. First, the data will provide a valuable reference point for designing functional experiments wherein a variety of candidate DE genes can be selected to further analyse their relationship to PCDH19 and its regulatory role in neurogenesis. Initial follow-up studies must first confirm whether candidate DE genes are differentially expressed via a separate system, such as qPCR. This can be achieved by extracting mRNA from WT, HET, and KO E11 cortical tissue, generating cDNA via reverse transcription, and then using the cDNA as the template for qPCR. However, a noticeable caveat of this experiment is that if cDNA is generated from whole cortical tissue, it is likely that genetic material from cells other than RGCs will be present, as demonstrated from the cluster analysis in this thesis (see section 4.2.4). One solution to ensure cDNA only comes from RGCs would be to aspirate RGCs from acute embryonic brain slices that can be cut and kept in *in vivo* conditions. RGCs can be pooled and the mRNA from these cells can then be converted to cDNA to be used for qPCR. Genes related to Wnt signalling that were dysregulated in HETs and KOs, such as *Ccnd2*, would be good initial candidates to confirm whether they are indeed dysregulated. Examining other genes related to the ribosomal pathway, including *Rps18* and *Ddx3x*, would also be useful to confirm whether they are dysregulated in the HETs, and thus validate that the striking decrease in ribosomal gene expression is genuine.

7.5.2. Translation analysis

Differences in the levels of ribosomal mRNAs may be indicative that translation is dysregulated in the HETs. This may suggest that the overall levels of translation are reduced in the HETs, or more interestingly, that the translation of specific mRNA targets is reduced, which may be an underlying cause of the unusual neurogenesis phenotype in these animals. To investigate this

further, methods to examine translation events would need to be employed (Dermit et al., 2017). For example, polysome profiling has been widely used as a proxy for measuring global translation *in vitro* and *in vivo* by isolating ribosome complexes attached to mRNAs and then sequencing the mRNAs to determine which genes were about to be translated (Arava et al., 2003). Likewise, proteomics analyses such as mass spectrometry would also be useful to identify and quantify nascent proteins directly, however a drawback of this technique is that it measures global protein levels after translation has occurred.

Another interesting experiment to perform would be single cell proteomics. Single cell studies typically focus on nuclei acids as a proxy for understanding the complex biology of samples. However, measuring gene expression does not necessarily provide information regarding the different chemistries, interactions, and dynamics of the translated proteins (Kelly, 2020). In this thesis, it was found that ribosomal expression was significantly decreased in the HETs, suggesting translation is reduced in these animals. However, what cannot be inferred from this data is which mRNAs are not translated (Perkel, 2021). For instance, it is possible that multiple genes are transcribed normally in the HETs, but the transcripts are not properly translated – something that would not be possible to detect using RNA-seq. Therefore, it would be interesting to perform single cell proteomics on HETs and integrate the data with the single cell RNA-seq measurements to infer which genes are expressed but are not properly translated. A new technology called single cell proteomics mass spectrometry (SCoPE-MS) has recently been developed that employs isobaric carriers which greatly increases the chance of detecting proteins by providing a tag-based framework in which proteins from single cells can be detected and quantified by mass spectrometry (Budnik et al., 2018). SCoPE-MS has been shown to quantify ~3000 proteins in ~1500 cells and has recently been used in conjunction with 10X chromium scRNA-seq to explore the interactions between the tumour suppressor p53 protein, its transcript, and the proteins that regulate p53 expression (Specht et al., 2021). Taken together, single cell proteomics would be a useful next step in understanding the role of PCDH19 in neurogenesis by providing insights into the proteome of HETs can exploring how disruption in protein mechanics may underlie the unusual neurogenesis phenotype in these animals.

7.5.3. PCDH19 and the Wnt signalling pathway

Data was generated in this thesis which suggests that PCDH19 may have a function in the Wnt signalling pathway, however further work is needed to elucidate this function. For instance, it would be interesting to examine whether PCDH19 and β-catenin form a complex through N-cadherin. This could be achieved by transfecting *Pcdh19* and *Ctnnb1* plasmids into HEK293 cells

while simultaneously transfecting shRNAs for *N-cadherin*. The CoIP and western analysis can then be performed in the same fashion as before (see section 6.2.1) to measure whether the IP band detected between PCDH19 and β-catenin is still present. If KD of N-cadherin reduces the intensity of the IP band, then it can be suggested that PCDH19 and β-catenin likely interact through N-cadherin. This would help in understanding whether PCDH19 is an important component of N-cadherin-mediated regulation of β-catenin signalling (Miyamoto et al., 2015; Sakane and Miyamoto, 2013; Zhang et al., 2013) and will highlight a mechanism in which PCDH19 may regulate neurogenesis (Zhang et al., 2013).

Moreover, it would be interesting to elaborate more on the functional role between RYK and PCDH19. Since it is possible that both proteins interact directly, then this provides an interesting avenue to explore the role of PCDH19 in a noncanonical Wnt pathway. For instance, since the proteolytic cleavage of RYK has been shown to be necessary for promoting neurogenic divisions in RGCs (Chang et al., 2017b; Lyu et al., 2008), and PCDH19 is also proteolytically cleaved by the same enzymes as RYK (Gerosa et al., 2022; unpublished data, 2023), this highlights a potential overlapping dynamic. Although it has been shown in the IMG lab that proteolytic cleavage of PCDH19 occurs less frequently at E11 than at P10 (data not shown), it is possible that the PCDH19-RYK interaction may be necessary for RYK proteolytic cleavage. One way to examine this would be to transfect *Pcdh19, Ryk,* and *Pcdh19-Ryk* together in HEK293 cells and treat with ionomycin, which has been shown to promote proteolytic processing (unpublished data, 2023). If the presence of PCDH19 and RYK reduces or enhances proteolytic processing, this may highlight a mechanism in which both proteins operate together to regulate their cleavage potential. Moreover, it would be interesting to determine whether the presence of PCDH19 with RYK reduces or enhances RYK processing after treatment with Wnt3a (Lyu et al., 2008).

Although PCDH19 cleavage occurs less frequently in RGCs compared to neurons, the cleavage that does occur in RGCs may play a specific and important function in neurogenesis. Therefore, it would be interesting to examine the role of the PCDH19 ICD in cortical development. For instance, it is known that the RYK ICD is transported to the nucleus by SMEK1/2 in RGCs and forms a transcription complex to drive expression of neurogenic genes (Chang et al., 2017b). Moreover, it is also known that the PCDH19 ICD is transported to the nucleus in neurons and regulates transcription (Gerosa et al., 2022; unpublished data, 2023). Interestingly, it has been predicted that the RYK ICD does not harbour an endogenous nuclear localisation sequence (NLS), and therefore must form a complex with chaperone proteins to be translocated (Chang et al., 2017b; Lyu et al., 2008). Therefore, since both RYK and PCDH19 full length proteins interact,

it would be interesting to examine whether their ICDs also interact and whether the PCDH19 ICD is important for the translocation of the RYK ICD during neurogenesis. This can be achieved by cloning the ICD sequences from both proteins and performing CoIP after transfection in HEK293 cells. Immunocytochemistry would also be useful to examine whether the presence of PCDH19 ICD is influences the translocation of the RYK ICD to the nucleus. If both ICDs do translocate to the nucleus, then it would be interesting to determine whether they form a transcription complex and drive expression of target genes. This can be examined by overexpressing both ICDs individually or together in dissociated neural progenitors *in vitro* and performing ChIP-seq to identify genes that are targeted by each ICD and genes that may be targeted only if both ICDs form a complex together. Interestingly, it is known that the localisation of RYK ICD to the nucleus is Wnt-dependent (Chang et al., 2017b; Lyu et al., 2008) but it is not known whether the same is true for PCDH19 ICD in RGCs. Therefore, it would be interesting to determine whether PCDH19 ICD is transported to the nucleus of neural progenitors after Wnt3a treatment.

7.5.4. How cell-cell forces may influence PCDH19 regulation of neurogenesis

Another important consideration is that the cell adhesive properties of PCDH19 may be crucial in its function to regulate neurogenesis (Chen et al., 2014; Kosodo et al., 2004). One way to examine this would be to use time lapse imaging in combination with ICC. For instance, it would be interesting to dissociate neural progenitors from E11 HET cortical tissue and keep these cells in culture. Time lapse analysis could then be used to record the first division the cells and afterwards, the cells could then be subjected to ICC and stained using β-GAL, PAX6, TBR2, and TBR1 antibodies. The recorded cells could then be identified using a confocal microscope and the identity of the daughter cell could be examined using the aforementioned markers. Moreover, β-GAL would be used to differentiate between WT and KO cells. Therefore, if the daughter cells of WT and KO progenitors show an increased chance to self-renew or differentiate, respectively, then it can be proposed that the neurogenesis phenotype in the HETs is partially cell autonomous. However, if the daughter cells produce progenitors and neurons evenly, then this would provide evidence that PCDH19 requires cell-cell contact to mediate its effect on neurogenesis.

7.6. Concluding remarks

This thesis has provided the first insight into the molecular mechanisms in which PCDH19 may regulate neurogenesis. For the first time, transcriptional data has been collected from WT, HET, and KO E11 cortical tissue *in vivo* and examined at the population (bulk RNA-seq) and single cell (scRNA-seq) levels. Particular attention was placed on the role of PCDH19 in the Wnt signalling

pathway, as demonstrated by multiple dysregulated Wnt genes found in the RNA-seq experiments. Moreover, PCDH19 may negatively regulate the canonical Wnt pathway and possibly operate through the noncanonical Wnt/RYK pathway to orchestrate neurogenesis, however further work will be needed to elucidate this mechanism further.

References

Aaku-Saraste, E., Hellwig, A., and Huttner, W.B. (1996). Loss of Occludin and Functional Tight Junctions, but Not ZO-1, during Neural Tube Closure—Remodeling of the Neuroepithelium Prior to Neurogenesis. Dev. Biol. *180*, 664–679.

Acebron, S.P., Karaulanov, E., Berger, B.S., Huang, Y.L., and Niehrs, C. (2014). Mitotic wnt signaling promotes protein stabilization and regulates cell size. Mol. Cell *54*, 663–674.

Alexandre, P., Reugels, A.M., Barker, D., Blanc, E., and Clarke, J.D.W. (2010). Neurons derive from the more apical daughter in asymmetric divisions in the zebrafish neural tube. Nat. Neurosci. *13*, 673–679.

Andre, P., Wang, Q., Wang, N., Gao, B., Schilit, A., Halford, M.M., Stacker, S.A., Zhang, X., and Yang, Y. (2012). The Wnt coreceptor Ryk regulates Wnt/planar cell polarity by modulating the degradation of the core planar cell polarity component Vangl2. J. Biol. Chem. *287*, 44518–44525.

Arai, Y., Pulvers, J.N., Haffner, C., Schilling, B., Nüsslein, I., Calegari, F., and Huttner, W.B. (2011). Neural stem and progenitor cells shorten S-phase on commitment to neuron production. Nat. Commun. *2*.

Arava, Y., Wang, Y., Storey, J.D., Liu, C.L., Brown, P.O., and Herschlag, D. (2003). Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U. S. A. *100*, 3889–3894.

Asami, M., Pilz, G.A., Ninkovic, J., Godinho, L., Schroeder, T., Huttner, W.B., and Götz, M. (2011). The role of Pax6 in regulating the orientation and mode of cell division of progenitors in the mouse cerebral cortex. Development *138*, 5067–5078.

Backman, M., Machon, O., Mygland, L., Van Den Bout, C.J., Zhong, W., Taketo, M.M., and Krauss, S. (2005). Effects of canonical Wnt signaling on dorso-ventral specification of the mouse telencephalon. Dev. Biol. *279*, 155–168.

Baek, C., Freem, L., Goïame, R., Sang, H., Morin, X., and Tozer, S. (2018). Mib1 prevents Notch Cis-inhibition to defer differentiation and preserve neuroepithelial integrity during neural delamination. PLOS Biol. *16*, e2004162.

Barker, N. (2008). The canonical wnt/β-catenin signalling pathway. Methods Mol. Biol. *468*, 5–15.

Bassani, S., Cwetsch, A.W., Gerosa, L., Serratto, G.M., Folci, A., Hall, I.F., Mazzanti, M.,

Cancedda, L., and Passafaro, M. (2018). The female epilepsy protein PCDH19 is a new GABAAR-binding partner that regulates GABAergic transmission as well as migration and morphological maturation of hippocampal neurons. Hum. Mol. Genet. *27*, 1027–1038.

Benjamini, Y., and Speed, T.P. (2012). Summarizing and correcting the GC content bias in high-throughput sequencing. Nucleic Acids Res. *40*.

Berletch, J.B., Yang, F., Xu, J., Carrel, L., and Disteche, C.M. (2011). Genes that escape from X inactivation. Hum Genet *130*, 237–245.

Bhanot, P., Brink, M., Samos, C.H., Hsieh, J.C., Wang, Y., Macke, J.P., Andrew, D., Nathans, J., and Nusse, R. (1996). A new member of the frizzled family from Drosophila functions as a Wingless receptor. Nature *382*, 225–231.

Bian, S., Hong, J., Li, Q., Schebelle, L., Pollock, A., Knauss, J.L., Garg, V., and Sun, T. (2013). MicroRNA cluster miR-17-92 regulates neural stem cell expansion and transition to intermediate progenitors in the developing mouse neocortex. Cell Rep. *3*, 1398–1406.

Binnerts, M.E., Kim, K.A., Bright, J.M., Patel, S.M., Tran, K., Zhou, M., Leung, J.M., Liu, Y., Lomas, W.E., Dixon, M., et al. (2007). R-Spondin1 regulates Wnt signaling by inhibiting internalization of LRP6. Proc. Natl. Acad. Sci. U. S. A. *104*, 14700–14705.

Bisogni, A.J., Ghazanfar, S., Williams, E.O., Marsh, H.M., Yang, J.Y.H., and Lin, D.M. (2018). Tuning of delta-protocadherin adhesion through combinatorial diversity. Elife *7*.

Biswas, S., Emond, M.R., and Jontes, J.D. (2010). Protocadherin-19 and N-cadherin interact to control cell movements during anterior neurulation. J. Cell Biol. *191*, 1029.

Biswas, S., Emond, M.R., Chenoweth, K.P., and Jontes, J.D. (2021). δ-Protocadherins regulate neural progenitor cell division by antagonizing Ryk and Wnt/β-catenin signaling. IScience *24*, 102932.

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

Boon, K., Caron, H.N., Van Asperen, R., Valentijn, L., Hermus, M.C., Van Sluis, P., Roobeek, I., Weis, I., Voûte, P.A., Schwab, M., et al. (2001). N-myc enhances the expression of a large set of genes functioning in ribosome biogenesis and protein synthesis. EMBO J. *20*, 1383.

Borello, U., Cobos, I., Long, J.E., Murre, C., and Rubenstein, J.L.R. (2008). FGF15 promotes

neurogenesis and opposes FGF8 function during neocortical development. Neural Dev. *3*.

Borghi, R., Magliocca, V., Petrini, S., Conti, L.A., Moreno, S., Bertini, E., Tartaglia, M., and Compagnucci, C. (2021). Dissecting the role of pcdh19 in clustering epilepsy by exploiting patient-specific models of neurogenesis. J. Clin. Med. *10*, 10.

Brasch, J., Goodman, K.M., Noble, A.J., Rapp, M., Mannepalli, S., Bahna, F., Dandey, V.P., Bepler, T., Berger, B., Maniatis, T., et al. (2019). Visualization of clustered protocadherin neuronal self-recognition complexes. Nature *569*, 280–283.

Brennecke, P., Anders, S., Kim, J.K., Kołodziejczyk, A.A., Zhang, X., Proserpio, V., Baying, B., Benes, V., Teichmann, S.A., Marioni, J.C., et al. (2013). Accounting for technical noise in single-cell RNA-seq experiments. Nat. Methods 2013 1011 *10*, 1093–1095.

Budnik, B., Levy, E., Harmange, G., and Slavov, N. (2018). SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. Genome Biol. *19*.

Bultje, R.S., Castaneda-Castellanos, D.R., Jan, L.Y., Jan, Y.N., Kriegstein, A.R., and Shi, S.H. (2009). Mammalian Par3 regulates progenitor cell asymmetric division via Notch signaling in the developing neocortex. Neuron *63*, 189.

Butler, M.G., Rafi, S.K., Hossain, W., Stephan, D.A., and Manzardo, A.M. (2015). Whole Exome Sequencing in Females with Autism Implicates Novel and Candidate Genes. Int. J. Mol. Sci *16*, 1312–1335.

Büttner, M., Miao, Z., Wolf, F.A., Teichmann, S.A., and Theis, F.J. (2018). A test metric for assessing single-cell RNA-seq batch correction. Nat. Methods 2018 161 *16*, 43–49.

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. J. Cell Sci. *116*, 4947–4955.

Calegari, F., Haubensak, W., Haffher, C., and Huttner, W.B. (2005). Selective lengthening of the cell cycle in the neurogenic subpopulation of neural progenitor cells during mouse brain development. J. Neurosci. *25*, 6533–6538.

Camacho, A., Simón, R., Sanz, R., Viñuela, A., Martínez-Salio, A., and Mateos, F. (2012). Cognitive and behavioral profile in females with epilepsy with PDCH19 mutation: Two novel mutations and review of the literature. Epilepsy Behav. *24*, 134–137.

Canzio, D., and Maniatis, T. (2019). The generation of a protocadherin cell-surface recognition code for neural circuit assembly. Curr. Opin. Neurobiol. *59*, 213–220.

Cárdenas, A., Villalba, A., de Juan Romero, C., Picó, E., Kyrousi, C., Tzika, A.C., Tessier-Lavigne, M., Ma, L., Drukker, M., Cappello, S., et al. (2018). Evolution of Cortical Neurogenesis in Amniotes Controlled by Robo Signaling Levels. Cell *174*, 590-606.e21.

Cayuso, J., Ulloa, F., Cox, B., Briscoe, J., and Marti, E. (2006). The Sonic hedgehog pathway independently controls the patterning, proliferation and survival of neuroepithelial cells by regulating Gli activity. Development *133*, 517–528.

Cen, C., Luo, L. Da, Li, W.Q., Li, G., Tian, N.X., Zheng, G., Yin, D.M., Zou, Y., and Wang, Y. (2018). PKD1 Promotes Functional Synapse Formation Coordinated with N-Cadherin in Hippocampus. J. Neurosci. *38*, 183–199.

Chaillou, T., Kirby, T.J., and Mccarthy, J.J. (2014). Ribosome biogenesis: emerging evidence for a central role in the regulation of skeletal muscle mass. J. Cell. Physiol. *229*, 1584.

Chalasani, K., and Brewster, R.M. (2011). N-cadherin-mediated cell adhesion restricts cell proliferation in the dorsal neural tube. Mol. Biol. Cell *22*, 1505–1515.

Chang, H., Hoshina, N., Zhang, C., Ma, Y., Cao, H., Wang, Y., Wu, D.D., Bergen, S.E., Landén, M., Hultman, C.M., et al. (2017a). The protocadherin 17 gene affects cognition, personality, amygdala structure and function, synapse development and risk of major mood disorders. Mol. Psychiatry 2018 232 *23*, 400–412.

Chang, W.H., Choi, S.H., Moon, B.S., Cai, M., Lyu, J., Bai, J., Gao, F., Hajjali, I., Zhao, Z., Campbell, D.B., et al. (2017b). Smek1/2 is a nuclear chaperone and cofactor for cleaved Wnt receptor Ryk, regulating cortical neurogenesis. Proc. Natl. Acad. Sci. U. S. A. *114*, E10717–E10725.

Chau, K.F., Shannon, M.L., Fame, R.M., Fonseca, E., Mullan, H., Johnson, M.B., Sendamarai, A.K., Springel, M.W., Laurent, B., and Lehtinen, M.K. (2018). Downregulation of ribosome biogenesis during early forebrain development. Elife *7*.

Chen, W. V., and Maniatis, T. (2013). Clustered protocadherins. Development *140*, 3297–3302.

Chen, B., Brinkmann, K., Chen, Z., Pak, C.W., Liao, Y., Shi, S., Henry, L., Grishin, N. V., Bogdan, S., and Rosen, M.K. (2014). The WAVE regulatory complex links diverse receptors to the actin cytoskeleton. Cell *156*, 195–207.

Chen, W., Ten Berge, D., Brown, J., Ahn, S., Hu, L.A., Miller, W.E., Caron, M.G., Barak, L.S., Nusse, R., and Lefkowitz, R.J. (2003). Dishevelled 2 recruits beta-arrestin 2 to mediate Wnt5A-stimulated endocytosis of Frizzled 4. Science *301*, 1391–1394.

Chen, W. V., Nwakeze, C.L., Denny, C.A., O'Keeffe, S., Rieger, M.A., Mountoufaris, G., Kirner, A., Dougherty, J.D., Hen, R., Wu, Q., et al. (2017). Pcdhαc2 is required for axonal tiling and assembly of serotonergic circuitries in mice. Science *356*, 406–411.

Chenn, A., and McConnell, S.K. (1995). Cleavage orientation and the asymmetric inheritance of notchl immunoreactivity in mammalian neurogenesis. Cell *82*, 631–641.

Chenn, A., and Walsh, C.A. (2002). Regulation of cerebral cortical size by control of cell cycle exit in neural precursors. Science *297*, 365–369.

Chitnis, A., Henrique, D., Lewis, J., Ish-Horowicz, D., and Kintner, C. (1995). Primary neurogenesis in Xenopus embryos regulated by a homologue of the Drosophila neurogenic gene Delta. Nature *375*, 761–766.

Chodelkova, O., Masek, J., Korinek, V., Kozmik, Z., and Machon, O. (2018). Tcf7L2 is essential for neurogenesis in the developing mouse neocortex. Neural Dev. *13*, 1–10.

Cimadamore, F., Amador-Arjona, A., Chen, C., Huang, C.T., and Terskikh, A. V. (2013). SOX2-LIN28/let-7 pathway regulates proliferation and neurogenesis in neural precursors. Proc. Natl. Acad. Sci. U. S. A. *110*.

Clapcote, S.J., and Roder, J.C. (2005). Simplex PCR assay for sex determination in mice. Biotechniques.

Clevers, H., and Nusse, R. (2012). Wnt/β-catenin signaling and disease. Cell *149*, 1192–1205.

Cohen, B., Shimizu, M., Izrailit, J., Ng, N.F.L., Buchman, Y., Pan, J.G., Dering, J., and Reedijk, M. (2010). Cyclin D1 is a direct target of JAG1-mediated Notch signaling in breast cancer. Breast Cancer Res. Treat. *123*, 113–124.

Cooper, S.R., Jontes, J.D., and Sotomayor, M. (2016). Structural determinants of adhesion by protocadherin-19 and implications for its role in epilepsy. Elife *5*.

Cotton, A.M., Price, E.M., Jones, M.J., Balaton, B.P., Kobor, M.S., and Brown, C.J. (2015). Landscape of DNA methylation on the X chromosome reflects CpG density, functional chromatin state and X-chromosome inactivation. Hum. Mol. Genet. *24*, 1528–1539.

Coumailleau, F., Fürthauer, M., Knoblich, J.A., and González-Gaitán, M. (2009). Directional Delta and Notch trafficking in Sara endosomes during asymmetric cell division. Nat. 2009 4587241 *458*, 1051–1055.

Cruciat, C.M., Dolde, C., De Groot, R.E.A., Ohkawara, B., Reinhard, C., Korswagen, H.C., and Niehrs, C. (2013). RNA helicase DDX3 is a regulatory subunit of casein kinase 1 in Wnt-β-catenin signaling. Science *339*, 1436–1441.

Dallosso, A.R., Øster, B., Greenhough, A., Thorsen, K., Curry, T.J., Owen, C., Hancock, A.L., Szemes, M., Paraskeva, C., Frank, M., et al. (2012). Long-range epigenetic silencing of chromosome 5q31 protocadherins is involved in early and late stages of colorectal tumorigenesis through modulation of oncogenic pathways. Oncogene *31*, 4409–4419.

Dave, R.K., Ellis, T., Toumpas, M.C., Robson, J.P., Julian, E., Adolphe, C., Bartlett, P.F., Cooper, H.M., Reynolds, B.A., and Wainwright, B.J. (2011). Sonic Hedgehog and Notch Signaling Can Cooperate to Regulate Neurogenic Divisions of Neocortical Progenitors. PLoS One *6*, e14680.

Davidson, G., Wu, W., Shen, J., Bilic, J., Fenger, U., Stannek, P., Glinka, A., and Niehrs, C. (2005). Casein kinase 1 gamma couples Wnt receptor activation to cytoplasmic signal transduction. Nature *438*, 867–872.

Delaunay, D., Cortay, V., Patti, D., Knoblauch, K., and Dehay, C. (2014). Mitotic Spindle Asymmetry: A Wnt/PCP-Regulated Mechanism Generating Asymmetrical Division in Cortical Precursors. Cell Rep. *6*, 400–414.

Dell'Isola, G.B., Vinti, V., Fattorusso, A., Tascini, G., Mencaroni, E., Di Cara, G., Striano, P., and Verrotti, A. (2022). The Broad Clinical Spectrum of Epilepsies Associated With Protocadherin 19 Gene Mutation. Front. Neurol. *12*.

Depienne, C., Bouteiller, D., Keren, B., Cheuret, E., Poirier, K., Trouillard, O., Benyahia, B., Quelin, C., Carpentier, W., Julia, S., et al. (2009). Sporadic Infantile Epileptic Encephalopathy Caused by Mutations in PCDH19 Resembles Dravet Syndrome but Mainly Affects Females. PLOS Genet. *5*, e1000381.

Depienne, C., Trouillard, O., Bouteiller, D., Gourfinkel-An, I., Poirier, K., Rivier, F., Berquin, P., Nabbout, R., Chaigne, D., Steschenko, D., et al. (2011). Mutations and deletions in PCDH19 account for various familial or isolated epilepsies in females. Hum. Mutat. *32*, E1959-75.

Dermit, M., Dodel, M., and Mardakheh, F.K. (2017). Methods for monitoring and measurement of

protein translation in time and space. Mol. Biosyst. *13*, 2477.

Dessaud, E., McMahon, A.P., and Briscoe, J. (2008). Pattern formation in the vertebrate neural tube: a sonic hedgehog morphogen-regulated transcriptional network. Development *135*, 2489–2503.

Dibbens, L.M., Tarpey, P.S., Hynes, K., Bayly, M.A., Scheffer, I.E., Smith, R., Bomar, J., Sutton, E., Vandeleur, L., Shoubridge, C., et al. (2008). X-linked protocadherin 19 mutations cause female-limited epilepsy and cognitive impairment. Nat. Genet.

Ding, J., Adiconis, X., Simmons, S.K., Kowalczyk, M.S., Hession, C.C., Marjanovic, N.D., Hughes, T.K., Wadsworth, M.H., Burks, T., Nguyen, L.T., et al. (2020). Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. Nat. Biotechnol. 2020 386 *38*, 737–746.

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

Dong, Z., Yang, N., Yeo, S.Y., Chitnis, A., and Guo, S. (2012). Intralineage directional Notch signaling regulates self-renewal and differentiation of asymmetrically dividing radial glia. Neuron *74*, 65–78.

Le Dréau, G. (2022). BuMPing Into Neurogenesis: How the Canonical BMP Pathway Regulates Neural Stem Cell Divisions Throughout Space and Time. Front. Neurosci. *15*, 1908.

Echelard, Y., Epstein, D.J., St-Jacques, B., Shen, L., Mohler, J., McMahon, J.A., and McMahon, A.P. (1993). Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. Cell *75*, 1417–1430.

Eckler, M.J., Nguyen, T.D., McKenna, W.L., Fastow, B.L., Guo, C., Rubenstein, J.L.R., and Chen, B. (2015). Cux2-positive radial glial cells generate diverse subtypes of neocortical projection neurons and macroglia. Neuron *86*, 1100–1108.

Elias, L.A.B., and Kriegstein, A.R. (2008). Gap junctions: multifaceted regulators of embryonic cortical development. Trends Neurosci. *31*, 243–250.

Emond, M.R., Biswas, S., and Jontes, J.D. (2009). Protocadherin-19 is essential for early steps in brain morphogenesis. Dev. Biol. *334*, 72–83.

Emond, M.R., Biswas, S., Blevins, C.J., and Jontes, J.D. (2011). A complex of Protocadherin-19

and N-cadherin mediates a novel mechanism of cell adhesion. J. Cell Biol. *195*, 1115–1121.

Emond, M.R., Biswas, S., Morrow, M.L., and Jontes, J.D. (2021). Proximity-dependent Proteomics Reveals Extensive Interactions of Protocadherin-19 with Regulators of Rho GTPases and the Microtubule Cytoskeleton. Neuroscience *452*, 26–36.

Englund, C., Fink, A., Lau, C., Pham, D., Daza, R.A.M., Bulfone, A., Kowalczyk, T., and Hevner, R.F. (2005). Brief Communication Pax6, Tbr2, and Tbr1 Are Expressed Sequentially by Radial Glia, Intermediate Progenitor Cells, and Postmitotic Neurons in Developing Neocortex.

Espinós, A., Fernández-Ortuño, E., Negri, E., and Borrell, V. (2022). Evolution of genetic mechanisms regulating cortical neurogenesis. Dev. Neurobiol. *82*, 428–453.

Esumi, S., Kakazu, N., Taguchi, Y., Hirayama, T., Sasaki, A., Hirabayashi, T., Koide, T., Kitsukawa, T., Hamada, S., and Yagi, T. (2005). Monoallelic yet combinatorial expression of variable exons of the protocadherin-alpha gene cluster in single neurons. Nat. Genet. *37*, 171–176.

Eze, U.C., Bhaduri, A., Haeussler, M., Nowakowski, T.J., and Kriegstein, A.R. (2021). Single-cell atlas of early human brain development highlights heterogeneity of human neuroepithelial cells and early radial glia. Nat. Neurosci. 2021 244 *24*, 584–594.

Fan, L., Lu, Y., Shen, X., Shao, H., Suo, L., and Wu, Q. (2018). Alpha protocadherins and Pyk2 kinase regulate cortical neuron migration and cytoskeletal dynamics via Rac1 GTPase and WAVE complex in mice. Elife *7*.

Fang, H., Disteche, C.M., and Berletch, J.B. (2019). X Inactivation and Escape: Epigenetic and Structural Features. Front. Cell Dev. Biol. *7*, 219.

Feng, Y., and Walsh, C.A. (2004). Mitotic Spindle Regulation by Nde1 Controls Cerebral Cortical Size. Neuron *44*, 279–293.

Fernández, V., Llinares-Benadero, C., and Borrell, V. (2016). Cerebral cortex expansion and folding: what have we learned? EMBO J. *35*, 1021–1044.

Ferrer-Vaquer, A., Piliszek, A., Tian, G., Aho, R.J., Dufort, D., and Hadjantonakis, A.-K. (2010). A sensitive and bright single-cell resolution live imaging reporter of Wnt/ß-catenin signaling in the mouse.

Fish, J.L., Kosodo, Y., Enard, W., Pääbo, S., and Huttner, W.B. (2006). Aspm specifically

maintains symmetric proliferative divisions of neuroepithelial cells. Proc. Natl. Acad. Sci. U. S. A. *103*, 10438–10443.

Flaherty, E., and Maniatis, T. (2020). The Role of Clustered Protocadherins in Neurodevelopment and Neuropsychiatric Diseases HHS Public Access. Curr Opin Genet Dev *65*, 144–150.

Fradkin, L.G., Dura, J.M., and Noordermeer, J.N. (2010). Ryks: new partners for Wnts in the developing and regenerating nervous system. Trends Neurosci. *33*, 84–92.

Franco, S.J., Gil-Sanz, C., Martinez-Garay, I., Espinosa, A., Harkins-Perry, S.R., Ramos, C., and Müller, U. (2012). Fate-restricted neural progenitors in the mammalian cerebral cortex. Science (80-.). *337*, 746–749.

Fujitani, M., Zhang, S., Fujiki, R., Fujihara, Y., and Yamashita, T. (2017). A chromosome 16p13.11 microduplication causes hyperactivity through dysregulation of miR-484/protocadherin-19 signaling. Mol. Psychiatry *22*, 364–374.

Furuta, Y., Piston, D.W., and Hogan, B.L.M. (1997). Bone morphogenetic proteins (BMPs) as regulators of dorsal forebrain development. Development *124*, 2203–2212.

Gai, M., Bianchi, F.T., Vagnoni, C., Vernì, F., Bonaccorsi, S., Pasquero, S., Berto, G.E., Sgrò, F., Chiotto, A.A., Annaratone, L., et al. (2017). ASPM and CITK regulate spindle orientation by affecting the dynamics of astral microtubules. EMBO Rep. *18*, 1870–1870.

Gaiano, N., Nye, J.S., and Fishell, G. (2000). Radial glial identity is promoted by Notch1 signaling in the murine forebrain. Neuron *26*, 395–404.

Gaitan, Y., and Bouchard, M. (2006). Expression of the δ-protocadherin gene Pcdh19 in the developing mouse embryo. Gene Expr. Patterns.

Galceran, J., Miyashita-Lin, E.M., Devaney, E., Rubenstein, J.L.R., and Grosschedl, R. (2000). Hippocampus development and generation of dentate gyrus granule cells is regulated by LEF1. Development *127*, 469–482.

Galindo-Riera, N., Adriana Newbold, S., Sledziowska, M., Llinares-Benadero, C., Griffiths, J., Mire, E., and Martinez-Garay, I. (2021). Cellular and behavioral characterization of PCDH19 mutant mice: subtle molecular changes, increased exploratory behavior and an impact of social environment. ENeuro *8*.

Gänzler-Odenthal, S.I.I., and Redies, C. (1998). Blocking N-Cadherin Function Disrupts the

Epithelial Structure of Differentiating Neural Tissue in the Embryonic Chicken Brain. J. Neurosci. *18*, 5415–5425.

Gao, P., Postiglione, M.P., Krieger, T.G., Hernandez, L., Wang, C., Han, Z., Streicher, C., Papusheva, E., Insolera, R., Chugh, K., et al. (2014). Deterministic progenitor behavior and unitary production of neurons in the neocortex. Cell *159*, 775–788.

García-Moreno, F., and Molnár, Z. (2015). Subset of early radial glial progenitors that contribute to the development of callosal neurons is absent from avian brain. Proc. Natl. Acad. Sci. U. S. A. *112*, E5058–E5067.

Gaspard, N., Bouschet, T., Hourez, R., Dimidschstein, J., Naeije, G., Van Den Ameele, J., Espuny-Camacho, I., Herpoel, A., Passante, L., Schiffmann, S.N., et al. (2008). An intrinsic mechanism of corticogenesis from embryonic stem cells. Nature *455*, 351–357.

Gerosa, L., Francolini, M., Bassani, S., and Passafaro, M. (2019). The Role of Protocadherin 19 (PCDH19) in Neurodevelopment and in the Pathophysiology of Early Infantile Epileptic Encephalopathy-9 (EIEE9). Dev. Neurobiol.

Gerosa, L., Mazzoleni, S., Rusconi, F., Longaretti, A., Lewerissa, E., Pelucchi, S., Murru, L., Giannelli, S.G., Broccoli, V., Marcello, E., et al. (2022). The epilepsy-associated protein PCDH19 undergoes NMDA receptor-dependent proteolytic cleavage and regulates the expression of immediate-early genes. Cell Rep. *39*.

Gil-Sanz, C., Espinosa, A., Fregoso, S.P., Bluske, K.K., Cunningham, C.L., Martinez-Garay, I., Zeng, H., Franco, S.J., and Müller, U. (2015). Lineage Tracing Using Cux2-Cre and Cux2-CreERT2 Mice. Neuron *86*, 1091–1099.

Goetz, R., and Mohammadi, M. (2013). Exploring mechanisms of FGF signalling through the lens of structural biology. Nat. Rev. Mol. Cell Biol. *14*, 166.

Gómez-Orte, E., Sáenz-Narciso, B., Moreno, S., and Cabello, J. (2013). Multiple functions of the noncanonical Wnt pathway. Trends Genet. *29*, 545–553.

Goodman, K.M., Rubinstein, R., Thu, C.A., Bahna, F., Mannepalli, S., Ahlsén, G., Rittenhouse, C., Maniatis, T., Honig, B., and Shapiro, L. (2016). Structural Basis of Diverse Homophilic Recognition by Clustered α- and β-Protocadherins. Neuron *90*, 709–723.

Goodman, K.M., Rubinstein, R., Dan, H., Bahna, F., Mannepalli, S., Ahlsén, G., Thu, C.A., Sampogna, R. V., Maniatis, T., Honig, B., et al. (2017). Protocadherin cis-dimer architecture and

recognition unit diversity. Proc. Natl. Acad. Sci. U. S. A. *114*, E9829–E9837.

Goodman, K.M., Katsamba, P.S., Rubinstein, R., Ahlsén, G., Bahna, F., Mannepalli, S., Dan, H., Sampogna, R. V., Shapiro, L., and Honig, B. (2022). How clustered protocadherin binding specificity is tuned for neuronal self-/ nonself-recognition. Elife *11*.

Götz, M., and Huttner, W.B. (2005). The cell biology of neurogenesis. Nat. Rev. Mol. Cell Biol. *6*, 777–788.

Götz, M., Stoykova, A., and Gruss, P. (1998). Pax6 controls radial glia differentiation in the cerebral cortex. Neuron *21*, 1031–1044.

Götz, M., Sirko, S., Beckers, J., and Irmler, M. (2015). Reactive astrocytes as neural stem or progenitor cells: In vivo lineage, In vitro potential, and Genome-wide expression analysis. Glia *63*, 1452–1468.

Grandori, C., Gomez-Roman, N., Felton-Edkins, Z.A., Ngouenet, C., Galloway, D.A., Eisenman, R.N., and White, R.J. (2005). c-Myc binds to human ribosomal DNA and stimulates transcription of rRNA genes by RNA polymerase I. Nat. Cell Biol. 2005 73 *7*, 311–318.

Green, J., Nusse, R., and van Amerongen, R. (2014). The Role of Ryk and Ror Receptor Tyrosine Kinases in Wnt Signal Transduction. Cold Spring Harb. Perspect. Biol. *6*.

Guarnieri, F.C., De Chevigny, A., Falace, A., and Cardoso, C. (2022). Disorders of neurogenesis and cortical development. Https://Doi.Org/10.31887/DCNS.2018.20.4/Ccardoso *20*, 255–266.

Guo, C., Eckler, M.J., McKenna, W.L., McKinsey, G.L., Rubenstein, J.L.R., and Chen, B. (2013). Fezf2 expression identifies a multipotent progenitor for neocortical projection neurons, astrocytes, and oligodendrocytes. Neuron *80*, 1167–1174.

Hadjantonakis, A.K., Gertsenstein, M., Ikawa, M., Okabe, M., and Nagy, A. (1998). Non-invasive sexing of preimplantation stage mammalian embryos [2]. Nat. Genet. *19*, 220–222.

El Hajj, N., Dittrich, M., and Haaf, T. (2017). Epigenetic dysregulation of protocadherins in human disease. Semin. Cell Dev. Biol. *69*, 172–182.

Han, Y., Gao, S., Muegge, K., Zhang, W., and Zhou, B. (2015). Advanced Applications of RNA Sequencing and Challenges. Bioinform. Biol. Insights *9*, 29.

Hanashima, C., Shen, L., Li, S.C., and Lai, E. (2002). Brain Factor-1 Controls the Proliferation and Differentiation of Neocortical Progenitor Cells through Independent Mechanisms. J. Neurosci.

22, 6526–6536.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573-3587.e29.

Haque, A., Engel, J., Teichmann, S.A., and Lönnberg, T. (2017). A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. Genome Med. *9*, 1–12.

Harrison-Uy, S.J., and Pleasure, S.J. (2012). Wnt Signaling and Forebrain Development. Cold Spring Harb. Perspect. Biol. *4*, 1–11.

Harrison, O.J., Bahna, F., Katsamba, P.S., Jin, X., Brasch, J., Vendome, J., Ahlsen, G., Carroll, K.J., Price, S.R., Honig, B., et al. (2010). Two-step adhesive binding by classical cadherins. Nat. Struct. Mol. Biol. 2010 173 *17*, 348–357.

Harrison, O.J., Brasch, J., Katsamba, P.S., Ahlsen, G., Noble, A.J., Dan, H., Sampogna, R. V., Potter, C.S., Carragher, B., Honig, B., et al. (2020). Family-wide Structural and Biophysical Analysis of Binding Interactions among Non-clustered δ-Protocadherins. Cell Rep. *30*, 2655-2671.e7.

Harrison, S.J., Nishinakamura, R., Jones, K.R., and Monaghan, A.P. (2012). Sall1 regulates cortical neurogenesis and laminar fate specification in mice: Implications for neural abnormalities in Townes-Brocks syndrome. DMM Dis. Model. Mech. *5*, 351–365.

Hasenpusch-Theil, K., Magnani, D., Amaniti, E.M., Han, L., Armstrong, D., and Theil, T. (2012). Transcriptional analysis of Gli3 mutants identifies Wnt target genes in the developing hippocampus. Cereb. Cortex *22*, 2878–2893.

Hatakeyama, J., Bessho, Y., Katoh, K., Ookawara, S., Fujioka, M., Guillemot, F., and Kageyama, R. (2004). Hes genes regulate size, shape and histogenesis of the nervous system by control of the timing of neural stem cell differentiation. Development *131*, 5539–5550.

Hatakeyama, J., Wakamatsu, Y., Nagafuchi, A., Kageyama, R., Shigemoto, R., and Shimamura, K. (2014). Cadherin-based adhesions in the apical endfoot are required for active Notch signaling to control neurogenesis in vertebrates. Dev.

Hatta, K., and Takeichi, M. (1986). Expression of N-cadherin adhesion molecules associated with early morphogenetic events in chick development. Nature *320*, 447–449.

Haÿ, E., Laplantine, E., Geoffroy, V., Frain, M., Kohler, T., Müller, R., and Marie, P.J. (2009). N-cadherin interacts with axin and LRP5 to negatively regulate Wnt/beta-catenin signaling, osteoblast function, and bone formation. Mol. Cell. Biol. *29*, 953–964.

Hayashi, S., and Takeichi, M. (2015). Emerging roles of protocadherins: From self-avoidance to enhancement of motility. J. Cell Sci. *128*, 1455–1464.

Hayashi, S., Inoue, Y., Hattori, S., Kaneko, M., Shioi, G., Miyakawa, T., and Takeichi, M. (2017). Loss of X-linked Protocadherin-19 differentially affects the behavior of heterozygous female and hemizygous male mice. Sci. Reports 2017 71 *7*, 1–15.

Henrique, D., Hirsinger, E., Adam, J., Roux, I. Le, Pourquié, O., Ish-Horowicz, D., and Lewis, J. (1997). Maintenance of neuroepithelial progenitor cells by Delta-Notch signalling in the embryonic chick retina. Curr. Biol. *7*, 661–670.

Herbert, Z.T., Kershner, J.P., Butty, V.L., Thimmapuram, J., Choudhari, S., Alekseyev, Y.O., Fan, J., Podnar, J.W., Wilcox, E., Gipson, J., et al. (2018). Cross-site comparison of ribosomal depletion kits for Illumina RNAseq library construction. BMC Genomics *19*.

Herculano-Houzel, S. (2009). The human brain in numbers: a linearly scaled-up primate brain. Front. Hum. Neurosci. *3*.

Hertel, N., and Redies, C. (2011). Absence of Layer-Specific Cadherin Expression Profiles in the Neocortex of the Reeler Mutant Mouse. Cereb. Cortex *21*, 1105–1117.

Hicks, S.C., Townes, F.W., Teng, M., and Irizarry, R.A. (2018). Missing data and technical variability in single-cell RNA-sequencing experiments. Biostatistics *19*, 562–578.

Higurashi, N., Takahashi, Y., Kashimada, A., Sugawara, Y., Sakuma, H., Tomonoh, Y., Inoue, T., Hoshina, M., Satomi, R., Ohfu, M., et al. (2015). Immediate suppression of seizure clusters by corticosteroids in PCDH19 female epilepsy. Seizure *27*, 1–5.

Hirabayashi, Y., Itoh, Y., Tabata, H., Nakajima, K., Akiyama, T., Masuyama, N., and Gotoh, Y. (2004). The Wnt/beta-catenin pathway directs neuronal differentiation of cortical neural precursor cells. Development *131*, 2791–2801.

Hirabayashi, Y., Suzki, N., Tsuboi, M., Endo, T.A., Toyoda, T., Shinga, J., Koseki, H., Vidal, M., and Gotoh, Y. (2009). Polycomb limits the neurogenic competence of neural precursor cells to promote astrogenic fate transition. Neuron *63*, 600–613.

Hirano, S., and Takeichi, M. (2012). Cadherins in Brain Morphogenesis and Wiring. Physiol Rev *92*, 597–634.

Hoch, R. V., Lindtner, S., Price, J.D., and Rubenstein, J.L.R. (2015). OTX2 Transcription Factor Controls Regional Patterning Within The Medial Ganglionic Eminence (MGE) And Regional Identity Of The Septum. Cell Rep. *12*, 482.

Homan, C.C., Pederson, S., To, T.H., Tan, C., Piltz, S., Corbett, M.A., Wolvetang, E., Thomas, P.Q., Jolly, L.A., and Gecz, J. (2018). PCDH19 regulation of neural progenitor cell differentiation suggests asynchrony of neurogenesis as a mechanism contributing to PCDH19 Girls Clustering Epilepsy. Neurobiol. Dis.

Hong, M., Tao, S., Zhang, L., Diao, L.T., Huang, X., Huang, S., Xie, S.J., Xiao, Z.D., and Zhang, H. (2020). RNA sequencing: new technologies and applications in cancer research. J. Hematol. Oncol. 2020 131 *13*, 1–16.

Hoshina, N., Johnson-Venkatesh, E.M., Hoshina, M., and Umemori, H. (2021). Female-specific synaptic dysfunction and cognitive impairment in a mouse model of PCDH19 disorder. Science (80-.). *372*.

Hoye, M.L., Calviello, L., Poff, A.J., Ejimogu, N.E., Newman, C.R., Montgomery, M.D., Ou, J., Floor, S.N., and Silver, D.L. (2022). Aberrant cortical development is driven by impaired cell cycle and translational control in a DDX3X syndrome model. Elife *11*.

Hrdlickova, R., Toloue, M., and Tian, B. (2017). RNA-Seq methods for transcriptome analysis. Wiley Interdiscip. Rev. RNA *8*, e1364.

Hsu, L.C.L., Nam, S., Cui, Y., Chang, C.P., Wang, C.F., Kuo, H.C., Touboul, J.D., and Chou, S.J. (2015). Lhx2 regulates the timing of β-catenin-dependent cortical neurogenesis. Proc. Natl. Acad. Sci. U. S. A. *112*, 12199–12204.

Hu, D.J.K., Baffet, A.D., Nayak, T., Akhmanova, A., Doye, V., and Vallee, R.B. (2013). Dynein recruitment to nuclear pores activates apical nuclear migration and mitotic entry in brain progenitor cells. Cell *154*, 1300.

Hulpiau, P., and van Roy, F. (2009). Molecular evolution of the cadherin superfamily. Int. J. Biochem. Cell Biol. *41*, 349–369.

Ibarra-Soria, X., Jawaid, W., Pijuan-Sala, B., Ladopoulos, V., Scialdone, A., Jörg, D.J., Tyser, R.C. V, Calero-Nieto, F.J., Mulas, C., Nichols, J., et al. (2018). Defining murine organogenesis at

single-cell resolution reveals a role for the leukotriene pathway in regulating blood progenitor formation. Nat. Cell Biol. *20*.

Ilicic, T., Kim, J.K., Kolodziejczyk, A.A., Bagger, F.O., McCarthy, D.J., Marioni, J.C., and Teichmann, S.A. (2016). Classification of low quality cells from single-cell RNA-seq data. Genome Biol. *17*.

Imayoshi, I., Isomura, A., Harima, Y., Kawaguchi, K., Kori, H., Miyachi, H., Fujiwara, T., Ishidate, F., and Kageyama, R. (2013). Oscillatory control of factors determining multipotency and fate in mouse neural progenitors. Science *342*, 1203–1208.

Inoue, T., Oz, H.S., Wiland, D., Gharib, S., Deshpande, R., Hill, R.J., Katz, W.S., and Sternberg, P.W. (2004). C. elegans LIN-18 is a Ryk ortholog and functions in parallel to LIN-17/Frizzled in Wnt signaling. Cell *118*, 795–806.

Inoue, T., Ota, M., Ogawa, M., Mikoshiba, K., and Aruga, J. (2007). Zic1 and Zic3 Regulate Medial Forebrain Development through Expansion of Neuronal Progenitors. J. Neurosci. *27*, 5461.

Islam, S., Kjällquist, U., Moliner, A., Zajac, P., Fan, J.B., Lönnerberg, P., and Linnarsson, S. (2011). Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. Genome Res. *21*, 1160–1167.

Juberg, R.C., and Hellman, C.D. (1971). A new familial form of convulsive disorder and mental retardation limited to females. J. Pediatr. *79*, 726–732.

Kalita, A., Gupta, S., Singh, P., Surolia, A., and Banerjee, K. (2013). IGF-1 stimulated upregulation of cyclin D1 is mediated via STAT5 signaling pathway in neuronal cells. IUBMB Life *65*, 462–471.

Kang, H.M., Subramaniam, M., Targ, S., Nguyen, M., Maliskova, L., McCarthy, E., Wan, E., Wong, S., Byrnes, L., Lanata, C.M., et al. (2017). Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. Nat. Biotechnol. 2017 361 *36*, 89–94.

Kang, J.S., Zhang, W., and Krauss, R.S. (2007). Hedgehog signaling: cooking with Gas1. Sci. STKE *2007*.

Kaplan, E.S., Ramos-Laguna, K.A., Mihalas, A.B., Daza, R.A.M., and Hevner, R.F. (2017). Neocortical Sox9+ radial glia generate glutamatergic neurons for all layers, but lack discernible evidence of early laminar fate restriction. Neural Dev. *12*, 1–10.

Katzmann, D.J., Odorizzi, G., and Emr, S.D. (2002). Receptor downregulation and multivesicular-

body sorting. Nat. Rev. Mol. Cell Biol. *3*, 893–905.

Kawauchi, T., Sekine, K., Shikanai, M., Chihama, K., Tomita, K., Kubo, K. ichiro, Nakajima, K., Nabeshima, Y. ichi, and Hoshino, M. (2010). Rab GTPases-Dependent Endocytic Pathways Regulate Neuronal Migration and Maturation through N-Cadherin Trafficking. Neuron *67*, 588–602.

Keeble, T.R., and Cooper, H.M. (2006). Ryk: a novel Wnt receptor regulating axon pathfinding. Int. J. Biochem. Cell Biol. *38*, 2011–2017.

Kelly, R.T. (2020). Single-cell Proteomics: Progress and Prospects. Mol. Cell. Proteomics *19*, 1739–1748.

Kharchenko, P. V., Silberstein, L., and Scadden, D.T. (2014). Bayesian approach to single-cell differential expression analysis. Nat. Methods 2014 117 *11*, 740–742.

Kiecker, C., and Niehrs, C. (2001). A morphogen gradient of Wnt/beta-catenin signalling regulates anteroposterior neural patterning in Xenopus. Development *128*, 4189–4201.

Kim, S.Y., Chung, H.S., Sun, W., and Kim, H. (2007). Spatiotemporal expression pattern of non-clustered protocadherin family members in the developing rat brain. Neuroscience *147*, 996–1021.

Kim, S.Y., Yasuda, S., Tanaka, H., Yamagata, K., and Kim, H. (2011). Non-clustered protocadherin. Cell Adh. Migr. *5*, 97–105.

Kirn-Safran, C.B., Oristian, D.S., Focht, R.J., Parker, S.G., Vivian, J.L., and Carson, D.D. (2007). Global growth deficiencies in mice lacking the ribosomal protein HIP/RPL29. Dev. Dyn. *236*, 447–460.

Kishi, Y., Fujii, Y., Hirabayashi, Y., and Gotoh, Y. (2012). HMGA regulates the global chromatin state and neurogenic potential in neocortical precursor cells. Nat. Neurosci. *15*, 1127–1133.

Klingler, E., and Jabaudon, D. (2020). Cortical development: Do progenitors play dice? Elife *9*.

Knoepfler, P.S., Cheng, P.F., and Eisenman, R.N. (2002). N-myc is essential during neurogenesis for the rapid expansion of progenitor cell populations and the inhibition of neuronal differentiation. Genes Dev. *16*, 2699–2712.

Kolc, K.L., Sadleir, L.G., Scheffer, I.E., Ivancevic, A., Roberts, R., Pham, D.H., and Gecz, J. (2019). A systematic review and meta-analysis of 271 PCDH19-variant individuals identifies

psychiatric comorbidities, and association of seizure onset and disease severity. Mol. Psychiatry *24*, 241–251.

Kolc, K.L., Møller, R.S., Sadleir, L.G., Scheffer, I.E., Kumar, R., and Gecz, J. (2020). PCDH19 Pathogenic Variants in Males: Expanding the Phenotypic Spectrum. Adv. Exp. Med. Biol. *1298*, 177–187.

Konno, D., Shioi, G., Shitamukai, A., Mori, A., Kiyonari, H., Miyata, T., and Matsuzaki, F. (2007). Neuroepithelial progenitors undergo LGN-dependent planar divisions to maintain self-renewability during mammalian neurogenesis. Nat. Cell Biol. 2008 101 *10*, 93–101.

Kosodo, Y., Röper, K., Haubensak, W., Marzesco, A.M., Corbeil, D., and Huttner, W.B. (2004). Asymmetric distribution of the apical plasma membrane during neurogenic divisions of mammalian neuroepithelial cells. EMBO J. *23*, 2314–2324.

Kowalczyk, T., Pontious, A., Englund, C., Daza, R.A.M., Bedogni, F., Hodge, R., Attardo, A., Bell, C., Huttner, W.B., and Hevner, R.F. (2009). Intermediate Neuronal Progenitors (Basal Progenitors) Produce Pyramidal–Projection Neurons for All Layers of Cerebral Cortex. Cereb. Cortex *19*, 2439–2450.

Kressmann, S., Campos, C., Castanon, I., Fürthauer, M., and González-Gaitán, M. (2015). Directional Notch trafficking in Sara endosomes during asymmetric cell division in the spinal cord. Nat. Cell Biol. *17*, 333–339.

Krishna-K, K., Hertel, N., and Redies, C. (2011). Cadherin expression in the somatosensory cortex: evidence for a combinatorial molecular code at the single-cell level. Neuroscience *175*, 37–48.

Kuwayama, N., Kishi, Y., Maeda, Y., Nishiumi, Y., Suzuki, Y., Koseki, H., Hirabayashi, Y., and Gotoh, Y. (2021). A role for Hmga2 in the early-stage transition of neural stem-progenitor cell properties during mouse neocortical development. BioRxiv 2020.05.14.086330.

Lacomme, M., Liaubet, L., Pituello, F., and Bel-Vialar, S. (2012). NEUROG2 Drives Cell Cycle Exit of Neuronal Precursors by Specifically Repressing a Subset of Cyclins Acting at the G1 and S Phases of the Cell Cycle. Mol. Cell. Biol. *32*, 2596.

Lancaster, M.A., Renner, M., Martin, C.A., Wenzel, D., Bicknell, L.S., Hurles, M.E., Homfray, T., Penninger, J.M., Jackson, A.P., and Knoblich, J.A. (2013). Cerebral organoids model human brain development and microcephaly. Nature *501*, 373–379.

Lange, C., and Calegari, F. (2010). Cdks and cyclins link G1 length and differentiation of embryonic, neural and hematopoietic stem cells. Https://Doi.Org/10.4161/Cc.9.10.11598 *9*, 1893–1900.

de Lange, I.M., Rump, P., Neuteboom, R.F., Augustijn, P.B., Hodges, K., Kistemaker, A.I., Brouwer, O.F., Mancini, G.M.S., Newman, H.A., Vos, Y.J., et al. (2017). Male patients affected by mosaic PCDH19 mutations: five new cases. Neurogenetics *18*, 147–153.

Lawson, A., Anderson, H., and Schoenwolf, G.C. (2001). Cellular Mechanisms of Neural Fold Formation and Morphogenesis in the Chick Embryo.

Lehtinen, M.K., Zappaterra, M.W., Chen, X., Yang, Y.J., Hill, A.D., Lun, M., Maynard, T., Gonzalez, D., Kim, S., Ye, P., et al. (2011). The cerebrospinal fluid provides a proliferative niche for neural progenitor cells. Neuron *69*, 893–905.

Leone, D.P., Srinivasan, K., Chen, B., Alcamo, E., and McConnell, S.K. (2008). The determination of projection neuron identity in the developing cerebral cortex. Curr. Opin. Neurobiol. *18*, 28–35.

Li, W., Cogswell, C.A., and LoTurco, J.J. (1998). Neuronal Differentiation of Precursors in the Neocortical Ventricular Zone Is Triggered by BMP. J. Neurosci. *18*, 8853–8862.

Li, X., Gordon, P.J., Gaynes, J.A., Fuller, A.W., Ringuette, R., Santiago, C.P., Wallace, V., Blackshaw, S., Li, P., and Levine, E.M. (2022). Lhx2 is a progenitor-intrinsic modulator of Sonic Hedgehog signaling during early retinal neurogenesis. Elife *11*.

Liang, D., Aygün, N., Matoba, N., Ideraabdullah, F.Y., Love, M.I., and Stein, J.L. (2022). Inference of putative cell-type-specific imprinted regulatory elements and genes during human neuronal differentiation. Hum. Mol. Genet.

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

Lim, L., Mi, D., Llorca, A., and Marín, O. (2018). Development and functional diversification of cortical interneurons. Neuron *100*, 294.

Linask, K.K., Knudsen, K.A., and Gui, Y.H. (1997). N-cadherin-catenin interaction: necessary component of cardiac cell compartmentalization during early vertebrate heart development. Dev. Biol. *185*, 148–164.

Liu, K., Lin, Q., Wei, Y., He, R., Shao, X., Ding, Z., Zhang, J., Zhu, M., Weinstein, L.S., Hong, Y.,

et al. (2015). Gαs regulates asymmetric cell division of cortical progenitors by controlling Numb mediated Notch signaling suppression. Neurosci. Lett. *597*, 97–103.

Llinares-Benadero, C., and Borrell, V. (2019). Deconstructing cortical folding: genetic, cellular and mechanical determinants. Nat. Rev. Neurosci. 2018 203 *20*, 161–176.

Llorca, A., Ciceri, G., Beattie, R., Wong, F.K., Diana, G., Serafeimidou-Pouliou, E., Fernández-Otero, M., Streicher, C., Arnold, S.J., Meyer, M., et al. (2019). A stochastic framework of neurogenesis underlies the assembly of neocortical cytoarchitecture. Elife *8*.

Loi, D.S.C., Yu, L., and Wu, A.R. (2021). Effective ribosomal RNA depletion for single-cell total RNA-seq by scDASH. PeerJ *9*.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014 1512 *15*, 1–21.

Lu, W., Yamamoto, V., Ortega, B., and Baltimore, D. (2004). Mammalian Ryk is a Wnt coreceptor required for stimulation of neurite outgrowth. Cell *119*, 97–108.

De Luca, A., Cerrato, V., Fucà, E., Parmigiani, E., Buffo, A., and Leto, K. (2016). Sonic hedgehog patterning during cerebellar development. Cell. Mol. Life Sci. *73*, 291–303.

Luecken, M.D., and Theis, F.J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. Mol. Syst. Biol. *15*.

Lv, X., Ren, S.Q., Zhang, X.J., Shen, Z., Ghosh, T., Xianyu, A., Gao, P., Li, Z., Lin, S., Yu, Y., et al. (2019). TBR2 coordinates neurogenesis expansion and precise microcircuit organization via Protocadherin 19 in the mammalian cortex. Nat. Commun. 2019 101 *10*, 1–15.

Lyon, M.F. (1961). Gene action in the X-chromosome of the mouse (Mus musculus L.). Nature *190*, 372–373.

Lyu, J., Yamamoto, V., and Lu, W. (2008). Cleavage of the Wnt Receptor Ryk Regulates Neuronal Differentiation during Cortical Neurogenesis. Dev. Cell *15*, 773–780.

MacDonald, B.T., Yokota, C., Tamai, K., Zeng, X., and He, X. (2008). Wnt signal amplification via activity, cooperativity, and regulation of multiple intracellular PPPSP motifs in the Wnt co-receptor LRP6. J. Biol. Chem. *283*, 16115–16123.

Macheda, M.L., Sun, W.W., Kugathasan, K., Hogan, B.M., Bower, N.I., Halford, M.M., Zhang, Y.F., Jacques, B.E., Lieschke, G.J., Dabdoub, A., et al. (2012). The Wnt receptor Ryk plays a role

in mammalian planar cell polarity signaling. J. Biol. Chem. *287*, 29312–29323.

Machon, O., Van Den Bout, C.J., Backman, M., Kemler, R., and Krauss, S. (2003). Role of β-catenin in the developing cortical and hippocampal neuroepithelium. Neuroscience *122*, 129–143.

Mah, K.M., Houston, D.W., and Weiner, J.A. (2016). The γ-Protocadherin-C3 isoform inhibits canonical Wnt signalling by binding to and stabilizing Axin1 at the membrane. Sci. Rep. *6*.

Maher, M.T., Flozak, A.S., Stocker, A.M., Chenn, A., and Gottardi, C.J. (2009). Activity of the beta-catenin phosphodestruction complex at cell-cell contacts is enhanced by cadherin-based adhesion. J. Cell Biol. *186*, 219–228.

Mallamaci, A., and Stoykova, A. (2006). Gene networks controlling early cerebral cortex arealization. Eur. J. Neurosci. *23*, 847–856.

Manuel, M.N., Mi, D., Masonand, J.O., and Price, D.J. (2015). Regulation of cerebral cortical neurogenesis by the Pax6 transcription factor. Front. Cell. Neurosci. *9*, 70.

Mao, J., Wang, J., Liu, B., Pan, W., Farr, G.H., Flynn, C., Yuan, H., Takada, S., Kimelman, D., Li, L., et al. (2001). Low-density lipoprotein receptor-related protein-5 binds to Axin and regulates the canonical Wnt signaling pathway. Mol. Cell *7*, 801–809.

Martínez-Cerdeño, V., Noctor, S.C., and Kriegstein, A.R. (2006). The Role of Intermediate Progenitor Cells in the Evolutionary Expansion of the Cerebral Cortex. Cereb. Cortex *16*, i152–i161.

Martinez-Garay, I., Gil-Sanz, C., Franco, S.J., Espinosa, A., Molnár, Z., and Mueller, U. (2016). Cadherin 2/4 signaling via PTP1B and catenins is crucial for nucleokinesis during radial neuronal migration in the neocortex. Dev. *143*, 2121–2134.

Martynoga, B., Morrison, H., Price, D.J., and Mason, J.O. (2005). Foxg1 is required for specification of ventral telencephalon and region-specific regulation of dorsal telencephalic precursor proliferation and apoptosis. Dev. Biol. *283*, 113–127.

McKenzie, M.G., Cobbs, L. V., Dummer, P.D., Petros, T.J., Halford, M.M., Stacker, S.A., Zou, Y., Fishell, G.J., and Au, E. (2019). Non-canonical Wnt Signaling through Ryk Regulates the Generation of Somatostatin- and Parvalbumin-Expressing Cortical Interneurons. Neuron *103*, 853-864.e4.

McShane, S.G., Molè, M.A., Savery, D., Greene, N.D.E., Tam, P.P.L., and Copp, A.J. (2015).

Cellular basis of neuroepithelial bending during mouse spinal neural tube closure. Dev. Biol. , 404 Pp. 113-124. *404*, 113–124.

Miyamoto, Y., Sakane, F., and Hashimoto, K. (2015). N-cadherin-based adherens junction regulates the maintenance, proliferation, and differentiation of neural progenitor cells during development. Cell Adh. Migr. *9*, 183–192.

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

Molnár, Z. (2011). Evolution of Cerebral Cortical Development. Brain. Behav. Evol. *78*, 94–107.

Molnár, Z., and Clowry, G. (2012). Cerebral cortical development in rodents and primates. In Progress in Brain Research, p.

Moncayo, J.A., Vargas, M.N., Castillo, I., Granda, P. V, Duque, A.M., Argudo, J.M., Matcheswalla, S., Lopez Dominguez, G.E., Monteros, G., Andrade, A.F., et al. (2022). Adjuvant Treatment for Protocadherin 19 (PCDH19) Syndrome. Cureus *14*.

Montanaro, L., Treré, D., and Derenzini, M. (2012). Changes in ribosome biogenesis may induce cancer by down-regulating the cell tumor suppressor potential. Biochim. Biophys. Acta *1825*, 101–110.

Morrow, E.M., Yoo, S.Y., Flavell, S.W., Kim, T.K., Lin, Y., Hill, R.S., Mukaddes, N.M., Balkhy, S., Gascon, G., Hashmi, A., et al. (2008). Identifying autism loci and genes by tracing recent shared ancestry. Science (80-.). *321*, 218–223.

Mulligan, K.A., and Cheyette, B.N.R. (2012). Wnt Signaling in Vertebrate Neural Development and Function. J. Neuroimmune Pharmacol. *7*, 774.

Munji, R.N., Choe, Y., Li, G., Siegenthaler, J.A., and Pleasure, S.J. (2011). Wnt Signaling Regulates Neuronal Differentiation of Cortical Intermediate Progenitors. J. Neurosci. *31*, 1676–1687.

Munsky, B., Neuert, G., and Van Oudenaarden, A. (2012). Using Gene Expression Noise to Understand Gene Regulation. Science *336*, 183.

Mutch, C.A., Schulte, J.D., Olson, E., and Chenn, A. (2010). Beta-Catenin Signaling Negatively Regulates Intermediate Progenitor Population Numbers in the Developing Cortex. PLoS One *5*,

e12376.

Nadarajah, B., Alifragis, P., Wong, R.O.L., and Parnavelas, J.G. (2003). Neuronal migration in the developing cerebral cortex: observations based on real-time imaging. Cereb. Cortex *13*, 607–611.

Nayak, R., and Hasija, Y. (2021). A hitchhiker's guide to single-cell transcriptomics and data analysis pipelines. Genomics *113*, 606–619.

Nelson, W.J. (2008). Regulation of cell-cell adhesion by the cadherin-catenin complex. Biochem. Soc. Trans. *36*, 149–155.

Niehrs, C., and Acebron, S.P. (2012). Mitotic and mitogenic Wnt signalling. EMBO J. *31*, 2705–2713.

Niehrs, C., and Shen, J. (2010). Regulation of Lrp6 phosphorylation. Cell. Mol. Life Sci. *67*, 2551–2562.

Noctor, S.C., Flint, A.C., Weissman, T.A., Dammerman, R.S., and Kriegstein, A.R. (2001). Neurons derived from radial glial cells establish radial units in neocortex. Nature *409*, 714–720.

Noctor, S.C., Martinez-Cerdeño, V., Ivic, L., and Kriegstein, A.R. (2004). Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases. Nat. Neurosci. 2004 72 *7*, 136–144.

Nowakowski, T.J., Fotaki, V., Pollock, A., Sun, T., Pratt, T., and Price, D.J. (2013). MicroRNA-92b regulates the development of intermediate cortical progenitors in embryonic mouse brain. Proc. Natl. Acad. Sci. U. S. A. *110*, 7056–7061.

Onishi, K., and Zou, Y. (2017). Sonic Hedgehog switches on Wnt/planar cell polarity signaling in commissural axon growth cones by reducing levels of Shisa2. Elife *6*.

Ornitz, D.M., and Itoh, N. (2015). The Fibroblast Growth Factor signaling pathway. Wiley Interdiscip. Rev. Dev. Biol. *4*, 215–266.

Osumi, N., Shinohara, H., Numayama-Tsuruta, K., and Maekawa, M. (2008). Concise review: Pax6 transcription factor contributes to both embryonic and adult neurogenesis as a multifunctional regulator. Stem Cells *26*, 1663–1672.

Panchision, D.M., Pickel, J.M., Studer, L., Lee, S.H., Turner, P.A., Hazel, T.G., and McKay, R.D.G. (2001). Sequential actions of BMP receptors control neural precursor cell production and

fate. Genes Dev. *15*, 2094–2110.

Pang, T., Atefy, R., and Sheen, V. (2008). Malformations of cortical development. Neurologist *14*, 181.

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

Paridaen, J.T., and Huttner, W.B. (2014). Neurogenesis during development of the vertebrate central nervous system. EMBO Rep. *15*, 351–364.

Pascual-Vargas, P., and Salinas, P.C. (2021). A Role for Frizzled and Their Post-Translational Modifications in the Mammalian Central Nervous System. Front. Cell Dev. Biol. *9*.

Patel, S.D., Ciatto, C., Chen, C.P., Bahna, F., Rajebhosale, M., Arkus, N., Schieren, I., Jessell, T.M., Honig, B., Price, S.R., et al. (2006). Type II Cadherin Ectodomain Structures: Implications for Classical Cadherin Specificity. Cell *124*, 1255–1268.

Patthy, L. (2000). The WIF module. Trends Biochem. Sci. *25*, 12–13.

Pauklin, S., and Vallier, L. (2013). XThe cell-cycle state of stem cells determines cell fate propensity. Cell *155*, 135.

Pearson, C.A., Moore, D.M., Tucker, H.O., Dekker, J.D., Hu, H., Miquelajáuregui, A., and Novitch, B.G. (2020). Foxp1 Regulates Neural Stem Cell Self-Renewal and Bias Toward Deep Layer Cortical Fates. Cell Rep. *30*, 1964-1981.e3.

Pederick, D.T., Homan, C.C., Jaehne, E.J., Piltz, S.G., Haines, B.P., Baune, B.T., Jolly, L.A., Hughes, J.N., Gecz, J., and Thomas, P.Q. (2016). Pcdh19 Loss-of-Function Increases Neuronal Migration In Vitro but is Dispensable for Brain Development in Mice. Sci. Rep.

Pederick, D.T., Richards, K.L., Piltz, S.G., Kumar, R., Mincheva-Tasheva, S., Mandelstam, S.A., Dale, R.C., Scheffer, I.E., Gecz, J., Petrou, S., et al. (2018). Abnormal Cell Sorting Underlies the Unique X-Linked Inheritance of PCDH19 Epilepsy. Neuron *97*, 59-66.e5.

Perkel, J.M. (2021). Single-cell proteomics takes centre stage. Nature *597*, 580–582.

Pham, D.H., Tan, C.C., Homan, C.C., Kolc, K.L., Corbett, M.A., McAninch, D., Fox, A.H., Thomas, P.Q., Kumar, R., and Gecz, J. (2017). Protocadherin 19 (PCDH19) interacts with paraspeckle

protein NONO to co-regulate gene expression with estrogen receptor alpha (ERα). Hum. Mol. Genet.

Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat. Methods 2013 1011 *10*, 1096–1098.

Pietro, F. di, Echard, A., and Morin, X. (2016). Regulation of mitotic spindle orientation: an integrated view. EMBO Rep. *17*, 1106.

Pilaz, L.J., Patti, D., Marcy, G., Ollier, E., Pfister, S., Douglas, R.J., Betizeau, M., Gautier, E., Cortay, V., Doerflinger, N., et al. (2009). Forced G1-phase reduction alters mode of division, neuron number, and laminar phenotype in the cerebral cortex. Proc. Natl. Acad. Sci. U. S. A. *106*, 21924–21929.

Pinto, L., and Götz, M. (2007). Radial glial cell heterogeneity—The source of diverse progeny in the CNS. Prog. Neurobiol. *83*, 2–23.

Pokutta, S., Drees, F., Yamada, S., Nelson, W.J., and Weis, W.I. (2008). Biochemical and structural analysis of alpha-catenin in cell-cell contacts. Biochem. Soc. Trans. *36*, 141–147.

Punovuori, K., Migueles, R.P., Malaguti, M., Blin, G., Macleod, K.G., Carragher, N.O., Pieters, T., van Roy, F., Stemmler, M.P., and Lowell, S. (2019). N-cadherin stabilises neural identity by dampening anti-neural signals. Dev. *146*.

Punovuori, K., Malaguti, M., and Lowell, S. (2021). Cadherins in early neural development. *78*, 4435–4450.

Rakic, P. (1972). Mode of cell migration to the superficial layers of fetal monkey neocortex. J. Comp. Neurol. *145*, 61–83.

Rakic, P. (1995). A small step for the cell, a giant leap for mankind: a hypothesis of neocortical expansion during evolution. Trends Neurosci. *18*, 383–388.

Rashid, D., Newell, K., Shama, L., and Bradley, R. (2006). A requirement for NF-protocadherin and TAF1/Set in cell adhesion and neural tube formation. Dev. Biol. *291*, 170–181.

Ratz, M., von Berlin, L., Larsson, L., Martin, M., Westholm, J.O., La Manno, G., Lundeberg, J., and Frisén, J. (2022). Clonal relations in the mouse brain revealed by single-cell and spatial transcriptomics. Nat. Neurosci. 2022 253 *25*, 285–294.

Redies, C., and Takeichi, M. (1993). Expression of N-cadherin mRNA during development of the mouse brain. Dev. Dyn. *197*, 26–39.

Redies, C., Vanhalst, K., and Van Roy, F. (2005). δ-Protocadherins: Unique structures and functions. Cell. Mol. Life Sci. *62*, 2840–2852.

Reillo, I., De Juan Romero, C., García-Cabezas, M.Á., and Borrell, V. (2011). A Role for Intermediate Radial Glia in the Tangential Expansion of the Mammalian Cerebral Cortex. Cereb. Cortex *21*, 1674–1694.

Romasko, E.J., DeChene, E.T., Balciuniene, J., Akgumus, G.T., Helbig, I., Tarpinian, J.M., Keena, B.A., Vogiatzi, M.G., Zackai, E.H., Izumi, K., et al. (2018). PCDH19-related epilepsy in a male with Klinefelter syndrome: Additional evidence supporting PCDH19 cellular interference disease mechanism. Epilepsy Res. *145*, 89–92.

Van Roy, F. (2014). Beyond E-cadherin: roles of other cadherin superfamily members in cancer. Nat. Rev. Cancer *14*, 121–134.

Ruan, X., Kang, B., Qi, C., Lin, W., Wang, J., and Zhang, X. (2021). Progenitor cell diversity in the developing mouse neocortex. Proc. Natl. Acad. Sci. U. S. A. *118*, e2018866118.

Rubinstein, R., Thu, C.A., Goodman, K.M., Wolcott, H.N., Bahna, F., Mannepalli, S., Ahlsen, G., Chevee, M., Halim, A., Clausen, H., et al. (2015). Molecular logic of neuronal self-recognition through protocadherin domain interactions. Cell *163*, 629–642.

Rubinstein, R., Goodman, K.M., Maniatis, T., Shapiro, L., and Honig, B. (2017). Structural origins of clustered protocadherin-mediated neuronal barcoding. Semin. Cell Dev. Biol. *69*, 140–150.

Sakane, F., and Miyamoto, Y. (2013). N-cadherin regulates the proliferation and differentiation of ventral midbrain dopaminergic progenitors. Dev. Neurobiol. *73*, 518–529.

Samanta, D. (2020). PCDH19-Related Epilepsy Syndrome: A Comprehensive Clinical Review. Pediatr. Neurol. *105*, 3–9.

Sano, K., Tanihara, H., Heimark, R.L., Obata, S., Davidson, M., St. John, T., Taketani, S., and Suzuki, S. (1993). Protocadherins: a large family of cadherin-related molecules in central nervous system. EMBO J. *12*, 2249–2256.

Sansom, S.N., Griffiths, D.S., Faedo, A., Kleinjan, D.J., Ruan, Y., Smith, J., Van Heyningen, V., Rubenstein, J.L., and Livesey, F.J. (2009). The level of the transcription factor Pax6 is essential

for controlling the balance between neural stem cell self-renewal and neurogenesis. PLoS Genet. *5*.

Sato, A., Kishida, S., Tanaka, T., Kikuchi, A., Kodama, T., Asashima, M., and Nishinakamura, R. (2004). Sall1, a causative gene for Townes-Brocks syndrome, enhances the canonical Wnt signaling by localizing to heterochromatin. Biochem. Biophys. Res. Commun. *319*, 103–113.

Sauer, F.C. (1935). Mitosis in the neural tube. J. Comp. Neurol. *62*, 377–405.

Schaarschuch, A., and Hertel, N. (2018). Expression profile of N-cadherin and protocadherin-19 in postnatal mouse limbic structures. J. Comp. Neurol. *526*, 663–680.

Scheffer, I.E., Turner, S.J., Dibbens, L.M., Bayly, M.A., Friend, K., Hodgson, B., Burrows, L., Shaw, M., Wei, C., Ullmann, R., et al. (2008). Epilepsy and mental retardation limited to females: An under-recognized disorder. Brain.

Sekine, K., Honda, T., Kawauchi, T., Kubo, K. ichiro, and Nakajima, K. (2011). The Outermost Region of the Developing Cortical Plate Is Crucial for Both the Switch of the Radial Migration Mode and the Dab1-Dependent “Inside-Out” Lamination in the Neocortex. J. Neurosci. *31*, 9426–9439.

Sekine, K., Kubo, K.I., and Nakajima, K. (2014). How does Reelin control neuronal migration and layer formation in the developing mammalian neocortex? Neurosci. Res. *86*, 50–58.

Shahi, M.H., Afzal, M., Sinha, S., Eberhart, C.G., Rey, J.A., Fan, X., and Castresana, J.S. (2010). Regulation of sonic hedgehog-GLI1 downstream target genes PTCH1, Cyclin D2, Plakoglobin, PAX6 and NKX2.2 and their epigenetic status in medulloblastoma and astrocytoma. BMC Cancer *10*, 1–20.

Shapiro, L., and Weis, W.I. (2009). Structure and biochemistry of cadherins and catenins. Cold Spring Harb. Perspect. Biol. *1*.

Shikata, Y., Okada, T., Hashimoto, M., Ellis, T., Matsumaru, D., Shiroishi, T., Ogawa, M., Wainwright, B., and Motoyama, J. (2011). Ptch1-mediated dosage-dependent action of Shh signaling regulates neural progenitor development at late gestational stages. Dev. Biol. *349*, 147–159.

Shimomura, Y., Agalliu, D., Vonica, A., Luria, V., Wajid, M., Baumer, A., Belli, S., Petukhova, L., Schinzel, A., Brivanlou, A.H., et al. (2010). APCDD1 is a novel Wnt inhibitor mutated in hereditary hypotrichosis simplex. Nat. 2010 4647291 *464*, 1043–1047.

Shtutman, M., Zhurinsky, J., Simcha, I., Albanese, C., D'Amico, M., Pestell, R., and Ben-Ze'ev, A. (1999). The cyclin D1 gene is a target of the β-catenin/LEF-1 pathway. Proc. Natl. Acad. Sci. *96*, 5522–5527.

Shu, Y., Xiang, M., Zhang, P., Qi, G., He, F., Zhang, Q., Zhang, Z., Lv, Z., Peng, X., Cai, H., et al. (2018). Wnt-5a Promotes Neural Development and Differentiation by Regulating CDK5 via Ca2+/Calpain Pathway. Cell. Physiol. Biochem. *51*, 2604–2615.

Da Silva, F., Zhang, K., Pinson, A., Fatti, E., Wilsch-Br€ Auninger, M., Herbst, J., Vidal, V., Schedl, A., Huttner, W.B., and Niehrs, C. (2021). Mitotic WNT signalling orchestrates neurogenesis in the developing neocortex. EMBO J. *40*, e108041.

Singh, I., Mehta, A., Contreras, A., Boettger, T., Carraro, G., Wheeler, M., Cabrera-Fuentes, H.A., Bellusci, S., Seeger, W., Braun, T., et al. (2014). Hmga2 is required for canonical WNT signaling during lung development. BMC Biol. *12*.

Smoller, J.W., Kendler, K., Craddock, N., Lee, P.H., Neale, B.M., Nurnberger, J.N., Ripke, S., Santangelo, S., Sullivan, P.S., Neale, B.N., et al. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: A genome-wide analysis. Lancet.

Snijders Blok, L., Madsen, E., Juusola, J., Gilissen, C., Baralle, D., Reijnders, M.R.F., Venselaar, H., Helsmoortel, C., Cho, M.T., Hoischen, A., et al. (2015). Mutations in DDX3X Are a Common Cause of Unexplained Intellectual Disability with Gender-Specific Effects on Wnt Signaling. Am. J. Hum. Genet. *97*, 343–352.

Soneson, C., and Robinson, M.D. (2018). Bias, robustness and scalability in single-cell differential expression analysis. Nat. Methods 2018 154 *15*, 255–261.

Sotomayor, M., Gaudet, R., and Corey, D.P. (2014). Sorting out a promiscuous superfamily: towards cadherin connectomics. Trends Cell Biol. *24*, 524–536.

Specchio, N., Marini, C., Terracciano, A., Mei, D., Trivisano, M., Sicca, F., Fusco, L., Cusmai, R., Darra, F., Bernardina, B.D., et al. (2011). Spectrum of phenotypes in female patients with epilepsy due to protocadherin 19 mutations. Epilepsia *52*, 1251–1257.

Specht, H., Emmott, E., Petelski, A.A., Huffman, R.G., Perlman, D.H., Serra, M., Kharchenko, P., Koller, A., and Slavov, N. (2021). Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. Genome Biol. *22*, 1–27.

Squair, J.W., Gautier, M., Kathe, C., Anderson, M.A., James, N.D., Hutson, T.H., Hudelle, R.,

Qaiser, T., Matson, K.J.E., Barraud, Q., et al. (2021). Confronting false discoveries in single-cell differential expression. Nat. Commun. 2021 121 *12*, 1–15.

Stegle, O., Teichmann, S.A., and Marioni, J.C. (2015). Computational and analytical challenges in single-cell transcriptomics. Nat. Rev. Genet. 2015 163 *16*, 133–145.

Stewart, M.J., and Denell, R. (1993). Mutations in the Drosophila gene encoding ribosomal protein S6 cause tissue overgrowth. Mol. Cell. Biol. *13*, 2524–2535.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. *102*, 15545–15550.

Tai, K., Kubota, M., Shiono, K., Tokutsu, H., and Suzuki, S.T. (2010). Adhesion properties and retinofugal expression of chicken protocadherin-19. Brain Res. *1344*, 13–24.

Takahashi, T., Nowakowski, R.S., and Caviness, V.S. (1993). Cell cycle parameters and patterns of nuclear movement in the neocortical proliferative zone of the fetal mouse. J. Neurosci. *13*, 820–833.

Takahashi, T., Nowakowski, R.S., and Caviness, V.S. (1996). The leaving or Q fraction of the murine cerebral proliferative epithelium: a general model of neocortical neuronogenesis. J. Neurosci. *16*, 6183–6196.

Takehara, T., Teramura, T., Onodera, Y., Frampton, J., and Fukuda, K. (2015). Cdh2 stabilizes FGFR1 and contributes to primed-state pluripotency in mouse epiblast stem cells. Sci. Rep. *5*.

Tamai, K., Semenov, M., Kato, Y., Spokony, R., Liu, C., Katsuyama, Y., Hess, F., Saint-Jeannet, J.P., and He, X. (2000). LDL-receptor-related proteins in Wnt signal transduction. Nature *407*, 530–535.

Tan, C., Shard, C., Ranieri, E., Hynes, K., Pham, D.H., Leach, D., Buchanan, G., Corbett, M., Shoubridge, C., Kumar, R., et al. (2015). Mutations of protocadherin 19 in female epilepsy (PCDH19-FE) lead to allopregnanolone deficiency. Hum. Mol. Genet. *24*, 5250–5259.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nat. Methods 2009 65 *6*, 377–382.

Tanriover, G., Kayisli, U.A., Demir, R., Pestereli, E., Karaveli, S., and Demir, N. (2004). Distribution of N-cadherin in human cerebral cortex during prenatal development. Histochem. Cell Biol. *122*, 191–200.

Tarabykin, V., Stoykova, A., Usman, N., and Gruss, P. (2001). Cortical upper layer neurons derive from the subventricular zone as indicated by Svet1 gene expression. Development *128*, 1983–1993.

Telley, L., Agirman, G., Prados, J., Amberg, N., Fièvre, S., Oberst, P., Bartolini, G., Vitali, I., Cadilhac, C., Hippenmeyer, S., et al. (2019). Temporal patterning of apical progenitors and their daughter neurons in the developing neocortex. Science (80-.). *364*.

Teng, T., Thomas, G., and Mercer, C.A. (2013). Growth control and ribosomopathies. Curr. Opin. Genet. Dev. *23*, 63–71.

Tenzen, T., Allen, B.L., Cole, F., Kang, J.S., Krauss, R.S., and McMahon, A.P. (2006). The cell surface membrane proteins Cdo and Boc are components and targets of the Hedgehog signaling pathway and feedback network in mice. Dev. Cell *10*, 647–656.

Touahri, Y., Escalas, N., Benazeraf, B., Cochard, P., Danesin, C., and Soula, C. (2012). Sulfatase 1 Promotes the Motor Neuron-to-Oligodendrocyte Fate Switch by Activating Shh Signaling in Olig2 Progenitors of the Embryonic Ventral Spinal Cord. J. Neurosci. *32*, 18018–18034.

Tozer, S., Baek, C., Fischer, E., Goiame, R., and Morin, X. (2017). Differential Routing of Mindbomb1 via Centriolar Satellites Regulates Asymmetric Divisions of Neural Progenitors. Neuron *93*, 542-551.e4.

Trivisano, M., Lucchi, C., Rustichelli, C., Terracciano, A., Cusmai, R., Ubertini, G.M., Giannone, G., Bertini, E.S., Vigevano, F., Gecz, J., et al. (2017). Reduced steroidogenesis in patients with PCDH19-female limited epilepsy. Epilepsia *58*, e91–e95.

Tsunekawa, Y., Britto, J.M., Takahashi, M., Polleux, F., Tan, S.S., and Osumi, N. (2012). Cyclin D2 in the basal process of neural progenitors is linked to non-equivalent cell fates. EMBO J. *31*, 1879.

Tuoc, T.C., Boretius, S., Sansom, S.N., Pitulescu, M.E., Frahm, J., Livesey, F.J., and Stoykova, A. (2013). Chromatin regulation by BAF170 controls cerebral cortical size and thickness. Dev. Cell *25*, 256–269.

Ulloa, F., Itasaki, N., and Briscoe, J. (2007). Inhibitory Gli3 Activity Negatively Regulates Wnt/β-

Catenin Signaling. Curr. Biol. *17*, 545–550.

Uzquiano, A., Gladwyn-Ng, I., Nguyen, L., Reiner, O., Götz, M., Matsuzaki, F., and Francis, F. (2018). Cortical progenitor biology: key features mediating proliferation versus differentiation. J. Neurochem.

Vaccarino, F.M., Schwartz, M.L., Raballo, R., Nilsen, J., Rhee, J., Zhou, M., Doetschman, T., Coffin, J.D., Wyland, J.J., and Hung, Y.T.E. (1999). Changes in cerebral cortex size are governed by fibroblast growth factor during embryogenesis. Nat. Neurosci. *2*, 246–253.

Vanhalst, K., Kools, P., Staes, K., Van Roy, F., and Redies, C. (2005). δ-Protocadherins: A gene family expressed differentially in the mouse brain. Cell. Mol. Life Sci. *62*, 1247–1259.

Vasistha, N.A., García-Moreno, F., Arora, S., Cheung, A.F.P., Arnold, S.J., Robertson, E.J., and Molnár, Z. (2015). Cortical and Clonal Contribution of Tbr2 Expressing Progenitors in the Developing Mouse Brain. Cereb. Cortex *25*, 3290–3302.

Vendrell, V., Summerhurst, K., Sharpe, J., Davidson, D., and Murphy, P. (2009). Gene expression analysis of canonical Wnt pathway transcriptional regulators during early morphogenesis of the facial region in the mouse embryo. Gene Expr. Patterns *9*, 296–305.

Vonica, A., Bhat, N., Phan, K., Guo, J., Iancu, L., Weber, J.A., Karger, A., Cain, J.W., Wang, E.C.E., DeStefano, G.M., et al. (2020). Apcdd1 is a dual BMP/Wnt inhibitor in the developing nervous system and skin. Dev. Biol. *464*, 71–87.

Wan, M., Yang, C., Li, J., Wu, X., Yuan, H., Ma, H., He, X., Nie, S., Chang, C., and Cao, X. (2008). Parathyroid hormone signaling through low-density lipoprotein-related protein 6. Genes Dev. *22*, 2968.

Wang, H., Ge, G., Uchida, Y., Luu, B., and Ahn, S. (2011). Gli3 Is Required for Maintenance and Fate Specification of Cortical Progenitors. J. Neurosci. *31*, 6440.

Wang, X., He, Y., Zhang, Q., Ren, X., and Zhang, Z. (2021). Direct Comparative Analyses of 10X Genomics Chromium and Smart-seq2. Genomics. Proteomics Bioinformatics *19*, 253–266.

Wei, Q., Yokota, C., Semenov, M. V., Doble, B., Woodgett, J., and He, X. (2007). R-spondin1 is a high affinity ligand for LRP6 and induces LRP6 phosphorylation and beta-catenin signaling. J. Biol. Chem. *282*, 15903–15911.

Whitley, S.K., Horne, W.T., and Kolls, J.K. (2016). Research Techniques Made Simple:

Methodology and Clinical Applications of RNA Sequencing. J. Invest. Dermatol. *136*, e77–e82.

Winden, K.D., Yuskaitis, C.J., and Poduri, A. (2015). Megalencephaly and Macrocephaly. Semin. Neurol. *35*, 277–287.

Wolf, J., Palmby, T.R., Gavard, J., Williams, B.O., and Gutkind, J.S. (2008). Multiple PPPS/TP motifs act in a combinatorial fashion to transduce Wnt signaling through LRP6. FEBS Lett. *582*, 255–261.

Woodhead, G.J., Mutch, C.A., Olson, E.C., and Chenn, A. (2006a). Cell-autonomous beta-catenin signaling regulates cortical precursor proliferation. J. Neurosci. *26*, 12620–12630.

Woodhead, G.J., Mutch, C.A., Olson, E.C., and Chenn, A. (2006b). Cell-autonomous β-catenin signaling regulates cortical precursor proliferation. J. Neurosci.

Wu, Q., and Maniatis, T. (1999). A striking organization of a large family of human neural cadherin-like cell adhesion genes. Cell *97*, 779–790.

Wu, Q., Zhang, T., Cheng, J.F., Kim, Y., Grimwood, J., Schmutz, J., Dickson, M., Noonan, J.P., Zhang, M.Q., Myers, R.M., et al. (2001). Comparative DNA sequence analysis of mouse and human protocadherin gene clusters. Genome Res. *11*, 389–404.

Wu, X., Yang, B., Udo-Inyang, I., Ji, S., Ozog, D., Zhou, L., and Mi, Q.S. (2018). Research Techniques Made Simple: Single-Cell RNA Sequencing and its Applications in Dermatology. J. Invest. Dermatol. *138*, 1004–1009.

Xiao, X., Zheng, F., Chang, H., Ma, Y., Yao, Y.G., Luo, X.J., and Li, M. (2017). The Gene Encoding Protocadherin 9 (PCDH9), a Novel Risk Factor for Major Depressive Disorder. Neuropsychopharmacol. 2018 435 *43*, 1128–1137.

Xu, X., Xiao, Z., Hou, Y., Yan, Z., Xu, Y., Lv, Y., Xiang, X., Yang, H., Qi, X., and Chu, L. (2022). TPPP3 Promote Epithelial-Mesenchymal Transition via Snail1 in Glioblastoma.

Yamagata, M., Lefebvre, J.L., Seuntjens, E., Pancho, A., Aerts, T., and Mitsogiannis, M.D. (2020). Protocadherins at the Crossroad of Signaling Pathways.

Yayon, A., Klagsbrun, M., Esko, J.D., Leder, P., and Ornitz, D.M. (1991). Cell surface, heparin-like molecules are required for binding of basic fibroblast growth factor to its high affinity receptor. Cell *64*, 841–848.

Yoshikawa, S., Bonkowsky, J.L., Kokel, M., Shyn, S., and Thomas, J.B. (2001). The derailed

guidance receptor does not require kinase activity in vivo. J. Neurosci. *21*.

Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). ClusterProfiler: An R package for comparing biological themes among gene clusters. Omi. A J. Integr. Biol. *16*, 284–287.

Yum, S., Lee, S.J., Piao, S., Xu, Y., Jung, J., Jung, Y., Oh, S., Lee, J., Park, B.J., and Ha, N.C. (2009). The role of the Ser/Thr cluster in the phosphorylation of PPPSP motifs in Wnt coreceptors. Biochem. Biophys. Res. Commun. *381*, 345–349.

Yun, K., Mantani, A., Garel, S., Rubenstein, J., and Israel, M.A. (2004). Id4 regulates neural progenitor proliferation and differentiation in vivo. Development *131*, 5441–5448.

Zeng, X., Tamai, K., Doble, B., Li, S., Huang, H., Habas, R., Okamura, H., Woodgett, J., and He, X. (2005). A dual-kinase mechanism for Wnt co-receptor phosphorylation and activation. Nature *438*, 873–877.

Zhang, J., Shemezis, J.R., McQuinn, E.R., Wang, J., Sverdlov, M., and Chenn, A. (2013). AKT activation by N-cadherin regulates beta-catenin signaling and neuronal differentiation during cortical development. Neural Dev. *8*.

Zhang, P., Wu, C., Liu, N., Niu, L., Yan, Z., Feng, Y., and Xu, R. (2014). Protocadherin 11 X Regulates Differentiation and Proliferation of Neural Stem Cell In Vitro and In Vivo. J. Mol. Neurosci. *54*, 199–210.

Zhang, Y., Sivasankar, S., Nelson, W.J., and Chu, S. (2009). Resolving cadherin interactions and binding cooperativity at the single-molecule level. Proc. Natl. Acad. Sci. U. S. A. *106*, 109.

Zhao, X., Garcia, J.Q., Tong, K., Chen, X., Yang, B., Li, Q., Dai, Z., Shi, X., Seiple, I.B., Huang, B., et al. (2021). Polarized endosome dynamics engage cytoplasmic Par-3 that recruits dynein during asymmetric cell division. Sci. Adv. *7*.

Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun. 2017 81 *8*, 1–12.

Zhong, J., Kim, H.T., Lyu, J., Yoshikawa, K., Nakafuku, M., and Lu, W. (2011). The Wnt receptor Ryk controls specification of GABAergic neurons versus oligodendrocytes during telencephalon development. Development *138*, 409–419.

Zhong, S., Zhang, S., Fan, X., Wu, Q., Yan, L., Dong, J., Zhang, H., Li, L., Sun, L., Pan, N., et al.

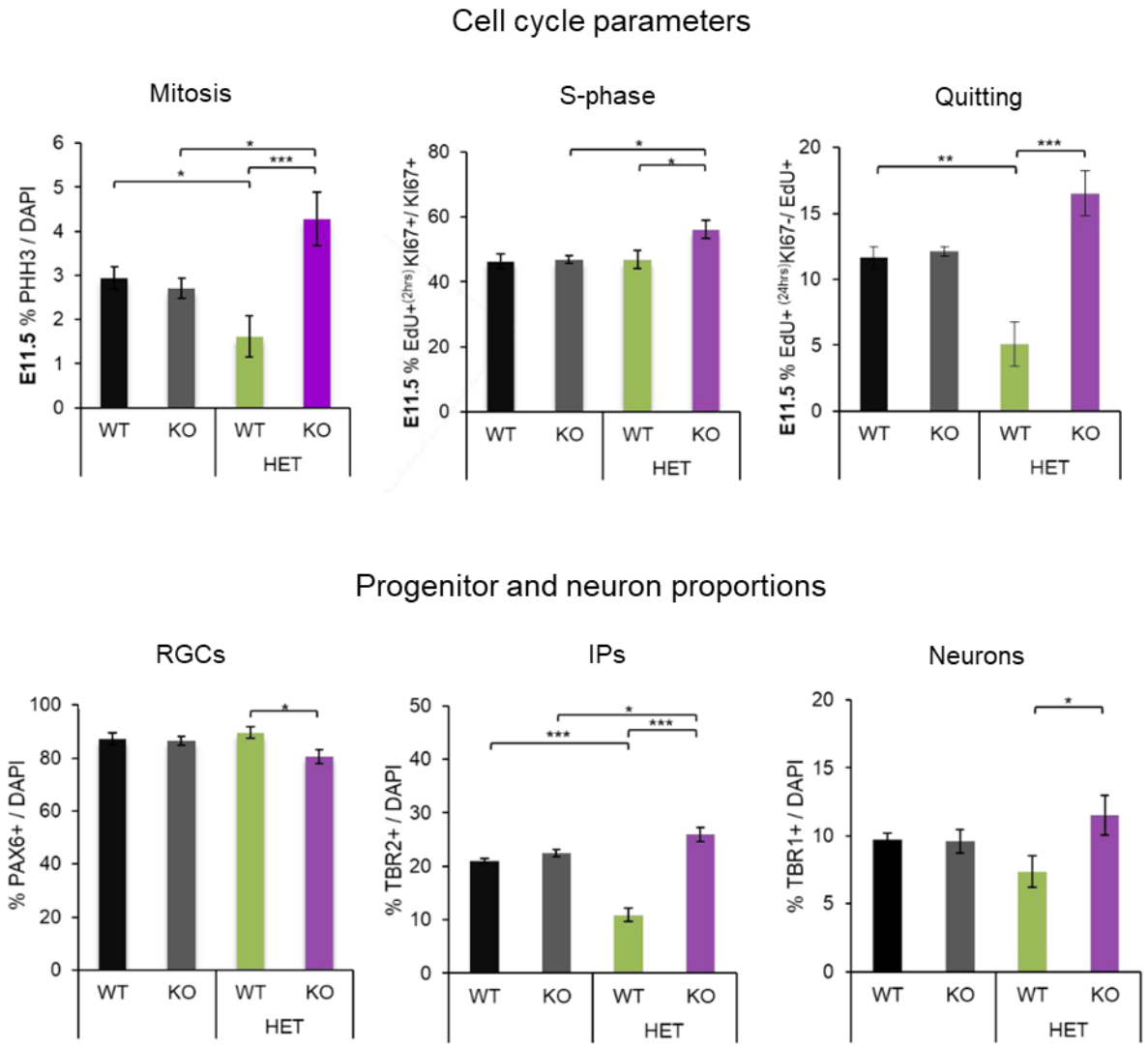
(2018). A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. Nat. 2018 5557697 *555*, 524–528.

Zhou, C.J., Borello, U., Rubenstein, J.L.R., and Pleasure, S.J. (2006). Neuronal production and precursor proliferation defects in the neocortex of mice with loss of function in the canonical Wnt signaling pathway. Neuroscience *142*, 1119–1131.

Zhou, Y., Chai, H., Guo, L., Dai, Z., Lai, J., Duan, J., Liu, Y., and Ding, Q. (2021). Knockdown of CENPW Inhibits Hepatocellular Carcinoma Progression by Inactivating E2F Signaling. Technol. Cancer Res. Treat. *20*.

Zou, C., Huang, W., Ying, G., and Wu, Q. (2007). Sequence analysis and expression mapping of the rat clustered protocadherin gene repertoires. Neuroscience *144*, 579–603.

Chapter 8: Appendices



Appendix 1. Proportional analysis of cell cycling parameters and progenitor and neuron proportions between WT, HET, and KO animals. Original data from Dr. Jessica Griffith's PhD thesis demonstrating the proportion of cell cycle parameters and the proportion of progenitors and neurons between WT, HET, and KO animals. The difference of these proportions between WT and KO cells within the HETs are also shown. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

Appendix 2. Table illustrating all the DE genes from the KO vs. WT female comparison.

Gene	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
Btbd17	84.66616	-2.41002	0.615689	-3.88839	0.000101	0.010242
Th	38.92205	2.533265	0.642948	3.851716	0.000117	0.011014
Galnt1	810.7304	-1.62798	0.490562	-3.31154	0.000928	0.029579
Mmp11	176.8674	1.517638	0.490965	3.085159	0.002034	0.044577
Gmcl1	589.2248	-2.23701	0.600434	-3.71137	0.000206	0.014114
Fbxo9	212.2848	-2.00781	0.534707	-3.76033	0.00017	0.013086
Ubl3	779.946	-2.14707	0.507672	-4.2184	2.46E-05	0.006224
Rac1	2021.793	-2.14727	0.57685	-3.69473	0.00022	0.014372
Emd	1562.012	1.316635	0.422022	3.115256	0.001838	0.042283
Ccne1	655.4714	-2.37004	0.576073	-4.10571	4.03E-05	0.007098
Dbf4	1331.405	-2.15744	0.578201	-3.74753	0.000179	0.013491
Nr2f6	176.8064	-1.94561	0.66714	-3.09227	0.001986	0.044154
Ocel1	137.1176	-2.19557	0.654374	-3.49236	0.000479	0.020335
Plekha3	254.9346	-2.08553	0.567979	-3.65807	0.000254	0.015164
Btbd6	129.3616	-1.89473	0.613622	-3.13479	0.00172	0.04093
Nab1	154.915	-1.71492	0.560236	-3.03736	0.002387	0.048734
Bcam	62.35199	2.513796	0.670472	3.848951	0.000119	0.011064
Prkar2b	518.2542	-2.42932	0.565434	-4.30431	1.68E-05	0.005173
Cnot11	346.7092	-2.64139	0.594502	-4.40546	1.06E-05	0.004942
Car11	28.10163	2.117514	0.629653	3.477554	0.000506	0.021039
Abhd17a	353.2262	-2.30921	0.65017	-3.64595	0.000266	0.015583
Grik5	170.9526	-2.04807	0.645686	-3.24394	0.001179	0.033692
Dll3	98.47004	-2.85565	0.643633	-4.35262	1.35E-05	0.005106
Cul3	2120.078	-2.40502	0.58463	-4.06757	4.75E-05	0.007468
Ppp1cc	632.2633	-2.60815	0.608757	-4.25306	2.11E-05	0.005822
Chn2	135.6926	-1.90531	0.536912	-3.55233	0.000382	0.017928
Ctcf	1935.596	-1.5745	0.511032	-3.07874	0.002079	0.045055
Rcn1	1678.918	-2.16148	0.593193	-3.66394	0.000248	0.015037
Srm	1319.498	-2.81995	0.629087	-4.52383	6.07E-06	0.00387
Mark3	1113.878	-1.7844	0.535055	-3.33079	0.000866	0.028128
Gatad1	3943.315	-2.05552	0.607261	-3.40835	0.000654	0.024216
Rplp1	5632.133	-1.83348	0.589511	-3.10123	0.001927	0.043711
Herpud2	562.4696	-2.20099	0.594265	-3.70896	0.000208	0.014114
Rala	1397.283	-2.61854	0.604752	-4.34883	1.37E-05	0.005106
Ppp2cb	536.7356	-1.76973	0.553138	-3.16926	0.001528	0.038341
Tusc2	318.0857	-1.93105	0.585612	-3.33629	0.000849	0.027734
Faf1	1399.341	-2.58979	0.584123	-4.42971	9.44E-06	0.004942
Gm266	24.90539	-2.95842	0.670301	-4.40009	1.08E-05	0.004942
Odc1	3598.461	-2.48002	0.638067	-4.00997	6.07E-05	0.008383
Mplkip	287.7547	-1.81472	0.570061	-3.21328	0.001312	0.035519

Pten	1477.459	-2.44261	0.588001	-4.16846	3.07E-05	0.006558
Naf1	292.1573	-2.47121	0.609742	-4.03547	5.45E-05	0.007676
Yes1	420.6575	-1.79473	0.541045	-3.29685	0.000978	0.030157
Fnta	1535.923	-2.34903	0.634772	-3.90094	9.58E-05	0.009925
Gtf3a	747.3291	-1.87908	0.49388	-3.80472	0.000142	0.011805
Mrps18c	1494.592	-1.68038	0.504283	-3.34239	0.000831	0.027496
Srsf6	5536.248	-2.50208	0.596476	-4.20493	2.61E-05	0.006224
Kctd5	620.2369	-1.69658	0.542905	-3.1334	0.001728	0.040932
Etaa1	451.412	-2.33812	0.587337	-3.96945	7.20E-05	0.008879
Nlk	457.1794	-1.80167	0.542275	-3.30461	0.000951	0.029905
Suz12	1863.908	-1.93932	0.538426	-3.59932	0.000319	0.016515
Serinc3	318.5616	-1.77606	0.566177	-3.13192	0.001737	0.040949
Crk	1761.417	-1.78876	0.532924	-3.34381	0.000826	0.027496
Rab5a	352.8584	-2.20831	0.616825	-3.56106	0.000369	0.01789
Mafk	71.5652	1.513535	0.458087	3.302459	0.000958	0.029905
Srsf1	5711.851	-1.95785	0.557345	-3.49675	0.000471	0.02019
C1qbp	1036.003	-2.35492	0.607082	-3.89961	9.63E-05	0.009925
Appbp2	1902.014	-1.42804	0.461627	-3.08957	0.002004	0.0442
Cenpv	470.9262	-2.79957	0.636952	-4.39637	1.10E-05	0.004942
Tmem160	177.6189	-2.11111	0.640087	-3.31655	0.000911	0.029257
Mtrf1l	91.41165	-2.37406	0.592982	-3.98576	6.73E-05	0.008724
Hdac2	3665.307	-2.41453	0.604949	-3.99511	6.47E-05	0.008583
Rwdd1	890.1511	-1.77201	0.514911	-3.43903	0.000584	0.022987
Nr2e1	1185.404	-1.553	0.499033	-3.10087	0.00193	0.043711
Aig1	182.7201	-2.18598	0.555211	-3.9401	8.14E-05	0.009141
Ppp1r12a	6186.216	-1.68801	0.54837	-3.09702	0.001955	0.043805
Ube2d1	1474.216	-2.05757	0.572972	-3.56644	0.000362	0.01782
Tmpo	6039.703	-2.0089	0.558537	-3.58575	0.000336	0.016933
Pex7	226.6378	-1.88238	0.63337	-3.13164	0.001738	0.040949
Ntn4	82.95826	-2.56379	0.643462	-3.94958	7.83E-05	0.009126
Cep83	597.2258	-2.17454	0.556283	-3.91984	8.86E-05	0.009631
Nudt4	2210.34	-2.24581	0.588216	-3.83731	0.000124	0.011223
Ascl1	199.7618	-2.2244	0.60952	-3.67384	0.000239	0.014767
Tysnd1	155.6405	-1.89669	0.624945	-3.09325	0.00198	0.044079
Peli1	811.5332	-2.095	0.572745	-3.63948	0.000273	0.015718
Actr2	18128.29	-2.05203	0.604181	-3.39067	0.000697	0.024986
Yeats4	1870.901	-2.17641	0.572638	-3.77238	0.000162	0.013086
Ppp2ca	1930.037	-1.83524	0.557629	-3.29266	0.000992	0.030379
Hnrnpab	8368.215	-2.46488	0.622538	-3.98609	6.72E-05	0.008724
Phykpl	5102.006	-2.22396	0.601408	-3.70839	0.000209	0.014114
Hspa4	2519.47	-1.71042	0.535362	-3.18703	0.001437	0.037367
Rtn4	2795.147	-1.92475	0.561136	-3.42914	0.000605	0.023544
Mrps24	689.4438	-3.16145	0.63384	-5.26434	1.41E-07	0.000324

Ccdc117	353.0997	-1.66147	0.529241	-3.13642	0.00171	0.040845
Ppm1d	484.8392	-2.63646	0.605582	-4.38345	1.17E-05	0.004942
Nampt	1805.531	-1.54241	0.503831	-3.05293	0.002266	0.047075
Rab10	1623.629	-1.79369	0.538235	-3.34172	0.000833	0.027496
Polg2	174.7019	-1.83398	0.560943	-3.28259	0.001029	0.031001
Mxra7	70.36424	-2.13734	0.590236	-3.62556	0.000288	0.016016
Pik3r5	11.56463	2.087055	0.660607	3.206639	0.001343	0.035788
Strn3	1038.585	-1.65078	0.508948	-3.24087	0.001192	0.033778
Klhdc2	2312.947	-2.38772	0.58793	-4.04807	5.16E-05	0.007468
Zc3h14	1896.423	-1.93024	0.565152	-3.43461	0.000593	0.023221
Ppp2r5e	1085.71	-2.20712	0.568769	-3.86952	0.000109	0.010561
Sgpp1	150.915	-2.43829	0.590723	-4.09775	4.17E-05	0.007098
Sav1	252.6948	-2.29748	0.607616	-3.71308	0.000205	0.014114
Ppm1a	1074.023	-1.94183	0.560966	-3.43443	0.000594	0.023221
Hif1a	1999.656	-1.56189	0.512224	-3.0423	0.002348	0.048156
Acot2	318.0773	-2.23603	0.63765	-3.55893	0.000372	0.017896
Evl	778.8529	-1.8407	0.589806	-3.13518	0.001717	0.04093
Yy1	847.1206	-1.62818	0.521307	-3.11582	0.001834	0.042274
Eif5	6191.953	-1.99942	0.599605	-3.4082	0.000654	0.024216
Trim27	1358.727	-1.95155	0.597049	-3.31181	0.000927	0.029579
Mcur1	347.0086	-1.74267	0.47485	-3.67093	0.000242	0.014767
Dek	8990.855	-2.20791	0.6029	-3.7182	0.000201	0.014114
Id4	6749.52	-2.59669	0.618474	-4.14039	3.47E-05	0.006741
Wrnip1	293.1872	-2.07652	0.587575	-3.56871	0.000359	0.017813
Pcbd2	36.89337	-2.05737	0.654661	-3.2946	0.000986	0.030302
Rasa1	1095.419	-2.09129	0.542849	-3.84353	0.000121	0.011087
X2210016F16Rik	253.3323	-1.78668	0.584708	-3.12165	0.001798	0.041863
Gkap1	494.029	-1.84367	0.489318	-3.76976	0.000163	0.013086
Nsun2	1330.184	-1.75927	0.542357	-3.25915	0.001117	0.032612
Slc30a5	358.4771	-1.74868	0.532243	-3.28525	0.001019	0.03086
Ankra2	267.1356	-1.71386	0.558556	-3.08228	0.002054	0.044797
Aggf1	804.2728	-1.7417	0.552667	-3.15592	0.0016	0.039089
Mrps30	494.514	-2.93988	0.638243	-4.64099	3.47E-06	0.002815
Kctd6	212.4107	-2.36682	0.617844	-3.8134	0.000137	0.011608
Nr1d2	69.11975	-1.57638	0.514807	-3.05833	0.002226	0.046623
Glud1	1058.576	-2.22476	0.57546	-3.8848	0.000102	0.010244
Ppif	463.844	-2.01368	0.583248	-3.46967	0.000521	0.021473
Slmap	402.7744	-1.66599	0.477548	-3.48415	0.000494	0.020651
Capn7	438.47	-1.64266	0.509545	-3.22679	0.001252	0.034669
Spryd7	271.6672	-1.99721	0.527597	-3.75501	0.000173	0.013219
Micu2	284.6899	-1.62719	0.507166	-3.21419	0.001308	0.035476
Mipep	267.1569	-1.85627	0.549186	-3.39579	0.000684	0.024664
Bnip3l	2887.028	-1.53894	0.474316	-3.2419	0.001187	0.033778

Ppp2r2a	1453.762	-1.71752	0.56234	-3.07208	0.002126	0.045423
Ppp3cc	34.94487	-2.14202	0.669894	-3.03142	0.002434	0.049338
Spry2	490.0434	-2.23668	0.620839	-3.64187	0.000271	0.015699
Fbxl3	264.4992	-1.54079	0.495316	-3.09893	0.001942	0.043711
Golph3	999.3054	-2.26879	0.550566	-4.11292	3.91E-05	0.007098
Nadk2	632.5221	-2.08056	0.596882	-3.48428	0.000493	0.020651
Khdrbs3	859.8433	-2.09689	0.580704	-3.58599	0.000336	0.016933
Cyrib	1370.359	-2.40119	0.574665	-4.15024	3.32E-05	0.006644
St13	26575.21	-2.17386	0.586331	-3.71873	0.0002	0.014114
Hes1	558.0749	-2.43419	0.623915	-3.87446	0.000107	0.010536
Kif21a	2255.216	-1.72673	0.541394	-3.19219	0.001412	0.03685
Yaf2	1142.759	-1.58857	0.520431	-3.02914	0.002453	0.049423
Nectin3	1969.059	-2.10749	0.562468	-3.73103	0.000191	0.013928
Atg3	379.3058	-1.87078	0.545661	-3.41417	0.00064	0.024195
Cpox	496.4809	-1.78697	0.532736	-3.37677	0.000733	0.025761
Tomm70a	1637.872	-1.74947	0.537148	-3.245	0.001174	0.033691
Dlg1	368.7984	-1.98179	0.567739	-3.50749	0.000452	0.019636
Yars2	318.9102	-2.3019	0.553748	-4.13348	3.57E-05	0.006762
Usp25	454.6425	-1.88308	0.558612	-3.36151	0.000775	0.026752
Brwd1	1985.459	-2.00453	0.502504	-3.97724	6.97E-05	0.008749
Donson	455.0077	-1.80005	0.574285	-3.12346	0.001787	0.041678
Nckap5l	123.9842	2.022363	0.624301	3.266539	0.001089	0.032228
Atf1	442.2568	-1.65715	0.481134	-3.43922	0.000583	0.022987
Nus1	774.2815	-2.04286	0.525243	-3.88524	0.000102	0.010244
Akirin1	485.268	-2.6066	0.600772	-4.30747	1.65E-05	0.005173
Ube2j2	441.424	-1.94208	0.589335	-3.35994	0.00078	0.0268
Marchf5	1009.711	-2.50244	0.607381	-4.12403	3.72E-05	0.006945
Rab12	288.823	-2.10426	0.578186	-3.61347	0.000302	0.016091
Chd1	1094.093	-1.47024	0.474722	-3.09174	0.00199	0.044161
Vegfa	120.8444	-2.85772	0.656962	-4.52888	5.93E-06	0.00387
Fbxl17	329.8517	-1.87062	0.580431	-3.22647	0.001253	0.034669
Pim1	342.8748	-1.64499	0.537086	-3.08372	0.002044	0.044671
Abcg1	1075.625	-2.275	0.623001	-3.72362	0.000196	0.014046
Spast	1013.919	-1.73561	0.551885	-3.13647	0.00171	0.040845
Yipf4	251.3335	-2.44528	0.613584	-3.95856	7.54E-05	0.009051
Strn	1828.972	-2.36615	0.627929	-3.92404	8.71E-05	0.00954
Vapa	2986.201	-2.38464	0.591962	-4.0046	6.21E-05	0.008476
Hnrnpll	2707.06	-2.06042	0.61365	-3.4387	0.000585	0.022987
Lrpprc	947.5621	-1.66954	0.545325	-3.06637	0.002167	0.046014
Bambi	76.34264	-2.11084	0.596889	-3.51002	0.000448	0.019573
Slc25a46	1045.423	-1.71978	0.553605	-3.10454	0.001906	0.043412
Ccny	312.372	-1.71414	0.538238	-3.18117	0.001467	0.037673
Spata24	46.83189	-1.5718	0.472368	-3.31456	0.000918	0.029397

Ndfip1	1723.983	-2.07781	0.622675	-3.41707	0.000633	0.02407
Ap3s1	606.083	-2.23017	0.592644	-3.74319	0.000182	0.013491
Dtwd2	128.8502	-2.19137	0.621794	-3.56648	0.000362	0.01782
Mbd2	253.0007	-2.47741	0.603402	-4.06655	4.77E-05	0.007468
Smad4	642.4732	-1.70047	0.536102	-3.16357	0.001558	0.038762
Isoc1	1300.953	-1.88838	0.537624	-3.51737	0.000436	0.019222
Osbp	108.4539	-2.11954	0.625689	-3.41836	0.00063	0.02406
Rfk	930.9909	-2.33619	0.579635	-4.05007	5.12E-05	0.007468
Carnmt1	119.0597	-2.3263	0.596663	-3.84715	0.000119	0.011071
Zfand5	1289.654	-2.56876	0.5936	-4.27896	1.88E-05	0.0054
Sac3d1	268.5555	-2.64446	0.618833	-4.37148	1.23E-05	0.00501
Minpp1	377.3653	-1.69877	0.547808	-3.09839	0.001946	0.043711
Vegfb	285.2135	-2.49603	0.633255	-3.98257	6.82E-05	0.008724
Pdcd4	3393.155	-2.12392	0.579406	-3.70209	0.000214	0.014278
Tm9sf3	1340.986	-1.63086	0.518861	-3.1465	0.001652	0.039877
Mxi1	438.4936	-1.45579	0.475155	-3.06922	0.002146	0.045751
Taf5	390.1113	-2.10665	0.559764	-3.76253	0.000168	0.013086
Pcgf6	457.0506	-2.14003	0.590715	-3.61146	0.000304	0.016091
Slk	482.6081	-1.87283	0.55764	-3.36301	0.000771	0.02674
Btbd1	559.6303	-2.14819	0.577468	-3.71656	0.000202	0.014114
Hdgfl3	1623.794	-1.88713	0.559578	-3.34855	0.000812	0.02742
Bhlhe22	41.21848	-2.05418	0.657084	-3.1789	0.001478	0.037828
Alyref	1661.681	-2.29382	0.595962	-3.83499	0.000126	0.011255
Atp23	184.153	-2.06594	0.611976	-3.45107	0.000558	0.022603
Larp1b	37.3552	-1.92039	0.560287	-3.43901	0.000584	0.022987
Iscu	228.1494	-2.56043	0.640785	-4.1333	3.58E-05	0.006762
Get4	304.9434	-1.74853	0.581734	-3.02946	0.00245	0.049423
Sgk3	200.0401	-1.59172	0.520849	-3.05449	0.002254	0.04701
Carf	663.121	-1.86371	0.608642	-3.10977	0.001872	0.04279
Tex30	2466.73	-1.70307	0.536965	-3.17149	0.001517	0.038273
Gls	1094.7	-2.10607	0.555721	-3.76385	0.000167	0.013086
Mff	1762.13	-2.22001	0.567096	-3.90031	9.61E-05	0.009925
Smap1	843.6188	-2.29852	0.573535	-4.00106	6.31E-05	0.008476
B3gat2	464.7154	-1.71827	0.536752	-3.19775	0.001385	0.036648
Agfg1	726.2541	-2.03598	0.558133	-3.62233	0.000292	0.016016
Septin2	217.0101	-2.69053	0.62414	-4.35075	1.36E-05	0.005106
Ilkap	1082.767	-1.57617	0.472842	-3.32727	0.000877	0.028419
Pam	373.6146	-2.33457	0.584152	-4.00038	6.32E-05	0.008476
Desi2	329.5925	-1.97286	0.602584	-3.30208	0.00096	0.029905
Opn3	74.06048	-2.10398	0.65024	-3.31834	0.000906	0.029206
Rgs7	37.13159	-2.0507	0.625066	-3.15646	0.001597	0.039089
Smyd2	344.8938	-2.10555	0.58622	-3.6161	0.000299	0.016091
Nmt2	11363.28	-1.98924	0.623761	-3.26129	0.001109	0.032504

Suv39h2	603.3908	-1.81707	0.55677	-3.26265	0.001104	0.032417
Bmi1	872.3661	-1.77736	0.52243	-3.38868	0.000702	0.024986
Spopl	173.3985	-1.84723	0.523884	-3.52048	0.000431	0.019161
Gad2	33.25633	3.375907	0.671647	4.51116	6.45E-06	0.00387
Olfm1	246.8552	-1.82815	0.581023	-3.15702	0.001594	0.039089
Zmynd19	386.2147	-1.68087	0.547632	-3.05929	0.002219	0.046544
Mtx2	860.5875	-2.06154	0.601643	-3.50461	0.000457	0.019788
Tcp11l1	149.9268	-1.76957	0.582997	-3.02559	0.002481	0.049716
Myef2	21118.88	-2.42093	0.638202	-3.90636	9.37E-05	0.009895
Jag1	551.701	-1.64583	0.504167	-3.25618	0.001129	0.032679
Crls1	305.5979	-1.75189	0.54908	-3.18309	0.001457	0.037667
Stard7	1422.68	-1.69402	0.55595	-3.05242	0.00227	0.047075
Snx5	2737.85	-2.18259	0.552967	-3.92944	8.51E-05	0.009403
Gnas	10889.82	-2.19361	0.595693	-3.69866	0.000217	0.014315
E2f5	737.2631	-2.11312	0.573012	-3.66573	0.000247	0.014998
Mrgbp	295.1123	-1.60504	0.506365	-3.15939	0.001581	0.038972
Gid8	528.7988	-2.0724	0.575863	-3.59451	0.000325	0.016554
Sla2	736.4606	-3.00186	0.644977	-4.82477	1.40E-06	0.001935
Rab5if	1007.818	-3.06618	0.651483	-4.89846	9.66E-07	0.001482
Ttc14	613.97	-1.76032	0.54591	-3.21596	0.0013	0.035328
Ccna2	4448.899	-2.26783	0.540668	-4.20461	2.62E-05	0.006224
Tsc22d2	305.7111	-1.61464	0.492501	-3.26613	0.00109	0.032228
Notch2	725.241	-1.91656	0.566236	-3.38891	0.000702	0.024986
X4933434E20Rik	3196.235	-2.21097	0.615084	-3.66129	0.000251	0.015127
Sfrp2	549.5	-2.08959	0.584795	-3.56393	0.000365	0.01782
Pla2g12a	296.2852	-1.75236	0.54692	-3.19758	0.001386	0.036648
Ppa2	406.7635	-1.8368	0.521269	-3.52013	0.000431	0.019161
Rnf115	918.1885	-1.84017	0.575643	-3.2213	0.001276	0.034953
Gclm	547.1231	-2.60886	0.597321	-4.38983	1.13E-05	0.004942
Rps20	7196.622	-2.09569	0.620843	-3.33473	0.000854	0.027797
Akirin2	1274.802	-3.16923	0.642853	-4.94616	7.57E-07	0.001306
Anp32b	5812.381	-1.8615	0.643538	-3.02677	0.002472	0.049594
Tmeff1	1098.198	-2.68356	0.612478	-4.33013	1.49E-05	0.005173
Ugcg	425.8936	-2.12241	0.559468	-3.76279	0.000168	0.013086
Bag1	1386.452	-2.96742	0.642731	-4.7966	1.61E-06	0.002018
Clta	1839.369	-1.96533	0.562734	-3.47522	0.00051	0.02116
Psip1	7119.883	-2.05352	0.597527	-3.45693	0.000546	0.022314
Mllt3	1663.532	-1.23425	0.404572	-3.05181	0.002275	0.047075
Cdkn2c	109.9763	-2.68159	0.632274	-4.30281	1.69E-05	0.005173
Rnf11	530.0142	-2.14394	0.578108	-3.71096	0.000206	0.014114
Usp1	2729.966	-2.47158	0.594583	-4.15552	3.25E-05	0.006588
Plaa	1701.005	-1.91613	0.565891	-3.39693	0.000681	0.024664
Caap1	243.3042	-1.82587	0.564624	-3.23276	0.001226	0.034296

Ybx1	6635.346	-2.34394	0.596427	-3.94842	7.87E-05	0.009126
Cap1	819.1766	2.116962	0.591199	3.555205	0.000378	0.017916
Pithd1	290.373	-2.30511	0.60628	-3.82172	0.000133	0.011433
Khdrbs1	2603.828	-1.88044	0.560263	-3.34718	0.000816	0.027433
Sfpq	6961.671	-2.16471	0.601268	-3.61389	0.000302	0.016091
Yrdc	284.4421	-2.8016	0.622127	-4.51969	6.19E-06	0.00387
Ctnnbip1	506.2147	-1.84907	0.606725	-3.05243	0.00227	0.047075
Tprgl	296.2455	-2.04954	0.580784	-3.53191	0.000413	0.018881
Guf1	580.9186	-1.774	0.5225	-3.38103	0.000722	0.025431
Slc30a9	744.9493	-2.08792	0.56586	-3.70055	0.000215	0.014278
Chic2	250.8251	-1.93802	0.563546	-3.41633	0.000635	0.02407
Tmem165	451.5336	-1.73382	0.523787	-3.30326	0.000956	0.029905
Coq2	433.1326	-2.54136	0.604901	-4.23711	2.26E-05	0.006092
Ppbp	87.81704	1.857753	0.651904	3.078139	0.002083	0.045055
Mthfd2l	376.1141	-2.49921	0.60314	-4.09532	4.22E-05	0.007098
Ccng2	847.0144	-1.66882	0.533398	-3.13409	0.001724	0.040932
Cdk2ap1	256.7258	-2.05812	0.607472	-3.34511	0.000823	0.027491
Gpc2	53.50159	2.283655	0.66648	3.444885	0.000571	0.022859
Cdk8	482.1148	-1.56686	0.510171	-3.06183	0.0022	0.046384
Tspan12	9248.551	-2.07753	0.639967	-3.34637	0.000819	0.027433
Wasl	565.0757	-1.75915	0.536889	-3.25444	0.001136	0.032811
Ezh2	9544.618	-2.44815	0.61494	-4.03998	5.35E-05	0.007607
Ppp1r35	351.8244	-1.89917	0.652124	-3.16303	0.001561	0.038764
Dlx5	38.42158	2.869809	0.663878	4.167336	3.08E-05	0.006558
Gars	2355.836	-2.16066	0.581194	-3.74576	0.00018	0.013491
Nt5c3	794.8066	-1.77855	0.525958	-3.38398	0.000714	0.025288
Luc7l2	1458.151	-1.68512	0.546405	-3.0678	0.002156	0.045866
Cbx3	973.6174	-1.90556	0.555595	-3.42134	0.000623	0.023894
Ino80b	49.17009	-2.01443	0.598167	-3.40179	0.000669	0.024447
Sumf1	136.123	-1.75997	0.560505	-3.13206	0.001736	0.040949
Ybx3	1256.695	-2.31923	0.588549	-3.94069	8.12E-05	0.009141
Strap	5434.713	-2.43263	0.604574	-4.04801	5.17E-05	0.007468
Golt1b	10520.7	-1.86437	0.629873	-3.0386	0.002377	0.048606
Cmas	420.1243	-2.20767	0.584026	-3.76644	0.000166	0.013086
Prmt3	531.4976	-1.98626	0.550905	-3.60675	0.00031	0.01622
Mylpf	434.0798	-1.89556	0.612768	-3.16093	0.001573	0.038906
Rab6a	1293.58	-2.02235	0.576278	-3.51991	0.000432	0.019161
Lipt2	95.70523	-2.18674	0.629373	-3.45133	0.000558	0.022603
Fus	7637.323	1.623805	0.508527	3.184521	0.00145	0.037622
Mettl9	1773.521	-2.13491	0.594159	-3.63381	0.000279	0.015799
Zranb1	722.5865	-1.66844	0.532298	-3.1306	0.001744	0.041023
Wee1	841.1142	-2.39222	0.573273	-4.15681	3.23E-05	0.006588
Eif3f	2987.677	-2.84533	0.643391	-4.51621	6.30E-06	0.00387

Hccs	494.0672	-2.06945	0.525842	-3.9355	8.30E-05	0.009243
Lamp1	1350.096	-2.33264	0.59087	-3.97853	6.93E-05	0.008749
Vegfc	208.8254	-2.02954	0.590285	-3.41196	0.000645	0.024216
Eri1	941.3959	-1.92361	0.577315	-3.35963	0.00078	0.0268
Smim19	164.306	-2.27187	0.618118	-3.683	0.000231	0.014746
Gsr	1099.977	-2.1882	0.591001	-3.74631	0.000179	0.013491
Rbpms	70.89464	-2.20548	0.603711	-3.62222	0.000292	0.016016
Sap30	555.9088	-2.48704	0.627539	-4.05268	5.06E-05	0.007468
Setd6	158.5756	-1.78941	0.591242	-3.05705	0.002235	0.04668
Dnaja2	4037.494	-1.81932	0.502227	-3.60375	0.000314	0.016339
Itfg1	1100.251	-1.98729	0.569573	-3.50011	0.000465	0.019999
Smarca5	3651.975	-2.18381	0.581417	-3.74307	0.000182	0.013491
Pgls	871.2397	-2.36567	0.621162	-3.87087	0.000108	0.010561
Cbfb	1696.879	-2.23613	0.561511	-3.98229	6.83E-05	0.008724
Panx1	399.4029	-2.24696	0.600236	-3.7393	0.000185	0.013622
Tmem170	953.1759	-1.91891	0.613404	-3.18252	0.00146	0.03767
Egln1	800.2867	-1.75436	0.550752	-3.19534	0.001397	0.036793
Dcun1d5	1086.193	-2.88969	0.605587	-4.76368	1.90E-06	0.002018
Oaf	164.3503	-1.69103	0.541016	-3.12604	0.001772	0.041383
Abhd12	212.3311	-1.69897	0.556271	-3.07383	0.002113	0.045227
Rdx	7073.445	-2.14881	0.544986	-3.95287	7.72E-05	0.009111
Fdx1	194.3871	-2.16398	0.616989	-3.55397	0.000379	0.017928
Lysmd2	87.7628	-2.16123	0.655226	-3.33643	0.000849	0.027734
Polr2m	2944.225	-2.43157	0.563955	-4.30307	1.68E-05	0.005173
Rp9	267.7134	-2.00147	0.578138	-3.49239	0.000479	0.020335
Irak1bp1	257.4226	-1.66605	0.512635	-3.25977	0.001115	0.03261
Ube2q2	402.3166	-2.36901	0.575716	-4.09502	4.22E-05	0.007098
Nptn	314.4666	-1.83187	0.559926	-3.24907	0.001158	0.033366
X2300009A05Rik	142.4366	-1.79657	0.567375	-3.13327	0.001729	0.040932
Crtap	128.367	-2.40723	0.635458	-3.82634	0.00013	0.011391
Lrrfip2	451.8043	-2.11427	0.60719	-3.48866	0.000485	0.020556
Ryk	725.9834	-2.06898	0.596649	-3.46024	0.00054	0.022106
Lmo2	456.821	-1.99545	0.55067	-3.69404	0.000221	0.014372
Mier3	2242.54	-2.0174	0.573282	-3.53071	0.000414	0.018881
Acsl3	1039.12	-2.06678	0.561379	-3.67908	0.000234	0.014767
Trim33	1069.754	-1.57444	0.502782	-3.1295	0.001751	0.041108
Dis3	735.6941	-1.67145	0.517791	-3.22074	0.001279	0.034953
Tmed7	1097.398	-2.58933	0.584742	-4.38841	1.14E-05	0.004942
Scaf11	2949.163	-1.63742	0.506332	-3.23245	0.001227	0.034296
Chst2	9396.531	-2.26498	0.609642	-3.81678	0.000135	0.011521
Pank1	375.3411	-2.2358	0.612018	-3.62899	0.000285	0.016001
Acbd6	403.4997	-2.08511	0.596512	-3.48608	0.00049	0.020651
Dnah8	15.58756	2.426161	0.668549	3.8563	0.000115	0.010985

Gde1	552.779	-1.80449	0.571301	-3.15975	0.001579	0.038972
Tpst1	286.204	-1.87681	0.576252	-3.27006	0.001075	0.032127
Srsf2	8506.413	-2.26657	0.572748	-3.95524	7.65E-05	0.009098
Lrrc58	16102.39	-2.40716	0.651699	-3.94512	7.98E-05	0.009141
Gas2l1	89.45255	-1.99731	0.601257	-3.31675	0.000911	0.029257
Senp6	1443.61	-1.88677	0.540463	-3.47216	0.000516	0.021338
Traf3ip1	537.3558	-2.24112	0.64553	-3.44147	0.000579	0.022987
Fam53c	1000.642	1.697894	0.517348	3.274706	0.001058	0.031741
Trim59	2024.311	-2.29444	0.55822	-4.1026	4.09E-05	0.007098
Spata6	189.9718	-1.61489	0.50849	-3.17689	0.001489	0.037984
Tiparp	258.1624	-1.82148	0.577881	-3.17562	0.001495	0.038057
Tbr1	641.3757	-2.02908	0.623661	-3.22054	0.001279	0.034953
Arhgap5	3403.18	-1.57262	0.480844	-3.26499	0.001095	0.032288
Ap2b1	1172.171	1.542864	0.50092	3.07683	0.002092	0.045055
Slc24a5	17796.64	-2.415	0.643363	-3.8806	0.000104	0.010348
Pdk3	1072.887	-1.77301	0.56858	-3.11696	0.001827	0.042274
Arx	740.0206	-1.98698	0.55751	-3.55289	0.000381	0.017928
Fbxo33	175.6937	-2.4015	0.637739	-3.71035	0.000207	0.014114
Cox18	190.1374	-1.94201	0.598104	-3.22072	0.001279	0.034953
Pawr	281.4539	-2.71785	0.65627	-4.08931	4.33E-05	0.00711
Fbxw2	889.5293	-1.7524	0.526139	-3.33603	0.00085	0.027734
Ube2r2	1834.617	-1.97536	0.581372	-3.38906	0.000701	0.024986
Naa30	611.3949	-2.22409	0.570641	-3.89389	9.86E-05	0.010087
Noa1	260.0884	-2.03335	0.622907	-3.28538	0.001018	0.03086
Ubac1	475.187	-1.88093	0.525372	-3.58127	0.000342	0.017164
Siah2	63.84019	-2.11373	0.637812	-3.26389	0.001099	0.032345
Thap11	765.1388	-1.9131	0.594598	-3.26792	0.001083	0.032228
Btg1	1356.068	-2.22138	0.58972	-3.805	0.000142	0.011805
Rfxap	168.7941	-1.91856	0.618723	-3.11166	0.00186	0.042588
Siah1a	346.2938	-2.07612	0.580979	-3.56411	0.000365	0.01782
X2410004B18Rik	546.5579	-1.97972	0.624617	-3.22666	0.001252	0.034669
Arhgap33	816.9	2.113988	0.578141	3.635872	0.000277	0.015738
Rap2b	1650.084	-1.6787	0.553803	-3.05115	0.00228	0.047108
Kdm1a	1684.458	-2.04348	0.565069	-3.61237	0.000303	0.016091
Kifbp	334.4442	-1.90044	0.599644	-3.19479	0.001399	0.036793
Zfp146	440.1385	1.925389	0.536263	3.595971	0.000323	0.016523
Sh3glb1	1099.426	-2.08829	0.546927	-3.80563	0.000141	0.011805
Rnf139	540.5114	-2.25228	0.548412	-4.07878	4.53E-05	0.007352
Rbm17	1254.451	-1.81819	0.502338	-3.62305	0.000291	0.016016
Mex3c	1440.93	-2.28517	0.594558	-3.83908	0.000123	0.011216
Uvssa	750.422	-1.84846	0.573586	-3.23936	0.001198	0.033818
Cib2	218.2381	-2.05276	0.608261	-3.43107	0.000601	0.023444
Cdkn1c	571.1349	-2.47512	0.600746	-4.10178	4.10E-05	0.007098

Cpe	1147.629	-1.75408	0.577441	-3.04884	0.002297	0.047331
Dexi	120.6724	-2.12034	0.634253	-3.35777	0.000786	0.026914
Cntln	564.3683	-1.838	0.538122	-3.4113	0.000647	0.024216
Kdm1b	1784.116	-1.31888	0.432001	-3.04988	0.002289	0.047238
Mfsd14b	634.7467	-1.57857	0.502422	-3.14292	0.001673	0.040227
Snx25	127.6922	-2.01612	0.60209	-3.30945	0.000935	0.029733
X1700066M21Rik	152.0486	-1.92641	0.605801	-3.17472	0.0015	0.038057
Zfp384	130.3291	-1.97183	0.603928	-3.24689	0.001167	0.033554
Abhd17c	795.7899	-2.33022	0.614287	-3.79729	0.000146	0.01202
Uqcrfs1	2015.226	-2.17553	0.622514	-3.5577	0.000374	0.017896
Grtp1	498.8961	-1.56377	0.487572	-3.20717	0.00134	0.035788
Ranbp9	590.2851	-2.19343	0.594869	-3.67083	0.000242	0.014767
Mcl1	1214.153	-1.85731	0.555483	-3.35278	0.0008	0.027269
Med30	326.8889	-2.91451	0.63738	-4.66265	3.12E-06	0.002699
Wdr26	1982.593	-2.05261	0.559275	-3.65888	0.000253	0.015164
Tspyl5	172.5357	-2.05971	0.558892	-3.68275	0.000231	0.014746
Camk1d	53.89051	-2.0933	0.643625	-3.07566	0.0021	0.04507
Akap7	106.3415	-2.14359	0.494504	-4.3141	1.60E-05	0.005173
Pcsk1n	76.18538	-2.37526	0.64639	-3.91005	9.23E-05	0.009895
Azi2	1001.058	-1.7228	0.532045	-3.2413	0.00119	0.033778
Igfbp2	2001.024	-2.08785	0.586962	-3.56521	0.000364	0.01782
Zdhhc2	493.0595	-1.64261	0.468045	-3.50779	0.000452	0.019636
Prrx2	39.73399	-2.09816	0.665886	-3.02797	0.002462	0.049542
Micu3	771.7059	-2.0836	0.582985	-3.59797	0.000321	0.016519
Cited2	176.0573	-1.76193	0.523856	-3.36346	0.00077	0.02674
Magi2	145.6825	-1.29963	0.420644	-3.0776	0.002087	0.045055
Rab11fip2	419.3121	-1.42041	0.450585	-3.15396	0.001611	0.039214
Elavl1	3749.419	-2.21673	0.581621	-3.8246	0.000131	0.011391
Plpp6	51.80461	-1.98943	0.594522	-3.34665	0.000818	0.027433
Mrps35	560.401	-1.42467	0.454574	-3.1266	0.001768	0.041375
Pnrc1	442.2034	-1.87297	0.588053	-3.20148	0.001367	0.036365
Hs2st1	1253.873	-1.9372	0.526786	-3.68864	0.000225	0.014611
Nsmce4a	976.1441	-2.58957	0.64397	-4.02144	5.78E-05	0.008066
Fam102b	367.752	-1.55621	0.508679	-3.0624	0.002196	0.046384
Ankib1	406.7121	-1.50156	0.493526	-3.03559	0.002401	0.048877
Abhd13	276.1708	-1.59894	0.518178	-3.08015	0.002069	0.04502
Arglu1	8428.184	-2.41749	0.605205	-4.05294	5.06E-05	0.007468
Rundc3b	128.5556	-2.05872	0.600876	-3.40803	0.000654	0.024216
Ofd1	237.8034	-1.45022	0.468696	-3.09361	0.001977	0.044079
Cop1	392.7154	-1.6966	0.530683	-3.19269	0.00141	0.03685
Dennd6a	547.644	-1.63501	0.51094	-3.19414	0.001402	0.036806
Gfer	355.2381	-2.15027	0.582606	-3.69598	0.000219	0.014372
Ripk2	58.59356	-1.91202	0.62824	-3.03098	0.002438	0.049339

Hectd2	161.6056	-1.8961	0.612136	-3.09329	0.001979	0.044079
Cdk13	660.7195	-1.8019	0.506776	-3.55011	0.000385	0.017948
Zfp281	883.7609	-1.50461	0.495609	-3.03563	0.0024	0.048877
Cep78	995.2724	-1.8021	0.574187	-3.18355	0.001455	0.037667
Pdcd7	328.9218	-2.2764	0.606211	-3.76383	0.000167	0.013086
Gpr45	1014.309	-1.99779	0.585338	-3.42379	0.000618	0.023745
Dlx1	223.1278	3.265935	0.648987	5.020061	5.17E-07	0.001019
Msantd2	490.0991	-2.11356	0.58967	-3.55048	0.000385	0.017948
Tent2	371.6844	-2.00079	0.568164	-3.52214	0.000428	0.019161
Bend6	33.59824	-1.94595	0.565615	-3.4259	0.000613	0.023745
Sanbr	877.7743	-2.21177	0.555569	-3.97478	7.04E-05	0.008761
Ammecr1	989.2925	-1.97397	0.58149	-3.41795	0.000631	0.02406
Cilp	30.70014	2.127201	0.619017	3.517606	0.000435	0.019222
Isl1	20.56001	2.233917	0.669656	3.07711	0.00209	0.045055
Pbrm1	2543.333	-1.76263	0.488059	-3.61205	0.000304	0.016091
Rbm45	78.36358	-1.80276	0.563759	-3.20045	0.001372	0.036425
Dmrt3	549.1394	-2.13201	0.621844	-3.53668	0.000405	0.018644
Radx	713.2125	-2.10966	0.646703	-3.39075	0.000697	0.024986
Usp22	4181.585	-1.8289	0.530182	-3.44366	0.000574	0.022896
Pwwp3b	513.457	-1.47564	0.466827	-3.14991	0.001633	0.039621
Bmt2	332.485	-1.83631	0.537361	-3.39898	0.000676	0.02457
Tmem64	144.5582	-2.25082	0.58431	-3.81968	0.000134	0.011457
Fbrsl1	279.9038	-1.93753	0.587773	-3.28511	0.001019	0.03086
Epop	161.4469	-1.95991	0.65251	-3.22003	0.001282	0.034953
Pgp	365.2987	-2.36518	0.58449	-4.09248	4.27E-05	0.007098
Pura	299.7961	-1.83468	0.572177	-3.20849	0.001334	0.035788
Mgat2	372.5773	-1.72944	0.523029	-3.29713	0.000977	0.030157
Als2cl	10.45842	1.308635	0.650072	3.116197	0.001832	0.042274
Rsbn1	449.2068	-1.70205	0.491897	-3.4464	0.000568	0.022797
Arf6	354.0537	-1.98021	0.571076	-3.46272	0.000535	0.021969
Foxo1	104.6036	-1.88098	0.579619	-3.22814	0.001246	0.034669
Grsf1	3045.71	-1.74624	0.536838	-3.25651	0.001128	0.032679
Dnajc21	297.7706	-2.1615	0.59853	-3.62566	0.000288	0.016016
Nhlrc1	32.82528	-1.69476	0.549834	-3.10052	0.001932	0.043711
Bod1	509.243	-2.09486	0.564737	-3.7268	0.000194	0.014015
H1f10	483.8118	-1.67041	0.672171	-3.17228	0.001512	0.038273
Zfp367	445.3808	-1.52879	0.481179	-3.16885	0.00153	0.038341
Pwwp2a	488.633	-2.05575	0.563663	-3.63726	0.000276	0.015718
Sft2d3	85.10656	-2.30362	0.586385	-3.94097	8.12E-05	0.009141
Megf8	243.5442	1.589326	0.518582	3.076247	0.002096	0.04507
Bola3	543.584	-2.17145	0.600658	-3.65732	0.000255	0.015164
Sox3	568.1374	-2.20447	0.66389	-3.67486	0.000238	0.014767
Zfp691	205.3421	-2.00589	0.628668	-3.22562	0.001257	0.034703

Insig1	951.8602	-2.03488	0.57147	-3.54848	0.000387	0.017948
Cxcr4	54.35538	-2.30082	0.634953	-3.76814	0.000164	0.013086
Epcam	149.5714	-2.11976	0.601835	-3.57708	0.000347	0.017377
Sys1	204.7267	-1.86603	0.587365	-3.1967	0.00139	0.03669
Sh3tc2	34.03447	1.981619	0.63385	3.090426	0.001999	0.0442
Gm4997	130.372	-2.01138	0.622057	-3.17175	0.001515	0.038273
Zfp36l2	138.8539	-2.40684	0.596579	-4.04761	5.17E-05	0.007468
Otulin	248.6311	-2.02054	0.622947	-3.26614	0.00109	0.032228
Eid2	124.7177	-1.96862	0.627305	-3.1809	0.001468	0.037673
Scand1	76.43346	-1.96784	0.656093	-3.07855	0.00208	0.045055
Lyrm4	1310.366	-2.32509	0.615053	-3.83166	0.000127	0.011263
Tapt1	175.0557	-1.7046	0.515054	-3.30378	0.000954	0.029905
X1110032F04Rik	35.70499	-2.29506	0.639307	-3.62843	0.000285	0.016001
D030056L22Rik	1054.794	-2.66767	0.630353	-4.29384	1.76E-05	0.00527
Lrrc4b	168.0238	-2.09946	0.642268	-3.37507	0.000738	0.025855
Dmrta2	584.1555	-2.09091	0.605723	-3.51492	0.00044	0.019338
Rab2a	1230.453	-2.00711	0.56674	-3.547	0.00039	0.017989
Mageh1	1018.891	1.268683	0.401306	3.162013	0.001567	0.03883
Abhd17b	470.5106	-2.45488	0.611707	-3.96378	7.38E-05	0.008933
Frat2	113.1416	-2.07941	0.61492	-3.40234	0.000668	0.024447
Fgfbp3	662.4584	-2.77954	0.64482	-4.20923	2.56E-05	0.006224
Msl3l2	57.42229	-2.05345	0.663083	-3.05986	0.002214	0.046527
Ppp1r2	740.3284	-2.08587	0.571526	-3.64186	0.000271	0.015699
Bri3	169.8813	-2.75121	0.647199	-4.38872	1.14E-05	0.004942
Hes5	854.328	-2.01637	0.596557	-3.42445	0.000616	0.023745
Isg20l2	988.973	-1.84783	0.561187	-3.28308	0.001027	0.031001
Zbtb33	210.4404	-1.67146	0.523709	-3.17495	0.001499	0.038057
Mcmbp	4555.474	-1.81289	0.579637	-3.14387	0.001667	0.040166
Bend7	32.43286	-1.97268	0.608507	-3.21958	0.001284	0.034953
Rnf149	170.1159	-2.24487	0.600656	-3.701	0.000215	0.014278
Nhlh2	356.5367	-1.98541	0.561834	-3.56189	0.000368	0.01789
Nog	96.5072	-1.99415	0.628957	-3.15152	0.001624	0.039473
Lemd3	140.6357	-2.29626	0.631993	-3.63843	0.000274	0.015718
Mex3d	346.4489	-2.0875	0.630235	-3.36165	0.000775	0.026752
Ccdc6	4875.461	-2.07236	0.599596	-3.52684	0.000421	0.019002
Tbc1d12	237.3353	-1.83101	0.599051	-3.08359	0.002045	0.044671
Sephs2	1180.437	-2.05012	0.556527	-3.68408	0.00023	0.014746
Gpr137c	137.4668	-1.6409	0.488631	-3.34258	0.00083	0.027496
Chchd10	30.66296	-1.92962	0.668157	-3.14705	0.001649	0.039871
Zfpm1	28.80639	-2.2589	0.674731	-3.50319	0.00046	0.019832
Zbtb45	633.2388	-1.74003	0.572324	-3.07534	0.002103	0.04507
Slc35c1	165.1691	1.754147	0.565954	3.101184	0.001927	0.043711
Plekhf2	304.4353	-1.75157	0.535702	-3.25648	0.001128	0.032679

Ubqln2	1662.122	-2.01168	0.586686	-3.41252	0.000644	0.024216
Snip1	340.0314	-2.11624	0.578024	-3.68	0.000233	0.014767
Foxc1	86.01017	-2.13626	0.656388	-3.30727	0.000942	0.02989
Lypd6	151.5596	-1.48778	0.467339	-3.17052	0.001522	0.038331
B3galt6	148.8128	-1.90568	0.625575	-3.08959	0.002004	0.0442
Camk2n2	76.50122	-2.31651	0.663022	-3.62195	0.000292	0.016016
Fam174a	128.9997	-2.08881	0.626887	-3.34295	0.000829	0.027496
Pcmtd1	365.0358	-1.62349	0.46943	-3.45177	0.000557	0.022603
Pcdh19	834.7027	-3.45908	0.511838	-6.84353	7.73E-12	2.13E-08
Ywhag	3726.871	1.635637	0.509454	3.206913	0.001342	0.035788
Crebzf	258.9908	-2.22679	0.608405	-3.6163	0.000299	0.016091
Irf2bp2	453.3781	-1.72086	0.557562	-3.094	0.001975	0.044079
Tbc1d22a	118.3789	1.480318	0.487817	3.041094	0.002357	0.048277
Tagap1	1262.239	-2.06351	0.62637	-3.41092	0.000647	0.024216
Pard3b	293.3659	-1.71965	0.54905	-3.15604	0.001599	0.039089
Ppp4r2	845.5866	-1.92321	0.579282	-3.30668	0.000944	0.02989
Acvr2a	645.0817	-1.78716	0.510501	-3.49299	0.000478	0.020335
Zfp622	378.6802	-2.71598	0.634601	-4.381	1.18E-05	0.004942
Serp2	69.04072	-2.26268	0.611286	-3.71049	0.000207	0.014114
Rnf103	201.7309	-1.88373	0.592995	-3.16933	0.001528	0.038341
Rap1b	1151.021	-2.36081	0.585528	-4.04675	5.19E-05	0.007468
Gas1	5053.6	-2.16894	0.603515	-3.61066	0.000305	0.016091
Adamts18	151.7233	-1.37647	0.387065	-3.55654	0.000376	0.017896
Kcnip1	68.58765	-1.97964	0.588644	-3.36855	0.000756	0.026407
Utp18	837.8375	-1.93496	0.602949	-3.27608	0.001053	0.031655
Cadps	406.2204	-1.50266	0.454702	-3.30419	0.000953	0.029905
Sytl5	14.39585	2.084604	0.648954	3.147078	0.001649	0.039871
Zfp771	179.5372	-2.69106	0.657198	-4.305	1.67E-05	0.005173
Set	63884.09	-2.56214	0.630862	-4.16373	3.13E-05	0.006558
Tmem158	38.28892	-2.47002	0.674502	-4.25799	2.06E-05	0.005811
Kcmf1	680.1863	-1.85529	0.508815	-3.63777	0.000275	0.015718
Srsf11	6006.605	-1.74281	0.553947	-3.16605	0.001545	0.038572
Zfp580	217.0933	-2.42533	0.644744	-3.80078	0.000144	0.011923
Dach1	999.985	-1.85405	0.591643	-3.13832	0.001699	0.040793
Rras2	286.5494	-2.2693	0.611913	-3.67843	0.000235	0.014767
Zbtb25	239.9765	-1.93649	0.568661	-3.38102	0.000722	0.025431
Cebpb	9.801939	-2.13968	0.655006	-3.35651	0.000789	0.02697
Ppp1r14b	432.332	-1.84321	0.621127	-3.03159	0.002433	0.049338
Ier5	635.6561	-2.38594	0.638553	-3.86872	0.000109	0.010561
Rpgrip1	3637.204	-1.73609	0.542923	-3.20755	0.001339	0.035788
E2f6	572.6441	-1.9759	0.510805	-3.85572	0.000115	0.010985
Zbtb8os	929.4516	-1.70281	0.517116	-3.30025	0.000966	0.030033
Gnai1	161.2336	-2.28968	0.571929	-3.96557	7.32E-05	0.008933

Bloc1s3	63.65705	-2.18278	0.642938	-3.42481	0.000615	0.023745
Mex3b	795.808	-2.44707	0.576098	-4.21247	2.53E-05	0.006224
Septin11	18544.94	-1.91645	0.565714	-3.40426	0.000663	0.024447
Palld	675.9444	-1.3767	0.452057	-3.04662	0.002314	0.04754
Sdhc	1865.736	1.13051	0.372661	3.034899	0.002406	0.048917
Phtf1	475.4731	-1.88275	0.561949	-3.34868	0.000812	0.02742
Znrf2	230.0664	-1.75444	0.563502	-3.08694	0.002022	0.04438
Memo1	1350.34	-1.62292	0.499129	-3.24454	0.001176	0.033691
Nap1l1	11641.07	-1.68975	0.512408	-3.29657	0.000979	0.030157
Rsl1	67.1727	1.871217	0.478709	3.905838	9.39E-05	0.009895
Sclt1	695.3713	-2.15907	0.586788	-3.65477	0.000257	0.015245
Unc5c	135.7365	-1.74133	0.543966	-3.18088	0.001468	0.037673
Taok2	936.5863	1.702954	0.563494	3.027222	0.002468	0.049592
Chpt1	181.0212	-1.70833	0.545765	-3.11279	0.001853	0.042496
Tpt1	11121.74	-2.82588	0.620596	-4.53303	5.81E-06	0.00387
Pop5	480.6789	-1.80507	0.578603	-3.16411	0.001556	0.03876
Golm2	677.9543	-1.9861	0.565431	-3.51302	0.000443	0.019415
Wtap	1238.505	-1.92506	0.619368	-3.12726	0.001764	0.041352
Bloc1s4	89.42492	-2.65196	0.634561	-4.16339	3.14E-05	0.006558
Tmem263	6741.408	-2.03021	0.622405	-3.32017	0.0009	0.029083
Rrs1	597.9542	-2.24173	0.580272	-3.85257	0.000117	0.011014
Ppm1b	472.744	-1.76667	0.537616	-3.28597	0.001016	0.03086
Exoc5	1372.996	-1.70431	0.553421	-3.10789	0.001884	0.042993
Sox21	711.2921	-2.11417	0.645746	-3.24197	0.001187	0.033778
Zic2	200.5121	-1.74806	0.563988	-3.09983	0.001936	0.043711
Med9	279.6148	-1.79404	0.553481	-3.2228	0.001269	0.034953
Rps12	7195.565	-2.77997	0.654012	-4.27914	1.88E-05	0.0054
Gm5088	14.55361	-1.99193	0.664371	-3.06901	0.002148	0.045751
Gspt1	1836.732	-2.42917	0.573595	-4.23407	2.30E-05	0.006092
Rapgef2	141.0878	-1.75937	0.568677	-3.06393	0.002185	0.04625
Tmem65	429.5983	-1.96849	0.576515	-3.40859	0.000653	0.024216
Tubb3	7113.567	1.561895	0.510012	3.047177	0.00231	0.047522
Mysm1	6089.453	-1.85405	0.614256	-3.07986	0.002071	0.04502
X9130023H24Rik	1531.066	-2.33359	0.625583	-3.78954	0.000151	0.012328
Mrpl42	9331.09	-1.74225	0.574195	-3.05721	0.002234	0.04668
Ccni	2399.055	-2.02479	0.577227	-3.52651	0.000421	0.019002
Ing2	174.8664	-1.89646	0.621113	-3.09099	0.001995	0.0442
Mapk1	1660.965	-1.66817	0.527115	-3.16626	0.001544	0.038572
Rnf217	145.9892	-2.13973	0.585355	-3.60663	0.00031	0.01622
Brpf3	158.4408	-2.08751	0.571835	-3.65004	0.000262	0.015402
Med14	793.9246	-1.65059	0.509462	-3.23377	0.001222	0.034277
Fam172a	334.5162	-1.50677	0.426372	-3.53079	0.000414	0.018881
Rab3ip	388.1334	-1.96543	0.586151	-3.36629	0.000762	0.026558

Lsm14a	1179.786	-1.68144	0.50265	-3.34853	0.000812	0.02742
Flvcr1	216.9947	-1.74599	0.470573	-3.70053	0.000215	0.014278
Nck2	78.28617	-2.09214	0.651726	-3.2725	0.001066	0.03192
Suds3	726.7196	-1.8104	0.576764	-3.15823	0.001587	0.039057
X4930444P10Rik	10.02773	-2.14021	0.665138	-3.21185	0.001319	0.035562
Insm1	214.5354	-2.10778	0.634039	-3.2399	0.001196	0.033818
Zfp467	165.7292	-1.82464	0.477018	-3.82419	0.000131	0.011391
Nr2f1	1044.285	-2.08393	0.565407	-3.72539	0.000195	0.01402
Gprin1	55.52628	-1.95242	0.621491	-3.08705	0.002022	0.04438
Fam8a1	416.9227	-1.82079	0.56939	-3.19215	0.001412	0.03685
Epc2	590.1034	-1.58074	0.490801	-3.21035	0.001326	0.035673
Hsf5	30.93654	-2.01854	0.6151	-3.30454	0.000951	0.029905
Nsmce3	232.4224	-2.64041	0.617899	-4.312	1.62E-05	0.005173
Eid2b	102.6072	-1.83909	0.58391	-3.15442	0.001608	0.039214
Cited4	39.69893	-2.21324	0.6738	-3.55697	0.000375	0.017896
Gad1	9.445418	3.22132	0.67243	4.399186	1.09E-05	0.004942
Klhl9	12524.82	-2.28068	0.616519	-3.757	0.000172	0.013187
Dnajc25	64.49461	-1.79794	0.579385	-3.09852	0.001945	0.043711
Ptges3	20957.92	-1.87691	0.59248	-3.18871	0.001429	0.037221
Jund	679.4948	-2.09596	0.673322	-3.61511	0.0003	0.016091
Pcnp	11932.22	-2.26435	0.620536	-3.67739	0.000236	0.014767
Nt5dc2	1890.261	-2.08851	0.603962	-3.48464	0.000493	0.020651
Nanos1	104.8374	-2.06594	0.597236	-3.4825	0.000497	0.020717
Rybp	76.05392	-1.91323	0.569798	-3.33961	0.000839	0.027577
Gpr27	115.3186	-2.05212	0.617873	-3.39614	0.000683	0.024664
Acot1	77.78849	-1.86364	0.607337	-3.09813	0.001947	0.043711
Nudt10	1045.309	-2.01454	0.594297	-3.44726	0.000566	0.02279
Nudt11	327.1073	-2.13934	0.603261	-3.57117	0.000355	0.01771
Mtarc2	373.8504	-2.60262	0.624337	-4.19722	2.70E-05	0.006322
Csnk1g3	574.5922	-1.85953	0.567369	-3.26932	0.001078	0.032142
Ap1ar	843.5022	-2.31874	0.568088	-4.06375	4.83E-05	0.007468
Mocs3	37.13205	-2.1437	0.6598	-3.34926	0.00081	0.02742
Kiz	342.6576	-1.5393	0.503494	-3.05223	0.002271	0.047075
Fam122a	112.3721	-2.63915	0.652988	-4.05384	5.04E-05	0.007468
Ptar1	2275.157	-1.92624	0.567407	-3.40329	0.000666	0.024447
Fjx1	371.6093	-1.82166	0.611092	-3.02945	0.00245	0.049423
Sp5	82.07443	-2.26691	0.615145	-3.84415	0.000121	0.011087
Kbtbd6	60.57458	-2.03347	0.587921	-3.4022	0.000668	0.024447
Gm12184	60.6136	-2.23715	0.609609	-3.61657	0.000299	0.016091
Nrarp	564.1365	-2.29902	0.633286	-3.71738	0.000201	0.014114
F8a	71.58797	-2.09741	0.617899	-3.38767	0.000705	0.025013
X1110065P20Rik	276.5859	-2.07791	0.566634	-3.6755	0.000237	0.014767
Msmp	15.83915	2.064611	0.671027	3.113573	0.001848	0.042454

Dact3	121.1852	-2.56829	0.650196	-4.14382	3.42E-05	0.006736
Gm4204	286.0434	-1.76077	0.535039	-3.28849	0.001007	0.030762
Gm11149	104.7267	1.786467	0.547083	3.238099	0.001203	0.033898
Rps12.ps5	17.0884	-2.45737	0.673317	-3.65376	0.000258	0.015245
Gm14398	23.24413	-2.15415	0.657314	-3.22866	0.001244	0.034669
Ccdc85c	710.7011	-2.11668	0.615398	-3.52227	0.000428	0.019161
A430018G15Rik	82.03473	-1.952	0.540896	-3.59901	0.000319	0.016515
X3010003L21Rik	100.8783	-1.50475	0.49139	-3.06123	0.002204	0.046384
Appbp2os	49.41331	-2.05231	0.574828	-3.54893	0.000387	0.017948
Gm15706	71.24825	-2.4534	0.578279	-4.19293	2.75E-05	0.006335
Gm11837	24.16787	-2.26791	0.62064	-3.73224	0.00019	0.013928
Hectd2os	70.84581	-1.86653	0.614481	-3.06457	0.00218	0.046222
Ier5l	73.14123	-1.9814	0.659629	-3.17862	0.00148	0.037828
Dlx6os1	116.7717	3.276631	0.667853	4.776373	1.78E-06	0.002018
Uba52	80.43405	-1.95126	0.546775	-3.5564	0.000376	0.017896
Mettl23	3520.026	-1.77444	0.513756	-3.44977	0.000561	0.022646
Ccdc71l	139.8974	-1.90113	0.604983	-3.08873	0.00201	0.044255
Serpine3	51.34526	-1.85921	0.609983	-3.06151	0.002202	0.046384
Eid1	2009.502	-2.24968	0.622597	-3.59682	0.000322	0.016523
Gm3636	32.27999	-1.87559	0.581993	-3.21181	0.001319	0.035562
Inafm1	50.50059	-2.34799	0.642884	-3.72802	0.000193	0.014015
Fancf	99.43488	-2.91396	0.671464	-4.71625	2.40E-06	0.002369
Zfp131	1214.023	-1.62287	0.517332	-3.13697	0.001707	0.040845
Erdr1	13.34071	-2.03727	0.673191	-3.11581	0.001834	0.042274
Sox1	755.269	-2.49544	0.636834	-3.90852	9.29E-05	0.009895
Erdr1.1	169.8346	-2.32736	0.647101	-3.67133	0.000241	0.014767
X9530082P21Rik	1052.128	-2.18406	0.572658	-3.83286	0.000127	0.011263
X1600020E01Rik	129.5588	-1.55765	0.476972	-3.25699	0.001126	0.032679
Mir155hg	20.43745	-2.26565	0.665964	-3.33958	0.000839	0.027577
Ppp1ccb	54.3993	-1.89785	0.605334	-3.12115	0.001801	0.041864
Chaserr	731.1495	-2.04629	0.581531	-3.52642	0.000421	0.019002
Zbtb11os1	26.27669	-2.13103	0.666492	-3.11689	0.001828	0.042274
Rpl9.ps8	32.74669	-2.00069	0.639701	-3.2364	0.00121	0.034031
Gm4332	44.27416	-2.28756	0.630647	-3.59156	0.000329	0.016682
Peg13	882.7972	1.796291	0.525082	3.400656	0.000672	0.024485
X0610012G03Rik	262.4339	-2.52842	0.630557	-4.09935	4.14E-05	0.007098
Spata5l1	57.86214	-1.8677	0.567864	-3.29738	0.000976	0.030157
Tusc3	952.9003	-2.08281	0.550757	-3.76126	0.000169	0.013086

Appendix 3. Table illustrating all the DE genes from the HET vs. WT male comparison.

Gene	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
Axl	835.5999	-2.01724	0.502843	-3.99653	6.43E-05	0.027251
Phlda2	12.11684	2.140569	0.670869	4.033485	5.50E-05	0.024238
Rspo3	1179.378	2.558687	0.497842	5.082747	3.72E-07	0.000568
Otx2	209.7757	2.218099	0.589174	3.815272	0.000136	0.047244
Gucy1b2	15.1066	-2.69998	0.634751	-4.05514	5.01E-05	0.024238
Sphkap	21.29811	-2.65961	0.648831	-4.11088	3.94E-05	0.020724
Celf3	577.152	-2.5833	0.613579	-4.27502	1.91E-05	0.014626
Cap1	819.1766	-2.43634	0.570881	-4.26377	2.01E-05	0.014626
Rspo1	396.2456	2.485004	0.52215	4.684649	2.80E-06	0.003061
Gpc2	53.50159	-2.66261	0.658209	-4.11499	3.87E-05	0.020724
Chl1	63.41976	2.028531	0.498005	4.031102	5.55E-05	0.024238
Calml4	12.4704	2.614404	0.648113	3.849935	0.000118	0.042987
Acsbg1	394.9169	-2.0779	0.414357	-4.99199	5.98E-07	0.00083
Caskin2	531.9785	-2.12114	0.547861	-3.85915	0.000114	0.042407
Megf11	31.69209	2.536449	0.596401	4.133118	3.58E-05	0.020724
Arhgap33	816.9	-2.45461	0.557098	-4.39717	1.10E-05	0.010581
Pcp4l1	102.4233	2.170161	0.558444	3.883115	0.000103	0.039397
Six3	27.65317	3.003101	0.634918	4.800284	1.58E-06	0.001862
Bmp6	65.95845	3.3411	0.52371	6.219797	4.98E-10	1.27E-06
Zfp503	79.85142	2.985264	0.57083	5.294141	1.20E-07	0.000203
Galnt12	62.96334	2.371897	0.59456	3.990325	6.60E-05	0.027251
Cited2	176.0573	1.972194	0.50171	3.939058	8.18E-05	0.032895
Rab9b	22.41292	-2.11655	0.538795	-3.88647	0.000102	0.039397
Prokr1	367.1444	-2.03847	0.494882	-4.10355	4.07E-05	0.020724
Trpm3	125.4852	2.344858	0.632181	4.037507	5.40E-05	0.024238
Ttr	175.0716	1.27162	0.626646	4.827974	1.38E-06	0.001756
Gjd2	18.84359	-2.5674	0.61655	-4.12715	3.67E-05	0.020724
Apcdd1	330.5401	2.190328	0.507738	4.268477	1.97E-05	0.014626
Apol7d	60.17982	-2.27971	0.598616	-3.81801	0.000135	0.047244
Cenpw	2072.055	1.386553	0.318288	4.345419	1.39E-05	0.012495
Erich2	18.28445	2.803625	0.650858	4.230396	2.33E-05	0.015873
X2610035F20Rik	183.2511	2.362958	0.573542	4.049809	5.13E-05	0.024238
Siah3	453.5127	-2.02688	0.492029	-4.10389	4.06E-05	0.020724
Six3os1	133.892	3.110304	0.64734	5.53778	3.06E-08	5.85E-05
Mir6236	141.7404	-2.52922	0.589198	-4.31207	1.62E-05	0.01373
Pcdhga11	96.84793	-2.52835	0.57648	-4.39497	1.11E-05	0.010581
Peg13	882.7972	-2.13087	0.502559	-4.22503	2.39E-05	0.015873
Rmst	666.0662	1.422935	0.342948	4.128857	3.65E-05	0.020724
Beta-gal	352.0223	2.178111	0.672899	5.790311	7.03E-09	1.53E-05

Appendix 4. Table illustrating the DE genes between WT, HET and KOs in the RGC cluster.

Gene	p_val	avg_log2FC	p_val_adj	Genotype
Rn18s	0	0.748629	0	WT
Rps29	0	0.658243	0	WT
Rpl38	0	0.617416	0	WT
Rps28	0	0.559867	0	WT
Rps21	0	0.552144	0	WT
Rpl37	0	0.541346	0	WT
Rps27	0	0.532969	0	WT
Rpl35	0	0.518711	0	WT
Rpl41	0	0.491324	0	WT
Rpl39	0	0.478348	0	WT
Rpl37a	0	0.47583	0	WT
Rpl36	0	0.474343	0	WT
Rpl34	0	0.384628	0	WT
Rps26	0	0.355509	0	WT
Rplp2	0	0.348732	0	WT
betagal	0	-2.7575	0	WT
Rpl35a	2.84E-297	0.371921	7.96E-293	WT
Rpl36a	8.14E-276	0.382985	2.28E-271	WT
mt-Nd1	5.62E-274	-0.43297	1.57E-269	WT
Rpl6	3.25E-268	0.323883	9.10E-264	WT
Cox7c	1.71E-267	0.443272	4.78E-263	WT
Tmsb10	1.40E-262	0.450746	3.93E-258	WT
Gm10076	1.69E-262	0.505677	4.74E-258	WT
Atp5k	2.52E-256	0.492432	7.07E-252	WT
Snrpg	1.71E-250	0.404094	4.78E-246	WT
Pcdh19	3.86E-246	0.51265	1.08E-241	WT
Hnrnpa0	6.14E-244	0.366211	1.72E-239	WT
Atp5md	1.01E-182	0.39517	2.83E-178	WT
Rpl31	2.14E-181	0.279655	6.00E-177	WT
Tomm7	2.04E-180	0.386948	5.71E-176	WT
Atp5e	1.13E-162	0.334402	3.16E-158	WT
Romo1	8.92E-162	0.397283	2.50E-157	WT
Marcks	1.00E-161	-0.28834	2.81E-157	WT
Gm47283	3.29E-161	-0.44943	9.22E-157	WT
H4c4	1.60E-159	0.869513	4.49E-155	WT
Atp5j2	1.88E-158	0.329663	5.25E-154	WT
H4c8	1.14E-156	0.48659	3.19E-152	WT
Ndufb1	3.05E-150	0.379098	8.54E-146	WT
Atp5mpl	2.49E-140	0.343522	6.97E-136	WT
Cox6c	9.58E-139	0.310632	2.68E-134	WT

Snrpf	3.60E-137	0.313148	1.01E-132	WT
Cbx3	8.82E-133	-0.3173	2.47E-128	WT
Uqcr11	1.68E-131	0.329671	4.71E-127	WT
mt-Nd4	7.25E-130	-0.29351	2.03E-125	WT
Ndufc1	2.65E-126	0.327099	7.43E-122	WT
Mir6236	1.85E-120	0.373421	5.18E-116	WT
Atpif1	1.47E-114	0.289049	4.11E-110	WT
Sem1	2.18E-110	0.251688	6.11E-106	WT
Bola2	7.55E-110	0.321929	2.11E-105	WT
mt-Rnr1	8.82E-110	-0.29534	2.47E-105	WT
H4c3	3.10E-109	0.255791	8.67E-105	WT
H3c3	1.14E-106	0.604623	3.18E-102	WT
Hmga2	3.60E-106	-0.27962	1.01E-101	WT
H2bc14	6.29E-105	0.368031	1.76E-100	WT
mt-Nd5	1.23E-104	-0.3323	3.44E-100	WT
H3c6	6.02E-104	0.461165	1.69E-99	WT
Ndufa3	1.54E-103	0.301345	4.30E-99	WT
Ndufa5	6.79E-102	0.303526	1.90E-97	WT
Tma7	1.56E-101	0.250937	4.38E-97	WT
Dpysl2	3.48E-99	-0.39398	9.75E-95	WT
Mid1	1.26E-97	-0.40973	3.54E-93	WT
mt-Co1	1.95E-96	-0.28205	5.45E-92	WT
Uqcr10	2.05E-96	0.271215	5.73E-92	WT
Eif3j1	5.97E-94	0.289902	1.67E-89	WT
Ddx3x	1.70E-93	-0.32006	4.75E-89	WT
mt-Atp8	2.51E-91	0.276598	7.03E-87	WT
Eif5b	1.30E-89	0.250984	3.63E-85	WT
Gli3	4.15E-89	-0.32245	1.16E-84	WT
Xist	5.29E-89	-0.38949	1.48E-84	WT
Hspa8	1.76E-88	0.277117	4.93E-84	WT
H2bc11	3.33E-88	0.296683	9.32E-84	WT
Ddx18	5.82E-86	0.277078	1.63E-81	WT
H3c11	1.57E-84	0.283486	4.40E-80	WT
Slc24a5	3.64E-83	-0.39646	1.02E-78	WT
Lars2	9.17E-83	0.294004	2.57E-78	WT
H2ac4	1.29E-82	0.398392	3.62E-78	WT
Snhg6	1.04E-80	0.256756	2.90E-76	WT
Gm1673	6.03E-80	0.25723	1.69E-75	WT
Malat1	1.12E-79	0.297417	3.15E-75	WT
Sec62	7.09E-79	0.267311	1.99E-74	WT
ENSMUSG00000115801	4.40E-78	0.476183	1.23E-73	WT
H2ac15	1.01E-77	0.254439	2.82E-73	WT
Mrpl52	7.31E-77	0.25188	2.05E-72	WT

C130071C03Rik	1.90E-76	-0.28295	5.32E-72	WT
H2ac24	1.13E-75	0.507604	3.18E-71	WT
Cops9	1.28E-73	0.250365	3.59E-69	WT
Ptms	1.01E-69	0.282512	2.83E-65	WT
Msi2	3.57E-64	-0.28306	9.99E-60	WT
Ndn	9.75E-64	-0.31372	2.73E-59	WT
H2ac20	4.48E-60	0.425299	1.25E-55	WT
Rpgrip1	1.98E-56	-0.37762	5.55E-52	WT
mt-Atp6	2.41E-54	-0.303	6.75E-50	WT
Zfp36l1	1.49E-51	-0.26045	4.18E-47	WT
H2bc18	3.81E-51	0.28536	1.07E-46	WT
H2ac8	2.34E-48	0.625881	6.55E-44	WT
H1f3	1.52E-47	0.312583	4.26E-43	WT
Zfp862-ps	4.88E-43	-0.2591	1.37E-38	WT
H1f2	1.72E-39	0.269394	4.83E-35	WT
Cdca3	1.16E-36	-0.27426	3.26E-32	WT
Nnat	4.68E-34	0.255872	1.31E-29	WT
H1f4	2.28E-32	0.477928	6.38E-28	WT
Shisa2	9.61E-32	-0.29072	2.69E-27	WT
H1f1	7.55E-26	0.311269	2.11E-21	WT
Id3	8.19E-26	0.436567	2.29E-21	WT
Igfbpl1	1.39E-22	0.302194	3.88E-18	WT
Wnt8b	1.30E-15	0.30016	3.63E-11	WT
H1f5	0.000374	0.33879	1	WT
betagal.1	0	1.854185	0	HET
Rpl6.1	0	-0.36522	0	HET
Rplp2.1	0	-0.37895	0	HET
Rps26.1	0	-0.40311	0	HET
Rpl35a.1	0	-0.4273	0	HET
Rpl34.1	0	-0.43873	0	HET
Rpl36a.1	0	-0.43903	0	HET
Snrpg.1	0	-0.47824	0	HET
Cox7c.1	0	-0.50549	0	HET
Tmsb10.1	0	-0.52814	0	HET
Rpl37a.1	0	-0.5412	0	HET
Gm10076.1	0	-0.54276	0	HET
Rpl36.1	0	-0.54779	0	HET
Rpl41.1	0	-0.55694	0	HET
Atp5k.1	0	-0.55737	0	HET
Rpl39.1	0	-0.56059	0	HET
Rpl35.1	0	-0.57998	0	HET
Rps27.1	0	-0.61929	0	HET
Rps21.1	0	-0.63565	0	HET

Rpl37.1	0	-0.63646	0	HET
Rps28.1	0	-0.64045	0	HET
Rpl38.1	0	-0.69793	0	HET
Rps29.1	0	-0.72224	0	HET
Rn18s.1	0	-0.77643	0	HET
mt-Nd1.1	1.61E-280	0.420651	4.51E-276	HET
Rps15a	1.66E-274	-0.27474	4.65E-270	HET
Rps17	3.40E-260	-0.26961	9.51E-256	HET
Rpl31.1	2.50E-258	-0.32532	7.01E-254	HET
Atp5md.1	5.93E-255	-0.45164	1.66E-250	HET
Xist.1	8.87E-250	0.688404	2.48E-245	HET
Tomm7.1	1.53E-244	-0.43684	4.30E-240	HET
Atp5e.1	1.22E-238	-0.39396	3.42E-234	HET
Hnrnpa0.1	6.10E-229	-0.34792	1.71E-224	HET
Gm47283.1	1.80E-214	0.489699	5.05E-210	HET
Cox6c.1	1.70E-212	-0.37736	4.76E-208	HET
Romo1.1	8.09E-210	-0.43623	2.26E-205	HET
Snrpf.1	8.89E-209	-0.37838	2.49E-204	HET
Ndufb1.1	9.44E-205	-0.43219	2.64E-200	HET
Atp5j2.1	1.30E-203	-0.36191	3.65E-199	HET
Atp5mpl.1	6.52E-202	-0.40277	1.83E-197	HET
H4c8.1	6.05E-186	-0.50221	1.69E-181	HET
H4c4.1	1.06E-183	-0.93057	2.96E-179	HET
Uqcr11.1	6.34E-175	-0.3703	1.78E-170	HET
Ndufc1.1	6.78E-175	-0.373	1.90E-170	HET
Mir6236.1	2.62E-172	-0.47495	7.34E-168	HET
Marcks.1	1.20E-162	0.275266	3.36E-158	HET
Elob	4.72E-158	-0.2874	1.32E-153	HET
Ndufa3.1	2.73E-157	-0.36511	7.65E-153	HET
Pcdh19.1	4.24E-153	-0.38148	1.19E-148	HET
Sem1.1	5.84E-150	-0.28394	1.64E-145	HET
Atpif1.1	4.19E-144	-0.31234	1.17E-139	HET
Hspa8.1	5.98E-141	-0.34364	1.67E-136	HET
H4c3.1	2.68E-138	-0.26991	7.50E-134	HET
Bola2.1	1.52E-132	-0.33975	4.26E-128	HET
Ndufa5.1	2.89E-132	-0.33263	8.09E-128	HET
Mid1.1	4.33E-131	0.45855	1.21E-126	HET
H3c3.1	1.67E-129	-0.6676	4.67E-125	HET
Uqcr10.1	2.48E-129	-0.30825	6.95E-125	HET
H3c6.1	1.26E-125	-0.47533	3.54E-121	HET
Ddx3x.1	5.80E-125	0.356228	1.62E-120	HET
mt-Nd4.1	7.27E-122	0.271749	2.03E-117	HET
mt-Nd5.1	3.00E-120	0.345048	8.39E-116	HET

H2bc14.1	1.66E-119	-0.36734	4.63E-115	HET
Hmga2.1	1.05E-118	0.286728	2.93E-114	HET
Snhg6.1	1.90E-117	-0.30317	5.33E-113	HET
Tma7.1	2.64E-117	-0.25958	7.40E-113	HET
Lars2.1	8.44E-115	-0.31642	2.36E-110	HET
Ddx18.1	1.92E-113	-0.30191	5.38E-109	HET
Mrpl52.1	8.99E-113	-0.29876	2.52E-108	HET
Cops9.1	9.14E-113	-0.29905	2.56E-108	HET
mt-Atp8.1	7.95E-112	-0.28635	2.23E-107	HET
Ndufa2	1.71E-111	-0.27602	4.78E-107	HET
mt-Co1.1	2.29E-111	0.294233	6.41E-107	HET
H2ac4.1	1.02E-110	-0.46079	2.85E-106	HET
Slirp	4.08E-110	-0.28972	1.14E-105	HET
Eif3j1.1	9.02E-109	-0.29844	2.53E-104	HET
Uqcrq	5.96E-107	-0.2629	1.67E-102	HET
H3c11.1	4.78E-106	-0.29701	1.34E-101	HET
Gm1673.1	2.61E-103	-0.2846	7.31E-99	HET
Pttg1	2.81E-100	-0.28949	7.87E-96	HET
H2bc11.1	4.98E-100	-0.28887	1.39E-95	HET
Cbx3.1	3.63E-98	0.266943	1.02E-93	HET
Sec62.1	1.11E-96	-0.2856	3.12E-92	HET
Ndufa7	2.10E-96	-0.25729	5.88E-92	HET
Fam98b	1.25E-95	-0.27343	3.51E-91	HET
H2ac24.1	5.79E-94	-0.55629	1.62E-89	HET
Lsm7	9.39E-94	-0.26706	2.63E-89	HET
mt-Rnr1.1	3.88E-93	0.253088	1.09E-88	HET
Tmem258	4.47E-93	-0.2718	1.25E-88	HET
Gli3.1	3.42E-92	0.318492	9.58E-88	HET
Anapc13	9.19E-91	-0.26834	2.57E-86	HET
Ndufa1	3.14E-87	-0.26179	8.80E-83	HET
Ptms.1	5.15E-86	-0.30445	1.44E-81	HET
Malat1.1	5.15E-86	-0.31353	1.44E-81	HET
Mrps21	9.73E-86	-0.25023	2.73E-81	HET
Ubb	1.40E-78	-0.25408	3.91E-74	HET
ENSMUSG00000115801.1	1.72E-75	-0.46012	4.81E-71	HET
C130071C03Rik.1	9.71E-74	0.271143	2.72E-69	HET
Dpysl2.1	1.31E-70	0.329102	3.66E-66	HET
H2ac20.1	6.96E-70	-0.44404	1.95E-65	HET
Tsix	1.10E-66	0.292464	3.09E-62	HET
Slc24a5.1	1.22E-61	0.335228	3.43E-57	HET
H2bc18.1	1.53E-60	-0.30115	4.27E-56	HET
mt-Nd3	2.18E-58	-0.29143	6.10E-54	HET
H2ac8.1	1.00E-57	-0.6902	2.80E-53	HET

H1f3.1	6.74E-54	-0.312	1.89E-49	HET
Nnat.1	3.09E-52	-0.34149	8.65E-48	HET
Zfp36l1.1	3.20E-52	0.253844	8.95E-48	HET
Cdca3.1	1.99E-40	0.283069	5.56E-36	HET
H1f2.1	2.01E-40	-0.25005	5.62E-36	HET
Creb5	8.21E-39	0.252442	2.30E-34	HET
mt-Atp6.1	2.04E-34	0.255969	5.70E-30	HET
Rpgrip1.1	1.51E-33	0.30646	4.24E-29	HET
Id3.1	7.26E-32	-0.45178	2.03E-27	HET
H1f4.1	1.99E-29	-0.45235	5.56E-25	HET
Rspo3	2.90E-29	-0.25229	8.13E-25	HET
H1f1.1	1.17E-26	-0.31229	3.28E-22	HET
Shisa2.1	1.63E-26	0.269507	4.56E-22	HET
Igfbpl1.1	8.37E-23	-0.2786	2.34E-18	HET
Fabp7	5.45E-20	-0.29505	1.53E-15	HET
Zic1	8.78E-19	-0.28849	2.46E-14	HET
Wnt8b.1	5.95E-18	-0.31398	1.67E-13	HET
Ddx3y	0	0.529857	0	KO
Eif2s3y	0	0.454849	0	KO
Kdm5d	0	0.378954	0	KO
Uty	0	0.279745	0	KO
Xist.2	2.15E-218	-3.72407	6.03E-214	KO
Mt2	1.56E-198	0.987117	4.37E-194	KO
Tsix.1	3.93E-148	-1.32316	1.10E-143	KO
Rpl30	1.50E-105	0.521186	4.19E-101	KO
Rpl37.2	2.63E-99	0.519455	7.37E-95	KO
Rps21.2	8.66E-99	0.477096	2.43E-94	KO
Mt1	1.20E-96	0.936275	3.36E-92	KO
Rps27.2	1.29E-92	0.48549	3.62E-88	KO
Rpl39.2	6.26E-89	0.462726	1.75E-84	KO
Rpl41.2	1.28E-87	0.399705	3.57E-83	KO
Pttg1.1	1.85E-81	0.611682	5.19E-77	KO
Rpl37a.2	1.10E-80	0.396502	3.07E-76	KO
Rps29.2	5.74E-77	0.411672	1.61E-72	KO
Rps28.2	9.81E-76	0.466819	2.75E-71	KO
Rpl36.2	1.02E-74	0.428139	2.87E-70	KO
Rpl38.2	1.31E-74	0.472456	3.67E-70	KO
Hbb-y	6.83E-71	-0.72943	1.91E-66	KO
Snrpg.2	7.84E-67	0.419779	2.20E-62	KO
Rps15a.1	3.17E-64	0.297562	8.88E-60	KO
Rpl34.2	1.11E-61	0.33604	3.11E-57	KO
Rpl35.2	4.07E-60	0.386057	1.14E-55	KO
Rpl35a.2	8.03E-60	0.338945	2.25E-55	KO

Rps26.2	3.47E-59	0.303735	9.70E-55	KO
Cox7c.2	1.43E-58	0.37895	4.01E-54	KO
mt-Nd3.1	9.27E-56	0.599755	2.60E-51	KO
Rpl36a.2	6.60E-54	0.343715	1.85E-49	KO
Rps19	3.08E-53	0.341334	8.64E-49	KO
Rpl27	1.61E-51	0.443941	4.50E-47	KO
Cox6c.2	3.00E-51	0.37289	8.40E-47	KO
Tmsb10.2	2.92E-50	0.439984	8.17E-46	KO
Snrpf.2	2.60E-49	0.367434	7.27E-45	KO
Atp5e.2	5.16E-49	0.348826	1.45E-44	KO
Rpl31.2	4.88E-47	0.280014	1.37E-42	KO
Rpl6.2	1.67E-44	0.271094	4.67E-40	KO
Atp5k.2	3.71E-44	0.397172	1.04E-39	KO
betagal.2	5.46E-43	0.596385	1.53E-38	KO
Atp5md.2	1.26E-42	0.347614	3.53E-38	KO
Zic3	2.61E-42	0.275832	7.30E-38	KO
Zic4	1.04E-39	0.452552	2.90E-35	KO
Hspa8.2	1.97E-39	0.364986	5.51E-35	KO
Pcdh19.2	2.22E-39	-0.65684	6.20E-35	KO
Atp5mpl.2	9.25E-39	0.349328	2.59E-34	KO
Ubb.1	9.16E-38	0.412932	2.56E-33	KO
Rn18s.2	1.52E-37	0.267853	4.25E-33	KO
Ndufa3.2	2.07E-37	0.359101	5.80E-33	KO
Rpl15	3.61E-37	0.355104	1.01E-32	KO
Tomm7.2	5.61E-36	0.319126	1.57E-31	KO
Fabp7.1	1.46E-35	0.709473	4.10E-31	KO
Eif5a	2.62E-35	-0.25129	7.34E-31	KO
Rpl27a	8.60E-35	0.286849	2.41E-30	KO
Rpl24	9.65E-35	0.413292	2.70E-30	KO
Rpl23a	2.39E-34	0.270465	6.70E-30	KO
Rpl22l1	2.67E-33	0.269653	7.47E-29	KO
Mir6236.2	1.36E-32	0.518454	3.79E-28	KO
Rps6	1.38E-32	0.292419	3.87E-28	KO
Elob.1	2.26E-32	0.266057	6.33E-28	KO
Ndufb1.2	3.41E-31	0.33088	9.56E-27	KO
Zic1.1	1.64E-30	0.694171	4.60E-26	KO
Mir100hg	2.15E-29	0.304205	6.01E-25	KO
Gm47283.2	4.33E-29	-0.38134	1.21E-24	KO
Rps7	6.10E-29	0.308104	1.71E-24	KO
Ndufc1.2	7.05E-29	0.291006	1.97E-24	KO
Cops9.2	3.69E-28	0.286111	1.03E-23	KO
Uba52	2.00E-27	0.268791	5.59E-23	KO
Serf2	1.64E-25	0.302969	4.61E-21	KO

Romo1.2	4.61E-25	0.274965	1.29E-20	KO
Gm10076.2	1.08E-24	0.284297	3.03E-20	KO
Snhg6.2	2.32E-24	0.278091	6.48E-20	KO
Mrpl52.2	3.38E-24	0.278922	9.46E-20	KO
Slirp.1	3.73E-24	0.262728	1.04E-19	KO
Uqcr11.2	4.87E-24	0.269288	1.36E-19	KO
Rps10	1.14E-21	0.414548	3.19E-17	KO
Fabp5	5.02E-20	-0.2509	1.40E-15	KO
Tuba1a	6.68E-20	0.275105	1.87E-15	KO
Mid1.2	2.82E-19	-0.41031	7.90E-15	KO
Siva1	1.13E-18	-0.26616	3.15E-14	KO
Ddx3x.2	4.63E-18	-0.30197	1.30E-13	KO
Ccnd2	7.07E-18	-0.26779	1.98E-13	KO
Hba-x	9.63E-18	-0.39729	2.70E-13	KO
Nr2f1	1.25E-17	-0.4164	3.51E-13	KO
Ndufa12	2.33E-17	0.258087	6.52E-13	KO
H2ac4.2	2.14E-16	0.372185	5.99E-12	KO
Rspo1	2.07E-15	0.317903	5.80E-11	KO
Mest	2.62E-15	0.434457	7.33E-11	KO
Pantr1	6.51E-15	0.320058	1.82E-10	KO
Pbdc1	1.66E-14	-0.27264	4.63E-10	KO
Nnat.2	4.23E-14	0.435129	1.18E-09	KO
Ptn	3.68E-13	0.275582	1.03E-08	KO
H3c3.2	4.04E-11	0.402779	1.13E-06	KO
Eif2s3x	1.16E-10	-0.25599	3.25E-06	KO
H4c4.2	2.06E-10	0.408687	5.76E-06	KO
H2ac24.2	1.19E-09	0.333272	3.34E-05	KO
Mpped2	2.17E-07	0.258437	0.006084	KO
H2ac8.2	3.28E-05	0.410179	0.918416	KO

Appendix 5. Table illustrating the DE genes between WT, HET and KOs in the IP cluster.

gene	p_val	avg_log2FC	p_val_adj	cluster
betagal	1.94E-99	-2.5988	5.42E-95	WT
Cox7c	5.50E-60	0.582915	1.54E-55	WT
Atp5k	4.88E-42	0.524182	1.37E-37	WT
Hmga2	5.00E-40	-0.95817	1.40E-35	WT
Atp5e	3.21E-39	0.42344	8.98E-35	WT
Atp5md	8.67E-37	0.445033	2.43E-32	WT
Hnrnpa0	1.78E-36	0.385012	4.98E-32	WT
Cox6c	8.19E-36	0.40228	2.29E-31	WT
Tmsb10	2.02E-34	0.496978	5.64E-30	WT
Ndufb1	2.59E-33	0.471798	7.26E-29	WT
Sfrp1	2.92E-31	-0.75024	8.19E-27	WT
Top1	8.16E-31	0.352205	2.28E-26	WT
Gli3	2.92E-30	-0.63258	8.17E-26	WT
Gm47283	5.49E-30	-0.49785	1.54E-25	WT
Ndufc1	2.76E-29	0.396113	7.73E-25	WT
Lrrn1	3.93E-29	-0.66796	1.10E-24	WT
Cox6b1	3.83E-28	0.31558	1.07E-23	WT
Malat1	1.43E-27	0.392394	4.02E-23	WT
Pbx1	2.09E-27	-0.4713	5.85E-23	WT
Ndufa5	4.38E-27	0.440339	1.23E-22	WT
Romo1	1.57E-26	0.428808	4.38E-22	WT
Stox2	1.75E-26	-0.51648	4.90E-22	WT
Gm10076	2.88E-26	0.430971	8.07E-22	WT
Creb5	4.16E-26	-0.60924	1.16E-21	WT
Tomm7	1.28E-25	0.361345	3.59E-21	WT
Ddx3x	1.67E-25	-0.38584	4.67E-21	WT
Zfp36l1	2.29E-25	-0.61816	6.42E-21	WT
Sox9	7.32E-25	-0.65388	2.05E-20	WT
Gas1	1.00E-24	-0.6435	2.80E-20	WT
Eif5a	2.27E-24	-0.32024	6.36E-20	WT
Sem1	3.35E-24	0.29352	9.37E-20	WT
Atpif1	5.89E-24	0.354024	1.65E-19	WT
Rbm39	3.68E-23	0.301873	1.03E-18	WT
Trp53i11	4.88E-23	0.529344	1.37E-18	WT
Id4	5.72E-23	-0.63135	1.60E-18	WT
Ndufa3	1.31E-22	0.368005	3.68E-18	WT
Mir6236	1.39E-22	0.34387	3.91E-18	WT
Elob	1.61E-22	0.314645	4.49E-18	WT
P4hb	1.91E-22	-0.41811	5.35E-18	WT
Srrm4	3.57E-22	0.538321	1.00E-17	WT

Plagl1	6.40E-22	-0.50624	1.79E-17	WT
Chst2	8.37E-22	-0.46938	2.34E-17	WT
Cbx3	1.13E-21	-0.31275	3.17E-17	WT
Cox7a2	1.64E-21	0.30959	4.59E-17	WT
Tma7	4.23E-21	0.323301	1.18E-16	WT
Sox3	5.10E-21	-0.45554	1.43E-16	WT
Tubb3	2.22E-20	0.54962	6.22E-16	WT
Atp5j2	2.39E-20	0.319432	6.70E-16	WT
Tmsb4x	4.63E-20	0.504657	1.30E-15	WT
Cox6a1	5.42E-20	0.314047	1.52E-15	WT
Elavl4	3.48E-19	0.471215	9.73E-15	WT
Atp5mpl	4.06E-19	0.319818	1.14E-14	WT
Mest	4.33E-19	-0.62654	1.21E-14	WT
Mid1	1.14E-18	-0.38629	3.19E-14	WT
Hspa8	1.65E-18	0.324205	4.62E-14	WT
Igfbpl1	1.86E-18	0.548087	5.22E-14	WT
Cops9	2.42E-18	0.319985	6.78E-14	WT
Ptges3	2.71E-18	-0.28028	7.59E-14	WT
Rexo2	3.47E-18	-0.39557	9.71E-14	WT
Efnb1	1.44E-17	-0.38526	4.02E-13	WT
Ubb	1.86E-17	0.292401	5.19E-13	WT
Uqcr10	2.25E-17	0.286127	6.31E-13	WT
Ndufa2	2.87E-17	0.301965	8.03E-13	WT
Sec62	3.25E-17	0.333794	9.09E-13	WT
Slc24a5	4.34E-17	-0.39493	1.21E-12	WT
Kdm1a	5.08E-17	0.347416	1.42E-12	WT
Pabpc1	5.28E-17	-0.27406	1.48E-12	WT
Ttyh1	6.07E-17	-0.36728	1.70E-12	WT
Gja1	2.60E-16	-0.35358	7.28E-12	WT
Eif4g1	4.08E-16	-0.30902	1.14E-11	WT
Gnai2	4.09E-16	-0.31252	1.15E-11	WT
Hmgb2	5.36E-16	0.357353	1.50E-11	WT
Nfia	5.83E-16	-0.44604	1.63E-11	WT
Fut9	8.39E-16	-0.33756	2.35E-11	WT
Psat1	1.36E-15	-0.37923	3.81E-11	WT
Btg2	1.40E-15	0.528986	3.92E-11	WT
Rgcc	1.49E-15	-0.3714	4.16E-11	WT
Stk39	1.49E-15	-0.3622	4.16E-11	WT
Fam98b	1.56E-15	0.308458	4.37E-11	WT
Ptp4a2	1.88E-15	-0.30085	5.28E-11	WT
Mat2a	2.56E-15	-0.29253	7.15E-11	WT
Son	2.99E-15	0.258612	8.37E-11	WT
Cox5b	4.16E-15	0.265343	1.17E-10	WT

Ubl5	4.36E-15	0.268446	1.22E-10	WT
Uqcr11	4.66E-15	0.2892	1.30E-10	WT
Sall1	4.86E-15	-0.33135	1.36E-10	WT
Uqcrq	5.00E-15	0.281336	1.40E-10	WT
Ywhaz	5.89E-15	-0.25263	1.65E-10	WT
Alcam	2.06E-14	-0.29684	5.78E-10	WT
Paxbp1	2.25E-14	0.290047	6.31E-10	WT
Slirp	2.50E-14	0.294707	7.00E-10	WT
Igsf8	3.49E-14	0.58901	9.78E-10	WT
Mcmbp	3.93E-14	-0.27378	1.10E-09	WT
Hes1	6.07E-14	-0.28652	1.70E-09	WT
Tpi1	6.57E-14	-0.29372	1.84E-09	WT
Eomes	7.92E-14	0.51287	2.22E-09	WT
Tenm3	8.19E-14	-0.30139	2.29E-09	WT
Qk	8.72E-14	-0.32716	2.44E-09	WT
Cox7b	9.48E-14	0.252768	2.66E-09	WT
Mrpl12	1.04E-13	-0.29194	2.90E-09	WT
Msi2	1.23E-13	-0.30033	3.44E-09	WT
Cspg5	1.44E-13	-0.26172	4.02E-09	WT
Nrarp	1.48E-13	-0.26023	4.16E-09	WT
Dpysl2	2.02E-13	-0.30841	5.65E-09	WT
Bpnt2	2.09E-13	-0.29878	5.85E-09	WT
Nhp2	2.12E-13	-0.31402	5.94E-09	WT
Tnfrsf19	2.30E-13	-0.29389	6.44E-09	WT
Bcl11a	2.43E-13	-0.34878	6.79E-09	WT
Cdon	2.66E-13	-0.32015	7.45E-09	WT
Siva1	3.07E-13	-0.29692	8.60E-09	WT
Timm8a1	4.53E-13	-0.32787	1.27E-08	WT
Nhlh1	5.05E-13	0.412272	1.41E-08	WT
Sox2	5.61E-13	-0.38671	1.57E-08	WT
Anapc13	6.31E-13	0.269733	1.77E-08	WT
Lix1	7.06E-13	-0.32646	1.98E-08	WT
Eef1d	7.17E-13	-0.26718	2.01E-08	WT
Ndufa1	7.42E-13	0.293589	2.08E-08	WT
Ptprz1	1.90E-12	-0.28346	5.33E-08	WT
Ndufa7	2.28E-12	0.26145	6.39E-08	WT
Tagln3	2.44E-12	0.384672	6.84E-08	WT
Pkp4	2.65E-12	-0.26011	7.41E-08	WT
Atxn7	3.46E-12	-0.3051	9.68E-08	WT
Mrps24	4.28E-12	-0.27778	1.20E-07	WT
Hes6	4.29E-12	0.569128	1.20E-07	WT
Foxp1	5.37E-12	-0.32848	1.50E-07	WT
Nr2e1	5.39E-12	-0.30729	1.51E-07	WT

Tacc1	6.51E-12	-0.28681	1.82E-07	WT
Neurog2	6.70E-12	0.631099	1.88E-07	WT
Mrpl52	6.76E-12	0.251333	1.89E-07	WT
Ppp1r14a	7.86E-12	0.354078	2.20E-07	WT
Atic	7.94E-12	-0.28207	2.22E-07	WT
Fam110a	7.95E-12	0.379583	2.23E-07	WT
Xist	8.96E-12	-0.37146	2.51E-07	WT
Tmem256	9.85E-12	0.275478	2.76E-07	WT
Cdkn1c	1.07E-11	0.838615	3.00E-07	WT
Mrps5	1.24E-11	0.252198	3.48E-07	WT
Mpped2	1.29E-11	-0.37841	3.61E-07	WT
Tenm4	1.31E-11	-0.32327	3.66E-07	WT
Rfc4	2.10E-11	0.302733	5.88E-07	WT
Trps1	2.47E-11	-0.27573	6.91E-07	WT
Cplx2	3.38E-11	0.564905	9.45E-07	WT
Rpgrip1	3.68E-11	-0.36668	1.03E-06	WT
Peg12	4.20E-11	-0.25775	1.17E-06	WT
Ndn	4.38E-11	-0.25418	1.23E-06	WT
Ahi1	4.41E-11	0.271214	1.23E-06	WT
Eif4a3	4.56E-11	0.283447	1.28E-06	WT
Brsk2	7.90E-11	0.258233	2.21E-06	WT
Itga6	1.10E-10	-0.25265	3.08E-06	WT
Ezh2	1.13E-10	0.291772	3.16E-06	WT
Nmral1	1.16E-10	0.325763	3.24E-06	WT
Fat1	1.18E-10	-0.28441	3.32E-06	WT
Syt11	1.25E-10	-0.2666	3.50E-06	WT
Tmem14c	1.46E-10	-0.25988	4.09E-06	WT
Epha4	1.57E-10	-0.30127	4.40E-06	WT
Ndufv3	1.58E-10	0.254591	4.43E-06	WT
Mycn	1.60E-10	-0.30285	4.47E-06	WT
Adgrl2	1.95E-10	-0.3064	5.46E-06	WT
Tmem258	2.02E-10	0.26504	5.66E-06	WT
Cachd1	2.60E-10	-0.30581	7.27E-06	WT
1810037I17Rik	3.82E-10	0.295329	1.07E-05	WT
Ipo5	4.83E-10	-0.27309	1.35E-05	WT
Dtl	4.94E-10	-0.31381	1.38E-05	WT
Alkbh5	5.03E-10	-0.27618	1.41E-05	WT
Insm1	5.08E-10	0.539473	1.42E-05	WT
Gm3764	5.21E-10	-0.28692	1.46E-05	WT
Prdx6	5.79E-10	-0.25329	1.62E-05	WT
Mphosph8	6.46E-10	0.25954	1.81E-05	WT
Spsb4	7.11E-10	0.304144	1.99E-05	WT
Fdps	7.14E-10	0.310844	2.00E-05	WT

Kat6b	7.24E-10	-0.2844	2.03E-05	WT
Mfng	7.26E-10	0.315649	2.03E-05	WT
Nectin3	7.35E-10	-0.27694	2.06E-05	WT
Auts2	7.45E-10	0.262237	2.08E-05	WT
Arpp19	1.15E-09	0.251726	3.23E-05	WT
Npas3	1.20E-09	-0.27171	3.35E-05	WT
Gadd45g	1.29E-09	0.541387	3.61E-05	WT
Ubald2	1.40E-09	0.301268	3.93E-05	WT
Kdelr2	3.03E-09	-0.25047	8.49E-05	WT
Sox21	3.27E-09	-0.28545	9.17E-05	WT
Dnajc1	3.91E-09	-0.27252	0.000109	WT
Uri1	4.48E-09	-0.25116	0.000125	WT
Ly6h	5.83E-09	0.265967	0.000163	WT
Flrt3	9.12E-09	-0.27171	0.000255	WT
Ccdc174	9.57E-09	0.254389	0.000268	WT
Diaph3	1.09E-08	0.289197	0.000304	WT
Ppp2r2b	2.34E-08	0.395891	0.000654	WT
Miat	2.65E-08	0.299518	0.000742	WT
Akna	3.91E-08	0.2571	0.001094	WT
Lzts1	5.30E-08	0.260518	0.001485	WT
Rgs16	5.60E-08	0.277131	0.001567	WT
Rbm15	6.27E-08	-0.27501	0.001757	WT
Ncapd3	7.57E-08	0.285785	0.002119	WT
Cotl1	9.96E-08	0.252668	0.002788	WT
Kif21a	1.07E-07	0.260613	0.003009	WT
Ndc80	1.43E-07	0.259833	0.004002	WT
Prim1	1.46E-07	0.255562	0.004082	WT
Rrm2	1.60E-07	0.368462	0.004483	WT
Btbd17	2.13E-07	0.312902	0.005976	WT
Spc25	2.15E-07	0.296198	0.006028	WT
Neurog1	2.41E-07	0.379086	0.006754	WT
Pcdh19	3.06E-07	0.268269	0.008576	WT
Dll3	3.45E-07	0.263981	0.009662	WT
Mfap4	4.36E-07	0.269875	0.012201	WT
Kif22	6.23E-07	0.250193	0.01743	WT
Cdkn1b	6.58E-07	0.257194	0.018418	WT
Dll1	6.67E-07	0.320235	0.018685	WT
Fbxo5	1.04E-06	0.258037	0.029026	WT
betagal.1	2.48E-85	1.850779	6.94E-81	HET
Cox7c.1	4.08E-73	-0.62013	1.14E-68	HET
Ddx3y	9.85E-56	-0.25434	2.76E-51	HET
Atp5k.1	1.56E-53	-0.56983	4.37E-49	HET
Atp5e.1	3.50E-52	-0.47508	9.79E-48	HET

Cox6c.1	5.72E-50	-0.50378	1.60E-45	HET
Hmga2.1	2.00E-49	0.9985	5.61E-45	HET
Tmsb10.1	2.70E-48	-0.60562	7.57E-44	HET
Atp5md.1	3.63E-48	-0.49808	1.02E-43	HET
Ndufb1.1	3.28E-44	-0.53646	9.18E-40	HET
Gm47283.1	1.79E-40	0.544186	5.00E-36	HET
Sfrp1.1	1.23E-39	0.793962	3.45E-35	HET
Ndufc1.1	1.50E-39	-0.45949	4.21E-35	HET
Top1.1	3.20E-39	-0.37303	8.95E-35	HET
Gli3.1	5.42E-38	0.6663	1.52E-33	HET
Hnrnpa0.1	1.96E-37	-0.36767	5.48E-33	HET
Lrrn1.1	2.30E-34	0.676599	6.44E-30	HET
Cox6b1.1	3.90E-34	-0.33685	1.09E-29	HET
Eif5a.1	4.27E-34	0.365852	1.20E-29	HET
Tomm7.1	1.46E-33	-0.40592	4.10E-29	HET
Romo1.1	1.74E-33	-0.45904	4.87E-29	HET
Ndufa5.1	4.59E-33	-0.46031	1.28E-28	HET
Creb5.1	1.48E-32	0.636403	4.13E-28	HET
Gas1.1	2.11E-32	0.694743	5.90E-28	HET
Elob.1	2.47E-32	-0.3625	6.93E-28	HET
Ddx3x.1	3.22E-32	0.4096	9.02E-28	HET
Atpif1.1	4.38E-32	-0.39094	1.23E-27	HET
Zfp36l1.1	1.07E-31	0.653074	3.01E-27	HET
Malat1.1	1.40E-31	-0.40011	3.91E-27	HET
Sem1.1	2.79E-31	-0.33444	7.81E-27	HET
Gm10076.1	5.63E-31	-0.44527	1.58E-26	HET
Hmgb1	1.04E-30	-0.27108	2.91E-26	HET
Ndufa3.1	2.01E-30	-0.40149	5.63E-26	HET
P4hb.1	5.88E-30	0.460395	1.65E-25	HET
Pbx1.1	7.18E-30	0.465811	2.01E-25	HET
Sox9.1	1.17E-29	0.672318	3.26E-25	HET
Xist.1	1.23E-29	0.662211	3.45E-25	HET
Plagl1.1	1.66E-29	0.555069	4.66E-25	HET
Mir6236.1	1.68E-29	-0.5314	4.71E-25	HET
Stox2.1	4.30E-29	0.505056	1.20E-24	HET
Id4.1	1.71E-28	0.662715	4.80E-24	HET
Atp5j2.1	2.36E-28	-0.3623	6.59E-24	HET
Cox7a2.1	2.80E-27	-0.32698	7.84E-23	HET
Rbm39.1	1.16E-26	-0.3072	3.26E-22	HET
Sox3.1	1.22E-26	0.488449	3.43E-22	HET
Trp53i11.1	1.92E-26	-0.53135	5.38E-22	HET
Hspa8.1	3.02E-26	-0.3786	8.46E-22	HET
Atp5mpl.1	4.29E-26	-0.37357	1.20E-21	HET

Tubb3.1	6.43E-26	-0.58493	1.80E-21	HET
Mid1.1	1.18E-25	0.433715	3.30E-21	HET
Ubb.1	2.59E-25	-0.39391	7.24E-21	HET
Chst2.1	3.79E-25	0.451537	1.06E-20	HET
Igfbpl1.1	2.88E-24	-0.59188	8.06E-20	HET
Rexo2.1	6.67E-24	0.433643	1.87E-19	HET
Uqcr10.1	2.66E-23	-0.31893	7.44E-19	HET
Elavl4.1	9.71E-23	-0.51046	2.72E-18	HET
Cops9.1	2.36E-22	-0.35772	6.61E-18	HET
Mat2a.1	7.89E-22	0.332774	2.21E-17	HET
Ptges3.1	9.79E-22	0.279876	2.74E-17	HET
Cox6a1.1	1.03E-21	-0.30245	2.88E-17	HET
Rbm3	1.32E-21	0.253974	3.70E-17	HET
Srrm4.1	1.38E-21	-0.52432	3.87E-17	HET
Tma7.1	1.95E-21	-0.3114	5.45E-17	HET
Tmsb4x.1	2.00E-21	-0.48602	5.59E-17	HET
Efnb1.1	2.77E-21	0.407193	7.76E-17	HET
Psat1.1	2.94E-21	0.422313	8.24E-17	HET
Gnai2.1	3.67E-21	0.340418	1.03E-16	HET
Npm1	5.40E-21	0.2608	1.51E-16	HET
Eif4g1.1	5.78E-21	0.33517	1.62E-16	HET
Ndufa2.1	6.63E-21	-0.31766	1.86E-16	HET
Ubl5.1	1.13E-20	-0.30102	3.17E-16	HET
Pabpc1.1	2.15E-20	0.286683	6.01E-16	HET
Gja1.1	2.64E-20	0.370215	7.39E-16	HET
Fut9.1	4.98E-20	0.369936	1.39E-15	HET
Hmgb2.1	5.93E-20	-0.40779	1.66E-15	HET
Nhp2.1	8.51E-20	0.363002	2.38E-15	HET
Tagln3.1	1.27E-19	-0.47506	3.56E-15	HET
Siva1.1	1.33E-19	0.349551	3.72E-15	HET
Ndufa1.1	1.33E-19	-0.33763	3.72E-15	HET
Fam98b.1	1.40E-19	-0.33457	3.91E-15	HET
Sec62.1	2.14E-19	-0.33737	6.00E-15	HET
Eif3a	2.92E-19	0.267694	8.17E-15	HET
Uqcrq.1	3.53E-19	-0.30079	9.88E-15	HET
Hnrnpa2b1	5.28E-19	-0.26342	1.48E-14	HET
Timm8a1.1	6.06E-19	0.375366	1.70E-14	HET
Mest.1	6.50E-19	0.543342	1.82E-14	HET
Uqcr11.1	7.60E-19	-0.3144	2.13E-14	HET
Tpi1.1	8.73E-19	0.330965	2.44E-14	HET
Ttyh1.1	1.21E-18	0.361569	3.39E-14	HET
Eef1d.1	1.31E-18	0.309535	3.66E-14	HET
Slirp.1	1.57E-18	-0.31513	4.41E-14	HET

Ptp4a2.1	1.67E-18	0.310454	4.67E-14	HET
Son.1	2.31E-18	-0.26692	6.48E-14	HET
Nfia.1	2.69E-18	0.453435	7.54E-14	HET
Mrpl12.1	3.76E-18	0.314763	1.05E-13	HET
Sec61g	5.14E-18	-0.27618	1.44E-13	HET
Atxn7.1	5.72E-18	0.354308	1.60E-13	HET
Anapc13.1	7.73E-18	-0.32524	2.16E-13	HET
Rgcc.1	1.10E-17	0.384119	3.07E-13	HET
Alcam.1	2.16E-17	0.283683	6.06E-13	HET
Cdkn1c.1	2.91E-17	-1.09929	8.16E-13	HET
Lix1.1	4.50E-17	0.362115	1.26E-12	HET
Cox5b.1	4.85E-17	-0.25782	1.36E-12	HET
Ncl	4.97E-17	0.289693	1.39E-12	HET
Sall1.1	5.53E-17	0.336097	1.55E-12	HET
Sox2.1	5.95E-17	0.415692	1.67E-12	HET
Bcl11a.1	9.51E-17	0.375392	2.66E-12	HET
Qk.1	1.03E-16	0.338688	2.89E-12	HET
Cdon.1	1.11E-16	0.331898	3.11E-12	HET
Nr2e1.1	1.69E-16	0.347208	4.74E-12	HET
Tnfrsf19.1	3.38E-16	0.307584	9.46E-12	HET
Stk39.1	3.91E-16	0.329583	1.09E-11	HET
Pgls	4.64E-16	0.288591	1.30E-11	HET
Eomes.1	4.74E-16	-0.53757	1.33E-11	HET
Hes1.1	5.55E-16	0.296118	1.55E-11	HET
Mcmbp.1	6.55E-16	0.277326	1.83E-11	HET
Kdm1a.1	8.32E-16	-0.31592	2.33E-11	HET
Tmem256.1	1.02E-15	-0.31806	2.86E-11	HET
Paxbp1.1	1.26E-15	-0.31638	3.53E-11	HET
Btg2.1	1.35E-15	-0.52747	3.79E-11	HET
Bpnt2.1	1.60E-15	0.292048	4.49E-11	HET
Tuba1a	1.76E-15	-0.32685	4.93E-11	HET
Cox7b.1	2.81E-15	-0.25217	7.87E-11	HET
Micos10	3.78E-15	-0.2558	1.06E-10	HET
Flrt3.1	4.15E-15	0.341121	1.16E-10	HET
Ptbp1	6.00E-15	0.262016	1.68E-10	HET
Tenm4.1	6.29E-15	0.343007	1.76E-10	HET
Polr2k	6.38E-15	-0.28968	1.79E-10	HET
0610012G03Rik	7.44E-15	0.285159	2.08E-10	HET
Smarcd1	8.03E-15	-0.26678	2.25E-10	HET
Pin4	8.13E-15	-0.25632	2.28E-10	HET
Tmem14c.1	8.75E-15	0.296664	2.45E-10	HET
Ndufa7.1	9.21E-15	-0.26309	2.58E-10	HET
Tenm3.1	9.79E-15	0.282028	2.74E-10	HET

Dync1i2	1.25E-14	-0.27589	3.51E-10	HET
Msi2.1	1.34E-14	0.291911	3.75E-10	HET
Epha4.1	1.68E-14	0.342991	4.71E-10	HET
Gm3764.1	2.11E-14	0.330151	5.90E-10	HET
Nrarp.1	2.22E-14	0.257215	6.22E-10	HET
Auts2.1	2.40E-14	-0.29568	6.71E-10	HET
Mrpl52.1	2.60E-14	-0.2761	7.27E-10	HET
Tacc1.1	2.76E-14	0.297744	7.72E-10	HET
Trps1.1	3.56E-14	0.287195	9.96E-10	HET
Tmem258.1	3.70E-14	-0.29758	1.04E-09	HET
Mpped2.1	3.93E-14	0.380085	1.10E-09	HET
Mycn.1	3.99E-14	0.318206	1.12E-09	HET
Ostc	5.26E-14	0.268664	1.47E-09	HET
Cachd1.1	8.47E-14	0.341966	2.37E-09	HET
Ndufb2	1.17E-13	-0.254	3.28E-09	HET
Sfxn1	1.20E-13	0.280863	3.35E-09	HET
Pkp4.1	1.21E-13	0.263894	3.38E-09	HET
Ndufs5	1.38E-13	-0.2538	3.86E-09	HET
Fat1.1	1.55E-13	0.306996	4.34E-09	HET
Itga6.1	1.63E-13	0.27433	4.58E-09	HET
Nectin3.1	1.72E-13	0.315044	4.82E-09	HET
Scg5	1.96E-13	-0.26385	5.48E-09	HET
Atic.1	1.97E-13	0.283504	5.52E-09	HET
Fam110a.1	2.10E-13	-0.37707	5.87E-09	HET
Mrps24.1	2.13E-13	0.271008	5.97E-09	HET
Aprt	2.59E-13	0.302143	7.26E-09	HET
Brsk2.1	2.75E-13	-0.27385	7.71E-09	HET
Igsf8.1	3.34E-13	-0.5201	9.35E-09	HET
Ipo5.1	3.38E-13	0.302803	9.46E-09	HET
Foxp1.1	3.77E-13	0.326775	1.06E-08	HET
Cplx2.1	4.11E-13	-0.6002	1.15E-08	HET
Prdx6.1	4.17E-13	0.280561	1.17E-08	HET
Cox20	4.62E-13	-0.27982	1.29E-08	HET
Tcf7l1	5.01E-13	0.282344	1.40E-08	HET
Ppp1r14a.1	5.08E-13	-0.3411	1.42E-08	HET
1810037I17Rik.1	5.66E-13	-0.30769	1.58E-08	HET
Norad	6.26E-13	0.261659	1.75E-08	HET
Insm1.1	7.60E-13	-0.59467	2.13E-08	HET
Mrps5.1	7.77E-13	-0.25268	2.17E-08	HET
Hes6.1	7.93E-13	-0.5648	2.22E-08	HET
Ubald2.1	8.81E-13	-0.32194	2.47E-08	HET
Ptprz1.1	9.07E-13	0.2545	2.54E-08	HET
Nhlh1.1	1.05E-12	-0.36899	2.95E-08	HET

Nmral1.1	1.19E-12	-0.34319	3.32E-08	HET
Dnajc1.1	1.28E-12	0.305749	3.60E-08	HET
Rsl1d1	1.33E-12	0.256241	3.73E-08	HET
Bex1	1.42E-12	0.262648	3.97E-08	HET
Alkbh5.1	1.68E-12	0.293747	4.71E-08	HET
Syt11.1	1.93E-12	0.273536	5.41E-08	HET
Gtf3c2	2.09E-12	0.255718	5.84E-08	HET
Hs6st2	2.71E-12	0.254506	7.60E-08	HET
Kat6b.1	3.05E-12	0.304235	8.55E-08	HET
Ahi1.1	3.96E-12	-0.27185	1.11E-07	HET
Npas3.1	4.06E-12	0.270609	1.14E-07	HET
Uri1.1	6.81E-12	0.276706	1.91E-07	HET
Canx	7.11E-12	0.255807	1.99E-07	HET
Fdps.1	8.00E-12	-0.31261	2.24E-07	HET
Tbl1x	8.14E-12	0.263944	2.28E-07	HET
Rest	8.18E-12	0.26351	2.29E-07	HET
Nek6	8.30E-12	0.252942	2.32E-07	HET
Spsb4.1	1.13E-11	-0.32219	3.18E-07	HET
Eif4a3.1	1.16E-11	-0.2631	3.24E-07	HET
Gcsh	1.49E-11	0.263337	4.17E-07	HET
Mrps21	1.78E-11	-0.25493	4.97E-07	HET
Plpp3	2.15E-11	0.280882	6.01E-07	HET
Sox12	2.60E-11	0.265446	7.29E-07	HET
Ezh2.1	3.21E-11	-0.27643	8.98E-07	HET
Arpp19.1	3.41E-11	-0.25303	9.55E-07	HET
Polr2h	6.18E-11	0.256984	1.73E-06	HET
Serpinh1	7.01E-11	0.265802	1.96E-06	HET
Rrm2.1	7.41E-11	-0.42912	2.08E-06	HET
Fbxo5.1	8.62E-11	-0.3356	2.41E-06	HET
Rfc4.1	1.03E-10	-0.28505	2.89E-06	HET
Rgs16.1	1.14E-10	-0.30965	3.20E-06	HET
Ndufa12	1.25E-10	-0.26317	3.50E-06	HET
Neurog2.1	1.27E-10	-0.59588	3.55E-06	HET
Tomm5	1.35E-10	0.266224	3.77E-06	HET
Irf2bpl	1.45E-10	0.253492	4.05E-06	HET
Cotl1.1	1.51E-10	-0.27474	4.23E-06	HET
Fam136a	1.65E-10	0.269422	4.62E-06	HET
Serf2	1.72E-10	-0.26028	4.80E-06	HET
Spc25.1	1.88E-10	-0.36971	5.26E-06	HET
Elovl6	2.30E-10	0.252828	6.43E-06	HET
Dll3.1	2.37E-10	-0.29779	6.64E-06	HET
Cdca7	2.61E-10	0.253275	7.30E-06	HET
Dtl.1	2.75E-10	0.299037	7.71E-06	HET

Lzts1.1	2.84E-10	-0.29238	7.94E-06	HET
Hba-x	3.56E-10	-0.43513	9.97E-06	HET
Sox21.1	3.80E-10	0.273871	1.07E-05	HET
Mfng.1	4.91E-10	-0.30297	1.37E-05	HET
Lars2	6.55E-10	-0.25469	1.83E-05	HET
Cyth2	7.02E-10	-0.25219	1.97E-05	HET
Ebf3	7.59E-10	-0.25784	2.12E-05	HET
Phykpl	1.05E-09	-0.26472	2.93E-05	HET
Ly6h.1	1.14E-09	-0.25436	3.20E-05	HET
Klhl7	1.42E-09	-0.25411	3.97E-05	HET
Uncx.1	1.60E-09	-0.38109	4.48E-05	HET
Miat.1	2.04E-09	-0.30484	5.72E-05	HET
Akna.1	3.16E-09	-0.25325	8.84E-05	HET
Slc1a2	4.08E-09	-0.25246	0.000114	HET
Adgrl2.1	4.53E-09	0.265044	0.000127	HET
Nfib	5.41E-09	0.262778	0.000151	HET
Diaph3.1	5.61E-09	-0.27455	0.000157	HET
Ndc80.1	5.62E-09	-0.27358	0.000157	HET
Pclaf.1	6.41E-09	-0.37424	0.000179	HET
Neurod4.1	9.07E-09	-0.31917	0.000254	HET
Rfc3	1.14E-08	-0.26864	0.000319	HET
Nnat.1	1.29E-08	-0.55853	0.00036	HET
Hmgn5	1.32E-08	-0.33587	0.000368	HET
Kif21a.1	1.39E-08	-0.26152	0.000388	HET
Coro1c.1	1.48E-08	-0.29671	0.000414	HET
Btbd17.1	1.83E-08	-0.32791	0.000512	HET
Ncapd3.1	1.97E-08	-0.27491	0.000553	HET
Rpgrip1.1	3.26E-08	0.27386	0.000914	HET
Gadd45g.1	3.41E-08	-0.47948	0.000954	HET
Mfap4.1	3.86E-08	-0.25997	0.00108	HET
Nusap1.1	4.09E-08	-0.43206	0.001145	HET
Neurog1.1	4.56E-08	-0.38637	0.001276	HET
Cdkn2d	5.54E-08	-0.28085	0.001551	HET
Ppp2r2b.1	6.69E-08	-0.4053	0.001872	HET
Cyfip2.1	1.29E-07	-0.30479	0.003615	HET
Id1	3.61E-07	-0.28973	0.010121	HET
Clvs1	7.15E-07	-0.25167	0.020009	HET
Eif2s3y	8.03E-72	0.519226	2.25E-67	KO
Kdm5d	2.08E-58	0.500579	5.83E-54	KO
Ddx3y.1	8.70E-55	0.431049	2.44E-50	KO
Xist.2	1.38E-27	-3.9518	3.86E-23	KO
Tsix	2.90E-23	-1.30467	8.12E-19	KO
Lhx1os	6.01E-12	0.302502	1.68E-07	KO

Cox6c.2	6.86E-12	0.640436	1.92E-07	KO
Hsp90ab1	7.40E-12	-0.39493	2.07E-07	KO
Tmsb10.2	9.00E-12	0.706378	2.52E-07	KO
Six3	3.76E-11	0.443184	1.05E-06	KO
Lhx5	1.13E-10	0.343493	3.16E-06	KO
Atp5e.2	2.48E-10	0.445015	6.95E-06	KO
Zic4	3.70E-10	0.809656	1.04E-05	KO
Cox7c.2	2.65E-09	0.444566	7.43E-05	KO
Elob.2	4.69E-09	0.381031	0.000131	KO
Atp5md.2	4.70E-09	0.460101	0.000132	KO
Naca	4.81E-09	-0.28788	0.000135	KO
Ndufb1.2	6.78E-09	0.520546	0.00019	KO
Anp32b	8.02E-09	-0.37746	0.000224	KO
Gm47283.2	8.17E-09	-0.54953	0.000229	KO
Atp5k.2	8.47E-09	0.459619	0.000237	KO
Eif5a.2	1.09E-08	-0.4318	0.000305	KO
Ndufc1.2	1.15E-08	0.48367	0.000322	KO
Npm1.1	1.16E-08	-0.381	0.000325	KO
Pttg1	1.35E-08	0.540565	0.000378	KO
Eif3a.1	2.18E-08	-0.43969	0.00061	KO
Onecut2	2.64E-08	0.314327	0.000739	KO
Btf3	2.73E-08	-0.30857	0.000766	KO
Tagln3.2	4.38E-08	0.589341	0.001227	KO
Flrt3.2	5.82E-08	-0.54568	0.00163	KO
Ubb.2	1.18E-07	0.592326	0.00331	KO
Ldha	1.39E-07	-0.60586	0.003883	KO
Fam162a	1.94E-07	-0.5492	0.005434	KO
Atp5j2.2	2.22E-07	0.361126	0.006204	KO
Hspa8.2	2.40E-07	0.413794	0.006714	KO
Tuba1a.1	2.52E-07	0.556903	0.007047	KO
Ndufs5.1	2.76E-07	0.457155	0.007719	KO
Aprt.1	2.80E-07	-0.51537	0.007833	KO
Nnat.2	3.00E-07	1.166368	0.008401	KO
Serf2.1	3.02E-07	0.455633	0.008448	KO
Ndufa1.2	3.13E-07	0.354585	0.008776	KO
Hmga2.2	3.63E-07	-0.83743	0.010176	KO
Pkm	3.66E-07	-0.3751	0.01024	KO
Slc1a2.1	4.31E-07	0.380112	0.012055	KO
Atpif1.2	4.34E-07	0.349931	0.012139	KO
Sec61g.1	5.07E-07	0.376646	0.014205	KO
Fth1	5.84E-07	-0.35945	0.016365	KO
Cdv3	6.37E-07	-0.28314	0.017827	KO
Ndufa3.2	6.48E-07	0.340587	0.018157	KO

C130093G08Rik	6.85E-07	0.508587	0.019186	KO
Top1.2	8.23E-07	0.275058	0.023035	KO
Tomm7.2	9.24E-07	0.38738	0.02587	KO
Sfrp1.2	9.31E-07	-0.7065	0.026062	KO
Plagl1.2	9.35E-07	-0.5696	0.026176	KO
Nhp2.2	1.14E-06	-0.44757	0.031965	KO
Siva1.2	1.16E-06	-0.45857	0.032496	KO
P4hb.2	1.25E-06	-0.47434	0.03503	KO
Gas1.2	1.39E-06	-0.67617	0.038913	KO
Ncl.1	1.43E-06	-0.37818	0.039962	KO
Mid1.2	1.66E-06	-0.4838	0.046468	KO

Appendix 6. Table illustrating the DE genes between WT, HET, and KOs in the neuron cluster

gene	p_val	avg_log2FC	p_val_adj	cluster
betagal	8.17E-103	-1.57048	2.29E-98	WT
Pcdh19	2.69E-50	0.729376	7.53E-46	WT
Dpysl2	3.75E-33	-0.49145	1.05E-28	WT
Atp5k	3.47E-31	0.435304	9.72E-27	WT
Hnrnpa0	3.35E-27	0.285142	9.39E-23	WT
Ndn	1.11E-24	-0.49974	3.11E-20	WT
Cox7c	4.27E-24	0.312755	1.20E-19	WT
Slc24a5	4.30E-21	-0.51662	1.20E-16	WT
Atp5e	6.18E-18	0.298076	1.73E-13	WT
Scrt1	7.45E-18	0.310625	2.08E-13	WT
Atp5md	7.50E-18	0.289361	2.10E-13	WT
Cbx3	7.75E-18	-0.34462	2.17E-13	WT
Atp5j2	1.04E-17	0.295665	2.91E-13	WT
Gm10076	1.37E-17	0.307693	3.83E-13	WT
Cox5b	4.88E-17	0.270661	1.37E-12	WT
Tomm7	1.38E-16	0.270785	3.86E-12	WT
Gnb1	1.17E-15	-0.27844	3.27E-11	WT
Son	1.78E-15	0.264461	4.98E-11	WT
Fam98b	1.69E-14	0.278283	4.74E-10	WT
Anapc13	2.09E-14	0.268713	5.84E-10	WT
Lars2	2.40E-14	0.25258	6.71E-10	WT
Bcl11a	2.61E-14	-0.27033	7.31E-10	WT
Sem1	2.89E-14	0.260523	8.09E-10	WT
Uqcr10	3.01E-14	0.267118	8.42E-10	WT
Myef2	9.28E-14	-0.34161	2.60E-09	WT
Ndufb1	1.53E-13	0.266415	4.27E-09	WT
Scg5	1.16E-12	0.281849	3.25E-08	WT
Ndufv3	1.48E-12	0.255251	4.13E-08	WT
Cops9	5.82E-12	0.268397	1.63E-07	WT
Ndufa3	5.82E-12	0.25978	1.63E-07	WT
Tmem258	3.12E-11	0.250802	8.73E-07	WT
Kdm6b	2.50E-10	-0.28295	7.00E-06	WT
Mir6236	1.06E-09	0.254811	2.97E-05	WT
Zfp862-ps	2.43E-09	-0.29503	6.81E-05	WT
Gm44235	2.63E-08	-0.36685	0.000736	WT
Hba-x	2.84E-08	-0.56009	0.000794	WT
Xist	3.43E-65	0.805572	9.59E-61	HET
Tmsb10	4.70E-57	-0.40377	1.32E-52	HET
Atp5k.1	3.40E-53	-0.52693	9.51E-49	HET

Cox7c.1	6.13E-48	-0.4185	1.72E-43	HET
Atp5e.1	4.30E-41	-0.43212	1.20E-36	HET
Atp5md.1	4.87E-41	-0.42777	1.36E-36	HET
betagal.1	5.50E-41	0.941844	1.54E-36	HET
Lars2.1	1.84E-34	-0.32564	5.14E-30	HET
Gm10076.1	1.17E-31	-0.37393	3.28E-27	HET
Elob	2.44E-31	-0.31974	6.84E-27	HET
Hnrnpa0.1	1.17E-29	-0.28247	3.27E-25	HET
Ndufb1.1	7.86E-29	-0.36092	2.20E-24	HET
Pcdh19.1	2.69E-28	-0.4467	7.54E-24	HET
Cox5b.1	3.24E-27	-0.32112	9.07E-23	HET
Cops9.1	1.78E-26	-0.36644	4.99E-22	HET
Tomm7.1	3.76E-26	-0.32076	1.05E-21	HET
Atp5j2.1	6.74E-26	-0.32338	1.89E-21	HET
Ndufa3.1	2.11E-25	-0.34125	5.90E-21	HET
Anapc13.1	2.20E-25	-0.30866	6.16E-21	HET
Scrt1.1	6.27E-25	-0.33773	1.76E-20	HET
Mir6236.1	9.33E-25	-0.38721	2.61E-20	HET
Cox6c	2.10E-24	-0.29645	5.88E-20	HET
Atp5mpl	4.67E-24	-0.31485	1.31E-19	HET
Uqcr10.1	2.80E-23	-0.31562	7.84E-19	HET
Tmem256	8.07E-23	-0.28226	2.26E-18	HET
Gm1673	1.08E-22	-0.3164	3.01E-18	HET
Son.1	2.68E-22	-0.29161	7.50E-18	HET
Hspa8	3.08E-22	-0.30078	8.62E-18	HET
Sem1.1	9.40E-22	-0.29299	2.63E-17	HET
Ndufa1	4.77E-21	-0.26884	1.33E-16	HET
Scg5.1	1.22E-20	-0.30892	3.40E-16	HET
Romo1	1.37E-20	-0.29634	3.83E-16	HET
Tmem258.1	6.92E-20	-0.28507	1.94E-15	HET
Bcl11a.1	1.61E-19	0.297322	4.50E-15	HET
Fam98b.1	3.72E-19	-0.27125	1.04E-14	HET
Ndufc1	2.58E-18	-0.27301	7.23E-14	HET
Uqcr11	7.71E-18	-0.26868	2.16E-13	HET
Ndufa5	7.83E-18	-0.25921	2.19E-13	HET
Tsix	3.46E-16	0.422746	9.69E-12	HET
Sorl1	2.59E-15	-0.26014	7.25E-11	HET
Tle4	2.90E-15	-0.2566	8.12E-11	HET
Gm28050	2.95E-15	-0.26413	8.25E-11	HET
Dpysl2.1	6.74E-14	0.287961	1.89E-09	HET
Gm13404	2.49E-13	-0.27587	6.98E-09	HET
Meis2	7.33E-13	-0.53756	2.05E-08	HET
Ebf2	1.09E-12	0.581257	3.06E-08	HET

Tshz2	1.18E-11	-0.27994	3.30E-07	HET
Mir100hg	3.05E-11	-0.29235	8.54E-07	HET
Slc24a5.1	3.32E-11	0.371978	9.29E-07	HET
Cnr1	6.54E-11	0.343014	1.83E-06	HET
Nav1	1.11E-10	0.278642	3.11E-06	HET
Ndn.1	1.56E-10	0.284506	4.35E-06	HET
Nfib	4.29E-10	0.430227	1.20E-05	HET
Crabp2	7.86E-10	0.287551	2.20E-05	HET
Pbx3	2.98E-09	-0.51346	8.35E-05	HET
Bcl11b	3.05E-09	0.298997	8.54E-05	HET
Zic1	2.83E-08	-0.45206	0.000793	HET
Eomes	3.72E-08	0.327008	0.001041	HET
Uncx	8.35E-07	-0.26629	0.023386	HET
Kdm5d	3.06E-123	0.422347	8.57E-119	KO
Eif2s3y	8.28E-123	0.488956	2.32E-118	KO
Xist.1	3.30E-112	-3.53829	9.25E-108	KO
Uty	3.39E-111	0.371245	9.49E-107	KO
Ddx3y	3.49E-110	0.405366	9.77E-106	KO
Tsix.1	8.14E-57	-1.36907	2.28E-52	KO
Tmsb10.1	7.12E-33	0.488196	1.99E-28	KO
Hsp90ab1	2.34E-32	-0.30591	6.56E-28	KO
Hbb-y	4.51E-32	0.773168	1.26E-27	KO
Ubb	2.85E-31	0.491807	7.97E-27	KO
8030451O07Rik	3.94E-31	0.538849	1.10E-26	KO
Zfhx3	7.73E-24	0.672535	2.17E-19	KO
Mab21l1	8.39E-24	0.456501	2.35E-19	KO
betagal.2	2.37E-23	0.343793	6.64E-19	KO
Pbx3.1	1.83E-21	1.185076	5.11E-17	KO
Tmsb4x	8.16E-21	-0.35639	2.28E-16	KO
Mir99ahg	8.58E-20	0.437504	2.40E-15	KO
Mef2c	1.58E-19	0.569392	4.44E-15	KO
Elmod1	5.63E-18	0.279051	1.58E-13	KO
Nxph1	1.86E-17	0.477359	5.22E-13	KO
Hspa8.1	3.21E-17	0.385399	8.99E-13	KO
Mir100hg.1	5.86E-17	0.56817	1.64E-12	KO
Gm13404.1	1.80E-16	0.544865	5.03E-12	KO
Crabp2.1	2.62E-16	-0.59795	7.34E-12	KO
Atp5md.2	2.98E-16	0.362634	8.34E-12	KO
Atp5e.2	3.44E-16	0.355731	9.62E-12	KO

Mn1	7.75E-16	0.407171	2.17E-11	KO
Rian	5.48E-15	0.43271	1.54E-10	KO
Cox7c.2	9.08E-15	0.30257	2.54E-10	KO
Dynll1	2.29E-14	-0.31269	6.41E-10	KO
Tshz3	2.57E-14	0.270222	7.19E-10	KO
Plk2	3.37E-13	0.370104	9.44E-09	KO
Fabp5	3.69E-13	-0.2792	1.03E-08	KO
Gria2	5.66E-13	0.415845	1.58E-08	KO
Ebf2.1	9.60E-13	-0.88738	2.69E-08	KO
C130073E24Rik	1.46E-12	0.276116	4.10E-08	KO
Dlk1	1.80E-12	0.306748	5.05E-08	KO
Atp5mpl.1	2.97E-12	0.297129	8.31E-08	KO
Meg3	3.45E-12	0.676636	9.67E-08	KO
Mir6236.2	3.75E-12	0.344351	1.05E-07	KO
Tshz2.1	4.31E-12	0.465164	1.21E-07	KO
Atp5k.2	6.01E-12	0.293575	1.68E-07	KO
Nfib.1	7.30E-12	-0.6907	2.04E-07	KO
Zfhx4	7.35E-12	0.544811	2.06E-07	KO
Elob.1	1.41E-11	0.254626	3.94E-07	KO
AI504432	1.65E-11	0.273449	4.63E-07	KO
Fxyd6	1.69E-11	0.277501	4.73E-07	KO
Cox6c.1	1.79E-11	0.282599	5.02E-07	KO
Nsg2	4.98E-11	0.302826	1.39E-06	KO
Neurod1	5.49E-11	-0.49877	1.54E-06	KO
Gm1673.1	7.39E-11	0.272431	2.07E-06	KO
Cnr1.1	1.04E-10	-0.53594	2.93E-06	KO
Ndufb1.2	1.07E-10	0.271196	3.00E-06	KO
Sorl1.1	1.34E-10	0.300703	3.77E-06	KO
Cops9.2	1.42E-10	0.278749	3.99E-06	KO
Slirp	1.64E-10	0.254876	4.58E-06	KO
Gm28050.1	1.94E-10	0.357137	5.42E-06	KO
Eomes.1	3.37E-10	-0.63914	9.44E-06	KO
Gpm6a	3.40E-10	0.309619	9.52E-06	KO
Nav1.1	3.75E-10	-0.39742	1.05E-05	KO
2410006H16Rik	8.68E-10	0.254689	2.43E-05	KO
Eif5a	8.95E-10	-0.26772	2.50E-05	KO
Mrpl52	1.46E-09	0.278646	4.10E-05	KO
Lcorl	1.87E-09	0.290417	5.22E-05	KO
Brinp2	1.96E-09	0.309548	5.50E-05	KO
Pbx1	2.04E-09	0.323907	5.70E-05	KO
Meis2.1	3.63E-09	0.767613	0.000102	KO
Psma7	4.63E-09	-0.26096	0.00013	KO
Dclk1	4.91E-09	0.340747	0.000137	KO

Bach2	5.74E-09	0.32793	0.000161	KO
Runx1t1	2.47E-08	0.623382	0.000692	KO
Socs2	2.52E-08	0.380048	0.000705	KO
Vim	2.59E-08	-0.50004	0.000724	KO
Mfap4	2.84E-08	-0.47162	0.000795	KO
Zbtb20	3.28E-08	0.294222	0.000917	KO
Barhl2	3.45E-08	0.277708	0.000966	KO
Uncx.1	4.06E-08	0.268295	0.001135	KO
Ncam1	8.08E-08	0.310345	0.002262	KO
Baz2b	1.27E-07	0.409677	0.003558	KO
Mapt	1.57E-07	0.395923	0.004404	KO
Cdh6	1.73E-07	0.261561	0.004845	KO
Id4	2.25E-07	0.465597	0.006307	KO
Mycbp2	3.17E-07	-0.29758	0.008877	KO
Zic1.1	3.76E-07	0.537135	0.010536	KO
Cnih2	4.31E-07	0.253915	0.012058	KO
Hba-x.1	9.76E-07	0.892515	0.027325	KO
Nrp2	1.50E-06	0.275648	0.041882	KO

Appendix 7. Table illustrating the DE genes between HETs and WT females in the RGC cluster.

gene	p_val	avg_log2FC	pct.1	pct.2	p_val_adj
betagal	0	-2.74967	0	0.699	0
Cox7c	1.84E-288	0.519145	0.998	0.974	5.17E-284
Atp5k	7.66E-269	0.579326	0.986	0.827	2.14E-264
Tmsb10	1.55E-265	0.517935	0.998	0.982	4.34E-261
Hnrnpa0	6.47E-262	0.424881	0.999	0.997	1.81E-257
Gm10076	2.56E-257	0.560709	0.984	0.836	7.16E-253
Marcks	6.57E-187	-0.34626	1	1	1.84E-182
Fau	2.06E-186	0.255149	1	1	5.76E-182
Atp5md	2.75E-185	0.440871	0.988	0.892	7.69E-181
Tomm7	1.99E-184	0.434886	0.99	0.903	5.58E-180
Atp5e	1.64E-169	0.38511	0.996	0.973	4.59E-165
Pcdh19	1.69E-168	0.457573	0.813	0.437	4.74E-164
Romo1	3.82E-161	0.439334	0.954	0.781	1.07E-156
Atp5j2	3.13E-153	0.362564	0.997	0.97	8.76E-149
Ndufb1	4.20E-147	0.416619	0.961	0.757	1.18E-142
Cox6c	2.98E-144	0.353753	0.995	0.95	8.34E-140
Atp5mpl	2.70E-135	0.380668	0.982	0.873	7.57E-131
Mir6236	3.38E-134	0.487147	0.822	0.512	9.45E-130
Ndufc1	6.92E-134	0.380763	0.977	0.871	1.94E-129
Uqcr11	6.99E-128	0.364654	0.982	0.871	1.96E-123
Atpif1	2.58E-120	0.335387	0.995	0.97	7.23E-116
Sem1	5.93E-113	0.283605	0.998	0.988	1.66E-108
Hspa8	2.65E-111	0.353411	1	0.986	7.42E-107
Elob	8.40E-111	0.275685	0.996	0.98	2.35E-106
Top1	2.02E-105	0.272235	0.998	0.997	5.66E-101
Ndufa3	3.12E-104	0.33983	0.883	0.626	8.74E-100
Tma7	1.08E-103	0.281897	0.998	0.983	3.03E-99
Ndufa5	4.80E-103	0.345747	0.918	0.697	1.35E-98
Gm47283	3.61E-101	-0.36449	0.972	0.971	1.01E-96
Lars2	7.28E-101	0.349453	0.748	0.442	2.04E-96
Bola2	3.71E-99	0.342826	0.926	0.724	1.04E-94
Cox6b1	1.85E-98	0.258751	0.996	0.98	5.17E-94
Cbx3	2.35E-98	-0.30561	0.999	0.996	6.58E-94
Uqcr10	6.93E-98	0.312928	0.983	0.904	1.94E-93
Ddx3x	9.31E-96	-0.35614	0.973	0.964	2.61E-91
Rbis	5.48E-93	0.258957	0.64	0.344	1.53E-88
Mid1	1.74E-92	-0.43136	0.656	0.723	4.86E-88
Sox11	5.08E-90	-0.31015	0.999	0.998	1.42E-85
Hmga2	2.16E-87	-0.27828	0.999	1	6.04E-83

Eif5b	1.48E-86	0.279478	0.993	0.952	4.15E-82
Gm1673	4.14E-86	0.304056	0.982	0.904	1.16E-81
Gli3	7.09E-86	-0.35405	0.952	0.945	1.99E-81
Uqcrq	6.98E-85	0.269229	0.991	0.951	1.96E-80
Rbm3	1.08E-84	-0.25729	0.991	0.998	3.02E-80
Dpysl2	1.21E-84	-0.4095	0.881	0.865	3.38E-80
Eif3j1	1.38E-84	0.304717	0.915	0.686	3.86E-80
Ndufa7	3.04E-84	0.283371	0.991	0.923	8.51E-80
Ndn	3.45E-84	-0.41181	0.499	0.691	9.65E-80
Ddx18	1.16E-81	0.298915	0.886	0.654	3.24E-77
Slirp	1.58E-80	0.285254	0.924	0.708	4.42E-76
Mrpl52	4.72E-80	0.291825	0.953	0.822	1.32E-75
Sec62	6.07E-80	0.299648	0.922	0.717	1.70E-75
Cox5b	3.84E-78	0.252033	0.995	0.96	1.07E-73
Ndufa2	9.15E-78	0.265383	0.991	0.926	2.56E-73
Tmem258	1.80E-77	0.297508	0.908	0.707	5.05E-73
Dpm3	1.35E-76	0.253465	0.727	0.457	3.78E-72
Cops9	2.91E-76	0.284569	0.912	0.702	8.16E-72
Slc24a5	1.56E-75	-0.4183	0.59	0.673	4.36E-71
Lsm7	1.79E-73	0.272075	0.933	0.747	5.01E-69
Fam98b	8.11E-71	0.275657	0.874	0.63	2.27E-66
Ubb	1.09E-70	0.266857	0.997	0.976	3.07E-66
Pttg1	9.94E-69	0.26223	0.46	0.241	2.78E-64
Vim	6.41E-68	0.293576	1	0.997	1.80E-63
Matr3	1.20E-67	-0.28364	0.979	0.962	3.36E-63
Mrps21	4.59E-67	0.258287	0.92	0.712	1.29E-62
Zfp36l1	2.30E-66	-0.33158	0.956	0.948	6.43E-62
Malat1	2.49E-63	0.298107	0.999	1	6.96E-59
Pbx1	3.06E-63	-0.28189	0.982	0.966	8.57E-59
Anapc13	3.79E-63	0.259809	0.863	0.644	1.06E-58
Nars	6.93E-63	0.26048	0.927	0.747	1.94E-58
Ndufa1	5.18E-62	0.258341	0.861	0.623	1.45E-57
Msi2	4.58E-61	-0.31129	0.918	0.902	1.28E-56
Ptms	2.48E-60	0.287027	0.993	0.953	6.94E-56
Ndufa6	1.69E-59	0.252478	0.93	0.747	4.74E-55
Prpf4b	2.52E-59	0.250882	0.986	0.901	7.07E-55
Nnat	1.39E-58	0.381436	0.969	0.868	3.90E-54
Mphosph8	1.05E-55	0.259226	0.949	0.816	2.93E-51
Bcl11a	2.07E-52	-0.2876	0.942	0.919	5.81E-48
Plagl1	2.14E-50	-0.28162	0.932	0.915	6.01E-46
Rpgrip1	1.26E-48	-0.39224	0.474	0.555	3.52E-44
Ccnd2	1.75E-46	-0.28351	0.993	0.995	4.90E-42
Creb5	4.49E-46	-0.32046	0.859	0.854	1.26E-41

Epha4	5.33E-44	-0.28941	0.908	0.878	1.49E-39
Zfp862-ps	1.52E-37	-0.27361	0.366	0.47	4.26E-33
Nfib	4.53E-37	-0.26765	0.956	0.938	1.27E-32
Celf2	5.19E-35	-0.25472	0.899	0.869	1.45E-30
Cdca3	2.55E-31	-0.27757	0.839	0.825	7.13E-27
Mcmbp	1.20E-30	-0.26047	0.687	0.67	3.36E-26
Id3	9.70E-28	0.526868	0.362	0.238	2.72E-23
Sall1	1.60E-27	-0.26021	0.665	0.677	4.48E-23
Shisa2	2.60E-24	-0.27798	0.614	0.656	7.27E-20
Rspo3	5.20E-21	0.260185	0.255	0.162	1.46E-16
Nfia	4.23E-19	-0.25798	0.725	0.704	1.18E-14
Hes5	1.34E-18	0.277731	0.806	0.697	3.75E-14
Igfbpl1	1.44E-17	0.324101	0.541	0.436	4.04E-13
Fabp7	9.43E-14	0.280563	0.886	0.811	2.64E-09
Wnt8b	2.04E-11	0.306265	0.338	0.27	5.72E-07

Appendix 8. Table illustrating the DE genes between the HETs and the WT females in the IP cluster.

gene	p_val	avg_log2FC	pct.1	pct.2	p_val_adj
betagal	4.70E-78	-2.60702	0	0.857	1.32E-73
Cox7c	2.94E-55	0.644995	1	0.988	8.24E-51
Atp5k	1.06E-36	0.569842	0.979	0.915	2.95E-32
Hmga2	3.62E-36	-1.02317	0.894	0.948	1.01E-31
Hnrnpa0	1.88E-35	0.444902	1	0.998	5.27E-31
Sfrp1	7.85E-35	-0.92775	0.788	0.92	2.20E-30
Atp5e	3.44E-33	0.442954	1	0.979	9.63E-29
Ndufb1	1.01E-31	0.530555	0.968	0.903	2.82E-27
Cox6c	1.32E-31	0.452601	1	0.979	3.70E-27
Top1	1.02E-30	0.412028	1	1	2.85E-26
Tmsb10	3.44E-30	0.551592	1	0.995	9.64E-26
Atp5md	4.03E-30	0.450953	1	0.948	1.13E-25
Gli3	5.81E-30	-0.70809	0.767	0.921	1.63E-25
Creb5	2.19E-28	-0.72449	0.534	0.838	6.14E-24
Gas1	8.60E-28	-0.77276	0.778	0.923	2.41E-23
Ndufc1	1.05E-27	0.444865	0.984	0.952	2.95E-23
Cox6b1	1.65E-27	0.361003	1	0.993	4.63E-23
Lrrn1	4.44E-27	-0.72832	0.577	0.848	1.24E-22
Pbx1	1.50E-26	-0.52005	0.958	0.979	4.20E-22
Ddx3x	2.12E-26	-0.44583	0.979	0.995	5.93E-22
Marcks	2.94E-26	-0.29959	1	1	8.23E-22
Zfp36l1	3.05E-26	-0.71402	0.661	0.883	8.53E-22
Ndufa5	3.45E-26	0.480456	0.963	0.818	9.67E-22
Stox2	4.02E-26	-0.58673	0.794	0.912	1.13E-21
Gm10076	5.05E-26	0.491942	0.968	0.886	1.41E-21
Plagl1	5.19E-26	-0.63613	0.72	0.88	1.45E-21
Tomm7	1.08E-25	0.425639	0.995	0.948	3.03E-21
Sox9	3.20E-25	-0.75182	0.608	0.871	8.95E-21
Mir6236	1.53E-24	0.519741	0.873	0.599	4.29E-20
Malat1	2.42E-24	0.425168	1	1	6.77E-20
Id4	1.47E-23	-0.72342	0.878	0.936	4.11E-19
Atpif1	3.46E-23	0.384696	1	0.991	9.69E-19
Tma7	4.62E-23	0.386293	1	0.991	1.29E-18
Romo1	5.37E-23	0.46111	0.979	0.879	1.50E-18
Eif5a	1.64E-22	-0.34884	1	1	4.59E-18
P4hb	2.09E-22	-0.47145	0.799	0.906	5.86E-18
Elob	2.18E-22	0.348527	1	0.988	6.11E-18
Trp53i11	3.03E-22	0.577278	0.937	0.782	8.49E-18
Sem1	4.50E-22	0.324641	1	0.995	1.26E-17

Ndufa3	1.15E-21	0.409413	0.921	0.749	3.23E-17
Hmgb1	2.87E-21	0.250377	1	1	8.04E-17
Sox3	6.36E-21	-0.51587	0.656	0.853	1.78E-16
Srrm4	7.22E-21	0.616907	0.974	0.903	2.02E-16
Eif4g1	8.44E-21	-0.39804	0.984	0.975	2.36E-16
Ubb	1.31E-20	0.40659	0.995	1	3.67E-16
Igfbpl1	6.09E-20	0.657089	0.984	0.979	1.71E-15
Chst2	9.19E-20	-0.4915	0.354	0.684	2.57E-15
Tubb3	1.07E-19	0.652822	0.989	0.991	2.99E-15
Nfia	1.41E-19	-0.57898	0.64	0.855	3.94E-15
Gm47283	2.89E-19	-0.41254	0.974	0.983	8.10E-15
Rexo2	2.17E-18	-0.44778	0.683	0.849	6.07E-14
Pabpc1	3.96E-18	-0.3324	0.995	0.993	1.11E-13
Tmsb4x	4.19E-18	0.532317	1	1	1.17E-13
Rbm3	5.13E-18	-0.27768	1	0.998	1.44E-13
Ptges3	5.23E-18	-0.31092	1	1	1.46E-13
Rbm39	5.93E-18	0.318195	1	1	1.66E-13
Ttyh1	6.12E-18	-0.43158	0.381	0.697	1.71E-13
Atp5j2	9.94E-18	0.33473	1	0.974	2.78E-13
Cox6a1	1.83E-17	0.338992	0.989	0.991	5.12E-13
Efnb1	2.08E-17	-0.43201	0.884	0.939	5.83E-13
Elavl4	2.23E-17	0.524337	0.989	0.979	6.26E-13
Cox7a2	2.69E-17	0.316511	1	0.991	7.52E-13
Hspa8	5.44E-17	0.371295	1	0.998	1.52E-12
Fut9	5.92E-17	-0.39588	0.444	0.732	1.66E-12
Atp5mpl	6.29E-17	0.364487	0.989	0.943	1.76E-12
Cops9	2.16E-16	0.35048	0.947	0.873	6.05E-12
Gja1	2.90E-16	-0.39771	0.45	0.722	8.13E-12
Cdon	4.17E-16	-0.40041	0.524	0.755	1.17E-11
Bcl11a	5.97E-16	-0.44882	0.937	0.961	1.67E-11
Hes1	8.50E-16	-0.36094	0.354	0.66	2.38E-11
Tpi1	1.23E-15	-0.35991	0.984	0.989	3.43E-11
Hnrnpa2b1	1.31E-15	0.284916	1	1	3.67E-11
Hba-x	2.24E-15	0.338268	0.794	0.571	6.28E-11
Gnas	2.33E-15	0.252099	1	0.995	6.52E-11
Mpped2	2.43E-15	-0.54147	0.614	0.798	6.82E-11
Ndn	4.06E-15	-0.38088	0.492	0.796	1.14E-10
Mid1	4.59E-15	-0.37313	0.667	0.796	1.29E-10
Sall1	5.06E-15	-0.3717	0.481	0.741	1.42E-10
Alcam	6.02E-15	-0.35434	0.296	0.592	1.69E-10
Cdkn1c	7.31E-15	1.165813	0.873	0.803	2.05E-10
Ptp4a2	8.24E-15	-0.33775	0.974	0.983	2.31E-10
Stk39	8.86E-15	-0.39104	0.571	0.783	2.48E-10

Fam98b	9.34E-15	0.353895	0.921	0.762	2.62E-10
Rgcc	1.26E-14	-0.43632	0.413	0.685	3.53E-10
Cbx3	1.35E-14	-0.28747	0.995	0.993	3.78E-10
Mat2a	1.61E-14	-0.32048	0.979	0.976	4.50E-10
Gnai2	2.32E-14	-0.34175	0.995	0.987	6.50E-10
Tacc1	2.44E-14	-0.3575	0.582	0.798	6.83E-10
Npm1	2.62E-14	-0.25285	1	1	7.34E-10
Ndufa2	2.78E-14	0.325697	0.989	0.969	7.77E-10
Slc24a5	2.95E-14	-0.37628	0.794	0.903	8.25E-10
Nhp2	4.65E-14	-0.35876	0.921	0.934	1.30E-09
Mest	6.52E-14	-0.53551	0.577	0.778	1.83E-09
Scg5	6.75E-14	0.329489	0.656	0.453	1.89E-09
Tnfrsf19	8.42E-14	-0.33484	0.423	0.689	2.36E-09
Fdps	1.08E-13	0.421348	0.915	0.782	3.02E-09
Epha4	1.50E-13	-0.39516	0.82	0.895	4.21E-09
Psat1	1.50E-13	-0.39055	0.698	0.822	4.21E-09
Paxbp1	1.58E-13	0.330326	0.868	0.716	4.41E-09
Atxn7	1.67E-13	-0.35708	0.413	0.664	4.68E-09
Hmgb2	1.72E-13	0.375254	1	1	4.80E-09
Cspg5	2.42E-13	-0.29048	0.275	0.551	6.77E-09
Uqcr10	2.55E-13	0.284095	0.989	0.947	7.14E-09
Uqcrq	2.67E-13	0.295653	1	0.976	7.47E-09
Mcmbp	3.02E-13	-0.29493	0.693	0.818	8.45E-09
Sec62	3.22E-13	0.331758	0.952	0.876	9.02E-09
Kdm1a	3.28E-13	0.356414	1	0.983	9.17E-09
Igsf8	4.23E-13	0.664044	0.942	0.889	1.19E-08
Btg2	6.01E-13	0.581659	0.958	0.901	1.68E-08
Timm8a1	6.40E-13	-0.35977	0.598	0.758	1.79E-08
Foxp1	7.00E-13	-0.39215	0.73	0.838	1.96E-08
Msi2	8.72E-13	-0.32832	0.942	0.954	2.44E-08
Cnbp	9.52E-13	-0.26718	0.995	0.995	2.67E-08
Sox2	1.00E-12	-0.43172	0.81	0.917	2.81E-08
Eomes	1.08E-12	0.556244	1	1	3.03E-08
Kat6b	1.19E-12	-0.36773	0.825	0.906	3.34E-08
Cox5b	1.32E-12	0.278513	1	0.978	3.69E-08
Fam110a	1.37E-12	0.461694	0.905	0.792	3.84E-08
Ywhaz	1.39E-12	-0.2625	0.984	0.999	3.90E-08
Cul1	1.60E-12	0.292889	1	0.991	4.49E-08
Lsm2	1.71E-12	-0.27574	0.995	0.979	4.79E-08
Lix1	1.80E-12	-0.36689	0.608	0.811	5.04E-08
Nrarp	1.99E-12	-0.28112	0.36	0.618	5.56E-08
Tenm4	2.25E-12	-0.3669	0.603	0.788	6.30E-08
Pum2	2.46E-12	-0.2912	0.947	0.956	6.88E-08

Dync1i2	2.59E-12	0.29496	0.968	0.928	7.24E-08
Mrpl52	2.98E-12	0.296395	0.931	0.862	8.34E-08
Celf1	4.20E-12	-0.27882	0.915	0.949	1.18E-07
Qk	4.36E-12	-0.34322	0.947	0.967	1.22E-07
Uqcr11	5.31E-12	0.297133	0.974	0.943	1.49E-07
Matr3	5.54E-12	-0.28233	0.995	0.987	1.55E-07
Slirp	6.46E-12	0.302169	0.926	0.825	1.81E-07
Nr2e1	6.46E-12	-0.34452	0.54	0.758	1.81E-07
Atic	6.74E-12	-0.31368	0.545	0.768	1.89E-07
Tuba1a	7.79E-12	0.323132	1	1	2.18E-07
Tcf7l1	8.19E-12	-0.31562	0.529	0.698	2.29E-07
Ncl	8.56E-12	-0.27278	1	1	2.40E-07
Tagln3	9.35E-12	0.427986	0.968	0.928	2.62E-07
Ptbp1	1.11E-11	-0.27554	0.825	0.881	3.10E-07
Tenm3	1.11E-11	-0.30445	0.418	0.679	3.11E-07
Ndufa1	1.34E-11	0.32538	0.894	0.785	3.76E-07
Spsb4	1.36E-11	0.409925	0.831	0.702	3.80E-07
Rac1	1.49E-11	-0.25383	0.958	0.945	4.17E-07
Nectin3	1.49E-11	-0.33533	0.656	0.81	4.17E-07
Eif2s2	1.57E-11	0.260167	0.995	0.988	4.39E-07
Gm3764	1.72E-11	-0.35418	0.905	0.934	4.81E-07
Zfp862-ps	1.73E-11	-0.27508	0.397	0.646	4.86E-07
Eif4a3	1.77E-11	0.338536	0.974	0.942	4.95E-07
Cachd1	1.79E-11	-0.35639	0.571	0.738	5.01E-07
Dpysl2	2.10E-11	-0.3127	0.958	0.966	5.87E-07
Trps1	2.13E-11	-0.30935	0.497	0.725	5.95E-07
Calm1	2.17E-11	0.273169	1	1	6.07E-07
Bpnt2	2.44E-11	-0.30528	0.873	0.912	6.82E-07
Siva1	3.15E-11	-0.29982	0.963	0.971	8.81E-07
Flrt3	3.82E-11	-0.35208	0.646	0.811	1.07E-06
Usp1	3.85E-11	-0.31185	0.952	0.953	1.08E-06
Mycn	4.22E-11	-0.34566	0.577	0.749	1.18E-06
Norad	4.51E-11	-0.2854	0.889	0.936	1.26E-06
Itga6	4.75E-11	-0.2841	0.455	0.688	1.33E-06
Mphosph8	4.94E-11	0.300371	0.942	0.922	1.38E-06
Tsix	5.59E-11	0.337946	0.937	0.906	1.57E-06
Nfib	6.89E-11	-0.36312	0.91	0.947	1.93E-06
Micos10	7.89E-11	0.258834	0.989	0.966	2.21E-06
Eif5b	8.12E-11	0.2656	1	0.967	2.27E-06
Tmem256	8.74E-11	0.313904	0.836	0.685	2.45E-06
Uri1	9.19E-11	-0.30718	0.598	0.771	2.57E-06
Cox7b	9.21E-11	0.256059	0.995	0.982	2.58E-06
Dnmt3a	9.95E-11	-0.29581	0.735	0.834	2.79E-06

Tmem14c	1.05E-10	-0.28489	0.857	0.912	2.93E-06
Insm1	1.08E-10	0.669683	0.852	0.73	3.02E-06
Psmd4	1.12E-10	0.266619	0.989	0.973	3.15E-06
Ndufa7	1.31E-10	0.266303	0.989	0.969	3.67E-06
Calm2	1.37E-10	0.255721	1	1	3.83E-06
Hes6	1.60E-10	0.611119	0.942	0.928	4.47E-06
Pin4	1.65E-10	0.268451	0.757	0.619	4.63E-06
Nhlh1	1.87E-10	0.385212	0.746	0.581	5.23E-06
Tsg101	2.02E-10	0.276527	0.979	0.934	5.65E-06
Ipo5	2.02E-10	-0.31793	0.852	0.9	5.65E-06
Pkp4	2.48E-10	-0.26408	0.471	0.678	6.93E-06
Tmx2	2.48E-10	-0.27396	0.534	0.756	6.96E-06
Noc2l	2.80E-10	-0.29106	0.677	0.787	7.84E-06
Cplx2	2.92E-10	0.622078	0.725	0.606	8.17E-06
Klhl5	2.98E-10	-0.26582	0.365	0.616	8.34E-06
Mrps24	3.24E-10	-0.28425	0.894	0.925	9.08E-06
Nmral1	3.57E-10	0.365084	0.921	0.864	9.99E-06
Rpgrip1	3.90E-10	-0.37726	0.513	0.703	1.09E-05
Canx	4.06E-10	-0.28712	0.915	0.961	1.14E-05
Eef1d	4.10E-10	-0.26188	0.984	0.985	1.15E-05
Brsk2	4.52E-10	0.286877	0.746	0.621	1.26E-05
Rfc4	4.61E-10	0.321049	0.963	0.942	1.29E-05
Sox21	4.86E-10	-0.33808	0.651	0.803	1.36E-05
Rgs16	4.89E-10	0.378406	0.593	0.409	1.37E-05
Mrpl12	4.95E-10	-0.27459	0.91	0.952	1.39E-05
Tmem258	5.60E-10	0.301447	0.915	0.82	1.57E-05
Ahi1	5.61E-10	0.28952	0.783	0.643	1.57E-05
Rsl1d1	5.69E-10	-0.2771	0.958	0.967	1.59E-05
Cyth2	6.09E-10	0.290726	0.952	0.867	1.70E-05
Sfxn1	6.11E-10	-0.28312	0.878	0.916	1.71E-05
Syt11	7.90E-10	-0.2856	0.963	0.971	2.21E-05
Hs6st2	8.87E-10	-0.26752	0.566	0.712	2.48E-05
Ppp1r14a	9.75E-10	0.408516	0.646	0.469	2.73E-05
0610012G03Rik	1.05E-09	-0.26948	0.873	0.889	2.95E-05
Mrps5	1.29E-09	0.258023	0.873	0.777	3.61E-05
Gtf3c2	1.34E-09	-0.26227	0.725	0.836	3.76E-05
Prdx6	1.36E-09	-0.27949	0.926	0.945	3.82E-05
Vars	1.46E-09	-0.25525	0.794	0.882	4.09E-05
Zfp26	1.58E-09	-0.26513	0.476	0.671	4.43E-05
Dnajc1	1.76E-09	-0.30852	0.757	0.838	4.94E-05
Lars2	1.82E-09	0.30012	0.735	0.577	5.10E-05
Plpp3	1.96E-09	-0.29419	0.757	0.856	5.48E-05
Nek6	2.15E-09	-0.26599	0.444	0.627	6.02E-05

Mbtd1	2.34E-09	-0.25128	0.889	0.935	6.56E-05
Fat1	2.41E-09	-0.2977	0.635	0.778	6.76E-05
Kif21a	2.45E-09	0.324231	0.995	0.979	6.85E-05
Rest	2.58E-09	-0.27713	0.466	0.678	7.21E-05
Dtl	2.67E-09	-0.34107	0.656	0.795	7.47E-05
Pgls	2.75E-09	-0.25267	0.937	0.959	7.69E-05
Auts2	3.49E-09	0.299263	1	0.983	9.76E-05
Rfx7	3.52E-09	-0.28515	0.73	0.815	9.87E-05
Neurog2	3.82E-09	0.658652	0.868	0.88	0.000107
Anapc13	3.95E-09	0.252395	0.862	0.779	0.000111
Bex1	4.02E-09	-0.25787	0.958	0.96	0.000113
Pnrc1	4.28E-09	0.251505	0.931	0.856	0.00012
Aprt	4.44E-09	-0.29685	0.873	0.917	0.000124
1810037I17Rik	4.65E-09	0.322648	0.947	0.896	0.00013
Pom121	4.93E-09	-0.26446	0.783	0.87	0.000138
Sox12	5.46E-09	-0.28103	0.667	0.808	0.000153
Alkbh5	5.49E-09	-0.29685	0.741	0.829	0.000154
Ide	5.94E-09	-0.25912	0.704	0.827	0.000166
Dab1	6.18E-09	-0.25534	0.476	0.676	0.000173
Polr2k	6.18E-09	0.26476	0.931	0.869	0.000173
Miat	6.25E-09	0.340289	0.968	0.936	0.000175
Gcsh	6.49E-09	-0.26918	0.757	0.858	0.000182
Cdca7	7.60E-09	-0.2792	0.683	0.805	0.000213
Ostc	1.14E-08	-0.25416	0.894	0.93	0.00032
Ash1l	1.17E-08	-0.28087	0.889	0.921	0.000327
Mfng	1.36E-08	0.353577	0.915	0.873	0.00038
Arpp19	1.38E-08	0.265106	0.984	0.942	0.000388
Ezh2	1.48E-08	0.286746	1	1	0.000414
Foxg1	1.70E-08	-0.25905	0.979	0.98	0.000476
Cdca3	1.86E-08	-0.2955	0.862	0.938	0.000521
Ubald2	1.89E-08	0.33288	0.947	0.902	0.000528
Irf2bpl	1.94E-08	-0.26338	0.556	0.716	0.000543
Adgrl2	1.94E-08	-0.30204	0.635	0.825	0.000544
Arx	1.99E-08	-0.27618	0.714	0.817	0.000557
Cops6	2.24E-08	0.250411	0.91	0.863	0.000626
Lzts1	2.33E-08	0.336407	0.661	0.513	0.000653
Gadd45g	2.53E-08	0.578874	0.926	0.902	0.000709
Ubc	2.70E-08	0.309533	0.989	0.983	0.000755
Dll3	2.86E-08	0.333602	0.667	0.518	0.0008
Serf2	3.26E-08	0.278055	0.974	0.927	0.000913
Cox20	3.31E-08	0.280268	0.741	0.637	0.000927
Gtf2i	3.37E-08	-0.25335	0.915	0.925	0.000945
Npas3	3.57E-08	-0.26379	0.55	0.728	0.001001

Cntnap2	3.94E-08	-0.26499	0.741	0.842	0.001103
Arhgap5	4.53E-08	-0.25506	0.921	0.945	0.001269
Zfp516	4.55E-08	-0.25132	0.603	0.749	0.001275
Kdelr2	5.27E-08	-0.25852	0.688	0.828	0.001474
Cacng4	5.39E-08	-0.27437	0.571	0.696	0.00151
Rrm2	5.65E-08	0.465029	0.984	0.979	0.001581
Prim1	5.78E-08	0.294966	0.905	0.801	0.001619
Larp1	6.02E-08	-0.25029	0.661	0.804	0.001686
Cyfip2	8.60E-08	0.345492	0.746	0.63	0.002408
Ago2	8.82E-08	-0.25096	0.915	0.92	0.002471
Btbd17	8.84E-08	0.378424	0.788	0.713	0.002474
Cotl1	8.99E-08	0.280178	0.82	0.726	0.002518
Serpinh1	1.03E-07	-0.25351	0.635	0.794	0.002875
Ebf3	1.05E-07	0.275396	0.492	0.309	0.002953
Ndc80	1.14E-07	0.293497	0.757	0.671	0.003205
Ndufa12	1.19E-07	0.25666	0.91	0.836	0.003321
Rfc3	1.26E-07	0.293249	0.915	0.85	0.003539
Wdr12	1.38E-07	-0.25805	0.64	0.733	0.00387
AI506816	1.48E-07	-0.25993	0.677	0.818	0.004153
Ckb	1.49E-07	0.268876	1	1	0.004168
Bach2	1.70E-07	-0.25988	0.603	0.738	0.004751
Dbi	2.55E-07	0.254008	1	0.993	0.007147
Ccdc174	2.77E-07	0.260082	0.862	0.769	0.007765
Fbxo5	2.87E-07	0.321126	0.921	0.889	0.008041
Fam136a	3.04E-07	-0.25702	0.72	0.811	0.008523
Coro1c	3.43E-07	0.31548	0.915	0.88	0.009601
Neurog1	4.23E-07	0.43552	0.714	0.599	0.011839
Cbfa2t2	4.43E-07	0.255126	0.788	0.691	0.012404
Ahsa2	4.52E-07	0.253273	0.915	0.886	0.012659
Dach1	5.28E-07	-0.26154	0.825	0.935	0.014787
Neurod4	5.46E-07	0.355035	0.312	0.17	0.015299
Ncapd3	7.02E-07	0.32006	0.921	0.809	0.019669
Brd8	9.47E-07	0.259842	0.926	0.887	0.026511
Rbm15	1.13E-06	-0.28419	0.656	0.748	0.031762

Appendix 9. Table illustrating the DE genes from the HET vs. WT female comparison in the neuron cluster.

gene	p_val	avg_log2FC	pct.1	pct.2	p_val_adj
betagal	4.44E-72	-1.57255	0	0.506	1.24E-67
Atp5k	1.01E-37	0.548622	0.938	0.587	2.83E-33
Tmsb10	4.40E-37	0.359519	1	1	1.23E-32
Pcdh19	1.73E-35	0.67653	0.476	0.159	4.85E-31
Cox7c	2.63E-31	0.416766	1	0.914	7.36E-27
Hnrnpa0	4.19E-28	0.333927	0.995	0.992	1.17E-23
Atp5md	6.25E-25	0.387681	0.982	0.793	1.75E-20
Lars2	9.56E-24	0.357086	0.613	0.28	2.68E-19
Atp5e	1.18E-23	0.3949	0.991	0.82	3.31E-19
Bcl11a	1.36E-23	-0.40095	0.989	0.996	3.82E-19
Gm10076	2.15E-23	0.404564	0.815	0.443	6.03E-19
Dpysl2	1.30E-22	-0.46356	0.973	0.948	3.65E-18
Elob	4.47E-21	0.319197	1	0.926	1.25E-16
Atpif1	2.45E-20	0.295387	0.998	0.99	6.87E-16
Atp5j2	3.74E-20	0.351817	0.984	0.88	1.05E-15
Scg5	8.53E-20	0.393741	0.796	0.478	2.39E-15
Tomm7	1.01E-19	0.338585	0.902	0.592	2.84E-15
Ndn	1.29E-19	-0.51056	0.7	0.909	3.61E-15
Scrt1	1.95E-19	0.378456	0.92	0.697	5.46E-15
Mir6236	4.52E-19	0.449167	0.785	0.48	1.26E-14
Uqcr10	4.87E-19	0.344894	0.867	0.548	1.36E-14
Malat1	7.19E-19	0.280238	1	1	2.01E-14
Cox5b	1.37E-18	0.320991	0.97	0.839	3.84E-14
Ndufb1	3.12E-18	0.350999	0.913	0.599	8.75E-14
Anapc13	6.16E-18	0.328765	0.737	0.411	1.72E-13
Cops9	7.37E-18	0.373239	0.911	0.641	2.06E-13
Slc24a5	3.77E-17	-0.53962	0.666	0.707	1.05E-12
Son	1.04E-16	0.307796	0.989	0.89	2.91E-12
Gm1673	4.70E-16	0.325627	0.998	0.919	1.32E-11
Fam98b	4.72E-16	0.326122	0.767	0.453	1.32E-11
Tmem256	1.03E-15	0.293535	0.773	0.455	2.88E-11
Tmem258	4.26E-15	0.319923	0.824	0.5	1.19E-10
Ndufa3	4.34E-15	0.316749	0.76	0.455	1.21E-10
Ndufa1	4.64E-15	0.274816	0.771	0.445	1.30E-10
Cox6c	6.98E-15	0.273958	0.984	0.825	1.95E-10
Sem1	1.54E-14	0.290873	0.986	0.877	4.31E-10
Ndufc1	3.04E-14	0.29752	0.952	0.775	8.50E-10
Atp5mpl	9.07E-14	0.278519	0.934	0.746	2.54E-09
Polr2i	1.02E-13	0.276177	0.833	0.531	2.86E-09

Dync1i2	1.71E-13	0.252945	0.993	0.929	4.79E-09
Ndufa7	2.03E-13	0.267874	0.973	0.821	5.68E-09
Ubl5	2.47E-13	0.258274	0.979	0.874	6.92E-09
Metap2	5.84E-13	0.27095	0.922	0.727	1.64E-08
Ndufa6	2.27E-12	0.266948	0.881	0.624	6.34E-08
Smarcd1	2.39E-12	0.270944	0.924	0.704	6.69E-08
Gnb1	2.61E-12	-0.29116	0.984	0.949	7.32E-08
Eif5b	2.67E-12	0.274334	0.952	0.771	7.49E-08
Hspa8	2.71E-12	0.253074	0.995	0.929	7.58E-08
Uqcr11	7.42E-12	0.259127	0.92	0.701	2.08E-07
Cbx3	1.44E-11	-0.32844	0.984	0.925	4.03E-07
Romo1	1.49E-11	0.263681	0.879	0.58	4.16E-07
Nrep	7.35E-11	-0.26138	0.998	0.988	2.06E-06
Nfib	7.93E-11	-0.64263	0.643	0.727	2.22E-06
Ddah2	1.16E-10	0.27343	0.989	0.952	3.25E-06
Myef2	1.35E-10	-0.36408	0.826	0.755	3.77E-06
Gm28050	2.22E-10	0.263032	0.584	0.366	6.23E-06
Tle4	1.19E-09	0.255357	0.632	0.406	3.34E-05
Aplp1	2.04E-09	0.255378	0.934	0.752	5.72E-05
Sec62	5.69E-09	0.257918	0.902	0.698	0.000159
Bcl11b	6.70E-09	-0.35031	0.856	0.863	0.000188
Kdm6b	4.00E-08	-0.30116	0.881	0.814	0.00112
Ebf2	4.68E-08	-0.54161	0.579	0.646	0.001311
Tcf4	5.85E-08	-0.35296	0.899	0.899	0.001639
Nav1	7.67E-08	-0.27826	0.979	0.977	0.002148
Nfia	8.01E-08	-0.68907	0.492	0.572	0.002242
Cnr1	8.49E-08	-0.36687	0.824	0.859	0.002377
Meis2	8.83E-08	0.438873	0.76	0.594	0.002474
Nnat	1.23E-07	0.321927	1	0.998	0.003435
Zfp862-ps	3.78E-07	-0.32867	0.213	0.307	0.010586
Zic1	7.97E-07	0.47957	0.302	0.189	0.022304

Appendix 10. Table illustrating the DE genes from the KO vs. WT male comparison in the RGC cluster.

gene	p_val	avg_log2FC	pct.1	pct.2	p_val_adj
betagal	1.27E-58	-2.74482	0	0.894	3.55E-54
Pcdh19	3.56E-43	1.032632	0.905	0.383	9.96E-39
Mt2	7.15E-31	-1.01977	0.246	0.798	2.00E-26
Gm10036	3.86E-29	0.425769	0.536	0.011	1.08E-24
Hbb-y	3.11E-25	0.562238	0.95	0.729	8.71E-21
Actb	4.00E-22	-0.44923	1	1	1.12E-17
Zic3	3.35E-16	-0.44379	0.101	0.473	9.39E-12
Zic1	6.77E-16	-0.96755	0.302	0.681	1.89E-11
Stmn1	1.95E-15	-0.31135	0.994	1	5.46E-11
Fabp7	2.48E-15	-0.93119	0.866	0.973	6.93E-11
Pttg1	1.51E-14	-0.58817	0.408	0.729	4.24E-10
Cbx3	1.98E-14	-0.33624	1	1	5.54E-10
Mt1	2.98E-14	-0.67936	0.592	0.862	8.35E-10
Mest	3.36E-14	-0.99297	0.844	0.904	9.40E-10
Hba-x	3.35E-12	0.319918	0.771	0.457	9.37E-08
Ulk4	3.72E-12	-0.3776	0.291	0.66	1.04E-07
Hmgn2	6.90E-12	-0.26113	0.994	1	1.93E-07
Efna5	1.39E-11	-0.25522	0.151	0.489	3.89E-07
Gm47283	5.62E-11	-0.46887	0.95	1	1.57E-06
Pdcd4	9.08E-11	0.359924	0.961	0.851	2.54E-06
Dpysl2	1.37E-10	-0.38722	0.888	0.979	3.83E-06
Cxxc4	3.19E-10	-0.28032	0.313	0.676	8.93E-06
Sfr1	3.61E-10	0.270546	0.989	1	1.01E-05
Zic4	4.63E-10	-0.53622	0.106	0.383	1.30E-05
Ncl	9.34E-10	0.295319	1	1	2.61E-05
Ubb	1.37E-09	-0.38009	0.989	0.995	3.83E-05
Ccnd2	1.43E-09	0.436598	0.994	0.995	4.01E-05
Pantr1	2.05E-09	-0.52573	0.922	0.984	5.74E-05
Nnat	1.20E-08	-0.53215	0.978	1	0.000337
Nr2f1	1.53E-08	0.366746	0.821	0.495	0.000429
Metrn	1.81E-08	-0.3798	0.637	0.894	0.000508
Dusp6	2.10E-08	-0.30524	0.156	0.399	0.000589
Ddx46	2.24E-08	0.281912	0.989	0.989	0.000626
Ldhb	2.47E-08	-0.31819	0.95	0.984	0.000691
Lhx2	3.84E-08	0.364977	1	1	0.001074
Cacybp	6.99E-08	0.252409	0.994	1	0.001957
Mpp6	7.67E-08	0.260596	0.983	0.995	0.002148
Hspa4	1.47E-07	0.292959	1	0.989	0.00412
Psma5	2.30E-07	-0.25858	0.955	0.989	0.006435

Ddx24	3.75E-07	0.256391	0.994	0.984	0.010493
Ptprz1	5.18E-07	-0.48572	0.922	0.952	0.014509
Cdk2ap1	5.44E-07	-0.2541	0.989	1	0.015242
Rpgrip1	5.58E-07	-0.35334	0.62	0.824	0.015615
Fabp5	6.25E-07	0.253067	0.994	1	0.017493
Dhx9	8.97E-07	0.260921	0.994	0.995	0.025111
Nom1	1.43E-06	0.261832	0.832	0.824	0.040117
Srsf4	1.60E-06	0.286827	0.95	0.931	0.04479
Mak16	1.70E-06	0.300741	0.939	0.931	0.047471

Appendix 11. Table illustrating the DE genes from the KO vs. WT male comparison in the neuron cluster

	p_val	avg_log2FC	pct.1	pct.2	p_val_adj
betagal	6.80E-11	-1.63954	0	0.819	1.90E-06
Ubb	1.71E-09	-0.82705	1	1	4.80E-05
Gm10036	3.17E-08	0.308013	0.308	0	0.000887
Meg3	1.07E-07	-2.07539	0.5	0.947	0.002982
Nnat	1.56E-07	-1.09936	1	1	0.004357
Hsp90ab1	1.73E-07	0.412295	1	1	0.004832
Nfib	3.91E-07	1.412757	0.923	0.511	0.010957
Nhlh1	4.12E-07	1.196771	0.846	0.574	0.011546
Ino80d	4.59E-07	-0.61545	0.115	0.745	0.012843
Tmsb4x	7.81E-07	0.513512	1	1	0.021872
Stmn1	7.81E-07	-0.61454	1	1	0.021872
Cecr2	1.48E-06	0.588179	0.885	0.596	0.041423
Neurod6	1.64E-06	1.114776	0.962	0.511	0.045967

Appendix 12. Table illustrating the enriched pathways from the Shiny Go analysis between HET and WT females within the RGC cluster

Enrichment FDR	nGenes	Pathway Genes	Fold Enrichment	Pathway
2.93E-06	9	118	18.3428	Oxidative phosphorylation
2.93E-06	7	58	29.0252	NADH dehydrogenase complex assembly
2.93E-06	7	58	29.0252	Mitochondrial respiratory chain complex I assembly
3.13E-06	9	138	15.68442	Electron transport chain
4.68E-06	11	261	10.13578	ATP metabolic process
7.42E-06	10	215	11.18579	Cellular respiration
1.02E-05	13	446	7.00993	Generation of precursor metabolites and energy
1.02E-05	9	171	12.65761	Aerobic respiration
3.17E-05	7	95	17.72065	Mitochondrial respiratory chain complex assembly
3.26E-05	18	1014	4.269133	Central nervous system development
9.03E-05	10	300	8.016484	Energy derivation by oxidation of organic compounds
0.000407	13	640	4.885045	Translation
0.000527	13	660	4.737013	Peptide biosynthetic process
0.000984	14	818	4.116043	Peptide metabolic process
0.000984	11	500	5.290879	Mitochondrion organization
0.002198	19	1552	2.944198	Organonitrogen compound biosynthetic process
0.002275	13	776	4.028903	Amide biosynthetic process
0.003563	5	87	13.82152	Respiratory electron transport chain
0.004157	6	150	9.61978	Proton transmembrane transport
0.004406	5	93	12.92981	Mitochondrial transmembrane transport
0.004406	3	17	42.44021	Mitochondrial ATP synthesis coupled proton transport
0.005858	12	758	3.807301	Brain development
0.005858	7	244	6.899433	Glial cell differentiation
0.005858	4	52	18.49958	Aerobic electron transport chain
0.005858	9	424	5.104836	Forebrain development
0.00594	4	55	17.49051	Forebrain generation of neurons
0.00594	8	335	5.743152	Gliogenesis
0.00594	8	337	5.709068	Renal system development
0.008055	4	63	15.26949	Glial cell migration
0.008055	3	24	30.06181	Energy coupled proton transport, down electrochemical gradient
0.008055	3	24	30.06181	ATP synthesis coupled proton transport
0.008055	6	186	7.757887	Central nervous system neuron differentiation
0.008055	4	63	15.26949	Mitochondrial ATP synthesis coupled electron transport
0.008055	14	1060	3.176343	Cellular amide metabolic process
0.008055	12	805	3.585011	Head development
0.009386	4	66	14.57542	ATP synthesis coupled electron transport

0.010171	8	381	5.049753	Urogenital system development
0.010171	3	27	26.72161	Adrenal gland development
0.011898	13	992	3.151642	Cellular protein-containing complex assembly
0.012229	6	209	6.904148	Mitochondrial transport
0.012631	2	6	80.16484	Forebrain radial glial cell differentiation
0.014248	6	217	6.649618	Regulation of neuron differentiation
0.01425	4	77	12.49322	Nephron epithelium morphogenesis
0.014337	4	78	12.33305	Mitochondrial translation
0.014337	3	32	22.54636	Regulation of astrocyte differentiation
0.014696	4	79	12.17694	Nephron morphogenesis
0.015054	3	33	21.86314	Cell proliferation in forebrain
0.015497	4	81	11.87627	Diencephalon development
0.015675	8	422	4.559138	Epithelial cell proliferation
0.015725	5	148	8.124814	Kidney epithelium development
0.016008	4	83	11.5901	Astrocyte differentiation
0.016225	3	35	20.61381	Proximal/distal pattern formation
0.016314	7	326	5.163992	Kidney development
0.016684	12	925	3.119929	Negative regulation of gene expression
0.016684	3	36	20.04121	Embryonic digestive tract development
0.016836	2	8	60.12363	Regulation of epithelial cell proliferation involved in lung morphogenesis
0.016836	4	87	11.05722	Central nervous system neuron development
0.016836	2	8	60.12363	Fat pad development
0.019623	4	91	10.57119	Metanephros development
0.022696	4	95	10.12608	Kidney morphogenesis
0.023574	5	169	7.115222	Gene silencing
0.023881	7	357	4.715579	Regulation of epithelial cell proliferation
0.024395	2	10	48.0989	Negative regulation of single stranded viral RNA replication via double stranded DNA intermediate
0.027841	2	11	43.72627	Mitochondrial electron transport, ubiquinol to cytochrome c
0.027841	8	483	3.983346	Posttranscriptional regulation of gene expression
0.027841	3	46	15.68442	Forebrain neuron differentiation
0.027841	2	11	43.72627	Epithelial cell proliferation involved in lung morphogenesis
0.027841	3	46	15.68442	Regulation of stem cell proliferation
0.029221	3	47	15.35071	Negative regulation of DNA binding
0.029781	4	107	8.990449	Mitochondrial gene expression
0.029781	3	48	15.03091	Negative regulation of gliogenesis
0.029781	16	1631	2.359235	Protein-containing complex subunit organization
0.029892	2	12	40.08242	Regulation of single stranded viral RNA replication via double stranded DNA intermediate
0.029892	2	12	40.08242	Radial glial cell differentiation
0.029892	2	12	40.08242	Cellular response to X-ray
0.030094	5	188	6.39613	Appendage development

0.030094	5	188	6.39613	Limb development
0.030294	6	282	5.116904	Regulation of mRNA metabolic process
0.030831	4	111	8.666469	Viral genome replication
0.03118	6	285	5.063042	Telencephalon development
0.032622	2	13	36.99915	Single stranded viral RNA replication via double stranded DNA intermediate
0.032784	4	114	8.438404	Nephron epithelium development
0.034536	4	116	8.292914	Regulation of gliogenesis
0.036398	2	14	34.35636	Positive regulation of autophagy of mitochondrion in response to mitochondrial depolarization
0.036398	7	406	4.146457	Wnt signaling pathway
0.036398	6	298	4.842171	Regulation of Wnt signaling pathway
0.036407	7	408	4.126131	Cell-cell signaling by wnt
0.036407	3	55	13.11788	Intracellular protein transmembrane transport
0.038455	2	15	32.06593	Regulation of autophagy of mitochondrion in response to mitochondrial depolarization
0.038455	3	57	12.65761	ATP biosynthetic process
0.038455	8	534	3.602914	Morphogenesis of an epithelium
0.038455	7	415	4.056534	Regulation of neurogenesis
0.043538	16	1735	2.217817	Positive regulation of RNA metabolic process
0.043538	3	60	12.02473	Protein transmembrane transport
0.046014	14	1421	2.369404	Negative regulation of RNA metabolic process
0.046886	9	686	3.155176	Negative regulation of cellular component organization
0.046986	2	17	28.29347	Negative regulation of astrocyte differentiation
0.047148	5	220	5.465784	Morphogenesis of a branching structure
0.047724	15	1595	2.261704	Negative regulation of cellular biosynthetic process
0.04845	16	1765	2.18012	Neurogenesis
0.048748	3	64	11.27318	Lung morphogenesis

Appendix 13. Table illustrating the enriched pathways from the Shiny Go analysis between KO and WT males within the RGC cluster

Enrichment FDR	nGenes	Pathway Genes	Fold Enrichment	Pathway
7.62E-05	13	1014	6.376748	Central nervous system development
0.000262	11	805	6.796584	Head development
0.000888	10	758	6.561825	Brain development
0.006461	11	1206	4.536692	Neuron development
0.006554	8	670	5.938942	Plasma membrane bounded cell projection morphogenesis
0.006554	2	6	165.7955	Stress response to copper ion
0.006554	7	445	7.824055	Axonogenesis
0.006554	2	6	165.7955	Detoxification of copper ion
0.006554	10	1064	4.674684	Neuron projection development
0.006554	8	656	6.065687	Neuron projection morphogenesis
0.006554	8	675	5.894949	Cell projection morphogenesis
0.006554	7	488	7.13464	Axon development
0.007188	8	704	5.652118	Cell part morphogenesis
0.011223	11	1470	3.721939	Neuron differentiation
0.011752	12	1765	3.381664	Neurogenesis
0.011752	2	9	110.5303	Detoxification of inorganic compound
0.01243	8	792	5.024105	Cellular component morphogenesis
0.012516	11	1531	3.573645	Plasma membrane bounded cell projection organization
0.012918	7	603	5.773971	Cell morphogenesis involved in neuron differentiation
0.013644	2	11	90.43388	Stress response to metal ion
0.013644	11	1570	3.484873	Cell projection organization
0.017426	11	1627	3.362784	Generation of neurons
0.020533	2	14	71.05519	Oxygen transport
0.022676	2	15	66.31818	Cellular response to zinc ion
0.025851	4	182	10.93157	Neural precursor cell proliferation
0.030628	6	524	5.695264	Regulation of neuron projection development
0.035826	7	762	4.569166	Cell morphogenesis involved in differentiation
0.035826	2	20	49.73864	Cellular response to copper ion
0.040918	2	22	45.21694	Gas transport
0.047891	8	1056	3.768079	Cell morphogenesis