Investigating the molecular basis of brain malformations causing severe infantile epilepsies

Matthew Coleman

ORCID: 0000-0001-8627-3322

This thesis is submitted in fulfilment of the degree of

Doctor of Philosophy (PhD)

August 2023

DEPARTMENT OF PAEDIATRICS

FACULTY OF MEDICINE, DENTISTRY AND HEALTH SCIENCE (MDHS)

THE UNIVERSITY OF MELBOURNE

BRUCE LEFROY CENTRE FOR NEUROGENETIC HEALTH (BLC)

MURDOCH CHILDREN'S RESEARCH INSTITUTE (MCRI)

Investigating the molecular basis of brain malformations causing severe infantile epilepsies

Abstract

Malformations of cortical development (MCD) encompass a wide spectrum of conditions that result from abnormal brain development, including polymicrogyria, periventricular nodular heterotopia and focal cortical dysplasia. MCD affect ~1 in 2,500 children and can result in epilepsy, developmental delay and cerebral palsy. Infantile epileptic spasms syndrome (IESS) is one of the most common forms of severe epilepsy of infancy and MCD are observed in ~50% of children with IESS. It is thought that the majority of cases of IESS and MCD have an underlying genetic cause. Although multiple molecular pathways have been identified, ~60% of cases still lack a genetic diagnosis after standard clinical diagnostic testing. Genetic diagnoses are important for prognosis, diagnostic testing, genetic counselling and precision medicine. Recent advances in genetic technologies allow for rapid, in-depth interrogation of the molecular mechanisms underlying MCD formation.

The hypotheses of this thesis are that the application of high-resolution genomic technologies will facilitate a better understanding of the genetic causes of IESS with MCD and that functional characterisation of genes of interest will assist to identify crucial molecular pathways in the normal development and function of the developing brain.

The primary research aims of this project are: to investigate the genetic landscape of IESS due to MCD by utilising brain tissue studies; to develop a deeper understanding of the molecular basis and phenotypic classifiers involved in the pathogenesis of brain somatic *SLC35A2* variation and its relationship to mild malformation of oligodendroglial hyperplasia in epilepsy (MOGHE); and to identify molecular mechanisms and biomarkers of MOGHE.

The genetic basis of IESS due to MCD was identified in 47/59 (80%) individuals. Germline pathogenic variants were identified in 27/59 (46%) individuals in *TSC2* (n=19), *DEPDC5* (2), *CDKL5* (2), *COL4A1* and 6p25.2 deletion (1), *NPRL3* (1), *FGFR1* (1) and *TSC1* (1). Pathogenic brain somatic variants were identified in 21/59 (36%) cases, in *SLC35A2* (n=9), *PIK3CA* (3), *MTOR* (2), *TSC2* (2), *AKT3* (2), *OFD1* (1), *TSC1* (1) and *DEPDC5* (1). This included one individual with a two-hit *DEPDC5* diagnosis, with a germline *DEPDC5* variant and a somatic *DEPDC5* variant. Multidisciplinary tools and integrated diagnosis review identified all nine individuals with *SLC35A2* variants as having mild malformation of cortical development with oligodendroglial hypoplasia in epilepsy (MOGHE). Multi-omic analysis of MOGHE tissue using single nuclei RNA-sequencing (snRNA-seq) and Tandem Mass Tag 16pro liquid chromatography–mass spectrometry (LC-MS/MS) showed that neurogenesis, neuron projection development, axo-dendritic development and neuron differentiation are dysregulated in MOGHE. snRNA-seq analysis identified oligodendrocyte progenitor cells (OPCs) and dividing oligodendrocyte progenitor cells (cOPCs) as the primary cells exhibiting aberrant expression profiles in lesional MOGHE tissues. ACTB, GNG7, MAP4, NEFL, SEMA3E and SOD1 were identified as potential biomarkers in MOGHE and may serve as disease biomarkers or druggable targets in the future.

The findings of this thesis contribute to the knowledge of the genetic landscape of paediatric brain malformations and severe epilepsies of infancy and broaden the understanding of cortical malformation development and epileptogenesis.

Declaration

This is to certify that:

I. The thesis comprises only my original work towards the Doctor of Philosophy except where indicated in the preface;

II. Due acknowledgement has been made in the text to all other material used; and

III. This thesis is fewer than 100,000 words in length, exclusive of tables, maps, bibliographies, and appendices.

_____

25 August 2023

Preface

This thesis contains work which will provide the basis for two original research papers, which are currently in preparation for submission to the peer-reviewed journals Neurology and Genetics in Medicine. The doctoral candidate is the first author on these two papers and contributed more than 50% of the work in the manuscript. This work was all carried out within the period of candidature and has not been submitted for other qualifications.

In addition, one first author research paper is currently under review in the journal of Neuropathology and Applied Neurobiology and nine co-authored publications are included as appendices to this thesis that were generated from the work carried out during the period of candidature but relate to studies that were peripheral to the aims of the thesis or, to which the doctoral candidate contributed less than 50%. The publications are listed below and the first author manuscript under review is included in the Appendix (Supplementary Section 8.5).

i. Kaspi A, Hildebrand MS, Jackson VE, Braden R, Van Reyk O, Howell T, Debono S, Lauretta M, Morison L, **Coleman MJ**, Webster R, Coman D, Goel H, Wallis M, Dabscheck G, Downie L, Baker EK, Parry-Fielder B, Ballard K, Harrold E, Ziegenfusz S, Bennett MF, Robertson E, Wang L, Boys A, Fisher SE, Amor DJ, Scheffer IE, Bahlo M, Morgan AT. 'Genetic aetiologies for childhood speech disorder: novel pathways co-expressed during brain development', *Mol. Psych*., 2022 Sept 18, doi: 10.1038/s41380-022-01764-8

ii. Feenstra B, Skotte L, Fadista J, Grauholm G, Appadurai V, **Coleman M**, Damiano JA, Burgess R, Schneider A, Hansen TF, Banasik K, Liu X, Pasternak B, Pers TH, Dreier JW, Christensen J, Hougaard DM, Buil A, Geller F, Hviid A, Ullum H, Mullen SA, Hildebrand MS, Scheffer IE, Berkovic SF, Werge T, Melbye M. 'Genome-wide association study of febrile seizures implicates fever response and neuronal excitability genes', *Brain*, 2022 Jan 12, doi: 10.1093/brain/awab260

iii. de Valles-Ibáñez G, Hildebrand M, Bahlo M, King C, **Coleman M**, Green T, Goldsmith J, Davis S, Gill D, Mandelstam S, Scheffer IE, Sadleir L. 'Infantile-onset Myoclonic Developmental and Epileptic Encephalopathy: a new RARS2 phenotype', *Epilepsia Open*, 2021 October 30, doi: 10.1002/epi4.12553

iv. Heron SE, Regan BM, Harris RV, Gardner AE, **Coleman MJ**, Bennett MF, Grinton BE, Helbig KL, Sperling MR, Haut S, Geller EB, Widdess-Walsh P, Pelekanos JT, Bahlo M, Petrovski S, Heinzen EL, Hildebrand MS, Corbett MA, Scheffer IE, Gécz J, Berkovic, SF. 'Association of *SLC32A1* missense variants with genetic epilepsy with febrile seizures plus', *Neurology*, 2021 Mar 23, doi: 10.1212/WNL.0000000000011855

v. Myers KA, Bennett MF, Hildebrand MS, **Coleman MJ**, Zhou G, Hollingsworth G, Cairns A, Riney K, Berkovic SF, Bahlo M, Scheffer IE, 'Transcriptome Analysis in Ring Chromosome 20 Reveals Broad Differential Gene Expression', *Epilepsia*, 2020 Nov 18, doi: 10.1111/epi.16766

vi. Palencia-Campos A, Aoto PC, Machal EMF, Rivera-Barahona A, Soto-Bielicka P, Bertinetti D, Baker B, Vu L, Piceci-Sparascio F, Torrente I, Boudin E, Peeters S, Van Hul W, Huber C, Bonneau D, Hildebrand MS, **Coleman M**, Bahlo M, Bennett MF, Schneider AL, Scheffer IE, Kibæk M, Kristiansen BS, Issa MY, Mehrez MI, Ismail S, Tenorio J, Li G, Skålhegg BS, Otaify GA, Temtamy S, Aglan M, Jønch AE, De Luca A, Mortier G, Cormier-Daire V, Ziegler A, Wallis M, Lapunzina P, Herberg FW, Taylor ST, Ruiz-Perez VL, 'Germline and Mosaic Variants in PRKACA and PRKACB Cause a Multiple Congenital Malformation Syndrome', *Am J Hum Genet*, 2020 Oct 14. doi: 10.1016/j.ajhg.2020.09.005

vii. Hildebrand M, Jackson V, Scerri T, Van Reyk O, **Coleman M**, Braden R, Turner S, Rigbye K, Boys A, Barton S, Webster R, Fahey M, Saunders K, Parry-Fielder B, Paxton G, Hayman M, Coman D, Goel H, Baxter A, Ma A, Davis N, Reilly S, Delatycki M, Liégeois F, Connelly A, Gecz J, Fisher S, Amor D, Scheffer I, Bahlo M, Morgan A, 'Severe childhood speech disorder: gene discovery highlights transcriptional dysregulation', *Neurology*, 2020 May 19, doi: 10.1212/WNL.0000000000009441

viii. Bennett MF, Oliver KL, Regan BM, Sikta N, Rafehi H, Bellows ST, Schneider AL, Crompton DE, Dolzhenko E, Eberle MA, **Coleman M**, Hildebrand MS, Corbett MA, Kroes T, Gecz J, Scheffer IE, Berkovic SF, Bahlo M, 'Familial adult myoclonic epilepsy type 1 SAMD12 TTTCA repeat expansion arose 17,000 years ago and is present in Sri Lankan and Indian families', *Eur J Hum Gen*, 2020 Mar 16. doi: 10.1038/s41431-020-0606-z

ix. Sadleir L, de Valles-Ibanez G, King C, **Coleman M**, Mossman S, Paterson S, Nguyen J, Berkovic S, Mullen S, Bahlo M, Hildebrand M, Mefford H, Scheffer I, 'Inherited *RORB* pathogenic variants: overlap of photosensitive genetic generalized and occipital lobe epilepsy', *Epilepsia*, 2020 Mar 12. doi: 10.1111/epi.16475

Chapter 3: The genetic landscape of Infantile Epileptic Spasms Syndrome with brain malformations

Contribution to chapter: I was involved in all aspects of this chapter. I was involved with study design with Dr Katherine B Howell, Prof Paul J Lockhart, Prof Rick J Leventer and Dr Sarah E M Stephenson. I performed the wet lab experimental work and aspects of bioinformatics work.

Bioinformatic processing not completed by myself was completed in collaboration with Dr Min Wang, Dr Cas Simons, Dr Sarah E M Stephenson and Dr Wei Shern Lee. Next-generation sequencing and library preparation was performed fee-for-service by the Translational Genomics Unit at the Victorian Clinical Genetic Services (Melbourne, Australia). Clinical data were provided by Dr Katherine B Howell, Dr Rick J Leventer and Dr A Simon Harvey.

- Unpublished material not yet submitted for publication.

Chapter 4: The relationship between somatic *SLC35A2* variation and MOGHE

Contribution to chapter: I was involved in all aspects of this chapter. I conceptualised and designed the study with Prof Paul J Lockhart, Prof Richard J Leventer, Dr Katherine B Howell and Dr Sarah E M Stephenson. I performed all the wet lab experimental work and dry lab bioinformatics analysis.

Clinical review was completed by Dr Colleen D'Arcy (neuropathology) and Dr Cristina Mignone (neuroradiology).

- Unpublished material not yet submitted for publication.

Chapter 5: Single cell RNA-sequencing and proteomics analyses reveal insights into MOGHE pathogenesis

Contribution to chapter: I was involved in all aspects of this chapter. I conceptualised and designed the study with Dr Sarah E M Stephenson, Dr Saskia Freytag, Prof Paul J Lockhart, Prof Rick J Leventer and Dr Katherine B Howell. I performed a significant portion of the experimental work and data analysis.

Nuclei isolation, GEM generation and barcoding were performed by me. Next-generation sequencing and library preparation was performed fee-for-service at the Translational Genomics Unit at the Victorian Clinical Genetic Services (Melbourne, Australia). Bioinformatic analysis of snRNA-seq data was performed by Dr Saskia Freytag.

Protein isolation and preparation was performed by me. Protein labelling, digestion and liquid chromatography and tandem mass spectrometry (LC-MS/MS) analysis was performed fee-for-service at the Monash Proteomics and Metabolomics Facility (MPMF) (Melbourne, Australia). Bioinformatic processing of proteomic data was performed by Dr Joel Steele.

- Unpublished material not yet submitted for publication.

I performed all downstream analyses and statistical analyses in all chapters of this thesis. No third-party editorial assistance was provided in the preparation of this thesis.

Acknowledgements

I have been fortunate to have met, worked with and befriended so many wonderful people throughout my PhD journey. I did not and could not have done this alone.

First and foremost, I would like to thank my supervisors, Prof Paul Lockhart, Prof Richard Leventer, Dr Sarah Stephenson and Dr Katherine Howell. Thank you, Paul, for your endless generosity and advice during my PhD. You took a chance on me, and I would not have been able to do a PhD without your support. I am so glad that I chose to undertake my PhD with you. Your mentorship and guidance have made me a better researcher and a better person. Thank you to Sarah, without whom I would not have had the confidence to even pursue a PhD. You have been my friend, ally and role model. I admire your passion for research and am very lucky to have worked and trained under your supervision. To Rick, thank you for being so generous with your expertise and advice. Your continued encouragement and support has been greatly appreciated. To Katherine, thank you for your kindness, enthusiasm and insight. You are one of the most intelligent people I have ever met and it has been an absolute pleasure getting to know you. I feel so fortunate to have had you all as my supervisors. Thank you for your guidance, support, feedback and encouragement.

I would also like to express my sincere gratitude to Dr. Ann Frazier, my PhD advisory committee chair, and Dr Peter Houweling, my committee member, for their extremely valuable contributions and guidance throughout the years.

I am grateful to all of the collaborators who have contributed to this thesis. I would like to thank Dr A Simon Harvey, Dr Colleen D'Arcy and Dr Cristina Mignone, for their guidance on the clinical aspects of this research. Thank you also to the large team who contributed to the success of this project, in particular Penny Snell, Kate Pope, Greta Gillies, Dr Min Wang and Dr Cas Simons. Special thanks to Dr Saskia Freytag and Dr Joel Steele for their insights and contributions to the bioinformatic analyses in this project.

I would like to thank all past and present members of the Bruce Lefroy Centre and my colleagues at the Murdoch Children's Research Institute for their support and company throughout the years. You are a wonderful bunch of people in and out of the lab. Particular thanks to Dr Wei Shern Lee, whose kindness, patience, and generosity of time has improved my research and this project immeasurably. To the participants and their families who generously donated their time and their tissues for this study.

This project was generously supported by the research funds of Prof Paul J Lockhart at the Murdoch Children's Research Institute. My studies were supported by an Australian Government Research Training Program fee offset scholarship. Overseas and interstate travelling activities were generously supported by: Harold Mitchell Foundation Travelling Fellowship provided by the Harold Mitchell Foundation; Henry and Rachel Ackman Travelling Scholarship awarded by the Henry and Rachel Ackman Bequest; the Lorna Stirling Fund; The International League Against Epilepsy (ILAE); and the Student Conference Support Award provided by the Murdoch Children's Research Institute.

Thank you also to Troye Sivan and Taylor Swift, whose music kept me company through many long hours in the lab and over the course of my studies.

I would like to thank my family and friends for their unconditional love and support. Heartfelt thanks to Ciara, Tom, Alice, David, Cam, Conor, Hollie, Bec, Ken and to the whole QueersInScience committee. To Zuko, the best chief parkland investigator and four-legged cuddle buddy. To my wonderful fiancé, Liam, thank you for your unwavering support and for brightening every one of my days. You are the reason I was able to accomplish any of this. I am so lucky and I cannot wait to marry you.

To anyone else not outright mentioned, I would like to thank you. If you are reading this, thank you and I hope this research helps you learn something new.

And finally, to my parents. You taught me to be ever curious, studious, and creative in everything I do. Thank you for your endless support and for always inspiring me to do the best that I can. I feel so fortunate to you as my parents. You have raised me to be the person that I am today, and I dedicate this thesis to you.

Abbreviations

Abbreviation	Definition
ACMG	American College of Medical Genetics and Genomics
ALS	Amyotrophic lateral sclerosis
ASM	Anti-seizure medication
BAM	Binary alignment map
BCs	Balloon cells
BSA	Bovine serum albumin
BWA	Burrow-Wheeler Aligner
CADD	Combined Annotation Dependent Depletion
CATs	Clinically accessible tissues
CDG	Congenital disorder of glycosylation
cDNA	Complementary deoxyribonucleic acid
CMA	Chromosomal microarray
CNS	Central nervous system
CNV	Copy number variant
cOPCs	Dividing oligodendrocyte progenitor cells
CREv2	Clinical Research Exome version 2
CSF	Cerebrospinal fluid
CxGO	C to Go translator
DAB	3,3'Diaminobenzidine
DAPI	4'-6-diamidino-2-phenylindole
ddPCR	Droplet digital polymerase chain reaction
DEG	Differentially expressed gene
DNs	Dysmorphic neurons
DNA	Deoxyribonucleic acid
DNET	Dysembryoplastic neuroepithelial tumour
DP	Differential protein
ECoG	Electrocorticography
EDTA	Ethylenediaminetetraacetic acid
EEG	Electroencephalogram
ES	Exome sequencing
FAM	Fluorescein amidite
FANS	Fluorescence activated nuclei sorting
FCD	Focal cortical dysplasia
FDR	False discovery ratio
FFPE	Formalin-fixed, paraffin-embedded
FLAIR	fluid-attenuated inversion recovery
FMCD	Focal malformation of cortical development
GATK	Genome Analysis Toolkit

gDNA	Genomic deoxyribonucleic acid
gnomAD	Genome Aggregation Database
GO	Gene ontology
HME	Hemimegalencephaly
IDR	Intrinsically disordered region
IESS	Infantile epileptic spasms syndrome
ILAE	International League Against Epilepsy
INDEL	Insertion and deletion mutation
iPSCs	Induced pluripotent stem cells
IS	Infantile spasms
kb	Kilobase pair
KEGG	Kyoto Encyclopedia of Genes and Genomes
LC-MS/MS	Liquid chromatography–tandem mass spectrometry
LITT	Laser interstitial thermal therapy
LOD	Limit of detection
LoF	Loss of function
LoH	Loss of heterozygosity
Mbp	Megabase pair
MCD	Malformation of cortical development
MCRI	Murdoch Children's Research Institute
MOGHE	Mild malformation of cortical development with oligodendroglial hyperplasia in epilepsy
MRI	Magnetic resonance imaging
mRNA	Messenger ribonucleic acid
MS	Multiple sclerosis
mTOR	Mammalian target of rapamycin
MTR	Missense Tolerance Ratio
NeuN	Neuronal nuclei
NMD	Nonsense mediated decay
NP-40	Nonidet P-40
NST	Nucleotide sugar transporter
Olig2	Oligodendrocyte transcription factor 2
OMIM	Online Mendelian Inheritance in Man database
OPCs	Oligodendrocyte progenitor cells
PBS	phosphate buffered saline
PCA	Principle component analysis
PCR	Polymerase chain reaction
PET	Position emission tomography
PMG	Polymicrogyria
PolyPhen	Polymorphism Phenotyping
pS6	Phosphorylated ribosomal protein S6
RCH	Royal Children's Hospital

RNA	Ribonucleic acid
RNase	Ribonuclease
RNA-seq	RNA sequencing
RVIS	Residual Variation Intolerance Score
scRNA-seq	Single cell RNA sequencing
SDS	Sodium dodecyl sulphate
SIFT	Sorting Intolerant From Tolerant
SMI	Sternberger Monoclonal
snRNA-seq	Single nuclei RNA sequencing
SNV	Single nucleotide variant
SSH	Sonic hedgehog
SVZ	Subventricular zone
TGU	Translational Genomics Unit
TMEM	Transmembrane domain
Tris	Tris(hydroxymethyl)aminomethane
TSC	Tuberous sclerosis complex
TWIST	Twist exome 2.0
UDP-galactose	Uridine diphosphate galactose
UMAP	Uniform Manifold Approximation and Projection
UMI	Unique molecular identifier
VAF	Variant allele frequency
VCGS	Victorian Clinical Genetics Services
VIC	2'-chloro-7'phenyl-1,4-dichloro-6-carboxy-fluorescein
VUS	Variant of unknown significance
VZ	Ventricular zone
WGS	Whole genome sequencing
XLD	X-linked dominant
XLR	X-linked recessive

Table of Contents

Table of Tables

Table of Figures

1 Chapter 1: Introduction

1.1 Epilepsy

Epilepsy is one of the most common neurological disorders, affecting an estimated 65 million people globally [1]. It is characterised by the ongoing susceptibility to have recurrent, epileptic seizures. Approximately 1 in 10 people will have a seizure at some point during their lifetime but individuals are not classified as having epilepsy until a second, unprovoked seizure is experienced [2]. It is estimated that 3-4% of the population in industrialised countries will develop epilepsy during their lifetime [3]. Epilepsy may result from acquired causes such as stroke, brain tumour, head injury or infection, may be the result of genetic or metabolic contributors, or the exact aetiology may remain unknown [4].

1.1.1 Classification

In 2017, the International League Against Epilepsy (ILAE) established a revised, three-level classification of epilepsy (Figure 1.1) [5]. This classification system replaced a collection of traditional terms and was implemented to reflect a greater understanding of epileptogenesis, improve epilepsy diagnoses and highlight the need to consider aetiology in diagnosis, as aetiology often directly influences treatment options and efficacy.

Level one serves to define seizure types as focal (i.e. seizure origin limited to one region or hemisphere of the brain), generalised (i.e. affecting both hemispheres simultaneously and through the brain networks), or seizures with unknown onset (i.e. cases where evidence is insufficient to classify seizure types as either focal or generalised).

The second level of classification addresses the diagnosis of the epilepsy type, for which there are four classes. These classes are focal epilepsy, generalised epilepsy, focal and generalised epilepsy, and unknown epilepsy. These diagnoses reflect the types of seizures an individual presents with and acknowledges that individuals may

present with both focal and generalised seizures by introducing the third category of combined focal and generalised epilepsy.

The third level addresses specific epilepsy syndromes and allows for more specific syndromic diagnoses to be applied, where appropriate. One such epilepsy syndrome is Infantile Epileptic Spasms Syndrome (IESS), which will be discussed in Chapter 1.1.5.

Aside from the three-level structure of the classification, the revised ILAE 2017 classification of the epilepsies also addresses the aetiologies and comorbidities (Figure 1.1).

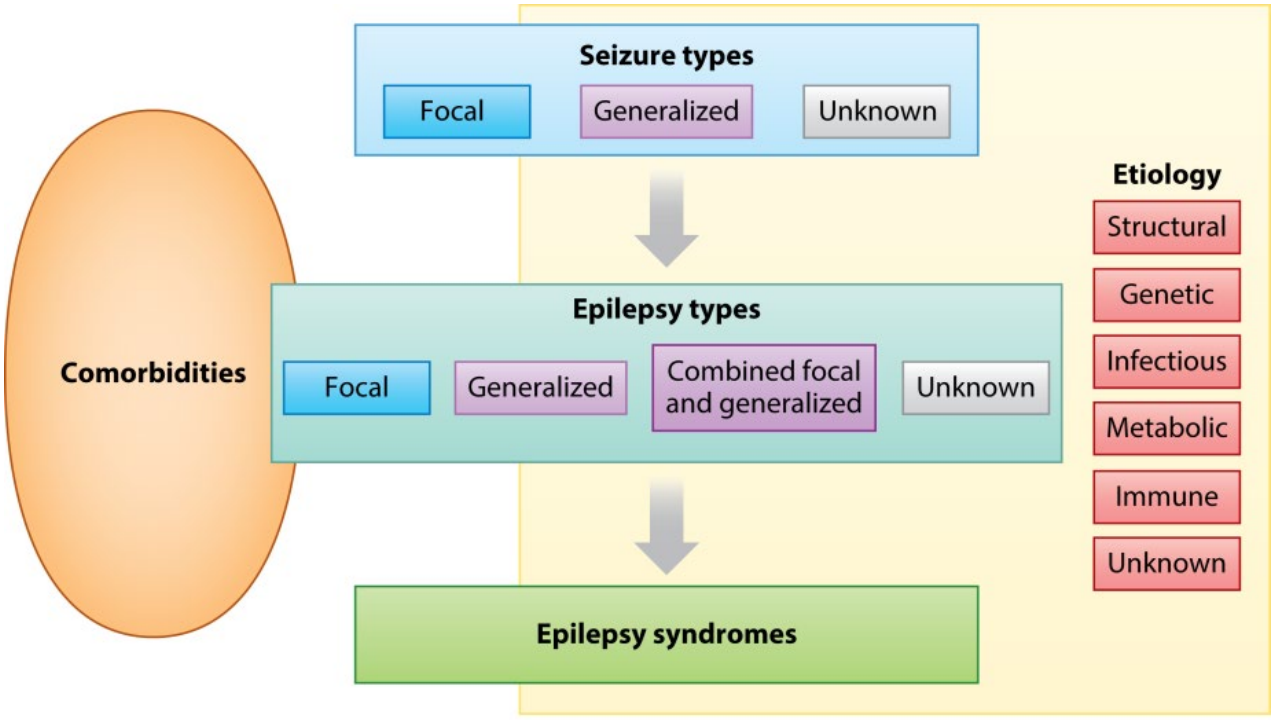


Figure 1.1: International League Against Epilepsy 2017 classification of the epilepsies. Figure adapted from Scheffer et al. 2017 and Perucca et al. 2020 [3, 5].

1.1.2 Symptoms and comorbidities

In addition to the primary symptom, unprovoked seizures, epilepsy is associated with a wide spectrum of symptoms and comorbidities. Electroencephalogram (EEG) abnormalities are observed in epilepsy. Individuals with epilepsy are more likely to

have learning difficulties, intellectual disability (ID) and developmental delay (DD) [6]. Developmental and epileptic encephalopathies (DEEs) are a collection of early-onset severe epilepsies that are characterised by severe DD or loss of developmental skills, and seizures that are usually drug-resistant [7]. Developmental abilities tend to decline with progression of seizures in DEEs. DD in DEEs may result from underlying aetiology, may be influenced by seizure activity, or result from some combination of the two [8].

Psychiatric and mood disorders such as depression, anxiety and bipolar disorder are also more common in individuals with epilepsy, with depression being the most common comorbidity [9]. One in three people with epilepsy will be diagnosed with a lifetime psychiatric condition and the risk of developing a psychiatric disorder is 2-5 times greater in people with epilepsy than the population average [10]. Other psychosocial factors contribute to wellbeing and quality of life for people with epilepsy such as loss of driving license, stigma and treatment cost [9]. Attention Deficit Hyperactivity Disorder (ADHD) and Autism Spectrum Disorder (ASD) also commonly co-occur with epilepsy [11]. Other possible comorbidities include movement disorders, tremor, physical problems including digestive complications and structural brain changes [12-14].

1.1.3 Genetic aetiology

The genetic landscape of the epilepsies is diverse, with a range of molecular pathways involved and genetic aetiologies including rare and common variants of varying penetrance [15]. Aside from Mendelian inheritance models, several other mechanisms have been associated with epilepsy including *de novo* mutation, repeat expansions, somatic mosaicism, X-linked cellular interference (in *PCDH19*, whereby hemizygous males are unaffected but heterozygous females are affected) and polygenic risk scores [16-19].

There are more than 85 genes associated with a primary diagnosis of epilepsy and almost 1000 broader epilepsy-associated genes (i.e. genes associated with epilepsy with other disorders or comorbidities) [20, 21]. Next generation sequencing (NGS)

approaches, such as exome sequencing (ES), whole genome sequencing (WGS) and targeted sequencing have been effective in identifying pathogenic variants in epileptic disorders and establishing novel genes in many epilepsies [3, 22]. In particular, the use of a parent-child trio design with NGS technologies has accelerated novel gene discovery for epilepsy. Trio design analyses have allowed high-throughput identification of *de novo* causal variants absent in parents and recessive or compound heterozygous variants on alternate alleles.

The largest NGS epilepsy study to date is the Epi25 Collaborative ES study, which sequenced 9,170 affected individuals and 8,436 controls from 40 research cohorts globally [23]. Affected individuals had epilepsy that was characterised as severe DEEs, genetic generalized epilepsy (GGE), or non-acquired focal epilepsy (NAFE). Three primary findings of the study were; an enrichment of missense variants in genes encoding $GABA_A$ receptors, an over representation of pathogenic variants in highly constrained genes, and genes encoding ion channels such as *CACNA1G*, *EEF1A2*, *GABRG2*, *LG11* and *TRIM3*.

Similarly, genome wide association studies (GWAS) such as the ILAE complex epilepsies GWAS, which used haplotype data from 15,212 complex epilepsy cases and 29,677 controls, have assisted in identifying regions and genes of interest [24]. This GWAS identified 16 loci of interest, including 11 novel loci, and confirmed association of genes which had been identified through linkage mapping and sequencing analysis, such as *SCN1A* and *SCN2A*.

Genetic studies of epilepsy have been broad and have identified the causes to be monogenic, polygenic, not genetic or unknown. This thesis will focus predominantly on monogenic epilepsies as monogenic causes are the most common causes identified in DEEs, accounting for at least 40% of cases, and pose the primary candidates for precision medicine and studies of molecular mechanisms in epilepsy [25].

1.1.4 Treatments

When appropriately selected, antiepileptic medications can control seizures in 60-70% of epilepsy cases [26]. Common anti-seizure medications (ASMs) include sodium valproate, carbamazepine, lamotrigine and, more recently, cannabidiol (for severe encephalopathies such as Lennox-Gastaut syndrome and Dravet syndrome) [27]. It is important to note that these treatments do not cure the epilepsy, nor treat the underlying cause, but rather suppress the seizures.

The remaining 30-40% of epilepsy patients, for whom ASMs are ineffective, have what is termed drug resistant epilepsy (also called refractory or intractable epilepsy). The ILAE defines drug resistant epilepsy as the failure to prevent seizure recurrence after adequate trials of two tolerated and appropriately chosen ASM (monotherapies or combination therapies) [28]. The implementation of a specific high-fat, low-carbohydrate diet, the ketogenic diet, is an alternate method sometimes used to limit seizures in drug resistant epilepsy with varying degrees of success [29].

Vagus nerve stimulation (VNS) has also proven to be effective in limiting seizures in some individuals with drug resistant epilepsy. There are two methods of VNS, surgically implanted and non-invasive [30]. Surgically implanted VNS involves the implantation of a small pacemaker-like device in the chest. Regularly spaced moderate electrical impulses are delivered by the device to the vagus nerve, which transmits signals to different regions of the brain. Newer non-invasive VNS devices that don't require surgical implementation are also available and have proved effective in some individuals [31].

Another alternative treatment option for some patients with drug resistant epilepsy is neurosurgical intervention. Epilepsy surgery serves to disrupt or resect seizure generating regions of the brain to curtail epileptic activity [32]. Seizure freedom is achieved in the majority of appropriately selected cases and research suggests that if performed before three years of age, surgical intervention may be associated with intelligence quotient (IQ) gains of up to 30 points and improved developmental

outcomes. When surgery is performed after infancy, surgical intervention is associated with halting decline in development but is rarely associated with improvement [32, 33]. Drug resistant epilepsy is commonly seen in epilepsy syndromes such as IESS and in comorbid disorders such as malformations of cortical development (MCD).

Although epilepsy surgery is closely linked to improved health related quality of life, there are factors limiting success and viability of surgical treatment. The average cost of conducting epilepsy surgery in Australia is over $60,000 per surgery and individuals requiring surgery need access to expert medical care [34, 35]. Some individuals may not be candidates for epilepsy surgery due to their age, comorbid disabilities or psychiatric conditions, or the brain region requiring disconnection or resection approximates eloquent cortex [35, 36]. These factors may indicate poor surgical outcomes or present inoperable conditions. Individuals with focal cortical malformations are the strongest surgical candidates.

1.1.5 Infantile Epileptic Spasms Syndrome

IESS is a relatively common DEE, typically arising between 4-8 months of age and characterised by epileptic spasms or infantile spasms (IS) and neurodevelopmental regression, with or without an abnormal, interictal EEG pattern known as hypsarrhythmia [37]. Spasms observed in IESS are characterised by clusters of flexion or extension limb and trunk spasms, predominantly observed shortly after waking [38]. ID is observed in over 80% of individuals with IESS [39]. Between 70-80% of IESS cases also meet the classification criteria for West Syndrome (WS), an epilepsy syndrome that is defined by the triad of epileptic spasms, neurodevelopmental regression and hypsarrhythmia. Together, IESS and WS are the most common forms of severe epilepsy of infancy (SEI), with a combined incidence of 32.7 per 100,000 live births [40]. IESS was previously referred to as Infantile Spasms Syndrome (ISS) but was formally renamed in a 2022 update to the ILAE classification of epilepsy syndromes [38].

There are myriad acquired, genetic, metabolic, structural, and undefined causes of IESS, with MCD being a major contributor [41, 42]. Focal cortical dysplasia (FCD) is the most

common MCD associated with IESS, accounting for up to 18% of all IESS cases [40, 43]. FCD is discussed further in Chapter 1.3.3. Current clinical practice identifies the underlying cause of IESS in up to 60% of cases [44]. Somatic variation has been implicated as a significant cause of MCD and recent research indicates that somatic variants and aberrant gene expression in the brain may represent a major, under-recognised cause of IESS with brain malformations in the undiagnosed cases.

Many causal genes have been identified in IESS (Table 1.1), including *CDKL5*, *STXBP1* and *ARX*, which are also associated with other epileptic disorders [45, 46]. While some molecular pathways are known, the understanding of the broader genetic landscape of IESS due to MCD is limited, largely due to the lack of accessibility of brain tissue in patients with IESS due to brain malformations.

Table 1.1: List of genes frequently associated with IESS (adapted from Pavone et al. 2020) [47].

Gene	Inheritance pattern	OMIM gene	OMIM phenotype
ALDH7A1	AR	107323	266100
ALG13	X-linked	300776	300884
ARX	X-linked	300382	308350
ATP2A2	AD	108740	101900
CDKL5	X-linked	300203	300672
CYFIP2	AD	606323	618008
DCX	X-linked	300121	300067
FOXG1	AD	164874	613454
GNB1	AD	139380	616973
GPT2	AR	138210	616281
GRIN2A	AD	138253	245570
HUWE1	X-linked	300697	309590
IARS2	AR	612801	616007
KCNQ2	AD	602235	613720
KMT2D	AD	602113	147920
MAGI2	AD	606382	617609
NEDD4L	AD	606384	617201
NOS3	AD	163729	163729
NSD1	AD	606681	117550
PAFAH1B1	AD	601545	607432
PHACTR1	AD	608723	618298
RARS2	AR	611524	611523
RYR1	AR	180901	117000
RYR2	AR	180902	600996
PURA	AD	600473	616158
SCN1A	AD	182389	619317
SCN2A	AD	182390	613721
SLC1A4	AR	600229	616657
SPTAN1	AD	182810	613477
STXBP1	AD	602926	612164
TAF1	X-linked	313650	313650
TECTA	AR	602574	603629
TUBA1A	AD	602529	611603
UBA5	AR	610552	617132
WDR45	X-linked	300526	300894
WWOX	AR	605131	616211

AD, autosomal dominant; AR, autosomal recessive; OMIM, Online Mendelian Inheritance in Man.

1.2 Development of the human cerebral cortex

Understanding the normal development of the brain is critical to comprehending possible aetiologies and mechanisms underlying aberrant development and pathogenesis of brain malformations and the resulting outcomes such as epilepsy,

developmental delay and cerebral palsy. The human brain begins development at approximately post conception week (pcw) three and continues developing until it reaches maturity at ~25 years of age [48]. Human brain development begins with primary neurulation and basic patterning. The neural tube forms from the primary ectoderm layer, one of the three primary embryonic germ cell layers, at pcw3 and eventually develops into the prosencephalon, mesencephalon and rhombencephalon. The prosencephalon forms the forebrain and will split into the telencephalon and diencephalon. The telencephalon will develop into the cerebral cortex, hippocampus, olfactory bulb and basal ganglia, while the diencephalon forms the forebrain structures (thalamus, hypothalamus, epithalamus and subthalamus) [49, 50].

The cerebral cortex is the outer layer of neural tissue of the cerebrum of the human brain and it is essential to cognition, motor function and sensory perception [51].

1.2.1 Cortical development

Approximately 16 billion neurons comprise the cerebral cortex, accounting for ~80% of the overall human brain mass [52]. The cerebral cortex plays a crucial role in cognitive function in humans and is comprised of the neocortex (also known as the isocortex) and the allocortex. The neocortex is made up of the hexilaminar (six-layered) cerebral mantle and is responsible for higher-order brain functions, including sensory perception, cognition, language and generation of motor commands. The allocortex is far smaller, accounting for approximately 10% of the cerebral cortex, and consisting of the olfactory cortex and the hippocampus. The allocortex is the ancestral region of the cerebral cortex and is responsible for limbic function of the body and sensory input [53]. The allocortex is comprised of the olfactory cortex which processes smell, the hippocampus which is involved in memory formation and spatial navigation, and the entorhinal cortex which serves as the main interface between hippocampus and other regions of the brain. The human cerebral cortex is significantly larger in size compared to other vertebrates, likely due to evolutionary increases in the volume and types of neural progenitor cells in humans [54].

Cortical development, or corticogenesis, is the process by which the cerebral cortex is formed and involves a series of complex and overlapping developmental stages (Figure 1.2). The four stages most relevant to cortical development and MCD are:

i. Progenitor division and neurogenesis
ii. Neuronal migration
iii. Neuritogenesis and synaptogenesis
iv. Cortical organisation (including apoptosis)

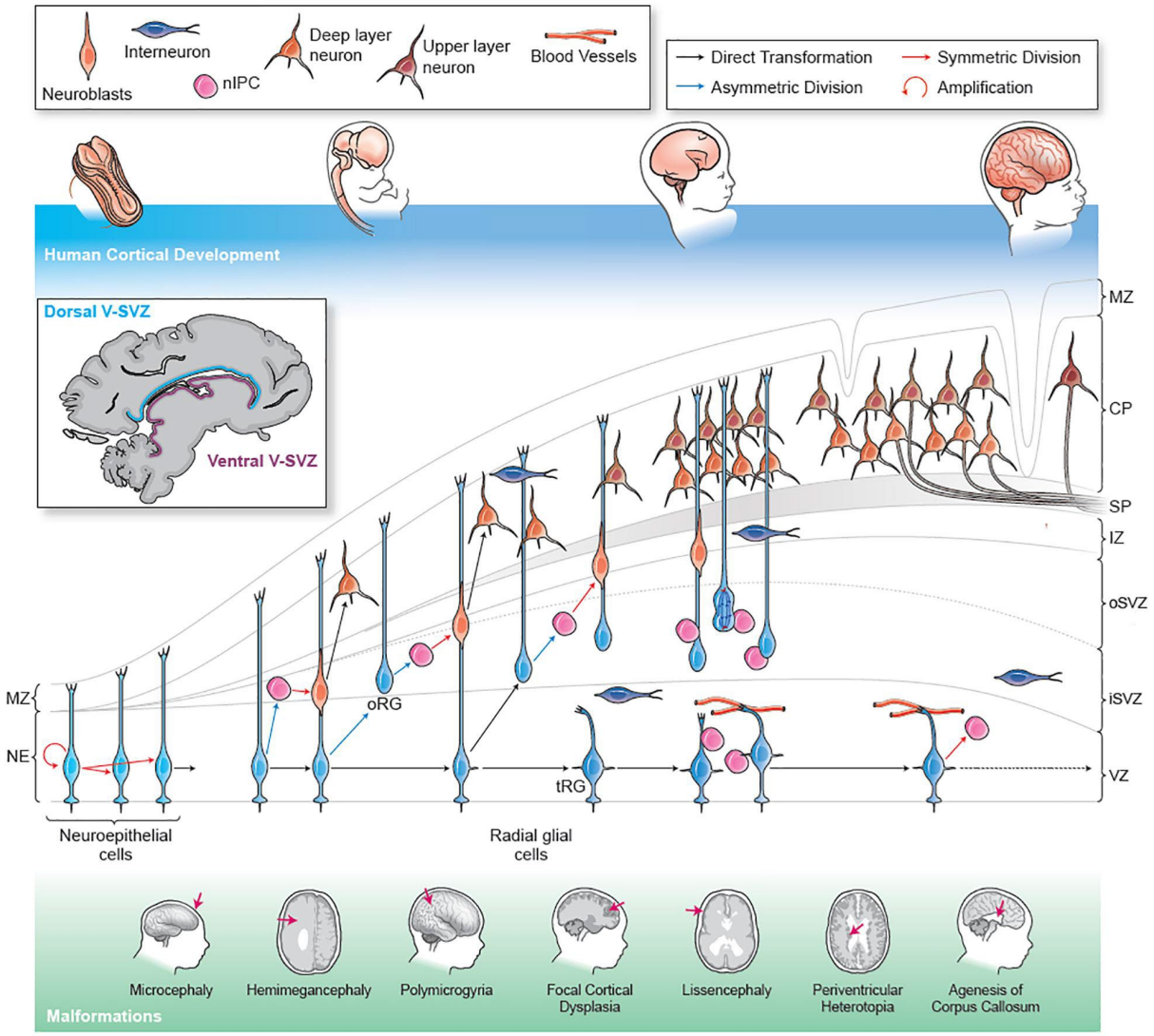


Figure 1.2: Human cortical development and stages of malformation. Figure adapted from Subramanian, Calcagnotto and Paredes, 2020 [54].
CP, cortical plate; iSVZ, inner subventricular zone; IZ, intermediate zone; MZ, marginal zone; NE, neuroepithelium; oRG, outer radial glia; oSVZ, outer subventricular zone; RG, radial glia; SP, subplate; tRG, truncated radial glia; V-SVZ, ventricular-subventricular zone; VZ, ventricular zone.

Neural progenitor cells known as outer radial glial (oRG) cells are found in the developing brain and are particularly important in the embryonic and fetal phases [55]. These cells aid the formation and growth of the neocortex. oRGs play a crucial role in generating the majority of cortical neurons and producing growth factors and signalling molecules to drive neocortex expansion. Increasing knowledge of the function of oRGs would build upon the current understanding of mechanisms of cortical development and may have implications for future studies of neurodevelopmental disorders.

Normal development at each stage of the cortical development process is required for structural integrity and function of the cerebral cortex. As such, malformations of cortical development (MCD) may result from abnormal development at any stage of corticogenesis.

1.2.2 Progenitor division and neurogenesis

Complex radial architecture of cortical neurons is established early in development [56, 57]. A parent population of neuroepithelial (NE) cells, the founder cells of the nervous system, gives rise to cortical excitatory neurons in humans over a period spanning pcw8 through pcw24 [48]. These mechanisms involve coordinated interactions between multiple molecular groups, including growth factors, transcription factors, intracellular signalling molecules and receptors. The first stages of this process are initiated during pcw4, when NE cells divide symmetrically to establish an increased population of NE cells in the ventricular zone (VZ) [58]. This population then generates apical radial glia (aRG, or progenitor) cells. aRG cells are primary progenitor cells in the mammalian cortex that may divide symmetrically to produce two daughter aRG cells that will undergo another cell division, or asymmetrically to generate one self-renewed aRG cell and either a neuron or basal progenitor cell. aRG cells provide pathways for neuronal migration to the cortical plate (CP) [59]. Radial glial (RG) cells have contact with both the basal and apical surfaces. Their basal processes get progressively longer and eventually form the radial scaffolding that provides cortical architecture and routes

for migrating neurons to travel along. The ability of progenitors to undergo symmetric cell division is central to cell proliferation and corticogenesis.

Cortical neurons that produce the neurotransmitter glutamate are described as glutamatergic and termed excitatory neurons. Distinct types of excitatory cortical neurons are sequentially generated from aRG and distributed in the hexilaminar cortical plate in an "inside-out" gradient whereby neurons that are generated earliest form layers VI and V, while neurons generated later in the pcw8-pcw24 time period contribute to layers IV, III and II [59, 60].

GABAergic (gamma-aminobutyric acid-containing, also known as inhibitory) interneurons arise in germinal zones of the ventral pallium and will migrate tangentially into the neocortex in order to establish shorter (local) circuits as the major GABAergic neurons of the cortex [61]. GABAergic neurons may also form the basal ganglia and midline grey matter structures.

Growth factors are exceedingly important to successful cortical development. The families of growth factors including fibroblast growth factor (FGF), epidermal growth factor (EGF) and insulin-like growth factor (IGF) play vital roles in neurogenesis and neuroblast differentiation [62, 63]. Growth factors are established during neural induction but continue to play varying roles in directing differentiation, survival, axon guidance and synaptogenesis [63].

MCD most closely associated with abnormalities of the progenitor division and neurogenesis stage include microcephaly (discussed further in Chapter 1.3.12), focal cortical dysplasia (FCD, Chapter 1.3.3), hemimegalencephaly (HME, Chapter 1.3.5) and tuberous sclerosis complex (TSC, Chapter 1.3.8) [54].

1.2.3 Neuroblast migration

During corticogenesis, newly-generated neurons migrate considerable distances within the developing brain to reach their final anatomical location [59]. The process of neuroblast migration is initiated by mitosis and involves the transport of young

neurons from their embryonic origin to their final laminar position. Neuronal migration may be radial or tangential. In the process of radial migration, post-mitotic daughter neurons travel from the VZ and subventricular zone (SVZ) guided by radial glial (RG) cells into the developing hexilaminar cortical plate, after which, the daughter neurons become the projection neurons [64].

There are two modes of radial migration that projection neurons may utilise: somal translocation (glia-independent), locomotion (glia-guided), or a mixture of both [64]. Somal translocation and locomotion are not specific to cell types. However, somal translocation appears to be more prevalent in the early stages of cortical development while locomotion appears to be favoured when longer migration distances are required [65].

Migrating neurons travel towards the pial surface and separate into their hexilaminar (L1-L6) layers. Once neurons have departed the VZ, they cease to divide [66]. Neuroblast migration typically occurs between pcw12 and pcw20 and the final laminar position is directly influenced by neuronal birthdate, internal and external transcriptional factors and molecular identity of the cortical neuron [67].

Homeobox transcription factors, such as ARX, play significant roles in both interneuron progenitor proliferation and subsequent neuronal migration during corticogenesis [68, 69]. Thus, mutation in the gene encoding ARX has been linked to MCD, particularly lissencephaly (LIS, Chapter 1.3.10) [70]. Additionally, disruption of the neuronal migration stage can lead to abnormal clustering of neurons or "heterotopias" (Chapter 1.3.11), most commonly periventricular nodular heterotopia (PNH) in the ventricular wall or subcortical band heterotopias (SBH) within the cortical layers.

1.2.4 Neuritogenesis and synaptogenesis

Following neurogenesis, neuritogenesis is the first step of neuronal maturation and involves the extension of axon and dendrites from embryonic neurons [71]. This process begins shortly after neurons are born and overlaps with neuronal migration [72]. Neurons typically extend one large axon and multiple smaller dendrites [73].

Axon-dendrite polarization is an important aspect of neuritogenesis. Neurons are one of the most highly polarised human cell types and it is this trait that allows the firing of synapses to transmit information [73]. Neuritogenesis is followed by and overlaps with synaptogenesis, the process whereby synapses form between neurons. Both correct neuritogenesis and synaptogenesis are vital for establishment of proper neuronal connectivity and correct neural circuit assembly [74].

Brain malformations associated with synaptogenesis and synaptic pruning include cortical connectivity disorders and atrophy, hypoplasia or agenesis of the corpus callosum [54]. Disorders of the corpus callosum and cortical connectivity are not discussed in detail in this thesis, as the vast majority of individuals with IESS due to a brain malformation in this study had a focal MCD.

Genes coding for cytoskeletal proteins are relevant to neuronal migration and neuritogenesis. Complete and uninterrupted migratory sequence and neuronal maturation is important for the normal development of the cerebral cortex and is mediated by microtubule and actin cytoskeleton-associated proteins. As such, disruptions in this process or mutation in these genes often results in a class of MCD called tubulinopathy [75, 76]. These can be severe MCD, as in the case of aberrant *TUBB3* expression leading to polymicrogyria (Chapter 1.3.9) or may produce an unviable embryo.

1.2.5 Cortical organisation (including apoptosis)

Cortical organisation establishes the foundation of the complex circuitry that distinguishes the human brain. The process begins as early as pcw12 and continues late into life. Cortical organisation comprises a number of subprocesses including alignment, orientation and layering of neurons, cerebral gyrification, glial proliferation and differentiation and apoptosis [54, 77-79]. Organisation and maintenance of cortical neural networks relies partially on synaptic activity [78].

Cerebral gyrification is the process of routine cortical infolding. It is thought that the abundance of progenitor cells generated during proliferative division in the early

stages of cortical development may have contributed to the gyrification of the cerebral cortex [54]. The vast majority of MCD involve some type of abnormality in cerebral gyrification, including LIS (Chapter 1.3.10) and FCD (Chapter 1.3.3).

Apoptosis (programmed cell death) results in the destruction of excess cells produced during and for the purpose of cortical development. The process is highly conserved and will selectively eliminate up to 50% of neuroblasts in the latter stages of corticogenesis [79]. Apoptosis of neuronal cells serves to reduce cell populations that are no longer required and to adjust maturing neuronal populations to an appropriate size [79].

In addition to the stages already discussed above, there are other important processes required for normal development of the human cerebral cortex, one such process being myelination. The myelination of the axons of neurons is the final developmental stage of corticogenesis and is progressive and directional in the brain. Myelination begins in the late gestational period but primarily occurs postnatally and continues well into adulthood [80]. Myelination is discussed in greater detail in Chapters 4 and 5 of this thesis. Other developmental processes of the cortex are not discussed in detail as this study focuses on structural brain malformations and these other processes fall outside the scope of this thesis.

Understanding the genetic, molecular and cellular processes involved at each level of cortical formation is vital to understanding MCD and providing accurate diagnoses, appropriate prognostic counselling and eventually aiding to develop therapies.

1.3 Malformations of cortical development

MCD encompass a wide spectrum of disorders associated with abnormal development of the cerebral cortex. MCD may result from genetic, infectious, vascular or metabolic causes [81]. Cortical development is an exceedingly complex process and abnormalities at any stage of development can generate varied MCD phenotypes [59].

Prior to the advent of magnetic resonance imaging (MRI), MCD were poorly understood and believed to be rare, as the majority of cases remained unidentified. Diagnosis was generally only made using low resolution CT scanning or following autopsy in severe cases. Following advancements in MRI and medical imaging, MCD have been recognised as a substantial cause of epilepsy, intellectual disability (ID), cerebral palsy and neurodevelopmental disorders (NDD), and recent studies have suggested MCD affect approximately 1 in 2500 children globally [82]. There is a wide range of MCD types with varying clinical manifestations and severity. The most common types of MCD are polymicrogyria (PMG), periventricular nodular heterotopia (PNH) and focal cortical dysplasia (FCD). Collectively, PMG, PNH and FCD, account for more than half of all MCD cases [83].

As subsequent stages of cortical development are dependent on earlier stages, disorders affecting early stages of cortical development often result in additional complications in downstream cortical development and increased severity in phenotypes. Loss of function mutations in genes that encode proteins important for basic patterning such as fibroblast growth factor (FGF8), sonic hedgehog (SHH), HOX and the Wnt signalling pathway often cause severe brain malformations and may be lethal [84]. Structural MCD may be visible macroscopically through clinical imaging or microscopically, using high resolution imaging or histopathological techniques. Subtle MCD may affect cortical structure or function but only be identifiable with specialised techniques [85].

1.3.1 Brain Malformations as a contributor to epilepsy

Epileptic seizures are a common clinical presentation associated with MCD and have been observed in 75% of MCD cases [86]. MCD are particularly associated with drug resistant epilepsy, accounting for ~40% of all drug resistant epilepsies, often requiring surgical intervention to remove lesional tissue for treatment of seizures [54]. Surgical intervention for MCD patients with drug resistant epilepsy has proved to be highly effective at reducing or eliminating seizures, particularly in individuals with focal MCD.

A 2011 study found 72% of all post-operative MCD patients were seizure free upon follow-up at 2 years post-surgery, 65% at 5 years and 67% at 10 years [32]. Genetic causes are thought to underlie the majority of MCD. However, although many genes and pathways have been associated, up to 60% of cases still remain without a genetic diagnosis following clinical diagnostic analysis [51].

1.3.2 Diagnosis and classification

The vast majority of MCD are initially diagnosed by MRI in children presenting with epilepsy, intellectual disability and neurological deficit, though definitive diagnoses and subclassification may require neuropathological findings [85]. The diagnosis rate and recognition of MCD continue to increase with advances in genetics, neuroradiology, electrophysiology and histopathological methods [87]. The complex nature of MCD, regular advances in diagnostic techniques and an evolving understanding of human cortical development make it difficult to maintain a consistent classification system [88-91]. Similarly, the overlapping relationship between clinical phenotype, neuroimaging, histopathological presentation, molecular and genetic aetiology require the use of multidisciplinary diagnostic tools for appropriate classification of MCD.

Approximately 200 genes have been associated with MCD pathogenesis to date [82]. Despite the identification of multiple causal genes and molecular pathways associated with MCD, the exact processes linking disrupted molecular mechanisms with clinical presentation are still poorly understood [59].

1.3.3 Focal cortical dysplasia

First described by Taylor et al. in 1971, focal cortical dysplasias (FCDs) are MCD characterised by cortical dyslamination, abnormal cell types and white matter abnormalities [92]. FCD is the most common form of MCD and individuals with FCD typically present with drug resistant seizures, delayed development and intellectual disability. FCDs account for 17% of all paediatric epilepsy surgeries and, as such, represent a substantial medical burden [93].

In 2018, a consensus classification of FCD subtypes was published by an ILAE task force, following an initial review of FCD classification in 2011 [94, 95]. There are three major forms of FCD with diagnosis and classification being predominantly based on neuropathological observations in a system proposed by the ILAE [94, 95]:

- Type I, characterised by cortical architecture abnormalities, specifically abnormal neuronal migration (i.e. microcolumns) and dyslamination, but with morphologically normal neurons and glial cells. Abnormal cortical lamination may be radial (FCD IA), tangential (FCD IB) or both (FCD IC).
- Type II, characterised by cortical dyslamination with the presence of dysmorphic neurons (FCD IIA), or cortical dyslamination with dysmorphic neurons and balloon cells (FCD IIB).
- Type III, characterised by cortical dyslamination in the temporal lobe and hippocampal sclerosis (FCD IIIA), cortical dyslamination adjacent to glial or glioneuronal tumour (FCD IIIB), cortical dyslamination adjacent to vascular malformation (FCD IIIC), or cortical dyslamination adjacent to any other non-specific brain lesion (FCD IIID).

FCD types I and II are both considered isolated, while FCD type III is associated with the principal lesion [96]. To date, the majority of studies investigating the genetic basis of focal cortical malformations have focused on FCD II due to the well-defined imaging and neuropathological hallmarks providing concise areas for research focus and the fact that individuals with FCD are often the most common candidates for surgery, accounting for >25% of all paediatric epilepsy surgery cases [97, 98].

Mutations in the PI3K-AKT-mTOR pathway are a well-established cause of focal cortical malformations, particularly FCD II [99, 100]. Mutation in *MTOR* and a number of upstream and downstream regulatory genes including *TSC1*, *TSC2*, *RHEB*, *DEPDC5*, *NPRL3, WDR62, AKT3, PIK3CA* and *TBC1D7* have been implicated in focal cortical malformations [93, 99, 101]. As such, the relationship between FCD II and the mechanistic target of rapamycin (mTOR) signalling pathway is well researched [87, 93].

The mTOR-signalling cascade plays a key role in the protein and lipid synthesis, cell growth, proliferation, and metabolism [102]. The presence of dysmorphic neurons (DNs) and balloon cells (BCs) in FCD II provided a strong focal point for such research. Less is known about the molecular aetiologies of FCD I and FCD III.

Disruption of the PIK3-AKT-mTOR pathway is closely associated with FCD, tuberous sclerosis complex (TSC), bottom-of-sulcus dysplasia (BOSD) and hemimegalencephaly (HME). As such, these conditions have been collectively terms mTORopathies [93].

1.3.4 Bottom-of-sulcus dysplasia

BOSD is a milder malformation on the spectrum of FCD II. Dysplastic features of BOSD are localised at the bottom of the sulcus, including cortical thickness and blurring of the grey-white matter boundary and widening of the sulcus [103]. These features fade to a normal gyral crown. There may be a line of high signal on MRI from the base of the affected sulcus to the lateral ventricle known as the "transmantle sign". Histopathological features are consistent with FCD II and can include the presence of DNs and BCs. BOSD is often more difficult to identify on MRI due to the relatively small region of malformation. Identification of BOSD can often be improved with the use of high resolution imaging or co-registered fluorodeoxyglucose positron emission tomography (FDG-PET) [104].

Genetic aetiology of BOSD is predominantly identified as low-level somatic *MTOR* variants or germline variants in *DEPDC5* or *NPRL3*. Mosaic variants are often restricted to BOSD regions and a mosaic gradient decreasing towards top of sulcus and adjacent healthy gyrus has been observed [104]. BOSD is one disorder in which the rare occurrence of two-hit *DEPDC5* mutations have been reported [105]. This mechanism is further discussed in Chapter 1.4.2.

1.3.5 Hemimegalencephaly

Hemimegalencephaly (HME) is the most severe subtype of FCD and shares many neurological, pathological and aetiological features with more restricted forms of FCD. It is characterised by abnormal overgrowth and malformation confined to one cerebral

hemisphere [106]. Seizures associated with HME are often drug resistant and individuals with HME may experience weakness in one side of the body, known as hemiparesis [106]. In the most recent classification of MCD, HME was reclassified as secondary to abnormal neuronal and glial proliferation or apoptosis [89].

Variation in the PI3K-AKT-mTOR pathway is an established mechanism of disease in HME, with the most common genetic causes being variation in *PIK3CA*, *AKT3* and *MTOR* [107-109]. A predominance for low-level somatic variants is observed in smaller malformations such as BOSD while a higher variant allele frequency (VAF), up to 30%, is typically observed in mosaic cases of HME [93, 110].

1.3.6 Mild malformation of cortical development

In 2022, an ad hoc task force of the ILAE diagnostic methods commission published a recommended update to the existing ILAE classification of focal cortical dysplasia, suggesting inclusion of mild malformation of cortical development (mMCD) as a specific histopathological entity discrete from FCD I, II and III [111]. This task force also proposed the inclusion of a formal classification for mild malformations of cortical development with oligodendroglial hyperplasia in epilepsy (MOGHE) and "no definite FCD on histopathology". mMCD are another a form of focal MCD which typically present with heterotopic neurons within or adjacent to layer 1 (mMCD type I) or heterotopic neuronal clusters in deep white matter (mMCD type II) [94].

1.3.7 Mild malformation of cortical development with oligodendroglial hyperplasia in epilepsy

Recently, a new pathological entity, MOGHE has been proposed. First described in 2017 by Schurr et al. [112], MOGHE is identified by an increase in oligodendroglial cells and heterotopic neurons in the white matter, a blurred grey-white matter boundary, and an absence of DNs and BCs. Several studies have since further described the classification and imaging hallmarks of MOGHE and identified somatic variation in *SLC35A2* as the predominant cause of this disorder [113, 114]. MOGHE was included in the recommended update proposed by the ad hoc task force of the ILAE diagnostic

methods commission in 2022 [111]. The clinical features of MOGHE are yet to be fully characterised. The relationship between MOGHE, FCD and somatic *SLC35A2* variation has been evolving over the past few years. The entity of MOGHE is of particular relevance to this thesis and as such, MOGHE will be discussed in depth in Chapters 4 and 5 of this thesis.

1.3.8 Tuberous sclerosis complex

Tuberous sclerosis complex (TSC) is a multi-system genetic disorder whereby hamartomas (i.e. benign tumours) grow on the brain, lung, heart and other systems within the body [115]. Individuals with TSC typically present with rounded, benign lesions known as subependymal nodules on MRI in addition to cortical hamartomas, and most individuals will develop drug resistant seizures. Some subependymal nodules may become subependymal giant cell astrocytomas. The incidence of TSC is estimated to be 1:10,000 live births but is sometimes observed in higher frequencies in some European populations [116].

TSC is caused by pathogenic variants in *TSC1* or *TSC2*, which encode hamartin (TSC1) and tuberin (TSC2) respectively [117]. The TSC1 and TSC2 proteins are upstream regulators of the mammalian target of rapamycin complex 1, mTORC1, a master growth regulator and major protein complex of the mTOR pathway [118]. Inheritance is typically dominant *de novo*, though somatic mosaicism (i.e. variants that arise in somatic cells as a result of postzygotic mutation) in *TSC1* and *TSC2* account for a considerable portion of TSC cases, including as rare second-hit mutations [105, 119]. Somatic mosaicism and the two-hit hypothesis model will be discussed in more detail in Chapters 1.4.1 and 1.4.2, respectively.

1.3.9 Polymicrogyria

Polymicrogyria (PMG) is a subtype of congenital MCD, defined by excessive microscopic folding of cortex, abnormal lamination and disruption of the brain surface with pial defects [51]. Patients present with convolutions to the cerebral cortex and an irregular brain surface as a result of overfolding and fusion of multiple miniature gyri.

Gyri (singular: gyrus) are the neuroanatomical ridges on the mammalian cerebral cortex [120]. The proposed mechanism of pathogenesis for PMG is disruption of cortical surface and leptomeninges interaction [120]. The clinical spectrum of PMG is varied, with common neurological symptoms including epilepsy, developmental delay, motor dysfunction and speech and language disorders [121]. Congenital muscular dystrophy and HME are two possible comorbidities of PMG [122, 123].

PMG is a very heterogenous disorder with many genes and comorbid syndromes reported. The most commonly reported genes associated primarily with PMG are *AKT3*, *C2CD3*, *COL3A1*, *DYNC1H1*, *FKTN*, *FIG4*, *KBP*, *NDE1*, *NHEJ1*, *NSDHL*, *OCLN*, *PI4KA*, *PIK3CA*, *RTTN*, *USP18*, *TMX2* and *WDR62*, many of which play roles in neuronal migration [51]. Most recently, somatic mosaic 1q trisomy in brain tissue has been identified as an emerging major cause of PMG [124]. PMG may also result from hypoxia-ischemia, trauma or congenital infection [125].

1.3.10 Lissencephaly

Lissencephaly (LIS) (also known as agyria/pachygyria or lissencephaly type I) is a spectrum of MCD characterised by the absence (agyria) or reduction (pachygyria) of brain gyri [126]. Together with the sulci (indentations in the cerebral mantle that surround gyri), gyri form the folds of the brain surface. As such, disrupted or incomplete development of these gyri results in an abnormally smooth brain [127]. More specifically, agyria is classified as gyri with sulci >3cm apart and pachygyria refers to sulci 1.5-3cm apart.

The phenotypic spectrum of LIS is variable, depending on the severity of the pachygyria, the underlying cause and the relative effect of additional comorbidities. Individuals with mild LIS may only present with mild ID or DD, while more severe LIS cases may have profound ID and limited survival [128]. Seizures are a feature of most LIS cases and typically present prior to six months of age.

LIS typically has a genetic aetiology, with more than 20 causal genes identified to date, the majority of which encode proteins that play a role in microtubule structure,

mediation or function [127]. Pathogenic variants in the gene encoding Platelet Activating Factor Acetylhydrolase 1b Regulatory Subunit 1 (*PAFAH1B1*), previously known as Lissencephaly-1 (*LIS1*), are the most common genetic cause of LIS [126]. Pathogenic variants in *PAFAH1B1* account for ~41% of LIS cases, and variants in *DCX*, *TUBA1A* and *DYNC1H1* are the next most common genes associated with LIS, accounting for 25%, 5% and 3% respectively [128]. Dinday et al. demonstrated tangential migration of GABAergic interneurons into the embryonic hippocampus is retarded in heterozygous knockout *Pafah1b1* mice [129]. In contrast, overexpression of wildtype *Pafah1b1* results in over-migration of cortical neurons in mice [130], further demonstrating the important role PAFAH1B1 plays in neuronal migration.

1.3.11 Heterotopia

Grey matter heterotopia are cortical malformations that are characterised by the presence of neuronal clusters in abnormal locations. Periventricular nodular heterotopia (PNH) and subcortical band heterotopia (SBH) are the most common heterotopia patterns and present as a grey matter heterotopia found in the periventricular and subcortical zones, respectively. This suggests a failure to complete migration of neurons destined for the cerebral cortex. Seizures are observed in 80-90% of patients with grey matter heterotopia, with more severe phenotypes being associated with more extensive heterotopia or additional MCD [131, 132].

SBH shares a number of similarities to LIS, which is discussed in Chapter 1.3.4 of this literature review. The most prevalent genetic causes of SBH are variation in *DCX* and *PAFAH1B1* genes [133]. While LIS typically consists of agyria or pachygyria, SBH is generally less severe and presents with bilateral smooth bands of grey matter in central or superficial white matter with shallow overlying sulcation in severe cases.

Pathogenic variants in *FLNA* account for the largest portion of PNH cases, including all X-linked cases [134, 135]. *FLNA* encodes Filamin A, an actin-binding protein which serves as a scaffolding protein and plays a role in cellular migration [136]. It is most highly expressed in utero and early infancy, two periods which are fundamentally

critical for cortical development. The phenotypic presentation associated with PNH is extremely varied in heterozygous females, and ranges from asymptomatic nodules to severe NDD, drug resistant epilepsy and large heterotopic regions [131]. X-linked variants associated with PNH are generally lethal in hemizygous males, though several cases of males surviving with partial loss of function variants in *FLNA* have been reported [137].

Variants in the gene encoding ADP Ribosylation Factor Guanine Nucleotide Exchange Factor 2 (ARFGEF2), are a rare cause of PNH. ADP ribosylation factor proteins play an important role in intracellular vesicle trafficking and fusion, providing further evidence linking PNH pathogenesis with aberrant neuronal migration and neural progenitor activity.

Aside from *FLNA* and *ARFGEF2*, pathogenic variants in other genes encoding cytoskeletal proteins including *TUBG1*, *KIF2A* and *MAP1B* have been identified in individuals with PNH [138, 139], in addition to recessive variants in *FAT4* and *DCHS1*. *FAT4* and *DCHS1* encode members of the protocadherin family and are highly expressed in the ventricular zone of the developing human brain [74]. Although less common, PNH may also result from microdeletions or chromosomal abnormalities on chromosomes 1, 5 and 15 [131].

1.3.12 Microcephaly

Microcephaly is defined as an occipito-frontal head circumference (OFC) that is at least two standard deviations (SD) below the mean value for an individual's age and gender [140]. If an individual's head circumference is 3 standard deviations below the mean, they would be classified as having true or severe microcephaly. Moreover, individuals with microcephaly at birth are classified as primary microcephaly (PM), and secondary (SM) if developing postnatally. Microcephaly may be acquired (also called progressive) or genetic in origin and most individuals with microcephaly have moderate to severe ID [138].

Variants in 25 causal genes are recognised as causing primary microcephaly through autosomal recessive inheritance [141, 142]. The majority of these genes encode proteins that play critical roles in centriole biogenesis or duplication and many of these proteins, including *WDR62*, *MCPH1* and *ASP,* commonly localise to the centrosome [140]. Microcephaly is also associated with disruption in mitotic spindle formation, transcription, cell signalling and DNA repair pathways [142]. Subsequently, it is generally accepted that microcephaly arises from cell cycle dysregulation and disruption of neurogenesis.

1.3.13 Cobblestone malformations

Cobblestone malformations describe the presence of an unusual "cobblestone-like" cortical appearance on MRI with a complete loss of normal lamination [51, 85]. Cobblestone malformations are typically bilateral and symmetric. The unusual patterning is a result of over-migration of neurons due to impaired glial limiting membrane function. Cobblestone malformations are associated with dystroglycanopathy affecting pial limiting membrane [143]. They may also be associated with congenital muscular dystrophy and ocular malformations.

Differing genetic aetiologies in histopathological presentations of cobblestone malformations have warranted the introduction of a subclassification of cobblestone malformation type A (caused by variants in *POMT1*, *POMT2* and *FKRP*), type B (associated with variants in *LARGE*) and type C (variation in *POMGNT1*) [144].

1.4 Genetic architecture of MCD

As evidenced above, MCD are a highly heterogeneous collection of disorders with many associated known genes and variants. Inheritance may be dominant *de novo*, recessive, compound heterozygous, somatic mosaic or X-linked. Historically, genetic testing has been performed on DNA derived from peripheral blood mononuclear cells (PBMCs) to identify germline variants (i.e. inherited variants or variants which arise *de*

novo in parental gametes) and parent-proband trios have been sequenced to investigate these modes of inheritance [3].

The pool of genes associated with MCD continues to grow with many publications identifying genes in case studies or small cohorts of family studies. Despite this growing collection of genes, MCD are highly heterogeneous and genetic studies of MCD commonly identify many variants of unknown significance (VUS), a classification applied to genetic variants that lack sufficient functional data required to establish causality or adequate association with disease [145]. Rodent models are most commonly used to investigate cortical development and associated brain disorders, however their relative lack of brain complexity renders them fundamentally different to the human cerebral cortex and limits their value for some research aspects [146].

Recent advances in sequencing technologies have enabled the in-depth study of surgical tissue [147]. This has led to a surge in genetic studies of brain malformations that test the relative importance of alternate modes of genetic pathogenesis that are of specific interest to MCD, including somatic mosaicism and the two-hit hypothesis.

1.4.1 Somatic mosaicism in the brain

Somatic mosaicism is the term that describes the rise of two or more genetically distinct populations of cells within a single individual, as a result of postzygotic mutation. Somatic variants may arise through cell division events at any time during development from conception through well into adulthood.

Somatic mosaicism has become a well-researched topic in human disease and has been implicated in the formation of brain lesions. Increasing research evidence suggests that somatic mosaicism in the brain may be a previously under-recognised cause of MCD [147, 148].

Somatic variants that occur earlier in development often have higher variant allele frequencies and may be detected across multiple tissues. Detection of somatic variants depends on the limit of detection (LOD) of technologies used and VAF in tissue. VAF

is the proportion of cells carrying the variant and can be indicative of the stage during which the variant arose. If a pathogenic postzygotic variant arises in a neural progenitor cell, that cell may give rise to a subpopulation of cells containing the somatic variant in the brain. This can happen at any time during development and though some early instances of postzygotic mutation may be detectable in tissues throughout the body, mutations that occur later in development may result in low-level mosaicism (i.e. low VAF) in the brain, which is often undetectable in blood-derived DNA [147]. Thus, somatic variants arising after gastrulation are difficult to identify in blood-derived DNA.

Similarly, due to the relatively low threshold of detection of conventional Sanger sequencing methods, somatic variants in brain tissue that have a VAF of less than 20% were unlikely to be captured with certainty prior to the development of more sensitive technologies such as next generation sequencing. Low-level mosaicism in brain tissue from MCD would previously not have been detected. Applying tools such as targeted sequencing, deep WGS or ES, or digital droplet PCR (ddPCR) can enable identification of somatic mosaic variants more easily [149, 150].

Single nucleotide variants (SNVs) constitute the majority of reported mosaic variants in human diseases but somatic mosaicism may also involve small insertions or deletions (indels), retrotransposition of long or short interspersed nuclear elements (LINEs/SINEs), copy number variants (CNV) or chromosomal aneuploidies [147].

Detection of somatic variants in the brain requires access to brain tissue, which is generally only available as resected brain tissue from neurosurgery or autopsy specimens [151]. Comparison between VAF in DNA in tissues from the brain and other regions of the body (e.g. skin or PBMCs) can assist in identification of pathogenic mosaic variants in the brain.

Some hypotheses have been proposed for minimally invasive methods to collect biological samples for somatic variant detection, such as using cell free DNA (cfDNA) from cerebrospinal fluid (CSF) or DNA from neuro-epithelium in the nasal passage

[147]. However, these methods are still in development and as such, current tissue collections are predominantly a result of epilepsy or tumour surgery.

1.4.2 Two-hit hypothesis

The two-hit hypothesis builds upon existing knowledge of postzygotic mutation in the developing brain. It was first hypothesised by Alfred G. Knudson in 1971 and states that a recessive disorder can be caused by two mutational events [152]. The first is a germline variant and the second occurs through postzygotic mutation arising on the alternate allele of the same gene, resulting in somatic mosaicism of compound heterozygous cells [153].

This mechanism of disease was initially postulated for cancer and first demonstrated in *Tsc2* in an Eker rat model but has since been observed in other disorders, including MCD [105, 154]. In the case of MCD, the two-hit hypothesis suggests a predisposition to the neurodevelopmental phenotype due to a germline variant but requires a second, localised somatic mutation to occur in the developing brain to cause the lesion to form. Lee et al. demonstrated that a somatic second hit *DEPDC5* variant identified in FCD IIA brain tissue was limited to dysmorphic neurons [105]. The two-hit model has also been identified as a mechanism of disease in rare cases of MCD due to mutations in *TSC2* and *STXBP1*, and one dual-pathway two-hit in *NPRL3* and *WNT2* [105, 155-157]. Second hits have proven rare and elusive, possibly due to technical inability to capture low VAF or second hits being restricted to very small regions that may not be captured in the tissues or cells sampled [105].

1.4.3 Brain malformation research in the single cell-omics era

Upon their introduction, NGS technologies such as WGS or ES significantly accelerated the identification of disease-causing variants and identification of disease-associated genes. Technological advancements have continued and since the advent of more recent single cell technologies, such as single cell on single nuclei RNA sequencing (scRNA-seq or snRNA-seq), single-cell chromatin accessibility sequencing (scATAC-

seq) and single cell isoform RNA seq (scISOr-seq), the landscape of gene discovery and molecular mapping has been rapidly changing [158-161].

Barcoded systems used in scRNA-seq allow interrogation of expression profiles from individual cells within a population. The use of scRNA-seq technologies enables tissue-specific or cell-specific variation to be identified and characterised. The scale and capabilities of these technologies are facilitating interrogation of genetic, epigenetic and molecular mechanisms of pathogenesis at an unprecedented depth.

Similarly, new technologies focused on long read sequencing (LRS) at a single-cell level, termed third-generation sequencing (TGS), are currently emerging, allowing cell-specific interrogation of genetic variants and transcriptomics [162].

1.4.4 Precision medicine and surgery in MCD

Developments in precision medicine have led to the trial of mTOR inhibitors as an ASM for mTORopathies in recent years [163]. Several mTOR inhibitors have been trialled and approved with varying success in patients [164, 165]. It is clear that dosage, timing of administration and individual genetic aetiology influence efficacy of mTOR inhibitor treatment in clinic. Precision medicine treatments require knowledge of the underlying cause of disease. As such, precision medicine treatment options are limited for children with severe infantile epilepsy as a result of MCD without a known genetic basis.

Surgical approaches for removing MCD, particularly FCD and TSC, have evolved over the years with limited resection guided by imaging and intraoperative electrophysiology now becoming more common. Laser interstitial thermal therapy (LITT) and stereo electroencephalography–guided radiofrequency thermocoagulation (SEEG-guided RF-TC) are emerging forms of ablative surgery that do not require resection of brain tissue [166]. Whilst resective techniques are more common, it is important to capitalise on researching the underlying basis of these MCD using brain tissue. An improved understanding of the genetic basis of MCD causing severe infantile epilepsies may assist in the identification of targeted treatments or precision medicine options that address the underlying cause of disease. The requirement for a

better understanding of the genetic basis of epilepsy necessitates further research into the molecular and cellular pathways dysregulated in mTORopathies and epileptogenesis more broadly.

1.5 Significance

Epilepsy is one of the most common neurological disorders, affecting an estimated 65 million people globally. Understanding the pathogenesis and dysregulated molecular pathways underlying MCD will not only provide crucial understanding of the genetic mechanisms involved in epileptogenesis, but also reveal molecular mechanisms underlying the normal development and function of the human brain.

Identifying a molecular aetiology signals an end to the diagnostic odyssey for patients and their families, informs genetic counselling and has the potential to affect clinical management. Additionally, broadening the library of genes associated with epilepsies and MCD allows for the development of more accurate diagnostic genetic testing and the potential for precision medicine and targeted therapies.

Similarly, identification of genes, molecular mechanisms or pathways of interest can also assist future genetic studies and provide a foundation for future pathway-based precision medicine, as in the case of the mTOR pathway, previously [147].

1.6 Project aims

The hypothesis of this project is that the application of genomic technologies will enable the investigation of the genetic landscape of common subtypes of MCD causing IESS and facilitate a better understanding of the genetic causes of IESS with MCD. Furthermore, that functional characterisation of these genes will assist to identify crucial molecular pathways in the normal development and function of the human brain.

To accomplish this, there are three major aims:

Aim 1: To investigate the genetic landscape of infantile epileptic spasms syndrome (IESS) in a cohort of paediatric patients with IESS and MCD.

Aim 2: To identify the molecular hallmarks and phenotypic classifiers in cases with brain mosaic *SLC35A2* variants and evaluate the relationship between mosaic *SLC35A2* and mild malformation of cortical development with oligodendroglial hyperplasia in epilepsy (MOGHE).

Aim 3: To interrogate the molecular mechanisms involved in the pathogenesis of MOGHE using snRNA-seq and proteomics data.

2 Chapter 2: Materials and methods

2.1 Standard protocols and consents

The Human Research Ethics Committee of The Royal Children's Hospital, Melbourne, Australia, approved this study (HREC 29077 and HREC 32288). Written informed consent was obtained from participants' parents or legal guardians. All individuals in this study were initially diagnosed with drug resistant epilepsy due to a brain malformation. Brain tissue samples were collected at the Royal Children's Hospital, Melbourne. Epilepsy surgeries were performed under general anaesthesia by a specialised clinical team with image guidance and intraoperative electrocorticography. Clinical, imaging and surgical details for each individual were reviewed and provided by neurologists at the Royal Children's Hospital. Brain tissue samples collected during surgery were stored as fresh-frozen specimens (-80 °C) in polypropylene containers, and archived formalin-fixed paraffin-embedded (FFPE) tissue blocks stored with the Royal Children's Hospital Pathology Department. Blood samples were collected in Vacuette 9 mL K3E K3EDTA tubes (Interpath, #454034) and stored at -30 °C.

2.2 Routine molecular techniques

A series of routine molecular techniques were used in the procurement of data for this thesis. All techniques were carried out in the Neurogenetic Research laboratory within the Bruce Lefroy Centre for Genetic Health at the Murdoch Children's Research Institute in Parkville, Victoria.

2.2.1 DNA, RNA and protein extraction

Genomic DNA and RNA was extracted from fresh frozen brain tissue using the AllPrep DNA/RNA Mini Kit (Qiagen, #80204), according to the manufacturer's instructions. Prior to DNA/RNA extraction, fresh frozen brain tissue specimens of 15 – 30 mg were cut from tissue blocks and homogenised in 600 µL Buffer RLT containing 1% β-

mercaptoethanol (β-ME) by manually disrupting tissue using 19G, 21G and 25G needles with 3 mL syringes.

DNA was eluted into 100 µL pre-heated (75°C) buffer EB and the frozen at -80°C for storage. RNA was eluted into 60 µL RNase-free water and stored at -80°C.

Following processing through the RNA shredder column of the AllPrep DNA/RNA Mini Kit, protein was purified from RNA flow-through elution via acetone precipitation. Pre-chilled (-20°C) acetone (80% final v/v) and 1M NaCl (10 mM final concentration) were added to RNA flow-through elution and precipitated overnight at -20°C. Precipitate was pelleted by centrifugation at 13,000 rcf for 15 minutes at 4°C and pellet was washed twice with 70% ethanol. Protein pellet was resuspended in 75 µL of pre-heated (95°C) 5% SDS/10 mM Tris-HCl, pH 7.5 buffer resulting in a final concentration of approximately 1 - 2 mg/mL.

Genomic DNA and RNA from saliva and blood samples were extracted using NucleoBond CB 20 and CB 100 midi kits (Macherey-Nagel, #740507 and #740508), respectively, according to manufacturer's instructions.

2.2.2 Nucleic acid quantification

1.5 µL purified DNA or RNA was used for quantification on a NanoDrop One (Thermo Scientific). In instances where nucleic acid quantity or purity was low, nucleic acid quantification was performed on a Qubit 4 Fluorometer (Invitrogen), using the Qubit dsDNA BroadRange Assay Kit (Invitrogen, #Q32850) or Qubit RNA BroadRange Assay Kit (Invitrogen, #Q10211), according to manufacturer's protocols.

2.2.3 Protein quantification

Protein quantification was performed on a FLUOstar Optima microplate reader (BMG LabTech) using the Pierce bicinchoninic acid (BCA) Protein Assay Kit (Thermo Scientific, #23225). Protein quantification analysis and calculations were carried out using Optima MARS v3.01 R2 software.

2.2.4 Gel electrophoresis

DNA and RNA integrity and PCR amplicon sizing were validated using gel electrophoresis. Separation and analysis of DNA, RNA, and PCR amplicons were carried out using electrophoresis on 0.8%, 1.5% and 1% (w/v) agarose gels respectively. Agarose gels consisted of agarose (Bio-Rad, #161-3102), 0.5X TBE Buffer (45 mM Tris, 45 mM borate, 1mM EDTA) and 1X GelRed Nucleic Acid Gel Stain (Biotium, #41002). 8 µL sample were combined with 2 µL 10X loading dye (30% glycerol, 0.5X PCR buffer, 0.1% bromophenol blue) and run at 85V for 20 – 40 minutes at 85 V. Gels were imaged using UV transillumination on a Gene Genius Bioimaging System (SynGene). Nucleic acid band sizing was determined using the 1 kb Plus DNA Ladder (Life Technologies, #10787-018).

2.2.5 Polymerase chain reaction

All polymerase chain reaction (PCR) experiments were performed using the following method unless otherwise specified. PCR and sequencing primers were designed using Primer3 web v4.1.0 [167]. PCR reaction mixes were assembled in 25 µL mixtures containing 1x MyTaq Reaction Buffer Colorless (Bioline, #BIO-37111), 0.33 units MyTaq DNA Polymerase (Bioline, #BIO-21105), 1 µM forward primer, 1 µM reverse primer and 20-50 ng DNA template. 5X MyTaq Reaction Buffer Colorless comprises 5 mM dNTPs, 15 mM MgCl2, stabilizers and enhancers. Reactions were run on a Veriti 96-well Thermal Cycler (Applied Biosystems) using the following cycling conditions at a ramp rate of 2°C per second: 2 minutes at 95°C; 35 – 40 cycles of 95°C for 30 seconds, 59 – 62°C for 30 seconds and 72°C for 1 – 2 minutes; 72°C for 5 minutes, then held at 4°C and stored at -20°C. PCR amplicons were purified using the ExoSAP-IT PCR product clean up reagent (Applied Biosystems, #78201.1.ML) prior to Sanger sequencing.

2.2.6 Sanger sequencing

Sanger sequencing reactions were performed in 10 µL reactions using the BigDye Terminator v3.2 Cycle Sequencing kit (Life Technologies, #4337455). Reaction mix consisted of 1X BigDye Terminator Buffer, 0.5 µL BigDye Terminator with ddNTPs, 0.25

µM primer, 3.5 µL dH_2O and 3.5 µL ExoSAP-IT reaction product. Reactions were run on a T-100 Thermal Cycler (Bio-Rad) using the following cycling conditions at a ramp rate of 1°C per second: 1 minute at 95°C; 25 cycles of 95°C for 1 minute, 50°C for 5 seconds, 60°C for 4 minutes; followed by a hold at 4°C. Sanger sequencing products were submitted to the Translational Genomics Unit (TGU) at the Victorian Clinical Genetics Services (VCGS) (Melbourne, Australia) for ethanol precipitation clean up and sequencing performed on a 3730xl DNA Analyzer (Thermo Fisher Scientific), on a fee for service basis. Sanger trace data were aligned and analysed using CodonCode Aligner v9.0.1 (CodonCode Corporation).

2.3 Next generation sequencing

2.3.1 HaloPlex^{HS} targeted panel sequencing

Genetic testing was performed on genomic DNA extracted from resected brain tissue. Individuals with TSC or FCD IIB, expected to carry variants in PI3K-AKT-mTOR pathway or GATOR1 complex genes, in this cohort received targeted panel sequencing using a previously described, custom-designed HaloPlex^{HS} gene panel using the SureDesign software (Agilent) [104, 168]. This panel consisted of 331 genes associated with malformations of cortical development, covering a combined target region of 1.334Mbp with a predicted coverage of 99.86% within target regions (Supplementary Table 8.2). DNA input of 50ng was used to generate paired-end, barcoded libraries using custom HaloPlex Target Enrichment Kits (Agilent). Paired-end sequencing was completed on NextSeq (Illumina) with a target sequencing depth of >500x coverage per sample. Sequencing was performed on a research basis at the VCGS, Melbourne, Australia.

2.3.2 Exome sequencing (ES)

Exome sequencing (ES) was performed on brain-derived DNA using either the SureSelect Clinical Research Exome version 2 (CREv2) or Twist Exome 2.0 (TWIST) library preparation. Sequencing was performed on an Illumina NovaSeq6000

sequencer (2 × 150 bp) at a 400x target read depth. Sequencing was performed by the Translational Genomics Unit (TGU) at the Victorian Clinical Genetics Services (VCGS) (Melbourne, Australia) on a fee for service basis.

2.3.3 Exome data analysis

A tier 1 IESS gene list was compiled which included 130 genes known or predicted to cause IESS in a dominant inheritance pattern and 50 MCD genes in which mosaicism is an established or suspected mechanism (Supplementary Table 8.1). Germline and somatic candidate variants were detected in ES data using HaplotypeCaller (joint-calling) and Mutect2 (tumour-only mode), respectively, following the GATK12 Best Practice Workflows (v4.2.0.0, Broad Institute). Variants were prioritised using seqr (v.0.3.0, https://github.com/broadinstitute/seqr) with the following inclusion criteria: (1) within the tier 1 IESS and MCD mosaicism gene list (180 genes), (2) coding/nonsynonymous, an insertion/deletion, or close proximity to a splice site, (3) Genome Aggregation Database (gnomAD) population allele frequency (AF) ≤0.001 in dominant genes or ≤0.01 in recessive genes, and (4) predicted to be damaging by two or more *in silico* tools (CADD [169], REVEL [170], PrimateAI [171], MPC [172], SpliceAI [173], EIGEN [174], PolyPhen [175], SIFT [176], MutTaster [177], FATHMM [178], MetaSVM [179]). Candidate variants were reviewed and classified as pathogenic (P), likely pathogenic (LP) or variant of unknown significance (VUS) according to American College of Medical Genetics and Genomics (ACMG) guidelines [145]. Inherited or somatic mosaic variants that were predicted damaging (CADD>25) in genes of unknown significance or genes relating to other pathologies were classified as VUS. Somatic mosaic VUS were only considered if they met the threshold of >2 bidirectional reads and present in a gene of relevance.

Analysis of potential copy number variants (CNV) within this cohort was carried out on singleton and trio ES data using C to Go translator, CxGo (v. 0.3.7, https://github.com/gotranspile/cxgo) [180]. CNVs were filtered as being *de novo* and

predicted to damaging by two or more in silico tools (ed, xhmm, cdx). Regions of interest were flagged based on existing clinical and neuropathological indications.

2.3.4 Trio exome analysis

If individuals remained unsolved following singleton ES and/or HaloPlex^{HS} panel analysis, parental gDNA extracted from saliva or blood was sent for standard clinical ES (100x depth) using Twist Exome 2.0 (TWIST) library preparation. Trio data was analysed using seqr. Predicted damaging variants were filtered for recessive, compound heterozygous or dominant *de novo* inheritance.

2.3.5 Bulk RNA sequencing

In addition to trio analysis, cDNA from lesional brain tissue from individuals without a genetic diagnosis via ES or HaloPlex^{HS} panel sequencing was submitted for bulk RNA sequencing through VCGS using TruSeq stranded mRNA preparation, pooling, and NovaSeq 6000 sequencing with a target of 80M reads. cDNA synthesis was performed using the SuperScript III First Strand Synthesis Kit (ThermoFisher #18080051). Bulk RNA-seq data files were processed into read count files using the Spliced Transcripts Alignment to a Reference (STAR) aligner module and the Subread-featureCounts pipeline [181, 182]. Read count files were uploaded into Degust web platform for analysis (https://degust.erc.monash.edu/) using limma/voom settings.

2.4 Droplet digital PCR

Variants with a variant allele frequency (VAF) detected below 35% in ES or HaloPlex^{HS} data was considered likely mosaic. Variants with a VAF greater than 35% were considered likely mosaic if detected on the X chromosome in hemizygous males. Mosaic variants that were novel or in genes which were not commonly associated with mosaicism were validated. Published, recurrent mosaic variants in mTOR pathway genes were only validated if VAF was less than 4% in ES or HaloPlex^{HS} data.

Mosaic VAF was confirmed via droplet digital PCR (ddPCR) using Custom TaqMan SNP Genotyping Assays (Thermo Fisher Scientific, #433139) with sequence-specific primers and probes (primer and probe details listed in Supplementary Table 8.3). Probes were designed with Applied Biosystems FAM™ and VIC™ dyes with fluorescence emission spectra of ~517 nm and ~554 nm, respectively. Droplet generation, PCR cycling, and droplet reading were performed on the QX200 ddPCR system (Bio-Rad) according to the manufacturer's recommendations.

Heterozygous variants (VAF >40%) were validated by Sanger sequencing using region-specific primers designed in-house.

2.5 Immunostaining

2.5.1 Immunohistochemistry

Immunohistochemical staining was completed on 3 µm FFPE tissue sections according to established protocols [183]. FFPE tissue slides were dewaxed in xylene twice for 5 minutes each then rehydrated in serially diluted ethanol, starting with 100% ethanol for 4 minutes, then 90%, 70%, 50% for 2 minutes each followed by 5 minutes in tap water. Antigen retrieval was achieved by completely submerging tissues in boiling citrate-based antigen unmasking solution (VectorLabs, #H3300) and microwaving for 15-20 minutes, without over-boiling, then allowing to cool to room temperature. Tissue samples were rinsed with tap water then incubated in methanol peroxide solution (ratio 5:1) at room temperature for 30 minutes. Tissues were rinsed again with tap water before incubation in pre-chilled 0.1M Tris pH 7.5 then blocking buffer (2% FBS in 0.1M Tris pH 7.5), on ice for 5 minutes each. Overnight incubation was completed with blocking buffer containing primary antibody at 4°C. Following overnight incubation, slides were rinsed in pre-chilled 0.1M Tris pH 7.5 then blocking buffer, on ice for 5 minutes each. Slides were incubated with blocking buffer containing secondary antibody (dilution 1:1000) for one hour at room temperature.

2.5.2 Chromogenic staining

Avidin-biotin treatment was performed with the Vectastain ABC-HRP kit (VectorLabs, #H4000) for one hour at room temperature, followed by treatment with 3,3'Diaminobenzidine (DAB) substrate kit (VectorLabs, #SK-4100) to allow visualisation of immunoreactivity. Solutions were removed by washing in 0.1M phosphate buffer three times for 10 minutes each. Tissues were counterstained with hematoxylin (VectorLabs, #H3401) then washed with tap water and dipped 6 times in 70% ethanol solution containing 0.37% HCl before rinsing with tap water and dehydrating with reverse dewaxing protocol. Coverslips were mounted with Ultramount No.4 (Fronine, #FNNII063C). Slides were imaged on a Zeiss Axio Imager Z2 microscope.

2.5.3 Antibodies

Slides were incubated overnight with primary antibodies at an appropriate concentration in blocking buffer, as follows; anti-Olig2 (EP112) rabbit monoclonal antibody (Olig2) (CellMarque, #387R-16, dilution 1:1000), anti-Phospho-S6 Ribosomal Protein (Ser235/236) (pS6) rabbit monoclonal antibody (Cell Signaling, #4858P, dilution 1:100), anti-neurofilament SMI-311R antibody (SMI-311) (BioLegend, #837801, dilution 1:1000), anti-NeuN rabbit monoclonal antibody clone 27-4 (NeuN) (Millipore, #MABN140, dilution 1:100) and anti-Map2 (2a+2b) mouse monoclonal antibody (Map2) (Sigma, #M1406, dilution 1:500). Hematoxylin staining was completed to identify nuclear components in the tissues. Table of antibodies used and concentrations listed in Supplementary Table 8.6.

2.5.4 Neuropathologic reassessment

Blinded neuropathologic reassessment was performed by a consultant neuropathologist on newly stained research brain FFPE tissue samples. If multiple specimens were available, the region closest to the most prominent malformation on MRI was selected. Individuals with genetic and histopathologic findings of a well-established phenotype-genotype correlation (i.e. TSC due to variants in *TSC1* or *TSC2*,

or FCD II due to a variant in an mTOR pathway gene) did not routinely undergo neuropathologic reassessment.

2.6 Single nuclei RNA sequencing

2.6.1 Nuclei isolation and sample preparation

Approximately 20 mg fresh frozen brain tissue was dissected from frozen surgical brain specimens on dry ice. Tissue dissociation was carried out by thawing tissue on ice in 1 mL lysis buffer (0.005% Nonidet P40, 10 mM Tris-HCl pH 7.5, 3 mM $MgCl_2$, 10 mM NaCl, 200 U/mL RNase inhibitor (Promega #N261B)). Tissue was manually disrupted by gently passaging 10 times sequentially through a 19G needle with 3 mL syringe and then incubated on ice for 20 minutes with gentle mixing with a 21G needle and 3 mL syringe every 5 minutes to further dissociate large clumps. 2.5 mL 1x PBS was added to dissociated tissue and filtered using a 30 µm cell strainer into a pre-chilled 15 mL tube. Lysate was centrifuged at 450 rcf for 5 minutes and 4°C. The supernatant was removed and discarded. The pellet was gently resuspended in 500 µL nuclei wash buffer (1x PBS containing 3% bovine serum albumin (BSA) and 4.5 µM DAPI (Sigma #D9542)) by flicking and transferred into a 5 mL polypropylene tube via a cell strainer. Independent nuclei were isolated into 1.5 mL Lo-Bind collection tubes and additional cellular debris was removed using fluorescence activated nuclei sorting (FANS) through a 70 µm nozzle at 20 psi with DAPI stain as a guide, targeting 100,000 events. 2 µL of RNase inhibitor was added to the collection tube prior to sorting to ensure nuclear RNA integrity and quality was not compromised. Nuclei isolation and preparation protocol was performed in an RNase free workspace.

2.6.2 Chromium Next GEM Single Cell 3 Gene Expression v3.1

Following FANS, nuclei concentration was estimated using a hemocytometer. Approximately 7000 nuclei were loaded onto the 10X Chromium System to generate barcoded cDNA libraries using the Chromium Next GEM Single Cell 3' Gene Expression kit v3.1 (dual index) (10X Genomics, #PN1000130) and library construction kit (10X

Genomics, #PN-1000196). GEM generation and barcoding were performed according to the manufacturer's protocols, using dual index kit TT set A (10X Genomics, #PN-1000215) and Chromium Next GEM Chip G Single Cell Kit (10X Genomics, #PN1000127).

2.6.3 Bioinformatic processing and analysis

Pipeline processing of snRNA-seq data was performed by Dr Saskia Freytag at the Brain Cancer Centre within the Walter and Eliza Hall Institute of Medical Research (WEHI), using a customised pipeline. Illumina sequencer base call library (BCL) files were processed and demultiplexed to produce fastq files in CellRanger v6.1.1 with default parameter for adapter trimming and UMI quantification counting [184].

An include introns requirement was set for pipeline analysis to ensure the capture of intronic reads and a pre-mRNA reference was created from the 10X Genomics refdata-cellranger-GR38-1.2.0 transcriptome. CellRanger data was examined for mitochondrial contamination and cleaning of nuclei was performed then exported into R for further analysis using DropletUtils v1.6.1, scater v1.14.6 and scran v1.14.6 [185-187]. Doublets were removed with scDblFinder and integration of multiple datasets was performed using Harmony v0.1.1 to eliminate batch effects [188, 189].

The integrated dataset was reduced to 5000 highly variable genes for principal component analysis (PCA) and annotation. The number of relevant dimensions in dataset was determined by maxLikGlobalDimEst and set to 17 [190]. The final dimension reduction was used for clustering and the generation of Uniform Manifold Approximation and Projection (UMAP) plots. Clustering was completed using Orchestrating Single-Cell Analysis with Bioconductor (OSCA) pipeline and the Leiden algorithm used with 50 neighbours, then a published set of markers were utilised to identify cluster cell type identities and subsequent dataset annotations [191, 192]. A pseudobulk analysis was completed for MOGHE_lesional and MOGHE_normal annotations separately and limma voom analysis was used to identify differentially

expressed genes (DEGs), followed by the selection of most significant markers based on p-values.

DEGs of interest were identified by applying a log fold change (logFC) threshold of a greater than 1 or less than -1 between MOGHE_normal and MOGHE_lesional tissues and an adjusted p-value (adj. p) less than 0.05.

2.7 Proteomics analysis

2.7.1 Protein isolation and enzymatic digestion

Entire proteomes isolated from brain tissue were analysed by liquid chromatography–mass spectrometry (LC-MS/MS) with Tandem Mass Tag 16pro (TMT-16pro) Quantitation. Samples were processed with a minimum input of 200 µg total protein. Processing and analysis were carried out by Monash Proteomic and Metabolomic Facility (MPMF) (Clayton, Australia) on a fee for service basis. The protein intensity values for each sample were generated by MPMF using Proteome Discoverer 2.5 [@RN46] software from ThermoFisher.

Samples were resuspended to a 5% sodium dodecyl sulphate (SDS) 10mM Tris HCL, heated at 95°C for 10 min and centrifuged at 13,000 g for 5 minutes. Samples were then processed using the S-trap protocol as per the manufacturer's instructions (https://pubs.acs.org/doi/10.1021/acs.jproteome.8b00505). Briefly, samples were reduced and alkylated with 10mM TCEP (Thermo, #77720) and 40mM chloroacetamide (Sigma, C0267-100G) with incubation at 55 ºC for 15 minutes. Enzymatic digestion was performed using Trypsin (Promega, V528X) at a 1:50 wt:wt ratio alongside Lys-C at a 1:25 wt:wt ratio (Promega, VA1170).at 37 ºC for 16 hours. Digestion efficiency was greater than 94% for this analysis.

2.7.2 TMT labelling and fractionation

Each sample was labeled with the TMTpro 16plex reagent set (Lot:XC34162, Thermo Scientific) according to the manufacturer's instructions and utilized a singular

reference channel (126) containing all samples pooled. Individual labelled samples were then pooled into plexes, and high-pH RP-HPLC was used to generate 12 fractions, which have been acquired individually by LC-MS/MS to maximize identifications. Labelling efficiency was determined to be greater than 97%.

2.7.3 Liquid Chromatography Mass Spectrometry Protocol

Liquid chromatography-mass spectrometric (LC-MS) analysis was conducted using the Mass Spectrometer and Nano LC system (Dionex Ultimate 3000 RSLCnano). The samples were loaded in an Acclaim PepMap RSLC (75 µm x 50 cm, nanoViper, C18, 2 µm, 100Å; Thermo Scientific) analytical column. The peptides were separated by increasing concentrations of buffer B (80% acetonitrile/0.1% formic acid) and analyzed via 2 kV nano-electrospray ionisation with an Orbitrap Eclipse Tribrid mass spectrometer (Thermo Scientific, Breman, Germany) operated in data-dependent acquisition mode using in-house optimized parameters with 120 minutes of chromatographic separation used for each fraction. Briefly, the acquisition used three FAIMS compensation voltages (-40, -55, -70) operated under standard resolution with an ion transfer tube temperature of 300℃ with a carrier gas flow rate of 4.6 L/min. Survey scans were performed at a resolution of 120,000 from 400-1,600 *m/z,* with a 250% AGC target and ion injection time set to auto. Fragmentation for peptide identification and reporter tag quantification were performed synchronously (10 per duty cycle per compensation voltage) with the fragmentation spectra generated in the ion trap using CID with turbo scan rate; MS3 reporter ion measurements were performed in the orbitrap with a resolution of 50,000. Dynamic exclusion was applied for 60 seconds across all compensation voltages with only one charge state per precursor selected for fragmentation. In addition, the use of real-time searching was performed with a close out of five peptides per protein within each injected fraction; this utilised the human SwissProt proteome with 1% false discovery filtering applied during acquisition.

2.7.4 Mass spectrometric data analysis

The raw data files were analyzed using Proteome Discoverer (v2.5.0.400, Thermo Scientific) to obtain protein identifications and their respective reporter ion intensities using in-house standard parameters with sequest. The human SwissProt proteome containing only reviewed sequences (accessed in June 2020) was used for protein identification at a 1% false discovery rate alongside a common contaminants database. Reporter ion quantifiers used a unique plus razor with analysis centred on protein groups for shared peptide sequences using all peptides for abundances determination, value output and quantitative values were corrected against stable isotope label impurities according to the manufacturer's values as per the lot number.

2.7.5 Proteomic data bioinformatic analysis

Before normalisation, proteomic data was filtered for high-confidence protein observations. In addition, contaminants, proteins that have been only identified by a single peptide and proteins not identified/quantified consistently across the experiment were removed. The remaining missing values were imputed using the missing-not-at-random (MNAR) method, assuming the missingness was due to low expression for such proteins, which are then normalised using the variance-stabilising-normalisation (VSN) method. Both imputations and VSN were conducted by the *DEP* package (PMID:29446774). Batch effects were corrected using the internal referencing scaling (IRS) method (PMID: 2832585) by the use of reference channels. The *limma* package (https://academic.oup.com/nar/article/43/7/e47/2414268) from R Bioconductor was used to generate a list of differentially expressed proteins for each pair-wise comparison. A cutoff of the adjusted p-value of 0.05 (Benjamini-Hochberg method) along with a fold change of 1.0 has been applied to determine significantly regulated proteins in the different pairwise comparisons.

2.8 Statistical analyses

Statistical analyses were carried out using GraphPad Prism 9 (GraphPad Software La Jolla California, USA) unless otherwise stated. P-values of less than 0.05 were considered statistically significant. The sample size in each experiment or analysis is indicated as 'n'. Unless stated otherwise, all results represent the mean ± standard error of the mean (SEM). Detailed information relating to statistical analyses is outlined in the corresponding sections of each chapter and in figure legends.

3 Chapter 3: The genetic landscape of Infantile Epileptic Spasms Syndrome with brain malformations

3.1 Introduction

This chapter seeks to address the first aim of this thesis: To investigate the genetic landscape of IESS in a cohort of paediatric patients with IESS and MCD.

This chapter utilises genomic sequencing technologies on gDNA extracted from surgically resected brain tissue to identify the genetic cause of IESS due to MCD and demonstrates the utility of deep sequencing tools on brain tissue. The findings in this study provide insights into the genetic spectrum of IESS due to MCD.

The work described in this chapter is currently being prepared as a manuscript for submission to Neurology. The final publication may differ slightly in inclusions or detail to what is presented in this chapter. Additional clinical analyses are underway in relation to this work. However, these additional data are beyond the scope of this thesis and will not be discussed in detail.

3.1.1 Population studies and prevalence

With a frequency of approximately 1:3000 live births each year, IESS is the most common severe epilepsy of infancy [38, 193, 194]. While myriad brain pathologies are reported, focal structural abnormalities, including MCD, particularly TSC and FCD [195-197], are the most frequently described causes of IESS. Over the period of 2011-2020, the epilepsy research unit at the Murdoch Children's Research Institute (MCRI) undertook a population-based study in Victoria examining birth rates and aetiologies of infants with IESS. This study identified that ~50% of IESS cases are caused by a structural brain change, with FCDs accounting for up to 16% of IESS cases born in Victoria during this period [198]. This is significantly higher than the previously reported incidence of 0.5-2.3% of infants [199, 200]. This study suggests that focal

malformations are underdiagnosed in infants with IESS; a likely contributing factor being that mild, focal MCD may be challenging to detect on brain imaging performed during infancy.

Seizure and developmental outcomes in individuals with IESS are typically poor; however, prompt and effective treatment of spasms can result in improved developmental and seizure outcomes [39, 201]. Anti-seizure medications (ASM) are generally ineffective for permanently treating IESS due to MCD. Combined prednisolone and vigabatrin treatment has shown to be somewhat successful at curtailing seizures in IESS, however individuals with IESS and structural malformations (including MCD) have a significantly higher relapse rate than individuals with IESS due to other causes [202]. As such, these patients often require epilepsy surgery for seizure control and to arrest the progressive developmental effects of drug resistant spasms [39]. A deeper understanding of the underlying genetic basis of IESS, and most effective methods of diagnosis, is required to make the best use of currently available targeted treatments. This may further guide development of novel therapies and inform future diagnoses.

3.1.2 Reported genetic causes of IESS and MCD

Over 30 genes have been causatively linked to IESS with many more associated with IESS and DEEs more broadly [203, 204]. In cases without MCD or other structural brain abnormalities, the most common reported genetic aetiologies are *STXBP1*, *ARX*, *SCN2A*, *CDKL5*, *SPTAN1* and trisomy 21 [47, 204].

With advances in genetic technologies, somatic mosaicism in the brain is increasingly recognised as a major genetic mechanism underlying focal brain malformations [147]. Pathogenic variants in genes in the PI3K/AKT/mTOR and GATOR1 complex signalling pathways have been identified in most individuals with TSC (*TSC1* and *TSC2* genes) and many individuals with FCD type 2 (FCD II) (predominantly in *AKT3*, *DEPCD5*, *MTOR NPRL2*, *NPRL3* and *PIK3CA* genes). These pathologies are collectively termed 'mTORopathies' [107, 205].

The largest study of an epilepsy surgery cohort of FCD to date identified the genetic aetiology in 54% of individuals [93]. The MRI and histopathologic diagnosis rather than the epilepsy syndrome has been the focus of the majority of studies on the genetic basis of focal brain malformations to date [206] [207]. The importance of each of these genes and the predominance of somatic variations in individuals with IESS due to brain malformations have not yet been investigated, despite the fact that infantile epileptic spasms are often reported in individuals with MCD caused by these genes [14, 97, 208]. This is due to a lack of understanding of MCD as a cause of IESS, in addition to a lack of access to brain tissue from affected individuals. Understanding the genetic basis of IESS due to focal brain malformations may aid in optimising diagnostic pathways to expedite diagnosis of a potentially surgically remediable cause for IESS, inform prognosis, expand understanding of the pathogenesis of IESS, and enable the development of novel therapies, improved surgical approaches or the identification of treatment alternatives to surgery.

3.1.3 Aim

The genetic basis of focal brain malformations causing IESS is reported to comprise both germline and somatic pathogenic variants, but the current understanding is incomplete. This study aimed to investigate the genetic landscape and histopathologic aetiologies of IESS due to focal brain malformations in a cohort of 59 individuals who underwent epilepsy surgery.

3.2 Methods

3.2.1 Surgical IESS study

This study examines the genetic landscape of IESS due to MCD in a cohort of paediatric individuals who required epilepsy surgery. The first major aim of this study was to investigate the genetic landscape of IESS in a cohort of children with IESS and MCD. To accomplish this, deep exome or panel sequencing was performed on resected brain

tissues from a cohort of 59 children with IESS and an MCD collected during surgery through the Royal Children's Hospital epilepsy neurosurgery program.

3.2.2 Cohort composition

The study cohort comprised 59 individuals with IESS and MCD who underwent resective epilepsy surgery at the Royal Children's Hospital (RCH), Melbourne, Australia, as a result of an MCD. Study participants presented with IESS prior to age 18 months and were born in Victoria, Australia, between 2011 and 2021 (Table 3.1). Participants were identified through the RCH Children's Epilepsy Program surgical database, and the databases of the Genetics of Brain Development and Severe Epilepsies of Infancy studies at the Murdoch Children's Research Institute (MCRI). Individuals were excluded if they were known to have an acquired basis for IESS (e.g. perinatal stroke). Clinical diagnoses were based on clinical and imaging findings, and histopathology review.

Of these 59 individuals, 26 (44%) were female and 33 (56%) were male. The 59 probands had a mean age of seizure onset at 5 months (median = 4 months) and mean age at first surgery of 3.4 years (median = 2.4 years). These individuals had initial clinical neuropathological diagnoses of TSC (22/59, 37%), FCD IIA (including six with HME) (15/59, 25%), FCD IIB (2/59, 3%), FCD I (4/59, 7%), non-specific findings with minor architectural and cytologic abnormalities (8/59, 14%), gliosis (3/59, 5%), with no architectural abnormalities detected (NAD) (3/59, 5%), dysembryoplastic neuroepithelial tumour (DNET) (1/59, 2%) and complex Aicardi syndrome-like MCD (1/59, 2%) (AA0068-01) (Table 3.1).

Table 3.1: Cohort and clinical data.

ID	Sex	Neuropathology	Age seizure onset (months)	Surgery age (years)	Surgery extent	Surgery type	Engel class	Testing completed					
								HaloPlex^{HS}	400x ES	Trio ES	Bulk RNA-seq	CMA	other
AA0068-01	Male	Complex MCD	9	8	MLo	R TPO	III		x			x	
AA0214-01	Female	DNET	4	1.6, 3	Lo, Lo	R F, R P	IV		x				
AA0221-01	Female	FCD I	1	0.8	MLo	L TPO	III		x			x	
AA0327-01	Male	FCD I	11	1.5	MLo	L TPO	I		x	x	x	x	
AA0148-01	Male	FCD IA	16	3,4,8	Lo, MLo, H	L F, L F/T/I, L H	I		x			x	
AA0228-01	Female	FCD IA	14	1.8, 4	ML, L	L TPO, L F	I		x			x	
AA0084-01	Male	FCD IIA	3	2.7, 7	Lo, Lo	L F, L F	I		x				
AA0161-01	Male	FCD IIA	6	3, 7	Lo, MLo	R T, R TPO	UK		x			x	
AA0171-01	Male	FCD IIA	2	3	Lo	R T	I	x					
AA0190-01	Female	FCD IIA	2	6	Lo	R F	I	x	x				
AA0193-01	Female	FCD IIA	4	2	Lo	L F	I	x	x	x	x		
AA0235-01	Male	FCD IIA	5	0.5, 2	MLo	L I/P, L I/P	IV		x				
AA0265-01	Male	FCD IIA	4	4	Lo	R F	I		x	x	x	x	
AA0294-01	Female	FCD IIA	6	3	H	L H	I	x	x	x	x	x	
AA1485-01	Male	FCD IIA	4	7	Lo	R F	I		x			x	
AA0096-01	Male	FCD IIA/HME	0	0.2, 0.6, 5, 9	MLo, MLo, Lo, MLo	L TPO, L TPO, L P, L TPO	I	x					
AA0097-01	Male	FCD IIA/HME	0	1.4, 3	H, Lo	R H (incomplete), R F	I	x					
AA0175-01	Male	FCD IIA/HME	0	0.3, 6	MLo, H	R TPO/R F, R H	II	x	x				
AA0192-01	Female	FCD IIA/HME	1	1.2, 4	MLo, Lo	R F-P-I, R F	II		x	x	x		
AA0342-01	Male	FCD IIA/HME	0	0.2	H	R H	I		x				x
AA0404-01	Male	FCD IIA/HME	0	1	H	R H	I		x			x	
AA0199-01	Female	FCD IIB	15	2, 3	Lo, Lo	R F, R F	I	x	x			x	
AA0936-01	Female	FCD IIB	7	1	Lo	R F	I		x				x
AA0164-01	Male	Gliosis	4	2, 4	Lo, H	R F, R H	I		x			x	
AA0937-01	Female	Gliosis	6	3	MLo, Lo	R TPO	IV		x	x	x	x	

AA2569-01	Male	Gliosis	1	3, 4	H, Lo	R H, R F	IV		x	x	x	x	
AA0203-01	Male	NAD	9	3	MLo	L TPO	I		x	x	x	x	
AA0236-01	Female	NAD	5	1.3	MLo	L TPO	II		x			x	
AA1391-01	Male	NAD	7	2	MLo	R TPO	I		x				
AA0087-01	Female	Non-specific	2	7	Lo	L T	IV		x			x	
AA0215-01	Female	Non-specific	6	2, 4	Lo, Lo	L F, L F-C	IV		x	x	x		
AA0239-01	Male	Non-specific	0	0.7	H	R H	I		x			x	
AA2537-01	Male	Non-specific	11	1.6	H	R H	I		x	x	x	x	
AA2561-01	Male	Non-specific	7	3	MLo	R TPO	I		x			x	
AA2575-01	Female	Non-specific	9	2.5	H	L H	IV		x			x	
AA2691-01	Male	Non-specific	0	0.1	H (incomplete)	L H (incomplete)	I		x				
AA0099-01	Female	Non-specific	4	11	Lo	L T	III		x	x	x	x	
AA0017-01	Male	TSC	5	16	Les	L F	I	x					
AA0033-01	Female	TSC	2	13	Les	R TP, L T, R TPO, R F	II	x					
AA0092-01	Male	TSC	6	1.5, 2, 6	Les	R F	II	x					
AA0093-01	Female	TSC	9	2.5, 4.9	Les	R T	II	x					
AA0140-01	Female	TSC	4	3.5	Les	R F	I	x					
AA0142-01	Male	TSC	9	3.8	Les	R FT	IV	x	x	x	x		
AA0143-01	Female	TSC	6	2.8	Les	L C	III	x					x
AA0145-01	Female	TSC	0	1, 1.4, 7	Les	L T	II	x					x
AA0151-01	Female	TSC	2	13	Les	R F	I	x	x				
AA0166-01	Male	TSC	8	4	Les	L F	II	x	x				
AA0178-01	Female	TSC	4	2.8	Les	L C	II	x					
AA0187-01	Male	TSC	1	0.7	Les	L T	III	x	x				
AA0195-01	Male	TSC	4	2.4	Les	R PFC	I	x	x			x	
AA0224-01	Male	TSC	6	2.6	Les	R T	III	x	x				
AA0264-01	Female	TSC	3	16	Les	L PCG	II	x	x				
AA0302-01	Female	TSC	1	2	Les	L P	I	x	x				

AA0313-01	Female	TSC	1	2, 3, 4, 5	Les	L F, R F, R P	IV	x					x	
AA0316-01	Male	TSC	3	3.4	Les	L F	II	x						
AA0329-01	Female	TSC	0	2	Les	R F	I	x						
AA0938-01	Female	TSC	4	0.6, 2, 3	Les	R T, L PO	III	x						
AA1027-01	Male	TSC	6	5	Les	R PO	II	x	x					
AA1408-01	Male	TSC	15	2	Les	L T	I	x						
							Total:	30	41	12	12	24	4	

C, corticectomy; CMA, chromosomal microarray; DNET, dysembryoplastic neuroepithelial tumour; ES, exome sequencing; F, frontal lobe; FCD, focal cortical dysplasia; H, hemispheric disconnection/hemispherectomy; HME, hemimegalencephaly; I, Engel class 1 (seizure free); II, Engel class 2; III, Engel class 3; IV, Engel class 4; I/P, inferior-parietal; L, left; Les, lesionectomy; Lo, lobar; MLo, multi-lobar; NAD, no abnormalities detected; PCG, pre-central gyrus; R, right; T, temporal; TPO, temporo-parieto-occipital; TSC, tuberous sclerosis complex; UK, unknown; x, testing completed. Extended clinical table included in Appendix (Supplementary Table 8.5)

3.2.3 Genetic investigation

Genomic DNA (gDNA) extracted from resected lesional brain tissue of the 59 individuals with IESS and MCD was used for genetic testing (Table 3.1). Sequencing was performed as described in Chapter 2.3.

3.2.4 HaloPlexHS targeted panel sequencing

Twenty-seven individuals with TSC or FCD IIB were deemed highly likely to carry variants in PI3K-AKT-mTOR pathway or GATOR1 complex genes. As such, targeted panel sequencing was performed on brain-derived DNA extracted from these individuals. The targeted panel was a custom-designed HaloPlex^{HS} gene panel using SureDesign software (Agilent, Santa Clara, CA) and consisted of 331 genes associated with MCD, the PI3K-AKT-mTOR pathway or GATOR1 complex. Individuals without a genetic diagnosis following Haloplex^{HS} sequencing also received 400x ES.

3.2.5 Exome sequencing

Thirty-two DNA samples were analysed using 400x ES to identify the genetic cause of IESS. Of these, 19 DNA samples were prepared using the SureSelect Clinical Research Exome V2 (CREv2) and 13 DNA samples were prepared using the Twist Exome 2.0. Sequencing was performed on an Illumina NovaSeq6000 sequencer (2 × 150 bp) at a 400x target read depth.

3.2.6 Copy number variation analysis

Individuals without a genetic diagnosis following ES single nucleotide variation (SNV) analysis were interrogated for copy number variation (CNV) in ES data. Analysis of potential copy number variants (CNV) within this cohort was carried out on trio ES data using CxGO, as described in Chapter 2.3.3.

3.2.7 Trio ES and bulk RNA sequencing

Twelve individuals remained unsolved following singleton ES or HaloPlex^{HS} analysis. Trio exome analysis was undertaken using blood-derived gDNA from unaffected

parents (n=24) of these 12 individuals and brain-derived cDNA was submitted for bulk RNA sequencing as described in Chapters 2.3.4 and 2.3.5, respectively.

3.2.8 Histopathological assessment

Brain tissue histopathology was reviewed in individuals whose genetic causes were unclear or those whose clinical histopathologic diagnosis did not match their genetic findings. Individuals who were found to have genetic and clinical histopathologic findings consistent with a well-established phenotype-genotype connection were not subjected to histopathologic review (i.e. individuals with FCD II due to a variant in an mTOR pathway gene or TSC due to variants in *TSC1* or *TSC2*).

Histopathologic reassessment was completed on newly stained research brain tissue samples. In cases where more than one specimen was available, the specimen nearest in origin to the region of most prominent abnormality on MRI was selected. Immunohistochemical staining was performed as described in Chapter 2.5. FFPE tissue slides were incubated overnight with anti-Phospho-S6 Ribosomal Protein (Ser235/236) (pS6) (Cell Signaling, #4858P), anti-NeuN rabbit monoclonal antibody clone 27-4 (NeuN) (Millipore, #MABN140) and anti-neurofilament SMI-311R antibody (SMI-311) (BioLegend, #837801) antibodies. To identify nuclear components in the tissues, hematoxylin staining was completed. Neuropathologic reassessment was carried out by a blinded clinical neuropathologist and findings were classified according to the ILAE Diagnostic Methods Commission classification [94, 95].

3.3 Results

Following genetic analysis, 49 putative pathogenic variants that met the criteria for ACMG class IV or V pathogenicity were identified in 47/59 (80%) individuals from the study cohort (Table 3.2). Candidate VUS were identified in a further three individuals (5%) (Supplementary Table 8.4; Supplementary Figure 8.1) and nine individuals remained without any strong candidate variants (15%).

3.3.1 Germline genetic findings

Germline variants were identified in 26/59 (42%) probands, in *TSC2* (n=19), *DEPDC5* (n=2), *CDKL5* (n=2), *COL4A1* and 6p25.2 deletion (n=1), *NPRL3* (n=1), *FGFR1* (n=1) and *TSC1* (n=1) (Figure 3.1, Table 3.2). This includes one individual with a blended phenotype as a result of a *COL4A1* variant and a 6p25.2 deletion that covered regions of *TUBB2A* and *TUBB2B* (AA0221-01). This includes one individual with a two-hit *DEPDC5* diagnosis, with a germline *DEPDC5* variant and a somatic *DEPDC5* variant (AA0171-01, previously published [209]).

3.3.2 Somatic genetic variants in the surgical cohort

Amongst these findings, pathogenic or likely pathogenic somatic variants in brain tissue accounted for 21 diagnoses. Somatic variants were identified in brain tissue in 21/59 (36%) individuals in this cohort in *SLC35A2* (n=9), *PIK3CA* (n=3), *MTOR* (n=2), *TSC2* (n=2), *AKT3* (n=2), *OFD1* (n=1), *TSC1* (n=1) and *DEPDC5* (n=1). ddPCR validation was carried out using custom-designed allele-specific Taqman probes in order to investigate variant allele frequencies (VAF) in additional brain tissue samples and peripheral tissues, where available. In brain tissue, somatic VAF ranged from 0.95-41%. No somatic mosaic variants were able to be detected in blood via ddPCR.

Table 3.2: Genetic variants identified in resected brain tissue from the study cohort of individuals with IESS and MCD.

Individual ID	Sex	Neuropathology	Gene	Transcript	Variant	AA Change	VAF	Inheritance	Solve method	Discordant
AA0097-01	M	FCD IIA/HME	*AKT3*	NM_181690.2	c.49G>A	E17K	4.90%	Mosaic	HaloPlex^{HS}	N
AA0404-01	M	FCD IIA/HME	*AKT3*	NM_005465	c.49G>A	E17K	5.10%	Mosaic	ES (seqr)	N
AA0087-01	F	Non-specific	*CDKL5*	NM_001323289.2	c.2706delG	Q902Hfs*25	-	Germline	ES (seqr)	N
AA2575-01	F	Non-specific	*CDKL5*		ChrX:18425546-18425744x1	*CDKL5* deletion	-	Germline	ES (CxGo)	N
AA0221-01	F	FCD I	*COL4A1*	NM_001303110.2	c.634G>A	G212S	-	Germline	ES (seqr)	N
			TUBB2A/TUBB2B		Chr6:3155598-3224833x1	6p25.2 deletion	-	Germline	CMA	N
AA0171-01	M	FCD IIA	*DEPDC5*	NM_001242896.3	c.2390delA	Q797Rfs*18	-	Germline	HaloPlex^{HS}	N
			DEPDC5	NM_001242896.3	c.3994C>T	R1332*	3.90%	Mosaic	ES (seqr)	N
AA0235-01	M	FCD IIA	*DEPDC5*	NM_001242896.3	c.2527C>T	R843*	-	Germline	ES (seqr)	N
AA0214-01	F	DNET	*FGFR1*		Chr8:38413607-38419737x3	*FGFR1* duplication	-	Germline	ES (CxGo)	N
AA0199-01	F	FCD IIB	*MTOR*	NM_001386500.1	c.7273_7275dup	P2425dup	4.10%	Mosaic	ES (seqr)	N
AA0342-01	M	FCD IIA/HME	*MTOR*	NM_001386500.1	c.4348_4359del	Y1450_L1453del	7.40%	Mosaic	ES (seqr)	N
AA0175-01	M	FCD IIA/HME	*NPRL3*	NM_001077350.3	c.1375_1376dupAC	S460Pfs*20	-	Germline	ES (seqr)	N
AA0068-01	M	Complex MCD, Aicardi-like	*OFD1*	NM_001330210.2	c.2192_2216dup	P740Qfs*18	21.00%	Mosaic	ES (seqr)	N
AA0096-01	M	FCD IIA/HME	*PIK3CA*	NM_006218.4	c.3140A>G	H1047R	29.00%	Mosaic	HaloPlex^{HS}	N
AA0239-01	M	Non-specific	*PIK3CA*	NM_006218.4	c.1624G>A	E542K	20.00%	Mosaic	ES (seqr)	Y (FCD IIA)
AA2691-01	M	Non-specific	*PIK3CA*	NM_006218.4	c.1624G>A	E542K	13.00%	Mosaic	ES (seqr)	Y (FCD IIA)
AA0084-01	M	FCD IIA	*SLC35A2*	NM_005660.3	c.136C>T	Q46*	20.00%	Mosaic	ES (seqr)	Y (other)
AA2561-01	M	Non-specific	*SLC35A2*	NM_005660.3	c.206C>T	T69I	1.00%	Mosaic	ES (seqr)	Y (other)
AA0148-01	M	FCD IA	*SLC35A2*	NM_005660.3	c.359_360delTC	L120Hfs*7	41.00%	Mosaic	ES (seqr)	Y (other)
AA1391-01	M	NAD	*SLC35A2*	NM_005660.3	c.502C>T	Q168*	7.60%	Mosaic	ES (seqr)	Y (other)
AA0161-01	M	FCD IIA	*SLC35A2*	NM_005660.3	c.511T>C	S171P	35.00%	Mosaic	ES (seqr)	Y (other)
AA0236-01	F	NAD	*SLC35A2*	NM_005660.3	c.547C>T	Q183*	1.80%	Mosaic	ES (seqr)	Y (other)
AA0164-01	M	Gliosis	*SLC35A2*	NM_005660.3	c.553C>T	Q185*	1.90%	Mosaic	ES (seqr)	Y (other)

AA0228-01	F	FCD IA	*SLC35A2*	NM_005660.3	c.626_628delCCT	S209del	5.60%	Mosaic	ES (seqr)	Y (other)
AA0190-01	F	FCD IIA	*SLC35A2*	NM_005660.3	c.665_667delAGA	K222del	20.00%	Mosaic	ES (seqr)	Y (other)
AA0224-01	M	TSC	*TSC1*	NM_000368.5	c.1498C>T	R500*	-	Germline	HaloPlex^{HS}	N
AA0151-01	F	TSC	*TSC1*	NM_000368.5	c.1886delA	K629Rfs*24	20.70%	Mosaic	HaloPlex^{HS}	N
AA1485-01	M	FCD IIA	*TSC2*	NM_000548.5	c.336+1G>A	SS variant	7.50%	Mosaic	HaloPlex^{HS}	N
AA0936-01	F	FCD IIB	*TSC2*	NM_000548.5	c.2071delC	R691Afs*7	13.00%	Mosaic	ES (seqr)	Y (TSC)
AA0017-01	M	TSC	*TSC2*	NM_000548.5	c.4527_4529delCTT	F1510del	-	Germline	HaloPlex^{HS}	N
AA0033-01	F	TSC	*TSC2*	NM_000548.5	c.2251C>T	R751*	-	Germline	HaloPlex^{HS}	N
AA0092-01	M	TSC	*TSC2*	NM_000548.5	c.4736G>T	G1579V	-	Germline	HaloPlex^{HS}	N
AA0093-01	F	TSC	*TSC2*	NM_000548.5	c.4991dupG	Q1665Pfs*41	-	Germline	HaloPlex^{HS}	N
AA0140-01	F	TSC	*TSC2*	NM_000548.5	c.5238_5255del	H1746_R1751del	-	Germline	HaloPlex^{HS}	N
AA0143-01	F	TSC	*TSC2*	NM_000548.5	c.4648+1 G>T	SS variant	-	Germline	HaloPlex^{HS}	N
AA0145-01	F	TSC	*TSC2*	NM_000548.5	c.5140C>T	Q1714*	-	Germline	HaloPlex^{HS}	N
AA0178-01	F	TSC	*TSC2*	NM_000548.5	c.3281C>A	S1094*	-	Germline	HaloPlex^{HS}	N
AA0187-01	M	TSC	*TSC2*	NM_000548.5	c.3685_3686insA	E1230Gfs*4	-	Germline	HaloPlex^{HS}	N
AA0264-01	F	TSC	*TSC2*	NM_000548.5	c.5024C>T	P1675L	-	Germline	HaloPlex^{HS}	N
AA0302-01	F	TSC	*TSC2*	NM_000548.5	c.1840-1G>A	SS variant	-	Germline	HaloPlex^{HS}	N
AA0316-01	M	TSC	*TSC2*	NM_000548.5	c.724_725insG	T242Sfs*96	-	Germline	HaloPlex^{HS}	N
AA0329-01	F	TSC	*TSC2*	NM_000548.5	c.5227C>T	R1743T	-	Germline	HaloPlex^{HS}	N
AA0938-01	F	TSC	*TSC2*	NM_000548.5	c.4573C>T	Q1525*	-	Germline	HaloPlex^{HS}	N
AA1408-01	M	TSC	*TSC2*	NM_000548.5	c.3574delC	Q1192Rfs*18	-	Germline	HaloPlex^{HS}	N
AA0166-01	M	TSC	*TSC2*		Chr16:2062961-2086850x1	16p13.3 deletion	-	Germline	CMA	N
AA0313-01	F	TSC	*TSC2*		Chr16:2090008-2100467x1	16p13.3 deletion	-	Germline	CMA	N
AA0195-01	M	TSC	*TSC2*		Chr16:2134591-2136455x1	16p13.3 deletion	-	Germline	CMA	N
AA1027-01	M	TSC	*TSC2*		Chr16:2046127-2050488x1	16p13.3 deletion	-	Germline	ES (CxGo)	N
AA0327-01	M	FCD I	…							N

AA0265-01	M	FCD IIA	…		Y (FCD I)
AA0294-01	F	FCD IIA	…		Y (FCD I)
AA0193-01	F	FCD IIA	…		N
AA0192-01	F	FCD IIA/HME	…		N
AA0937-01	F	Gliosis	…		N
AA2569-01	M	Gliosis	…		N
AA0203-01	M	NAD	…	-	N
AA0099-01	F	Non-specific	…		N
AA0215-01	F	Non-specific	…		N
AA2537-01	M	Non-specific	…		N
AA0142-01	M	TSC	…		N

AA, amino acid; ACMG, American College of Medical genetics and Genomics classification; CMA, chromosomal microarray; DNET, dysembryoplastic neuroepithelial tumour; ES, exome sequencing; F, female; FCD, focal cortical dysplasia; HME, hemimegalencephaly; M, male; MCD, malformations of cortical development; N, no; NAD, no abnormality detected; VAF, variant allele frequency; Y, yes; '…' indicates no genetic diagnosis made.

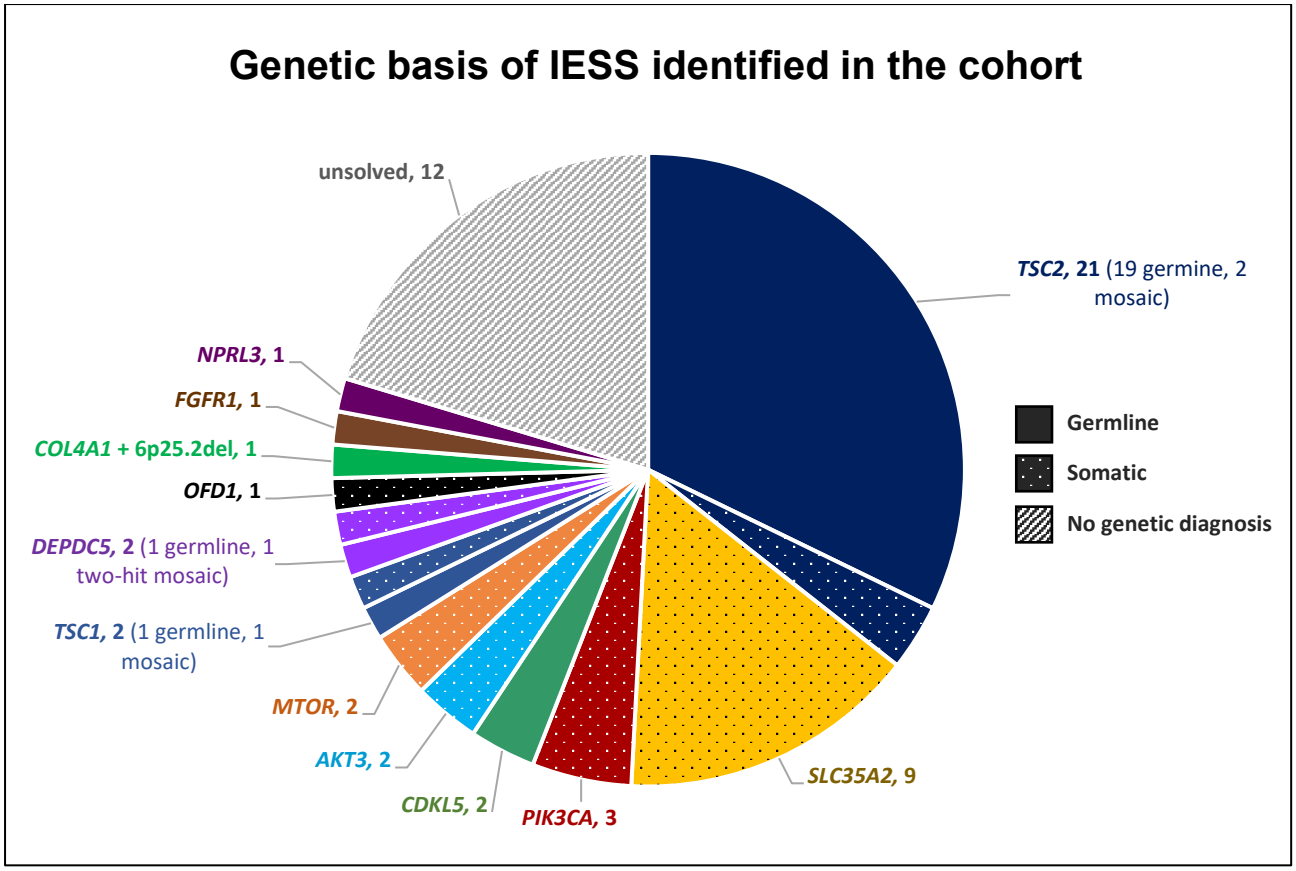


Figure 3.1: Genetic landscape of IESS with MCD in this cohort.
Solid regions indicate germline cases, dotted regions indicate somatic cases, striped region indicates cases without a genetic diagnosis identified. n=59 total.

3.3.3 Histopathological findings

Following histopathological reassessment, nine individuals with brain somatic variants in *SLC35A2* were found to be discordant with initial histopathological diagnoses and these nine individuals were labelled as "other" and will be discussed in detail in Chapters 4 and 5 of this thesis.

In addition to the nine individuals with brain somatic *SLC35A2* who were found to be discordant, five other individuals were found to have neuropathological diagnoses discordant with initial diagnosis following genetic or revised neuropathologic analysis. One individual with FCD IIB (AA0936-01) was reclassified as TSC after secondary lesions were identified on MRI and histopathological reassessment. This classification was consistent with the mosaic *TSC2* variant (c.2071delC; p.Arg691Alafs*7, 13% VAF) identified in this individual (Table 3.2). Two individuals (AA0265-01, AA0294-01) were

classified from FCD IIA to FCD I based on histopathological analysis and two individuals (AA0239-01, AA2691-01) were reclassified from non-specific to FCD IIA based on histopathological staining and the presence of *PIK3CA* variants in brain tissue (Figure 3.2).

Following identification of discordant diagnoses and subsequent reclassification, the cohort consisted of TSC (23/59, 39%), FCD IIA (including six with HME) (12/59, 20%), "other" (9/59, 15%), non-specific findings with minor architectural and cytologic abnormalities (5/59, 8%), FCD I (4/59, 7%), gliosis (2/59, 3%), FCD IIB (1/59, 2%), with no architectural abnormalities detected (NAD) (1/59, 2%), dysembryoplastic neuroepithelial tumour (DNET) (1/59, 2%) and complex Aicardi syndrome-like MCD (1/59, 2%) (Table 3.2).

All individuals with an initial classification of FCD IIA and HME in the broader cohort remained unchanged following an integrated diagnostic reclassification.

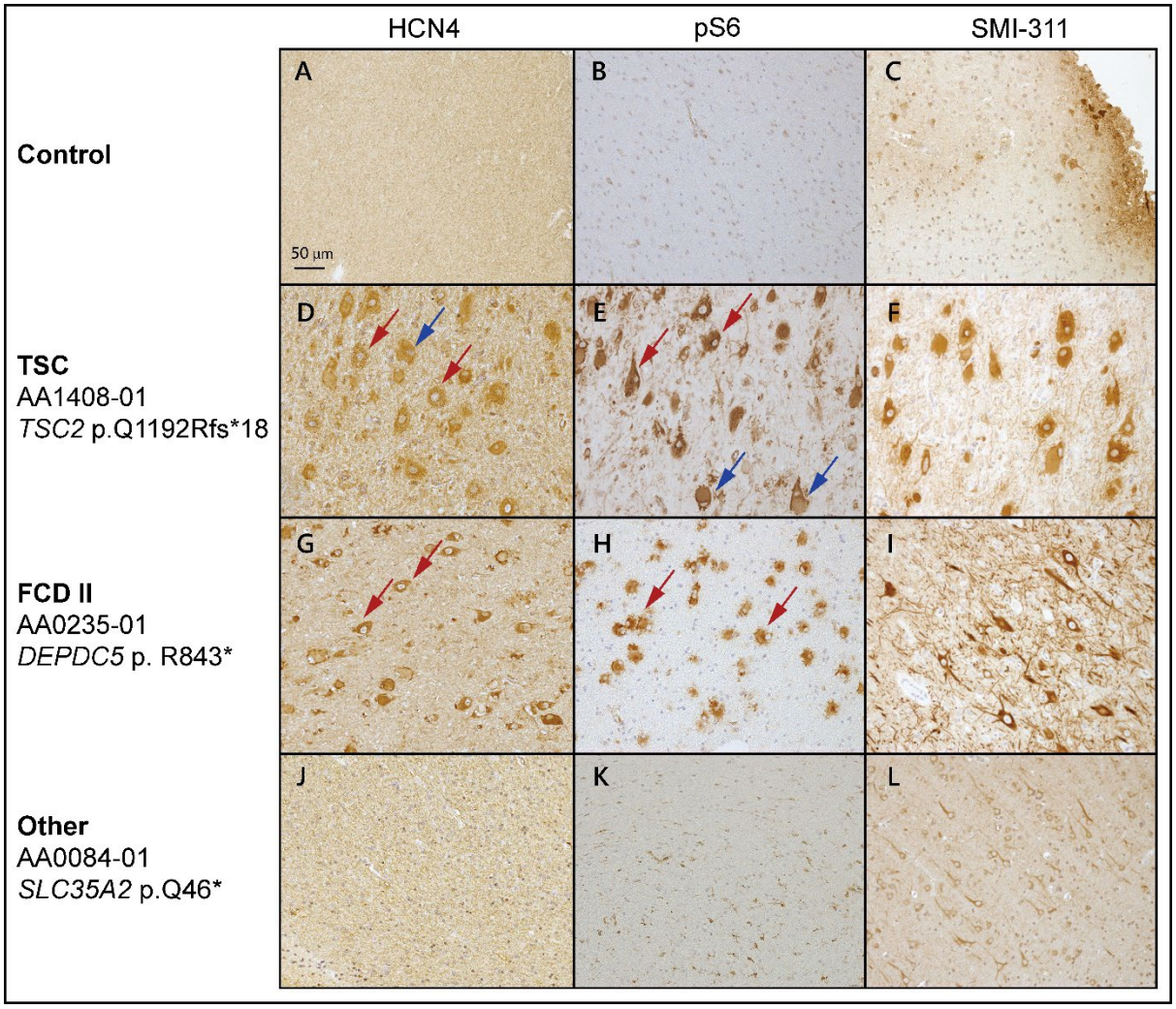


Figure 3.2: HCN4, pS6 and SMI-311 immunohistochemistry staining.
Staining performed on 3µm FFPE tissue sections tissue from Control (A-C), TSC (AA1408-01) (D-F), FCD II (AA0235-01) (G-I) and "Other" (AA0084-01) (J-L) at 20x magnification. Red arrows indicate examples of dysmorphic neurons and blue arrows indicate examples of balloon cells. Dysmorphic neurons and balloons cells are revealed by HCN4 and pS6 staining. FCD, focal cortical dysplasia; TSC, tuberous sclerosis complex.

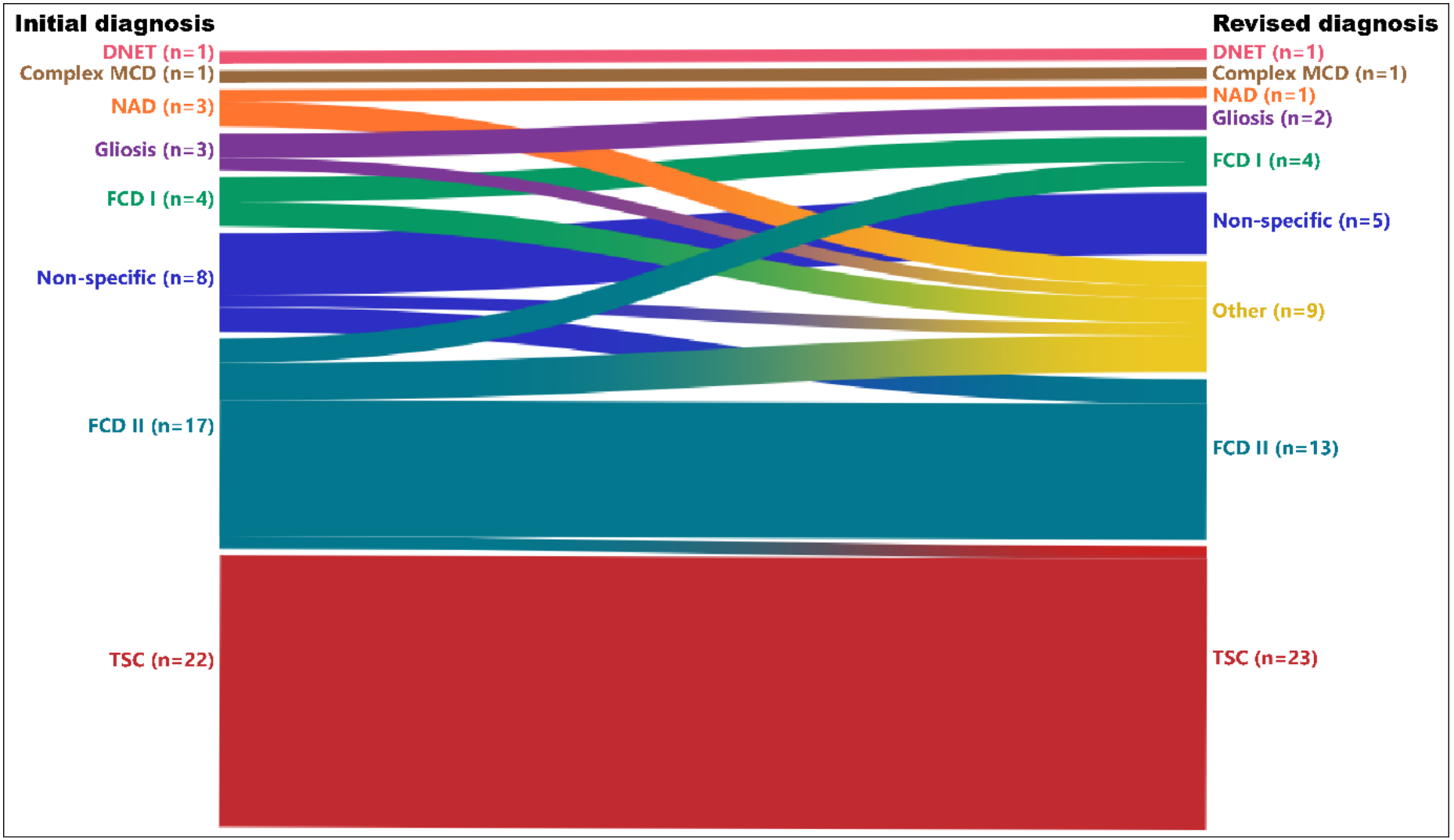


Figure 3.3: Alluvial diagram showing changes between initial diagnosis and revised diagnosis of IESS with MCD cohort.
Complex MCD, complex Aicardi syndrome-like malformation; DNET, dysembryoplastic neuroepithelial tumour; FCD, focal cortical dysplasia; HME, hemimegalencephaly; NAD, no abnormalities detected; TSC, tuberous sclerosis complex.

3.3.4 Genetic diagnosis rate by neuropathology

The genetic basis was identified in 22/23 (96%) individuals with TSC (being *TSC1*, n=2; *TSC2*, n=20), 11/13 (85%) individuals with FCD II (*PIK3CA*, n=3; *AKT3*, n=2; *MTOR*, n=2; *DEPDC5*, n=2; *TSC2*, n=1; *NPRL3*, n=1), 3/12 (53%) individuals with FCD I/non-specific findings/NAD/gliosis (*CDKL5*, n=2; *COL4A1* & 6p25.2del, n=1), 1/1 (100%) with a complex MCD (*OFD1*), 1/1 (100%) with DNET (*FGFR1*) and 9/9 (100%) "other" (*SLC35A2*, n=9) (Figure 3.4).

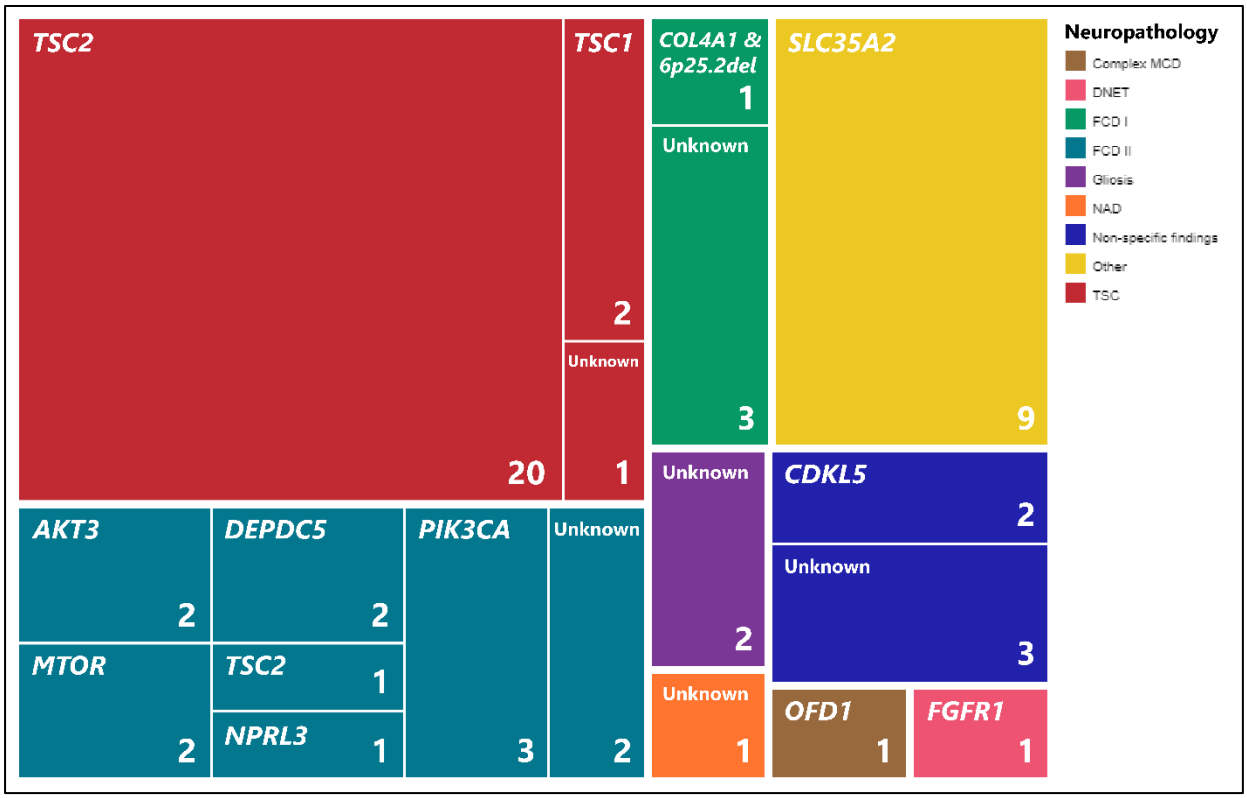


Figure 3.4: Tree-map of genetic causes by neuropathology.
Number in bottom right corner of each box indicates number of individuals with that genetic diagnosis. Unknown denotes no genetic diagnosis identified. Complex MCD, complex Aicardi syndrome-like malformation; DNET, dysembryoplastic neuroepithelial tumour; FCD, focal cortical dysplasia; HME, hemimegalencephaly; NAD, no abnormalities detected; TSC, tuberous sclerosis complex.

3.3.5 Diagnostic rate by analysis method

Of the pathogenic genetic variants identified in this cohort, 21 were identified via HaloPlex^{HS} analysis (in 21/30 (70%) individuals that received HaloPlex^{HS} sequencing), 21 via seqr analysis of singleton ES data and three by CxGo analysis of trio ES data (Figure 3.5) (in 24/41 (59%) individuals that received ES), and four diagnoses were

made via clinical chromosomal microarray (CMA) analysis on peripheral blood samples in parallel to this study (Table 3.2). These numbers include the two-hit *DEPDC5* diagnosis with the germline variant identified via HaloPlex^{HS} and the somatic variant identified via ES seqr analysis; and the individual (AA0221-01) with a variant in *COL4A1* identified through ES seqr analysis and a 6p25.2 deletion identified via clinical CMA where both variants contribute to the phenotype. No further genetic diagnoses were made using seqr analysis of trio ES data or Degust analysis of bulk RNA-seq data, however several candidate VUS were subsequently excluded by segregation data obtained from trio ES data (Figure 3.6).

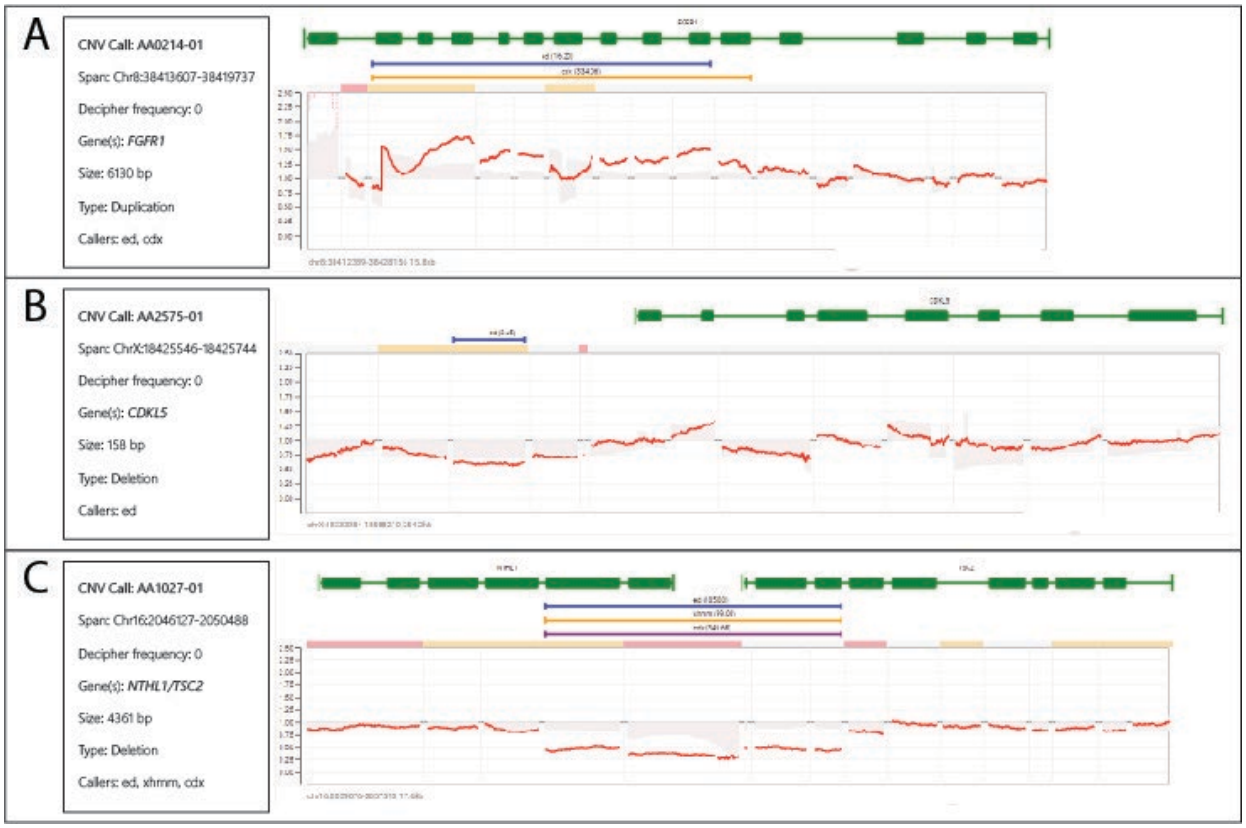


Figure 3.5: CxGO analysis plots of the three copy number variants (CNVs) detected in ES data in individuals AA0214-01 (Panel A), AA2575-01 (B) and AA1027-01 (C).
Variant details are included inside box on left of each plot. Green box and line diagrams indicate genes. Variant caller (ed, cdx and xhmm) detected CNV regions indicated with blue, yellow and purple bars. Population average indicated in grey on plot. ES read coverage at each site indicated by red line. CNVs <1Mbp in size were validated by subsequent high resolution chromosomal microarray (CMA).

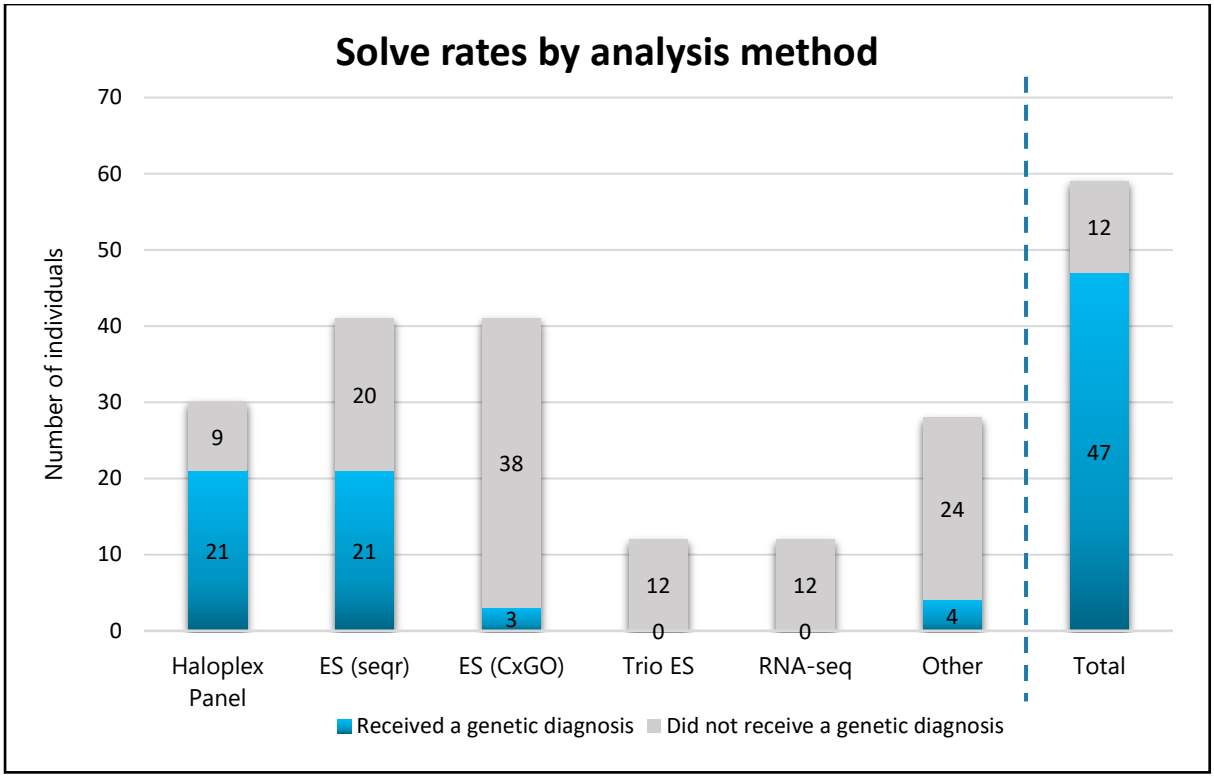


Figure 3.6: Stacked bar plot of solve rates by analysis method in the cohort.
Y-axis indicates number of individuals. Analysis methods are listed along the X-axis. Blue area indicates individuals who received a genetic diagnosis. Grey area indicates individuals who received that analysis method but did not receive a genetic diagnosis. ES, exome sequencing; other includes four clinical gene screens and 24 chromosomal microarrays (CMA).

3.4 Discussion

3.4.1 The genetic landscape of IESS with MCD is accurately determined by deep genetic sequencing on brain tissue

Germline and somatic variants comprise the genetic landscape of IESS due to MCD. This study presents the most exhaustive analysis of surgical brain samples from individuals with IESS due to MCD to date. The genetic diagnosis was identified in almost 80% (47/59) of individuals in this cohort. The diagnostic rate in individuals with IESS with MCD in this study is significantly higher than the average clinical diagnostic rate identified in early onset epilepsies of close to 40% [210-212]. Similarly, the diagnostic rate is high compared to the majority of genetic studies of MCD which have reported diagnostic rates ranging from 9-53% [93, 99, 213-215]. This high diagnostic yield is likely a result of deep sequencing technologies and the use of brain tissue available for in this study.

The most efficient methods of diagnosis were HaloPlex^{HS} panel sequencing and seqr analysis of 400x singleton ES, resulting in 21 (70% diagnostic rate) and 21 (51% diagnostic rate) genetic diagnoses respectively.

Only three individuals were genetically solved by CxGO analysis of ES data. The diagnostic rate from this method was low, likely due to the four other individuals with CNVs already having been identified prior via clinical CMA. Neither trio exome analysis nor bulk RNA sequencing analysis resulted in any further genetic diagnoses. This may be due to the fact that the majority of genetic diagnoses in this cohort were already identified and that the remaining individuals have acquired causes of IESS with MCD; or the possibility that these individuals have been affected by repeat expansions or deep intronic genetic variants and these variants would not have been captured in this study.

It is likely that the diagnoses made using HaloPlex^{HS} data would have also been captured by 400x ES analysis. The samples selected for HaloPlex^{HS} analysis were predominantly selected as they were expected to have mTOR pathway variants. The seqr gene lists included all of the genes identified in the HaloPlex^{HS} cohort and the 400x ES captured these genes at a similar or deeper level of coverage. In this cohort, 8/9 (89%) individuals received a genetic diagnosis from ES analysis following a negative result on HaloPlex^{HS} analysis. This is likely due to the broader gene list able to be screened using seqr analysis of ES data, in addition to CNV screening using CxGO. 400x ES analysis on surgical brain tissue is a valuable tool for investigating the genetic causes of epilepsy and brain malformation disorders.

3.4.2 Mosaicism as an under-recognised contributor to IESS with MCD

In this cohort, somatic variants constituted less than half (44.6%, 21/47) of all genetic diagnoses identified and only 1/21 (4%) diagnoses in the TSC group. However, somatic variants accounted for the majority (20/28, 71.4%) of findings in the non-TSC cohort.

Current clinical practice typically uses gDNA derived from peripheral tissues such as blood or saliva for genetic testing at moderate depths of coverage (<100x depth ES),

significantly lower than those utilised in this study. As a result, low level somatic mosaicism would be difficult to detect, particularly if mosaicism is restricted to the brain [147, 216]. Pre-surgical diagnostic yield in somatic mosaic cases may be improved by the use of CSF liquid biopsies or less invasive procedures for detecting brain mosaic variants brought about by future technological advances [217, 218]. It is likely that brain somatic mosaicism, particularly in FCD I and FCD II cases, is an under-recognised aetiology of IESS based on the observation of 21 cases with somatic variations in lesional brain tissue in this cohort. Furthermore, given the underdiagnosis of small focal malformations in the IESS population, and the challenges of genetic screening for brain mosaic variants in routine clinical practice, it is possible that somatic mosaicism may represent a more prevalent genetic basis of IESS than previously understood [198]. The majority of clinical testing utilises peripheral tissues or clinically accessible tissues (CATs) which are not appropriate tissues for genetic screening or detection of brain mosaic variants.

The most striking finding from this cohort is the disproportionate number of individuals with IESS and MCD who carry somatic *SLC35A2* variants in the brain. These nine individuals account for 15% (9/59) of the overall cohort and almost half (43%, 9/21) of the mosaic cases. The incidence of *SLC35A2* brain somatic mosaicism warrants further discussion and elevates *SLC35A2* as a prime gene candidate for further interrogation. *SLC35A2* encodes solute carrier family 35 member A2, a nucleotide-sugar transporter active in the golgi apparatus. In-depth interrogation and discussion of this gene are detailed in Chapters 4 and 5 of this thesis.

3.4.3 Limited success of mosaic variant detection in peripheral tissues

Only one of the 21 brain somatic variants that was present in brain at 13% (AA0936-01, *TSC2* p.R691Afs*7) was also able to be detected in blood via clinical testing. This individual was reclassified as TSC following histopathologic, genetic and clinical review. TSC involves multisystem tuber development, and as such, variant presence outside of the brain would be concordant with this diagnosis. The other 20 brain mosaic variants

were not able to be detected in blood or saliva (where available). Clinical diagnostic methods routinely use peripheral tissues for primary genetic screening, as these are the most readily available CATs and do not require invasive collection methods. Clinical diagnosis rates are limited by the use of peripheral tissues, as these are unlikely to capture mosaic genetic variants. Individuals with mosaic variants are therefore likely to remain without a genetic diagnosis following clinical testing. Appropriate samples (i.e. brain tissue-derived gDNA) are required to increase diagnostic rates in clinical testing.

3.4.4 pS6 and HCN4 identify mTORopathies

Ribosomal Protein S6 Phosphorylation (pS6) and HCN4 immunochemistry showed increased immunoreactivity in abnormally large and cytomegalic neurons in FCD II and TSC tissues, indicating DNs and BCs. pS6 is commonly used as a marker of mTOR pathway activation [219]. More recently, Hseih et al. identified upregulated HCN4 in *RHEB*-related FCM as a key driver of mTOR-dependent epilepsy in mice [220]. HCN4 was shown to be present exclusively in FCM neurons and it was found that increasing cyclic adenosine monophosphate (cAMP) in these neurons resulted in repetitive firing, while disruption of HCN4 activity resulted in a rescue of the seizure phenotype in mice. The authors further demonstrated abnormal HCN4 steady-state levels in DNs in surgically resected human brain tissue from individuals with TSC and FCD IIB, without reporting the genetic basis in these individuals, and proposed HCN4 as a candidate therapeutic target. HCN4 was similarly observed as upregulated in DNs and BCs in brain tissue from individuals with FCD II and TSC in this cohort. Neither pS6 nor HCN4 immunoreactivity was observed in the three individuals with initial histopathology of FCD IIA with somatic *SLC35A2* variants (AA0084-01, AA0161-01 and AA0190-01). In total, somatic *SLC35A2* variants were identified in brain tissue from nine individuals with initial histopathological diagnoses of FCD I, FCDIIA and gliosis. The genetic diagnosis in these individuals was found to be discordant with initial histopathological diagnosis in these individuals and they were removed from these histopathological classifications (Table 3.2). The nine individuals with somatic *SLC35A2* variants in brain will be discussed in depth in Chapters 4 and 5.

Both pS6 and HCN4 act as markers for mTOR pathway activation in this study and accurately distinguish FCD II and TSC tissues from other neuropathologies (Figure 3.2).

3.4.5 mTORopathies and non-mTOR FCMs

The genetic basis of IESS and MCD was identified in 22/23 (96%) individuals with TSC, as either *TSC1* (n=2) or *TSC2* (n=20), and 11/13 (85%) individuals with FCD II (*PIK3CA*, n=3; *AKT3*, n=2; *MTOR*, n=2; *DEPDC5*, n=2; *TSC2*, n=1; *NPRL3*, n=1) (Table 3.2).

All genetic variants identified in the TSC and FCD II cases were in genes involved in the PI3K-AKT3-mTOR pathway. These results are consistent with the well-established genetic aetiology of mTORopathies [93, 205, 221]. Consistent with previous studies, *TSC2* was identified as the more common genetic cause of TSC in this cohort, compared to *TSC1* [222]. This data suggests that IESS due to TSC, FCD IIA and FCD IIB are similar in genetic aetiology to TSC, FCD IIA and FCD IIB causing broader epilepsy disorders, and that the genetic basis of IESS with mTORopathies is not unique to IESS, with no predominance of any specific mTOR pathway genes. Children presenting with IESS with TSC or FCD II have the same genetic causes as older children presenting with TSC and FCD II with later seizure onset. It is not currently known why some individuals with these MCD develop IS and others do not, despite sharing the same genetic basis. This discrepancy in onset may be due to variant type or somatic variant location but the sample size in this study is too small to adequately answer this question.

3.4.6 Diagnostic yield is low in FCD I and non-mTOR-related MCD

The genetic diagnostic yield was comparatively low in individuals with FCD I, non-specific findings, NAD and more subtle neuropathologic characteristics. Of the 13 cases in this group, a genetic diagnosis was only achieved in four (33%) individuals. Two individuals (AA0087-01 and AA2575-01) carried variants in *CDKL5*, which will be discussed further in Chapter 3.4.7; one individual (AA0221-01) carried a variant in *COL4A1* and a deletion at 6p25.2 which is discussed further in Chapter 3.4.9; and one individual with DNET (AA0214-01) was identified to carry a CNV in *FGFR1*, a known gene associated with DNET [223]. These diagnoses consisted of only germline

pathogenic variants that could have been detected in peripheral tissues prior to surgery. Importantly, the bulk of these patients experienced poor surgical outcomes, with over 60% (8/13) experiencing continued seizures after surgery or only very brief periods of seizure freedom (Supplementary Table 8.5).

The lack of a genetic diagnosis in many individuals with FCD I, non-specific malformations or NAD in this study may be due to these individuals having non-genetic causes. It is understood that some epilepsies result from environmental factors or acquired causes [224]. It has been hypothesised that epilepsies due to brain malformations including FCD I and non-specific malformations may also have an acquired basis such as perinatal stroke, traumatic brain injury or vascular lesions [225]. If these cases are due to non-genetic causes, surgical intervention may not be the ideal treatment. It is also possible that individuals without a genetic diagnosis in this study may have somatic mosaic genetic causes that are present at a VAF that is too low to be detected in the brain region used for testing [150, 217]. Alternatively, IESS with MCD may have resulted from polygenic causes, deep intronic variants, repeat expansions or variation in a gene yet to be associated with MCD and epilepsy in the individuals without a genetic diagnosis following analysis in this study.

3.4.7 *CDKL5* and the case for pre-operative genetic screening

Two individuals with non-specific neuropathological findings carried germline variants in *CDKL5* that were not detected by clinical testing prior to surgery. These individuals received CMA and single gene screening on gDNA derived from peripheral tissues prior to surgery but they did not receive exome sequencing. *CDKL5* encodes cyclin-dependent kinase-like 5, a phosphorylated member of the Ser/Thr protein kinase family and disruption of this gene is typically associated with CDKL5 deficiency disorder (OMIM: 300203), an infantile epilepsy syndrome that is not commonly considered to cause focal brain malformations [226].

Individual AA0087-01 carried a germline *de novo* *CDKL5* variant (c.2706delG, p.Gln902Hisfs*25), which was identified via seqr exome analysis of gDNA derived from

brain tissue, following surgery. This individual had drug resistant focal and tonic seizures which did not improve after surgery – a left temporal lobectomy at 8 years of age – and they continue to have severe ID at last review (Table 3.1). This individual had received limited genetic testing for metabolic and genetic disorders prior to surgery, including Angelman syndrome and *MECP2* genetic panels. These tests were unsuccessful at identifying the genetic cause in this individual.

A second individual (AA2575-01) with a *CDKL5* variant was identified during CNV analysis using CxGO. This individual presented with multiple daily seizures, loss of vision and developmental delay. They had a left hemispherotomy at 2.5 years of age to resect epileptogenic tissue and have since been diagnosed with Lennox Gastaut syndrome and hemispatial neglect, following persistent seizures. Clinical genetic testing prior to surgery consisted of a standard microarray test (Illumina Infinium GSA-24 v1.0, 0.20Mb resolution) on blood, which did not detect any abnormalities. The deletion, detected via CxGo and subsequently validated using a high-resolution array (Illumina Infinium GDACyto-8 v1.0), encompasses the first exon of *CDKL5*, a non-coding exon. The deletion is less than 1kb in size (estimated 158bp) and was therefore below the resolution and not detected via standard CMA. Historically, deletions only in the 5′ UTR of *CDKL5* have been difficult to interpret, however the deletion identified in this individual removes the entire exon and was classified as pathogenic [226, 227]. This diagnosis could have been achieved using the same methods used in this study on peripheral tissues prior to surgery and would have informed surgical decisions. This finding would likely have excluded the patient as surgical candidate and avoided the additional disability of having a hemispherotomy.

Though the phenotypes of both individuals were not entirely typical of *CDKL5* deficiency disorder, the *CDKL5* diagnosis does explain the ongoing seizures, dyskinesias and delayed development [227]. Insertion of a vagus nerve stimulator was eventually decided upon for both individuals following their respective surgeries, in an

effort to minimise seizures. This treatment has proven to be somewhat effective at reducing seizures.

Some of the genetic findings in this cohort would not have been identified through pre-surgical clinical genetic testing. Somatic variants, for example, are seldom detected or reported on by clinical testing. One contributing factor is that mosaic sequences are discarded during the curation process due to extremely stringent filtering used for clinical testing. Clinical filtering aims to ascertain only certain results and not potential artefacts, particularly as these findings often inform potential reproductive counselling. As a result, many low read count variants (<20% VAF), real mosaic or artefact, are filtered out [228]. Similarly, CNVs are difficult to detect on ES data as they often appear as low read depth regions and most forms of CMA are limited by probe number. Higher density CMA raise cost issues for diagnostic application.

However, this would not explain the single nucleotide germline variants like the *CDLK5* variants identified in individuals AA0087-01 and AA2575-01. All 26 of the germline variants identified in this cohort could have been detected in peripheral tissues prior to surgery. A change in treatment or clinical management would have occurred for some of these patients prior to surgery. In particular, the two individuals with *CDKL5* variants (AA0087-01 and AA2575-01) underwent major neurosurgery prior to a genetic diagnosis but would likely have been identified as poor surgical candidates based on their *CDKL5* diagnosis. Earlier identification of these variants could have informed clinical decisions regarding extent and necessity of surgery in these cases. This demonstrates the importance of pre-operative genetic screening in IESS with MCD. These findings are consistent with recent studies on the value and utility of pre-surgical genetic testing [229, 230]. This presents implications for surgical candidates more broadly. Genetic testing should be performed prior to surgery in infant onset epilepsies, as genetic knowledge can inform clinical decision making and may avoid unnecessary surgeries.

3.4.8 Blended phenotypes in complex diseases

A de novo missense variant was identified in *COL4A1* in individual AA0221-01 (*COL4A1* c.634G>A; p.Gly212Ser). *COL4A1* encodes a type IV collagen alpha protein and is a well-established gene for leukoencephalophy and microangiopathy [231]. Pathogenic genetic missense variants in *COL4A1* tend to be glycine substitutions, as this variant is. Furthermore, the same variant (p.Gly212Ser) was previously reported as de novo in two other patients. The first was a 9-year-old male with white matter hyperintensities that were later reclassified as leukoencephalopathy and the other a newborn male with porencephaly, a brain malformation disorder characterised by intracerebral cystic cavities [232, 233]. Individual AA0221-01 has clinical features of leukoencephalopathy that match the phenotype of these published *COL4A1* cases.

Interestingly, this individual (AA0221-01) also had a finding via chromosomal microarray (CMA) of a microdeletion at [GRCh37] 6p25.2(3155598_3224833)x1, encompassing the promoter and first exon of *TUBB2A*, in addition to the final exon and 3'UTR of *TUBB2B*. Tubulin Beta 2A Class IIa (*TUBB2A*) and Tubulin Beta 2B Class IIb (*TUBB2B*) have been established in MCD pathogenesis [234] but did not adequately explain the phenotype of this individual on their own. Only when the two genetic variants are considered in tandem does the genetic diagnosis account for all aspects of the phenotype of this individual. This individual had imaging features consistent with both of these genetic disorders. Therefore, the conclusion of this research is that both the *COL4A1* missense variant and the 6p25.2 deletion are contributing to a blended phenotype in this individual.

MCD are complex diseases and due to the unique nature of each clinical case, it is important to consider alternatives to the monogenic model. Post-zygotic mutation, two hit variants and multi-gene blended phenotypes have been observed in this cohort. It is possible that polygenic risk scores (PRS), repeat expansions, intronic changes or other genetic modes of inheritance are also at play in malformative IESS.

3.4.9 Mosaic *OFD1* as a novel cause of IESS in males

Another finding of note in this cohort is a male with an Aicardi syndrome-like complex brain malformation with a mosaic *OFD1* variant (individual AA0068-01). This individual presented with Aicardi syndrome-like features and was found to have a mosaic heterozygous frameshift *OFD1* variant (c.2192_2216dup; p.Pro740GlnfsTer18) at a VAF of 21% in lesional brain tissue. Aicardi syndrome is a MCD disorder characterised by infantile spasms, agenesis of the corpus callosum and an eye disorder known as chorioretinal lacunae. Aicardi syndrome is predominantly observed in females and it has been historically extremely difficult to identify a genetic basis for Aicardi syndrome. To date, sporadic cases of Aicardi syndrome have been identified in XXY males ut no XY males have been reported [235]. *OFD1* is an X-linked gene that has been show to escape X-chromosome inactivation (XCI).

OFD1 encodes the OFD1 Centriole And Centriolar Satellite protein and has been previously associated with X-linked dominant (XLD) Orofaciodigital syndrome 1 (OMIM: 311200), X-linked recessive (XLR) retinitis pigmentosa 23 (OMIM: 300424), XLR Simpson-Golabi-Behmel syndrome (OMIM: 300209) and XLR Joubert syndrome (OMIM: 300804) [236]. Pathogenic variants are generally considered embryonic lethal for hemizygous males, with only a few exceptions [237]. Aberrant accumulation of OFD1 by impaired autophagy has been found to affect neuronal ciliogenesis underlying cortical dyslamination [238].

The identification of mosaic *OFD1* variant in this individual is consistent with the phenotype observed, despite a lack of classical craniofacial features associated with Orofaciodigital syndrome 1, as it has been established that *OFD1* phenotype in males differs to females [237]. Following classification and clinical reporting of this variant as likely pathogenic and the indication of a new phenotype for *OFD1* mosaic males, a subsequent report was published of an *OFD1* hemizygous male with a similar phenotype. In 2022, Gangaram et al. reported a male with a hemizygous variant in *OFD1*, who did not share the same phenotype as previously reported heterozygous

females or other hemizygous males with *OFD1* variants [239]. Gangaram et al. expanded the clinical phenotype to include features consistent with Aicardi syndrome, including thinning of the corpus callosum and a molar tooth sign on MRI. Similarly, Gangaram et al. reported PMG in this case, which was also observed in the *OFD1* mosaic male in this study. Venkatesan et al. subsequently reported further overlapping phenotypes between Aicardi syndrome and Orofaciodigital syndrome 1 in one female individual with an *OFD1* variant, diagnosed to have Orofaciodigital syndrome 1, and three females with Aicardi syndrome, providing further strength to the phenotypic similarity between *OFD1* variation and Aicardi syndrome [240].

3.4.10 Health economic burden and genetic counselling implications

As evidenced by the large proportion of cases with germline variants in this cohort, pre-operative genetic screening is a valuable tool and could have informed clinical care decisions prior to surgery. This raises the question of whether the health benefits of pre-operative testing offset the economic burden of its application.

The current cost of a singleton clinical exome through the Victorian Clinical Genetics Services (VCGS) is AU$1800 (February 2023). This includes variant curation analysis with a focused gene list such as the ones utilised in this study or the PanelApp Genetic Epilepsy panels. The Genetic Epilepsy PanelApp gene list currently evaluates 820 genes associated with epilepsy. Notably, 12 of 14 genes identified in this cohort are captured by the Genetic Epilepsy PanelApp gene list. Only *OFD1* and *FGFR1* are not included in this gene list.

Likewise, seven individuals in this cohort carried germline CNVs in *TSC2* (n=4), *TUBB2A/TUBB2B* (n=1), *CDKL5* (n=1) and *FGFR1* (n=1). These were either identified in resected brain tissue by CMA on brain tissue or CxGo analysis and were subsequently validated on CMA, demonstrating that all seven variants were detectable using CMA. CMA is the recommended initial test for many neurodevelopmental disorders including intellectual disability (ID), developmental delay (DD) and autism spectrum disorders (ASD) [241, 242]. Medicare is the federal health insurance scheme in

Australia. It is publicly funded and universal. The Medicare Benefits Schedule is a list of healthcare and services subsidised by Medicare in Australia. Standard clinical CMA is often subsidised by the Medicare Benefits Schedule (item 73292) in Australia, making it a more readily accessible method, with some restrictions regarding inclusions and candidates. However, standard CMA has a relatively low resolution and would not have captured all of the CNVs identified in this cohort.

In specific cases of suspected Mendelian disorders in childhood, a Medicare Benefits Schedule subsidy for exome or genome sequencing may be available for singleton (item #73358) or trio (item #73359) analysis. Moreover, reanalysis of existing exome data costs only AU$425 and the information gained from this analysis can drastically impact the clinical management and treatment decisions [243]. Additional genes and novel variants are being associated with epilepsy every week [21]. Genes or variants of unknown significance may be reclassified as causative based on new evidence or knowledge gained in subsequent years.

It is true that in some cases, the turnaround time of exome analysis would negate its value, particularly if early seizure onset and severity necessitates urgent resective surgery. Equally, the 21 cases in this cohort with somatic mosaicism would not have been detected by way of pre-operative CMA or exome analysis on peripheral tissues. However, pre-operative genetic testing on peripheral tissues could have resulted in genetic diagnosis in up to 44% (26/59) individuals.

It should be considered that individuals with TSC represented the majority of germline cases in this cohort and it is unlikely that a genetic diagnosis would have significantly altered surgical decisions in individuals with TSC. However, even if the yield is low in FCD and other MCD, it is still valuable to perform genetic testing prior to surgery, as informed clinical decisions may result in improved developmental outcomes in addition to the financial benefits. Similarly, personalised medical treatments may be devised, or other medical treatments may be considered such as vagus nerve

stimulation as an alternative to resective or disconnective neurosurgery or ASMs, as in the case of *CDKL5*.

Another financial consideration is surgical cost burden. The mean cost of conducting epilepsy surgery in Australia was reported as AU$61,417 per person in 2019 [34]. An analysis of cost-effectiveness of epilepsy surgery at the Royal Children's Hospital in Melbourne, Australia, found that the cost of conducting epilepsy surgery was significantly more expensive than medical treatment with ASMs in individuals with drug-resistant epilepsy [34]. This study also found that seizure control was more commonly achieved by epilepsy surgery than continued treatment with ASMs. Patient outcome and healthcare implications are important factors in epilepsy surgery and may be informed by a genetic diagnosis.

A genetic diagnosis prior to surgery could assist in determining clinical prognosis and inform decisions around reproductive counselling for parents of individuals with IESS and MCD [244]. Aside from the proband, genetic diagnoses can have implications for families, particularly in early onset disorders such as IESS [245]. A genetic diagnosis signifies the end of the diagnostic odyssey for many families and subsequent genetic counselling may improve sense of empowerment and alleviate guilt and anxiety for patients, parents and families [246].

In some of these cases, the benefits of informed clinical care decisions and prognosis that genetic testing would have provided prior to surgery, would have justified the cost of its implementation. Moving forward, pre-surgical genetic testing should be employed in non-urgent candidates for epilepsy surgery. Evidence for the value of pre-surgical is building in IESS [15, 46, 194, 215] and the outcomes of this study add further evidence to support implementation of genetic testing in all individuals presenting with IESS.

3.4.11 Limitations

It is important to note that the data collected in this study cannot be used to accurately estimate the incidence of somatic mosaicism in focal brain malformations that do not

require surgery or are inoperable. Technological advancements that allow for non-invasive, in-depth analyses of brain mosaicism would be required to accurately investigate the prevalence of somatic mosaicism in non-surgical IESS cases. Furthermore, clinical observations may be skewed due to the ascertainment bias for epileptic spasms that has been introduced within the neuropathological phenotype groups, as recruitment was made on the basis of IESS, focal brain malformation diagnosis, and surgery.

It's possible that pathogenic variations in genes that haven't yet been linked to epilepsy or brain abnormalities were overlooked as a result of the methodology choice to utilise a panel-based analysis of exome data. An unbiased singleton ES approach may have assisted in identifying additional genetic causes in this cohort but would likely have resulted in much higher rates of VUS detection. This study was also constrained by the limited coverage of exome sequencing which makes it challenging to detect repeat expansions and intronic variants. The diagnostic yield in the unsolved cases of this cohort may be increased with the use of whole genome sequencing (WGS) [247]. Singleton WGS, however, would likely have resulted in the identification of many more VUS.

It is also possible that there are an even greater number of somatic mosaic cases in this cohort than were detected in this study. As surgical techniques continue to improve, smaller resections are being made [248]. The tissue received may not have been optimal for genetic assessment. In cases with larger disconnections, the resected tissue was not necessarily from the region of most prominent malformation. If the MCD resulted from a somatic variant, it is possible that the region used for genetic testing does not contain the genetic variant in a high enough allele frequency to be detected by current methods [104, 217]. This may be assuaged by utilising more sensitive or targeted technologies such as ddPCR, which have a lower LOD [216]. Though lower LOD would not resolve this issue if resected tissues do not capture variant-carrying cells.

3.5 Conclusions

This study serves to provide an overview of the genetic landscape of IESS associated with MCD and demonstrates the value of multimodal methods of diagnosis in achieving a more precise diagnosis in complex disorders. The genetic causes of IESS with FCD II and TSC is largely similar to other mTORopathies but the underlying genetic causes of IESS due to other FCMs is considerably more diverse. An improved understanding of the genetics of IESS affords a deeper insight into the pathogenesis of this disorder. This may, in turn, enable identification of surgically remediable focal malformations at earlier stages and improve genetic diagnoses and prognoses in clinical practice. This knowledge may guide development of novel medicines for IESS.

Furthermore, as shown by the 26 cases with germline variants in this cohort, and in particular the two cases of germline *CDKL5* variants, pre-operative genetic screening is a valuable tool for informing patient management in clinic and avoiding ineffective surgeries.

Finally, *SLC35A2* accounted for 9/59 (15%) diagnoses in this cohort and, as such, was identified as a gene of interest in IESS that warrants deeper investigation.

4 Chapter 4: The relationship between somatic *SLC35A2* variation and MOGHE

4.1 Introduction

This chapter seeks to address the second aim of this thesis: To identify the molecular hallmarks and phenotypic classifiers in cases with brain mosaic *SLC35A2* variants.

This chapter utilises a multi-disciplinary approach to diagnosis by combining clinical, imaging, histopathologic and genetic data, and demonstrates the importance and clinical utility of employing multimodal methods of analysis in order to achieve accurate diagnoses in complex disorders.

The work described in this chapter, in conjunction with the work described in Chapter 5 of this thesis, is currently being prepared as a manuscript for submission to Genetics in Medicine. The final publication may differ slightly in inclusions or detail to what is presented in this chapter. Additional laboratory-based validation experiments and bioinformatic analyses are underway in relation to this work. However, these additional data are beyond the scope of this thesis and will not be discussed in detail.

4.1.1 Solute Carrier Family 35 Member A2

SLC35A2 (OMIM: 314375) is located on the short arm of chromosome X at Xp11.22-Xp11.23 and encodes Solute Carrier Family 35 Member A2 [249]. The canonical transcript of *SLC35A2* (ENST00000247138.11) contains 5 exons and encodes a 396 amino acid protein that is a member of the nucleotide sugar transporter (NST) family. SLC35A2 is predicted to consist of 10 transmembrane helical domains (positions 3-23, 37-57, 65-85, 97-117, 140-160, 169-189, 200-220, 238-258, 269-289 and 315-335), a disordered region (358-379), an endoplasmic reticulum (ER) retention motif (392-396) and a 307aa nucleotide sugar transporter domain (32-399) (Figure 4.2). SLC35A2 is the major Golgi-localized uridine diphosphate galactose (UDP-galactose) transporter and is required for proper protein and lipid glycosylation [250]. SLC35A2 plays a major role

in N-glycosylation and is located in the Golgi apparatus and endoplasmic reticulum [251].

The Gene Ontology (GO) terms associated with *SLC35A2* include 'UDP-galactose transmembrane transporter activity', 'protein binding', 'carbohydrate transport' and 'galactose metabolic process'.

4.1.2 *SLC35A2* phenotype in the germline state

Germline pathogenic variants in *SLC35A2* (i.e. heterozygous variants in females or hemizygous variants in males) are associated with X-linked dominant congenital disorder of glycosylation (CDG) type IIm (OMIM: 300896), sometimes also called SLC35A2-CDG, early infantile epileptic encephalopathy-22 (EIEE22) or developmental and epileptic encephalopathy 22 (DEE22) [252]. CDGs are a collection of rare genetic conditions resulting from defects in glycan synthesis, processing or transport [253]. Most CDGs involve N-glycosylation defects or disruptions in the formation or transfer of the N-linked glycans lipid-linked oligosaccharide precursor [253, 254].

SLC35A2-CDG is a rare, neurodevelopmental disorder [255]. Symptoms include muscular hypotonia, epilepsy, hypsarrhythmia, feeding difficulties and developmental delay. Phenotypic severity is reported to vary among affected individuals with many individuals presenting with skeletal abnormalities and occasional brain malformations, typically cerebral or cerebellar atrophy [255, 256]. Dysmorphic or coarse facial features have also been reported in some individuals [256, 257]. SLC35A2-CDG is rarely identified via transferrin isoform analysis, the predominant testing method for individuals with CDG, as such, the diagnosis relies on genetic confirmation [254].

4.1.3 Brain-restricted somatic *SLC35A2* variants

More recently, somatic variants in the *SLC35A2* gene identified in brain tissue have been reported in some individuals with focal cortical dysplasia type I (FCD I) and mild malformation of cortical development (mMCD) [113]. A 2018 study by Winawer et al. compared paired blood and brain-derived gDNA from38 individuals with MCD and 18 individuals with non-lesional focal epilepsy (NLFE) [113]. The study identified five

somatic *SLC35A2* variants in lesional brain tissue from three individuals with NFLE who were subsequently classified as FCD I (following neuropathological re-evaluation) and two FCD I cases. Emerging data suggests that somatic variants in *SLC35A2* are a major cause of FCD I and increasing somatic variant load (ranging from 1%-40% of cells within the lesion) has been linked with severity of electrophysiologic and radiographic findings in individuals with FCD I [148]. SLC35A2 has not been associated with the mTOR signalling pathway which is commonly dysregulated in other FCDs [258].

Parallel to this, in 2017, a newly recognized kind of focal malformation, mild malformation of cortical development with oligodendroglial hyperplasia in epilepsy (MOGHE) was first described [112]. A retrospective study led by Bonduelle et al then identified brain somatic *SLC35A2* variation as the major cause of MOGHE in 2021 [114].

Bonduelle et al. described some of the recognisable histopathological features of MOGHE as clusters of oligodendrocytes in the white matter, a blurred grey-white matter boundary, patchy hypomyelination and the presence of heterotopic neurons in the white matter (Figure 4.1). They hypothesized that loss of SLC35A2 transporter function resulted in abnormal glycosylation and produced these cellular markers. However, the exact clinical and phenotypic hallmarks of MOGHE are still to be fully defined.

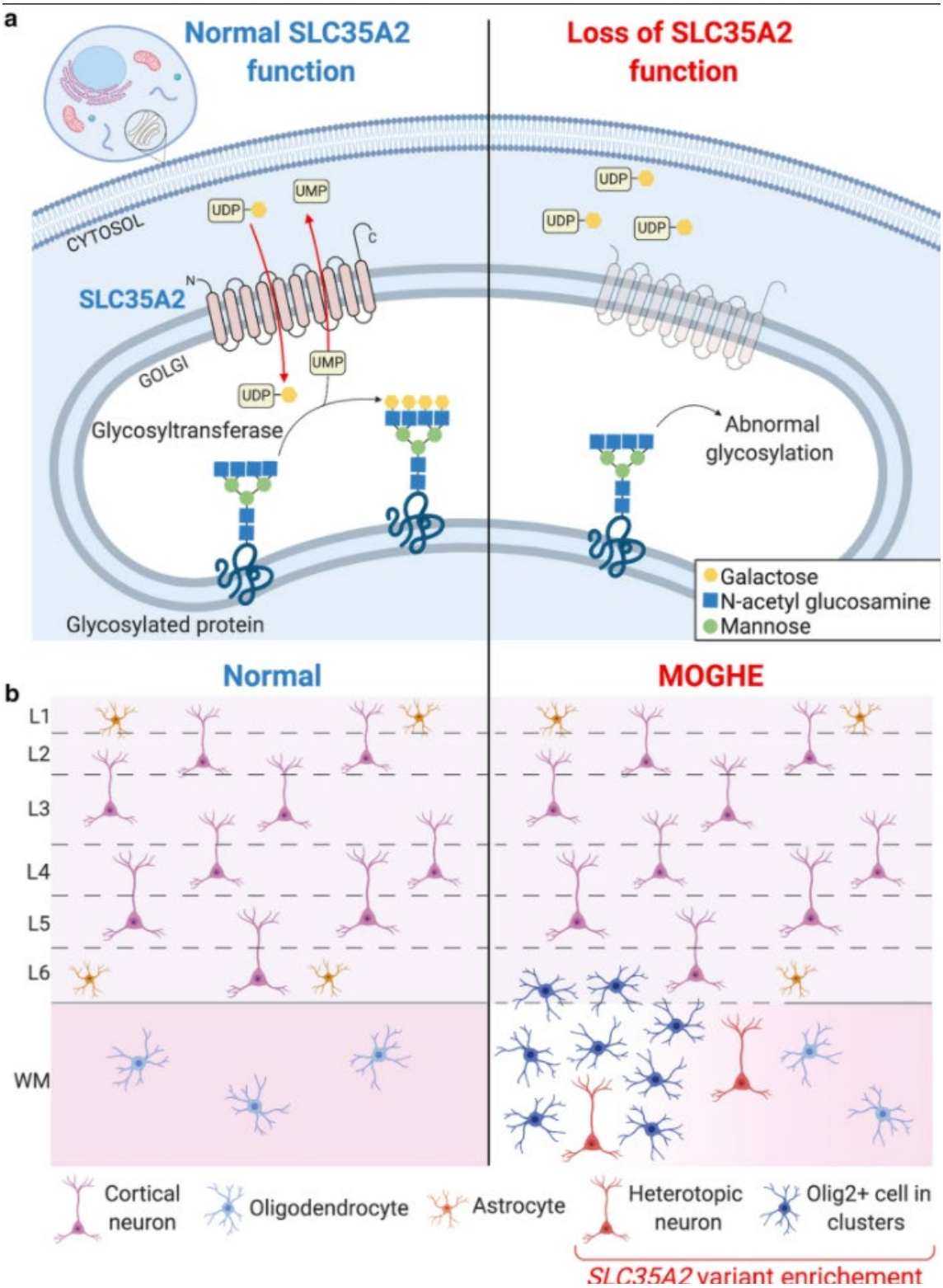


Figure 4.1: Representation of neuropathological findings associated with loss of SLC35A2 function. Figure adapted from Bonduelle et al. 2021 [114].

a) Normal SLC35A2 function in healthy cells (left) and abnormal glycosylation as a result of loss of SLC35A2 function (right). b) Loss of function in SLC35A2 results in clusters of Olig2-positive cells in the white matter, a blurred grey-white matter boundary (represented by the dotted line), patchy hypomyelination and the presence of heterotopic neurons in the white matter. L1-L6, cortical layers 1 to 6; MOGHE, mild malformation of cortical development with oligodendroglial hyperplasia in epilepsy; SLC35A2, Solute Carrier Family 35 Member A2; WM, white matter.

As discussed, pathogenic variants in *SLC35A2* in a germline state, cause SLC35A2-CDG with an X-linked dominant inheritance pattern. It appears that brain somatic variants in *SLC35A2* produce a significantly different phenotype to SLC35A2-CDG but the clinical knowledge and molecular understanding of this pathology are incomplete. The current understanding of this genetic entity and its predominance to cause MOGHE or other MCD is limited, in part due to its relative recent discovery. Similarly, the clinical features of MOGHE are still being characterised and reliable methods of recognising and diagnosing MOGHE in clinical settings are not yet established.

Furthermore, the mechanisms by which somatic *SLC35A2* variants cause focal brain malformations and epilepsy are not clear. The study described in this thesis chapter aimed to expand the understanding of the relationship between *SLC35A2* brain somatic variants and MOGHE.

4.1.4 Aims

This study aimed to determine the molecular hallmarks and phenotypic features in nine individuals with malformative IESS due to brain somatic variants in *SLC35A2*, identified in Chapter 3 of this thesis, and to re-evaluate the histopathological diagnoses for these individuals. This was achieved by reviewing and expanding phenotypic, histopathological, imaging and genetic data for these nine individuals. Additionally, this study aimed to review the aptness of existing diagnostic classification for MOGHE.

4.2 Methods

4.2.1 Cohort and specimen collection

The nine individuals with IESS and MCD with a mosaic *SLC35A2* variant identified in brain tissue were assessed for neuropathological, clinical and imaging features indicative of MOGHE or other MCD subclassifications. Clinical neuropathologists and radiologists were blinded to patient genotype.

The 12 individuals without a genetic basis identified in Chapter 3 (Table 3.2) also underwent neuropathological assessment and clinical re-evaluation, in tandem with the *SLC35A2* cases.

Brain tissue specimens from individuals with definitive clinical diagnoses of FCD I, FCD IIA, FCD IIB and TSC with an established genetic basis identified in Chapter 3 of this thesis were used as disorder-specific positive controls in immunohistochemical analysis. Post-mortem non-malformation brain tissue from the unaffected hemisphere of one age-matched individual with epilepsy and hemispheric PMG due to a heterozygous variant in *ATP1A3* (Control) was used as an unaffected control [259]. This individual did not meet the criteria for IESS, and no structural or histopathological brain abnormalities were identified in the tissue used for analysis.

Individuals with a histopathological, clinical and imaging classification of FCD II and FCD I, NAD, gliosis or non-specific findings comprised the clinical cohorts used for phenotypic comparison.

4.2.2 DNA extraction

Blood and brain-derived gDNA were extracted as previously described in Chapter 2.2.1. gDNA was extracted from all available brain tissue specimens from each individual, including additional lesional, peri-lesional and normal brain tissue samples where available. "Lesional tissues" were classified based on tissue closest to the region of most prominent malformation on MRI, or tissue that had been indicated as lesional from clinical histopathological analysis using H&E staining. "Peri-lesional tissue" was abnormal tissue from regions adjacent to the region of most prominent malformation, and "normal tissue" was tissue from regions distant to malformation and without any histological abnormalities identified via initial clinical pathology testing. Normal brain tissue was collected if it was required to be resected in order to gain surgical access to the lesion.

Cell-free DNA (cfDNA) was extracted from cerebrospinal fluid (CSF) collected during neurosurgery where possible. cfDNA was centrifuged at 1000 × g at 4°C for 10 min

before extraction using the QIAamp Circulating Nucleic Acid Kit (Qiagen, #55114) according to the manufacturer's instructions. cfDNA was stored at -80°C prior to use.

4.2.3 Variant allele analysis, validation and quantification

Variant pathogenicity was evaluated with *in silico* tools. Genomic build GRC38/hg38 was used for alignment and canonical transcript NM_0005660.3 was used for analyses unless otherwise specified.

Variant-specific validation and absolute quantification in additional tissues was performed using droplet digital polymerase chain reaction (ddPCR). Custom TaqMan SNP Genotyping Assays (Thermo Fisher Scientific, #433139) with sequence-specific primers and probes were used for variant detection as described in Chapter 2.4 of this thesis. DNA from all available specimens were tested for the pathogenic variant in each individual, including gDNA from all available brain tissue specimens collected, blood-derived gDNA and cfDNA extracted from CSF. Probe sequences are provided in Supplementary Table 8.3.

VAF levels were compared between tissue type groups using a one-sided Wilcoxon sign-ranked test. VAF quantification was compared between 400x ES data analysis prediction and ddPCR absolute quantification using a paired t-test, two sample for means, two-tailed.

4.2.4 Molecular graphics and protein modelling

Molecular graphics and protein mutagenesis simulation figures were modelled in PyMOL (The PyMOL Molecular Graphics System, Version 2.5.4, Schrödinger, LLC) using AlphaFold SLC35A2 model (ID: P78381). The interaction prediction region was limited to five angstroms from mutated residue.

4.2.5 Immunochemistry analysis

Immunohistochemical staining was completed on 3 µm FFPE tissue sections as described in Chapter 2.5. A neuropathologist, blinded to the clinical histopathology report and genetic findings, reviewed the stained specimens of all individuals, and

classified the findings according to the ILAE Diagnostic Methods Commission classification [94, 95, 111]. A formal histopathologic diagnosis of MOGHE was assigned if clusters of oligodendroglial cells were identified at a density of >2000/mm^2 at the grey-white matter junction, an excess (>20mm^2) of heterotopic neurons was observed in the white matter and a blurring of the grey-white matter was observed, as previously described [111]. A diagnosis of 'possible MOGHE' was assigned if sufficient heterotopic neurons were observed, clusters of oligodendroglial cells were identified and an increased density of oligodendrocytes was observed but cell counts did not reach the threshold of >2000/mm^2 at the grey-white matter junction. A possible MOGHE diagnosis required absence of evidence suggesting another clear classification (e.g. pS6 immunoreactivity and presence of balloon cells indicating FCD IIB).

Primary antibodies utilised were anti-oligodendrocyte transcription factor 2 (Olig2) (EP112) rabbit monoclonal antibody (CellMarque, #387R-16), anti-Phospho-S6 Ribosomal Protein (Ser235/236) (pS6) rabbit monoclonal antibody (Cell Signaling, #4858P), anti-neurofilament SMI-311R antibody (BioLegend, #837801) and anti-MAP2 (2a+2b) mouse monoclonal antibody (Sigma, #M1406).

4.2.6 Imaging review

A neuroradiologist blinded to genotype reviewed the MRI scans from all individuals in the cohort (Chapter 3), excluding those with a definitive TSC diagnosis. The resulting cohort of 37 individuals included the MRI scans of all nine individuals with brain mosaic *SLC35A2* variants. Individuals with TSC were not re-examined due to the stringent classification of this disorder resulting in a clear diagnosis. All cases reviewed were screened for broad MCD features, in addition to the two major magnetic resonance imaging (MRI) features that have been established in published MOGHE cases; i) an increased laminar T2 and fluid attenuated inversion recovery (FLAIR) signal at the corticomedullary junction, and ii) reduced corticomedullary differentiation because of increased signal of the adjacent white matter [174]. Both features were described in a

recent paper by Hartleib *et al.* and classified as either MOGHE subtype 1 or MOGHE subtype 2, respectively [260].

4.2.7 Clinical and surgical outcome data

Participants were recruited via the Royal Children's Hospital (RCH) Epilepsy Surgery program as described in Chapter 2.1. Clinical and surgical details for each individual were reviewed and provided by neurologists at the Royal Children's Hospital. Additional data was collated from participant records via the Murdoch Children's Research Institute (MCRI) Research Electronic Data Capture (REDCap) system and the RCH Electronic Medical Records (EMR).

4.2.8 Integrated diagnosis reclassifications

Individuals were evaluated for MOGHE based on existing and new clinical, neuroimaging, histopathological, surgical and genetic data. Details of strength of MOGHE association used for classification is detailed in Table 4.1. The criteria for a MOGHE reclassification included a genetic diagnosis of a brain mosaic *SLC35A2* variant, neuroimaging consistent with either MOGHE subtype 1 or subtype 2, and histopathological and clinical features that were definite of MOGHE or suggestive of MOGHE and inconsistent with current or alternate possible classifications (e.g. FCD I, FCD IIA, FCD IIB, DNET, TSC, etc). In order to be classified 'MOGHE', a genetic diagnosis of *SLC35A2* was required and other common MCD subtypes needed to have been definitively ruled out. If any individuals without a genetic diagnosis shared two or more of the other MOGHE features (i.e. clinical, neuroimaging, histopathological or surgical), they were considered to be 'possible MOGHE'.

Table 4.1: Integrated diagnosis reclassification requirements for MOGHE.

	Not MOGHE ←		Possible MOGHE →	MOGHE
	Against	**Supporting**	**Moderate**	**Strong**
Histopathological data	pS6 staining indicative of mTORopathy No clusters of Olig2 positive cells Absence of heterotopic neurons in white matter	Blurred grey-white matter boundary	>20 Heterotopic neurons in white matter Absence of pS6 staining indicative of mTORopathy	>2000 Olig2 cells/mm2 in white matter Clusters of Olig2 positive cells in white matter
Genetic data	Pathogenic genetic variant identified in a different gene		Mosaic *SLC35A2* variant identified in brain	Recurrent mosaic *SLC35A2* variant identified in brain
Imaging data	MRI phenotype consistent with alternate, established MCD		MOGHE subtype I or II identified on MRI	
Clinical data		Epileptic spasms/IESS		

4.3 Results

4.3.1 Variant analysis and in silico predictions of pathogenicity

SLC35A2 is predicted to be intolerant to mutation. *SLC35A2* has a Residual Variation Intolerance Score (RVIS) of -0.16 (41.64%), suggesting that *SLC35A2* is intolerant to mutation, and a gnomAD probability of loss-of-function (LoF) intolerance (pLI) score of 0.718 (o/e = 0.12), suggesting that *SLC35A2* is intolerant to protein truncating variants. All nine somatic genetic variants identified in *SLC35A2* in this cohort were predicted damaging by multiple *in silico* tools (Table 4.2).

SLC35A2 encodes a UDP-galactose transporter and as such, contains 10 transmembrane helical (TMEM) domains and one large nucleotide sugar transporter (NST) domain. The nine somatic variants are all located within the main NST domain and a slight trend towards variants on the boundaries of the TMEM domains was observed, with 7/9 variants located within TMEM domains. Four of the variants identified in this cohort (p.Thr69Ile, p.L120Hfs*7, p.Gln168Ter and p.Gln185Ter) are

recurrent variants that have previously been reported in individuals with MOGHE [114]. In terms of variants, cytosine (C) to thymine (T) substitutions were most frequent, and glutamine to termination amino acid changes were observed more commonly. This diversity and spread of mosaic variant locations is consistent with the previously published mosaic cases, that had been classified as MOGHE (Figure 4.2).

Protein mutagenesis modelling using PyMOL indicated that all nine variants resulted in altered protein products (Figure 4.3). The two deletion mutations (p.Ser209del and p.Lys222del) and two missense changes (p.Thr69Ile and p.Ser171Pro) interrupt α-helices in the final protein product. The other five variants are predicted to result in a premature stop codon (PTC).

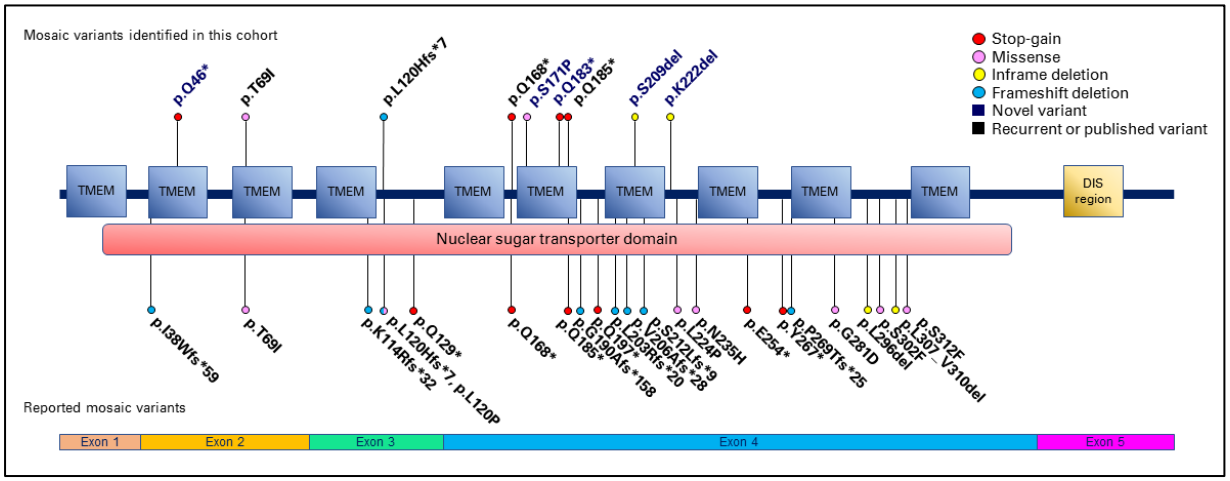


Figure 4.2: Variant map of brain mosaic *SLC35A2* variants identified in this cohort and published to date.
Novel variants are indicated in blue text. Variants above map were identified in this cohort. Variants below map have been previously reported. Red markers indicate Stop-Gain variants, pink markers indicate missense variants, yellow markers indicate inframe deletions and blue markers indicate frameshift deletions. Variants are mapped to transcript NM_0005660.3 (ENST00000247138).

Table 4.2: Cohort of IESS individuals with brain mosaic *SLC35A2* variants.

Individual ID	Sex	Initial diagnosis	Location (hg38)	Variant	AA change	Previously reported variant	Lesional tissue VAF (ES)	Lesional tissue VAF (ddPCR)	Normal tissue VAF (ddPCR)	Blood VAF (ddPCR)	CSF VAF (ddPCR)	Exon	Domains	MuTaste	SIFT	PPH2	CADD	PTC	gnomAD
AA0084-01	M	FCD IIA	X:48909952_G>A	c.136C>T	Q46*		20%	18.70%	14.20%	0%	ns	2	TMEM, NST domain	DC	n/a	n/a	36	Yes	0
AA2561-01	F	Non-specific	X:48909882_G>A	c.206C>T	T69I	[255]	0.95%	2.19%	0%	0%	0%	2	TMEM, NST domain	DC	D	PD (0.810)	25.2	No	0
AA0148-01	M	FCD IA	X:48906457_TGA>T	c.359_360delTC	L120Hfs*7	[114]	41%	62%	3.70%	0%	1.30%	3	NST domain	DC			n/a	Yes	0
AA1391-01	M	NAD	X:48905407_G>A	c.502C>T	Q168*	[258]	7.60%	7.60%	ns	0%	ns	4	NST domain	DC	n/a	n/a	36	Yes	0
AA0161-01	M	FCD IIA	X:48905398_A>G	c.511T>C	S171P		35%	36.30%	11%	0%	ns	4	TMEM, NST domain	DC	D	D (0.999)	27	No	0
AA0236-01	F	NAD	X:48905362_G>A	c.547C>T	Q183*		1.80%	1.65%	ns	0%	ns	4	TMEM, NST domain	DC	n/a	n/a	36	Yes	0
AA0164-01	M	Gliosis	X:48905356_G>A	c.553C>T	Q185*	[114, 150, 258]	1.90%	16.50%	0.64%	0%	ns	4	TMEM, NST domain	DC	n/a	n/a	35	Yes	0
AA0228-01	F	FCD IA	X:48905280_CAGG>C	c.626_628delCCT	S209del		5.60%	5.88%	ns	0%	0%	4	TMEM, NST domain	DC			n/a	No	0
AA0190-01	F	FCD IIA	X:48905241_ATCT>A	c.665_667delAGA	K222del		20%	23%	0.63%	0%	0%	4	NST domain	DC			n/a	No	0

AA, amino acid; ACMG, American College of Medical genetics and Genomics classification; CADD, combined annotated dependent depletion score; CSF, cerebrospinal fluid; D, damaging; DC, disease causing; ddPCR, droplet digital PCR; ES, exome sequencing; F, female; gnomAD, gnomAD population frequency; M, male; MuTaste, Mutation Taster prediction; n/a = not applicable; NAD, no abnormalities detected; NMD, predicted nonsense mediated decay; ns, no sample; NST domain, nucleotide sugar transporter; PD, probably damaging; PPH2, PolyPhen2 score; PTC, premature Stop codon; SIFT, sorting intolerant from tolerant score; TMEM, transmembrane domain; VAF, variant allele frequency.

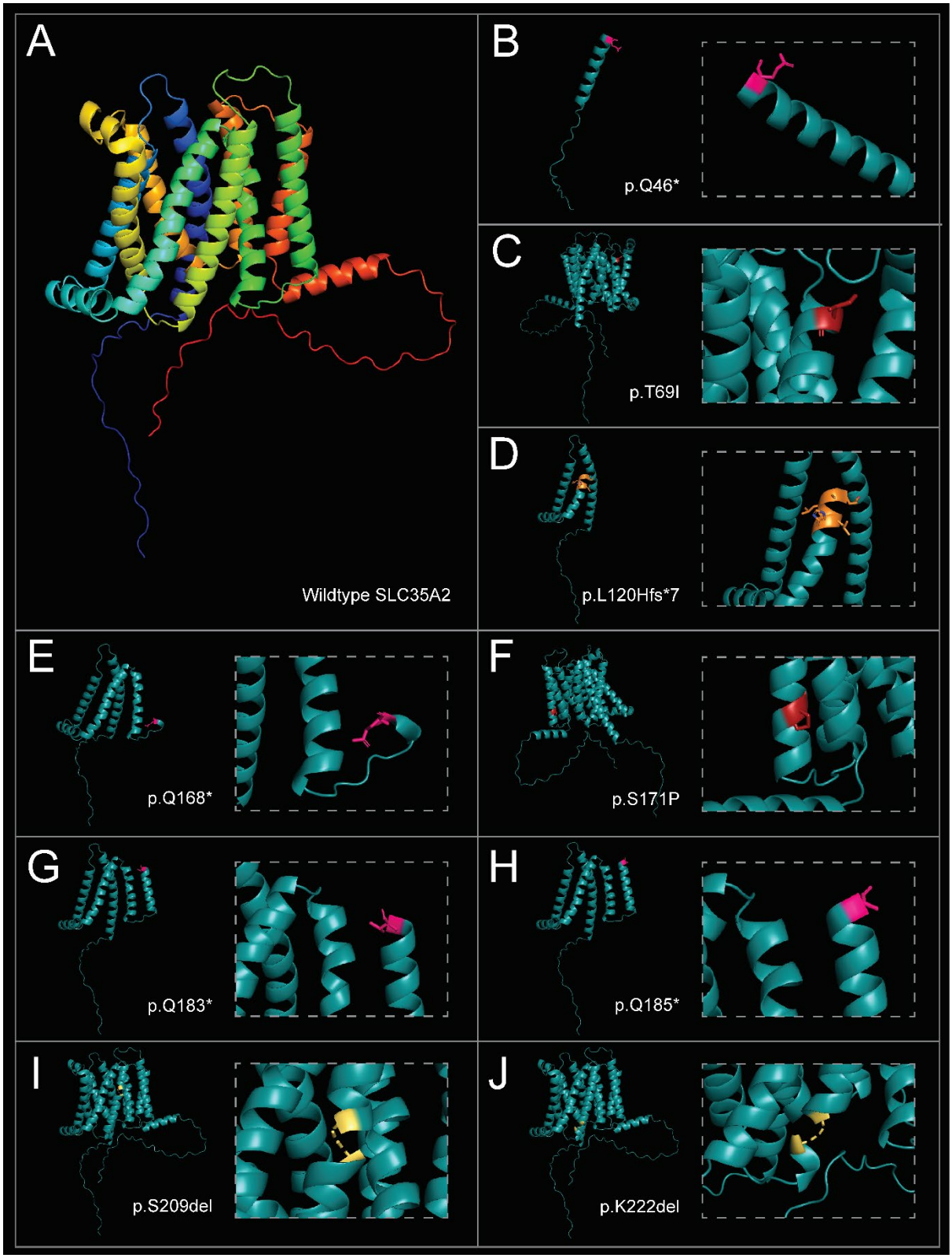


Figure 4.3: PyMOL visualisations of *SLC35A2* protein interruptions.
Panel A shows wildtype SLC35A2 protein (AlphaFold ID: P78381) with individual components identified in different colours. Panels B-J show mutant protein products identified in this cohort. Variants modelled include SLC35A2 p.Q46* (B), p.T69I (C), p.L120Hfs*7 (D), p.Q168* (E), p.S171P (F), p.Q183* (G), p.Q185* (H), p.S209del (I) and p.K222del (J). Wildtype amino acids are coloured teal. Pink indicates Stop-Gain variants. Red indicates missense variants. Orange indicates frameshift termination changes and yellow highlights deletions. Dotted box shows close up view

of each disrupted amino acid. Interaction prediction region modelling was limited to five angstroms from mutated residue.

4.3.2 Variant allele frequencies

Variant allele frequency (VAF) estimation in lesional tissues via MuTect2 analysis of ES data was highly concordant with absolute quantification via ddPCR and no significant difference was observed between the two detection methods overall ($p=0.13$, correlation coefficient (R^2) = 0.865) (Figure 4.4A-B). VAF was observed to be significantly higher in lesional tissue than predicted by MuTect2 in ES data in two individuals (AA0148-01, ES VAF = 41%, ddPCR VAF = 62%; AA0164-01, ES VAF = 1.90%, ddPCR VAF = 16.5%) via ddPCR, however this was not significant across the cohort.

VAF tended to be highest at regions indicated to be the primary lesion via MRI or histopathology. In each patient, VAF at the lesional regions were significantly higher than VAF at adjacent normal brain tissue ($p<0.05$). In some cases, VAF in tissues from peri-lesional areas was observed to be equal to or higher than primary lesional tissue. In no instances was VAF observed to be higher in normal tissue than in primary lesional tissue.

Variants were not detected in blood-derived gDNA from any individual via ddPCR and were only detected in CSF-derived cfDNA from one individual (AA0148-01, 1.3% VAF in CSF). However, CSF was unavailable for 5/9 individuals, limiting ability to draw conclusions about potential clinical utility of CSF as a sample for genomic testing.

Overall, a significant increase in VAF was observed ($p<0.05$) in primary lesional tissue compared with adjacent normal brain tissue, and a mosaicism gradient was observed between primary lesional brain tissue, peri-lesional tissue, normal tissue and CSF (Figure 4.4C-D). The observed mosaicism gradient had decreased to zero before reaching blood in all cases and as early as normal brain tissue in some (Table 4.2).

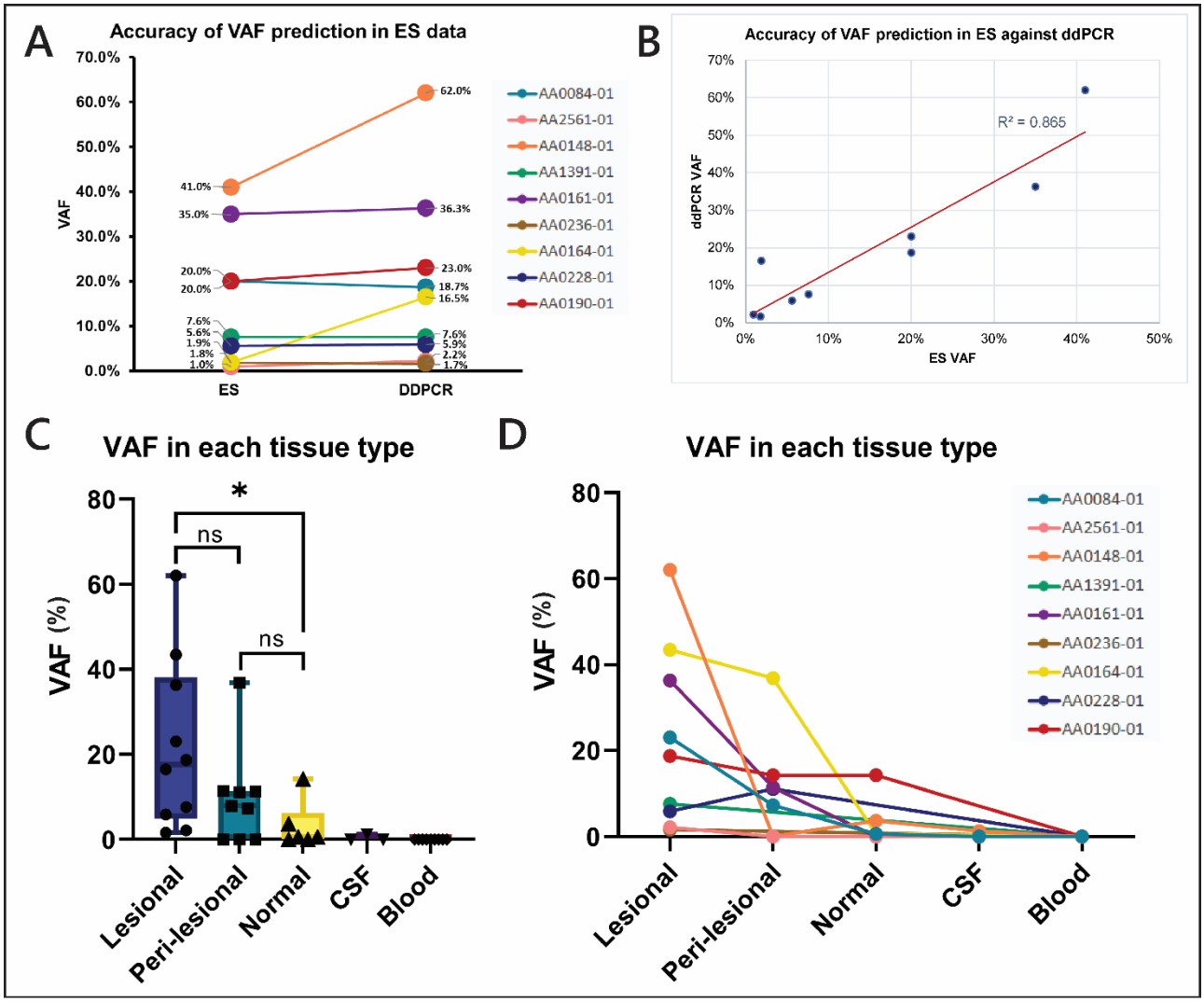


Figure 4.4: *SLC35A2* variant allele detection in tissues.
Panel A) Lesional *SLC35A2* VAF predicted by MuTect2 using ES data compared with ddPCR TaqMan absolute quantification in gDNA derived from primary lesion tissue. B) Correlation graph of predicted *SLC35A2* VAF from ES data (X-axis) and measured VAF via ddPCR (Y-axis). Blue data points indicate individual samples. Red line indicateds trendline. Correlation co-efficient (R^2) = 0.865. C) Box plot of ddPCR *SLC35A2* VAF detected DNA from primary lesion (lesional), peri-lesional and histopathologically normal (normal) brain tissues, cerebrospinal fluid (CSF) and blood. Box indicates mean and standard deviation. Error bars indicate minimum and maximum. Paired wilcoxon rank sum test. NS = not significant, * = $p<0.05$. D) Linked data from ddPCR *SLC35A2* VAF in each sample across lesional, peri-lesional and normal brain tissues, CSF and blood.

4.3.3 Histopathology

To evaluate if lesional tissue from individuals with *SLC35A2* brain mosaic variants met the histopathological criteria for MOGHE, immunostaining was performed to visualise the tissue architecture and composition. Anti-Olig2 immunostaining was performed to identify oligodendrocyte lineage cells and visualise oligodendrocyte cell density. pS6

(Ser235/236) immunoreactivity was used as a marker for cells that exhibit dysregulation of the mTOR pathway. NeuN immunostaining was employed to locate postmitotic neurons and Map2 was utilised to stain dendrites and perikarya of neurons.

Olig2 immunostaining revealed a marked increase in Olig2-positive cell density in the white matter across all nine brain tissue specimens from individuals with *SLC35A2* brain somatic variants, compared to control tissue or other MCD (Figure 4.5; Supplementary Figure 8.5). Clusters of Olig2-positive cells were evident throughout the white matter of lesional tissue and decreased in density towards peri-lesional tissue.

NeuN staining revealed heterotopic neurons in the deep white matter (marked in red arrows) and evidence of a blurred grey-white matter boundary, which was supported by Map2 staining. pS6 staining showed absence of mTOR activation in 7/9 *SLC35A2* cases. In the two SLC35A2 tissues with evidence of pS6 immunoreactivity in neurons, the intensity was considerably less than that observed in individuals with confirmed genetic FCD II pS6 activation and neurons were not sufficiently cytomegalic.

Blinded neuropathological assessment classified MOGHE in 3/9 *SLC35A2* cases and concluded there was insufficient white matter tissue for a definitive diagnosis in 6/9 *SLC35A2* cases (Table 4.3). An FCD II diagnosis was ruled out in all cases and these six cases were classified as 'possible MOGHE' via histopathology.

In addition, two individuals (AA0265-01 and AA0294-01) lacking a genetic diagnosis were classified as possible MOGHE based on neuropathological assessment of Olig2 and NeuN staining. These two individuals were also categorised as MOGHE candidates during neuroimaging review.

Although a correlation was observed between high VAF and more severe histopathological findings, the trend was not found to be significant between lesional and peri-lesional tissues.

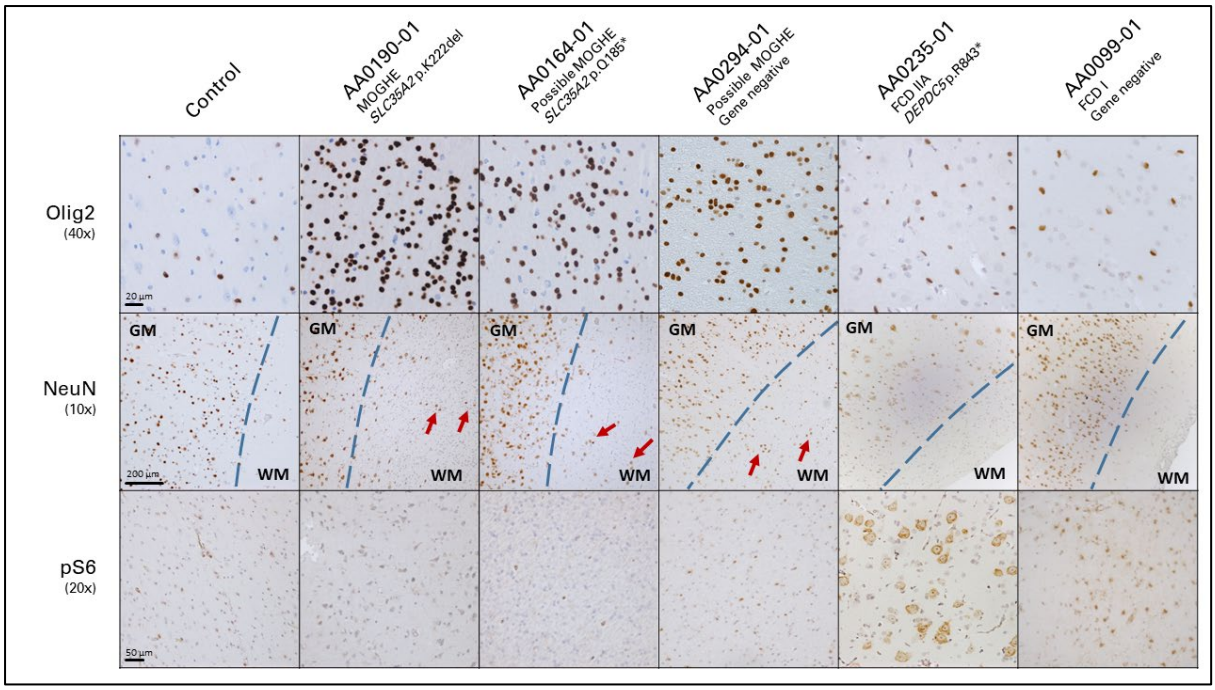


Figure 4.5: Histopathological features in *SLC35A2* brain mosaic cases.
Representative histopathology stains from an unaffected individual, two *SLC35A2* brain mosaic cases, one gene negative "possible MOGHE" case, one FCD IIA case and one FCD I case. Images of tissues have been chromogenically stained against anti-Olig2, anti-NeuN and anti-pS6. Blue dotted line indicates grey-white matter boundary. Red arrows indicate heterotopic neurons. GM = grey matter; WM = white matter. 40x, 10x and 20x indicate magnification. Black bar indicates scale.

4.3.4 MRI findings

Clinical review of the MRI scans of the nine individuals with brain-mosaic *SLC35A2* variants showed that six (AA0148-01, AA0164-01, AA0190-01, AA0084-01, AA0236-01 and AA0228-01) presented with features consistent with MOGHE subtype 1, including presence of FLAIR signal at the corticomedullary junction (Figure 4.6A). Four individuals displayed poor grey-white differentiation due to hyperintensity of the white matter and features of MOGHE subtype 2 (AA1391-01, AA0161-01, AA2561-01 and AA0228-01) (Figure 4.6B). This included one individual (AA0228-01, Figure 4.6C) with features of both MOGHE imaging subtypes, subtype 1 anteriorly and subtype 2 posteriorly. All nine individuals were classified as MOGHE by an independent, blinded neuroradiologist (Table 4.3, Supplementary Figure 8.6). No significant difference was

observed in mean age at time of surgery between MOGHE Subtype 1 and Subtype 2 (Supplementary Figure 8.6K).

Two additional individuals (AA0265-01 and AA0294-01) from the genetically unsolved cohort were identified as having features consistent with MOGHE subtype 1 (Table 4.3). These two individuals had also been reclassified as possible MOGHE during neuropathology reassessment.

No other cases from the broader IESS cohort (Chapter 3) were indicative of MOGHE upon MRI review.

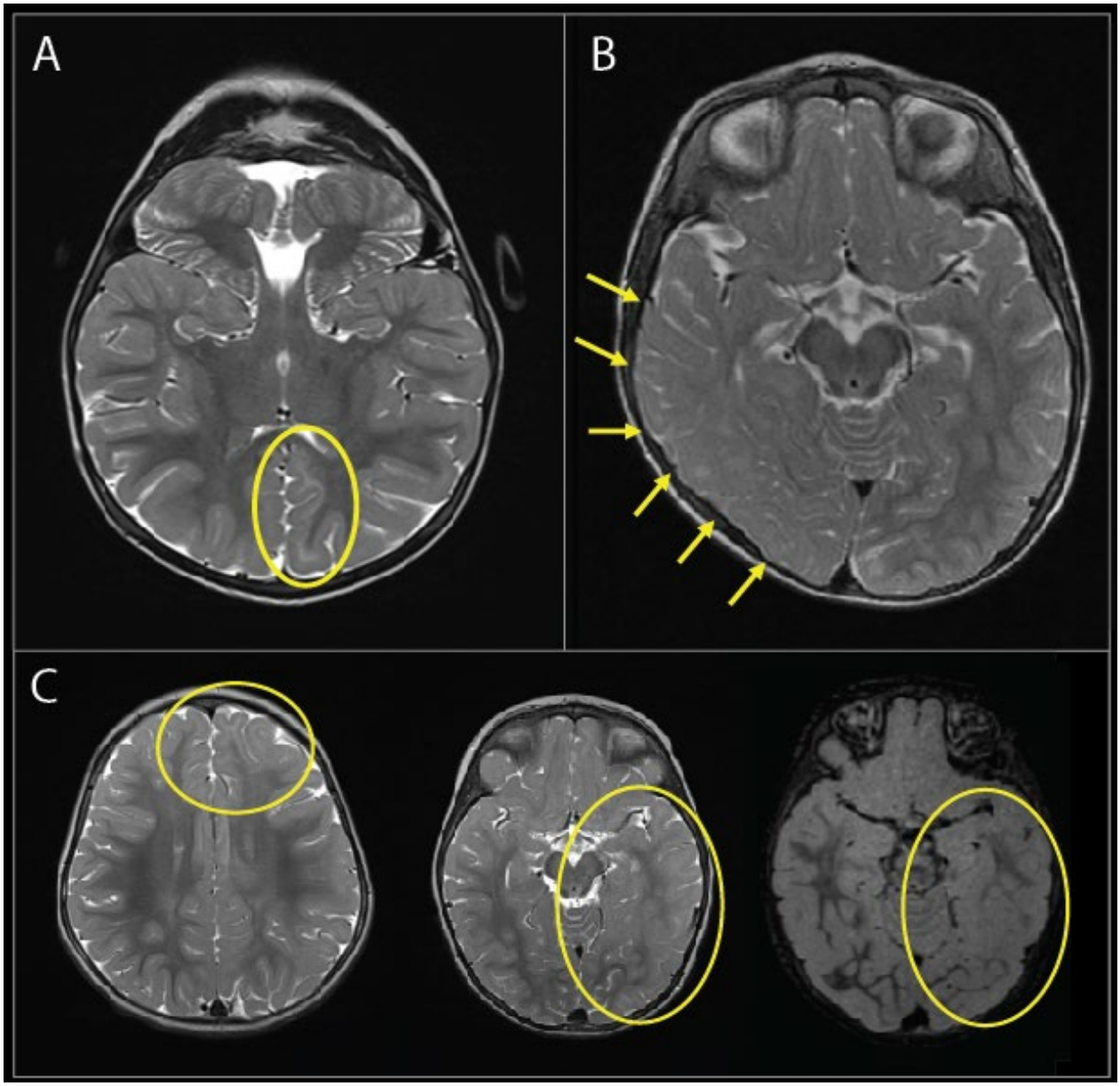


Figure 4.6 [previous page]: Representative MRI scans from this cohort demonstrating features of MOGHE subtypes 1 and 2.
Panel A shows axial view MRI of individual AA0190-01. Evidence of fluid attenuated inversion recovery (FLAIR) signal at the corticomedullary junction (consistent with MOGHE subtype 1) is indicated in the yellow circled area. Panel B shows axial view MRI of individual AA01391-01. Arrows indicate reduced corticomedullary differentiation because of increased signal of the adjacent white matter in the right temporal lobe (MOGHE Subtype 2). Panel C shows three successive sections of an axial scan of individual AA0228-01 with features of MOGHE subtype 2 in the posterior region, and subtype 1 in anterior regions, indicated by yellow circles. MRI scans of the full cohort are included in Appendix (Supplementary Figure 8.6).

4.3.5 Integrated diagnosis reclassification

Combined reassessment of clinical and neuropathological phenotype following genetic analysis, neuroimaging and histopathological review with new data resulted in the nine individuals with brain mosaic *SLC35A2* variants being reclassified as MOGHE (AA0084-01, AA0148-01, AA0161-01, AA0164-01, AA0190-01, AA0228-01, AA0236-01, AA1391-01 and AA2561-01) and two gene-negative individuals (AA0294-01 and AA0265-01) being reclassified as "possible MOGHE" (Table 4.3). Formal multi-modal reclassification as MOGHE would require a genetic finding of a somatic brain *SLC35A2* variant in the two gene-negative individuals, which has not occurred. The nine individuals with an integrated diagnosis of MOGHE were previously diagnosed as FCD I (n=5) FCD IIA (n=3) and gliosis (n=1) (Table 4.3).

Table 4.3: Table of clinical review findings and integrated diagnosis outcomes in *SLC35A2* brain mosaic cohort.

ID	Sex	Initial diagnosis	Integrated diagnosis	Histopathology			Revised histopathological diagnosis	Revised MRI diagnosis	Age at seizure onset (months)	Age at surgery (years)	Surgery extent	Surgery type	Engel Class	Seizure type at onset	Other seizure types	Age at last review (years)
				Olig2	NeuN	pS6										
AA0084-01	Male	FCD IIA	MOGHE	++	++	-	MOGHE	MOGHE subtype 1	3	2.7, 7	Lo, Lo	L F, L F	I	Spasms	Focal	13
AA2561-01	Female	Non-specific	MOGHE	+	++	-	Possible MOGHE*	MOGHE subtype 2	7	3	MLo	R TPO	I	Spasms	Tonic	4
AA0148-01	Male	FCD IA	MOGHE	+	++	-	Possible MOGHE*	MOGHE subtype 1	16	3,4,8	Lo, MLo, H	L F, L F/T, L H	I	Spasms	Tonic	12
AA1391-01	Male	NAD	MOGHE	+	++	-	Possible MOGHE*	MOGHE subtype 2	7	2	MLo	R TPO	I	Spasms	n/a	18
AA0161-01	Male	FCD IIA	MOGHE	+	++	-	Possible MOGHE*	MOGHE subtype 2	6	3, 7	Lo, MLo	R T, R TPO	UK	Spasms	Focal	10
AA0236-01	Female	NAD	MOGHE	+	++	-	Possible MOGHE*	MOGHE subtype 1	5	1.3	MLo	L TPO	II	Spasms	n/a	7
AA0164-01	Male	Gliosis	MOGHE	+	++	-	Possible MOGHE*	MOGHE subtype 1	4	2, 4	Lo, H	R F, R H	I	Spasms	n/a	5
AA0228-01	Female	FCD IA	MOGHE	+	++	-	Possible MOGHE*	MOGHE subtype 1 (ant) & 2 (post)	14	1.8, 4	MLo, Lo	L TPO, L F	I	Spasms	Tonic	6
AA0190-01	Female	FCD IIA	MOGHE	++	++	-	MOGHE	MOGHE subtype 1	2	6	Lo	R F	I	Spasms	Focal	8
AA0294-01	Female	FCD IIA	Possible MOGHE	+	++	-	Possible MOGHE*	MOGHE subtype 1	6	3	H	L H	I	Spasms	n/a	5
AA0265-01	Male	FCD IIA	Possible MOGHE	+	++	-	Possible MOGHE*	MOGHE subtype 1	4	4	Lo	R F	I	Spasms	n/a	6

F, frontal lobe; FCD, focal cortical dysplasia; H, hemispheric disconnection; L, left; Lo, lobar; MLo, multi-lobar; MOGHE, mild malformation of cortical development with oligodendroglial hyperplasia in epilepsy; n/a, not applicable; NAD, no abnormalities detected; R, right; T, temporal; TPO, temporo-parieto-occipital; UK, unknown; +, met requirments for clustering of Olig2 cells but not threshold increase (i.e. <2000 cells/mm2); ++, met all requirements for MOGHE classification [Olig2: clustering and >2000 cells/mm2 white matter density/NeuN: >20 white matter neurons per mm2]; -, cytomegalic neurons and pS6 immunorectivity not observed; *, noted limited sample depth for analysis.

4.3.6 Clinical and developmental phenotyping

Clinically, all of the individuals with *SLC35A2* mosaic variants and classified as MOGHE or possible MOGHE were similar in presentation and were distinct from FCD II presentations (Figure 4.7). Individuals with mosaic *SLC35A2*/MOGHE had 1-3 surgeries and often ultimately required multilobar, quadrantic or hemispheric disconnections. MOGHE cases had good post-operative outcomes after the last surgery, with 7/9 being seizure free at last review (i.e. Engel Class I). One MOGHE individual (AA0263-01) had ongoing episodes although it is unclear if these are actually seizures and the seizure outcome of another individual (AA0161-01) is unknown. In the two individuals classified as possible MOGHE, the extent of surgical resection was similar (hemispheric and lobar disconnection), and seizure freedom was achieved in both individuals.

In all nine individuals with MOGHE with mosaic *SLC35A2* variants, epilepsy began with IS, with median onset age of six months (range 2-16 months). Presentation with drug resistant IS was the indication for surgery in 7/9 (78%) individuals with MOGHE, and 2/2 (100%) possible MOGHE. Only two individuals with MOGHE (AA0084-01 and AA0190-01) had drug resistant focal seizures as the predominant indication for surgery. These two individuals were initially classified as FCD IIA, and had IS at seizure onset, however these settled and were replaced by focal seizures. In total, 6/9 individuals with MOGHE had another seizure type at some point, presenting as either focal (reflex) or tonic seizures. Developmental outcomes following surgery in the MOGHE cases ranged from normal to severely impaired with global developmental delay.

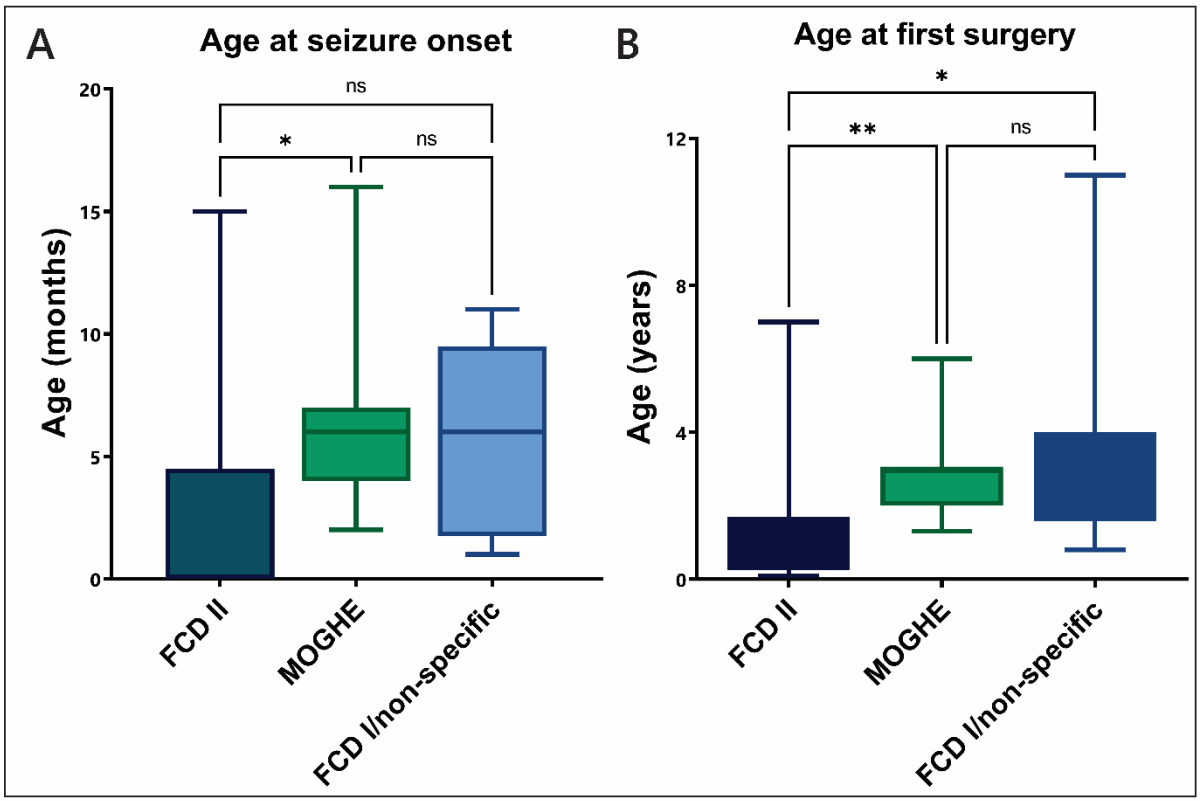


Figure 4.7: Phenotypic comparison between MOGHE, FCD II and FCD I or broader cohorts. Panel A) box plot of seizure onset between FCD II, MOGHE and FCD I/non-specific groups. B) box plot of age at first surgery between FCD II, MOGHE and FCD I/non-specific groups. Box indicates mean and interquartile range. Error bars indicate minimum and maximum. Kruskal-Wallis test performed for significance. ns = not significant, * = $p<0.05$, ** = $p<0.005$.

4.4 Discussion

4.4.1 MOGHE as an emerging phenotype

MOGHE is an entity that has only been described relatively recently, yet it appears to account for a significant proportion of IESS cases that were previously genetically unsolved. Of the nine individuals with IESS and mosaic *SLC35A2* variants identified in brain tissue, five had initial histopathological diagnoses of FCD I, three were classified as likely FCD IIA and one was classified broadly as gliosis (Table 4.3). Notably, none of these patients had initially been diagnosed as MOGHE. Review using genetic, histopathologic, clinical and imaging data in combination led to an integrated diagnosis with all nine cases being reclassified as MOGHE (Table 4.3). This multimodal method of reclassification demonstrates that while some biological hallmarks bear similarities to other MCD, and in particular FCD I and FCD IIA, MOGHE is a distinct

clinicopathological entity. This exemplifies the importance of multidisciplinary diagnostic pathways and the utility of post-surgical investigation. Furthermore, this highlights the need for review of historical diagnoses where appropriate as new genetic or diagnostic entities are described. Five of the individuals with MOGHE in this cohort required more than one surgery to address seizures and prior knowledge of MOGHE diagnosis may have informed surgical decisions and resulted in larger disconnections during the first surgery to minimise repeat surgeries and the likelihood of protracted seizures due to incomplete resection and the associated neurodevelopmental impacts. MOGHE is associated with distinctive surgical prognoses. Without an accurate diagnosis, clinical management is not informed, and clinical decisions may result in less effective outcomes.

The finding of this study that 9/9 individuals with brain somatic SLC35A2 variants were reclassified as MOGHE is concordant with previous studies and further suggests that somatic variants in *SLC35A2* are distinct from mTORopathies. Previously, somatic pathogenic variants in *SLC35A2* have been reported in MOGHE, mMCD and FCD I, although several of these FCD I and mMCD cases have since been reclassified as MOGHE following better characterisation of the disorder [113, 114, 261, 262]. It is increasingly clear that MOGHE is the predominant *SLC35A2*-related pathology [261, 262]. The presence of four recurrent variants in this cohort, previously reported in MOGHE, further suggests that a MOGHE diagnosis is accurate.

It is likely that brain mosaic *SLC35A2* variation invariably causes MOGHE and that these earlier FCD I and mMCD cases were undiagnosed MOGHE, as had initially occurred with the nine *SLC35A2* MOGHE cases in this cohort. It is clear that there are distinct clinicopathological features in MOGHE that are distinguishable from other MCD types associated with IESS such as FCD II, FCD I, gliosis, and TSC.

Upon histopathology and neuroimaging review, two additional 'possible MOGHE' cases (AA0265-01 and AA0294-01) were identified, who remain without a genetic diagnosis (Table 4.3). Despite their consistent neuropathological, imaging and clinical

features, exhaustive analysis of 400x ES data has not yielded any candidate genetic variants in *SLC35A2* in brain tissue in these individuals. This may be due to the fact that available tissue samples were distant from the area of most prominent malformation, given individual AA0294-01 had a hemispheric disconnection rather than a large resection and individual AA0265-01 had a right frontal lobe disconnection. It is also possible that ES coverage has missed deep intronic variants or CNVs that could affect *SLC35A2* expression or regulation. Alternatively, these individuals may carry variants in another gene that result in a phenotype similar to MOGHE. A better understanding of *SLC35A2* activity and protein interactors in the brain may assist in identifying other potential molecular causes of MOGHE. This will be discussed further in Chapter 5.

Following neuropathology reassessment, imaging review, genetic testing and clinical review, 11/59 (19%) individuals from the broader IESS with MCD cohort (Chapter 3) were ultimately classified as having MOGHE or possible MOGHE using an integrated diagnosis. Brain somatic variants in *SLC35A2* were identified in 9/59 individuals (15%), all of whom had an integrated diagnosis of MOGHE. As such, brain somatic *SLC35A2* variants are considered a major cause of IESS with MCD, and likely a major genetic cause of IESS more broadly. These findings are consistent with a report published by Barba et al. earlier this year [262]. MOGHE should be considered when performing imaging and histology review as it is becoming increasingly apparent that MOGHE is a substantial contributor to IESS with focal brain malformations.

Despite reports of unique imaging and neuropathological characteristics in MOGHE, the exact mechanism of pathogenesis remains unknown. The relationship between N-glycosylation and the emergence of a brain malformation and epilepsy is yet to be determined. A trial of galactose as a potential therapy for this condition is currently underway [263]. Increasing knowledge of this unique condition and investigating the underlying mechanisms of pathogenicity in MOGHE are both necessary for earlier detection of MOGHE cases in clinical settings and potential identification of precision therapies.

4.4.2 Molecular distinguishers of MOGHE

The current ILAE proposed criteria for histopathological classification of MOGHE may be excessively stringent and result in the omission of some MOGHE cases due to lack of appropriate tissue. Only 2/9 *SLC35A2* brain mosaic individuals were formally categorised as MOGHE upon blinded clinical neuropathology review, despite histopathologic features in all individuals with *SLC35A2* somatic variants being reminiscent of published MOGHE cases and were not indicative of other MCD classifications. The histopathological hallmarks of MOGHE are distinguishable when highlighted by the appropriate tools.

Although some cases did not meet the Olig2-positive cell density criteria recently proposed for MOGHE (>2000/mm2 at the grey-white matter junction), an increase in Olig2 immunoreactive cells (indicating an increase in oligodendroglial lineage cells) was observed in the white matter of *SLC35A2* positive tissues, compared to control tissue or other MCD. Despite the inconsistency of Olig2-positive cell density levels in the white matter of these samples, Olig2-positive cell clustering was observed in the white matter across all nine individuals with *SLC35A2* brain mosaic variants, as well as the additional two possible MOGHE cases without a genetic diagnosis. This conclusion was supported by NeuN staining which revealed heterotopic neurons in the deep white matter and evidence of a blurred grey-white matter boundary, further supported by Map2 immunoreactivity. Similarly, pS6 immunoreactivity, a marker of mTOR activation which is closely tied to FCD II and other mTORopathies, was weak in all MOGHE cases and no clear cytomegalic or dysmorphic neurons were identified.

These observations are consistent with published cases, which have listed a broad range in Olig2-positive cell density and histopathological features required for a definitive MOGHE diagnosis, from >1100/mm^2 to >2200/mm^2, or sometimes only requiring an increase in Olig2-positive cell density "compared to neighbouring regions" [111, 114, 264]. This study uses the most recent and stringent classification of MOGHE histopathological diagnosis (>2000 Olig2-immunoreactive cells/mm^2) from

the 2022 proposed classification requirements by the ad hoc task force of the ILAE diagnostic methods commission [111]. However, this may not be the most appropriate classification as it did not accurately identify 7/9 (78%) MOGHE cases in this cohort. These seven *SLC35A2* positive cases that did not meet the proposed ILAE criteria are clearly MOGHE in pathology. Their genetic, clinical and imaging presentations match those of MOGHE cases.

However, it is likely that histopathological diagnosis was inconclusive in these individuals due to a lack of appropriate tissue for analysis. As discussed in Chapter 3.4.11, tissue resection sizes are decreasing with advances in surgical accuracy and techniques [248]. In MOGHE in particular, most individuals underwent disconnective surgeries. As a result, tissue resections specimens are not generally taken from the region of most prominent malformation. As such, the tissues collected from surgery may not have been optimal for genetic or histopathologic assessment.

The lack of a definitive histopathological diagnosis of MOGHE in 7/9 individuals with somatic *SLC35A2* variants, could be due to the failure of overly stringent requirements and the fact that the resected tissue was unlikely to originate from the region of most prominent malformation in cases with larger disconnections. However, all six of these individuals were identified to carry somatic variants in *SLC35A2* in perilesional tissues. This suggests that molecular genetic testing is a valuable tool for identifying MOGHE in cases with suboptimal tissue samples.

Nevertheless, the histopathological features of MOGHE are often highlighted by Olig2 immunostaining, and the findings of this study highlight the importance of incorporating routine Olig2 screening in clinic for IESS cases following surgery. Three MOGHE cases in this cohort were initially classified as FCD IIA based upon clinical neuropathology assessment with only hematoxylin and eosin (H&E) stain, and neuroimaging. Two of these cases (AA084-01 and AA0190-01) clearly met even the most stringent histopathological criteria and were identified as MOGHE upon blinded neuropathological review following Olig2 staining.

Olig2 expression levels in MOGHE oligodendrocyte lineage cells may be informative if tested via laser capture microdissection (LCM) and RNAseq or running a comparison of snRNAseq expression levels in MOGHE Olig2-expressing cells and control Olig2-expressing cells.

4.4.3 Mosaicism gradient

All nine *SLC35A2* variants identified in this cohort fall within the main NST domain, and there is a trend towards variants within or on the boundaries of the TMEM helical domains. This diversity of variant type and spread of variant locations is consistent with the previously published mosaic *SLC35A2* cases (Figure 4.2). Pathogenic variants outside the NST domain have only been reported in germline form in SLC35A2-CDG (e.g. ClinVar VCV001686209.1).

The mean VAF in the lesional tissue was 19.31% (range 1.65%-62%) compared to a mean VAF of 3.20% in normal tissue (range 0%-14.2%). The VAF in lesional tissue was higher in all cases compared to paired normal tissue. Although ddPCR is the gold standard for absolute quantification of VAF, MuTect2 analysis of ES data reliably detected VAF within 1-2% margin in 7/9 cases. The differences observed in VAF between MuTect2 ES prediction and ddPCR absolute quantification from the same gDNA sample in individuals AA0148-01 and AA0164-01 were likely limited by read depth at the variant site. The VAF observed in ES data may also have been impacted by amplification bias of one allele during sequencing [228]. A mosaicism gradient was clearly observed in some cases but was inconsistent in others, with respective peri-lesional samples having similar or higher VAFs than primary lesional tissue, or lower VAFs than normal brain tissue.

4.4.4 MRI reveals developmental changes in MOGHE imaging characteristics

Features of MOGHE subtype 1 (n=6) or subtype 2 (n=4) were reported in all nine individuals upon blinded neuroimaging review (Table 4.3). A previous publication reported that MOGHE subtype 1 imaging is predominantly observed in younger

individuals and MOGHE subtype 2 is primarily observed in individuals who are older [260]. Although a slight trend was observed in this cohort, individuals with MOGHE imaging subtype 2 (mean age of 26 months at time of MRI, range 15-36 months) were not significantly older than individuals with MOGHE imaging subtype 1 (mean age of 19 months at time of MRI, range 10-31 months). This trend may have reached significance in a larger cohort. Hartlieb et al. postulated that the imaging features change with progression of myelination [260]. Interestingly, one individual (AA0228-01) in this study had both patterns present in one scan; subtype 2 posteriorly, and subtype 1 in the later-myelinating anterior regions. This individual was 1.8 years of age at time of MRI and first surgery.

White matter myelination occurs progressively, beginning prenatally, with a large portion occurring between 14 and 16 months after birth. Myelination progresses in the posterior to anterior direction and myelination of terminal U-fibres typically occurs between 18 to 24 months. Myelination in the brain is mostly complete by two years of age with terminal zone myelination continuing for a further 12 months [265, 266]. The intracortical fibres of the cerebral cortex are the final site of the adult brain to be myelinated [267]. IESS has been linked with delayed myelination [268].

The presence of both MOGHE subtypes on MRI in individual AA0228-01 at 1.8 years of age supports the theory of a progressive phenotypic spectrum in MOGHE, linked to myelination stage and progression.

It is also likely that myelination is abnormal in MOGHE. The increased T2 signal observed in MOGHE could indicate myelination loss or changes. Changes in myelination would ideally be examined via diffusion weighted MRI [269], however this data was unavailable for these patients.

4.4.5 Surgical outcomes in MOGHE vs FCD cases

Increasing knowledge and understanding of surgical outcomes in MOGHE patients has the potential to improve prognosis and genetic counselling.

The spectrum of developmental outcomes in individuals with MOGHE in this study was broad at last review, ranging from normal to severely impaired with global developmental delay. Sample size in this study is too small to accurately compare factors associated with developmental outcomes, however favourable outcomes were observed in individuals AA0084-01 and AA0190-01 following surgery. Both individuals presented with small malformations, remission of epileptic spasms during infancy, and experienced a prolonged period of seizure freedom prior to onset of drug resistant focal seizures.

Overall, MOGHE cases had good seizure outcomes after their last surgeries, with 7/9 (78%) individuals with MOGHE and 2/2 (100%) possible MOGHE achieving Engel Class I (i.e. complete seizure freedom post-surgery). The post-operative outcome of one individual with MOGHE was unknown and only one MOGHE individual reported ongoing seizures (Engel Class II).

Individuals with FCD II had similar post-operative outcomes, 10/13 (77%) FCD II patients achieved Engel Class I. Similar proportions of MOGHE and FCD II required multiple surgeries (6/13 (46%) in FCD II cohort, 5/11 (45%) in MOGHE and possible MOGHE cohort), and surgeries required were often multilobar and hemispheric in both MOGHE and FCD II. In contrast, FCD I cases generally had less extensive surgeries but were less likely to achieve seizure freedom (3/10 (30%) Engel Class I) (Supplementary Table 8.5). Pre-surgical screening for MOGHE in individuals with IESS and MCD may inform surgical decisions with the knowledge that MOGHE is often present in most or all of one hemisphere. There is the possibility that this knowledge could result in fewer but more extensive surgeries in the first instance in individuals with MOGHE.

4.4.6 Clinical presentation of MOGHE vs FCD

To date, the majority of studies investigating the genetic basis of focal cortical malformations have focused on FCD II due to the well-defined imaging and neuropathological hallmarks (such as DNs and BCs). These have provided key focus areas for research [93, 206]. Although recent studies of MOGHE brain tissues have

begun to identify neuropathological hallmarks and potential focal areas of MOGHE, these can often only be identified at a post-surgical stage [111, 114]. MOGHE does not yet have a well-described clinical phenotype. While the sample size of this cohort is small (n=11 MOGHE or possible MOGHE, n = 13 FCD II), it is clear that individuals with MOGHE were clinically quite similar to each other and distinct from FCD II phenotypes. Establishing the phenotypic spectrum and clinical features of MOGHE may allow earlier suspicion of MOGHE diagnoses which in turn may aid in surgical planning, direct pathological analysis including the use of Olig2 staining, inform genomic testing and aid in prognostic and genetic counselling.

Individuals with MOGHE were identified to have several clinical indicators that assist to distinguish MOGHE from FCD II. Individuals with a MOGHE diagnosis tended to have an older median age of onset for seizures, and focal seizures were less frequently observed in addition to spasms, compared with FCD II (Figure 4.7, Supplementary Table 8.5). Focal seizures are often the primary clinical concern and the indication for surgery in FCD II but, if present, focal seizures are often subsequent to epileptic spasms in MOGHE.

Genetic diagnostic yield achieved was similarly high in MOGHE/possible MOGHE (9/11 (82%)) and FCD II (12/13 (92%)) groups. Cortical dyslamination was more frequently reported in FCD II (13/13 individuals (100%)) than MOGHE/possible MOGHE (4/11 individuals (36%)), while gliosis was more commonly reported in MOGHE/possible MOGHE (7/11 individuals (64%)) than in FCD II (2/13 individuals (15%)) (Supplementary table 8.5). Intense pS6 staining, which is a marker of mTOR pathway activation, was observed in all FCD II cases but was typically absent or minimal in MOGHE tissues. This further indicates that MOGHE is pathologically distinct from FCD II and other mTORopathies.

Similarly, MRI features provide clues to distinguish FCD II and MOGHE diagnoses. Most frequently, in MOGHE cases, neuroimaging changes are identified in the white matter, with no observed changes in cortex and no difference in hemisphere size. FCD II

however, tends to involve cortical thickening and altered subcortical signal such as the transmantle sign [87]. Hemispheric changes may also be observed in FCD II. Though blurring of the grey-white matter is common to both MOGHE and FCD II.

MOGHE cases are distinct and often distinguishable from FCD II cases via multiple methods. Although the two entities share similarities in overall presentation, this study identifies MOGHE as a distinct clinicopathological entity from FCD II and other neuropathological diagnoses associated with IESS. MOGHE cases have distinct genetic, neuropathologic and imaging features when compared with other MCD, but the clinical presentation is not yet well defined. It is particularly important to be able to differentiate between MOGHE and FCD I or FCD II phenotypes, due to the potential difference in clinical management and impact of diagnosis on surgical decisions. As evidenced by the cases requiring multiple surgeries to achieve seizure freedom in this cohort and the extent of the surgeries in all nine cases, MOGHE cases generally required more surgeries and often ultimately larger disconnections or resections than many other FCDs. This is consistent with other studies [262]. As such, pre-surgical identification of MOGHE diagnoses would likely result in fewer and more accurate surgeries.

4.4.7 Multi-modal diagnostic tools and clinical reclassification of MOGHE

Multidisciplinary diagnostic tools are necessary to accurately identify MOGHE cases. No individual method – genetic analysis, histopathologic analysis, clinical phenotyping nor neuroimaging – would have reliably identified all nine *SLC35A2* mosaic MOGHE cases and the two gene-negative possible MOGHE cases. All individuals with MOGHE or possible MOGHE required two or more diagnostic tools prior to formal clinical reclassification.

Regardless of diagnostic capability, reliable prognostication was still elusive. Neither genetic variant nor VAF, lesion size nor lesion location were indicative of phenotypic severity or surgical outcomes in this cohort. Similarly, MRI lesion size and Olig2-positive cell density, as identified by IHC staining with anti-Olig2, did not accurately

reflect patient phenotypes in this cohort. This is consistent with recent findings by Barba et al. in a retrospective, multi-site cohort of 47 published cases with brain-mosaic *SLC35A2* variants [262]. Identification of molecular predictors of phenotypic severity in MOGHE would not only be important for informing clinical prognostication but may also assist in the selection of the best medications for the relevant molecular pathways affected in MOGHE.

4.4.8 Implementing tools for earlier and more accurate diagnoses

As discussed earlier in this thesis, none of the MOGHE or possible MOGHE cases were identified clinically prior to this study. These initial misdiagnoses were due, in part, to a lack of clinician awareness of MOGHE. MOGHE is a relatively recently described entity, first described in 2017, and the ILAE classification was not updated to include MOGHE until 2022 [111, 112]. Furthermore, surgical specimens collected during disconnected surgery are often not ideal for rigid criteria required for formal histopathological classifications. These shortfalls highlight the need for increased clinician knowledge of MOGHE and, more importantly, the need for more robust clinical and imaging criteria to enable earlier identification of MOGHE prior to surgery.

When clinicians are knowledgeable of neuroimaging hallmarks of MOGHE, they are able to recognisable these hallmarks in clinic. Similarly, the clear delineation of neuropathology provided by pS6 and Olig2 immunoreactivity demonstrate the value of incorporating these markers in routine clinical pathology assessment following epilepsy surgery. Additionally, review of historical FCD I or non-specific malformation cases may result in the identification of further MOGHE cases. This may have important impacts on clinical decisionmaking for many historical IESS patients who are identified to have MOGHE both within Australia and globally.

Although three FCD IIA cases were reclassified as MOGHE, none of these individuals were observed to have hemimegalencephaly. No FCD IIA cases with an additional diagnosis of HME were reclassified following multi-modal reassessment. To date, there have not been any MOGHE cases reported with features consistent with HME.

Although there are similarities between MOGHE and the FCD IIA phenotypes, these similarities do not extend to FCD IIA with HME. The three individuals with MOGHE with initial diagnoses of FCD IIA were reasonably soft diagnoses with no clear dysmorphic neurons being reported, only dyslamination and potential MRI features of FCD II.

In most MCD, regions with higher VAF are predicted to present with more severe histopathological features and, as such, are often selected for genetic analysis thereby theoretically increasing the chance of finding the pathogenic somatic variation [270]. This hypothesis was tested in MOGHE as part of this study and although there was a trend towards higher VAF in some primary MOGHE lesions, this trend was not significant and was not consistent across all samples, as discussed in Chapter 4.4.3 (Figure 4.4B). Retrospective examination of electrophysiological data and histopathological findings may be used to guide selection of brain specimens with detectable variant allele, but VAF may not necessarily correlate with electrophysiological data and histopathological findings in MOGHE.

4.4.9 Mosaicism and a novel phenotype for glycosylation disorder

These observations beg the question, "how does a presumed glycosylation disorder cause a brain malformation when in a mosaic form?". There are several possible reasons that could relate to variant severity, tissue localisation or some unknown interactor in the brain.

It is possible that the difference in phenotype is due to the severity of the individual variants in *SLC35A2*. Some of the genetic variants that are observed in a mosaic form may not be viable in a germline form. X-linked dominant disorders can be embryonic lethal in hemizygous males and the excess of female cases of SLC35A2-CDG lends support to this hypothesis though rare hemizygous males have been reported [252]. There is little crossover between the genetic variants reported in SLC35A2-CDG and those reported in MOGHE. The majority of mosaic variants reported in MOGHE cases have been novel or only previously reported as mosaic in MOGHE.

Lesion size or location in the brain may contribute to the phenotype by interrupting important networks in the brain. The fact that all MOGHE cases in this cohort and the majority of reported MOGHE cases present as IESS, regardless of overall lesion size or location suggests that size or location are unlikely to be clear predictors of severity, outcome or malformation genesis. Additionally, the link between MOGHE and IESS is still not understood.

Alternatively, other interactors of SLC35A2 function in the developing brain may be integral to oligodendrocyte development and neural formation. Little is currently known about the interactors of SLC35A2 in the brain. Proteomics studies on brain or disease models of MOGHE may assist to delineate such pathways and interactors. This possibility will be discussed further in Chapter 5.

4.4.10 Limitations

For this study, initial recruitment required a clinical indication of IESS and neurosurgery. Therefore, it is possible that an ascertainment bias for epileptic spasms has occurred, resulting in an over-representation of IESS and MOGHE. Although spasms in infancy have typically been described in published MOGHE cases, previous studies have reported a subset of cases without epileptic spasms [262]. As such, the MOGHE phenotypic spectrum may be wider than represented by this study. As MRI techniques improve and MOGHE neuroimaging hallmarks become more widely known, detection of additional MOGHE cases without infantile spasms may become more frequent.

A second limitation is the capacity to detect brain somatic variants, which is constrained in cases with limited tissue (discussed in Chapters 3.4.10 and 4.4.2 of this thesis).

Another possible limitation may be the underdiagnosis of MOGHE more broadly. Several MOGHE cases in this cohort and reported previously had initially presented with mild malformations that were not detected on routine MRI screening. Upon detailed review, these individuals were found to have hallmarks of the MOGHE imaging subtypes. Therefore, it is possible that other cases have been missed previously. These

cases would not likely have been identified as surgical candidates, leading to a deficit in cases being identified at subsequent post-surgical stages, such as brain mosaic *SLC35A2* variant detection via genetic testing in brain tissue or classification via neuropathological assessment.

4.5 Conclusions

As demonstrated in Chapter 3 of this thesis and expanded upon in this chapter, brain limited somatic *SLC35A2* variants are a major cause of IESS with brain malformations. Affected individuals present with a distinct clinicopathological entity now termed MOGHE. Clinical diagnostic methods are evolving with this emerging phenotype, and there may be an opportunity to identify a greater portion of MOGHE cases at earlier stages through improved imaging and histopathological review. Targeted neuroimaging analysis prior to surgery may improve detection rates in MOGHE cases, and subsequently lead to fewer surgeries and improved outcomes in these patients. Additionally, routine Olig2 immunostaining in all surgical IESS cases may assist to detect MOGHE cases and inform clinical decisions post-surgery.

Genetic diagnosis would be the optimal method for identifying MOGHE cases, given the limited success in identifying MOGHE cases through histopathological or imaging analysis alone. However, genetic diagnosis in MOGHE currently requires surgical brain tissue for analysis. Although pre-surgical genetic testing is currently not available, it may be possible to detect brain somatic *SLC35A2* variants via less invasive methods in the future, such as genetic screening using CSF. In this study, only one genetic variant was able to be identified using cfDNA derived from CSF. This yield may be increased with improved technologies and screening methods.

Determining the pathogenic mechanisms in MOGHE remains a challenge to date. Although a connection to N-glycosylation has been established, the exact mechanism of pathogenesis is not yet understood. Examining the molecular profiles of MOGHE

tissues with cutting-edge single-cell sequencing or proteomic technologies may provide some insight into the molecular drivers of epileptogenicity in this disorder and help to explain how a glycosylation disorder & oligodendrocyte abnormality result in a focal brain malformation.

5 Chapter 5: Single cell RNA-sequencing and proteomic analysis reveal insights into MOGHE pathogenesis

5.1 Introduction

This chapter describes single-nuclei transcriptomic and high-resolution proteomic analyses of surgical brain tissue specimens from individuals identified as having mild malformation of cortical development with oligodendroglial hyperplasia in epilepsy (MOGHE). The work described in this chapter addresses the third major aim of this thesis; to interrogate the molecular mechanisms involved in the pathogenesis of MOGHE using powerful, cutting-edge molecular tools.

Improved understanding of the molecular mechanisms underlying MOGHE is important for researchers, clinicians and patients. Early intervention with effective treatment of spasms reduces the severity of developmental impacts and is associated with better neurodevelopmental outcomes. Current medical treatments are only effective in some individuals and the development of more effective treatments requires a deeper understanding of the molecular causes of MOGHE. This study utilises single nuclei RNA sequencing and high-resolution proteomics analyses to elucidate potential transcriptional and proteomic changes in MOGHE that may contribute to the IESS and brain malformation phenotypes observed in individuals with MOGHE.

Data and analysis discussed in this chapter is currently being prepared as a manuscript for submission to Genetics in Medicine, together with work described in Chapter 4 of this thesis. The final publication may differ slightly in inclusions or detail to what is presented in this chapter. Additional wet laboratory experiments and bioinformatic analyses are underway in relation to this work. However, these additional studies extend beyond the timeline of this thesis and will not be discussed in detail.

5.1.1 SLC35A2 and N-glycosylation research

SLC35A2 is a transmembrane protein that transports nucleotide sugars across cellular membranes and is involved in glycosylation. Glycosylation is the most complex and ubiquitous post translational modification (PTM) [271]. There are five major forms of glycosylation, N-glycosylation, O-glycosylation, C-glycosylation, phosphoglycosylation and glypiation. SLC35A2 activity is associated with N-glycosylation, which involves the attachment of an oligosaccharide – a carbohydrate made up of various sugar molecules – to a nitrogen atom, to form an N-linked glycan or N-glycan [272]. N-glycosylation occurs in the Golgi apparatus and endoplasmic reticulum (ER) of mammalian cells, and plays a crucial role in correct protein folding, stability, trafficking, and function.

SLC35A2 plays a crucial role in transporting UDP-galactose, the necessary substrate for the synthesis of N-glycans involved in N-glycosylation, from the cytosol into the lumen of the Golgi apparatus [250]. If SLC35A2 transporter activity is interrupted in the Golgi apparatus, subsequent N-glycosylation is impaired, affecting downstream cellular function. SLC35A2 is X-linked and germline loss of function (LoF) defects result in a congenital disorder of glycosylation (SLC35A2-CDG), first described in 2013 [255]

As discussed in Chapter 4, somatic variants in *SLC35A2* in the brain have recently been reported as causing a newly recognised focal MCD termed mild malformation of cortical development with oligodendroglial hyperplasia in epilepsy (MOGHE) [114]. The molecular mechanisms underpinning MOGHE remain poorly understood and *SLC35A2* mosaicism presents pathogenesis with no increased activation of the mTOR pathway, the predominant pathway disrupted in FCD [148].

Ng et al. have recently developed a method to measure SLC35A2-dependent UDP-galactose transport into the Golgi of fibroblasts, as a diagnostic test for SLC35A2-CDG [256]. An equivalent biochemical test has not yet been developed for MOGHE and transferrin isoform testing has shown to be normal in some individuals with *SLC35A2* variants but abnormal in others [256, 273]. Similarly, galactose supplementation has

been reported to show benefit in some individuals with SLC35A2-CDG [274] but conclusive evidence supporting the application of this treatment in MOGHE is yet to be established.

Although a connection to N-glycosylation has been established, the exact mechanism of pathogenesis in MOGHE is not yet understood. High resolution analyses such as single cell RNA sequencing (scRNA-seq) and high-resolution proteomics may enable identification of the molecular drivers of epileptogenicity in MOGHE or provide insights into dysregulated pathways. The way in which somatic, brain-specific dysregulation of N-glycosylation impacts IESS and brain malformations is yet to be determined.

5.1.2 Application of single cell RNA sequencing technologies on brain tissue

MOGHE has been associated with disruption of the N-glycosylation pathway and loss of SLC35A2 function results in abnormal glycosylation, an overabundance of Olig2-positive cells in the white matter, heterotopic neurons in white matter and blurred grey-white matter boundary (Figure 4.2) [114, 275]. Due to the heterogeneity of cell types in the brain, and particularly the knowledge of cell types of interest as heterotopic neurons and oligodendroglial cells, traditional 'bulk' RNA sequencing (RNA-seq) is not adequately sensitive to explore differences in gene expression at the necessary resolution. Next-generation molecular tools such as scRNA-seq are required to delineate the pathogenesis of mosaic *SLC35A2* and MOGHE further.

scRNA-seq is a method of transcriptomic analysis that allows examination of a sample at the cellular level. scRNA-seq technology was first used to describe mammalian cell in 2009 [276] and has since revolutionised transcriptomic analysis as it facilitates high-throughput interrogation of gene expression data at the level of individual cells. scRNA-seq reveals expression profiles specific to individual cell types and enables the identification, characterisation and interrogation of disease associated cells. The technique was named Science magazine's Breakthrough of the Year in 2018. Single nuclei RNA-sequencing (snRNA-seq) has become a widely used alternative in cells with

low transcription levels or originating from precious tissues. Whole cells, particularly neurons, are difficult to capture intact using single cell methods due to morphology and tissue composition in brain. Nuclei provide a more accurate capture of cell expression in the brain than whole cells and snRNA-seq presents an appropriate method for capturing neurons and rare cell types in the brain [277].

There are two major techniques utilised in scRNA-seq: 1) droplet-based methods such as the 10X Genomics system, which operates by capturing individual cells within droplets containing barcoded labels, and 2) microfluidics-based systems such as the Fluidigm C1system, which require an integrated fluidics chip (IFC) to process up to 96 cells per preparation at one time. During snRNA-seq barcoding, nuclei are passed through a microfluidics system and each nucleus is barcoded with a unique molecular identifier (UMI). Each transcript originating from that nucleus contains the specific UMI. RNA transcripts are sequenced with respective UMI labels until the transcriptome of each nucleus is derived. The data is processed into clusters by similarity and plotted into a uniform manifold approximation and projection (UMAP) graph for visualisation.

Historically, snRNA-seq analyses on MCD brain tissue have focused on FCD and TSC due to the presence of known disease specific cells, such as dysmorphic neurons (DNs) and balloon cells (BCs), which have provided a focal point for analysis. These studies have been successful at identifying dysregulated pathways and potential biomarkers of disease in these disorders [278]. Work has been done to increase knowledge of the brain at the transcriptomic level in several other disorders. In 2022, a single-cell atlas of human brain vasculature was described, with paired normal and malformed tissues from cerebrovascular diseases [279].

Transcriptomic studies of SLC35A2-related disorders (i.e. SLC35A2-CDG and MOGHE) are extremely limited with only two studies being reported and both consisting of only bulk RNA-seq analysis on *SLC35A2* mosaic brain tissues [264, 275]. These studies found that *SLC35A2* messenger RNA (mRNA) expression did not correlate with mutant VAF in the same regions and that *SLC35A2* expression was extremely low in brain [264,

275]. Miller et al. hypothesised that nonsense mediated decay (NMD) may have contributed to low expression of the mutant allele in brain [264]. To date, no published studies have utilised scRNA-seq or snRNA-seq to interrogate MOGHE tissues.

5.1.3 Protein interactors with SLC35A2

SLC35A2 is one of multiple nucleotide sugar transporters (NSTs) [271]. One major study was published in 2021 investigating novel potential interactors of SLC35A2, along with interactors of the closely related NSTs SLC35A3 and SLC35A4 [271]. SLC35A2 was found to interact with two ATPases, ATP2A2 and ATP2C1, Golgi pH regulator B (GPR89B) and calcium channel (TMCO1) [271]. Limited study has been performed to identify additional protein interactors of SLC35A2 and such information is vital to improving the molecular understanding of MOGHE.

SnRNA-seq and high-resolution proteomics may enable identification of the molecular drivers of epileptogenicity in MOGHE and potential biomarkers of disease. High resolution proteomics analyses have the capacity to enable investigation of protein interactors, trends and potential biomarkers in disease tissues. Tandem mass tag (TMT) pro 16-plex liquid chromatography tandem mass spectrometry (LC-MS/MS) is a high-resolution method of proteomics that allows a greater number and variety or range of proteins to be identified in a sample [280, 281]. It uses a tandem labelling and multiplexing method which enables fast throughput, high sensitivity, and deep proteome coverage. This is particularly useful in cases such as MOGHE where tissue is limited and originates from a rare and precious resource. These brain tissue specimens are limited, and it is important to ensure the greatest amount of data is collected from these specimens when they are used in research testing.

Furthermore, multi-omic analysis methods, combining more than one high resolution method such as snRNA-seq and proteomics may provide complementary data which can be explored to a greater level. Multi-omics analyses could enable the identification of the molecular drivers of MOGHE and will form critical pre-clinical knowledge in the search for improved treatment strategies for epilepsy.

A detailed single nuclei transcriptomic dataset complemented with high resolution proteomics may elucidate the molecular mechanisms of MOGHE by highlighting dysregulated pathways and providing clues as to the identity of unusual cell types in MOGHE. This knowledge may provide insight into how a glycosylation disorder and oligodendrocyte abnormality result in a focal brain malformation. Further to this, the identification of dysregulated genes and proteins in MOGHE may assist in the identification of prospective biomarkers in MOGHE and subsequently, proteins that may serve as druggable targets in the future.

5.1.4 Aims

This study hypothesised that dysfunction within cells harbouring somatic variants in *SLC35A2* are the pathological catalyst for epileptic seizure activity in MOGHE and that protein biomarkers of disease will be present in lesional MOGHE tissues.

This study aimed to determine the molecular signatures of tissues and cells within resected brain tissue that harbour an *SLC35A2* somatic variant, and to identify potential biomarkers of disease in MOGHE. To accomplish this, single nuclei RNA sequencing and proteomic analysis were used to compare lesional MOGHE tissues and adjacent healthy normal MOGHE tissues from the same individuals. Multi-omics analysis of these data was performed to identify potential biomarkers of disease in MOGHE.

5.2 Methods

5.2.1 Cohort composition

The nine MOGHE cases identified in aims 1 and 2 of this Thesis were included in this analysis. MOGHE samples were separated into normal (low VAF) and lesional (high VAF) MOGHE samples, based on histopathological findings and VAF. Two additional likely MOGHE cases identified in Chapter 4 without a mosaic *SLC35A2* genetic diagnosis were included in proteomic analysis (AA0294-01-04 and AA0265-01-04).

SnRNA-seq analysis was performed on 11 tissues from individuals with MOGHE with pathogenic *SLC35A2* somatic variants in the brain (Table 5.1). Tissues were collected during epilepsy surgery and consisted of seven lesional specimens with high variant allele frequencies (VAF) (MOGHE_Lesional) and five paired adjacent normal brain tissue specimens (MOGHE_Normal) from the cohort of seven patients IESS with MOGHE. Normal MOGHE samples had a lower VAF (mean VAF = 3.93%, median VAF = 2.17%) and lesional MOGHE samples had a higher VAF (mean VAF = 24.6%, median VAF = 20.0%).

Proteomic analysis was performed on a total of 30 samples. This cohort included 17 samples originating from lesional MOGHE brain tissue from 11 individuals (MOGHE lesional = MOGHE_L) and seven samples from normal tissue from seven individuals with MOGHE (MOGHE normal = MOGHE_N). All MOGHE tissues were collected during epilepsy surgery from paediatric patients and included specimens derived from the specimens used for snRNA-seq analysis. Protein extraction from tissue was carried out in conjunction with RNA and DNA extraction used for analysis in Chapters 3 and 4 of this thesis. As such, proteomic data is directly related to genomic data discussed in these chapters, including variant allele frequencies (VAF).

As discussed in Chapter 4.4, it cannot be said with certainty that normal MOGHE tissues with a detected VAF of 0% were completely absent of *SLC35A2* variants, but rather that any variant in these tissues were below the LOD in ddPCR experiments (LOD >0.01%).

One individual (AA0177-01) with hemispheric polymicrogyria (PMG) due to a variant in *ATP1A3* was used as a disease control and healthy control in the proteomics analyses. Control samples comprised three tissue specimens derived from lesional brain tissues from an individual with hemispheric PMG (PMG tissues = CONTROL_P) and three paired healthy tissue specimens from matched regions on the unaffected hemisphere (healthy tissues = CONTROL_H) from the same individual (Table 5.2).

Table 5.1: snRNA-seq samples.

Case	Sample ID	Library ID	Tissue type	Sex	Age at surgery	Surgery	Gene	Variant	AA Change	VAF	Total reads	# of cells
1	AA0148-01C	ML212988	MOGHE_Lesional	Male	4.2 years	L frontal lesionectomy	*SLC35A2*	c.359_360delTC	Leu120HisfsTer7	62%	260,067,584	4,094
2	AA0161-01A	ML212989	MOGHE_Lesional	Male	3.6 years	R temporal corticectomy	*SLC35A2*	c.511T>C	Ser171Pro	35%	232,380,634	5,149
3	AA0190-01D	ML212990	MOGHE_Lesional	Female	6.9 years	R frontal corticectomy	*SLC35A2*	c.749_751delAGA	Lys222del	23%	266,508,901	4,170
4	AA1391-01A	ML212991	MOGHE_Lesional	Male	7.6 years	R TPO disconnection	*SLC35A2*	c.502C>T	Gln168Ter	7.6%	279,940,744	4,554
5	AA0228-01C	ML212992	MOGHE_Lesional	Female	1.8 years	L TPO disconnection	*SLC35A2*	c.626_628delCCT	Ser209del	5.6%	289,491,795	6,826
6	AA0164-01A	ML223051	MOGHE_Lesional	Male	2.8 years	R frontal corticectomy	*SLC35A2*	c.553C>T	Gln185Ter	43%	227,985,898	3,128
8	AA0084-01A	ML223537	MOGHE_Lesional	Male	7.7 years	L frontal corticectomy	*SLC35A2*	c.136C>T	Gln46Ter	18.7%	270,037,764	4,266
1	AA0148-01B	ML212993	MOGHE_Normal	Male	8.6 years	L hemispherotomy	*SLC35A2*	c.359_360delTC	Leu120HisfsTer7	0.7%	266,435,288	4,785
2	AA0161-01B	ML212994	MOGHE_Normal	Male	3.6 years	R temporal corticectomy	*SLC35A2*	c.511T>C	Ser171Pro	0%	257,213,617	7,807
3	AA0190-01A	ML212995	MOGHE_Normal	Female	6.9 years	R frontal corticectomy	*SLC35A2*	c.65_667delAGA	Lys250del	0.6%	274,157,161	9,010
7	AA2561-01C	ML223536	MOGHE_Normal	Female	3.6 years	R TPO disconnection	*SLC35A2*	c.206C>T	Thr69Ile	0%	248,639,287	2,645

L, left; R, right; TPO, temporo-parieto-occipital; AA amino acid; MOGHE, mild malformation of cortical development with oligodendroglial hyperplasia in epilepsy; SLC35A2, solute carrier family 35 member A2; VAF, variant allele frequency.

Table 5.2: Samples used for proteomics analysis.

#	Sample ID	Disorder	Label	Sex	Gene	Variant	VAF	Tissue	Age At Surgery
1	AA0148-01A	MOGHE	MOGHE_L	M	*SLC35A2*	Leu120Hisfs*7	3.7%	LESIONAL	3.9 years
2	AA0148-01C	MOGHE	MOGHE_L	M	*SLC35A2*	Leu120Hisfs*7	62%	LESIONAL	4.2 years
3	AA0161-01A	MOGHE	MOGHE_L	M	*SLC35A2*	Ser171Pro	35%	LESIONAL	3.6 years
4	AA0161-01C	MOGHE	MOGHE_L	M	*SLC35A2*	Ser171Pro	11.4%	LESIONAL	7.2 years
5	AA0190-01D	MOGHE	MOGHE_L	F	*SLC35A2*	Lys222del	23%	LESIONAL	6.9 years
6	AA0190-01E	MOGHE	MOGHE_L	F	*SLC35A2*	Lys222del	7.9%	LESIONAL	6.9 years
7	AA0084-01A	MOGHE	MOGHE_L	M	*SLC35A2*	Gln74*	18.7%	LESIONAL	7.7 years
8	AA0084-01B	MOGHE	MOGHE_L	M	*SLC35A2*	Gln74*	14.2%	LESIONAL	7.7 years
9	AA1391-01A	MOGHE	MOGHE_L	M	*SLC35A2*	Gln196*	7.6%	LESIONAL	7.6 years
10	AA0228-01A	MOGHE	MOGHE_L	F	*SLC35A2*	Ser237del	11.1%	LESIONAL	1.8 years
11	AA0164-01B	MOGHE	MOGHE_L	M	*SLC35A2*	Gln213*	16.5%	LESIONAL	2.8 years
12	AA0164-01A	MOGHE	MOGHE_L	M	*SLC35A2*	Gln213*	43%	LESIONAL	2.8 years
13	AA0236-01A	MOGHE	MOGHE_L	F	*SLC35A2*	Gln211*	1.7%	LESIONAL	1.3 years
14	AA2561-01D	MOGHE	MOGHE_L	F	*SLC35A2*	Thr96Ile	2.2%	LESIONAL	3.6 years
15	AA0294-01B	MOGHE	MOGHE_L	F	*SLC35A2*	unknown	n/a	LESIONAL	3.6 years
16	AA0265-01C	MOGHE	MOGHE_L	M	*SLC35A2*	unknown	n/a	LESIONAL	4.9 years
17	AA0228-01C	MOGHE	MOGHE_L	F	*SLC35A2*	Ser237del	5.9%	LESIONAL	4.7 years
18	AA0148-01B	MOGHE	MOGHE_N	M	*SLC35A2*	Leu120Hisfs*7	0.7%	NORMAL	8.6 years
19	AA0161-01B	MOGHE	MOGHE_N	M	*SLC35A2*	Ser171Pro	0.0%	NORMAL	3.6 years
20	AA0190-01A	MOGHE	MOGHE_N	F	*SLC35A2*	Lys222del	0.6%	NORMAL	6.9 years
21	AA0164-01D	MOGHE	MOGHE_N	M	*SLC35A2*	Gln213*	0.6%	NORMAL	4.6 years
22	AA2561-01C	MOGHE	MOGHE_N	F	*SLC35A2*	Thr96Ile	0.0%	NORMAL	3.6 years
23	AA0294-01E	MOGHE	MOGHE_N	F	*SLC35A2*	unknown	n/a	NORMAL	3.6 years
24	AA0265-01D	MOGHE	MOGHE_N	M	*SLC35A2*	unknown	n/a	NORMAL	4.9 years
25	AA0177-01D	n/a	CONTROL_H	M	*ATP1A3*	Leu924Pro	50%	CONTROL	0.5 years
26	AA0177-01E	n/a	CONTROL_H	M	*ATP1A3*	Leu924Pro	50%	CONTROL	0.5 years
27	AA0177-01G	n/a	CONTROL_H	M	*ATP1A3*	Leu924Pro	50%	CONTROL	0.5 years
28	AA0177-01A	PMG	CONTROL_P	M	*ATP1A3*	Leu924Pro	50%	POLYMICROGYRIA	0.5 years
29	AA0177-01C	PMG	CONTROL_P	M	*ATP1A3*	Leu924Pro	50%	POLYMICROGYRIA	0.5 years
30	AA0177-01F	PMG	CONTROL_P	M	*ATP1A3*	Leu924Pro	50%	POLYMICROGYRIA	0.5 years

Ctrl, control; F, female; M, Male; MOGHE, mild malformation of cortical development with oligodendroglial hyperplasia in epilepsy; n/a, not applicable; PMG, polymicrogyria; VAF, variant allele frequency.

5.2.2 Nuclei Isolation

Nuclei were selected rather than whole cells as nuclei provide a more accurate capture of cell expression in the brain, and because intact neurons are difficult to capture using scRNAseq. Nuclei were isolated as described in Chapter 2.6. Fluorescence-activated nuclei sorting (FANS) was performed on a BD Influx Cell Sorter (Becton Dickinson Biosciences, New Jersey). Nuclei sorting quality control (QC) is included as Appendix 8.6. Following FANS, nuclei concentration was estimated using a hemocytometer. Viability scores greater than 95% were observed in all nuclei suspension samples when viewed under hemocytometer. Nuclei suspensions required a cell count >500 nuclei/µL for snRNA-seq application.

Nuclei isolation buffer initially contained 1% BSA, however this was raised to 3% BSA content after difficulties with nuclei processing during the FANS.

5.2.3 Single nuclei RNA sequencing and bioinformatic sequencing

Nuclei suspensions were loaded onto the 10X Chromium System to generate barcoded cDNA libraries using the Chromium Next GEM Single Cell 3' Gene Expression kit v3.1 (dual index) (10X Genomics, #PN1000130) and library construction kit (10X Genomics, #PN-1000196) according to the manufacturer's protocol. Sequencing was performed on a research basis at the Victorian Clinical Genetics Services (VCGS, Melbourne), targeting ~5000 nuclei at a depth of ~40,000 reads per cell.

Bioinformatic processing including clustering and annotation was performed by Dr Saskia Freytag at the Brain Cancer Centre within the Walter and Eliza Hall Institute of Medical Research (WEHI), as described in Chapter 2.6. Differentially expressed genes (DEGs) were identified as dysregulated to a log fold change (logFC) greater than 1 or less than -1 with an adjusted p-value (adj. p) less than 0.05.

5.2.4 TMT-Labelled Mass Spectrometry and proteomic analysis

Parallel to snRNA-seq analysis, high resolution proteomic analysis was performed on 24 MOGHE tissues of varying VAFs, three healthy control brain tissues and three disease control brain tissues (PMG). For proteomic analysis, 30 samples of brain-

derived whole protein were submitted for tandem mass tag (TMT) -16pro liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis to quantify protein, as described in Chapter 2.7.

Differential proteins (DPs) were identified at a logFC greater than 2.5 or less than -2.5 and an adj. $p<0.05$ between MOGHE_L and CONTROL_H tissues. LogFC cut off was reduced to ±1.5 in paired tissue analyses (i.e. MOGHE_N vs MOGHE_L tissues from individuals with MOGHE) due to the high quality of data and the internal control of being paired brain tissues. Proteins relating to blood coagulation and bleeding were removed from analysis, as these were considered artefacts deriving from specimen collection method between MOGHE_N/MOGHE_L tissues collected during epilepsy surgery, and CONTROL_H/CONTROL_P collected post-mortem.

5.2.5 Biological process and pathway analyses

Analyses of impacted biological processes and dysregulated pathways were carried out using KEGG pathway (https://www.genome.jp/kegg/pathway.html), String database (https://String-db.org), Reactome (https://reactome.org/PathwayBrowser/#/) and Gene Ontology (GO) pathway (http://geneontology.org/) tools. Subsequent multi-tool analyses and figure creation were completed using WebGestalt (https://Webgestalt.org) and GoFigure (https://gitlab.com/evogenlab/GO-Figure). WebGestalt settings: Over-representation Analysis (ORA) with GO term 'Biological Process' and 'Molecular Function' filters. Pathways were sorted by count in network and filtered to exclude terms with a false discovery rate (FDR) greater than 0.05. GO term lists were cross-referenced between snRNA-seq data and proteomic data to identify robust dysregulated pathways.

5.3 Results

5.3.1 Single nuclei RNA-sequencing cohort and read quality

There were 51,071 independent nuclei captured in this experiment in total across the 11 samples analysed. Of these, 29,018 nuclei originated from MOGHE_Lesional tissues with high VAF (56.8%) (n=7) and 22,053 originated from MOGHE_Normal tissues of low VAF (43.2%) (n=4). Total reads per sample in all cases was above 225,000,000 (mean = 248,887,650, median = 248,639,287, range = 227,985,898 - 289,491,795). Estimated nuclei captured was greater than 2,600 in all samples (mean = 5,130, median = 4,554, range = 2,645 - 9,010). Mean reads per nuclei count ranged between 30,428 and 94,004 (mean = 56,881, median = 64,471) and mean genes per nucleus ranged between 1,607 and 2,604 (mean = 1,997, median = 1,986). Fraction of reads mapped to antisense was >10% in most samples though this was not unexpected for single nuclei analyses (mean fraction of reads mapped to antisense = 30.4%). A low fraction of reads in nuclei was observed in most cases but this is not atypical of brain tissue (mean fraction of reads in nucleus = 66.0%). High nuclear origin (>99%) was identified during mitochondrial contamination filtering.

5.3.2 Characterisation of MOGHE cell clusters in snRNA-seq data

A published dataset of markers was used for initial cell annotation and clustering [191]. Separation in clustering was observed between MOGHE_Lesional and MOGHE_Normal tissues in the snRNA-seq data, particularly in astrocytes, oligodendrocyte progenitor cells (OPCs) and dividing oligodendrocyte progenitor cells (cOPCs). Although oligodendrocytes accounted for the largest proportion of nuclei in lesional tissues overall and were overrepresented in MOGHE_Lesional (59% total nuclei) compared to MOGHE_Normal (29% total nuclei) tissues, this trend was not significant at the individual tissue sample level (Figure 5.1, Supplementary Figure 8.7). Oligodendrocytes have been well documented as overabundant in MOGHE lesional tissues, as further exemplified by the histopathologic data discussed in Chapter 4.4. As oligodendrocytes were a major source of difference between the two tissue groups, it was expected that

oligodendrocytes would be the major cell of interest and would harbour the most differentially expressed genes (DEGs). Despite their relative overrepresentation in the dataset, no DEGs were identified between oligodendrocytes originating from MOGHE_Lesional and MOGHE_Normal tissues.

SLC35A2 transcript level was consistently too low (*SLC35A2* expression <2.5 log counts (DAT)) to accurately compare *SLC35A2* expression between normal and lesional MOGHE tissues within snRNA-seq data (Figure 5.2). This was not unexpected as it is established that *SLC35A2* is generally lowly expressed in brain after infancy, and a previous study has shown *SLC35A2* mRNA expression to be low in brain tissue [264]. The data was interrogated for other dysregulated genes that may be contributing to the MOGHE phenotype.

Unbiased differential expression analysis of all lesional and normal cell types revealed 302 DEGs between the two tissues, consisting of 251 DEGs in cOPCs, 45 DEGs in OPCs, five DEGs in astrocytes and one DEG in layer 5/6 *THEMIS* expressing neurons. Of these 302 DEGs, six were observed in more than one cell type. Four genes were upregulated in both cOPCs and OPCs (*NPIPA1*, *PALM2-AKAP2*, *ANKRD55* and *SEMA3E*), one gene was downregulated in both cOPCs and OPCs (*PGRMC1*) and one gene was upregulated in both cOPCs and astrocytes (*TPTEP1*). No other dysregulated genes were shared between cell clusters. Several transcripts of mitochondrial genes were identified as DEGs in cOPCs but were disregarded due to irrelevance in analysis and subsequent filtering.

A 'Mixed' nuclei population was identified via clustering annotation analysis (Figure 5.1A). This population consists of nuclei which clustered together and did not meet expression marker criteria for any one cell group. Further to this, several satellite clusters of these Mixed cells were identified on the fringes of several other cell groups, including astrocytes, OPCs, oligodendrocytes, Layer 5/6 *THEMIS*-expressing excitatory neurons, Ca2+-binding protein parvalbumin (PV) and neuropeptide somatostatin (SST) GABAergic (inhibitory) interneurons. This Mixed cell population is not typically

observed in snRNA-seq data from healthy brain tissue specimens and accounted for a greater portion of nuclei originating from lesional MOGHE tissues than normal MOGHE tissues. The Mixed cell population accounts for 4.2% (1225) nuclei in MOGHE_Lesional tissues and 1.6% (370) nuclei in MOGHE_Normal tissues.

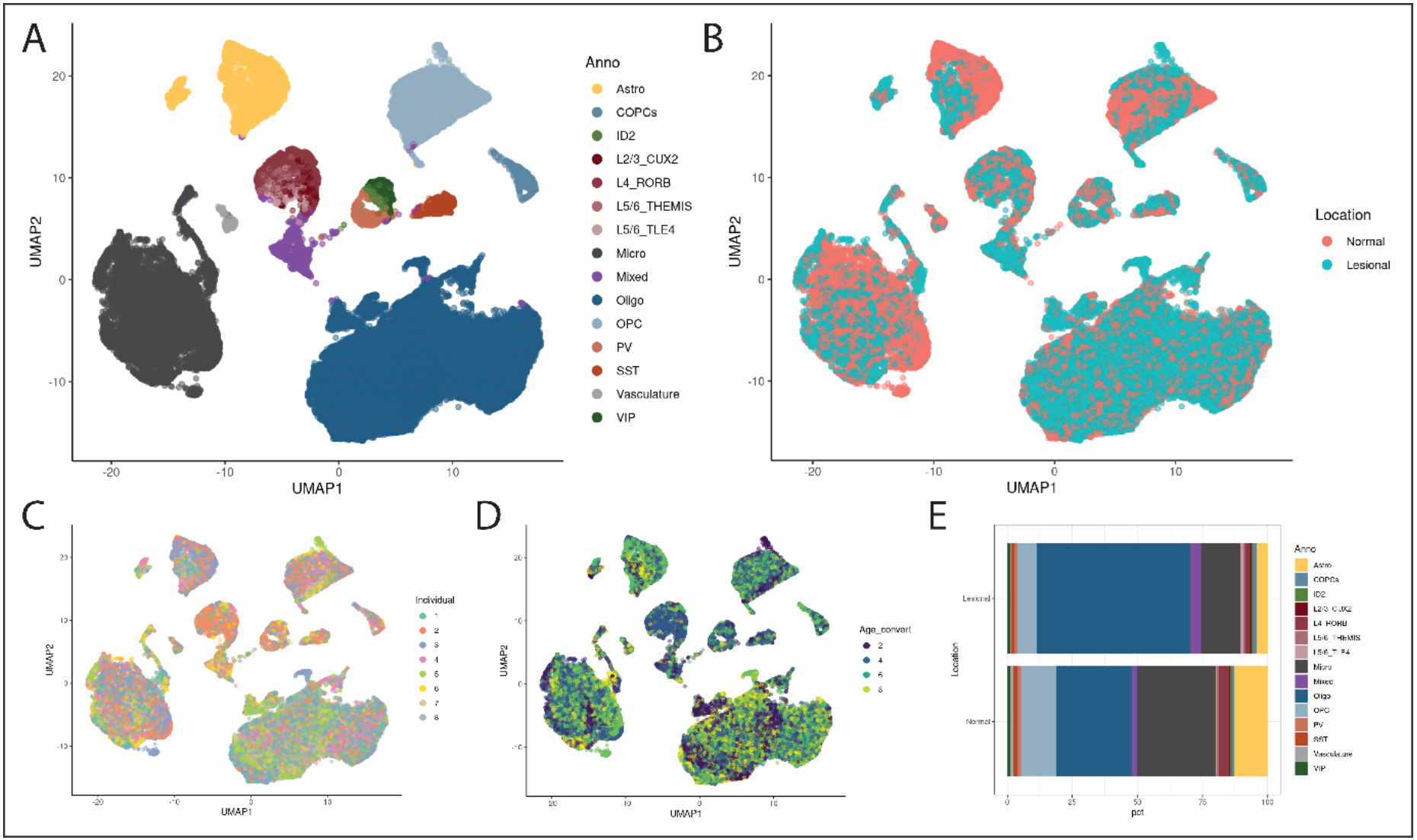


Figure 5.1: Full snRNA-seq dataset.

A) Uniform manifold approximation and projection (UMAP) graph of snRNA-seq data with annotated nuclei clusters labelled. Mixed population indicated in purple. B) UMAP showing MOGHE_Lesional (blue) and MOGHE_Normal (pink) tissue separation. C) UMAP separated by donor origin. D) UMAP coloured by age of donor at time of surgery. E) Proportional abundance of cell types in snRNA-seq data between MOGHE lesional and normal tissues.

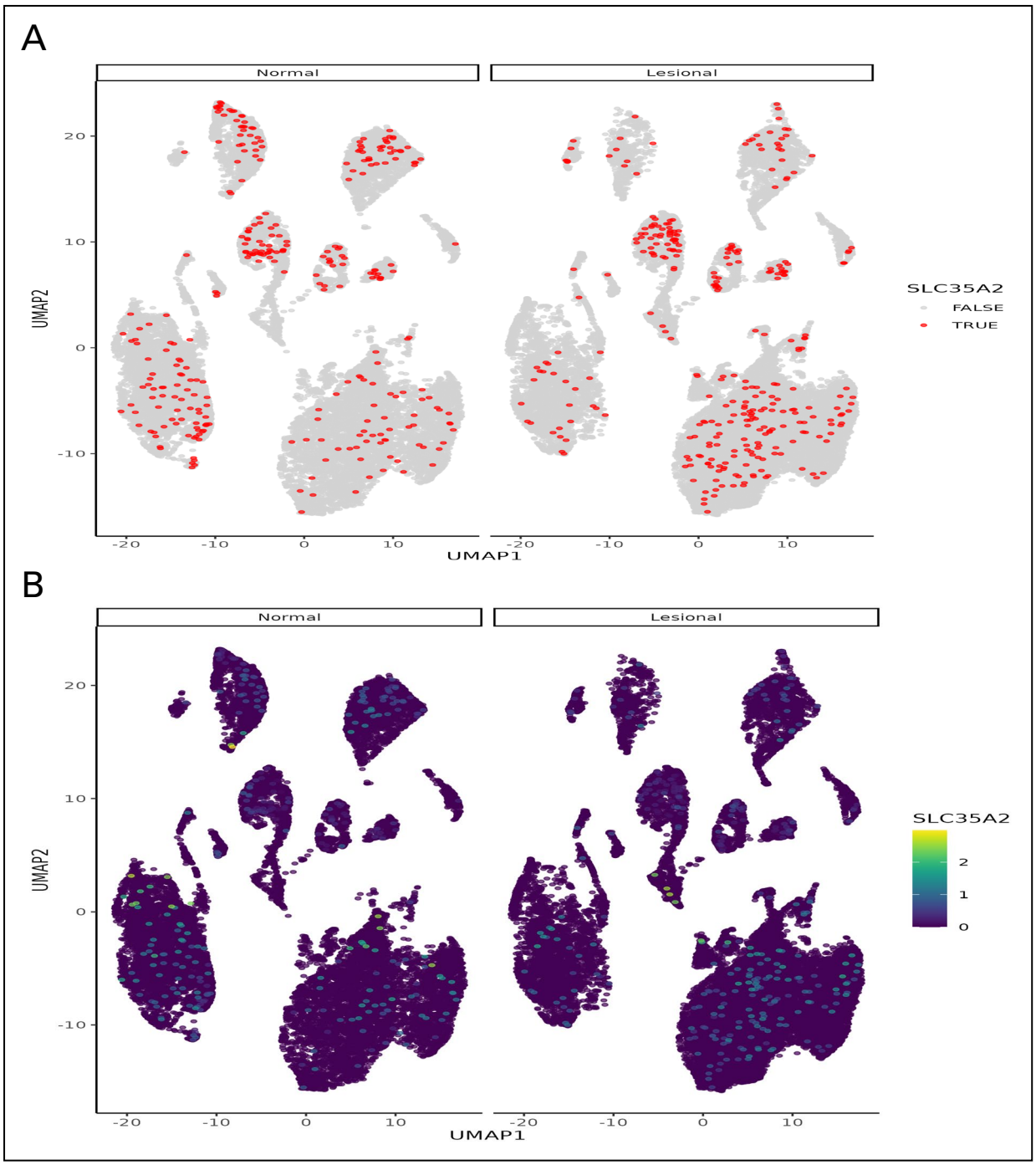


Figure 5.2: *SLC35A2* detection in the snRNA-seq dataset.
A) UMAP of MOGHE_Normal (left) and MOGHE_Lesional (right) snRNA-seq data with cells with detectable expression of *SLC35A2* highlighted in rose (true). B) Levels of *SLC35A2* expression detected in snRNA-seq dataset was consistently too low (<2.5 DAT) to be adequately evaluated in this cohort.

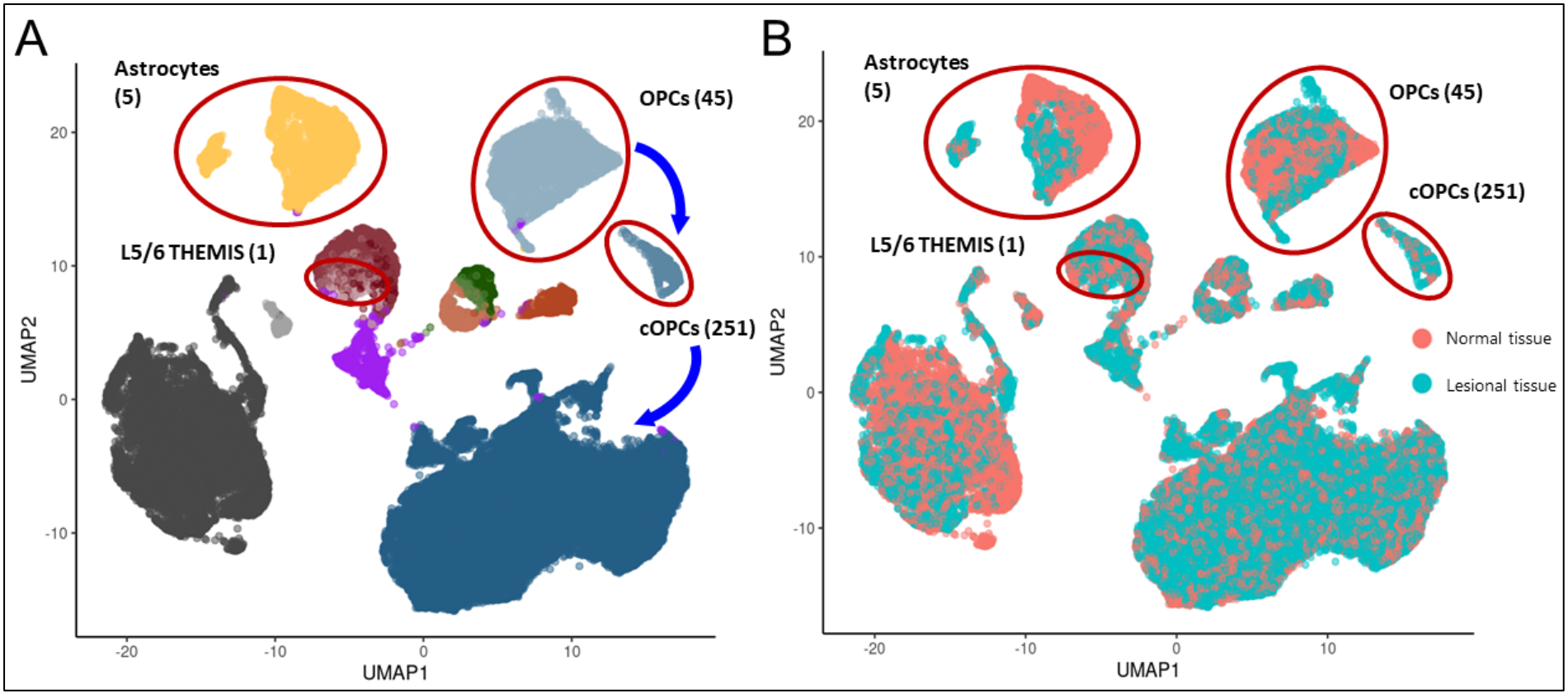


Figure 5.3: DEGs identified in snRNA-seq data.

A) annotated UMAP of full dataset adapted from Figure 5.1A. B) UMAP of full dataset showing MOGHE_lesional tissue (blue) and MOGHE_Normal tissue (pink). Red ovals indicate nuclei populations with DEGs identified between MOGHE_Lesional and MOGHE_Normal tissues. Number in brackets indicates number of DEGs identified in each population. Blue arrows indicate developmental pathway of oligodendrocyte progenitor cells (OPCs) to dividing oligodendrocyte progenitor cells (cOPCs) and mature oligodendrocytes. L5/6 THEMIS = layer 5/6 *THEMIS*-expressing neurons.

5.3.3 Reclustering of oligodendrocytes, OPCs and cOPCs

A total of 17,115 oligodendrocyte nuclei were identified in the MOGHE_Lesional dataset, encompassing 59.0% of all nuclei captured in the lesional dataset. 6,420 oligodendrocyte nuclei were captured in the MOGHE_Normal dataset, constituting 29.1% of nuclei originating from the normal dataset. Reclustering analysis was performed on these oligodendrocyte populations to elucidate molecular processes that resulted in proliferation of these cells in high VAF MOGHE_Lesional tissues. Secondary reclustering analysis revealed no further differences between oligodendrocytes of lesional MOGHE origin compared with oligodendrocytes originating from normal MOGHE tissues (Supplementary Figure 8.7).

There were 251 DEGs identified within cOPCs and 45 DEGs were identified in OPCs. Due to the presence of a large number of DEGs in these cell populations and the relationship between OPCs, cOPCs and oligodendrocytes, reclustering analysis was completed on these cells to further delineate potential differences in these cells. Reclustering analysis of cOPCs and OPCs resulted in the formation of six unbiased cell clusters, however no significant differences were observed between these six populations and no clear separation between MOGHE_Lesional and MOGHE_Normal tissues was observed (Figure 5.4).

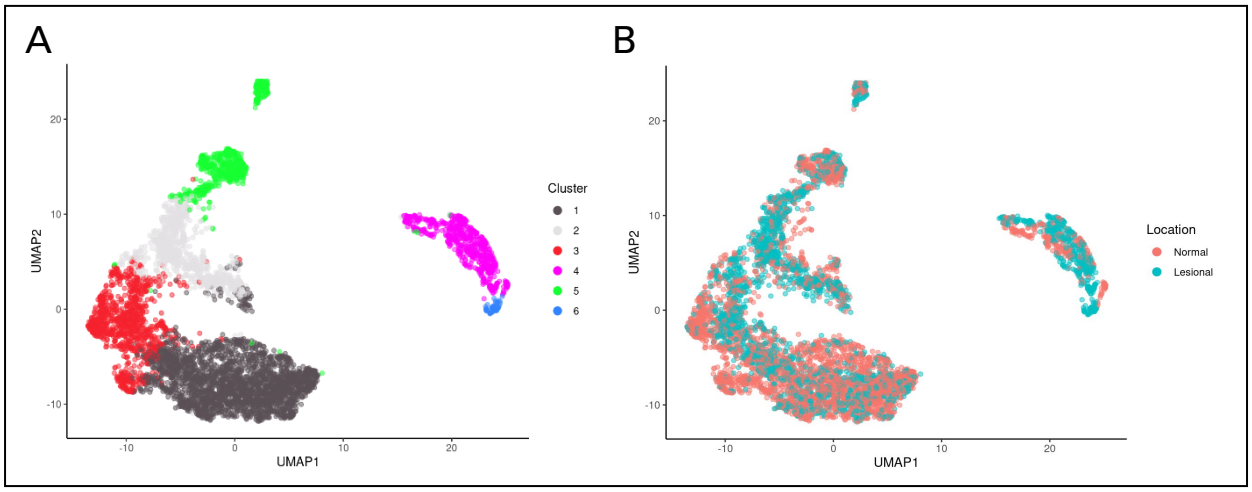


Figure 5.4: Reclustering analysis of OPCs and cOPCs.
A) UMAP of unbiased reclustering analysis of OPCs and cOPCs. B) UMAP annotated with sample origin as either MOGHE_Lesional (blue) or MOGHE_Normal (pink).

5.3.4 Proteomic profile of MOGHE tissue

The run efficiency of the TMT-16pro LC-MS/MS was observed to be >97% (i.e high efficiency of protein labelling). From the cohort of 30 brain specimens that underwent TMT-16pro LC-MS/MS, 8,843 protein groups were successfully identified. Of these, 7,655 proteins were able to be quantified consistently across all 30 samples. Six pairwise comparisons were initally performed which revealed 777 proteins that differed significantly between samples using adj. $p < 0.05$ and LogFC of ±2.0 cut-off (Figure 5.5). Analysis revealed 123 differential proteins (DPs) between MOGHE_L and CONTROL_H tissues (Figure 5.6) and one DP was identified between MOGHE_L and MOGHE_N cohorts.

A single protein, SOD1, was identified as differing significantly between paired lesional (MOGHE_L) and normal (MOGHE_N) tissues in the MOGHE cohort above a LogFC of 2.5, at an adjusted p-value of <0.001. As MOGHE_N tissues originated from the same individuals as the MOGHE_L tissues and carried the same variants at lower VAFs, it was appropriate that a reduced threshold be applied for analysis of proteomics data to identify broader DPs for subsequent analysis. Using a raw p-value of <0.05, a total of 97 proteins were identified to be differential between MOGHE lesional (MOGHE_L) and MOGHE normal (MOGHE_N) tissues (Figure 5.7). Of these, 82 were upregulated >1.0 LogFC and 15 were decreased by greater than >-1.0 LogFC in MOGHE_L samples. Only two proteins, GNG7 and NEFL, were detected as significantly increased in both MOGHE_L vs CONTROL_H and MOGHE_L vs MOGHE_N analyses (Figure 5.9).

Figure 5.5 [next page]: Heatmap of differential proteins identified in MOGHE cohort.
Heatmap 777 differential proteins (DPs) identified in pairwise comparisons of MOGHE lesional (MOGHE_L, indicated in lavender), MOGHE normal (MOGHE_N, honey), healthy control (CONTROL_H, ice-blue) and lesional control (CONTROL_P, leaf-green) tissue samples within proteomic data. Each column represents an individual sample and each row of the heatmap represents a different protein. Clustering analysis performed using TMT-Analyst online browser software (http://118.138.235.40:3838/TMT-Analyst/).

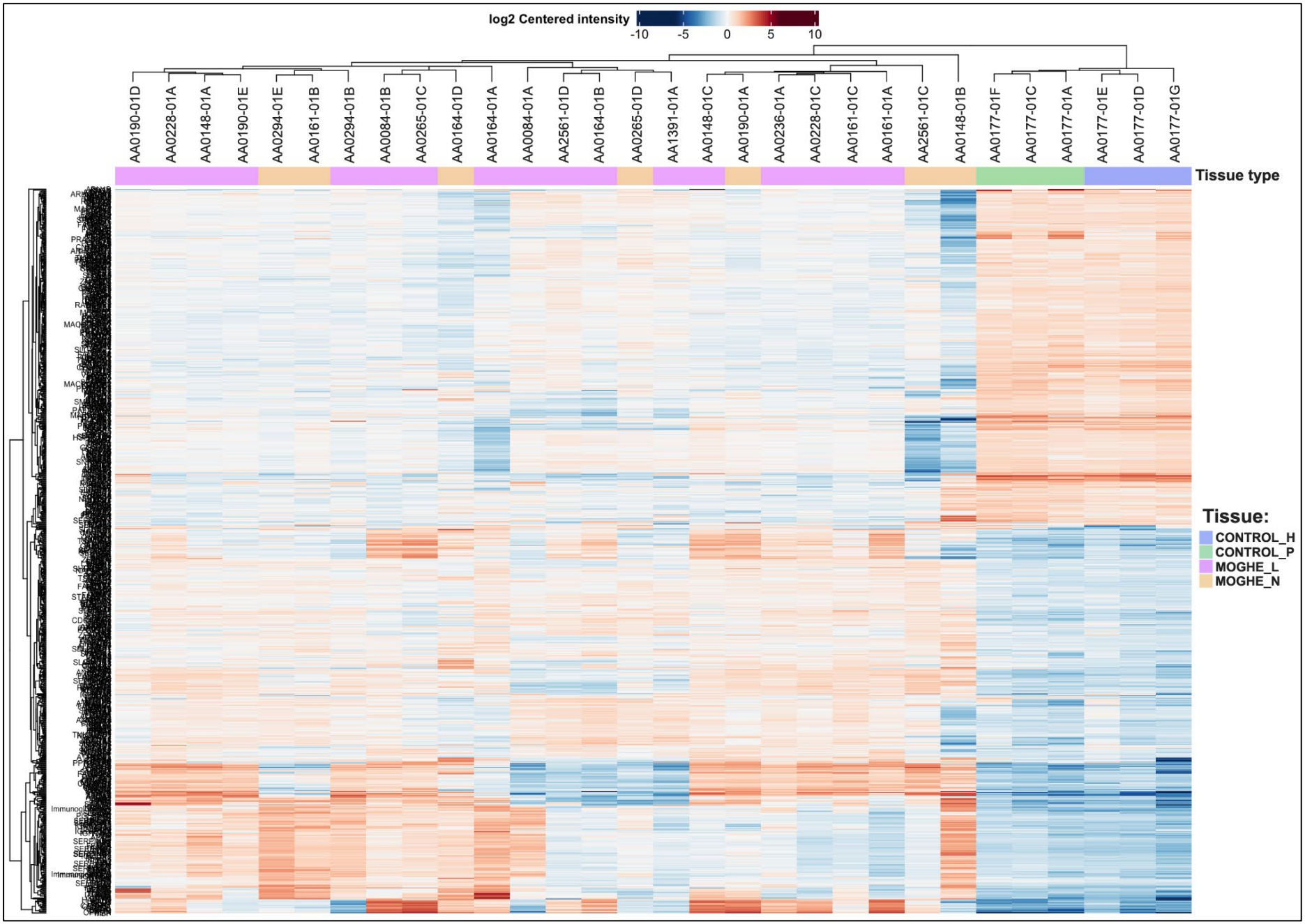

log2 Centered intensity
-10
-5
0
5
10
AA0190-01D
AA0228-01A
AA0148-01A
AA0190-01E
AA0294-01E
AA0161-01B
AA0294-01B
AA0084-01B
AA0265-01C
AA0164-01D
AA0164-01A
AA0084-01A
AA2561-01D
AA0164-01B
AA0265-01D
AA1391-01A
AA0148-01C
AA0190-01A
AA0236-01A
AA0228-01C
AA0161-01C
AA0161-01A
AA2561-01C
AA0148-01B
AA0177-01F
AA0177-01C
AA0177-01A
AA0177-01E
AA0177-01D
AA0177-01G
Tissue type
Tissue:
CONTROL_H
CONTROL_P
MOGHE_L
MOGHE_N

Figure 5.6: Volcano plot of differential proteins (DPs) identified between healthy control and MOGHE tissues within proteomic data.

A) Volcano plot of DPs between healthy control (CONTROL_H) and MOGHE lesional (MOGHE_L) tissues within proteomic data. Each dot represents one protein. Pink indicates proteins differentially abundant with a log2 fold change of <-2.0 and an adjusted p-value of less than 0.05 (i.e increased in MOGHE_L compared to CONTROL_H). Blue indicates proteins which were differentially abundant to a log2 fold change of >2.0 and an adjusted p-value of less than 0.05 (i.e decreased in MOGHE_L compared to CONTROL_H). Grey points indicate proteins which did not meet the threshold for significance. Percentages indicate percentage of imputed values. B) Volcano plot of DPs between CONTROL_H and MOGHE normal (MOGHE_N) tissues within proteomic data. Each dot represents one protein. Honey indicates proteins differentially abundant with a log2 fold change of <-2.0 and an adjusted p-value of less than 0.05. Blue indicates proteins which were increased to a log2 fold change of >2.0 and an adjusted p-value of less than 0.05. Grey points indicate proteins which did not meet this threshold. Percentages indicate percentage of imputed values. Volcano plot was created using TMT-Analyst online browser software (http://118.138.235.40:3838/TMT-Analyst/

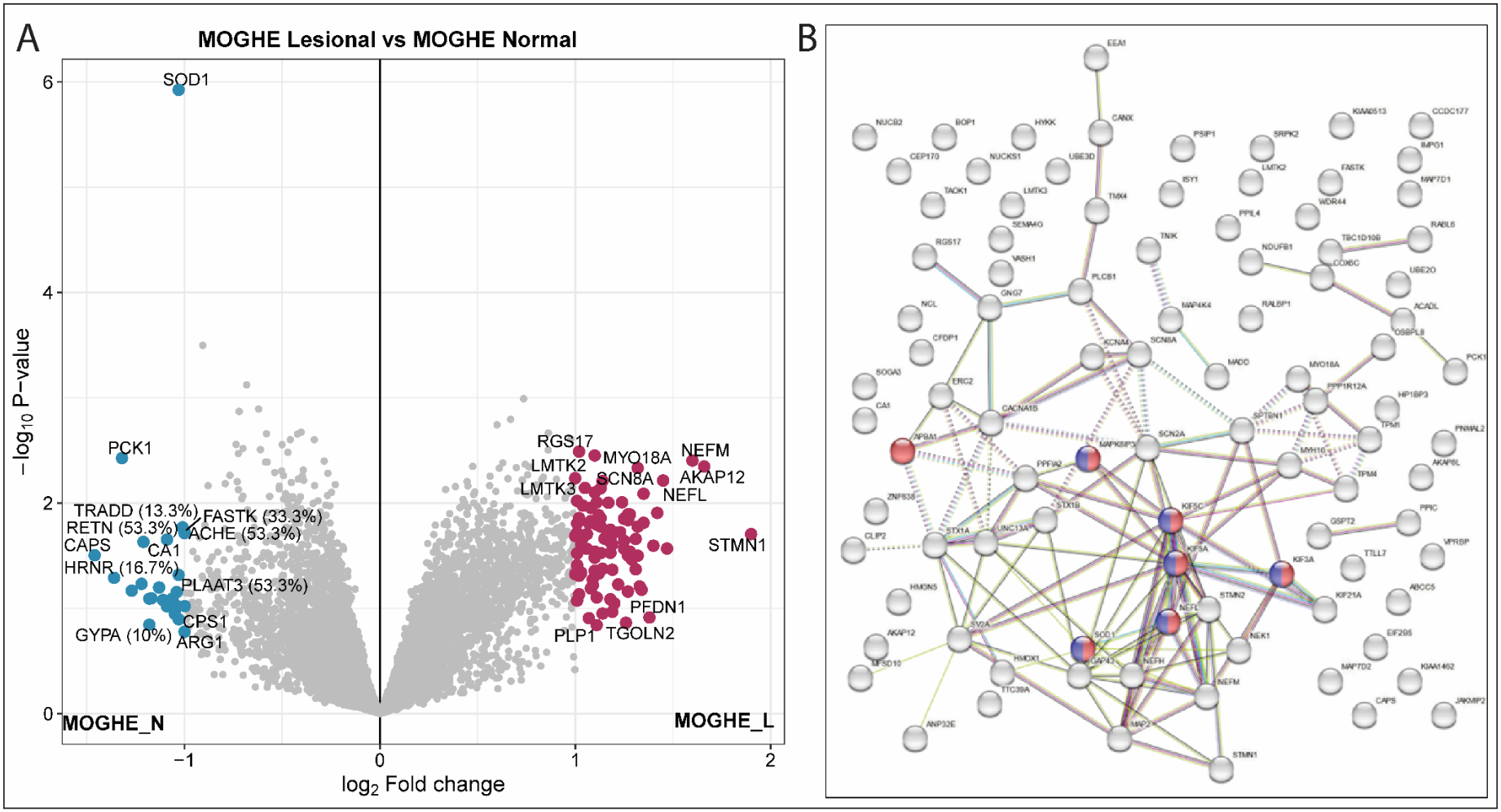


Figure 5.7: Differential proteins (DPs) identified between MOGHE lesional (MOGHE_L) and MOGHE normal (MOGHE_N) tissues within proteomic data. A) Volcano plot of DPs identified between MOGHE_L and MOGHE_N tissues within proteomic data. Each dot represents one protein. Plum coloured proteins indicate proteins differentially abundant with a log2 fold change of >1.0 and a raw p-value of less than 0.05. Teal indicates proteins differentially abundant with a log2 fold change of <-1.0 in lesional MOGHE tissues. Grey indicates proteins which did not meet this threshold. Percentages indicate percentage of imputed values. Volcano plot was created using TMT-Analyst online browser software (http://118.138.235.40:3838/TMT-Analyst/). B) String database network analysis of the top 97 DPs between MOGHE_L and MOGHE_N (https://version-11-5.string-db.org/cgi/network?networkId=bRi3Za3PyOuG).

5.3.5 Multi-omic analysis of dysregulated pathways in transcriptomic and proteomic data

The top dysregulated pathways identified between lesional and normal MOGHE cOPCs during snRNA-seq analysis were neurogenesis (GO: 0022008) (46 DEGs of 1290 genes in network, FDR = 8.13e-05 in snRNA-seq data), generation of neurons (GO: 0048699) (40 DEGs of 1131 genes in network, FDR = 0.00058 in snRNA-seq data), neuron differentiation (GO: 0030182) (38 DEGs of 1062 genes in network, FDR = 0.00064 in snRNA-seq data) and neuron development (GO:0048666) (32 DEGs of 840 genes in network, FDR = 0.0011 in snRNA-seq data). The major biological processes identified as dysregulated in the proteomic data between MOGHE_L and MOGHE_N tissues were microtubule-based process (GO: 0007017) (18 differential proteins of 803 proteins in network, FDR = 0.00047 in proteomics data), axo-dendritic transport (GO: 0008088) (7 DPs of 73 in network, FDR = 0.0005 in proteomics data) and anterograde axonal transport (GO:0008089) (6 DPs of 45 in network, FDR = 0.0005 in proteomics data). Two biological process GO terms were identified in both datasets, neuron projection development (GO:0031175) (41 total of 674 proteins in network, FDR = 5.21e-07 in full dataset) and cell morphogenesis (GO:0000902) (40 total of 708 proteins in network, FDR = 2.36e-06 in full dataset) (Table 5.3).

A further 21 protein coding genes were identified as DEGs in the snRNA-seq data that were also present in the proteomic data as differential between MOGHE_L and CONTROL_H tissue cohorts. A further five protein-coding genes were identified as DEGs in the snRNA-seq data and detected as DPs in the proteomics data analysis of MOGHE_L and MOGHE_N tissues. These were *SCN2A* (cOPC LogFC = -1.034), *MAP4K4* (cOPC LogFC = 0.543), *SPTBN1* (cOPC LogFC = 0.735), *UBE2O* (cOPC LogFC = 0.879) and *AKAP8L* (cOPC LogFC = 1.099). However, these proteins were not significant between CONTROL_H and MOGHE_L or MOGHE_N tissues. All 26 of these protein-coding genes were detected as DEGs in cOPCs only. The major dysregulated pathways common to the proteomics and snRNA-seq analyses primarily involved axon and

dendrite development, neurogenesis and cytoskeletal formation (Figure 5.8; Figure 5.11).

DEGs that were sustained at the proteomic level were prioritised. Although SEMA3E was not identified as a DP, PLXND1, a close interactor of SEMA3E was identified and SEMA3E remained a candidate due to its roles in dysregulated molecular pathways. Curation of multi-omic analysis data resulted in the identification of 11 potential biomarkers of interest; ACTB, AP3B2, GNG7, KIF5C, MAP2, MAP4, MOBP, MOG, NEFL, SEMA3E and SOD1 (Figure 5.9). Co-expression analysis of these top 11 protein coding genes of interest identified in proteomics and snRNA-seq data revealed an association between *ACTB*, *AP3B2*, *KIF5C*, *MAP2*, *MAP4*, *MOBP*, *MOG*, *NEFL* and *SOD1* in the MOGHE_Lesional Mixed cell population (Figure 5.10).

Combined analysis of single-nuclei expression data, protein abundance and GO term pathway activity of the 11 top protein coding genes of interest identified six proteins – ACTB, GNG7, MAP4, NEFL, SEMA3E and SOD1 – as possible biomarkers for MOGHE.

Figure 5.8: Differential genes, proteins, and dysregulated pathways in snRNA-seq and proteomics data. (A) Area graph of 302 differentially expressed genes (DEGs) identified between MOGHE_Lesional and MOGHE_Normal tissues in snRNA-seq data. LogFC along the x-axis. Colour indicates the nuclei cluster that the DEG was identified in. (B) Map of dysregulated pathways identified from the 302 DEGs in snRNA-seq data. (C) Area graph of 97 differential proteins (DPs) identified in proteomic data between MOGHE_L and MOGHE_N tissues. LogFC along the x-axis. D) Map of dysregulated pathways identified from the 97 DPs in proteomics data. Dysregulation maps created with reactome (https://reactome.org/PathwayBrowser/#/).

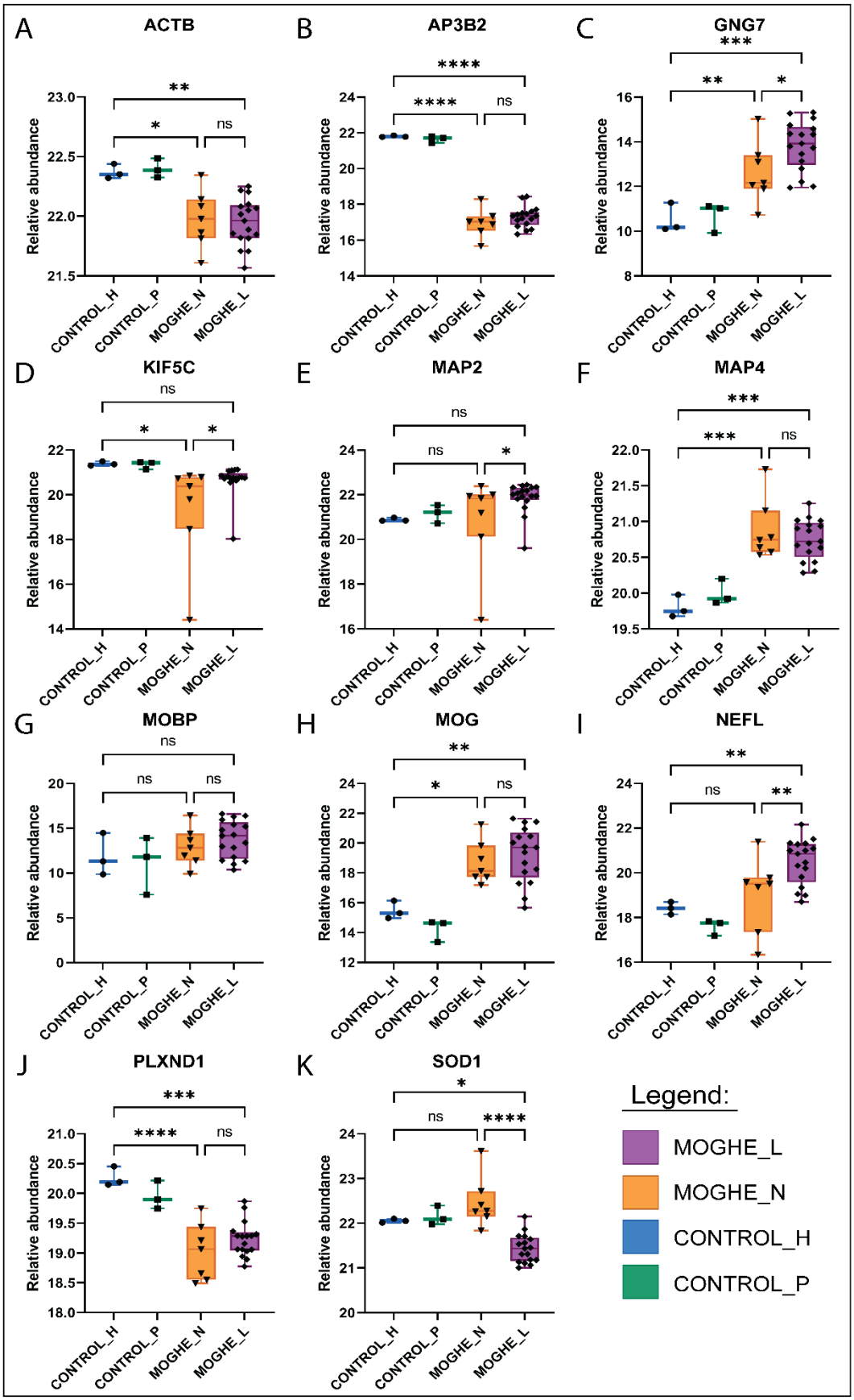


Figure 5.9: Protein levels in control and MOGHE tissues.

Box plots of proteins levels in CONTROL_H, CONTROL_P, MOGHE_N and MOGHE_L tissues for ACTB (A), AP3B2 (B), GNG7 (C), KIF5C (D), MAP2 (E), MAP4 (F), MOBP (G), MOG (H), NEFL (I), SEMA3E (J) and SOD1 (K) as detected by TMT-16pro LC-MS/MS analysis. Error bars indicate minimum and maximum. Statistical test performed was one-way analysis of variance (ANOVA) with multiple comparisons. ns = not significant, * = p<0.05, ** = p<0.01, *** = P<0.005, **** = p<0.001.

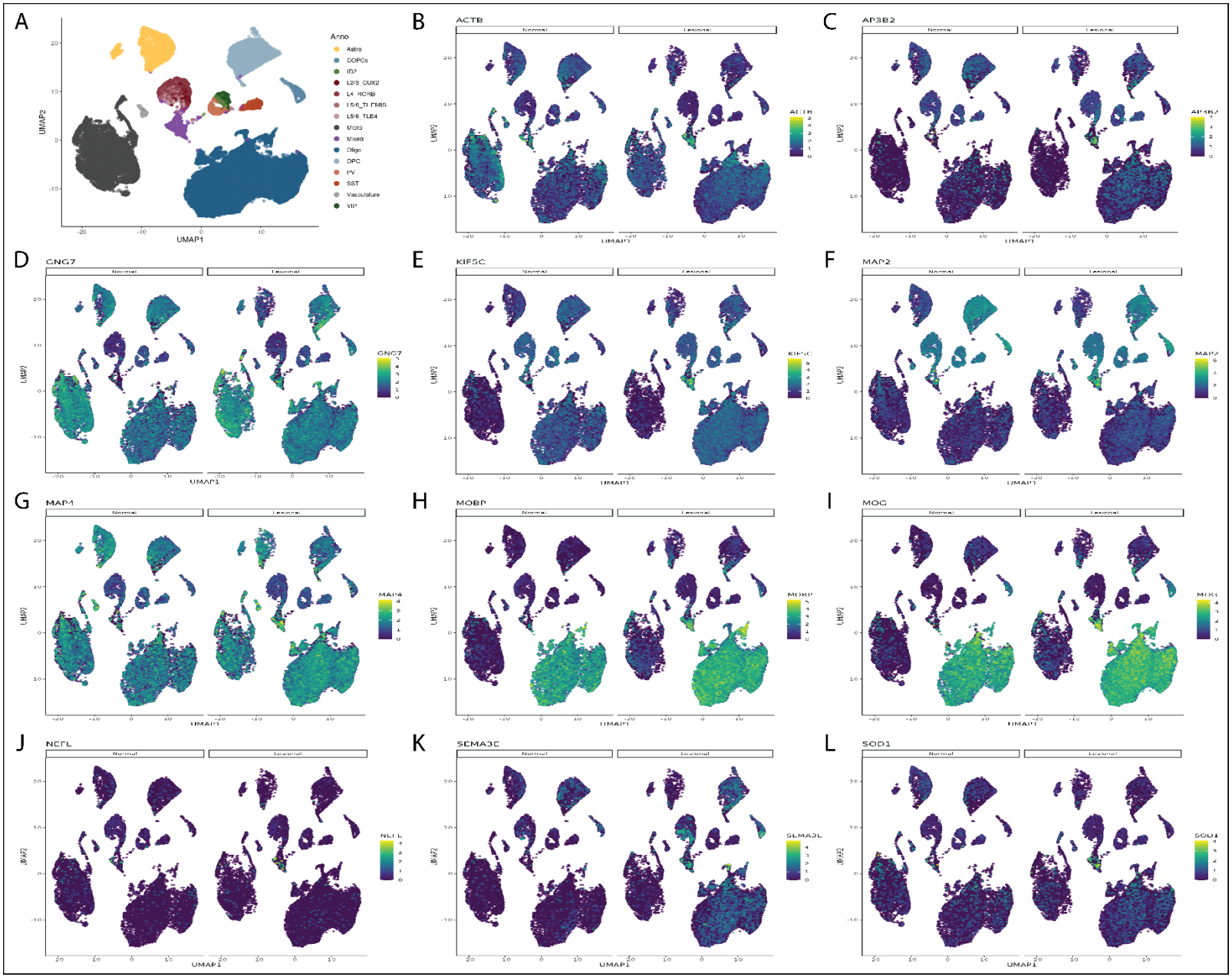


Figure 5.10: Biomarker expression in snRNA-seq dataset.

A) Uniform manifold approximation and projection (UMAP) graph of snRNA-seq data with annotated nuclei clusters labelled. Purple indicates Mixed population. B-I) Side-by-side UMAPs showing heatmap expression in MOGHE_Normal (left) and MOGHE_Lesional (right) tissues of *ACTB* (B), *AP3B2* (C), *GNG7* (D), *KIF5C* (E), *MAP2* (F), *MAP4* (G), *MOBP* (H), *MOG* (I), *NEFL* (J), *SEMA3E* (K) and *SOD1* (L). Blue indicates nuclei with low expression. Yellow indicates high expression.

Table 5.3: GO term analysis with shortlisted biomarkers listed.

GO term ID	Term description	Analysis identified via	Observed gene count			Background gene count	Strength	FDR	Biomarkers present
			snRNAseq	Proteomics	Total				
GO:0022008	Neurogenesis	T	46	15	61	1290	0.41	4.03E-07	ACTB, KIF5C, MAP2, MAP4, NEFL, SEMA3E, SOD1
GO:0031175	Neuron projection development	T/P	27	15	41	674	0.52	5.21E-07	ACTB, KIF5C, MAP2, MAP4, NEFL, SEMA3E, SOD1
GO:0048666	Neuron development	T	32	14	46	840	0.47	6.05E-07	ACTB, KIF5C, MAP2, MAP4, NEFL, SEMA3E, SOD1
GO:0030182	Neuron differentiation	T	38	14	52	1062	0.43	1.15E-06	ACTB, KIF5C, MAP2, MAP4, NEFL, SEMA3E, SOD1
GO:0048699	Generation of neurons	T	40	14	54	1131	0.41	1.15E-06	ACTB, KIF5C, MAP2, MAP4, NEFL, SEMA3E, SOD1
GO:0000902	Cell morphogenesis	T/P	28	13	40	708	0.49	2.36E-06	ACTB, KIF5C, MAP2, NEFL, SEMA3E, SOD1
GO:0016043	Cellular component organization	P	102	52	154	5436	0.19	2.42E-06	ACTB, AP3B2, KIF5C, MAP2, MAP4, NEFL, SEMA3E, SOD1
GO:0022603	Regulation of anatomical structure morphogenesis	T	32	14	46	920	0.43	3.68E-06	MAP2, NEFL, SEMA3E
GO:0061564	Axon development	T	19	9	28	402	0.58	8.53E-06	ACTB, KIF5C, NEFL, SEMA3E
GO:0007399	Nervous system development	T	60	19	79	2188	0.29	8.53E-06	MAP4, MOBP, MOG
GO:0120036	Plasma membrane bounded cell projection organization	P	31	18	49	1112	0.38	2.95E-05	ACTB, KIF5C, MAP2, MAP4, NEFL, SEMA3E, SOD1
GO:0048667	Cell morphogenesis involved in neuron differentiation	P	18	10	28	438	0.54	3.22E-05	ACTB, KIF5C, MAP2, NEFL, SEMA3E
GO:0048812	Neuron projection morphogenesis	P	17	12	29	477	0.52	4.39E-05	ACTB, KIF5C, MAP2, NEFL, SEMA3E
GO:0032989	Cellular component morphogenesis	P	19	13	32	603	0.46	0.00015	ACTB, KIF5C, MAP2, NEFL, SEMA3E

GO:0048731	System development	T	86	26	112	3867	0.2	0.00022	ACTB, KIF5C, MAP2, NEFL, SEMA3E, SOD1
GO:0008088	Axo-dendritic transport	P	4	7	11	73	0.91	0.00023	AP3B2, KIF5C, NEFL, SOD1
GO:0008089	Anterograde axonal transport	P	3	6	9	45	1.04	0.00031	KIF5C, NEFL, SOD1
GO:0030705	Cytoskeleton-dependent intracellular transport	P	9	8	17	200	0.67	0.00031	AP3B2, KIF5C, NEFL, SOD1
GO:0007010	Cytoskeleton organization	P	29	20	49	1229	0.34	0.00031	ACTB, MAP2, MAP4, NEFL, SOD1
GO:0007275	Multicellular organism development	T	92	26	118	4209	0.18	0.00036	ACTB, KIF5C, MAP2, MAP4, MOBP, MOG, NEFL, SEMA3E, SOD1
GO:0050793	Regulation of developmental process	T	60	17	77	2492	0.23	0.0021	ACTB, MAP2, NEFL, SEMA3E, SOD1
GO:0007264	Small GTPase mediated signal transduction	T	15	3	18	273	0.55	0.0031	none
GO:0006996	Organelle organization	P	62	36	98	3470	0.19	0.0031	ACTB, AP3B2, KIF5C, MAP2, MAP4, NEFL, SOD1
GO:0048856	Anatomical structure development	T	105	26	131	5117	0.14	0.0061	ACTB, KIF5C, MAP2, MAP4, MOBP, MOG, NEFL, SEMA3E, SOD1
GO:0032502	Developmental process	T	111	30	141	5657	0.13	0.0081	ACTB, KIF5C, MAP2, MOBP, MOG, NEFL, SEMA3E, SOD1
GO:0007017	Microtubule-based process	P	15	18	33	803	0.35	0.0084	AP3B2, KIF5C, MAP2, MAP4, NEFL, SOD1
GO:0099504	Synaptic vesicle cycle	P	5	6	11	125	0.68	0.0123	ACTB, AP3B2
GO:0016081	Synaptic vesicle docking	P	1	3	4	10	1.34	0.0245	none
GO:0033693	Neurofilament bundle assembly	P	0	3	3	3	1.74	0.0288	NEFL
GO:0016482	Cytosolic transport	P	5	6	11	147	0.61	0.0324	none

FDR, false discovery ratio; GO, Gene Ontology; P, proteomics; T, transcriptomics.

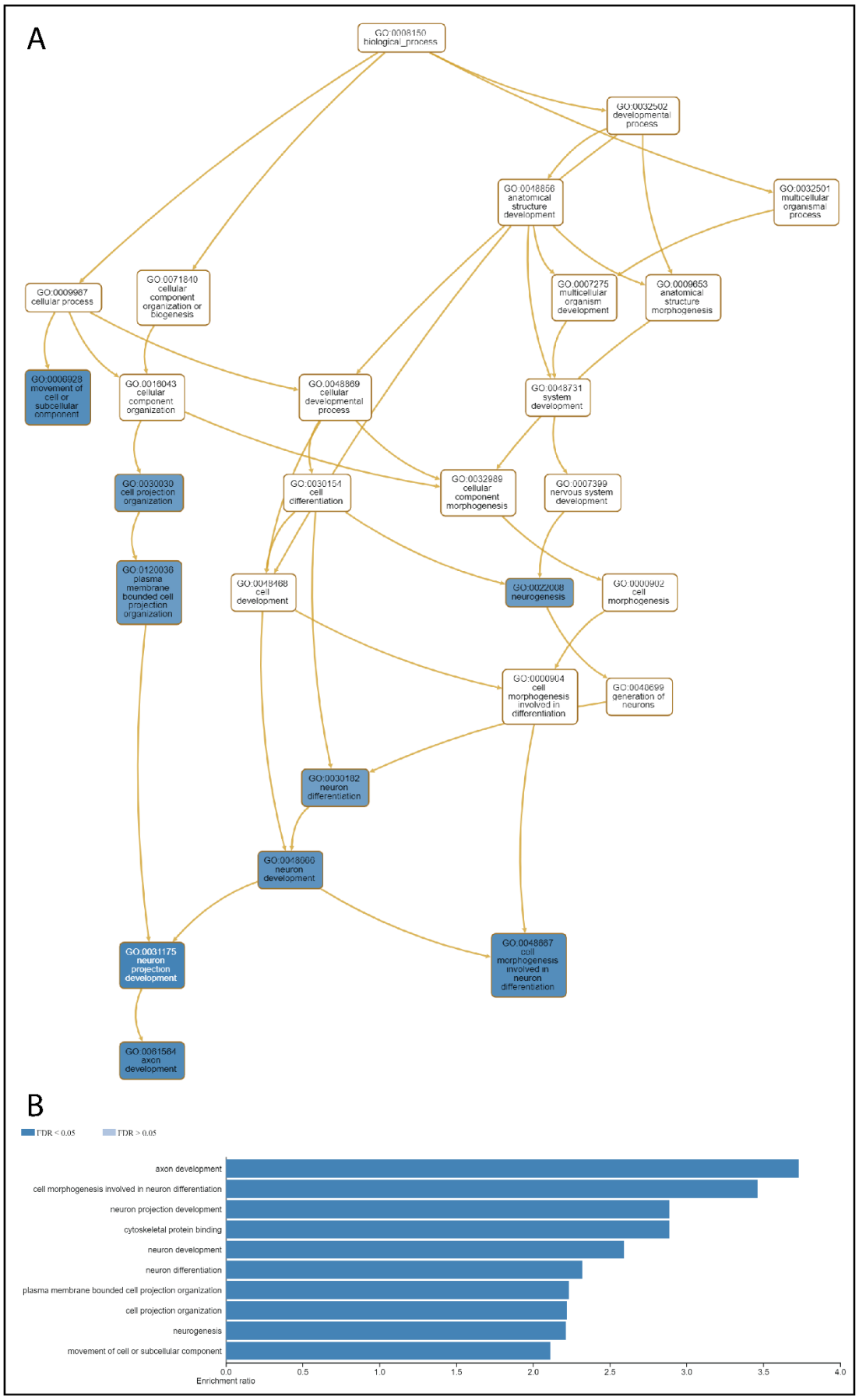


Figure 5.11: Combined KEGG and GO term overlap analysis of top dysregulated pathways in single nuclei transcriptomic data and proteomic data.

A) Directed acyclic graph (DAG) showing relationships among the GO terms associated with differentially expressed gene set and differential protein set identified between MOGHE lesional and MOGHE normal tissues. B) Bar graph of enriched GO terms. Over representation analysis (ORA) with enrichment categories of Gene Ontology biological process and molecular function. Created using the Webgestalt tool (https://webgestalt.org/).

5.4 Discussion

This study presents the first single cell dataset derived from MOGHE brain tissues and the first high-resolution proteomic dataset of MOGHE brain specimens derived from epilepsy surgery. The snRNA-seq dataset contains 51,071 single-nuclei transcriptomes and the proteomic profiles constitute 7,655 consistently quantified proteins originating from MOGHE tissues. The findings of this study have broad reaching research and clinical outcomes as they assist to elucidate the mechanisms of brain somatic *SLC35A2* disruption and highlight future pathways of research focus and clinical targets.

5.4.1 Single nuclei RNA-sequencing reveals cell types of interest in MOGHE

Comparison between the seven lesional MOGHE snRNA-seq samples (MOGHE_Lesional) and the four healthy MOGHE snRNA-seq samples (MOGHE_Normal) identified 302 DEGs between nuclei from the two tissue types, of which six were common between two cell populations and 296 were unique. These 302 DEGs consisted of 251 DEGs in cOPCs, 45 DEGs in OPCs, five DEGs in astrocytes and one DEG in layer 5/6 *THEMIS*-expressing neurons. Two hundred and ninety-one unique DEGs were identified between lesional and normal MOGHE tissues across cOPCs and OPCs, including five that were observed as significantly dysregulated in both cOPCs and OPCs. These cell types are inherently linked. OPCs begin dividing and become cOPCs which eventually mature into oligodendrocytes. However, these DEGs are only observed in the earlier stages of cell development and are no longer retained by the time these cells form mature oligodendrocytes. Due to their relation to mature oligodendrocytes and the large numbers of DEGs in these cell clusters, these cells became cells of interest.

Layer 5/6 *THEMIS*-expressing neurons are glutamatergic, excitatory neurons. *GRB14* was the only DEG identified in this nuclei cluster and was downregulated in lesional MOGHE tissue. *GRB14* encodes Growth Factor Receptor Bound Protein 14 which has a role in signalling pathways and forms a major component of the miR-4709-3p/GRB14/PDGFRα pathway. It is a known modulator of insulin signalling and glucose uptake [282]. Platelet-derived growth factor receptor-alpha (PDGFRα) has been identified as a marker of putative OPCs in MOGHE in an early study [112]. The relationship between GRB14 and PDGFRα may be indicative of the relationship between the clustering of Olig2-positive cells and heterotopic neurons in the white matter of MOGHE brain tissue. The Layer 5/6 *THEMIS*-expressing neuron nuclei population could be the heterotopic neurons observed in the immunohistochemical staining. Immunofluorescence co-staining of MOGHE tissues with THEMIS, GRB14 and PDGFRα or Olig2 antibodies may assist to spatially resolve these neurons in tissue. Spatial resolution may identify the location of these Layer 5/6 *THEMIS*-expressing neurons relative to oligodendrocyte-rich white matter or identity this population as the heterotopic neurons observed in white matter in MOGHE. This would be an appropriate next step for validation of these findings and potential identification of the heterotopic neurons in future studies.

Alternatively, the identity of the Mixed population is unclear. The lesional Mixed nuclei cluster could indicate disease-specific cells in MOGHE and may include the heterotopic neurons observed in MOGHE. The Mixed nuclei cluster was a hotspot for upregulation of *ACTB*, *AP3B2*, *KIF5C*, *MAP2*, *MAP4*, *MOBP*, *MOG*, *NEFL* and *SOD1*, particularly in lesional tissue (Figure 5.10). Spatial resolution with one or more of these markers may assist to determine whether this population represents heterotopic neurons. SOD1 would pose a particularly good validation candidate for the Mixed cell population due to its global downregulation in MOGHE lesional tissues and upregulation in the Mixed cell population.

5.4.2 Proteomic profile of MOGHE brain tissue

There were 777 differential proteins (DPs) identified between healthy control (CONTROL_H) and all MOGHE (normal, MOGHE_N / lesional, MOGHE_L) tissues. MOGHE profiles were distinct from MCD controls (CONTROL_P) and healthy controls (CONTROL_H). The heatmap of the 777 DPs shows a clear distinction between the molecular profiles of MOGHE (MOGHE_N and MOGHE_L) tissues and control (CONTROL_H and CONTROL_P) tissues (Figure 5.5). The two groups, MOGHE and CONTROL clearly segregated during analysis as represented by the dendrogram, demonstrating the unique proteomic profile of MOGHE. No significant association was identified between MOGHE_N tissues with a low *SLC35A2* VAF (mean VAF = 0.99%, median VAF = 0.63%) and MOGHE_L tissues with a high VAF (mean VAF = 17.4%, median VAF = 11.4%). Proteins that were detected as dysregulated in MOGHE tissues overall were only significantly greater in MOGHE_L tissues than MOGHE_N tissues in GNG7 and NEFL (Figure 5.9). Although 97 DPs and 301 DEGs were identified between the two groups overall, these two cohorts did not completely segregate in the snRNA-seq UMAP or the protein abundance heatmap (Figure 5.1B, Figure 5.5). Likewise, while an overrepresentation of oligodendrocytes was observed in lesional MOGHE samples in snRNA-seq data overall, this was not consistent across the cohort at an individual sample level. Similar to trends observed in the proteomics data, though differences were observed between MOGHE_Lesional (high VAF) and MOGHE_Normal (low VAF) tissues, VAF in the snRNA-seq data did not appear to correlate closely with oligodendrocyte capture proportion.

5.4.3 Neurogenesis and axo-dendritic development are dysregulated in MOGHE

The top dysregulated pathways identified between MOGHE_Lesional and MOGHE_Normal cOPCs during snRNA-seq analysis were neurogenesis, axo-dendritic transport, neuron projection development, neuron development and generation of neurons (Table 8.5). Notably, multiple genes and respective proteins in these pathways were detected as being dysregulated. Neurons are the primary epileptogenic cells in

epilepsy and dysfunctional neuronal subtypes are believed to underly seizure activity [283]. Neurogenesis, axon and dendrite development, and myelination are all likely affected in MOGHE and may contribute to seizure generation. Dysregulation of axon development and myelination could result in the disruption of neuronal circuitry and signalling [284, 285]. Interestingly, the top dysregulated pathways were closely associated with neuron generation, migration and myelination but only one DEG was actually identified in a neuronal population. This is likely due to the fact that many of the DEGs identified in OPCs and cOPCs related to signalling factors and proteins related to externally influencing neurogenesis and migration. DPs involved in these processes may be candidates for biomarkers of disease in MOGHE.

5.4.4 Multi-omics analyses enable identification of potential biomarkers of disease

SOD1 was identified as the only DP between MOGHE_L and MOGHE_N tissues at a LogFC greater than 2.5 (adj. p-value of <0.001). Copper/zinc superoxide dismutase 1 (SOD1) is an antioxidant enzyme, encoded by *SOD1*. SOD1 is involved in apoptosis and oxidative stress response. In the brain, SOD1 protects neurons from damage acquired from build-up of reactive oxygen species generated from normal cell function [286]. Pathogenic variants in the *SOD1* gene are associated with amyotrophic lateral sclerosis 1 (ALS) (OMIM: 105400) and progressive spastic tetraplegia and axial Hypotonia (STAHP), and disruption of SOD1 activity is known to affect motor neuron function [286, 287]. Deficiency of SOD1 has also been implicated in oxidative stress response in peripheral nervous system and has been reported to cause a progressive distal motor axonopathy [288].

Interestingly, SOD1 was identified as decreased in MOGHE tissues through the proteomics data but was identified as increased in the Mixed nuclei population in the snRNA-seq analysis. It is known that transcriptomic levels do not always accurately reflect proteomic abundance in tissues [289]. Factors such as molecular half-lives and PTMs may influence mRNA transcript expression and protein abundance. Additionally,

the Mixed population is very small and only accounts for 4.2% cells (1225) in MOGHE_Lesional tissues and 1.6% (370 nuclei) in MOGHE_Normal tissues. These Mixed cells are rare cell types and it's possible that the differences between the Mixed cell populations in lesional and normal tissues would not be as readily apparent in the overall dataset at the proteomics level. The increase of *SOD1* expression in these cells may not be sustained at total protein level. However, the fact that SOD1 is decreased overall but *SOD1* mRNA is highly expressed in only the Mixed nuclei population further suggests that these are cells of interest in MOGHE and may be crucial to mechanisms of pathogenesis. SOD1 is involved in many of the top dysregulated pathways identified in this dataset and has been identified as a key potential biomarker for MOGHE.

Similarly, Semaphorin 3E (*SEMA3E*) was dysregulated in both cOPCs (LogFC = 2.389) and OPCs (LogFC = 2.927) and was present in 14/17 top dysregulated pathways identified in the transcriptomic data and 18/30 dysregulated pathways identified overall (Table 5.3). *SEMA3E* was the top dysregulated gene common to both cOPCs and OPCs. SEMA3E plays a crucial role in axon guidance, neuronal development, and the formation of neural circuits in the brain [290]. Neurons migrate during brain development from their site of neurogenesis to their final laminar destination to form the various brain structures. SEMA3E is known to influence neuronal migration by repelling migrating neurons, guiding them to their correct positions. By creating repulsive barriers, SEMA3E helps establish the layered organization of the brain. SEMA3E is also known to function as a chemorepellent, directing axon growth and repulsively guiding extension to establish connectivity and circuitry of neural circuits in the developing brain. SEMA3E is closely related to Plexin D1 (PLXND1) and together, these proteins have been shown to play a role in determining laminar positioning of heterotopically projecting neurons in the corpus callosum [291]. Furthermore, SEMA3E has been shown to be expressed in the marginal zone and subplate of the developing cortex [290]. These regions are important for correct establishment of cortical layers and neuronal migration. The levels of overexpression of *SEMA3E* observed in this MOGHE dataset may contribute to the presence of heterotopic neurons observed in

the white matter of MOGHE tissues. As established through the pathway analyses, neurogenesis, neuronal migration and axon development and guidance are major dysregulated pathways in MOGHE. *SEMA3E* is a major contributor to these pathways and was upregulated in both cOPCs and OPCs in lesional MOGHE tissues.

Another two proteins of interest identified in proteomic data were microtubule-associated protein 2 (MAP2) and 4 (MAP4). MAP2 and MAP4 are involved in stabilizing and organising essential components of neuronal cytoskeleton called microtubules [292, 293]. They regulate the structure and plasticity of dendrites including growth, development and maintenance and neurite morphology [294]. These proteins are involved in several of the major dysregulated pathways identified in MOGHE and interact closely with several other dysregulated proteins identified between MOGHE_L and MOGHE_N samples including neurofilaments, NEFL, NEFM and NEFH [295], and the kinesin motor proteins, KIF5A and KIF5C [296, 297]. Collectively, these proteins have critical roles in neuronal development, maintenance, and synaptic function. MAP2, MAP4 and KIF5C are involved in 20, 13 and 21 of the top 30 dysregulated pathways identified in the GO term analysis, respectively (Table 5.3). Dysregulation of these proteins can lead to impaired intracellular transport and contribute to neurological disorders and neurodegenerative diseases. MAP4 was significantly increased in both MOGHE_L and MOGHE_N tissues compared to CONTROL_H tissues (Figure 5.9) and was most highly expressed in lesional Mixed cell population in the snRNA-seq data (Figure 5.10). MAP4 is present in the dendritic spines and dendrites of neurons [293]. Consequently, MAP4 was identified as a biomarker of interest in MOGHE. However, the observed increase in MAP2 and decrease in KIF5C were not sustained to a significant level between MOGHE_L and CONTROL_H tissues (Figure 5.9). Furthermore, MAP2 is commonly used in immunolabelling studies to identify neurons due to its role in neuronal morphology and dendritic development, in addition to its localisation in dendrites and neuronal cell bodies and thus, does not serve as an optimal biomarker for MOGHE. Dysregulated MAP2, and its interactor KIF5C, are likely

involved in MOGHE pathogenesis but MAP2 and KIF5C are not reliable global biomarkers for MOGHE.

Other dysregulated proteins include neurofilament light chain (NEFL) which was dysregulated in both MOGHE_L vs CONTROL_H analysis and MOGHE_L vs MOGHE_N analysis. NEFL is primarily expressed in axons and provides structural support to neurons. It forms part of the neurofilament complex with neurofilament medium (NEFM) and heavy (NEHM) chains and regulates axon diameter. NEFL is present in all three of the top dysregulated pathways identified in the proteomics data, microtubule-based process, axodendritic transport and neuron projection development, and 25/30 dysregulated pathways identified overall (Table 5.3). NEFL dysregulation has been associated with disruption of SOD1, the top downregulated protein identified in the proteomics analysis [298]. NEFL has previously been reported as disrupted in other MCD and has been identified as a biomarker of neurodegenerative diseases previously. NEFL levels in cerebrospinal fluid (CSF) and blood have been found to be altered in several neurodegenerative diseases, including Alzheimer's disease, Parkinson's disease, CLN3-Batten disease and ALS – the latter condition being strongly linked to SOD1 [286, 299, 300]. An increase in *NEFL* expression has been associated with axonal disrepair and neurodegeneration [301, 302]. While the *NEFL* signal in the snRNA-seq data was not strong enough to validate this gene as a potential biomarker for MOGHE, the broad and current use of NEFL as a biomarker in other disorders limits its clinical utility in MOGHE. However, NEFL remains a gene of interest and its application in conjunction with other biomarkers identified in MOGHE may prove useful as a screening tool to potentially aid in the identification of individuals with MOGHE.

ACTB encodes Beta-Actin (β-actin), a protein crucial for actin cytoskeletal development within cells in the brain. Actin filaments in neurons provide structural support for dendrites, axons and synapses and influence neuronal migration and axon guidance [303]. ACTB interacts closely with many of the other proteins identified in the proteomics analysis of MOGHE_N and MOGHE_L tissues and is central to the three

major dysregulated pathways in lesional MOGHE tissues, identified from the proteomics data. *ACTB* was downregulated in cOPCs and was detected as significantly decreased in both MOGHE_N and MOGHE_L tissues compared to controls. Although *ACTB* was downregulated in cOPCs, it was highly expressed by the MOGHE_Lesional Mixed nuclei population in the snRNA-seq data, suggesting that dysregulated ACTB may be involved in the processes leading to generation of the Mixed nuclei population. The single-nuclei transcriptomic data and proteomic data demonstrate that axon development and neuronal differentiation and migration are impacted in MOGHE. The role of ACTB in these processes further suggests dysfunction of ACTB may be involved in MOGHE pathogenesis. Dysregulation of axon development may disrupt neuronal circuits and signalling, potentially contributing to the seizure phenotype observed in MOGHE. A human induced pluripotent stem cell (iPSC) model of MOGHE would provide a strong candidate to validate disruption of axon development in MOGHE.

G Protein Subunit Gamma 7 (GNG7) was one of the few proteins which was identified as significantly increased in all MOGHE tissues (MOGHE_L and MOGHE_N) compared with control tissues, and further significantly increased in MOGHE_L tissues compared with MOGHE_N tissues (Figure 5.9). GNG7 was not detected as a DEG in the snRNA-seq data. GNG7 belong to the family of G protein subunits and has a crucial role in transmembrane signalling in cells [304]. GNG7 has been previously associated with cancers in the breast, lung, head, neck and oesophagus, and specifically identified as a biomarker in gastric and pancreatic cancers [304, 305]. Mouse model studies of *Gng7* in brain identified regional expression in the midbrain, hippocampus, striatum, cortex of mice and suggested a role in the signal transduction pathway of striatal GABAergic neurons [306]. Though the role of GNG7 in the human brain is not well understood, it is clearly a robust protein marker of MOGHE and warrants further investigation. *GNG7* expression should be evaluated in disease models of MOGHE in future studies.

The data suggests *ACTB*, *GNG7*, *MAP4*, *NEFL*, *SEMA3E* and *SOD1* are genes of interest in this disorder and dysregulation of these genes may serve as a potential biomarker

for MOGHE in the future. Multi-omic analysis of MOGHE tissues led to the identification of six potential biomarkers of disease; ACTB, GNG7, MAP4, NEFL, SEMA3E and SOD1.

Although both transcriptomic and proteomic approaches were informative in identifying candidates in isolation, the combination and overlap between single nuclei transcriptomic analysis and proteomic analysis provided the most effective method for identifying robust dysregulated pathways and potentially identifying appropriate biomarkers of disease in MOGHE. Multi-omic analyses assisted in filtering and removal of other biomarker candidates. For example, Adaptor Protein 3 Beta 2 subunit (AP3B2) is a subunit of the Adaptor Protein 3 (AP-3) complex. The AP-3 complex is involved in intracellular vesicle trafficking and membrane trafficking pathways. Disruption of AP3B2 activity is associated with another form of severe epilepsy of infancy, DEE48 (OMIM: 617276), and as such, AP3B2 was a strong candidate biomarker for MOGHE. AP3B2 was identified as decreased significantly in MOGHE_N and MOGHE_L tissues compared to controls. However, this decrease was not detected at the transcriptomic level and the role of AP3B2 was peripheral to the major dysregulated pathways identified in subsequent analyses. As such, AP3B2 was considered insufficiently robust to be utilised as a biomarker of MOGHE.

Galactose supplementation is currently being trialled in *SLC35A2*-mosaic MOGHE cases with partial success but there is currently no cure for MOGHE [274]. The severity of the disease necessitates the identification of clinically valid biomarkers and molecular targets for treatment development. ACTB, GNG7, MAP4, NEFL, SEMA3E and SOD1 may pose additional druggable targets in MOGHE which could be an alternative to current methods and may, in the future pose an alternative to surgery in some cases.

5.4.5 Oligodendrocyte developmental markers are inappropriate biomarkers for MOGHE

Oligodendrocytes have not been identified as a cell type associated with epilepsy and seizure generation or propagation, so it is important to understand the potential

effects of oligodendrocyte proliferation on neuronal function and disease pathogenesis in MOGHE. *MOBP*, encoding myelin-basic oligodendrocyte basic protein, was the most significantly upregulated gene identified in cOPCs (LogFC = 5.145, adj. p = 0.005). MOBP is a crucial protein involved in myelin formation, maintenance, stability, and organization in the central nervous system (CNS). Myelin sheath is a specialized membrane that wraps around neuronal axons, providing insulation and facilitating efficient signal transmission. MOBP is expressed in developing oligodendrocytes and is involved in their maturation and differentiation. It promotes the formation of myelin internodes and the proper wrapping of myelin membranes around axons [307, 308]. MOBP further regulates axonal diameter, plays a role in myelin-associated signalling, and drives oligodendrocyte development. Dysregulated MOBP has been shown to disrupt myelin structure and function, leading to impaired neuronal signalling and neurological symptoms, particularly in multiple sclerosis (MS) [309].

Closely related to MOBP, MOG or myelin oligodendrocyte glycoprotein is found primarily in the myelin sheath in the brain and throughout the CNS and was identified as increased in MOGHE tissues compared to control tissues but was not identified in transcriptomic data. MOG has been implicated in disorders such as MS, acute disseminated encephalomyelitis (ADEM) and seizures, encephalitis, anti-aquaporin-4-antibody (AQP4-Ab)-seronegative neuromyelitis optica spectrum disorder (NMOSD) [310, 311]. It is already widely used as a biomarker for these disorders [312]. However, neither MOBP nor MOG were identified as differentially abundant upon comparative proteomic analysis of MOGHE lesional (MOGHE_L) and normal (MOGHE_N) tissues. Therefore, the identification of increased MOBP and MOG in the lesional MOGHE tissues within the CONTROL_H/MOGHE_L analysis (Figure 5.6) is likely due to the overrepresentation of oligodendrocytes present in MOGHE tissues and does not necessarily indicate dysregulation of these proteins as causal or a primary manifestation of MOGHE pathogenesis.

Though both MOBP and MOG are highly dysregulated in MOGHE and play a potential role in MOGHE pathogenesis, they are unlikely to serve as valuable biomarkers of the condition as they are generic markers of OPC, cOPC and oligodendrocyte activity and already serve as clinical biomarkers of MS, ADEM and NMOSD [312]. Furthermore, these genes are included in the current list of markers used for snRNA-seq annotation. Therefore, an unbiased and independent analysis would be required to accurately gauge their clinical utility as a biomarker of MOGHE. While these genes and their respective proteins remain candidates of interest in MOGHE due to their well-established association with oligodendrocyte function, they are not considered clinically viable biomarkers for MOGHE at this time.

Furthermore, the broader DEGs identified in the snRNAseq data suggest cell differentiation and localisation may be disrupted in cortex development. This may relate to the increase in Olig2-positive cells observed in MOGHE and suggest disruption at an earlier stage of development and oligodendrocyte cell lineage.

5.4.6 Mitochondrial transcripts in single nuclei data

Several transcripts of mitochondrial origin (*MT-ND1*, *MT-ND4*, *MT-ND2* and *MT-CO2*) were identified as DEGs in cOPCs, all of which were determined to be downregulated. Nuclear transcripts of mitochondrial origin have been observed previously and are not unexpected in snRNA-seq data [191]. It is possible that these are circulating mRNAs that have been captured by accident in this dataset. However, as these were detected in all nuclei clusters and were tagged with UMIs relating to these cells, it is more likely that mRNA transcripts of mitochondrial genes have been transported into the nucleus. It is possible that mitochondrial function is affected in MOGHE but little to no evidence has been published to support this hypothesis. As it not possible to definitively exclude the possibility that these mitochondrial transcripts are circulating RNAs that have been captured incidentally, and the transcriptomic and proteomic analyses did not identify mitochondrial function as a pathway of interest in MOGHE, these DEGs were not classified as genes of interest and were removed from subsequent analysis.

5.4.7 Limitations and challenges

Low levels of *SLC35A2* expression in brain rendered differential expression analysis of *SLC35A2* impossible in this dataset. Therefore, it has not been possible to identify individual cells containing *SLC35A2* somatic variants. A sequencing technique which enabled identification of transcripts containing *SLC35A2* variants and allowed subsequent identification of cell origin would inform this study and enable additional specificity in identifying molecular pathways and cell types relevant to disease pathogenesis. Unfortunately, such a technique is not currently available but may be developed in the future.

Similarly, SLC35A2 is known to be low in abundance in brain tissue and indeed the protein was not detected consistently across the 30 tissue extracts used in this analysis. As such, it is difficult to draw conclusions about potential protein relationship activity of SLC35A2 in MOGHE tissues. The low abundance of *SLC35A2* transcripts and SLC35A2 protein molecules have limited the ability to draw conclusions about dysregulated SLC35A2 in these tissues.

The differential expression analysis performed on snRNA-seq data highlighted differences in cOPCs, OPCs, astrocytes and Layer 5/6 *THEMIS*-expressing excitatory neurons. Layer 5/6 *THEMIS*-expressing neurons accounted for only 0.69% total nuclei captured. This consisted of 0.79% nuclei (n = 228) in the MOGHE_Lesional samples and 0.57% nuclei (n = 126) in the MOGHE_Normal samples. These are rare cell types and, as such, it is possible that the ability to accurately assay these neurons in this dataset is limited. *GBR14* was identified as a single DEG between lesional and normal MOGHE nuclei but it is possible that other DEGs may be identified in a study that captures a greater number of nuclei from this cell type. Furthermore, if the heterotopic neurons observed in immunohistochemical analysis of MOGHE tissues are indeed Layer 5/6 *THEMIS*-expressing neurons, it is possible that the proportional abundance of these disease-specific cell types may be too small to interrogate meaningfully using currently employed methods and technology. Transcriptomic analysis of these heterotopic

neurons isolated using laser capture microdissection (LCM) or spatial platforms such as the Nanostring GeoMx Digital Spatial Profiler (Nanostring Technologies, Seattle) may assist to delineate the transcriptomic identity and morphological location of these cells in the future.

5.4.8 Future directions

Future studies should build upon findings from this thesis using alternate methods and animal models. Indeed, the findings of this study will be valuable to inform initial research in mouse models of MOGHE.

This study has identified *ACTB*, *GNG7*, *MAP4*, *NEFL*, *SEMA3E* and *SOD1* as candidate biomarkers in MOGHE using proteomic and single-nuclei transcriptomic analyses. Subsequent histological validation of potential biomarkers in resected brain tissue is necessary for adequate authentication of these genes and their respective proteins. Immunofluorescence co-staining of these markers with known cellular markers and maturity markers for cOPCs, OPCs, oligodendrocytes and neurons would assist to identify co-expression trends and spatial proximity of these cell types. These proteins may act as biomarkers of disease or may serve as druggable targets in the future. These studies are currently being performed but are outside the timeframe of this thesis.

For transcriptomic analysis, this study utilised snRNA-seq. This method was appropriate for neuronal tissue as nuclei have been shown to result in more accurate data. Recently, other forms of single-cell-omics such as the Nanostring single-cell spatial solution or RNAscope have enabled a spatial level to be added to single cell transcriptomics data analysis [313, 314]. This technique aims to provide a deeper understanding of the roles and activity of distinct cell types in tissue regions, based on their spatial location. Spatial analysis is predominantly facilitated by in situ hybridisation (ISH) methods or a spatially resolved transcriptomic method which incorporates oligonucleotide microarrays in the capture and processing of RNA transcripts across a tissue section [315]. Spatially resolved transcriptomics provides single cell transcriptomic data which is linked to a high-resolution tissue map. In future,

spatial single-cell transcriptomics may assist to further delineate the origins and location of dividing OPCs and Layer 5/6 *THEMIS*-expressing neurons in MOGHE. It would be valuable to spatially resolve the Mixed population and examine the spatial relationship between this Mixed population, dividing OPCs and heterotopic neurons. Spatial single-cell transcriptomics would be particularly useful to establish if these Layer 5/6 *THEMIS*-expressing neurons originate from OPC, cOPC or oligodendrocyte-rich neighbourhoods. Furthermore, spatial resolution may assist in the identification and characterisation of the Mixed cell population identified in this study.

5.5 Conclusions

Determining the exact mechanisms of pathogenesis in MOGHE remain a challenge to date. The N-glycosylation pathway still needs to be characterised better to comprehensively understand its impact on IESS and brain malformations. This study serves to provide insights into how a glycosylation disorder and oligodendrocyte abnormality result in a focal brain malformation.

The results presented in this chapter demonstrate that multi-omics analyses are a valuable tool in interrogating complex disorders like MOGHE and IESS more broadly. Combined snRNA-seq and proteomics analyses can assist to elucidate clues regarding molecular mechanisms of pathogenesis in brain tissue from MOGHE cases. snRNA-seq enables characterisation of disease cells and can provide insights into transcriptional dysregulation in disease-specific cells. Disease-specific cells were identified as oligodendrocyte progenitor cells (OPCs), dividing oligodendrocyte progenitor cells (cOPCs), astrocytes, Layer 5/6 *THEMIS*-expressing neurons and the Mixed population of nuclei. Major pathways dysregulated in MOGHE include axon and dendrite development, and neuron growth and localisation. The high-resolution proteomics data produced by TMT-16pro LC-MS/MS analysis of MOGHE brain tissue specimens enabled the detection of dysregulated proteins in lesional MOGHE tissues and the identification of a distinct MOGHE proteomic profile. As the proteomic analysis was performed on bulk tissue, the power of this analysis is limited when used in isolation.

In this study, the combination of snRNA-seq and proteomics into a multi-omic approach has addressed this limitation to some extent. Comparison between overlapping transcriptomic and proteomic data allowed the detection of dysregulated biomarkers and pathways that were sustained at both the mRNA and protein level. Most of the major overlapping dysregulated pathways identified across the datasets involved neurogenesis and axon and dendrite development (Table 5.3). Specifically, neuron projection development (GO:0031175) and cell morphogenesis (GO:0000902) were the two major biological processes common between the datasets and suggest these are robust dysregulated pathways in MOGHE.

Collectively, the multi-omic analyses described have identified six dysregulated protein-coding genes – *ACTB*, *GNG7*, *MAP4*, *NEFL*, *SEMA3E* and *SOD1* – in MOGHE which warrant validation and further investigation. These genes are potential biomarkers for MOGHE and may have utility as druggable targets in future treatment trials for MOGHE.

6 Chapter 6: Concluding remarks and outcomes

IESS is the most common form of severe epilepsy of infancy and represents a substantial medical burden. IESS resulting from MCD account for ~50% of all IESS cases and seizures are often drug resistant in these individuals. An improved understanding of the genetic basis of IESS due to MCD is necessary to accurately diagnose and treat this disorder. This study represents a substantial contribution to the knowledge of IESS mediated by MCD and the molecular pathogenesis of MOGHE.

In Chapter 3, the genetic basis of IESS due to MCD was identified in 80% of cases, comprising single gene variants in *TSC2*, *SLC35A2*, *PIK3CA*, *AKT3*, *MTOR*, *TSC1*, *CDKL5*, *DEPDC5*, *OFD1*, *FGFR1*, *NPRL3* and a blended case with a *COL4A1* variant and a 6p25.2 deletion. Novel genetic variants have been identified in this study in *SLC35A2*, *CDKL5*, *OFD1* and *FGFR1*. Somatic mosaic variants accounted for a much larger proportion of IESS cases with MCD than reported previously, particularly in individuals with malformations other than TSC. An important outcome of this study was the observation that ~40% of diagnoses were achievable using peripheral tissues prior to surgery. In some cases, this knowledge may have informed clinical decision making and subsequent surgery.

The genetic landscape of IESS due to MCD is highly heterogenous and comprised germline and somatic variants in a range of genes associated with epilepsy and MCD. Mosaicism is clearly an under-recognised cause of IESS with MCD and genomic sequencing on brain tissue is the most effective way to accurately identify these mosaic variants. In the context of the most efficient pathway to a genetic diagnosis, the data did not demonstrate a significant increase in diagnostic yield by performing either bulk RNA-seq and trio ES analyses on cases which remained negative following deep exome or panel sequencing. Additional diagnoses were only achieved via CNV analysis in this cohort. This is consistent with other studies of paediatric disorders utilising brain tissue including speech disorders, though trio ES have often demonstrated increased diagnostic yield over singleton ES in peripheral tissue studies [46, 93, 278, 316].

Alternative methods may assist to identify genetic diagnoses in the remaining 20% of cases. WGS and genome wide association studies have proven to be effective in identifying the genetic basis of epileptic disorders [22, 317, 318]. Similarly, methylation analysis on blood-derived gDNA has proven to be effective at identifying genetic signatures of different neurodevelopmental disorders [319-321]. One recent study found methylation analysis to be informative when delineating different MCD [322]. These tools may assist in the identification of a genetic basis in the remaining 20% in this cohort. Alternatively, these individuals may have IESS due to an MCD with an acquired basis.

In Chapter 4, a comprehensive study of nine cases with brain somatic variants in *SLC35A2* determined that these individuals had initially been clinically misdiagnosed as FCD I, FCD IIA or gliosis, due to phenotypic overlaps between these clinicopathological entities and MOGHE, and a lack of appropriate histopathological data for MOGHE classification. The combination of genetic, histopathological, imaging and developmental investigations resulted in an integrated diagnosis and reclassification of these nine individuals as having MOGHE. Absolute quantification of variant allele frequency in brain demonstrated a positive correlation between high variant allele frequency and region of most prominent malformation. Two additional individuals were identified as possibly having MOGHE via subsequent histopathological and MRI review. Therefore, this study added further weight to the association between brain somatic *SLC35A2* variation and MOGHE, confirmed MOGHE as a distinct clinicopathological entity, and illustrated the way in which multidisciplinary methods of diagnosis are necessary to accurately diagnose complex disorders and inform clinical management. This research has considerable clinical utility and indeed changed clinical practice at the Royal Children's Hospital. It is possible that implementation of a similar approach at other equivalent tertiary centres may occur following the publication of the results of this study.

Chapter 5 of this thesis presented a newly generated snRNA-seq and high-resolution proteomic dataset derived from brain tissue specimens of individuals with MOGHE. This snRNA-seq dataset consisted of 51,071 single-nuclei transcriptomes and the proteomics dataset constituted 7,655 proteins, likely representing the largest and most comprehensive MOGHE multi-omic dataset generated to date. Bioinformatic analysis identified several pathways as dysregulated in MOGHE, involving axon and dendrite development, and neuron growth and localisation. Molecular characterisation has provided insights into the molecular signatures of disease cells and has identified an unknown Mixed cell population. This study identified six protein-coding genes – *ACTB*, *GNG7*, *MAP4*, *NEFL*, *SEMA3E* and *SOD1* – as potential biomarkers for MOGHE which warrant validation and further investigation. These potential biomarkers may serve as potential druggable targets in the future. This study demonstrated that snRNA-seq and high-resolution proteomics can provide insights into the molecular drivers of epileptogenesis and can be used to identify potential biomarkers of disease. Furthermore, that multi-omic analyses are a valuable tool in delineating the molecular basis of rare disorders, with the technologies providing complementary strengths (e.g. the single cell nature of snRNA-seq compared with the bulk tissue analysis of the proteomics data). The proteomics analysis provides insight into the functional outcomes of transcriptomic alterations and provides potential for additional levels of analyses in the future. Although not performed as part of this study, proteomics allows the assessment of PTMs. This is particularly important due to the function of SLC35A2 in the N-glycosylation pathway, one of the major PTMs. Future analysis of MOGHE proteomic data should investigate the effect of SLC35A2 dysregulation on PTMs.

Collectively, the outcomes of this thesis have demonstrated that while IESS due to focal MCD is highly heterogenous, there are two genes which account for a significant proportion of cases, namely *TSC2* and *SLC35A2*, causing TSC/FCD II and MOGHE respectively. Additionally, this study has identified clinical and molecular hallmarks of MOGHE and provided insight into the potential dysregulated pathways contributing to MOGHE pathogenesis.

Although this study achieved a diagnostic yield of 80% in IESS due to MCD, there remain a portion of individuals (n=12) without a genetic basis. It is possible that the tissues or techniques utilised in this study did not capture extremely low-level somatic mosaic variants causing these unsolved cases, or there may be other contributing causes not investigated in this study. Future research into IESS due to MCD should investigate the contribution of deep intronic variants, polygenic risk scores, repeat expansions or acquired causes such as perinatal stroke, traumatic brain injury or vascular lesions, to disease aetiology. Long read sequencing techniques such as the Pacific Biosciences single-molecule real-time (SMRT) sequencing technology or the Oxford Nanopore Technologies (ONT) sequencing platform may assist to elucidate alternate genetic aetiologies of IESS with MCD, in particular the potential contribution of repeat expansions and regulatory or intronic variants [323]. Long-read sequencing technologies are capable of generating reads >10 kb in length, compared to the short-read sequencing technologies used in this study, which are able to generate reads >600bp in length [323]. In addition, the advent of real-time software-controlled target enrichment (adaptive sequencing [324]) and equivalent technologies is likely to considerably improve the ability to sequence pathway gene lists at great depth. The ability to analyse and select or reject reads in real time not only produces rapid genetic testing data, but also reduces sequencing costs and provides more complexity with long reads [325, 326].

As surgical techniques improve, smaller tissue resections are being collected and ablative techniques, such as laser interstitial thermal therapy (LITT) or stereo electroencephalography–guided radiofrequency thermocoagulation (SEEG-guided RF-TC), negate the need for resective surgery [327, 328]. As such, the opportunity to use brain tissue for genetic studies in the future may be limited. Non-invasive techniques for genetic screening or alternative methods of tissue collection may prove necessary in the future. Specifically, identifying methods to increase yield of cfDNA collection from CSF or DNA from trace tissue on stereo-electroencephalography

(SEEG) depth electrodes would be beneficial [217, 218, 329]. Subsequent investigations should examine the utility and viability of these alternate tissue collection methods.

Despite the findings regarding the molecular hallmarks and characterisation of MOGHE presented in this thesis, aspects of the disease mechanisms underlying epileptogenesis in MOGHE remain unclear. Future investigations into the pathogenesis in MOGHE are necessary to fully understand these mechanisms. Specifically, future studies should further address how a mosaic glycosylation disorder can result in a focal MCD and proliferation of oligodendrocytes; and what specific mechanisms in MOGHE may serve as druggable targets in clinical studies. Disease models of MOGHE will be an important tool to address these questions. An *Slc35a2* knockout mouse model may assist in the development of morphological and functional assays for a preclinical model of MOGHE. Mouse knockout models have proven historically useful in other neurological and epileptic disorders [330-332]. However, a brain somatic model of MOGHE is considerably more difficult to generate and would require the use of *in utero* electroporation (IUE) [220]. Similarly, iPSCs can provide a model for therapeutic screens and functional disease mechanism investigation in specific neuronal cell types [333-337]. Generation of a model for discovery and validation of targeted therapies in MOGHE would be valuable and an appropriate future direction for research. The findings in this thesis provide a foundation for future pathway-based precision medicine in MOGHE.

The overall results of this study are significant in terms of both research and clinical contributions. This project has provided genetic diagnoses to 43 individuals affected by IESS due to an MCD, in addition to four individuals who received clinical genetic diagnoses in parallel to this study that was able to be validated. As such, this study achieved a diagnostic yield of 80% (47/59). Not only have these genetic diagnoses signalled the end of the diagnostic odyssey for these children and their families, but a positive genetic diagnosis may also assist to improve clinical management, inform genetic and reproductive counselling, and assist with future diagnoses.

This study has also proved evidence of the value of an integrated diagnosis in clinical settings. Using multidisciplinary methods of diagnosis, this study has resulted in the reclassification of nine cases as MOGHE, increased knowledge of MOGHE, and potentially informed clinical decision making for future children presenting with MOGHE symptoms. Moreover, the snRNA-seq and proteomic data generated in the study have provided novel insights into the molecular profile of MOGHE and present a comprehensive dataset of MOGHE when paired with the genomic sequencing, histopathological, imaging and clinical data contained in this thesis.

Data and findings generated in this study have facilitated a better understanding of the genetic landscape of IESS due to MCD, and an improved characterisation of the relationship between brain somatic *SLC35A2* variation and MOGHE. This knowledge is not only important for IESS and MCD research but also informs the broader understanding of the normal development and function of the human brain.

7 Bibliography

1. Ngugi, A.K., et al., *Estimation of the burden of active and life-time epilepsy: a meta-analytic approach.* Epilepsia, 2010. **51**(5): p. 883-890.
2. Hesdorffer, D.C., et al., *Estimating risk for developing epilepsy: a population-based study in Rochester, Minnesota.* Neurology, 2011. **76**(1): p. 23-27.
3. Perucca, P., M. Bahlo, and S.F. Berkovic, *The Genetics of Epilepsy.* Annual Review of Genomics and Human Genetics, 2020: p. 205.
4. Berkovic, S.F., et al., *Human epilepsies: interaction of genetic and acquired factors.* Trends in Neurosciences, 2006(7): p. 391.
5. Scheffer, I.E., et al., *ILAE classification of the epilepsies: Position paper of the ILAE Commission for Classification and Terminology.* Epilepsia, 2017. **58**(4): p. 512-521.
6. Devinsky, O., et al., *Delivery of epilepsy care to adults with intellectual and developmental disabilities.* Neurology, 2015. **85**(17): p. 1512-21.
7. Specchio, N. and P. Curatolo, *Developmental and epileptic encephalopathies: what we do and do not know.* Brain, 2021. **144**(1): p. 32-43.
8. Wirrell, E.C., et al., *Methodology for classification and definition of epilepsy syndromes with list of syndromes: Report of the ILAE Task Force on Nosology and Definitions.* Epilepsia, 2022. **63**(6): p. 1333-1348.
9. Marco, M., et al., *Psychiatric Comorbidities in People With Epilepsy.* Neurology: Clinical Practice, 2021. **11**(2): p. e112.
10. Fiest, K.M., et al., *Depression in epilepsy: a systematic review and meta-analysis.* Neurology, 2013. **80**(6): p. 590-599.
11. Levisohn, P.M., *The autism-epilepsy connection.* Epilepsia, 2007. **48**(s9): p. 33-35.
12. Ottman, R., et al., *Comorbidities of epilepsy: Results from the Epilepsy Comorbidities and Health (EPIC) survey.* Epilepsia, 2011. **52**(2): p. 308-315.
13. Zinkin, N.T. and M.A. Peppercorn, *Abdominal epilepsy.* Best Practice & Research Clinical Gastroenterology, 2005. **19**(2): p. 263-274.
14. Whelan, C.D., et al., *Structural brain abnormalities in the common epilepsies assessed in a worldwide ENIGMA study.* Brain, 2018. **141**(2): p. 391-408.
15. Ellis, C.A., S. Petrovski, and S.F. Berkovic, *Epilepsy genetics: clinical impacts and biological insights.* The Lancet. Neurology, 2020. **19**(1): p. 93-100.
16. Bennett, M.F., et al., *Familial adult myoclonic epilepsy type 1 SAMD12 TTTCA repeat expansion arose 17,000 years ago and is present in Sri Lankan and Indian families.* European Journal of Human Genetics, 2020. **28**(7): p. 973.
17. Palencia-Campos, A., et al., *Germline and Mosaic Variants in PRKACA and PRKACB Cause a Multiple Congenital Malformation Syndrome.* The American Journal of Human Genetics, 2020.

18. Khera, A.V., et al., *Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations.* Nature genetics, 2018. **50**(9): p. 1219-1224.
19. Pederick, D.T., et al., *Abnormal cell sorting underlies the unique X-linked inheritance of PCDH19 epilepsy.* Neuron, 2018. **97**(1): p. 59-66. e5.
20. Wang, J., et al., *Epilepsy-associated genes.* Seizure, 2017. **44**: p. 11-20.
21. Oliver, K.L., et al., *Genes4Epilepsy: An epilepsy gene resource.* Epilepsia, 2023. **64**(5): p. 1368-1375.
22. International League Against Epilepsy Consortium on Complex, E., *Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies.* Nature communications, 2018. **9**(1): p. 5269-5269.
23. *Ultra-Rare Genetic Variation in the Epilepsies: A Whole-Exome Sequencing Study of 17,606 Individuals.* Am J Hum Genet, 2019. **105**(2): p. 267-282.
24. *Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies.* Nature communications, 2018. **9**(1): p. 5269.
25. Renzo, G., et al., *Monogenic Epilepsies.* Neurology, 2021. **97**(17): p. 817.
26. Shijun, Y., W. Bin, and H. Xiong, *Models for predicting treatment efficacy of antiepileptic drugs and prognosis of treatment withdrawal in epilepsy patients.* Acta Epileptologica, 2021. **3**(1): p. 1-6.
27. Abou-Khalil, B.W., *Update on antiepileptic drugs 2019.* CONTINUUM: Lifelong Learning in Neurology, 2019. **25**(2): p. 508-536.
28. Kwan, P., et al., *Definition of drug resistant epilepsy: consensus proposal by the ad hoc Task Force of the ILAE Commission on Therapeutic Strategies.* Epilepsia, 2010. **51**(6): p. 1069-77.
29. Ułamek-Kozioł, M., et al., *Ketogenic Diet and Epilepsy.* Nutrients, 2019. **11**(10): p. 2510.
30. Ben-Menachem, E., et al., *Surgically implanted and non-invasive vagus nerve stimulation: a review of efficacy, safety and tolerability.* European journal of neurology, 2015. **22**(9): p. 1260-1268.
31. Thompson, S.L., et al., *A review of parameter settings for invasive and non-invasive vagus nerve stimulation (VNS) applied in neurological and psychiatric disorders.* Frontiers in Neuroscience, 2021. **15**: p. 709436.
32. Edward F, C., et al., *Predictors of seizure freedom after surgery for malformations of cortical development.* Annals of Neurology, 2011. **70**(1): p. 151-162.
33. Loddenkemper, T., et al., *Developmental outcome after epilepsy surgery in infancy.* Pediatrics, 2007. **119**(5): p. 930-935.
34. Catchpool, M., et al., *Cost-effectiveness of epileptic surgery compared with medical treatment in children with drug-resistant epilepsy.* Epilepsy Behav, 2019. **97**: p. 253-259.
35. Samanta, D., et al., *Underutilization of epilepsy surgery: Part II: Strategies to overcome barriers.* Epilepsy & Behavior, 2021. **117**: p. 107853.
36. Rolston, J., *Surgical strategies for epilepsy in eloquent areas.* J Epilepsy, 2016. **2**(103): p. 2472-0895.1000103.
37. Cowan, L.D. and L.S. Hudson, *The Epidemiology and Natural History of Infantile Spasms.* Journal of Child Neurology, 1991. **6**(4): p. 355-364.

38. Zuberi, S.M., et al., *ILAE classification and definition of epilepsy syndromes with onset in neonates and infants: Position statement by the ILAE Task Force on Nosology and Definitions.* Epilepsia, 2022. **63**(6): p. 1349-1397.
39. Shields, W.D., *Infantile spasms: little seizures, BIG consequences.* Epilepsy currents, 2006. **6**(3): p. 63-69.
40. Howell, K.B., et al., *The severe epilepsy syndromes of infancy: A population-based study.* Epilepsia. **n/a**(n/a).
41. Guerrini, R., F. Sicca, and L. Parmeggiani, *Epilepsy and malformations of the cerebral cortex.* Epileptic Disorders, 2003. **5**(2): p. 9-26.
42. Carmant, L., *Infantile spasms: West syndrome.* Archives of neurology, 2002. **59**(2): p. 317-318.
43. Velíšek, L. and J. Velíšková, *Modeling epileptic spasms during infancy: Are we heading for the treatment yet?* Pharmacology & therapeutics, 2020. **212**: p. 107578.
44. Salar, S., S.L. Moshé, and A.S. Galanopoulou, *Metabolic etiologies in West syndrome.* Epilepsia open, 2018. **3**(2): p. 134-166.
45. Peng, J., et al., *Novel West syndrome candidate genes in a Chinese cohort.* CNS neuroscience & therapeutics, 2018. **24**(12): p. 1196-1206.
46. McTague, A., et al., *The genetic landscape of the epileptic encephalopathies of infancy and childhood.* The Lancet Neurology, 2016(3): p. 304.
47. Pavone, P., et al., *West syndrome: a comprehensive review.* Neurological Sciences, 2020. **41**(12): p. 3547-3562.
48. Silbereis, J.C., et al., *The Cellular and Molecular Landscapes of the Developing Human Central Nervous System.* Neuron, 2016. **89**(2): p. 248-268.
49. Ayala, R., T. Shu, and L.-H. Tsai, *Trekking across the brain: the journey of neuronal migration.* Cell, 2007. **128**(1): p. 29-43.
50. Kriegstein, A.R. and S.C. Noctor, *Patterns of neuronal migration in the embryonic cortex.* Trends in neurosciences, 2004. **27**(7): p. 392-399.
51. Brock, S., F. Cools, and A.C. Jansen, *Neuropathology of genetically defined malformations of cortical development – a systematic literature review.* Neuropathology and Applied Neurobiology. **n/a**(n/a).
52. Frederico A.C, A., et al., *Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain.* Journal of Comparative Neurology, 2009. **513**(5): p. 532-541.
53. Cadwell, C.R., et al., *Development and arealization of the cerebral cortex.* Neuron, 2019. **103**(6): p. 980-1004.
54. Subramanian, L., M.E. Calcagnotto, and M.F. Paredes, *Cortical Malformations: Lessons in Human Brain Development.* Frontiers in Cellular Neuroscience, 2020. **13**(576).
55. Pollen, A.A., et al., *Molecular identity of human outer radial glia during cortical development.* Cell, 2015. **163**(1): p. 55-67.
56. Rakic, P. and P.J. Lombroso, *Development of the cerebral cortex: I. Forming the cortical structure.* Journal of the American Academy of Child and Adolescent Psychiatry, 1998(1): p. 116.
57. Rakic, P., *Evolution of the neocortex: a perspective from developmental biology.* Nature Reviews Neuroscience, 2009(10): p. 724.

58. Subramanian, L., et al., *Dynamic behaviour of human neuroepithelial cells in the developing forebrain*. 2017.
59. Klingler, E., et al., *Mapping the molecular and cellular complexity of cortical malformations.* Science, 2021. **371**(6527): p. 361.
60. Jabaudon, D., *Fate and freedom in developing neocortical circuits.* Nature communications, 2017. **8**(1): p. 1-9.
61. Lim, L., et al., *Development and functional diversification of cortical interneurons.* Neuron, 2018. **100**(2): p. 294-313.
62. Guillemot, F. and C. Zimmer, *From cradle to grave: the multiple roles of fibroblast growth factors in neural development.* Neuron, 2011. **71**(4): p. 574-588.
63. Mason, I., *Initiation to end point: the multiple roles of fibroblast growth factors in neural development.* Nature Reviews Neuroscience, 2007. **8**(8): p. 583-596.
64. Nadarajah, B., et al., *Two modes of radial migration in early development of the cerebral cortex.* Nature Neuroscience, 2001. **4**(2): p. 143-150.
65. Wood, H., *Do the locomotion?* Nature Reviews Neuroscience, 2001. **2**(3): p. 153-153.
66. Cowan, W.M., *The development of the brain.* Scientific American, 1979. **241**(3): p. 112-133.
67. Guerrini, R. and E. Parrini, *Neuronal migration disorders.* Neurobiology of disease, 2010. **38**(2): p. 154-166.
68. Friocourt, G., et al., *Cell-autonomous roles of ARX in cell proliferation and neuronal migration during corticogenesis.* Journal of Neuroscience, 2008. **28**(22): p. 5794-5805.
69. Friocourt, G.M. and J.G. Parnavelas, *Identification of Arx targets unveils new candidates for controlling cortical interneuron migration and differentiation.* Frontiers in cellular neuroscience, 2011. **5**: p. 28.
70. Friocourt, G. and J.G. Parnavelas, *Mutations in ARX result in several defects involving GABAergic neurons.* Frontiers in cellular neuroscience, 2010. **4**: p. 4.
71. Komuro, H. and P. Rakic, *Distinct modes of neuronal migration in different domains of developing cerebellar cortex.* J Neurosci, 1998. **18**(4): p. 1478-90.
72. Xu, C., et al., *Radial glial cell–neuron interaction directs axon formation at the opposite side of the neuron from the contact site.* Journal of Neuroscience, 2015. **35**(43): p. 14517-14532.
73. Barnes, A.P. and F. Polleux, *Establishment of axon-dendrite polarity in developing neurons.* Annual review of neuroscience, 2009. **32**.
74. Cappello, S., et al., *Mutations in genes encoding the cadherin receptor-ligand pair DCHS1 and FAT4 disrupt cerebral cortical development.* Nature genetics, 2013. **45**(11): p. 1300-1308.
75. Francis, F., et al., *Doublecortin Is a Developmentally Regulated, Microtubule-Associated Protein Expressed in Migrating and Differentiating Neurons.* Neuron, 1999. **23**(2): p. 247-256.
76. Bahi-Buisson, N., et al., *The wide spectrum of tubulinopathies: what are the key features for the diagnosis?* Brain, 2014. **137**(6): p. 1676-1700.

77. Jäkel, S. and L. Dimou, *Glial Cells and Their Function in the Adult Brain: A Journey through the History of Their Ablation.* Frontiers in Cellular Neuroscience, 2017. **11**(24).
78. Katz, L.C. and C.J. Shatz, *Synaptic activity and the construction of cortical circuits.* Science, 1996. **274**(5290): p. 1133-1138.
79. Langley, B.C., *Chapter 6 - The Cell Cycle and Oxidative Neuronal Cell Death*, in *Oxidative Stress and Neurodegenerative Disorders*, G.A. Qureshi and S.H. Parvez, Editors. 2007, Elsevier Science B.V.: Amsterdam. p. 165-182.
80. Williamson, J.M. and D.A. Lyons, *Myelin Dynamics Throughout Life: An Ever-Changing Landscape?* Frontiers in Cellular Neuroscience, 2018. **12**(424).
81. Raybaud, C. and E. Widjaja, *Development and dysgenesis of the cerebral cortex: malformations of cortical development.* Neuroimaging clinics of North America, 2011. **21**(3): p. 483.
82. Mancini, G.M. and N.-M. Network, *Neuro-MIG: A European network on brain malformations.* European journal of medical genetics, 2018. **61**(12): p. 741-743.
83. Leventer, R., et al., *Clinical and imaging features of cortical malformations in childhood.* Neurology, 1999. **53**(4): p. 715-715.
84. Petryk, A., D. Graf, and R. Marcucio, *Holoprosencephaly: signaling interactions between the brain and the face, the environment and the genes, and the phenotypic variability in animal models and humans.* Wiley interdisciplinary reviews. Developmental biology, 2015. **4**(1): p. 17-32.
85. Severino, M., et al., *Definitions and classification of malformations of cortical development: practical guidelines.* Brain, 2020. **143**(10): p. 2874-2894.
86. Leventer, R.J., et al., *Clinical and imaging features of cortical malformations in childhood*. 1999. p. 715-722.
87. Tahta, A. and M. Turgut, *Focal cortical dysplasia: etiology, epileptogenesis, classification, clinical presentation, imaging, and management.* Child's Nervous System, 2020. **36**(12): p. 2939.
88. Barkovich, A.J. and R.I. Kuzniecky, *Neuroimaging of focal malformations of cortical development.* Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society, 1996. **13**(6): p. 481-494.
89. Barkovich, A.J., et al., *A developmental and genetic classification for malformations of cortical development: update 2012.* Brain, 2012. **135**(5): p. 1348-1369.
90. Barkovich, A.J., et al., *A developmental and genetic classification for malformations of cortical development.* Neurology, 2005. **65**(12): p. 1873-1887.
91. Barkovich, A.J., et al., *Classification system for malformations of cortical development: update 2001.* Neurology, 2001. **57**(12): p. 2168-2178.
92. Taylor, D.C., et al., *Focal dysplasia of the cerebral cortex in epilepsy.* Journal of Neurology, Neurosurgery & Psychiatry, 1971. **34**(4): p. 369.
93. Baldassari, S., et al., *Dissecting the genetic basis of focal cortical dysplasia: a large cohort study.* Acta Neuropathologica, 2019. **138**(6): p. 885-900.

94. Blümcke, I. and A. Mühlebner, *Neuropathological work-up of focal cortical dysplasias using the new ILAE consensus classification system - practical guideline article invited by the Euro-CNS Research Committee.* Clinical neuropathology, 2011. **30**(4): p. 164-177.
95. Najm, I.M., H.B. Sarnat, and I. Blümcke, *Review: The international consensus classification of Focal Cortical Dysplasia - a critical update 2018.* Neuropathology and applied neurobiology, 2018. **44**(1): p. 18-31.
96. Kim, S.H. and J. Choi, *Pathological Classification of Focal Cortical Dysplasia (FCD) : Personal Comments for Well Understanding FCD Classification.* Journal of Korean Neurosurgical Society, 2019. **62**(3): p. 288-295.
97. Blumcke, I., et al., *Histopathological Findings in Brain Tissue Obtained during Epilepsy Surgery.* New England Journal of Medicine, 2017. **377**(17): p. 1648-1656.
98. López-Rivera, J.A., et al., *Incidence and prevalence of major epilepsy-associated brain lesions.* Epilepsy Behav Rep, 2022. **18**: p. 100527.
99. Jansen, L.A., et al., *PI3K/AKT pathway mutations cause a spectrum of brain malformations from megalencephaly to focal cortical dysplasia.* Brain, 2015. **138**(6): p. 1613-1628.
100. Mühlebner, A., et al., *New insights into a spectrum of developmental malformations related to mTOR dysregulations: challenges and perspectives.* Journal of anatomy, 2019. **235**(3): p. 521-542.
101. Scerri, T., et al., *Familial cortical dysplasia type IIA caused by a germline mutation in DEPDC 5.* Annals of clinical and translational neurology, 2015. **2**(5): p. 575-580.
102. Marsan, E. and S. Baulac, *Review: Mechanistic target of rapamycin (mTOR) pathway, focal cortical dysplasia and epilepsy.* Neuropathol Appl Neurobiol, 2018. **44**(1): p. 6-17.
103. Harvey, A.S., et al., *The surgically remediable syndrome of epilepsy associated with bottom-of-sulcus dysplasia.* Neurology, 2015. **84**(20): p. 2021.
104. Lee, W.S., et al., *Genetic characterization identifies bottom-of-sulcus dysplasia as an mTORopathy.* Neurology, 2020. **95**(18): p. e2542.
105. Wei Shern, L., et al., *Second-hit DEPDC5 mutation is limited to dysmorphic neurons in cortical dysplasia type IIA.* Annals of Clinical and Translational Neurology, 2019. **6**(7): p. 1338-1344.
106. Massimi, L. and C. Di Rocco, *Hemimegalencephaly.* Textbook of Pediatric Neurosurgery, 2020: p. 1049-1080.
107. Crino, P.B., *mTOR signaling in epilepsy: insights from malformations of cortical development.* Cold Spring Harbor perspectives in medicine, 2015. **5**(4): p. a022442.
108. Garcia, C.A., et al., *mTOR pathway somatic variants and the molecular pathogenesis of hemimegalencephaly.* Epilepsia open, 2020. **5**(1): p. 97-106.
109. Lee, J.H., et al., *De novo somatic mutations in components of the PI3K-AKT3-mTOR pathway cause hemimegalencephaly.* Nature genetics, 2012. **44**(8): p. 941-945.
110. D'Gama, A.M. and C.A. Walsh, *Somatic mosaicism and neurodevelopmental disease.* Nature neuroscience, 2018. **21**(11): p. 1504-1514.

111. Najm, I., et al., *The ILAE consensus classification of focal cortical dysplasia: An update proposed by an ad hoc task force of the ILAE diagnostic methods commission.* Epilepsia, 2022. **63**(8): p. 1899-1919.
112. Schurr, J., et al., *Mild Malformation of Cortical Development with Oligodendroglial Hyperplasia in Frontal Lobe Epilepsy: A New Clinico-Pathological Entity.* Brain Pathology, 2017(1): p. 26.
113. Winawer, M.R., et al., *Somatic SLC35A2 variants in the brain are associated with intractable neocortical epilepsy.* Annals of Neurology, 2018. **83**(6): p. 1133-1146.
114. Bonduelle, T., et al., *Frequent SLC35A2 brain mosaicism in mild malformation of cortical development with oligodendroglial hyperplasia in epilepsy (MOGHE).* Acta neuropathologica communications, 2021. **9**(1): p. 3.
115. Caban, C., et al., *Genetics of tuberous sclerosis complex: implications for clinical practice.* The Application of Clinical Genetics, 2020.
116. C, J., et al., *Learning disability and epilepsy in an epidemiological sample of individuals with tuberous sclerosis complex.* Psychological Medicine, 2003. **33**(2): p. 335-344.
117. Marom, D., *Genetics of tuberous sclerosis complex: an update.* Child's Nervous System, 2020(10): p. 2489.
118. Dunlop, E.A. and A.R. Tee, *Mammalian target of rapamycin complex 1: Signalling inputs, substrates and feedback mechanisms.* Cellular Signalling, 2009. **21**(6): p. 827-835.
119. Qin, W., et al., *Analysis of TSC cortical tubers by deep sequencing of TSC1, TSC2 and KRAS demonstrates that small second-hit mutations in these genes are rare events*, in *Brain pathology (Zurich, Switzerland).* 2010, International Society of Neuropathology: Switzerland. p. 1096-1105.
120. Squier, W. and A. Jansen, *Polymicrogyria: pathology, fetal origins and mechanisms.* Acta neuropathologica communications, 2014. **2**(1): p. 1-16.
121. Leventer, R.J., et al., *Clinical and imaging heterogeneity of polymicrogyria: a study of 328 patients.* Brain, 2010. **133**(5): p. 1415-1427.
122. Rivière, J.-B., et al., *De novo germline and postzygotic mutations in AKT3, PIK3R2 and PIK3CA cause a spectrum of related megalencephaly syndromes.* Nature genetics, 2012. **44**(8): p. 934-940.
123. Moore, S.A., et al., *Deletion of brain dystroglycan recapitulates aspects of congenital muscular dystrophy.* Nature, 2002. **418**(6896): p. 422-425.
124. Kobow, K., et al., *Mosaic trisomy of chromosome 1q in human brain tissue associates with unilateral polymicrogyria, very early-onset focal epilepsy, and severe developmental delay.* Acta Neuropathologica, 2020. **140**(6): p. 881-891.
125. Jansen, A. and E. Andermann, *Genetics of the polymicrogyria syndromes.* Journal of medical genetics, 2005. **42**(5): p. 369-378.
126. Pramparo, T. and T. Wynshaw-Boris, *Lissencephaly*, in *Encyclopedia of Neuroscience*, L.R. Squire, Editor. 2009, Academic Press: Oxford. p. 495-501.
127. Di Donato, N., et al., *Lissencephaly: expanded imaging and clinical classification.* American Journal of Medical Genetics Part A, 2017. **173**(6): p. 1473-1488.

128. Di Donato, N., et al., *Analysis of 17 genes detects mutations in 81% of 811 patients with lissencephaly.* Genetics in Medicine: Official journal of the American College of Medical Genetics and Genomics, 2018. **20**(11): p. 1354.
129. Dinday, M.T., et al., *PAFAH1B1 haploinsufficiency disrupts GABA neurons and synaptic E/I balance in the dentate gyrus.* Sci Rep, 2017. **7**(1): p. 8269.
130. Katayama, K.I., et al., *Enhanced expression of Pafah1b1 causes over-migration of cerebral cortical neurons into the marginal zone.* Brain Struct Funct, 2017. **222**(9): p. 4283-4291.
131. Srour, M., et al., *The clinical spectrum of nodular heterotopias in children: report of 31 patients.* Epilepsia, 2011. **52**(4): p. 728-737.
132. d'Orsi, G., et al., *Clinical features and long term outcome of epilepsy in periventricular nodular heterotopia. Simple compared with plus forms.* Journal of Neurology, Neurosurgery & Psychiatry, 2004. **75**(6): p. 873.
133. Haverfield, E.V., et al., *Intragenic deletions and duplications of the LIS1 and DCX genes: a major disease-causing mechanism in lissencephaly and subcortical band heterotopia.* Eur J Hum Genet, 2009. **17**(7): p. 911-8.
134. Sheen, V., et al., *Filamin A mutations cause periventricular heterotopia with Ehlers-Danlos syndrome.* Neurology, 2005. **64**(2): p. 254-262.
135. Fox, J.W., et al., *Mutations in filamin 1 prevent migration of cerebral cortical neurons in human periventricular heterotopia.* Neuron, 1998. **21**(6): p. 1315-1325.
136. Fumihiko, N., S. Thomas P, and H. John H, *The filamins: Organizers of cell structure and function.* Cell Adhesion & Migration, 2011. **5**(2): p. 160-169.
137. Volney L, S., et al., *Mutations in the X-linked filamin 1 gene cause periventricular nodular heterotopia in males as well as in females.* Human Molecular Genetics, 2001. **10**(17): p. 1775-1783.
138. Poirier, K., et al., *Mutations in TUBG1, DYNC1H1, KIF5C and KIF2A cause malformations of cortical development and microcephaly.* Nature genetics, 2013. **45**(6): p. 639-647.
139. Heinzen, E.L., et al., *De novo and inherited private variants in MAP1B in periventricular nodular heterotopia.* PLoS genetics, 2018. **14**(5): p. e1007281.
140. Jayaraman, D., B.-I. Bae, and C.A. Walsh, *The genetics of primary microcephaly.* Annual review of genomics and human genetics, 2018. **19**: p. 177-200.
141. Jayaraman, D., et al., *Microcephaly proteins Wdr62 and Aspm define a mother centriole complex regulating centriole biogenesis, apical complex, and cell fate.* Neuron, 2016. **92**(4): p. 813-828.
142. Francesca, J., S. Amanda, and T.-G. Maja, *Dissecting the Genetic and Etiological Causes of Primary Microcephaly.* Frontiers in Neurology, 2020. **11**.
143. Desikan, R.S. and A.J. Barkovich, *Malformations of cortical development.* Annals of neurology, 2016. **80**(6): p. 797-810.
144. Devisme, L., et al., *Cobblestone lissencephaly: neuropathological subtypes and correlations with genes of dystroglycanopathies.* Brain, 2012. **135**(2): p. 469-482.

145. Richards, S., et al., *Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology.* Genetics in Medicine, 2015. **17**(5): p. 405-423.
146. Clowry, G., et al., *New insights into the development of the human cerebral cortex.* Journal of Anatomy, 2019(3): p. 432.
147. Ye, Z., et al., *Somatic mutation: The hidden genetics of brain malformations and focal epilepsies.* Epilepsy Research, 2019. **155**.
148. Miller, K.E., et al., *Somatic SLC35A2 mosaicism correlates with clinical findings in epilepsy brain tissue.* Neurology Genetics, 2020. **6**(4): p. e460.
149. Hildebrand, M.S., et al., *Somatic GNAQ mutation in the forme fruste of Sturge-Weber syndrome.* 2018.
150. Sim, N.S., et al., *Precise detection of low-level somatic mutation in resected epilepsy brain tissue.* Acta neuropathologica, 2019. **138**(6): p. 901-912.
151. Aicher, J.K., et al., *Mapping RNA splicing variations in clinically accessible and nonaccessible tissues to facilitate Mendelian disease diagnosis using RNA-seq.* Genet Med, 2020. **22**(7): p. 1181-1190.
152. Knudson, A.G., Jr., *Mutation and cancer: statistical study of retinoblastoma.* Proceedings of the National Academy of Sciences of the United States of America, 1971. **68**(4): p. 820-823.
153. Hino, O. and T. Kobayashi, *Mourning Dr. Alfred G. Knudson: the two-hit hypothesis, tumor suppressor genes, and the tuberous sclerosis complex.* Cancer science, 2017. **108**(1): p. 5-11.
154. Ribierre, T., et al., *Second-hit mosaic mutation in mTORC1 repressor DEPDC5 causes focal cortical dysplasia–associated epilepsy.* The Journal of clinical investigation, 2018. **128**(6): p. 2452-2458.
155. Goswami, S. and J. Hsieh, *One-Hit Wonders and 2-Hit Tubers: A Second-Hit to TSC2 Causes Tuber-Like Cells in Spheroids.* Epilepsy currents, 2019. **19**(1): p. 49-50.
156. Uddin, M., et al., *Germline and somatic mutations in STXBP1 with diverse neurodevelopmental phenotypes.* Neurology Genetics, 2017. **3**(6): p. e199.
157. Bennett, M.F., et al., *Evidence for a Dual-Pathway, 2-Hit Genetic Model for Focal Cortical Dysplasia and Epilepsy.* Neurol Genet, 2022. **8**(1): p. e652.
158. Ding, J., et al., *Systematic comparison of single-cell and single-nucleus RNA-sequencing methods.* Nature Biotechnology, 2020. **38**(6): p. 737-746.
159. Myers, K.A., et al., *Transcriptome analysis of a ring chromosome 20 patient cohort.* Epilepsia. **n/a**(n/a).
160. Yu, W., et al., *scATAC-pro: a comprehensive workbench for single-cell chromatin accessibility sequencing data.* Genome Biology, 2020. **21**(1): p. 94.
161. Gupta, I., et al., *Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells.* Nat Biotechnol, 2018.
162. Spyros, O., et al., *Methodologies for Transcript Profiling Using Long-Read Technologies.* Frontiers in Genetics, 2020. **11**.
163. Xu, Q., et al., *mTOR inhibitors as a new therapeutic strategy in treatment resistant epilepsy in hemimegalencephaly: a case report.* Journal of Child Neurology, 2019. **34**(3): p. 132-138.

164. D'Gama, A.M. and A. Poduri, *Precision therapy for epilepsy related to brain malformations.* Neurotherapeutics, 2021. **18**(3): p. 1548-1563.
165. Davies, M., A. Saxena, and J.C. Kingswood, *Management of everolimus-associated adverse events in patients with tuberous sclerosis complex: a practical guide.* Orphanet Journal of Rare Diseases, 2017. **12**(1): p. 1-14.
166. Kerezoudis, P., et al., *Surgical Outcomes of Laser Interstitial Thermal Therapy for Temporal Lobe Epilepsy: Systematic Review and Meta-analysis.* World Neurosurgery, 2020. **143**: p. 527-536.e3.
167. Kõressaar, T., et al., *Primer3_masker: integrating masking of template sequence with primer design software.* Bioinformatics, 2018. **34**(11): p. 1937-1938.
168. Marsh, A.P., et al., *Mutations in DCC cause isolated agenesis of the corpus callosum with incomplete penetrance.* Nature genetics, 2017. **49**(4): p. 511-514.
169. Rentzsch, P., et al., *CADD: predicting the deleteriousness of variants throughout the human genome.* Nucleic acids research, 2019. **47**(D1): p. D886-D894.
170. Ioannidis, N.M., et al., *REVEL: an ensemble method for predicting the pathogenicity of rare missense variants.* The American Journal of Human Genetics, 2016. **99**(4): p. 877-885.
171. Sundaram, L., et al., *Predicting the clinical impact of human mutation with deep neural networks.* Nature genetics, 2018. **50**(8): p. 1161-1170.
172. Samocha, K.E., et al., *Regional missense constraint improves variant deleteriousness prediction.* BioRxiv, 2017: p. 148353.
173. Jaganathan, K., et al., *Predicting splicing from primary sequence with deep learning.* Cell, 2019. **176**(3): p. 535-548. e24.
174. Ionita-Laza, I., et al., *A spectral approach integrating functional genomic annotations for coding and noncoding variants.* Nature genetics, 2016. **48**(2): p. 214-220.
175. Adzhubei, I.A., et al., *A method and server for predicting damaging missense mutations.* Nature methods, 2010. **7**(4): p. 248-249.
176. Ng, P.C. and S. Henikoff, *SIFT: Predicting amino acid changes that affect protein function.* Nucleic acids research, 2003. **31**(13): p. 3812-3814.
177. Schwarz, J.M., et al., *MutationTaster2: mutation prediction for the deep-sequencing age.* Nature methods, 2014. **11**(4): p. 361-362.
178. Shihab, H.A., et al., *Ranking non-synonymous single nucleotide polymorphisms based on disease concepts.* Human genomics, 2014. **8**(1): p. 1-6.
179. Dong, C., et al., *Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies.* Human molecular genetics, 2015. **24**(8): p. 2125-2137.
180. Sadedin, S.P., et al., *Ximmer: a system for improving accuracy and consistency of CNV calling from exome data.* Gigascience, 2018. **7**(10): p. giy112.
181. Liao, Y., G.K. Smyth, and W. Shi, *featureCounts: an efficient general purpose program for assigning sequence reads to genomic features.* Bioinformatics, 2014. **30**(7): p. 923-930.

182. Liao, Y., G.K. Smyth, and W. Shi, *The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote.* Nucleic acids research, 2013. **41**(10): p. e108-e108.
183. Baybis, M., et al., *mTOR cascade activation distinguishes tubers from focal cortical dysplasia.* Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society, 2004. **56**(4): p. 478-487.
184. Zheng, G.X., et al., *Massively parallel digital transcriptional profiling of single cells.* Nature communications, 2017. **8**(1): p. 14049.
185. Lun, A.T.L., et al., *EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data.* Genome Biology, 2019. **20**(1): p. 63.
186. McCarthy, D.J., et al., *Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R.* Bioinformatics, 2017. **33**(8): p. 1179-1186.
187. Lun, A.T., D.J. McCarthy, and J.C. Marioni, *A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor.* F1000Res, 2016. **5**: p. 2122.
188. Germain, P.-L., et al., *Doublet identification in single-cell sequencing data using scDblFinder.* F1000Research, 2021. **10**.
189. Korsunsky, I., et al., *Fast, sensitive and accurate integration of single-cell data with Harmony.* Nature Methods, 2019. **16**(12): p. 1289-1296.
190. Germain, P.L., A. Sonrel, and M.D. Robinson, *pipeComp, a general framework for the evaluation of computational pipelines, reveals performant single cell RNA-seq preprocessing tools.* Genome Biol, 2020. **21**(1): p. 227.
191. Herring, C.A., et al., *Human prefrontal cortex gene regulatory dynamics from gestation to adulthood at single-cell resolution*, in *Cell*. 2022, Cell Press: United States. p. 4428.
192. Freytag, S., et al., *Comparison of clustering tools in R for medium-sized 10x Genomics single-cell RNA-sequencing data.* F1000Research, 2018. **7**.
193. Cowan, L.D. and L.S. Hudson, *The epidemiology and natural history of infantile spasms.* J Child Neurol, 1991. **6**(4): p. 355-64.
194. Howell, K.B., et al., *A population-based cost-effectiveness study of early genetic testing in severe epilepsies of infancy.* Epilepsia, 2018. **59**(6): p. 1177-1187.
195. Ebrahimi-Fakhari, D., et al., *Incidence of tuberous sclerosis and age at first diagnosis: new data and emerging trends from a national, prospective surveillance study.* Orphanet J Rare Dis, 2018. **13**(1): p. 117.
196. Wilmshurst, J.M., et al., *Summary of recommendations for the management of infantile seizures: Task Force Report for the ILAE Commission of Pediatrics.* Epilepsia, 2015. **56**(8): p. 1185-1197.
197. Howell, K.B., *The epidemiology and aetiologies of the severe epilepsies of infancy*. 2016.
198. Howell, K.B., et al., *The severe epilepsy syndromes of infancy: A population-based study.* Epilepsia, 2021. **62**(2): p. 358.
199. Osborne, J.P., et al., *The underlying etiology of infantile spasms (West syndrome): Information from the United Kingdom Infantile Spasms Study*

(UKISS) on contemporary causes and their classification 2. Epilepsia, 2010. **51**(10): p. 2168-2174.
200. Osborne, J.P., et al., *The underlying etiology of infantile spasms (West syndrome): Information from the International Collaborative Infantile Spasms Study (ICISS).* Epilepsia, 2019. **60**(9): p. 1861-1869.
201. Jacqueline, W. and N.M. Sunita, *Developmental Outcomes of Infants Treated With Combination Therapy for Infantile Spasms.* Pediatric Neurology Briefs, 2019. **33**(0).
202. Dzau, W., et al., *Response to sequential treatment with prednisolone and vigabatrin in infantile spasms.* Journal of Paediatrics and Child Health, 2022. **58**(12): p. 2197-2202.
203. Muir, A.M., et al., *Genetic heterogeneity in infantile spasms.* Epilepsy Research, 2019. **156**: p. 106181.
204. Liu, L.-Y., et al., *Diagnostic yield of a multi-strategy genetic testing procedure in a nationwide cohort of 728 patients with infantile spasms in China.* Seizure, 2022. **103**: p. 51-57.
205. Crino, P.B., *mTORopathies: A Road Well-Traveled.* Epilepsy Currents, 2020. **20**(6_suppl): p. 64S-66S.
206. Parrini, E., et al., *Genetic Basis of Brain Malformations.* Mol Syndromol, 2016. **7**(4): p. 220-233.
207. Mancini, G.M.S., et al., *Multidisciplinary interaction and MCD gene discovery. The perspective of the clinical geneticist.* European Journal of Paediatric Neurology, 2021. **35**: p. 27-34.
208. Bast, T., et al., *Focal cortical dysplasia: prevalence, clinical presentation and epilepsy in children and adults.* Acta Neurologica Scandinavica, 2006. **113**(2): p. 72-81.
209. Lee, W.S., et al., *Second-hit DEPDC5 mutation is limited to dysmorphic neurons in cortical dysplasia type IIA.* ANNALS OF CLINICAL AND TRANSLATIONAL NEUROLOGY, 2019. **6**(7): p. 1338-1344.
210. Carvill, G.L., et al., *Targeted resequencing in epileptic encephalopathies identifies de novo mutations in CHD2 and SYNGAP1.* Nature genetics, 2013. **45**(7): p. 825-830.
211. Lemke, J.R., et al., *Targeted next generation sequencing as a diagnostic tool in epileptic disorders.* Epilepsia, 2012. **53**(8): p. 1387-1398.
212. Allen, A.S., et al., *De novo mutations in epileptic encephalopathies.* Nature, 2013. **501**(7466): p. 217-221.
213. D'Gama, A.M., et al., *Somatic mutations activating the mTOR pathway in dorsal telencephalic progenitors cause a continuum of cortical dysplasias.* Cell reports, 2017. **21**(13): p. 3754-3766.
214. Lim, J.S., et al., *Somatic mutations in TSC1 and TSC2 cause focal cortical dysplasia.* The American Journal of Human Genetics, 2017. **100**(3): p. 454-472.
215. López-Rivera, J.A., et al., *The genomic landscape across 474 surgically accessible epileptogenic human brain lesions.* Brain, 2023. **146**(4): p. 1342-1356.
216. Lee, W.S., R.J. Leventer, and P.J. Lockhart, *Droplet digital PCR as a first-tier molecular diagnostic tool for focal cortical dysplasia type II.* Brain, 2022. **145**(12): p. e119-e121.

217. Ye, Z., et al., *Cerebrospinal fluid liquid biopsy for detecting somatic mosaicism in brain*. Brain Communications, 2021. **3**(1).
218. Ye, Z., et al., *Improving Specificity of Cerebrospinal Fluid Liquid Biopsy for Genetic Testing*. Annals of Neurology, 2021. **90**(4): p. 693-694.
219. Biever, A., E. Valjent, and E. Puighermanal, *Ribosomal Protein S6 Phosphorylation in the Nervous System: From Regulation to Function*. Frontiers in Molecular Neuroscience, 2015. **8**.
220. Hsieh, L.S., et al., *Ectopic HCN4 expression drives mTOR-dependent epilepsy in mice*. SCIENCE TRANSLATIONAL MEDICINE, 2020. **12**(570): p. eabc1492.
221. Abdijadid, S., et al., *Basic Mechanisms of Epileptogenesis in Pediatric Cortical Dysplasia*. CNS Neuroscience & Therapeutics, 2015. **21**(2): p. 92-103.
222. Au, K.S., et al., *Genotype/phenotype correlation in 325 individuals referred for a diagnosis of tuberous sclerosis complex in the United States*. Genetics in Medicine, 2007. **9**(2): p. 88-100.
223. Rivera, B., et al., *Germline and somatic FGFR1 abnormalities in dysembryoplastic neuroepithelial tumors*. Acta Neuropathologica, 2016. **131**(6): p. 847-863.
224. Golomb, M.R., et al., *Perinatal stroke and the risk of developing childhood epilepsy*. The Journal of pediatrics, 2007. **151**(4): p. 409-413. e2.
225. Thom, M., et al., *Temporal lobe sclerosis associated with hippocampal sclerosis in temporal lobe epilepsy: neuropathological features*. Journal of Neuropathology & Experimental Neurology, 2009. **68**(8): p. 928-938.
226. Olson, H.E., et al., *Cyclin-Dependent Kinase-Like 5 Deficiency Disorder: Clinical Review*. Pediatr Neurol, 2019. **97**: p. 18-25.
227. Hector, R.D., et al., *CDKL5 variants: Improving our understanding of a rare neurologic disorder*. Neurology Genetics, 2017. **3**(6).
228. Corominas, J., et al., *Clinical exome sequencing-Mistakes and caveats*. Hum Mutat, 2022. **43**(8): p. 1041-1055.
229. Moloney, P.B., et al., *Genomics in the presurgical epilepsy evaluation*. Epilepsy Research, 2022. **184**: p. 106951.
230. Boßelmann, C.M., et al., *Genetic testing before epilepsy surgery – An exploratory survey and case collection from German epilepsy centers*. Seizure, 2022. **95**: p. 4-10.
231. Yuriko, Y., et al., *Phenotypic Spectrum of COL4A1 Mutations: Porencephaly to Schizencephaly*. Annals of Neurology, 2013. **73**(1): p. 48-57.
232. Matías-Pérez, D., et al., *Identification of novel pathogenic variants and novel gene-phenotype correlations in Mexican subjects with microphthalmia and/or anophthalmia by next-generation sequencing*. Journal of human genetics, 2018. **63**(11): p. 1169-1180.
233. Elliott, A.M., et al., *RAPIDOMICS: rapid genome-wide sequencing in a neonatal intensive care unit—successes and challenges*. European journal of pediatrics, 2019. **178**(8): p. 1207-1218.
234. Eid, M., et al., *Further Insights into Developmental Brain Malformations and Leukoencephalopathy Associated with 6p25.3 Deletion*. 2020. p. 76-82.

235. Shetty, J., et al., *Aicardi syndrome in a 47 XXY male–a variable developmental phenotype?* European Journal of Paediatric Neurology, 2014. **18**(4): p. 529-531.
236. William, B.H., et al., *The expanding phenotype of OFD1-related disorders: Hemizygous loss-of-function variants in three patients with primary ciliary dyskinesia.* Molecular Genetics & Genomic Medicine, 2019. **7**(9).
237. Sakakibara, N., et al., *Clinical spectrum of male patients with OFD1 mutations.* Journal of human genetics, 2019. **64**(1): p. 3-9.
238. Park, S.M., et al., *Brain Somatic Mutations in MTOR Disrupt Neuronal Ciliogenesis, Leading to Focal Cortical Dyslamination.* Neuron, 2018. **99**(1): p. 83-97.
239. Gangaram, B., W.P. Devine, and A. Slavotinek, *Expanding the phenotype of males with OFD1 pathogenic variants-a case report and literature review*, in *European journal of medical genetics*. 2022, Elsevier: Netherlands. p. 104496.
240. Venkatesan, C., et al., *Imaging Similarities Between Oral-Facial-Digital Syndrome Type 1 and Aicardi Syndrome: Prenatal and Postnatal Magnetic Resonance Imaging (MRI) Findings in 4 Patients.* Journal of Child Neurology, 2022: p. 1.
241. Miller, D.T., et al., *Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies.* The American Journal of Human Genetics, 2010. **86**(5): p. 749-764.
242. Battaglia, A., et al., *Confirmation of chromosomal microarray as a first-tier clinical diagnostic test for individuals with developmental delay, intellectual disability, autism spectrum disorders and dysmorphic features.* European Journal of Paediatric Neurology, 2013. **17**(6): p. 589-599.
243. Wenger, A.M., et al., *Systematic reanalysis of clinical exome data yields additional diagnoses: implications for providers.* Genetics in Medicine, 2017. **19**(2): p. 209-214.
244. Mei, D., et al., *The impact of next-generation sequencing on the diagnosis and treatment of epilepsy in paediatric patients.* Molecular diagnosis & therapy, 2017. **21**(4): p. 357-373.
245. Poulton, A., et al., *Family Communication about Diagnostic Genetic Testing for Younger-Onset Dementia.* Journal of Personalized Medicine, 2023. **13**: p. 621.
246. Vlaskamp, D.R.M., et al., *Changes in empowerment and anxiety of patients and parents during genetic counselling for epilepsy.* European Journal of Paediatric Neurology, 2021. **32**: p. 128-135.
247. Lionel, A.C., et al., *Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test.* Genetics in Medicine, 2018. **20**(4): p. 435-443.
248. Macdonald-Laurs, E., et al., *One-Stage, Limited-Resection Epilepsy Surgery for Bottom-of-Sulcus Dysplasia.* Neurology, 2021. **97**(2): p. 65-65.
249. Hara, T., et al., *The UDP-galactose translocator gene is mapped to band Xp11.23-p11.22 containing the Wiskott-Aldrich syndrome locus.* Somat Cell Mol Genet, 1993. **19**(6): p. 571-5.

250. Maszczak-Seneczko, D., et al., *UDP-N-acetylglucosamine transporter and UDP-galactose transporter form heterologous complexes in the Golgi membrane.* FEBS Lett, 2012. **586**(23): p. 4082-7.
251. Maszczak-Seneczko, D., T. Olczak, and M. Olczak, *Subcellular localization of UDP-GlcNAc, UDP-Gal and SLC35B4 transporters.* Acta Biochim Pol, 2011. **58**(3): p. 413-9.
252. Vals, M.A., et al., *Clinical, neuroradiological, and biochemical features of SLC35A2-CDG patients.* Journal of Inherited Metabolic Disease, 2019. **42**.
253. Milena, G., et al., *Perinatal manifestations of congenital disorders of glycosylation—A clue to early diagnosis*, in *Frontiers in Genetics*. 2022, Frontiers Media S.A.
254. Westenfield, K., et al., *Mosaicism of the UDP-Galactose transporter SLC35A2 in a female causing a congenital disorder of glycosylation: a case report.* BMC Medical Genetics, 2018. **19**(1): p. 100.
255. Ng, Bobby G., et al., *Mosaicism of the UDP-Galactose Transporter SLC35A2 Causes a Congenital Disorder of Glycosylation.* The American Journal of Human Genetics, 2013. **92**(4): p. 632-636.
256. Ng, B., et al., *SLC35A2-CDG: Functional Characterization, Expanded Molecular, Clinical, and Biochemical Phenotypes of 30 Unreported Individuals.* Human Mutation, 2019. **40**.
257. Kodera, H., et al., *De Novo Mutations in SLC35A2 Encoding a UDP-Galactose Transporter Cause Early-Onset Epileptic Encephalopathy.* Human Mutation, 2013. **34**(12): p. 1708-1714.
258. Sim, N.S., et al., *Brain somatic mutations in SLC35A2 cause intractable epilepsy with aberrant N-glycosylation.* Neurology: Genetics, 2018. **4**(6).
259. Miyatake, S., et al., *De novo ATP1A3 variants cause polymicrogyria.* Science Advances, 2022. **7**(13): p. eabd2368.
260. Hartlieb, T., et al., *Age-related MR characteristics in mild malformation of cortical development with oligodendroglial hyperplasia and epilepsy (MOGHE).* Epilepsy and Behavior, 2019.
261. Goldman, A.M. and K.L.L. Thio, *SLC35A2–Related Epilepsy.* Neurology (Ovid), 2023. **100**(5): p. 225-226.
262. Barba, C., et al., *Clinical Features, Neuropathology, and Surgical Outcome in Patients With Refractory Epilepsy and Brain Somatic Variants in the SLC35A2 Gene.* Neurology, 2023. **100**(5): p. e528-e542.
263. Aledo-Serrano, A., et al., *D-galactose supplementation for the treatment of patients with mild malformation of cortical development with oligodendroglial hyperplasia in epilepsy (MOGHE): An interim analysis of a proof-of-concept trial.* EPILEPSIA, 2022. **63**: p. 22-23.
264. Miller, K.E., et al., *Somatic SLC35A2 mosaicism correlates with clinical findings in epilepsy brain tissue.* Neurology: Genetics, 2020. **6**(4): p. e460-e460.
265. Prabhu, S., et al., *Myelination clock: a simplified step-by-step approach to normal myelination.* imaging, 2016. **1**(3): p. 4.
266. Rademacher, J., et al., *Measuringin vivomyelination of human white matter fiber tracts with magnetization transfer MR.* Neuroimage, 1999. **9**(4): p. 393-406.

267. YAKOVLEV, P.I. and S. LOCKE, *Limbic nuclei of thalamus and connections of limbic cortex: III. Corticocortical connections of the anterior cingulate gyrus, the cingulum, and the subcallosal bundle in monkey.* Archives of neurology, 1961. **5**(4): p. 364-400.
268. Staudt, M., et al., *MRI assessment of myelination: an age standardization.* Pediatric Radiology: Roentgenology, Nuclear Medicine, Ultrasound, CT, MRI, 1994. **24**(2): p. 122-127.
269. Fadeel, S.R.A., et al., *The role of diffusion weighted magnetic resonance imaging in assessment of normal myelination in infantile brain.* Alexandria Journal of Medicine, 2015. **51**(3): p. 271-276.
270. Kim, J.H., et al., *Ultra-Low Level Somatic Mutations and Structural Variations in Focal Cortical Dysplasia Type II.* Annals of Neurology, 2023. **93**(6): p. 1082-1093.
271. Wiktor, M., et al., *Identification of novel potential interaction partners of UDP-galactose (SLC35A2), UDP-N-acetylglucosamine (SLC35A3) and an orphan (SLC35A4) nucleotide sugar transporters.* Journal of proteomics, 2021: p. 104321.
272. Reily, C., et al., *Glycosylation in health and disease.* Nature Reviews Nephrology, 2019. **15**(6): p. 346-366.
273. Xia, B., et al., *Serum N-glycan and O-glycan analysis by mass spectrometry for diagnosis of congenital disorders of glycosylation.* Analytical biochemistry, 2013. **442**(2): p. 178-185.
274. Witters, P., et al., *Clinical and biochemical improvement with galactose supplementation in SLC35A2-CDG.* Genetics in Medicine, 2020. **22**(6): p. 1102-1107.
275. Goldman, A.M., *What does a defect in N-glycosylation mean for neuronal migration and function?* Neurology Genetics, 2020. **6**(4): p. e490.
276. Tang, F., et al., *mRNA-Seq whole-transcriptome analysis of a single cell.* Nature methods, 2009. **6**(5): p. 377-382.
277. Cuevas-Diaz Duran, R., et al., *Single-cell and single-nuclei RNA sequencing as powerful tools to decipher cellular heterogeneity and dysregulation in neurodegenerative diseases.* Frontiers in Cell and Developmental Biology, 2022. **10**: p. 884748.
278. Lee, W.S., et al. *Cortical Dysplasia and the mTOR Pathway: How the Study of Human Brain Tissue Has Led to Insights into Epileptogenesis.* International Journal of Molecular Sciences, 2022. **23**, DOI: 10.3390/ijms23031344.
279. Winkler, E.A., et al., *A single-cell atlas of the normal and malformed human brain vasculature.* Science, 2022. **375**(6584): p. eabi7377.
280. Wang, Z., et al., *High-throughput and deep-proteome profiling by 16-plex tandem mass tag labeling coupled with two-dimensional chromatography and mass spectrometry.* JoVE (Journal of Visualized Experiments), 2020(162): p. e61684.
281. Li, J., et al., *TMTpro reagents: a set of isobaric labeling mass tags enables simultaneous proteome-wide measurements across 16 samples.* Nature methods, 2020. **17**(4): p. 399-404.

282. Ding, X., et al., *Inhibition of Grb14, a negative modulator of insulin signaling, improves glucose homeostasis without causing cardiac dysfunction.* Scientific reports, 2020. **10**(1): p. 1-10.
283. Pfisterer, U., et al., *Identification of epilepsy-associated neuronal subtypes and gene expression underlying epileptogenesis.* Nature Communications, 2020. **11**(1): p. 5038.
284. de Curtis, M., R. Garbelli, and L. Uva, *A hypothesis for the role of axon demyelination in seizure generation.* Epilepsia, 2021. **62**(3): p. 583-595.
285. Proietti Onori, M., et al., *RHEB/mTOR hyperactivity causes cortical malformations and epileptic seizures through increased axonal connectivity.* PLoS Biology, 2021. **19**(5): p. e3001279.
286. Xu, J., et al., *Nuclear SOD1 in Growth Control, Oxidative Stress Response, Amyotrophic Lateral Sclerosis, and Cancer.* Antioxidants, 2022. **11**(2): p. 427.
287. de Souza, P.V.S., et al., *Progressive spastic tetraplegia and axial hypotonia (STAHP) due to SOD1 deficiency: is it really a new entity?* Orphanet Journal of Rare Diseases, 2021. **16**(1).
288. Fischer, L.R., et al., *Absence of SOD1 leads to oxidative stress in peripheral nerve and causes a progressive distal motor axonopathy.* Experimental neurology, 2012. **233**(1): p. 163-171.
289. Haider, S. and R. Pal, *Integrated analysis of transcriptomic and proteomic data.* Curr Genomics, 2013. **14**(2): p. 91-110.
290. Oh, W.-J. and C. Gu, *The role and mechanism-of-action of Sema3E and Plexin-D1 in vascular and neural development.* Seminars in Cell and Developmental Biology, 2013. **24**(3): p. 156-162.
291. Velona, T., et al., *PlexinD1 and Sema3E determine laminar positioning of heterotopically projecting callosal neurons.* Molecular and Cellular Neuroscience, 2019. **100**.
292. Sánchez, C., J. Dıaz-Nido, and J. Avila, *Phosphorylation of microtubule-associated protein 2 (MAP2) and its relevance for the regulation of the neuronal cytoskeleton function.* Progress in neurobiology, 2000. **61**(2): p. 133-168.
293. Tokuraku, K., et al., *Distinct neuronal localization of microtubule-associated protein 4 in the mammalian brain.* Neuroscience Letters, 2010. **484**(2): p. 143-147.
294. Nishida, K., et al., *Effects of three microtubule-associated proteins (MAP2, MAP4, and Tau) on microtubules' physical properties and neurite morphology.* Scientific Reports, 2023. **13**(1): p. 8870.
295. Khalil, M., et al., *Neurofilaments as biomarkers in neurological disorders.* Nature Reviews Neurology, 2018. **14**(10): p. 577-589.
296. Nakajima, K., et al., *Molecular motor KIF5A is essential for GABAA receptor transport, and KIF5A deletion causes epilepsy.* Neuron, 2012. **76**(5): p. 945-961.
297. Willemsen, M.H., et al., *Involvement of the kinesin family members KIF4A and KIF5C in intellectual disability and synaptic function.* Journal of medical genetics, 2014. **51**(7): p. 487-494.

298. Meyer, T., et al., *Neurofilament light-chain response during therapy with antisense oligonucleotide tofersen in SOD1-related ALS: Treatment experience in clinical practice.* Muscle & Nerve, 2023. **67**(6): p. 515-521.
299. Dang Do, A.N., et al., *Cerebrospinal Fluid Protein Biomarker Discovery in CLN3.* Journal of Proteome Research, 2023. **22**(7): p. 2493-2508.
300. Wingo, A.P., et al., *Shared proteomic effects of cerebral atherosclerosis and Alzheimer's disease on the human brain.* Nature Neuroscience, 2020. **23**(6): p. 696-700.
301. Yum, S.W., et al., *A novel recessive NEFL mutation causes a severe, early-onset axonal neuropathy.* Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society, 2009. **66**(6): p. 759-770.
302. Alami, N.H., et al., *Axonal transport of TDP-43 mRNA granules is impaired by ALS-causing mutations.* Neuron, 2014. **81**(3): p. 536-543.
303. Russell, S.A. and G.J. Bashaw, *Axon guidance pathways and the control of gene expression.* Developmental Dynamics, 2018. **247**(4): p. 571-580.
304. Zhang, Y., et al., *GNG7 and ADCY1 as diagnostic and prognostic biomarkers for pancreatic adenocarcinoma through bioinformatic-based analyses.* Scientific Reports, 2021. **11**(1): p. 20441.
305. Duan, H., et al., *Identification of GNG7 as a novel biomarker and potential therapeutic target for gastric cancer via bioinformatic analysis and in vitro experiments.* Aging (Albany NY), 2023. **15**(5): p. 1445-1474.
306. Schwindinger, W.F., et al., *Loss of G Protein γ₇ Alters Behavior and Reduces Striatal α_{olf} Level and cAMP Production **.* Journal of Biological Chemistry, 2003. **278**(8): p. 6575-6579.
307. Yoshikawa, H., *Myelin-associated oligodendrocytic basic protein modulates the arrangement of radial growth of the axon and the radial component of myelin.* Medical Electron Microscopy, 2001. **34**: p. 160-164.
308. Schäfer, I., et al., *MOBP levels are regulated by Fyn kinase and affect the morphological differentiation of oligodendrocytes.* Journal of cell science, 2016. **129**(5): p. 930-942.
309. Kaushansky, N., et al., *The myelin-associated oligodendrocytic basic protein (MOBP) as a relevant primary target autoantigen in multiple sclerosis.* Autoimmunity reviews, 2010. **9**(4): p. 233-236.
310. Reindl, M. and P. Waters, *Myelin oligodendrocyte glycoprotein antibodies in neurological disease.* Nature Reviews Neurology, 2019. **15**(2): p. 89-102.
311. Reindl, M., et al., *Antibodies against the myelin oligodendrocyte glycoprotein and the myelin basic protein in multiple sclerosis and other neurological diseases: a comparative study.* Brain, 1999. **122**(11): p. 2047-2056.
312. Juryńczyk, M., et al., *Myelin oligodendrocyte glycoprotein (MOG) antibody-associated disease: practical considerations.* Practical neurology, 2019. **19**(3): p. 187-195.
313. Asp, M., J. Bergenstråhle, and J. Lundeberg, *Spatially resolved transcriptomes—next generation tools for tissue exploration.* BioEssays, 2020. **42**(10): p. 1900221.
314. Duncan, D.J., et al., *Assessment of PD-L1 mRNA and protein expression in non-small cell lung cancer, head and neck squamous cell carcinoma and*

urothelial carcinoma tissue specimens using RNAScope and immunohistochemistry. PLoS One, 2019. **14**(4): p. e0215393.
315. Williams, C.G., et al., *An introduction to spatial transcriptomics for biomedical research*. Genome Medicine, 2022. **14**(1): p. 1-18.
316. Michael, S.H., et al., *Severe childhood speech disorder*. Neurology, 2020. **94**(20): p. e2148.
317. Skotte, L., et al., *Genome-wide association study of febrile seizures implicates fever response and neuronal excitability genes*. Brain, 2022. **145**(2): p. 555-568.
318. International League Against Epilepsy Consortium on Complex, E., et al., *Genome-wide meta-analysis of over 29,000 people with epilepsy reveals 26 loci and subtype-specific genetic architecture*. medRxiv, 2022: p. 2022.06.08.22276120.
319. Sadikovic, B., et al., *Clinical epigenomics: genome-wide DNA methylation analysis for the diagnosis of Mendelian disorders*. Genetics in Medicine, 2021. **23**(6): p. 1065-1074.
320. Haghshenas, S., et al., *Diagnostic utility of genome-wide DNA methylation analysis in mendelian neurodevelopmental disorders*. International journal of molecular sciences, 2020. **21**(23): p. 9303.
321. Hood, R.L., et al., *The defining DNA methylation signature of Floating-Harbor Syndrome*. Scientific reports, 2016. **6**(1): p. 38803.
322. Jabari, S., et al., *DNA methylation-based classification of malformations of cortical development in the human brain*. Acta Neuropathologica, 2022. **143**(1): p. 93.
323. Amarasinghe, S.L., et al., *Opportunities and challenges in long-read sequencing data analysis*. Genome Biology, 2020. **21**(1): p. 30.
324. Martin, S., et al., *Nanopore adaptive sampling: a tool for enrichment of low abundance species in metagenomic samples*. Genome biology, 2022. **23**(1): p. 11.
325. Loose, M., S. Malla, and M. Stout, *Real-time selective sequencing using nanopore technology*. Nature methods, 2016. **13**(9): p. 751-754.
326. Shih, P.J., et al., *Efficient real-time selective genome sequencing on resource-constrained devices*. GigaScience, 2023. **12**: p. giad046.
327. Bourdillon, P., et al., *Stereo electroencephalography–guided radiofrequency thermocoagulation (SEEG-guided RF-TC) in drug-resistant focal epilepsy: Results from a 10-year experience*. Epilepsia, 2017. **58**(1): p. 85-93.
328. Medvid, R., et al., *Current applications of MRI-guided laser interstitial thermal therapy in the treatment of brain neoplasms and epilepsy: a radiologic and neurosurgical overview*. American Journal of Neuroradiology, 2015. **36**(11): p. 1998-2006.
329. Checri, R., et al., *Detection of brain somatic mutations in focal cortical dysplasia during epilepsy presurgical workup*. Brain Communications, 2023. **5**(3).
330. Li, M., et al., *Antisense oligonucleotide therapy reduces seizures and extends life span in an SCN2A gain-of-function epilepsy model*. The Journal of clinical investigation, 2021. **131**(23).

331. Doyle, A., et al., *The construction of transgenic and gene knockout/knockin mouse models of human disease.* Transgenic Research, 2012. **21**(2): p. 327-349.
332. Hedrich, U.B., et al., *Impaired action potential initiation in GABAergic interneurons causes hyperexcitable networks in an epileptic mouse model carrying a human NaV1. 1 mutation.* Journal of Neuroscience, 2014. **34**(45): p. 14874-14889.
333. Doss, M.X. and A. Sachinidis, *Current Challenges of iPSC-Based Disease Modeling and Therapeutic Implications.* Cells, 2019. **8**(5): p. 403.
334. Higurashi, N., et al., *A human Dravet syndrome model from patient induced pluripotent stem cells.* Molecular brain, 2013. **6**: p. 1-12.
335. Liu, J., et al., *CRISPR/Cas9 facilitates investigation of neural circuit disease using human iPSCs: mechanism of epilepsy caused by an SCN1A loss-of-function mutation.* Translational psychiatry, 2016. **6**(1): p. e703-e703.
336. Liu, Y., et al., *Dravet syndrome patient-derived neurons suggest a novel epilepsy mechanism.* Annals of neurology, 2013. **74**(1): p. 128-139.
337. Jiao, J., et al., *Modeling Dravet syndrome using induced pluripotent stem cells (iPSCs) and directly converted neurons.* Human molecular genetics, 2013. **22**(21): p. 4241-4252.

8 Appendices

8.1 Supplementary Tables

Supplementary Table 8.1: List of genes on infantile epileptic spasms syndrome (IESS) panel (130 genes) and malformations of cortical development (MCD) mosaicism panel (50 genes).

IESS Genes:				
AKT1	*EEF1A2*	*HRAS*	*NDP*	*SIK1*
AKT3	*EHMT1*	*HUWE1*	*NEDD4L*	*SIX3*
ALG13	*EIF2S3*	*IFIH1*	*NF1*	*SLC16A2*
ARID1B	*FGF12*	*IKBKG*	*NPRL2*	*SLC1A2*
ARX	*FGFR1*	*IQSEC2*	*NPRL3*	*SLC35A2*
ATP2A2	*FLNA*	*IRF2BPL*	*NRAS*	*SLC9A6*
ATP7A	*FOXG1*	*KANSL1*	*NSD1*	*SMC1A*
BRAF	*GABRA1*	*KCNA2*	*NSDHL*	*SPTAN1*
CACNA1A	*GABRB2*	*KCNB1*	*OFD1*	*STX1B*
CASK	*GABRB3*	*KCNC2*	*PAFAH1B1*	*STXBP1*
CCND2	*GABRG2*	*KCNJ11*	*PDHA1*	*SYNGAP1*
CDK19	*GFAP*	*KCNQ2*	*PHACTR1*	*TBL1XR1*
CDKL5	*GLI3*	*KCNQ5*	*PIGA*	*TCF4*
CLCN4	*GLUD1*	*KCNT1*	*PIK3CA*	*TREX1*
CLCN6	*GNAO1*	*KIAA2022*	*PIK3R2*	*TRIM8*
COL4A1	*GNAQ*	*KIF1A*	*PPP3CA*	*TSC1*
COL4A2	*GNB1*	*KIF2A*	*PRRT2*	*TSC2*
CYFIP1	*GRIA2*	*KIF5C*	*PTEN*	*TUBA1A*
CYFIP2	*GRIN1*	*KMT2D*	*PURA*	*TUBB2A*
DCX	*GRIN2A*	*KRAS*	*SCN1A*	*TUBB2B*
DDX3X	*GRIN2B*	*MAP2K1*	*SCN2A*	*TUBB3*
DEPDC5	*GRIN2D*	*MAP2K2*	*SCN8A*	*TUBG1*
DNM1	*HCCS*	*MECP2*	*SETBP1*	*UBE3A*
DNM1L	*HCFC1*	*MEF2C*	*SETD1B*	*WDR45*
DYNC1H1	*HECW2*	*MTOR*	*SETD5*	*YWHAG*
DYRK1A	*HNRNPU*	*NACC1*	*SHH*	*ZEB2*
MCD Mosaicism Genes:				
AKT1	*LAMTOR2*	*PIK3CG*	*RRAGA*	*SLC35A2*
AKT3	*LAMTOR3*	*PIK3R2*	*RRAGB*	*SLC35A4*
BRAF	*LAMTOR4*	*PIK3R2*	*RRAGC*	*STRADA*
CCND2	*LAMTOR5*	*PRKAA1*	*RRAGD*	*SZT2*
DEPDC5	*MIOS*	*PTEN*	*SCN2A*	*TBC1D7*
EIF4EBP1	*MLST8*	*RHEB*	*SEC13*	*TSC1*
FKBP1A	*MTOR*	*RICTOR*	*SEH1L*	*TSC2*
IKBKB	*NPRL2*	*RPS6KA1*	*SESN1*	*WDR24*
LAMP1	*NPRL3*	*RPS6KB1*	*SESN2*	*WDR59*
LAMTOR1	*PIK3CA*	*RPTOR*	*SESN3*	*YWHAB*

Supplementary Table 8.2: List of genes targeted by HaloPlex^{HS} panel sequencing.

ACE	CEP41	ERMARD	KCNJ15	NF2	PIK3CB	RARS2	TCTN2
ACTB	CEP55	ETV6	KCTD17	NHEJ1	PIK3CD	RBBP8	TCTN3
ACTG1	CEP63	EXOSC3	KDM6A	NHLRC1	PIK3CG	RCAN1	TEAD4
AGL	CHD8	FARP2	KIAA0556	NIN	PIK3IP1	RELN	TELO2
AGT	CHMP1A	FAT4	KIAA0586	NKX2-1	PIK3R1	RHEB	TGIF1
AHI1	CIZ1	FBXO7	KIAA1279	NOV	PIK3R2	RICTOR	TH
AKT1	CNTNAP2	FECH	KIAA1549	NPHP1	PIK3R3	RIN2	THAP1
AKT1S1	COL18A1	FGF8	KIF11	NPHP3	PIK3R4	ROBO1	TMEM107
AKT3	COL4A1	FGFR1	KIF14	NPPB	PIK3R5	RPGRIP1L	TMEM138
ALDOA	COL4A2	FGFR2	KIF1A	NPRL2	PIK3R6	RPS6KA1	TMEM216
AMPD1	COL6A3	FGFR3	KIF1C	NPRL3	PIKFYVE	RPS6KB1	TMEM231
AMPD2	CSPP1	FH	KIF20A	NRAS	PINK1	RPTOR	TMEM237
ARFGEF2	CTGF	FIG4	KIF21A	NSD1	PLA2G6	RRAGA	TMEM5
ARHGAP31	CXCL2	FKRP	KIF2A	NTN1	PLK4	RRAGB	TMEM67
ARHGEF7	CXCR4	FKTN	KIF5C	OAT	PNKD	RRAGC	TNN
ARHGEF9	DCC	FLNA	KIF7	OCLN	PNKP	RRAGD	TOR1A
ARID1A	DCHS1	FOXG1	KMT2D	OFD1	POLG	RRP12	TRAPPC9
ARID1B	DCLK2	FOXP2	KRAS	OLIG2	POMGNT1	RTTN	TRMT10A
ARL13B	DCTN1	FST	L1CAM	OPHN1	POMGNT2	SAP130	TSC1
ARX	DCTN2	G6PC	LAMA1	ORC1	POMT1	SCARB1	TSC2
ASNS	DCX	G6PC3	LAMA2	ORC4	POMT2	SEC13	TSEN2
ASPM	DEPDC1	G6PD	LAMB1	ORC6	PPP1R3A	SEH1L	TSEN34
ASXL3	DEPDC1B	GAA	LAMC3	PACRG	PPP1R3B	SEPSECS	TSEN54
ATP13A2	DEPDC4	GANC	LAMP2	PAFAH1B1	PPP1R3C	SESN1	TTC21B
ATP6AP2	DEPDC5	GAS7	LAMTOR1	PANK2	PPP1R3D	SESN2	TTI1
ATP6V0A2	DEPDC7	GBA	LAMTOR2	PARK2	PPP2CA	SESN3	TTI2
ATR	DEPTOR	GBE1	LAMTOR3	PARK7	PQBP1	SGCE	TUBA1A
ATRX	DLGAP4	GCH1	LAMTOR4	PAX6	PREX1	SHH	TUBA8
B3GALNT2	DMD	GIGYF2	LAMTOR5	PCNT	PREX2	SIX3	TUBB
B3GNT1	DMRTA2	GLI2	LARGE	PDE6D	PRICKLE1	SLC12A6	TUBB2A
B9D1	DNAH11	GLI3	LDHA	PEX1	PRKAA1	SLC17A3	TUBB2B
B9D2	DNAJC13	GMPPB	LGI1	PEX10	PRKAA2	SLC25A19	TUBB3
BAX	DNAJC6	GNAL	LIG4	PEX12	PRKAB1	SLC2A1	TUBB4A
BMPR1B	DNAL4	GNAQ	LPIN1	PEX13	PRKAB2	SLC2A1	TUBB4A
BRAF	DRD2	GPR56	LRRK2	PEX14	PRKAG1	SLC2A2	TUBG1
BTD	DRD5	GPSM2	MAK16	PEX16	PRKAG2	SLC2A4	TUBG1
C5orf42	DYNC1H1	GPSM2	MAPKAP1	PEX19	PRKAG2	SLC2A7	TUBGCP6
CABIN1	DYNC1I2	GRIK1	MAPT	PEX2	PRKAG3	SLC37A4	UBE3A
CACNA1B	DYNC1U1	GSK3A	MAPT	PEX26	PRKAG3	SLC6A3	UCHL1
CACNA1G	DYNC1U2	GSK3B	MAST4	PEX3	PRKAR1B	SLC9A6	USP54
CASC5	DYNLRB1	GYG1	MBD5	PEX5	PRKCA	SNAP29	UTRN
CASK	DYNLRB2	GYS1	MCPH1	PEX6	PRKRA	SNCA	VLDLR
CC2D2A	DYNLT1	GYS2	MDM2	PFKM	PRR5	SPINK1	VPS35
CCDC147	DYNLT3	HAMP	MECP2	PGAM2	PRR5L	SPR	VRK1
CCND1	DYRK1A	HEPACAM	MED17	PGK1	PRRT2	SRD5A3	VWF
CCND2	EBLN1	HNRNPK	MGAM	PGM1	PTCH1	SRPX2	WDR24
CDC6	EBLN2	HPCA	MIOS	PHC1	PTEN	STAMBP	WDR59
CDK5	EGFR	HRAS	MKLN1	PHF20L1	PTK2	STIL	WDR62
CDK5RAP2	EIF2AK2	HSD11B1	MKS1	PHKA1	PXN	STK25	WDR73
CDK6	EIF4B	HSD17B4	MLST8	PHKA2	PYGL	STRADA	YWHAE
CDKL5	EIF4E	HTRA2	MTOR	PHKB	PYGM	STRADA	ZEB2
CDON	EIF4EBP1	HYLS1	MYC	PHKG2	RAB18	STRADB	ZIC2
CDT1	EIF4G1	IER3IP1	MYCN	PI4KA	Rab39b	SYNJ1	ZNF335
CENPE	EML1	IL8	MYO18B	PI4KA	RAB3GAP1	TAF1	
CENPF	EML4	INPP5E	MYOF	PICK1	RAB3GAP2	TAOK1	
CENPJ	EML4	INS	NBN	PIK3C2A	RAB7L1	TBC1D20	
CEP104	EMX2	INTU	NCOA2	PIK3C2B	RAD50	TBC1D7	
CEP135	ENO3	IQGAP3	NDE1	PIK3C2G	RAD51	TCF3	
CEP152	EOMES	ISPD	NDEL1	PIK3C3	RAPGEF3	TCF4	
CEP290	EPM2A	KATNB1	NF1	PIK3CA	RAPGEF4	TCTN1	

Supplementary Table 8.3: List of custom Taqman probes used for Droplet Digital PCR (ddPCR) single nucleotide polymorphism (SNP) genotyping assays in this study.

Custom Assay ID	Gene	Transcript	Variant (hg38)	Protein Change	Catalog ID
ANAAUAN	*SLC35A2*	NM_005660	c.136C>T	p.Q46*	4331349
AN2XPTG	*SLC35A2*	NM_005660	c.206C>T	p.T69I	4331349
ANXG72P	*SLC35A2*	NM_005660	c.359_360delTC	p.L120Hfs*7	4331349
ANT2RCX	*SLC35A2*	NM_005660	c.502C>T	p.Q168*	4331349
AN9H3UJ	*SLC35A2*	NM_005660	c.511T>C	p.S171P	4331349
ANWDEGT	*SLC35A2*	NM_005660	c.547C>T	P.Q183*	4331349
ANU7JWV	*SLC35A2*	NM_005660	c.553C>T	p.Q185*	4331349
AN9H2WX	*SLC35A2*	NM_005660	c.626_628delCCT	p.S209del	4331349
ANYM2MM	*SLC35A2*	NM_005660	c.665_667delAGA	p.K222del	4331349

Note: All probes available for order at Thermo Fisher Scientific using Assay IDs.

Supplementary Table 8.4: Variants of unknown significance (VUS) identified in IESS cohort.

Individual ID	Sex	Neuropathology	Gene	Transcript	Variant	AA Change	VAF	Inheritance	Solve method
AA0192-01	F	FCD IIA	*GNA01*	NM_138736.3	c.448A>C	N150H	18.0%	Mosaic	ES (seqr)
AA0193-01	F	FCD IIA	*SCN2A*	NM_001040142.2	c.386+2T>C	SS variant	8.3%	Mosaic	ES (seqr)
AA2569-01	M	Gliosis	*SCN8A*	NM_014191.4	c.3563G>A	R1188Q	51.0%	Germline (Pat)	ES (seqr)
AA2569-01	M	Gliosis	*FOXG1*	NM_005249.5	c.460delG	E154Rfs*38	3.5%	Mosaic	ES (seqr)

AA, amino acid; F, female; FCD IIA, focal cortical dysplasia type IIA; M, male; Pat, paternal; SS, splice site; VAF, variant allele frequency.

Supplementary Table 8.5: Extended clinical table of IESS with MCD cohort.

ID	Sex	Revised Diagnosis	Initial Diagnosis	Age seizure onset (months)	Surgery age (years)	Surgery extent	Surgery type	Engel class	Seizure type at onset	Other seizure types	Development prior to seizure onset	Development at last review
AA0068-01	Male	Complex MCD	Complex MCD	9	8	MLo	R TPO	III	Spasms	Focal	Delayed	ID, ASD
AA0214-01	Female	DNET	DNET	4	1.6, 3	Lo, Lo	R F, R P	IV	Focal	Spasms	Normal	Global DD, ASD
AA0221-01	Female	FCD I	FCD I	1	0.8	MLo	L TPO	III	Focal	Spasms	Normal	Borderline
AA0265-01	Male	FCD I	FCD IIA	4	4	Lo	R F	I	Spasms	N/A	Normal	Learning difficulties
AA0294-01	Male	FCD I	FCD IIA	6	3	H	L H	I	Spasms	N/A	UK	Global DD
AA0327-01	Male	FCD I	FCD I	11	1.5	MLo	L TPO	I	Spasms	N/A	Normal	Global DD
AA0171-01	Male	FCD IIA	FCD IIA	2	3	Lo	R T	I	Spasms	Focal	Normal	Global DD, ASD
AA0235-01	Female	FCD IIA	FCD IIA	5	0.5, 2	MLo	L I/P, L I/P	IV	Focal	Spasms	Normal	Learning difficulties
AA0239-01	Female	FCD IIA	Non-specific	0	0.7	H	R H	I	Focal	Spasms	N/A (neonatal onset)	Global DD
AA1485-01	Female	FCD IIA	FCD IIA	4	7	Lo	R F	I	Spasms	Focal	Normal	Normal
AA2691-01	Female	FCD IIA	Non-specific	0	0.1	H (incomplete)	L H (incomplete)	I	Spasms, Tonic	N/A	N/A (neonatal onset)	Global DD
AA0096-01	Male	FCD IIA/HMEG	FCD IIA/HMEG	0	0.2, 0.6, 5, 9	MLo, MLo, Lo, MLo	L TPO, L TPO, L P, L TPO	I	Focal	Spasms	N/A (neonatal onset)	ID (mod), ASD
AA0097-01	Male	FCD IIA/HMEG	FCD IIA/HMEG	0	1.4, 3	H, Lo	R H (incomplete), R F	I	Focal	Spasms	N/A (neonatal onset)	ID, ASD
AA0175-01	Male	FCD IIA/HMEG	FCD IIA/HMEG	0	0.3, 6	MLo, H	R TPO/R F, R H	II	Focal	Spasms	N/A (neonatal onset)	Global DD

AA0192-01	Male	FCD IIA/HMEG	FCD IIA/HMEG	1	1.2, 4	MLo, Lo	R F-P-I, R F	II	Spasms	Focal	UK	ID (mild)
AA0342-01	Female	FCD IIA/HMEG	FCD IIA/HMEG	0	0.2	H	R H	I	Focal	Spasms	N/A (neonatal onset)	Global DD
AA0404-01	Male	FCD IIA/HMEG	FCD IIA/HMEG	0	1	H	R H	I	Focal	Spasms	N/A (neonatal onset)	Global DD
AA0199-01	Male	FCD IIB	FCD IIB	15	2, 3	Lo, Lo	R F, R F	I	Spasms	Focal	Normal	ID (mild), ASD
AA0937-01	Male	Gliosis	Gliosis	6	3	MLo, Lo	R TPO	IV	Spasms	Tonic	Normal	Global DD
AA2569-01	Female	Gliosis	Gliosis	1	3, 4	H, Lo	R H, R F	IV	Focal	Spasms, myoclonic	Normal	Global DD
AA0084-01	Male	MOGHE	FCD IIA	3	2.7, 7	Lo, Lo	L F, L F	I	Spasms	Focal (reflex)	Normal	Normal
AA0148-01	Male	MOGHE	FCD IA	16	3,4,8	Lo, MLo, H	L F, L F/T/I, L H	I	Spasms	Tonic	Delayed	ID (mod-severe)
AA0161-01	Male	MOGHE	FCD IIA	6	3, 7	Lo, MLo	R T, R TPO	UK	Spasms	Focal	Normal	ID (mod), ASD
AA0164-01	Female	MOGHE	Gliosis	4	2, 4	Lo, H	R F, R H	I	Spasms	N/A	Normal	Global DD, ASD
AA0190-01	Female	MOGHE	FCD IIA	2	6	Lo	R F	I	Spasms	Focal (reflex)	Normal	Normal
AA0228-01	Female	MOGHE	FCD IA	14	1.8, 4	ML, L	L TPO, L F	I	Spasms	Tonic	Normal	ASD (without ID)
AA0236-01	Male	MOGHE	NAD	5	1.3	MLo	L TPO	II	Spasms	N/A	Normal	ASD
AA1391-01	Female	MOGHE	NAD	7	2	MLo	R TPO	I	Spasms	N/A	UK	ID
AA2561-01	Male	MOGHE	Non-specific	7	3	MLo	R TPO	I	Spasms	Tonic	Normal	Global DD
AA0203-01	Male	NAD	NAD	9	3	MLo	L TPO	I	Spasms	Tonic	Normal	ID, ASD
AA0087-01	Male	Non-specific	Non-specific	2	7	Lo	L T	IV	Spasms	Focal, tonic	Delayed	ID (severe)
AA0215-01	Female	Non-specific	Non-specific	6	2, 4	Lo, Lo	L F, L F-C	IV	Spasms	Focal	Normal	Global DD
AA2537-01	Male	Non-specific	Non-specific	11	1.6	H	R H	I	Spasms	N/A	Delayed	Language delay
AA2575-01	Male	Non-specific	Non-specific	9	2.5	H	L H	IV	Focal, Spasms	Tonic	Delayed	Global DD
AA0099-01	Male	Non-specific	Non-specific	4	11	Lo	L T	III	Focal, Tonic	Spasms, tonic	Delayed	ID (severe)
AA0193-01	Female	Non-specific	Non-specific	4	2	Lo	L F	I	Spasms	Myoclonic, atonic	Normal	Normal
AA0017-01	Male	TSC	TSC	5	16	Les	L F	I	Spasms	UK	UK	Global DD
AA0033-01	Female	TSC	TSC	2	13	Les	R TP, L T, R TPO, R F	II	UK	Spasms	UK	UK
AA0092-01	Male	TSC	TSC	6	1.5, 2, 6	Les	R FC	II	UK	Spasms	UK	UK
AA0093-01	Female	TSC	TSC	9	2.5, 4.9	Les	R T	II	UK	Spasms	UK	UK

AA0140-01	Female	TSC	TSC	4	3.5	Les	R F	I	Spasms	UK	Normal	Global DD
AA0142-01	Male	TSC	TSC	9	3.8	Les	R FT	IV	Focal	Spasms	UK	UK
AA0143-01	Female	TSC	TSC	6	2.8	Les	L C	III	Spasms	Focal, generalised	UK	UK
AA0145-01	Female	TSC	TSC	0	1, 1.4, 7	Les	L T	II	Focal, Spasms	Focal, Spasms	UK	UK
AA0151-01	Female	TSC	TSC	2	13	Les	R F	I	Spasms	Tonic, focal	Normal	Global DD
AA0166-01	Male	TSC	TSC	8	4	Les	L F	II	Spasms	UK	UK	Global DD
AA0178-01	Female	TSC	TSC	4	2.8	Les	L C	II	Spasms	UK	UK	Global DD
AA0187-01	Male	TSC	TSC	1	0.7	Les	L T	III	Spasms	Febrile convulsion	Normal	ID
AA0195-01	Male	TSC	TSC	4	2.4	Les	R PFC	I	UK	Spasms	UK	UK
AA0224-01	Male	TSC	TSC	6	2.6	Les	R T	III	Spasms	UK	UK	UK
AA0264-01	Female	TSC	TSC	3	16	Les	L PCG	II	UK	Spasms	UK	UK
AA0302-01	Female	TSC	TSC	1	2	Les	L P	I	Spasms	UK	UK	Global DD
AA0313-01	Female	TSC	TSC	1	2, 3, 4, 5	Les	L F, R F, R P	IV	UK	Spasms	UK	UK
AA0316-01	Male	TSC	TSC	3	3.4	Les	L F	II	UK	Spasms	UK	UK
AA0329-01	Female	TSC	TSC	0	2	Les	R F	I	UK	Spasms	UK	UK
AA0936-01	Female	TSC	FCD IIB	7	1	Lo	R F	I	Spasms	N/A	Normal	Global DD
AA0938-01	Female	TSC	TSC	4	0.6, 2, 3	Les	R T, L PO	III	UK	Spasms	UK	UK
AA1027-01	Male	TSC	TSC	6	5	Les	R PO	II	Spasms	UK	Normal	Global DD
AA1408-01	Male	TSC	TSC	15	2	Les	L T	I	UK	Spasms	UK	UK

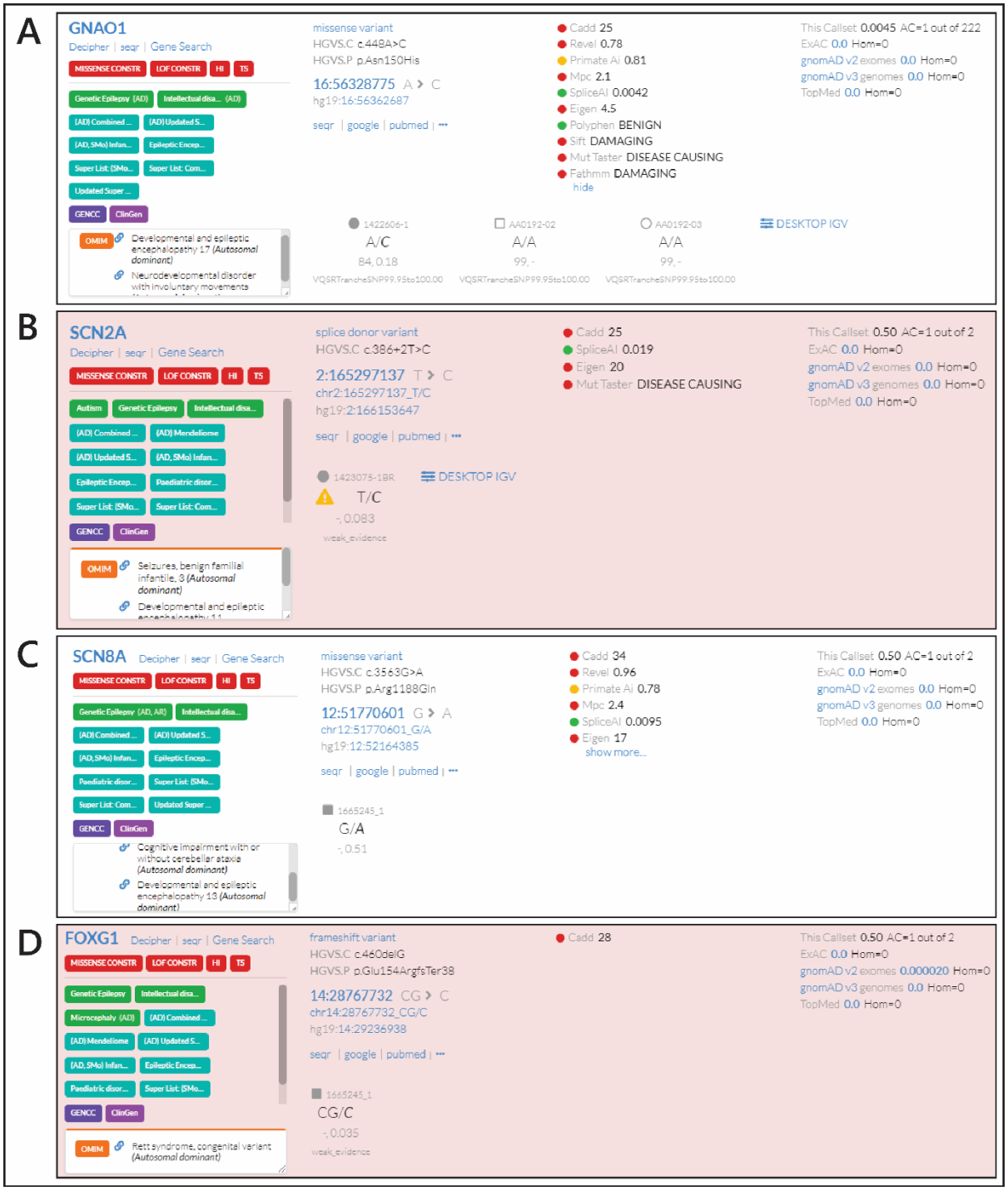
ASD, autism spectrum disorder; C, corticectomy; CMA, chromosomal microarray; DD, developmental delay; DNET, dysembryoplastic neuroepithelial tumour; ES, exome sequencing; F, frontal lobe; FCD, focal cortical dysplasia; H, hemispheric disconnection/hemispherectomy; HME, hemimegalencephaly; I, Engel class 1 (seizure free); ID, intellectual disability; II, Engel class 2; III, Engel class 3; IV, Engel class 4; I/P, inferior-parietal; L, left; Les, lesionectomy; Lo, lobar; MLo, multi-lobar; N/A, not applicable; NAD, no abnormalities detected; PCG, pre-central gyrus; R, right; T, temporal; TPO, temporo-parieto-occipital; TSC, tuberous sclerosis complex; UK, unknown.

Supplementary Table 8.6: Antibody concentrations (diluted in blocking buffer (TRIS/FBS).

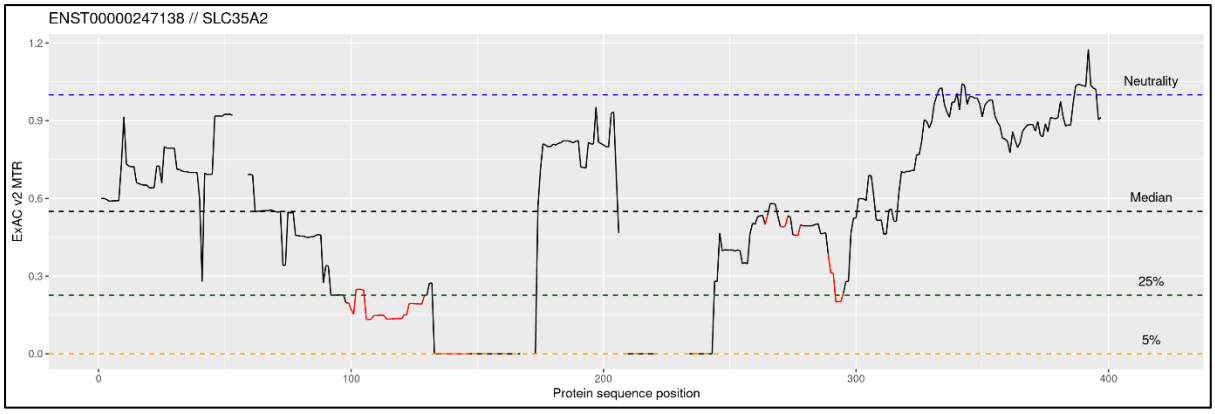
Antibody	Company	Cat #	Host	Storage	Concentration
pS6	Cell signalling	4858P	Rabbit	-20°C	1:100
SMI-311	Biolegend	837801	Mouse	-20°C	1:1000
SMI-311	Biolegend	837801	Mouse	-20°C	1:1000
HCN4	Alomone	APC-052	Rabbit	-20°C	1:500
Olig2	Ventana	EP112	Rabbit	4°C	Pre-diluted (1:1000)
NeuN	Millipore	MABN140	Rabbit	4°C	1:100
Map2	Sigma	M1406	Mouse	4°C	1:500

NB: secondary antibodies (Vector biotinylated secondary antibodies)

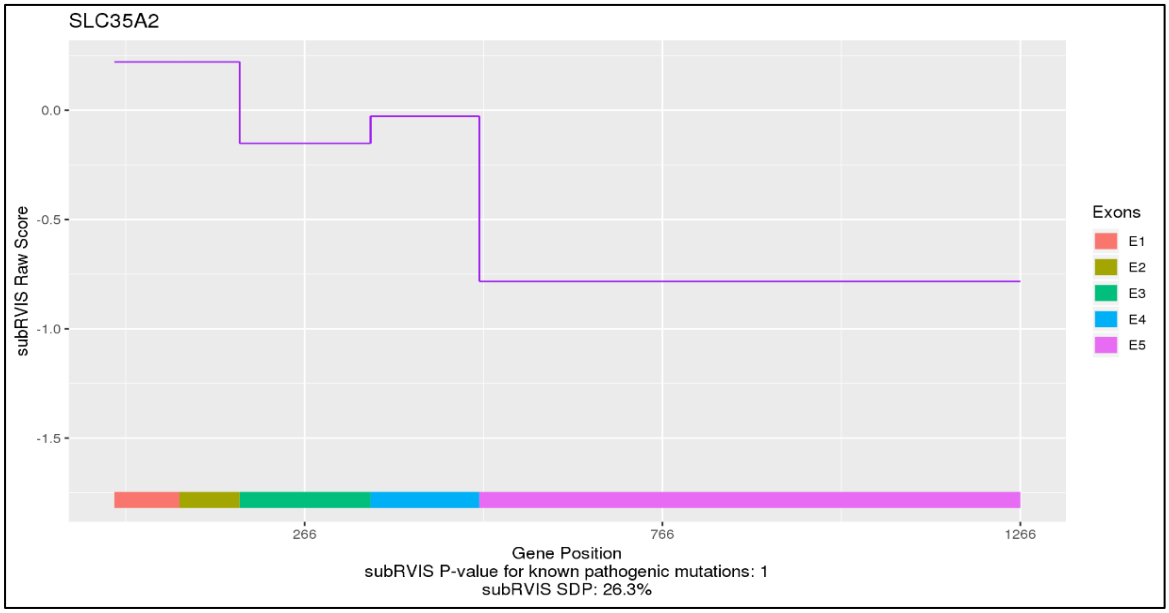
8.2 Supplementary Figures



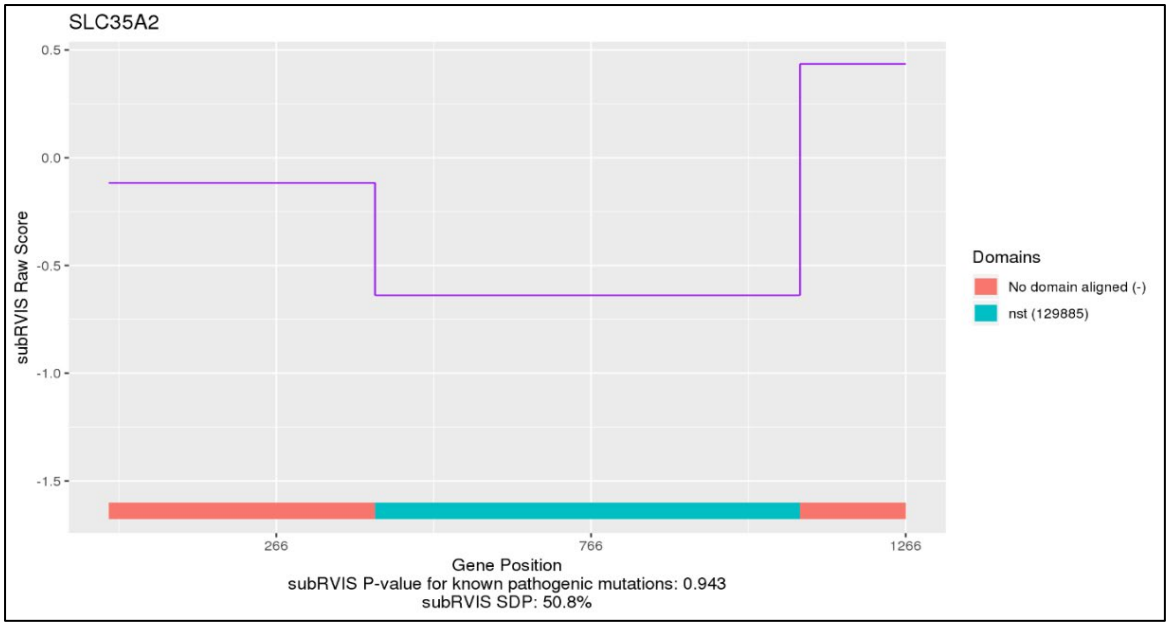
Supplementary Figure 8.1: Seqr captures of VUS from the IESS cohort.
A) Mosaic GNAO1 variant in individual AA0192-01. B) Mosaic SCN2A variant in AA0193-01. C) Paternally inherited SCN8A variant in AA2569-01. D) Mosaic FOXG1 variant in AA2569-01.



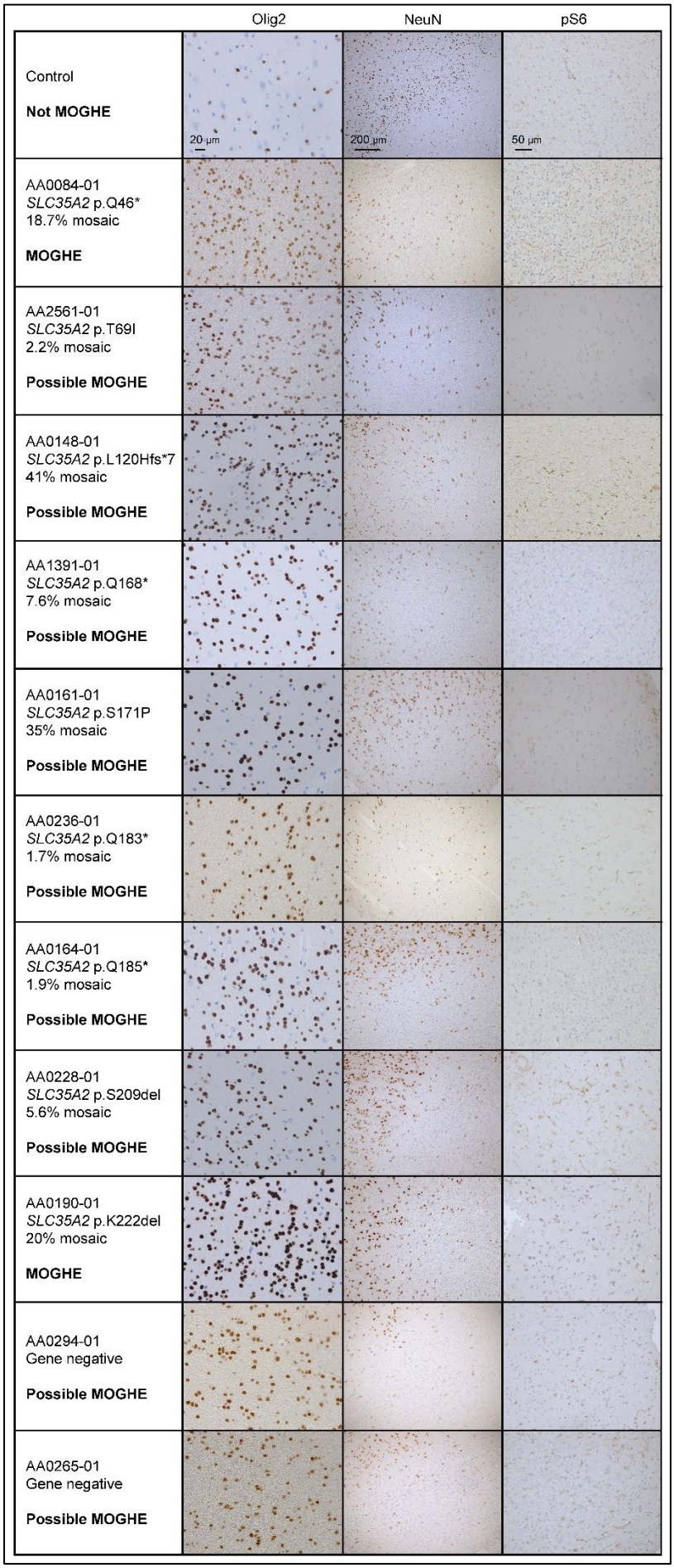
Supplementary Figure 8.2: Missense tolerance ratio (MTR) plot of *SLC35A2*.



Supplementary Figure 8.3: Exon-based Residual Variant Intolerance Score (RVIS) plot of *SLC35A2*.

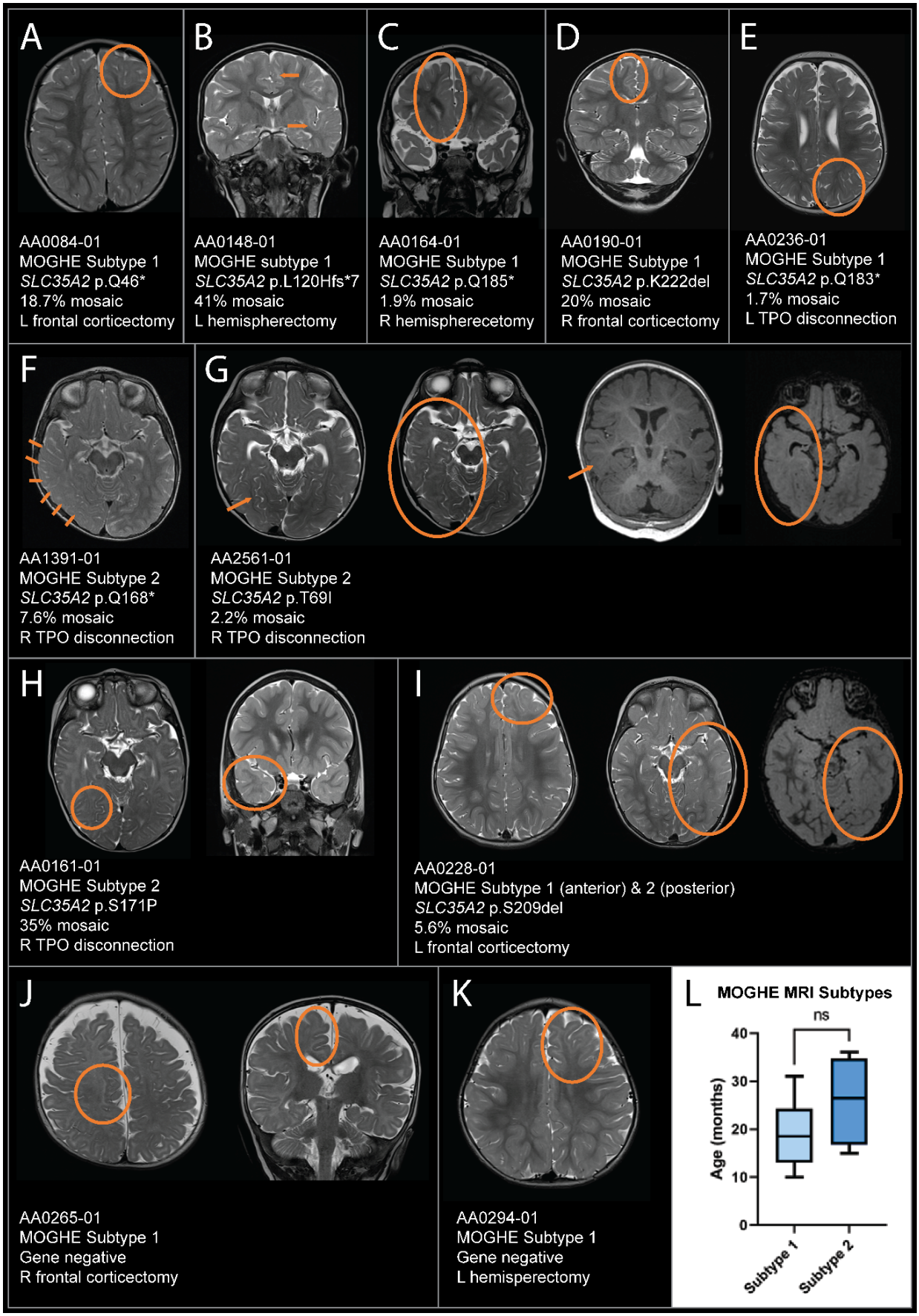


Supplementary Figure 8.4: Domain-based Residual Variant Intolerance Score (RVIS) plot of *SLC35A2*.



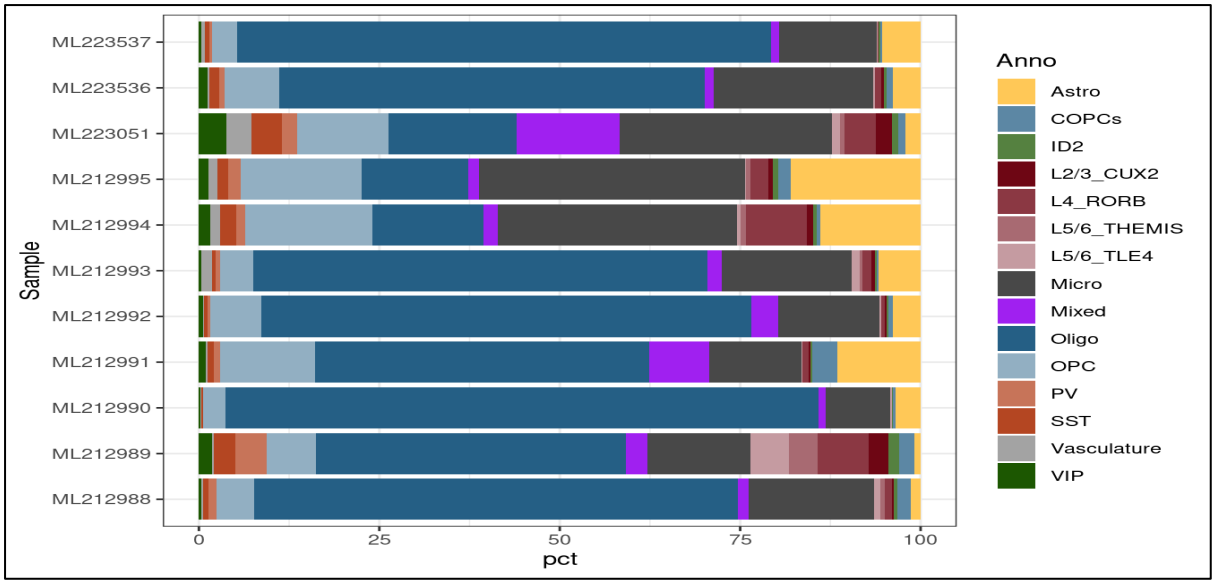
Olig2, NeuN and pS6 IHC staining for Control and all 11 MOGHE and possible MOGHE cases. Black bar denotes scale. Staining was performed on 3µm FFPE tissue sections.

Supplementary Figure 8.5: Olig2, NeuN and pS6 stains from full MOGHE cohort.

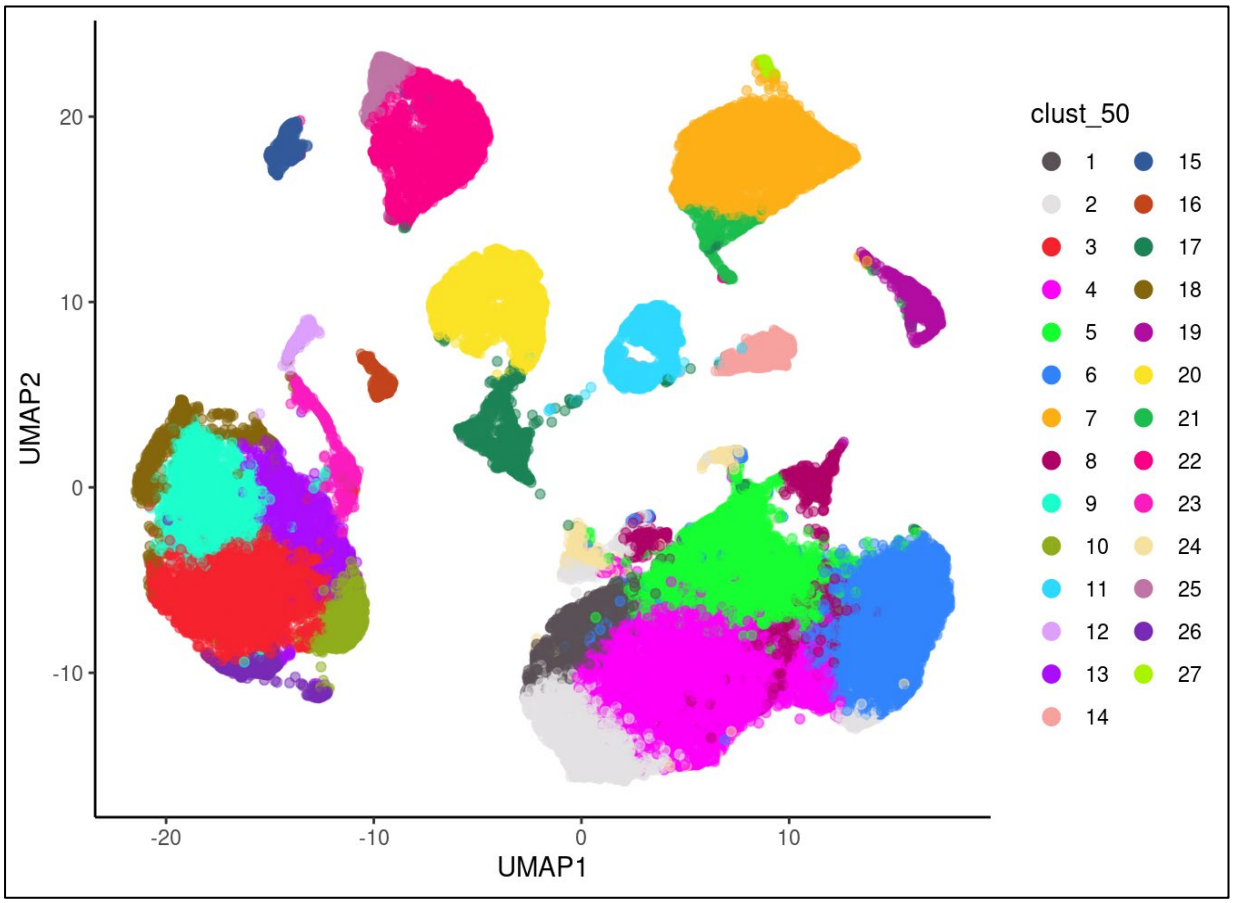


Supplementary Figure 8.6: MRI scans from full MOGHE cohort.

A-K) MRI scans for the 11 MOGHE and possible MOGHE cases. L) Box plot of ages at time of MRI.



Supplementary Figure 8.7: Proportional abundance of cell types in snRNA-seq data separated into individual samples.



Supplementary Figure 8.8: UMAP depicting unbiased clustering of snRNA-seq data.

8.3 Papers published during candidature

1. Kaspi A, Hildebrand MS, Jackson VE, Braden R, Van Reyk O, Howell T, Debono S, Lauretta M, Morison L, **Coleman MJ**, Webster R, Coman D, Goel H, Wallis M, Dabscheck G, Downie L, Baker EK, Parry-Fielder B, Ballard K, Harrold E, Ziegenfusz S, Bennett MF, Robertson E, Wang L, Boys A, Fisher SE, Amor DJ, Scheffer IE, Bahlo M, Morgan AT. 'Genetic aetiologies for childhood speech disorder: novel pathways co-expressed during brain development', *Mol. Psych*., 2022 Sept 18, doi: 10.1038/s41380-022-01764-8

2. Feenstra B, Skotte L, Fadista J, Grauholm G, Appadurai V, **Coleman M**, Damiano JA, Burgess R, Schneider A, Hansen TF, Banasik K, Liu X, Pasternak B, Pers TH, Dreier JW, Christensen J, Hougaard DM, Buil A, Geller F, Hviid A, Ullum H, Mullen SA, Hildebrand MS, Scheffer IE, Berkovic SF, Werge T, Melbye M. 'Genome-wide association study of febrile seizures implicates fever response and neuronal excitability genes', *Brain*, 2022 Jan 12, doi: 10.1093/brain/awab260

3. de Valles-Ibáñez G, Hildebrand M, Bahlo M, King C, **Coleman M**, Green T, Goldsmith J, Davis S, Gill D, Mandelstam S, Scheffer IE, Sadleir L. 'Infantile-onset Myoclonic Developmental and Epileptic Encephalopathy: a new RARS2 phenotype', *Epilepsia Open*, 2021 October 30, doi: 10.1002/epi4.12553

4. Heron SE, Regan BM, Harris RV, Gardner AE, **Coleman MJ**, Bennett MF, Grinton BE, Helbig KL, Sperling MR, Haut S, Geller EB, Widdess-Walsh P, Pelekanos JT, Bahlo M, Petrovski S, Heinzen EL, Hildebrand MS, Corbett MA, Scheffer IE, Gécz J, Berkovic, SF. 'Association of *SLC32A1* missense variants with genetic epilepsy with febrile seizures plus', *Neurology*, 2021 Mar 23, doi: 10.1212/WNL.0000000000011855

5. Myers KA, Bennett MF, Hildebrand MS, **Coleman MJ**, Zhou G, Hollingsworth G, Cairns A, Riney K, Berkovic SF, Bahlo M, Scheffer IE, 'Transcriptome Analysis in Ring Chromosome 20 Reveals Broad Differential Gene Expression', *Epilepsia*, 2020 Nov 18, doi: 10.1111/epi.16766

6. Palencia-Campos A, Aoto PC, Machal EMF, Rivera-Barahona A, Soto-Bielicka P, Bertinetti D, Baker B, Vu L, Piceci-Sparascio F, Torrente I, Boudin E, Peeters S, Van Hul W, Huber C, Bonneau D, Hildebrand MS, **Coleman M**, Bahlo M, Bennett MF, Schneider AL, Scheffer IE, Kibæk M, Kristiansen BS, Issa MY, Mehrez MI, Ismail S, Tenorio J, Li G, Skålhegg BS, Otaify GA, Temtamy S, Aglan M, Jønch AE, De Luca A, Mortier G, Cormier-Daire V, Ziegler A, Wallis M, Lapunzina P, Herberg FW, Taylor ST, Ruiz-Perez VL, 'Germline and Mosaic Variants in PRKACA and PRKACB Cause a Multiple Congenital Malformation Syndrome', *Am J Hum Genet*, 2020 Oct 14. doi: 10.1016/j.ajhg.2020.09.005

7. Hildebrand M, Jackson V, Scerri T, Van Reyk O, **Coleman M**, Braden R, Turner S, Rigbye K, Boys A, Barton S, Webster R, Fahey M, Saunders K, Parry-Fielder B, Paxton G, Hayman M, Coman D, Goel H, Baxter A, Ma A, Davis N, Reilly S, Delatycki M, Liégeois F, Connelly A, Gecz J, Fisher S, Amor D, Scheffer I, Bahlo M, Morgan A, 'Severe childhood speech disorder: gene discovery highlights transcriptional dysregulation', *Neurology*, 2020 May 19, doi: 10.1212/WNL.0000000000009441

8. Bennett MF, Oliver KL, Regan BM, Sikta N, Rafehi H, Bellows ST, Schneider AL, Crompton DE, Dolzhenko E, Eberle MA, **Coleman M**, Hildebrand MS, Corbett MA, Kroes T, Gecz J, Scheffer IE, Berkovic SF, Bahlo M, 'Familial adult myoclonic epilepsy type 1 SAMD12 TTTCA repeat expansion arose 17,000 years ago and is present in Sri Lankan and Indian families', *Eur J Hum Gen*, 2020 Mar 16. doi: 10.1038/s41431-020-0606-z

9. Sadleir L, de Valles-Ibanez G, King C, **Coleman M**, Mossman S, Paterson S, Nguyen J, Berkovic S, Mullen S, Bahlo M, Hildebrand M, Mefford H, Scheffer I, 'Inherited *RORB* pathogenic variants: overlap of photosensitive genetic generalized and occipital lobe epilepsy', *Epilepsia*, 2020 Mar

12. doi: 10.1111/epi.16475

https://scholar.google.com/citations?user=cfJRpSwAAAAJ&hl=en

8.4 List of conference presentations during candidature

8.4.1 Oral Presentations:

1. **Matthew Coleman,** *SLC35A2* and brain mosaicism in infantile spasms with focal brain malformations, Workshop on the Neurobiology of Epilepsy, Talloires, France, 5 July 2022.
2. **Matthew Coleman,** Somatic mosaicism in infantile spasms with brain malformations, European Society of Human Genetics Conference, Vienna, Austria, 13 June 2022.
3. **Matthew Coleman,** Piecing together a mosaic of the brain, Australian LGBTQIA+ STEMM Day Symposium, 18 November 2021.
4. **Matthew Coleman,** Katherine B Howell. Brain mosaicism in infantile spasms with brain malformations surgical cohort, Translational Neurogenetics 2020, Melbourne, Australia, 20 September 2021.
5. **Matthew Coleman,** The genetic basis of malformative infantile spasms, Melbourne Children's Campus Research Student Symposium, Melbourne, Australia, 30 September 2021.
6. **Matthew Coleman,** Genetic Landscape of Infantile Spasms, 34th International Epilepsy Congress, Paris, France, 30 August 2021.
7. **Matthew Coleman,** Brain mosaicism in infantile spasms with brain malformations: a surgical cohort, Translational Neurogenetics 2021, Melbourne, Australia.
8. **Matthew Coleman,** Brain somatic mosaicism and severe epilepsies of infancy, Epilepsy Melbourne 2020, Melbourne, Australia.
9. **Matthew Coleman,** Genetic diagnoses in a cohort of individuals with cortical malformations, MDHS Graduate Research Conference 2020, Melbourne, Australia.
10. **Matthew Coleman,** Genetic diagnoses in a cohort of paediatric brain malformations, Students of Brain Research (SOBR) Symposium 2020, Melbourne, Australia.
11. **Matthew Coleman,** Genetic diagnostic outcomes in a cohort of individuals with cortical malformations, Australian Society of Medical Research Victorian Symposium 2020, Melbourne, Australia.
12. **Matthew Coleman,** Genetic diagnoses in a cohort of individuals with brain malformations, Melbourne Children's Campus Research Student Symposium 2020, Melbourne, Australia.

8.4.2 Invited Speaker Talks:

1. **Matthew Coleman,** Mosaicism and malformations of the brain, Students of Brain Research (SOBR) Symposium 2022, Melbourne – 11/11/2022
2. **Matthew Coleman,** Diversity and Inclusion in Research, Melbourne Medical School Research Symposium 2022, Melbourne – 03/11/2022
3. **Matthew Coleman,** The genetic landscape of infantile spasms with malformations of cortical development, Institut du Cerveau (ICM), Paris, France – 12/07/2022
4. **Matthew Coleman,** Somatic mosaicism in infantile epilepsies with brain malformations, Institut für Humangenetik, University Hospital, Essen, Germany – 15/06/2022
5. **Matthew Coleman,** Scholarly Community in the Time of COVID, University of Melbourne Researcher Development Unit Seminar, Melbourne, Australia – 19/10/2021
6. **Matthew Coleman,** Diversity, Inclusion and Allyship in STEMM, Australian Mathematics and Sciences Institute (AMSI) BioInfoSummer Conference, Melbourne, Australia – 03/12/2020
7. **Matthew Coleman,** Meaningful LGBTQIA+ Diversity & Inclusion in STEMM, ASMR National Scientific Conference 2020, virtual – 18/10/2020.

8.4.3 Poster Presentations:

1. **Coleman M**, Wang M, Stephenson SEM, Simons C, Lee WS, D'Arcy C, Freytag S, Maixner WJ, Harvey AS, Leventer RJ, Howell KB, Lockhart PJ. *SLC35A2* and Brain Mosaicism in Infantile Spasms with Focal Brain Malformations, Lorne Genome Conference 2022, Lorne, Australia
2. **Coleman M**, Hildebrand M, Jackson V, Scerri T, Van Reyk O, Braden R, Turner S, Rigbye K, Boys A, Barton S, Webster R, Fahey M, Saunders K, Parry-Fielder B, Paxton G, Hayman M, Coman D, Goel H, Baxter A, Ma A, Davis N, Reilly S, Delatycki M, Liégeois F, Connelly A, Gecz J, Fisher S, Amor D, Scheffer I, Bahlo M, Morgan A. Transcriptional dysregulation underlies severe childhood speech disorder, Lorne Genome Conference 2020, Melbourne, Australia.
3. **Coleman M**, Hildebrand M, Jackson V, Scerri T, Van Reyk O, Braden R, Turner S, Rigbye K, Boys A, Barton S, Webster R, Fahey M, Saunders K, Parry-Fielder B, Paxton G, Hayman M, Coman D, Goel H, Baxter A, Ma A, Davis N, Reilly S, Delatycki M, Liégeois F, Connelly A, Gecz J, Fisher S, Amor D, Scheffer I, Bahlo M, Morgan A. Transcriptional dysregulation underlies severe childhood speech disorder, First Australian LGBT STEM Day Symposium 2019, Melbourne, Australia.

8.5 Manuscript currently under review at Neuropathology and Applied Neurobiology

See next page.

Title: Ectopic HCN4 provides a target biomarker for the genetic spectrum of mTORopathies

Authors: Matthew Coleman[1,2], Paulo Pinares-Garcia[3], Sarah E. M. Stephenson[1,2], Wei Shern Lee[1], Daniz Kooshavar[1,2], Catriona A. McLean[4], Katherine B. Howell[1,2,5], Richard J. Leventer[1,2,5], Christopher A. Reid[3,6], Paul J. Lockhart[*1,2]

Affiliations:

1 Murdoch Children's Research Institute; Parkville, Victoria, Australia

2 Department of Paediatrics, University of Melbourne; Parkville, Victoria, Australia

3 The Florey Institute of Neuroscience and Mental Health; Parkville, Victoria, Australia

4 Alfred Hospital, Prahran, Victoria, Australia

5 Department of Neurology, The Royal Children's Hospital, Parkville, Victoria, Australia

6 Epilepsy Research Centre, Department of Medicine, University of Melbourne, Austin Health, Heidelberg, 3084, VIC, Australia.

*Corresponding author. Email: paul.lockhart@mcri.edu.au

Keywords:

Epilepsy, tuberous sclerosis complex, focal cortical dysplasia, neuropathology, MTOR, HCN4.

Word count: 2692

Figures: Two

Tables: One

Key points:

- Elevated hyperpolarization-activated cyclic nucleotide-gated potassium channel isoform 4 (HCN4) has been identified as a key driver of epileptogenic activity in dysmorphic neurons in an mTOR-dependent epilepsy mouse model.
- Elevated levels of HCN4 were observed to be highly restricted to dysmorphic neurons and balloon cells in brain tissue from individuals with tuberous sclerosis complex and focal cortical dysplasia type IIB.
- Aberrant HCN4 profile was observed in the full range of genetic mTORopathies tested (eight genes), regardless of variant allele frequency.
- HCN4 provides an alternate biomarker for the genetic spectrum of focal cortical dysplasia type IIB and tuberous sclerosis complex in humans.

Abstract:

Aims

To validate elevated hyperpolarization-activated cyclic nucleotide-gated potassium channel isoform 4 (HCN4) as a key marker in brain malformations across the genetic spectrum of mTORopathies in humans.

Methods

Mutations in PI3K-AKT-mTOR pathway and GATOR1 complex genes resulting in hyperactivation of mechanistic target of rapamycin (mTOR) complex 1 are a major cause of drug-resistant epilepsy and focal cortical malformations (FCM). Resective neurosurgery is often required to achieve seizure control in patients with mTORopathies, due to lack of effectiveness of non-surgical therapies, including antiseizure medication and mTOR inhibitors. Our study

investigated the relative steady state levels and cellular localization of HCN4 in resected human brain tissue from 18 individuals with mTORopathies (three tuberous sclerosis complex (TSC) due to *TSC2* variants, five individuals with focal cortical dysplasia type IIA (FCD IIA) due to variants in *MTOR*, *AKT3* and *PIK3CA*, and ten individuals with FCD IIB with pathogenic genetic variants in *TSC1*, *MTOR, RHEB*, *DEPDC5* or *NPRL3*).

Results

Elevated HCN4 was observed to be highly restricted to abnormal cell types (dysmorphic neurons and balloon cells) in brain tissue from all individuals. Elevated HCN4 was not observed in controls or individuals with non-mTOR related focal epilepsy due to a pathogenic variants in *ATP1A3*, *SLC35A2* or *FGFR1*.

Discussion

HCN4 provides a biomarker for the genetic spectrum of mTORopathies and may present a potential therapeutic target for seizure control in mTOR-related epilepsy.

List of abbreviations

BC = balloon cell; BOSD = bottom-of-sulcus dysplasia; cAMP = cyclic adenosine monophosphate; CAPOS = ; ddPCR = droplet digital polymerase chain reaction; DEE = developmental and epileptic encephalopathy; DN = dysmorphic neuron; ECoG = electrocorticogram; FCD = focal cortical dysplasia; FCD II = focal cortical dysplasia type II; FCM = focal cortical malformations; FFPE = formalin-fixed paraffin-embedded; HME = hemimegalencephaly; IHC = immunohistochemistry; HCN4 = hyperpolarization-activated cyclic nucleotide-gated potassium channel isoform 4; MOGHE = mild malformation of cortical development with oligodendroglial hyploplasia in epilepsy; MRI = magnetic resonance imaging;

mTOR, mechanistic target or rapamycin; TSC = tuberous sclerosis complex; VAF = variant allele frequency.

Main Text:

INTRODUCTION

Mutations in the PI3K-AKT-mTOR pathway and GATOR1 complex genes resulting in hyperactivation of mechanistic target of rapamycin (mTOR) complex 1 are a well-established cause of focal cortical malformations (FCM) causing drug resistant epilepsy. Collectively, these disorders are known as mTORopathies and include tuberous sclerosis complex (TSC), hemimegalencephaly (HME) and focal cortical dysplasia (FCD) type II (FCD II) [1].

FCD is the most common form of FCM. Individuals with FCD typically present with seizures and may have secondary developmental impairments and intellectual disability if drug resistant epilepsy presents in early childhood. FCD II are FCM that are characterized by cortical dyslamination and the presence of dysmorphic neurons (DNs), either without (FCD IIA) or with (FCD IIB) balloon cells (BCs) [2-4]. TSC is a multi-system disorder, one feature of which is epileptogenic brain lesions known as cortical tubers, which are glioneuronal hamartomas with similar histopathologic features to FCD IIB [5]. The incidence of TSC is estimated to be 1:10,000 live births but has been observed at a higher frequency in the European population [6]. Together, TSC and FCD II are the most common disorders requiring surgical intervention to control seizures and, as such, represent a substantial medical burden, particularly in childhood [7].

Germline or somatic variants in mTOR pathway genes, including in *MTOR*, *TSC1*, *TSC2*, *RHEB*, *DEPDC5, AKT3, PIK3CA* and *NPRL3*, are well-established causes of FCD IIB [8, 9]. Somatic variants in *TSC1* and *TSC2* can cause TSC but are far less common than germline *TSC1* and *TSC2* variants. In FCM with brain somatic mosaicism, a gradient of higher variant allele frequency (VAF) has been observed in abnormal cells such as DNs and BCs in FCD IIB and in the central core of tubers in TSC, compared to adjacent non-dysplastic tissue [10, 11].

Despite the genetic basis of mTORopathies being relatively well established, the pathophysiologic and epileptogenic mechanisms are still not fully understood. Epilepsy neurosurgery, which is both invasive and expensive, is often the only effective treatment to control seizures and prevent serious developmental consequences in these individuals [12, 13]. As such, there is a pressing need to identify novel therapeutic targets for seizure control.

Hyperpolarization-activated cyclic nucleotide-gated potassium channel isoform 4 (HCN4) is a member of the hyperpolarization-activated, cyclic nucleotide-gated cation (HCN) channel family. HCN-mediated currents regulate cardiac and neuronal pacemaker activity [14]. Hseih *et al.* recently identified upregulated HCN4 in *RHEB*-related FCM as a key driver of mTOR-dependent epilepsy in mice [15]. HCN4 was shown to be increased in FCM neurons and it was found that increasing cyclic adenosine monophosphate (cAMP) in these neurons resulted in repetitive firing, while disruption of HCN4 activity resulted in a rescue of the seizure phenotype in mice. The authors further demonstrated abnormal HCN4 steady-state levels in DNs in surgically resected human brain tissue from individuals with TSC and FCD IIB, without reporting the genetic basis in these individuals, and proposed HCN4 as a candidate therapeutic target. This study examined

the distribution and relative steady state levels of HCN4 in brain tissue derived from individuals with distinct genetic mTORopathies to determine if aberrant *HCN4* expression is a consistent feature across the genetic spectrum of FCM-related mTORopathies, mediated by either germline or brain-specific somatic variants.

MATERIALS AND METHODS

Study design and cohort

This study was approved by the Human Research Ethics Committee at the Royal Children's Hospital (ID 29077). Written informed consent was obtained from participants or their parents. Brain tissue samples used for immunohistochemical (IHC) and immunofluorescence analysis were retrieved from the Royal Children's Hospital Pathology Service, the tissue bank associated with the Neurogenetics Group at the Murdoch Children's Research Institute, and the Victorian Brain Bank, all in Melbourne, Australia.

Brain tissue had been collected during epilepsy surgery or post-mortem and stored as formalin-fixed paraffin-embedded (FFPE) tissue. Individuals with mTORopathies consisted of three individuals with TSC with variants in the *TSC2* gene (cases 1-3), ten individuals with FCD IIB (cases 4-13, including one individual with bottom-of-sulcus dysplasia (BOSD)) with germline variants in *DEPDC5* (n=1) and *NPRL3* (n=1) or brain somatic variants in *MTOR* (n=6, VAF = 2.1 – 5.7%), *RHEB* (n=1, VAF = 13.0%) or *TSC1* (n=1, VAF = 2.8%), and five individuals with FCD IIA and HME (cases 14-18) with brain somatic variants in *MTOR* (n=1, VAF = 7.40%), *AKT3* (n=2, VAF = 4.9 – 5.1%) or *PIK3CA* (n=2, VAF = 20 – 29%). Genetic variants had previously been identified using deep targeted sequencing on brain-derived DNA, and variant allele

frequencies were validated by droplet digital PCR (ddPCR) as previously described [11]. Healthy control brain tissue samples were collected from four adults obtained post-mortem from the Victorian Brain Bank (controls 1-4). Additionally, post-mortem brain tissue from one age-matched individual with epilepsy due to a heterozygous variant in *ATP1A3* (control 5), unrelated to the PI3K-AKT-mTOR pathway, and brain tissue from four individuals with mild malformation of cortical development with oligodendroglial hyperplasia in epilepsy (MOGHE) due to a somatic *SLC35A2* variants (control 6-9) and two individuals with dysembryoplastic neuroepithelial tumours (DNET) due to variants in *FGFR1*, collected during epilepsy surgery, were used as a non-mTORopathy epilepsy and FCM controls (Table 1).

Immunohistochemical analysis

Staining was completed according to established protocols. FFPE tissue slides were incubated overnight with anti-HCN4 (Alomone, #APC-052, dilution 1:500), anti-Phospho-S6 Ribosomal Protein (Ser235/236) (pS6) (Cell Signaling, #4858P, dilution 1:100) and anti-neurofilament marker (SMI-311) (BioLegend, #837801, dilution 1:1000) antibodies. Hematoxylin staining was completed to identify nuclear components in the tissues. Details of the antibodies utilized are provided in Supplementary Table 1.

HCN4 quantification and dysmorphic neuron identification

Quantification of HCN4 staining was performed using ImageJ analysis software. Regions of interest (ROIs) were manually drawn, capturing 15 cells from each slide, and normalised against five background regions of equal size to account for background signal intensity. Neuronal cells that met the criteria of >25 μm cell diameter with 15–28 μm nucleus diameter were identified as

cytomegalic and labelled as dysmorphic neurons or balloon cells [16]. Statistical analyses were performed using GraphPad Prism 9 (GraphPad Software La Jolla California, USA).

Immunofluorescence staining

Staining was completed according to established protocols. FFPE tissue slides were incubated overnight with anti-HCN4 (Alomone, #APC-052, dilution 1:500), anti-Phospho-S6 Ribosomal Protein (Ser235/236) (pS6) (Cell Signaling, #E2R10, dilution 1:500) and anti-neurofilament H (SMI-32) (Invitrogen, #PA110002, dilution 1:1000) antibodies. DAPI (Abacus, #VEH1500) staining was completed to identify nuclear components in the tissues. Details of the antibodies utilised is provided in Supplementary Table 2.

Droplet digital PCR mRNA quantification

Messenger RNA (mRNA) was available from 12/18 cases and 6/7 non-mTOR epilepsy controls (Table 1). mRNA was converted to cDNA and absolute quantification of mRNA was performed using droplet digital PCR (ddPCR) gene expression analysis with TaqMan gene expression probes specific for *HCN4* (Product code Hs00975492_m1). Reactions were run on a BioRad C1000 Touch Thermal Cycler (Bio-Rad) and droplets were analysed using the QX100 Droplet Reader (BioRad) and Quantasoft software.

RESULTS

Histopathology and genetic findings

To assess the steady-state level and distribution of HCN4 in dysplastic tissue derived from the genetic spectrum of mTORopathies, IHC analysis of human surgically resected brain tissue was

189 performed. The cohort consisted of specimens from 18 individuals with causal genetic variants in
190 mTOR pathway genes (Table 1). Brain tissue specimens from four deceased adult individuals
191 without FCM were used as healthy controls. Seven age-matched individuals with non-mTOR-
192 related epilepsy and brain malformation due to variants in *ATP1A3* (control 5) , *SLC35A2* (controls
193 6-9) or *FGFR1* (Controls 10-11), were used as unrelated epilepsy controls.
194
195 **Strong HCN4 immunoreactivity was observed in dysplastic tissues from individuals with**
196 **mTOR-related FCM**
197 To investigate if aberrant HCN4 in DNs and BCs was a common feature in the genetic spectrum
198 of mTORopathies, IHC was performed on FFPE tissues from individuals with genetically
199 confirmed mTORopathy. Representative results for two cases, one control and two non-mTOR
200 epilepsy/malformation controls are shown in Figure 1. Results for all cases and controls are
201 presented in Supplementary Figure 1. Immunostaining of TSC, FCD IIA and FCD IIB tissues with
202 an anti-HCN4 antibody (Alomone, #APC-052) demonstrated strong immunoreactivity, indicating
203 aberrant HCN4 levels, in individuals with TSC and FCD II (Figure 1J, 1M). In all TSC and FCD
204 II cases, immunoreactivity was restricted to enlarged neurons with irregular shape and increased
205 soma size, features consistent with DNs and BCs. In contrast, adjacent neuronal cells with normal
206 shape and soma size, showed only weak HCN4 immunoreactivity. Lesional tissue from individuals
207 with FCD IIA/B caused by somatic mTOR pathway variants showed the same pattern of strong
208 HCN4 immunoreactivity in abnormal neurons as seen in lesional tissue from individuals with
209 germline variants. Consistent with previous studies [14], HCN4 immunoreactivity was very weak
210 or absent in control post-mortem specimens (Figure 1A). Similarly, HCN4 and pS6
211 immunoreactivity very weak in brain tissue from the individuals with epilepsy due to non-mTOR

related causes (including *ATP1A3* (control 5) (Figure 1D), *SLC35A2* (controls 6-9) (Figure 1G, Supplementary Figure 1) and *FGFR1* (controls 10-11) (Supplementary Figure 1).

Quantitative analysis showed that FCD II and TSC cases had significantly elevated HCN4 in dysmorphic neurons and balloon cells, compared to epilepsies unrelated to the mTOR pathway (Figure 1P) ($p<0.0001$). Furthermore, HCN4 levels as measured by staining intensity did not exhibit any correlation with VAF in this cohort ($R^2=0.63$) (Figure 1Q)

To determine the identity of HCN4-positive cells, IHC was performed using an antibody directed against Phospho-S6 Ribosomal Protein (Ser235/236) (pS6) (Cell Signaling, #4858P). This protein is a key marker of mTOR kinase activity and can be utilized to identify the abnormal cell types associated with TSC and FCD II (DNs and BCs). Immunostaining of brain sections from affected individuals showed strong immunoreactivity for pS6 in abnormal and enlarged cells with irregular shape, indicating an abundance of DNs and BCs in these tissues (Figure 1K, 1N). The density of pS6-positive cells varied between individuals, but all TSC and FCD II tissues exhibited DNs, and BCs were observed in most individuals. Limited and weak immunoreactivity against pS6 was observed in control post-mortem specimens (Figure 1B) and from the non-mTOR epilepsy control individuals with PMG, MOGHE or DNET (Figure 1E, 1H, Supplementary Figure 1). In addition, IHC was performed using an antibody directed against pan-neuronal neurofilaments (SMI-311) (BioLegend, #837801). SMI-311 enables visualization of neurofilaments that accumulate in the soma and dendrites of dysmorphic nerve cells and is utilized as a specific marker of DNs in FCM tissue. Similar to the results observed for pS6 immunoreactivity, strong specific SMI-311 signal was only observed in tissue from individuals with FCD II and TSC and large, abnormal neuronal cells (Figure 1L, 1O). Strong SMI-311 immunoreactivity was absent from control tissues (Figure

1C). Collectively, the distribution and localization of HCN4 immunoreactivity was consistent with the abnormal cell types present in FCD II and TSC.

Immunofluorescence staining shows strong HCN4 signal in FCM neurons

To further demonstrate HCN4 was elevated primarily in abnormal cell types, colocalization of HCN4 and pS6 was determined by immunofluorescence analysis of specimens from cases 1 (TSC) and 4 (FCD IIB) and control 5 (Figure 2). Immunofluorescence co-staining of HCN4 (Alomone, #APC-052) in red, pS6 (Cell Signaling, #E2R10) in purple and SMI-32 (Invitrogen, #PA110002) showed clear overlap in the cellular localization of HCN4 and pS6 in both DNs in case 1 and 4 (Figure 2). Presence of HCN4 and pS6 with absence of SMI-32 identified BCs. Strong co-expression of HCN4 and pS6 was not observed in control non-mTOR-related epilepsy tissues (Figure 2A) or adjacent normal cortical neurons in the FCM specimens.

***HCN4* mRNA expression is not indicative of HCN4 levels**

Expression of HCN4 is reported to be low in brain tissue and our analysis of mRNA expression in total RNA derived from bulk brain tissue revealed a similar outcome. Very low droplet counts were observed for all samples that had sufficient material for RNA isolation (Supplementary Figure 2), although cDNA quality was high, as evidenced by robust droplet counts for control genes (data not shown). The low droplet counts and therefore low quality ddPCR results meant we were unable to correlate HCN4 expression from this analysis to the HCN4 levels observed in individuals with TSC and FCD II.

DISCUSSION

We report histopathological data and elevated HCN4 steady state levels in FCM tissue from a cohort of individuals with a spectrum of genetically distinct mTORopathies with varying mutant allele frequencies in brain tissue. The strong co-expression of pS6 and HCN4 observed using immunofluorescence staining in brain tissue specimens from TSC and FCD IIB individuals, confirms that aberrant HCN4 expression is highly localized to cells demonstrating abnormal mTOR pathway activation.

These findings are consistent with initial data published by Hsieh *et al.* reporting in *RHEB* and *TSC1*-related FCM [15]. We build on their work to show that increased abundance of HCN4 in DNs and BCs is a common feature across the genetic spectrum of mTOR-pathway-related FCM, including FCM due to pathogenic variants in *MTOR*, *TSC1*, *TSC2*, *RHEB*, *DEPDC5, AKT3*, *PIK3CA* and *NPRL3*. Furthermore, elevated HCN4 was not observed in brain tissue from healthy controls, nor from the individuals with *ATP1A3*-related epilepsy (control 5, supplementary figure 1) nor with MOGHE (control 6-9), nor DNET (controls 10-11). *ATP1A3* encodes the α3 subunit of Na+/K+-ATPase and has been linked to alternating hemiplegia, cerebellar ataxia, areflexia, pes cavus, optic atrophy, and sensorineural hearing loss (CAPOS) syndrome, rapid-onset dystonia-parkinsonism and developmental and epileptic encephalopathy (DEE) [17, 18]. *ATP1A3* has no known associations with the PI3K-AKT3-mTOR pathway and is not reported to have an impact on pS6 levels [18]. Similarly, individuals with MOGHE typically present with severe drug resistant epilepsy within the first few months of life, and their FCMs may be amenable to surgery. However, they are understood to be clinically and genetically distinct from mTORopathies, with somatic *SLC35A2* variants being identified in the majority of cases (20). FGFR1 plays a role in the MAPK-ERK pathway [19]. Notably, DNs and BCs are not a feature of MOGHE, PMG or DNET (21).

Absence of dysregulated HCN4 in FCM tissue from these controls suggests that HCN4 upregulation is likely unique to FCM tissues associated with mTORopathies.

The strong co-localization of pS6 and HCN4 observed via immunofluorescence staining confirms that elevated HCN4 is highly restricted to DNs and BCs, providing further evidence that HCN4 represents a potential target for seizure control in TSC and FCD IIB with mTOR-pathway variants more broadly.

Brain somatic mosaicism is common in mTORopathies, particularly in FCD IIB. In this study, we observed similar levels of aberrant HCN4 in FCM neurons in specimens from individuals with germline or somatic variants, even at levels of mosaicism as low as 2.1%, establishing that this aberrant activity is not exclusive to germline mTORopathies. Previously, HCN4 immunoreactivity and channel activity was observed in FCM neurons in a cellular mosaic $Rheb^{CA}$ mouse model of FCD II, achieved by *in utero* electroporation [15, 20]. In this study we assessed human tissue with variant allele frequencies ranging from 2.1% to 13.0% to demonstrate the steady distribution and expression of HCN4 in DNs and BCs in human mosaic cases. This is relevant due to the high incidence of low-level brain-specific somatic variants in the broader FCD IIB and TSC patient population and suggests that this mechanism is not exclusive to germline mTORopathies.

The exact mechanism and association between HCN4 increase and mTOR mutation is not yet understood. Is it possible that mTOR kinase hyperactivation may influence HCN channel activity, transcriptional regulation or steady-state protein level of HCN4 in abnormal cell types within brain tissue, resulting in aberrant HCN4 levels. It is not yet known whether a similar increase in HCN4

abundance is sustained at the mRNA expression level. However, *HCN4* is lowly expressed in brain tissue and mRNA are often not reflective of protein levels. This may limit the utility of *HCN4* mRNA expression as a biomarker in RNA analyses. Further studies at the single cell level of resolution are required to fully address this issue.

Two limitations of this work are: 1) this study uses a relatively restricted set of control tissues with four adult specimens and seven paediatric epilepsy specimens. Closer age-matched controls would eliminate any effect of age on HCN4 expression but are difficult to source. However, the absence of HCN4 staining in adjacent 'non-dysplastic' cells in the individuals with mTORopathies provides some confidence that HCN4 staining in the dysplastic tissue differs from age-expected levels, with the non-dysplastic cells acting as an internal control; and 2) this study has only demonstrated dysregulated protein abundance and has not measured HCN-mediated channel activity in human tissue. However, given results observed in the mouse model of TSC- and FCDII-associated FCM, we suggest that HCN4 channel activity in dysplastic tissue may be a major contributor to seizure activity in FCD II. This study has used a HCN4 antibody that has previously been validated in a conditional mouse knockout model [14] but has not examined other HCN channel proteins.

Our results are concordant with the initial findings of Hsieh *et al.*, [15] and collectively suggest that dysregulated and overactive HCN4 is a biomarker of germline and somatic mTORopathies. Recent studies have identified that abnormal cell types appear to be most abundant in the centre of tubers and bottom of sulcus dysplasias, corresponding to the sites of maximal magnetic resonance imaging (MRI) and electrocorticogram (ECoG) abnormalities [21, 22]. Moreover, DNs are

reported to be the major epileptogenic cell type in FCD and TSC [23], and somatic variants causing FCD may be restricted to DNs and BCs [11]. It has previously been shown that surgical resections limited to the tuber centre, targeting a nidus of dysmorphic and intrinsically epileptogenic neurons, are sufficient for seizure freedom in TSC [21]. As such, establishment of an effective, targeted non-surgical therapy for epilepsy in mTORopathies could bridge a major treatment gap, particularly in individuals with multifocal seizures or dysplasia in eloquent cortex, who are poor candidates for resective surgery.

CONCLUSION

Therapeutic targeting of aberrant HCN4 activity may represent a potential new strategy for seizure control in FCM, particularly for drug resistant epilepsies and in cases where individuals have multifocal seizures or dysplasia in eloquent cortex, meaning they are poor candidates for resective surgery. Ectopic HCN4 in DNs is thought to underlie hyperexcitability through a cAMP-mediated activation of the channel causing depolarization and aberrant firing [15] The expression of non-functioning HCN4 channels in DNs is sufficient to stop seizures in the $Rheb^{CA}$ mouse model of FCD II [15]. This suggests that HCN4 channel activation is central to excitability leading to seizures and that 'blocking' these channels may be an effective therapeutic strategy. The use of gene technologies, including anti-sense oligonucleotides, targeted to reducing *HCN4* expression may be worth considering. Importantly, brain specific HCN4 channel knockout is well tolerated in mice arguing that this approach may be feasible in humans [24]. Alternatively, a small molecule approach aimed at developing brain penetrant HCN4 selective blockers could be an alternative targeting approach.

In summary, an elevated abundance of HCN4 in FCM neurons in this genetically distinct spectrum of mTORopathies suggests that it provides a biomarker for mTORopathies and should be investigated as a potential therapeutic target for seizure control in TSC, FCD IIB and other mTORopathies.

References*:*

1. Marsan, E. and S. Baulac, *Review: Mechanistic target of rapamycin (mTOR) pathway, focal cortical dysplasia and epilepsy.* Neuropathol Appl Neurobiol, 2018. **44**(1): p. 6-17.
2. Blümcke, I. and A. Mühlebner, *Neuropathological work-up of focal cortical dysplasias using the new ILAE consensus classification system - practical guideline article invited by the Euro-CNS Research Committee.* Clinical neuropathology, 2011. **30**(4): p. 164-177.
3. Najm, I.M., H.B. Sarnat, and I. Blümcke, *Review: The international consensus classification of Focal Cortical Dysplasia - a critical update 2018.* Neuropathology and applied neurobiology, 2018. **44**(1): p. 18-31.
4. Bast, T., et al., *Focal cortical dysplasia: prevalence, clinical presentation and epilepsy in children and adults.* Acta Neurologica Scandinavica, 2006. **113**(2): p. 72-81.
5. Caban, C., et al., *Genetics of tuberous sclerosis complex: implications for clinical practice.* The Application of Clinical Genetics, 2020.
6. C, J., et al., *Learning disability and epilepsy in an epidemiological sample of individuals with tuberous sclerosis complex.* Psychological Medicine, 2003. **33**(2): p. 335-344.
7. Baldassari, S., et al., *Dissecting the genetic basis of focal cortical dysplasia: a large cohort study.* Acta Neuropathologica, 2019. **138**(6): p. 885-900.
8. Sim, N.S., et al., *Precise detection of low-level somatic mutation in resected epilepsy brain tissue.* Acta neuropathologica, 2019. **138**(6): p. 901-912.
9. Ye, Z., et al., *Somatic mutation: The hidden genetics of brain malformations and focal epilepsies.* Epilepsy Research, 2019. **155**.
10. Lee, W.S., et al. *Gradient of brain mosaic RHEB variants causes a continuum of cortical dysplasia*. 2021. John Wiley & Sons Ltd.
11. Lee, W.S., et al., *Second-hit DEPDC5 mutation is limited to dysmorphic neurons in cortical dysplasia type IIA.* ANNALS OF CLINICAL AND TRANSLATIONAL NEUROLOGY, 2019. **6**(7): p. 1338-1344.
12. Dwivedi, R., et al., *Surgery for Drug-Resistant Epilepsy in Children.* New England Journal of Medicine, 2017. **377**(17): p. 1639-1647.
13. Catchpool, M., et al., *Cost-effectiveness of epileptic surgery compared with medical treatment in children with drug-resistant epilepsy.* Epilepsy Behav, 2019. **97**: p. 253-259.
14. Zobeiri, M., et al., *The Hyperpolarization-Activated HCN4 Channel is Important for Proper Maintenance of Oscillatory Activity in the Thalamocortical System.* Cerebral Cortex, 2019. **29**(5): p. 2291-2304.
15. Hsieh, L.S., et al., *Ectopic HCN4 expression drives mTOR-dependent epilepsy in mice.* SCIENCE TRANSLATIONAL MEDICINE, 2020. **12**(570): p. eabc1492.
16. Najm, I., et al., *The ILAE consensus classification of focal cortical dysplasia: An update proposed by an ad hoc task force of the ILAE diagnostic methods commission.* Epilepsia, 2022. **63**(8): p. 1899-1919.

17. Miyatake, S., et al., *De novo ATP1A3 variants cause polymicrogyria.* Science Advances. **7**(13): p. eabd2368.
18. Arystarkhova, E., et al., *Misfolding, altered membrane distributions, and the unfolded protein response contribute to pathogenicity differences in Na,K-ATPase ATP1A3 mutations.* Journal of Biological Chemistry, 2021. **296**.
19. Rivera, B., et al., *Germline and somatic FGFR1 abnormalities in dysembryoplastic neuroepithelial tumors.* Acta Neuropathologica, 2016. **131**(6): p. 847-863.
20. Nguyen, L.H., T. Mahadeo, and A. Bordey, *mTOR Hyperactivity Levels Influence the Severity of Epilepsy and Associated Neuropathology in an Experimental Model of Tuberous Sclerosis Complex and Focal Cortical Dysplasia.* J Neurosci, 2019. **39**(14): p. 2762-2773.
21. Stephenson, S.E.M., et al., *Resection of tuber centers only for seizure control in tuberous sclerosis complex.* Epilepsy Research, 2021. **171**: p. 106572.
22. Lee, W.S., et al., *Genetic characterization identifies bottom-of-sulcus dysplasia as an mTORopathy.* Neurology, 2020. **95**(18): p. e2542.
23. Abdijadid, S., et al., *Basic Mechanisms of Epileptogenesis in Pediatric Cortical Dysplasia.* CNS Neuroscience & Therapeutics, 2015. **21**(2): p. 92-103.
24. Kharouf, Q., et al., *The hyperpolarization-activated cyclic nucleotide-gated 4 channel as a potential anti-seizure drug target.* Br J Pharmacol, 2020. **177**(16): p. 3712-3729.

DECLARACTIONS

Acknowledgments:

We thank the families involved in this research and acknowledge the assistance of Kate Pope and Greta Gillies in patient recruitment and sample collection.

Ethics approval and consent to participate:

This study was approved by the Human Research Ethics Committee at the Royal Children's Hospital (ID 29077). Written informed consent was obtained from participants or their parents. Brain tissue samples used for immunohistochemical (IHC) and immunofluorescence analysis were retrieved from the Royal Children's Hospital Pathology Service, the tissue bank associated with the Neurogenetics Group at the Murdoch Children's Research Institute, and the Victorian Brain Bank, all in Melbourne, Australia.

Consent for publication:

Not applicable.

Availability of data and materials:

All data are available in the main text or the supplementary materials. All original images are included in supplementary and all clinical data are described in the main text.

Competing interests:

CAR was principle academic lead on a project funded by BioCurate investigating HCN channels in epilepsy. This program has ceased. All other authors have no competing interests to declare.

Funding:

This work was supported in part by the Australian Government National Health and Medical Research Council grant GNT1161549 to PJL and RJL and an industry grant from BioCurate (PJL, CAR). Additional funding was provided by the Independent Research Institute Infrastructure Support Scheme, the Victorian State Government Operational Infrastructure Program and the Murdoch Children's Research Institute.

Author contributions:

Conceptualization: MC, PPG, CAR, PJL, RJL

Methodology: MC, PPG, SEMS, WSL

Investigation: MC, WSL, KBH

Funding acquisition: RJL, CAR, PJL

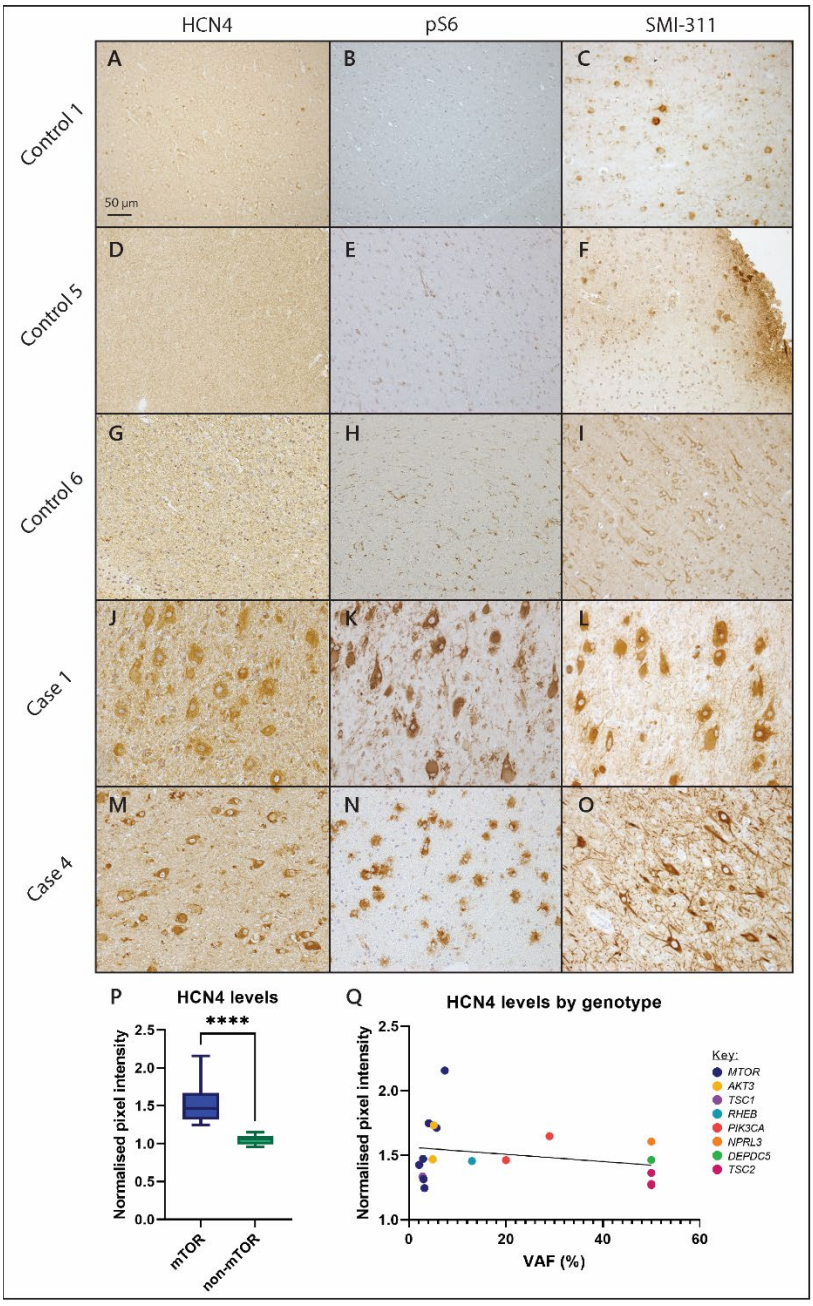
Project administration: KBH, CAR, PJL

Resources: CAM, DK

Supervision: SEMS, RJL, CAR, PJL

Writing – original draft: MC, PJL, KBH

Writing – review & editing: MC, PJL, CAR, KBH, RJL, SEMS



456

457 **Fig. 1. HCN4, SMI311 and pS6 IHC staining for control 1 (A-C), control 5 (D-F), control 6 (MOGHE)**
458 **(G-I), case 1 (TSC) (J-L) and case 4 (FCD IIB) (M-O) at 20x magnification. (P) Box plot of**
459 **HCN4 quantification in dysmorphic neurons and balloon cells in case cohort vs neurons**
460 **in control cohort and (Q) association of HCN4 staining intensity in DNs and BCs with**
461 **variant allele frequency (VAF) in case cohort.** Staining was performed on 3µm FFPE tissue
462 sections. Dysmorphic neurons and balloons cells are revealed by HCN4 and pS6 staining.
463 Increased HCN4 is observed in FCD IIB and TSC tissues compared to focal epilepsies
464 caused by non-mTOR genes ($p<0.0001$). No correlation was observed between VAF and
465 HCN4 levels ($R^2=0.63$). Box represents median and interquartile range. Error bars indicate
466 minimum and maximum values.

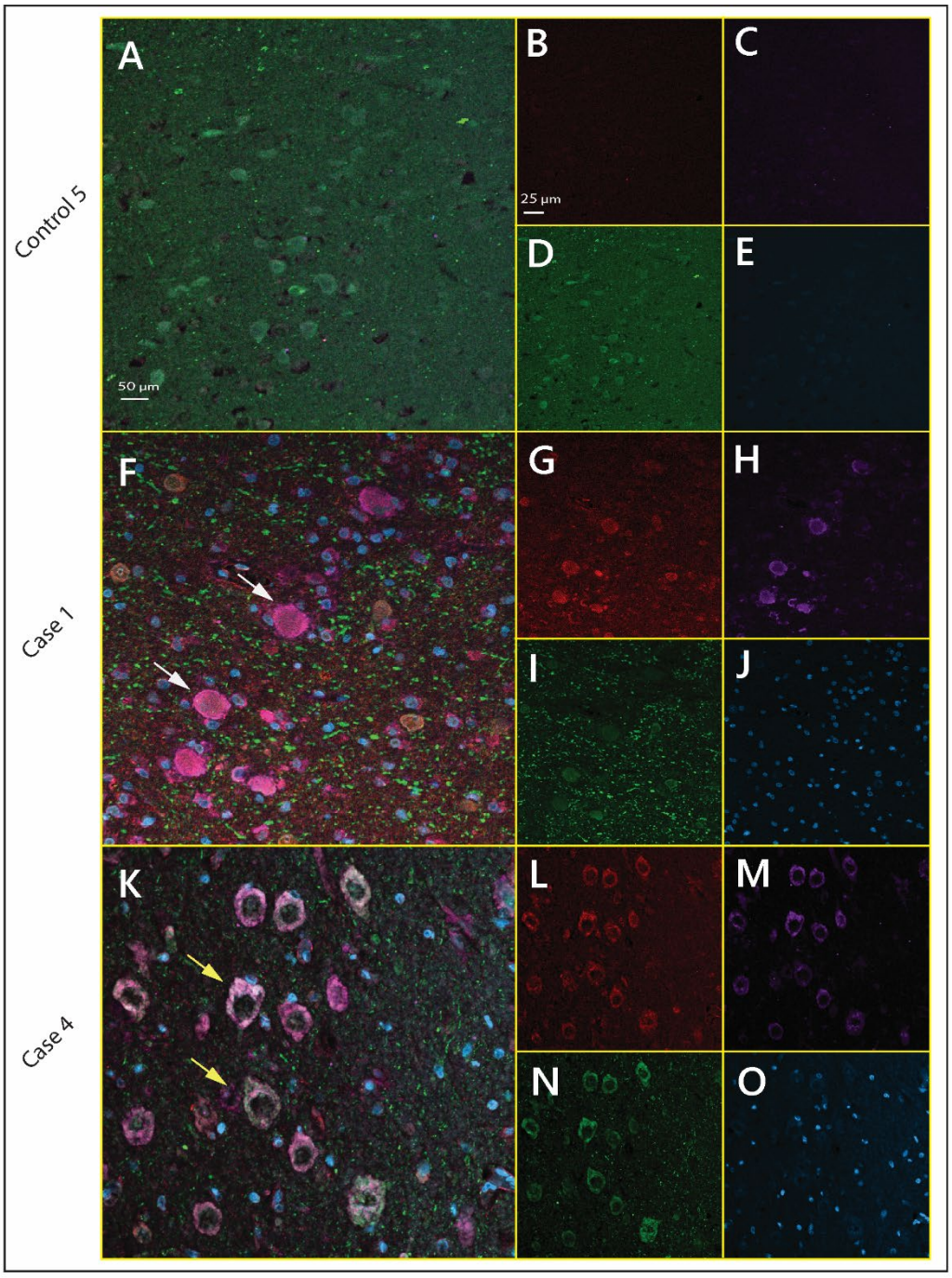


Fig. 2. HCN4 and pS6 immunofluorescence co-staining for control 5, case 1 (TSC) and case 4 (FCD IIB) at 20x magnification. A, F, K: combined HCN4, pS6 and SMI-32 immunofluorescence. B, G, L: HCN4 staining, indicated by red channel. C, H, M: pS6 staining, indicated by purple channel. D, I, N: SMI-32 staining, indicated by green channel. E, J, O; DAPI stain indicated by blue channel. Staining was performed on 3µm FFPE tissue sections. Immunofluorescence staining shows strong overlap between HCN4 and pS6 immunofluorescence in Cases 1 and 4, suggesting that HCN4 is in higher abundance in dysmorphic neurons (SMI-32 positive, marked by yellow arrows) and balloon cells (SMI-32 negative, marked by white arrows).

Table 1. Cohort of brain tissue specimens Individual	**Age at collection (y)**	**Sex**	**Age seizure onset (y)**	**Seizure types**	**Pathology**	**Gene**	**Amino Acid Change**	**VAF**	**Tissue region**	**Specimen type**	RNA
Control 1	43	M	n/a	n/a	n/a	n/a	n/a	n/a	L temporal cortex	Post-mortem	-
Control 2	49	M	n/a	n/a	n/a	n/a	n/a	n/a	frontal cortex	Post-mortem	-
Control 3	59	F	n/a	n/a	n/a	n/a	n/a	n/a	temporal cortex	Post-mortem	-
Control 4	60	F	n/a	n/a	n/a	n/a	n/a	n/a	frontal cortex	Post-mortem	-
Control 5	0.3	M	0	Fo	PMG	*ATP1A3*	L924P	Heterozygous	L frontal cortex	Post-mortem	+
Control 6	4	M	1.5	Sp	MOGHE	*SLC35A2*	L120Hfs*7	41.00%	L perisylvian cortex	Surgical	+
Control 7	2.7	M	0.2	Sp, Fo	MOGHE	*SLC35A2*	Q46*	20%	L superior temporal gyrus	Surgical	+
Control 8	1.2	F	0.5	Sp	MOGHE	*SLC35A2*	Q183*	1.80%	L superior temporal gyrus	Surgical	+
Control 9	2	M	0.3	Sp	MOGHE	*SLC35A2*	Q185*	16.50%	R frontal cortex	Surgical	+
Control 10	3	F	0.3	Sp, Fo	DNET	*FGFR1*	CNV	Heterozygous	L inferior frontal cortex	Surgical	+
Control 11	17	F	10.5	Fo	DNET	*FGFR1*	N546K	30%	L temporal cortex	Surgical	-
Case 1	2	M	0.9	Sp, Fo	TSC	*TSC2*	Q1192Rfs*18	Heterozygous	L temporal tuber	Surgical	+
Case 2	2	F	0.3	Sp, Fo	TSC	*TSC2*	Q1281*	Heterozygous	R cuneus tuber	Surgical	+
Case 3	2	F	1.3	Fo	TSC	*TSC2*	Q204*	Heterozygous	R posterior temporal occipital tuber	Surgical	+
Case 4	1.9	M	1.8	Fo	FCD IIB	*RHEB*	Y35L	13.00%	L frontal cortex	Surgical	-
Case 5	2	F	0.8	Fo	FCD IIB	*TSC1*	F581Hfs*6	2.80%	R temporal cortex	Surgical	-
Case 6	0.5	M	0.5	Sp, Fo	FCD IIB	*DEPDC5*	R843*	Heterozygous	L insular cortex	Surgical	0.5

Case 7	6	M	0	Sp, Fo	FCD IIB	*NPRL3*	T459Nfs*21	Heterozygous	R anterior insula long gyrus	Surgical	6
Case 8	12	M	2	Fo	FCD IIB	*MTOR*	T1977K	5.70%	R frontal cortex	Surgical	+
Case 9	3	F	1.4	Sp, Fo	FCD IIB	*MTOR*	P2425dup	4.10%	R frontal medial cortex	Surgical	+
Case 10	9	M	3	Fo	FCD IIB	*MTOR*	S2215Y	3.20%	R subcentral cortex	Surgical	+
Case 11	18	M	13	Fo	FCD IIB	*MTOR*	C1483R	3.00%	L insular cortex	Surgical	+
Case 12	17	M	6	Fo	FCD IIB	*MTOR*	S2215Y	2.10%	R precentral cortex	Surgical	-
Case 13	6	F	0	Fo	FCD IIB/ BOSD	*MTOR*	T1977K	2.90%	R frontal medial cortex	Surgical	+
Case 14	0.1	M	0	Sp, Fo	FCD IIA/ HME	*MTOR*	Y1450_L1453del	7.40%	R supramarginal gyrus	Surgical	+
Case 15	1	M	0	Sp, Fo	FCD IIA/ HME	*AKT3*	E17K	5.10%	R superior temporal gyrus	Surgical	+
Case 16	3	M	0	Fo	FCD IIA/ HME	*AKT3*	E17K	4.90%	R frontal cortex	Surgical	-
Case 17	0.6	M	0	Sp, Fo	FCD IIA/ HME	*PIK3CA*	E542K	20.00%	R anterior frontal cortex	Surgical	+
Case 18	5	M	0	Sp, Fo	FCD IIA/ HME	*PIK3CA*	H1047R	29.00%	L inferior parietal cortex	Surgical	-

DNET = dysembryoplastic neuroepithelial tumour; FCD IIA = focal cortical dysplasia type IIA; FCD IIB = focal cortical dysplasia type IIB; HME = hemimegalencephaly; MOGHE = mild malformation of cortical development with oligodendroglial hyploplasia in epilepsy; PMG = polymicrogyria; TSC = tuberous sclerosis complex; VAF = variant allele frequency; y= years; M = male; F = female; L = left; R = right; Fo= focal seizures; Sp = epileptic spasms; n/a = not applicable.

8.6 snRNA-seq quality control web summaries

See next page.

10x GENOMICS Cell Ranger • count

ML223537

Alerts

The analysis detected 1 error.

	Alert	Value	Detail
⊗	High Fraction of Reads Mapped Antisense to Genes	34.9%	Ideal < 10%. This can indicate use of an unsupported chemistry type (e.g. using Single Cell V(D)J for gene counting). Application performance is likely to be affected.

Summary | Analysis

4,266
Estimated Number of Cells

63,300
Mean Reads per Cell

1,607
Median Genes per Cell

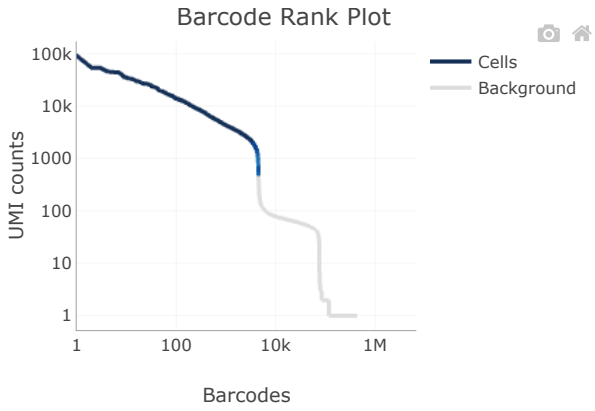
Sequencing

Number of Reads	270,037,764
Number of Short Reads Skipped	0
Valid Barcodes	97.0%
Valid UMIs	99.9%
Sequencing Saturation	82.7%
Q30 Bases in Barcode	94.7%
Q30 Bases in RNA Read	93.3%
Q30 Bases in UMI	94.2%

Mapping

Reads Mapped to Genome	96.5%
Reads Mapped Confidently to Genome	93.3%
Reads Mapped Confidently to Intergenic Regions	7.7%
Reads Mapped Confidently to Intronic Regions	0.0%
Reads Mapped Confidently to Exonic Regions	85.6%
Reads Mapped Confidently to Transcriptome	47.7%
Reads Mapped Antisense to Gene	34.9%

Cells



Estimated Number of Cells	4,266
Fraction Reads in Cells	79.2%
Mean Reads per Cell	63,300
Median Genes per Cell	1,607
Total Genes Detected	27,178
Median UMI Counts per Cell	2,936

Sample

Sample ID	ML223537
Sample Description	
Chemistry	Single Cell 3' v3
Include introns	False
Reference Path	…ger/GRCh38-1.2.0_premrna
Transcriptome	GRCh38-1.2.0_premrna-
Pipeline Version	cellranger-5.0.0

10x GENOMICS Cell Ranger • count

ML223051

Alerts

The analysis detected 2 errors.

	Alert	Value	Detail
⊗	High Fraction of Reads Mapped Antisense to Genes	35.1%	Ideal < 10%. This can indicate use of an unsupported chemistry type (e.g. using Single Cell V(D)J for gene counting). Application performance is likely to be affected.
⊗	Low Fraction Reads in Cells	49.6%	Ideal > 70%. Application performance may be affected. Many of the reads were not assigned to cell-associated barcodes. This could be caused by high levels of ambient RNA or by a significant population of cells with a low RNA content, which the algorithm did not call as cells. The latter case can be addressed by inspecting the data to determine the appropriate cell count and using --force-cells.

Summary | Analysis

3,128
Estimated Number of Cells

72,886
Mean Reads per Cell

1,937
Median Genes per Cell

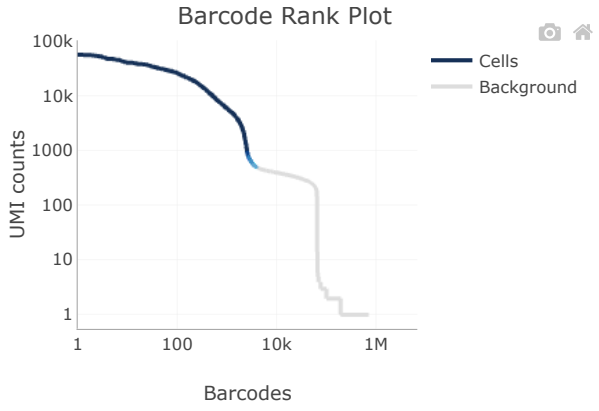
Sequencing

Number of Reads	227,985,898
Number of Short Reads Skipped	0
Valid Barcodes	97.0%
Valid UMIs	99.9%
Sequencing Saturation	60.8%
Q30 Bases in Barcode	96.5%
Q30 Bases in RNA Read	94.4%
Q30 Bases in UMI	96.2%

Mapping

Reads Mapped to Genome	96.8%
Reads Mapped Confidently to Genome	92.1%
Reads Mapped Confidently to Intergenic Regions	9.8%
Reads Mapped Confidently to Intronic Regions	0.0%
Reads Mapped Confidently to Exonic Regions	82.4%
Reads Mapped Confidently to Transcriptome	44.6%

Cells



Estimated Number of Cells	3,128
Fraction Reads in Cells	49.6%
Mean Reads per Cell	72,886
Median Genes per Cell	1,937
Total Genes Detected	27,962
Median UMI Counts per Cell	3,948

Sample

Sample ID	ML223051
Sample Description	
Chemistry	Single Cell 3' v3
Include introns	False
Reference Path	…ger/GRCh38-1.2.0_premrna
Transcriptome	GRCh38-1.2.0_premrna-

10x GENOMICS Cell Ranger • count

ML223536

Alerts

The analysis detected ⚠ 2 warnings.

	Alert	Value	Detail
⚠	High Fraction of Reads Mapped Antisense to Genes	18.5%	Ideal < 10%. This can indicate use of an unsupported chemistry type (e.g. using Single Cell V(D)J for gene counting). Application performance is likely to be affected.
⚠	Low Fraction Reads in Cells	61.8%	Ideal > 70%. Application performance may be affected. Many of the reads were not assigned to cell-associated barcodes. This could be caused by high levels of ambient RNA or by a significant population of cells with a low RNA content, which the algorithm did not call as cells. The latter case can be addressed by inspecting the data to determine the appropriate cell count and using --force-cells.

Summary | Analysis

2,645
Estimated Number of Cells

94,004
Mean Reads per Cell

2,604
Median Genes per Cell

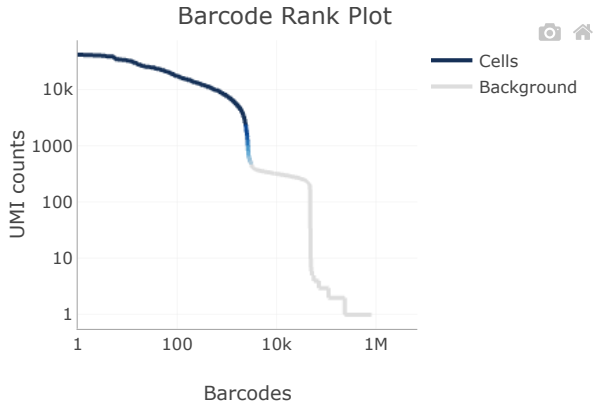
Sequencing

Number of Reads	248,639,287
Number of Short Reads Skipped	0
Valid Barcodes	97.5%
Valid UMIs	99.9%
Sequencing Saturation	78.7%
Q30 Bases in Barcode	94.8%
Q30 Bases in RNA Read	93.7%
Q30 Bases in UMI	94.3%

Mapping

Reads Mapped to Genome	96.9%
Reads Mapped Confidently to Genome	93.1%
Reads Mapped Confidently to Intergenic Regions	7.8%
Reads Mapped Confidently to Intronic Regions	0.0%
Reads Mapped Confidently to Exonic Regions	85.3%
Reads Mapped Confidently to Transcriptome	63.7%

Cells



Estimated Number of Cells	2,645
Fraction Reads in Cells	61.8%
Mean Reads per Cell	94,004
Median Genes per Cell	2,604
Total Genes Detected	26,685
Median UMI Counts per Cell	6,353

Sample

Sample ID	ML223536
Sample Description	
Chemistry	Single Cell 3' v3
Include introns	False
Reference Path	…ger/GRCh38-1.2.0_premrna
Transcriptome	GRCh38-1.2.0_premrna-

10x GENOMICS Cell Ranger • count

ML212988

Alerts

The analysis detected 1 error and 1 warning.

	Alert	Value	Detail
⊗	High Fraction of Reads Mapped Antisense to Genes	31.8%	Ideal < 10%. This can indicate use of an unsupported chemistry type (e.g. using Single Cell V(D)J for gene counting). Application performance is likely to be affected.
⚠	Low Fraction Reads in Cells	66.5%	Ideal > 70%. Application performance may be affected. Many of the reads were not assigned to cell-associated barcodes. This could be caused by high levels of ambient RNA or by a significant population of cells with a low RNA content, which the algorithm did not call as cells. The latter case can be addressed by inspecting the data to determine the appropriate cell count and using --force-cells.

Summary Analysis

4,094
Estimated Number of Cells

63,524
Mean Reads per Cell

1,986
Median Genes per Cell

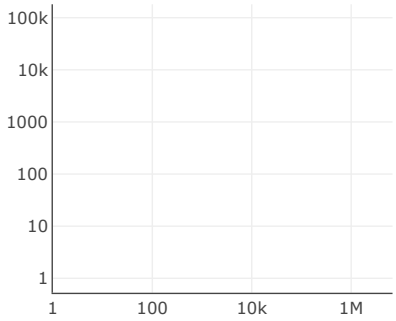
Sequencing

Number of Reads	260,067,584
Number of Short Reads Skipped	0
Valid Barcodes	95.7%
Valid UMIs	99.9%
Sequencing Saturation	68.8%
Q30 Bases in Barcode	95.1%
Q30 Bases in RNA Read	93.3%
Q30 Bases in UMI	94.7%

Mapping

Reads Mapped to Genome	95.7%
Reads Mapped Confidently to Genome	90.5%
Reads Mapped Confidently to Intergenic Regions	7.8%
Reads Mapped Confidently to Intronic Regions	0.0%
Reads Mapped Confidently to Exonic Regions	82.7%
Reads Mapped Confidently to Transcriptome	48.0%

Cells



Estimated Number of Cells	4,094
Fraction Reads in Cells	33%
Mean Reads per Cell	63,524
Median Genes per Cell	1,986
Total Genes Detected	27,748
Median UMI Counts per Cell	3,956

Sample

Sample ID	ML212988
Sample Description	
Chemistry	Single Cell 3' v3
Include introns	False
Reference Path	…ger/GRCh38-1.2.0_premrna
Transcriptome	GRCh38-1.2.0_premrna-

10x GENOMICS Cell Ranger • count

ML212989

Alerts

The analysis detected 1 error and 1 warning.

	Alert	Value	Detail
⊗	High Fraction of Reads Mapped Antisense to Genes	34.3%	Ideal < 10%. This can indicate use of an unsupported chemistry type (e.g. using Single Cell V(D)J for gene counting). Application performance is likely to be affected.
⚠	Low Fraction Reads in Cells	66.7%	Ideal > 70%. Application performance may be affected. Many of the reads were not assigned to cell-associated barcodes. This could be caused by high levels of ambient RNA or by a significant population of cells with a low RNA content, which the algorithm did not call as cells. The latter case can be addressed by inspecting the data to determine the appropriate cell count and using --force-cells.

Summary | Analysis

5,149
Estimated Number of Cells

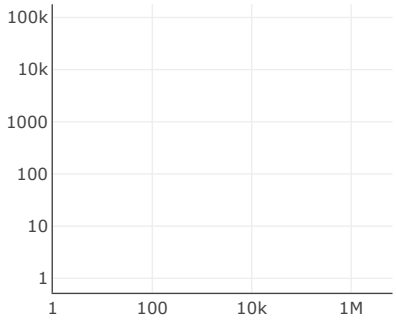
45,131
Mean Reads per Cell

1,883
Median Genes per Cell

Sequencing

Number of Reads	232,380,634
Number of Short Reads Skipped	0
Valid Barcodes	95.9%
Valid UMIs	99.9%
Sequencing Saturation	45.6%
Q30 Bases in Barcode	95.1%
Q30 Bases in RNA Read	93.4%
Q30 Bases in UMI	94.6%

Cells



Estimated Number of Cells	5,149
Fraction Reads in Cells	33.3%
Mean Reads per Cell	45,131
Median Genes per Cell	1,883
Total Genes Detected	28,418
Median UMI Counts per Cell	3,406

Mapping

Reads Mapped to Genome	95.9%
Reads Mapped Confidently to Genome	91.4%
Reads Mapped Confidently to Intergenic Regions	9.3%
Reads Mapped Confidently to Intronic Regions	0.0%
Reads Mapped Confidently to Exonic Regions	82.1%
Reads Mapped Confidently to Transcriptome	45.2%

Sample

Sample ID	ML212989
Sample Description	
Chemistry	Single Cell 3' v3
Include introns	False
Reference Path	…ger/GRCh38-1.2.0_premrna
Transcriptome	GRCh38-1.2.0_premrna-

10x GENOMICS Cell Ranger • count

ML212990

Alerts

The analysis detected 1 error and 1 warning.

	Alert	Value	Detail
⊗	High Fraction of Reads Mapped Antisense to Genes	32.1%	Ideal < 10%. This can indicate use of an unsupported chemistry type (e.g. using Single Cell V(D)J for gene counting). Application performance is likely to be affected.
⚠	Low Fraction Reads in Cells	62.4%	Ideal > 70%. Application performance may be affected. Many of the reads were not assigned to cell-associated barcodes. This could be caused by high levels of ambient RNA or by a significant population of cells with a low RNA content, which the algorithm did not call as cells. The latter case can be addressed by inspecting the data to determine the appropriate cell count and using --force-cells.

Summary | Analysis

4,170
Estimated Number of Cells

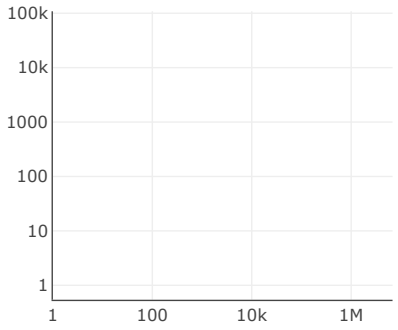
63,911
Mean Reads per Cell

2,207
Median Genes per Cell

Sequencing

Number of Reads	266,508,901
Number of Short Reads Skipped	0
Valid Barcodes	95.7%
Valid UMIs	100.0%
Sequencing Saturation	67.9%
Q30 Bases in Barcode	95.1%
Q30 Bases in RNA Read	93.2%
Q30 Bases in UMI	94.7%

Cells



Estimated Number of Cells	4,170
Fraction Reads in Cells	37.6%
Mean Reads per Cell	63,911
Median Genes per Cell	2,207
Total Genes Detected	27,166
Median UMI Counts per Cell	4,950

Mapping

Reads Mapped to Genome	95.4%
Reads Mapped Confidently to Genome	89.9%
Reads Mapped Confidently to Intergenic Regions	7.0%
Reads Mapped Confidently to Intronic Regions	0.0%
Reads Mapped Confidently to Exonic Regions	82.9%
Reads Mapped Confidently to Transcriptome	48.0%

Sample

Sample ID	ML212990
Sample Description	
Chemistry	Single Cell 3' v3
Include introns	False
Reference Path	…ger/GRCh38-1.2.0_premrna
Transcriptome	GRCh38-1.2.0_premrna-

10x GENOMICS Cell Ranger • count

ML212991

Alerts

The analysis detected 1 error.

	Alert	Value	Detail
⊗	Low Fraction Reads in Cells	43.2%	Ideal > 70%. Application performance may be affected. Many of the reads were not assigned to cell-associated barcodes. This could be caused by high levels of ambient RNA or by a significant population of cells with a low RNA content, which the algorithm did not call as cells. The latter case can be addressed by inspecting the data to determine the appropriate cell count and using --force-cells.

Summary | Analysis

4,554
Estimated Number of Cells

61,471
Mean Reads per Cell

1,906
Median Genes per Cell

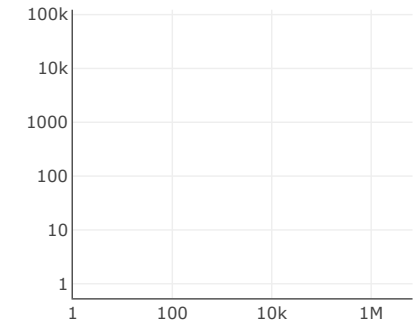
Sequencing

Number of Reads	279,940,744
Number of Short Reads Skipped	0
Valid Barcodes	96.3%
Valid UMIs	100.0%
Sequencing Saturation	67.2%
Q30 Bases in Barcode	95.0%
Q30 Bases in RNA Read	93.7%
Q30 Bases in UMI	94.6%

Mapping

Reads Mapped to Genome	94.4%
Reads Mapped Confidently to Genome	84.8%
Reads Mapped Confidently to Intergenic Regions	7.5%
Reads Mapped Confidently to Intronic Regions	0.0%
Reads Mapped Confidently to Exonic Regions	77.3%
Reads Mapped Confidently to Transcriptome	64.7%
Reads Mapped Antisense to Gene	9.4%

Cells



Estimated Number of Cells	4,554
Fraction Reads in Cells	56.8%
Mean Reads per Cell	61,471
Median Genes per Cell	1,906
Total Genes Detected	26,995
Median UMI Counts per Cell	3,821

Sample

Sample ID	ML212991
Sample Description	
Chemistry	Single Cell 3' v3
Include introns	False
Reference Path	…ger/GRCh38-1.2.0_premrna
Transcriptome	GRCh38-1.2.0_premrna-
Pipeline Version	cellranger-5.0.0

10x GENOMICS Cell Ranger • count

ML212992

Alerts

The analysis detected 1 error and 1 warning.

	Alert	Value	Detail
⊗	High Fraction of Reads Mapped Antisense to Genes	30.6%	Ideal < 10%. This can indicate use of an unsupported chemistry type (e.g. using Single Cell V(D)J for gene counting). Application performance is likely to be affected.
⚠	Low Fraction Reads in Cells	64.4%	Ideal > 70%. Application performance may be affected. Many of the reads were not assigned to cell-associated barcodes. This could be caused by high levels of ambient RNA or by a significant population of cells with a low RNA content, which the algorithm did not call as cells. The latter case can be addressed by inspecting the data to determine the appropriate cell count and using --force-cells.

Summary | Analysis

6,826
Estimated Number of Cells

42,410
Mean Reads per Cell

2,021
Median Genes per Cell

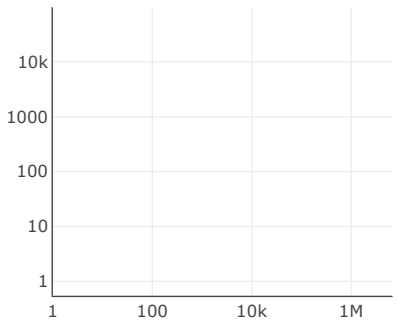
Sequencing

Number of Reads	289,491,795
Number of Short Reads Skipped	0
Valid Barcodes	95.8%
Valid UMIs	100.0%
Sequencing Saturation	60.6%
Q30 Bases in Barcode	95.2%
Q30 Bases in RNA Read	93.1%
Q30 Bases in UMI	94.7%

Mapping

Reads Mapped to Genome	95.6%
Reads Mapped Confidently to Genome	91.3%
Reads Mapped Confidently to Intergenic Regions	7.4%
Reads Mapped Confidently to Intronic Regions	0.0%
Reads Mapped Confidently to Exonic Regions	83.9%
Reads Mapped Confidently to Transcriptome	50.5%

Cells



Estimated Number of Cells	6,826
Fraction Reads in Cells	1%
Mean Reads per Cell	42,410
Median Genes per Cell	2,021
Total Genes Detected	27,946
Median UMI Counts per Cell	4,356

Sample

Sample ID	ML212992
Sample Description	
Chemistry	Single Cell 3' v3
Include introns	False
Reference Path	…ger/GRCh38-1.2.0_premrna
Transcriptome	GRCh38-1.2.0_premrna-

10x GENOMICS Cell Ranger • count

ML212993

Alerts

The analysis detected 1 error and 1 warning.

	Alert	Value	Detail
⊗	High Fraction of Reads Mapped Antisense to Genes	37.1%	Ideal < 10%. This can indicate use of an unsupported chemistry type (e.g. using Single Cell V(D)J for gene counting). Application performance is likely to be affected.
⚠	Low Fraction Reads in Cells	69.1%	Ideal > 70%. Application performance may be affected. Many of the reads were not assigned to cell-associated barcodes. This could be caused by high levels of ambient RNA or by a significant population of cells with a low RNA content, which the algorithm did not call as cells. The latter case can be addressed by inspecting the data to determine the appropriate cell count and using --force-cells.

Summary | Analysis

4,785
Estimated Number of Cells

55,681
Mean Reads per Cell

1,631
Median Genes per Cell

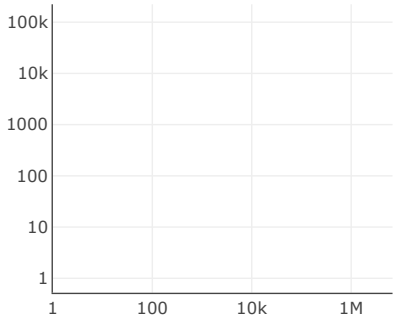
Sequencing

Number of Reads	266,435,288
Number of Short Reads Skipped	0
Valid Barcodes	95.6%
Valid UMIs	100.0%
Sequencing Saturation	71.0%
Q30 Bases in Barcode	95.1%
Q30 Bases in RNA Read	93.3%
Q30 Bases in UMI	94.7%

Mapping

Reads Mapped to Genome	95.8%
Reads Mapped Confidently to Genome	90.6%
Reads Mapped Confidently to Intergenic Regions	7.9%
Reads Mapped Confidently to Intronic Regions	0.0%
Reads Mapped Confidently to Exonic Regions	82.7%
Reads Mapped Confidently to Transcriptome	42.8%

Cells



Estimated Number of Cells	4,785
Fraction Reads in Cells	30.9%
Mean Reads per Cell	55,681
Median Genes per Cell	1,631
Total Genes Detected	27,880
Median UMI Counts per Cell	2,968

Sample

Sample ID	ML212993
Sample Description	
Chemistry	Single Cell 3' v3
Include introns	False
Reference Path	…ger/GRCh38-1.2.0_premrna
Transcriptome	GRCh38-1.2.0_premrna-

10x GENOMICS Cell Ranger • count

ML212994

Alerts

The analysis detected 1 error.

	Alert	Value	Detail
⊗	High Fraction of Reads Mapped Antisense to Genes	35.3%	Ideal < 10%. This can indicate use of an unsupported chemistry type (e.g. using Single Cell V(D)J for gene counting). Application performance is likely to be affected.

Summary Analysis

7,807
Estimated Number of Cells

32,947
Mean Reads per Cell

2,038
Median Genes per Cell

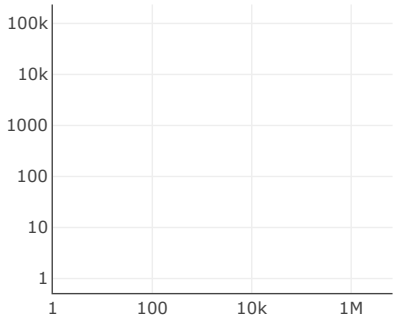
Sequencing

Number of Reads	257,213,617
Number of Short Reads Skipped	0
Valid Barcodes	95.8%
Valid UMIs	100.0%
Sequencing Saturation	42.6%
Q30 Bases in Barcode	95.1%
Q30 Bases in RNA Read	93.4%
Q30 Bases in UMI	94.7%

Mapping

Reads Mapped to Genome	96.1%
Reads Mapped Confidently to Genome	90.9%
Reads Mapped Confidently to Intergenic Regions	9.0%
Reads Mapped Confidently to Intronic Regions	0.0%
Reads Mapped Confidently to Exonic Regions	82.0%
Reads Mapped Confidently to Transcriptome	44.2%
Reads Mapped Antisense to Gene	35.3%

Cells



Estimated Number of Cells	7,807
Fraction Reads in Cells	3%
Mean Reads per Cell	32,947
Median Genes per Cell	2,038
Total Genes Detected	28,861
Median UMI Counts per Cell	3,808

Sample

Barcodes

Sample ID	ML212994
Sample Description	
Chemistry	Single Cell 3' v3
Include introns	False
Reference Path	…ger/GRCh38-1.2.0_premrna
Transcriptome	GRCh38-1.2.0_premrna-
Pipeline Version	cellranger-5.0.0

10x GENOMICS Cell Ranger • count

ML212995

Alerts

The analysis detected 1 error.

	Alert	Value	Detail
⊗	High Fraction of Reads Mapped Antisense to Genes	34.1%	Ideal < 10%. This can indicate use of an unsupported chemistry type (e.g. using Single Cell V(D)J for gene counting). Application performance is likely to be affected.

Summary Analysis

9,010
Estimated Number of Cells

30,428
Mean Reads per Cell

2,147
Median Genes per Cell

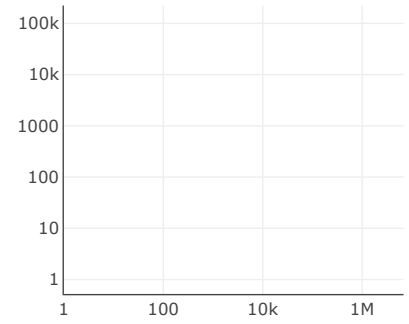
Sequencing

Number of Reads	274,157,161
Number of Short Reads Skipped	0
Valid Barcodes	95.8%
Valid UMIs	100.0%
Sequencing Saturation	43.9%
Q30 Bases in Barcode	95.0%
Q30 Bases in RNA Read	93.4%
Q30 Bases in UMI	94.6%

Mapping

Reads Mapped to Genome	96.0%
Reads Mapped Confidently to Genome	91.3%
Reads Mapped Confidently to Intergenic Regions	8.0%
Reads Mapped Confidently to Intronic Regions	0.0%
Reads Mapped Confidently to Exonic Regions	83.3%
Reads Mapped Confidently to Transcriptome	46.3%
Reads Mapped Antisense to Gene	34.1%

Cells



Estimated Number of Cells	9,010
Fraction Reads in Cells	95.8%
Mean Reads per Cell	30,428
Median Genes per Cell	2,147
Total Genes Detected	28,865
Median UMI Counts per Cell	4,232

Sample

Sample ID	ML212995
Sample Description	
Chemistry	Single Cell 3' v3
Include introns	False
Reference Path	…ger/GRCh38-1.2.0_premrna
Transcriptome	GRCh38-1.2.0_premrna-
Pipeline Version	cellranger-5.0.0