THE UNIVERSITY of ADELAIDE

The Genetic Basis of Malformation of Cortical Development Syndromes:

Primary Focus on Aicardi Syndrome

Thuong Thi Ha
B. Sc, M. Bio

Neurogenetics Research Group
The University of Adelaide

Thesis submitted for the degree of
Doctor of Philosophy

In

Discipline of Genetics and Evolution
School of Biological Sciences
Faculty of Science
The University of Adelaide

June 2018

Table of Contents

Abstract

Aicardi Syndrome is a rare neurodevelopmental disorder recognized by a classical triad of chorioretinal lacunae, infantile spasms and agenesis of the corpus callosum. The revised diagnostic criteria of Aicardi Syndrome have been broadened to include additional phenotypes outside of the classical triad that are frequently observed. Early investigations into the genetics of Aicardi Syndrome were predominantly focused on chromosome X for two reasons. Firstly, the first chromosomal aberration reported in a suspected Aicardi female was an X/3 translocation. Secondly, an X-linked male-lethal cause would best explain the predominance of the disease in females, which are represented in more than 99% of cases reported in literature. Despite 70 years of genetic and genomic investigations, an X-linked cause has yet to be established.

We performed whole exome and genome sequencing on a cohort of 13 individuals diagnosed with or suspected of Aicardi Syndrome. The phenotypes displayed by of our cohort were clinically heterogeneous, which enabled us to query whether there are: (1) different genetic causes underlying different clinical subsets of Aicardi Syndrome or (2) mutations in a recurring gene that is pleiotropic and with variable penetrance or expressivity. We also investigated the molecular consequences of candidate genes using protein-specific *in vitro* assays and/or the phenotypic consequences of morpholino-mediated knockdown in zebrafish embryos.

We utilised a list-based approach to enrich for variants in genes associated with eye and brain development. To test the efficiency of this list-based approach in finding a genetic diagnosis, we performed whole exome sequencing in a family with autosomal dominant disorder comprising of porencephaly, focal epilepsy and lens opacities. From the affected individuals in this multi-generation family, who were negative for mutations in *COL4A1*, we found a rare variant in *COL4A2*. Mutations in *COL4A1* are well described and result in brain abnormalities manifesting

with severe neurological deficits. Our findings expanded the phenotypic spectrum associated with *COL4A2* and highlight the increasing overlap with phenotypes associated with *COL4A1*.

We applied the same list-based approach to our Aicardi Syndrome cohort study. In five unrelated individuals, we identified *de novo* variants in *HCN1*, *KMT2B*, *SLF1*, *SZT2* and *WNT8B* respectively. Next, we assessed the likely pathogenicity of the autosomal variants we identified using a combination of: pre-existing in vitro assays (*HCN1* and *WNT8B*), published expression and phenotype studies in human or mice and morpholino knockdown in zebrafish (*Danio rerio*) embryos. Our findings show that causes underlying AIC are genetically heterogeneous but converge on molecular pathways central to cortical development. We highlighted the importance of utilising genetic studies to guide differential diagnosis of syndromes like Aicardi Syndrome that involve multiple complex traits. This will lead to better understanding of how these brain disorders arise, accurate genetic diagnosis and potentially gene-tailored treatments.

Thesis Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works. I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time. I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Master of Biotechnology (Biomedical)

Student Number:

Signed: Thuong Ha

Date: 26/06/2018

Acknowledgements

Firstly, I would like to thank the families and patients for giving consent and providing biological samples to this PhD study. You are the foundation and biggest motivation for the completion of this thesis.

I am grateful to the clinicians, scientists and academics that have contributed to the medical, practical and intellectual aspects of my work. I would especially like to thank the members of the Epilepsy Research Retreat for providing constructive feedback and helping me mature into an early career researcher.

I would like to thank past and present members of the Neurogenetics Research Group for the useful feedback during Monday meeting and casual banter. The scientific knowledge I have acquired from the senior members and the hilarious memories I made with the younger crew will forever be embedded in my head and heart. Special thank you to my coffee gang (Atma, Dee, Cbear, Tilda, Tessa and Sayaka) for the much-needed peer pressure to write and submit.

This PhD journey would not be possible without the love and support from my wonderful family and friends. Thank you, mum, for the food, dad for the beers and Quoc for the company. Thank you to the tripod + boothangs and PhanHaTranLe family for the social and comedic relief. I would also like to dedicate this thesis to my late grandpa and grandma who oversaw the completion of this thesis from above.

I would especially like to acknowledge my two supervisors, Prof. Jozef Gecz and Dr. Mark Corbett for giving me the opportunity to pursue a PhD. Thank you Jozef for blessing me with strong bones and fertility every year around Easter. Thank you Mark for your healthy brain

stimulations and beer support at conferences. I will forever be your student and in debt to you both for the knowledge and guidance you have provided throughout my whole PhD.

Last but not least, I would like to thank the two beloved people in my life. To my partner, Tong Bok, thank you for putting up with my mood and food swings. You have been my pillar of support through this difficult journey. Finally to my darling baby, Madeleine Bok, this thesis was written up during my entire pregnancy with you. I dedicate this second biggest achievement of my life to you baby girl, who makes me thrive to be a better scientist and person.

Publications

1. **Ha, T.T**., Sadleir, L.G., Mandelstam, S.A., Paterson, S.J., Scheffer, I.E., Gecz, J., and Corbett, M.A. (2016). A mutation in COL4A2 causes autosomal dominant porencephaly with cataracts. Am J Med Genet A 170A, 1059-1063.

2. Kumar, R.*, **Ha, T.***, Pham, D., Shaw, M., Mangelsdorf, M., Friend, K.L., Hobson, L., Turner, G., Boyle, J., Field, M., et al. (2016). A non-coding variant in the 5' UTR of DLG3 attenuates protein translation to cause non-syndromic intellectual disability. Eur J Hum Genet 24, 1612-1616. *Co-first authors

3. Corbett, M.A., Turner, S.J., Gardner, A., Silver, J., Stankovich, J., Leventer, R.J., Derry, C.P., Carroll, R., **Ha, T**., Scheffer, I.E., et al. (2017). Familial epilepsy with anterior polymicrogyria as a presentation of COL18A1 mutations. Eur J Med Genet 60, 437-443.

Abbreviations

A - E

a-CGH	Array Comparative Genomic Hybridisation
ACC	Agenesis of Corpus Callosum
AED	Anti-epileptic Drugs
AIC	Aicardi Syndrome
BCF	Binary Variant Call Format
BP	Base Pair
BSA	Bovine Serum Albumin
BWA	Burrows-wheeler Aligner
cDNA	Complementary DNA
ChrX	Chromosome X
CNV	Copy Number Variant
COL IV	Collagen Type IV
CRL	Chorioretinal Lacunae
DEL	Deletion
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
dpc	Days Post Conception
DUP	Duplication
ECM	Extracellular Matrix
EDTA	Ethylenediaminetetraacetic Acid
EE	Epileptic Encephalopathies
EEIM	Eye, Epilepsy, Intellectual Disability and Malformation of Cortical Development

F - J

GATK	Genome Analysis Toolkit
GC	Guanine-Cytosine
HCN1	Hyperpolarisation Activated Cyclic Nucleotide Gated Potassium Channel 1
HEK293T	Human Embryonic Kidney 293T Cell Line
HGNC	HUGO Gene Nomenclature Committee
hpf	Hours Post Fertilisation
INS	Insertion
INV	Inversion
IS	Infantile Spasm

K - O

Kb	Kilo-base Pair
KMT2B	Histone-lysine N-methyltransferase 2B
LB	Luria Broth
LCL	Lymphoblastoid Cell Lines
LN_2	Liquid Nitrogen
Mb	Mega-base Pair
MCD	Malformations of Cortical Development
MEI	Mobile Element Insertion
MO	Morpholino Antisense Oligonucleotides
ORF	Open Reading Frame

P - T

PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
RLA	Relative Luciferase Activity

RNA	Ribonucleic Acid
RPMI	Roswell Park Memorial Institute Medium
SLF1	SMC5-SMC6 Complex Localisation Factor Protein 1
SOC	Super Optimal Broth
STR	Short Tandem Repeats
SV	Structural Variant
SZT2	Seizure Threshold 2
TBS-T	Tris Buffered Saline with Tween20
TEAD1	Transcriptional Enhancer Factor TEF-1
TRA	Translocation
Tris-HCl	Tris Hydrochloric Acid

U - Z

VCF	Variant Call Format
VN	Vagus Nerve
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
WNT8B	Wingless-type MMTV Integration Site Family, Member 8b
XIA	X-inactivation

CHAPTER I: Introduction

1.1 Overview of Malformations of Cortical Development (MCD)

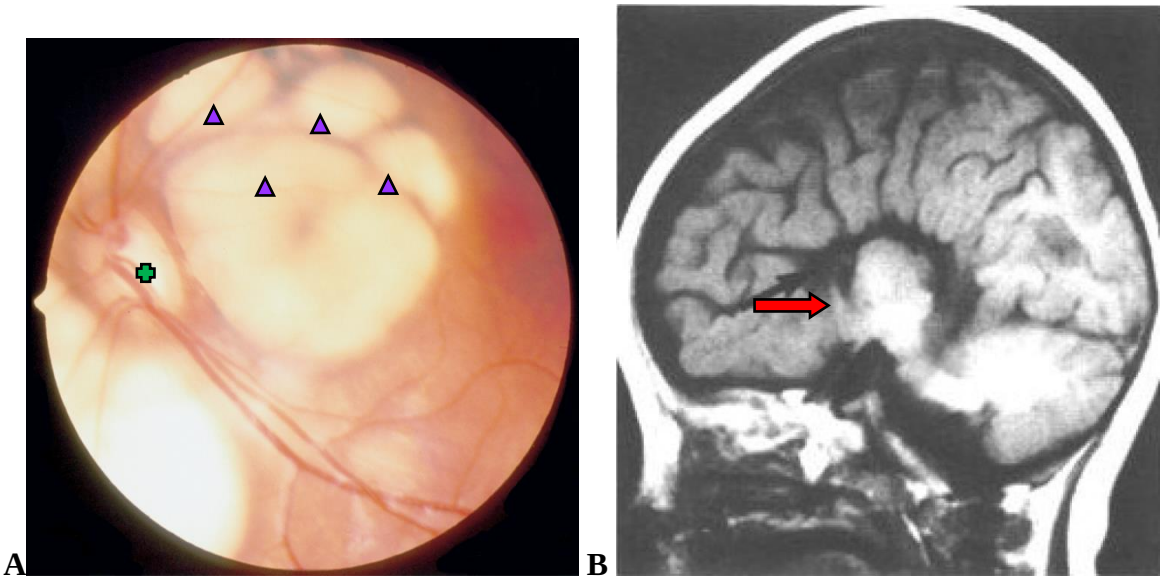
Cortical development is a dynamic and multiplex process; involving the proliferation, differentiation, migration and organisation of neuronal cells (Guerrini and Dobyns, 2014). In humans, cortical development occurs in a series of overlapping morphological events during the first two trimesters of pregnancy (Leventer et al., 2008, Baffet et al., 2016). The process begins with the formation of the neural plate at 18 days post conception (dpc), followed by differentiation of the early embryonic brain into three divisions (prosencephalon, mesencephalon and rhombencephalon at 33 dpc (Barkovich et al., 2005). Next, the proliferation of cortical precursor cells starting in the ventricular zone. Between the 10-16th week of gestation, cell migration occurs in serial waves starting from the intermediate zone to the marginal zone and eventually reaching the cortical plate. By the 22nd week of gestation, radial or tangential movements of cortical cells form 6 distinctive layers of the cerebral cortex. Finally, cortical organisation commences with the following molecular processes: neurite extension, synaptogenesis and neuronal maturation (Barkovich et al., 2005).

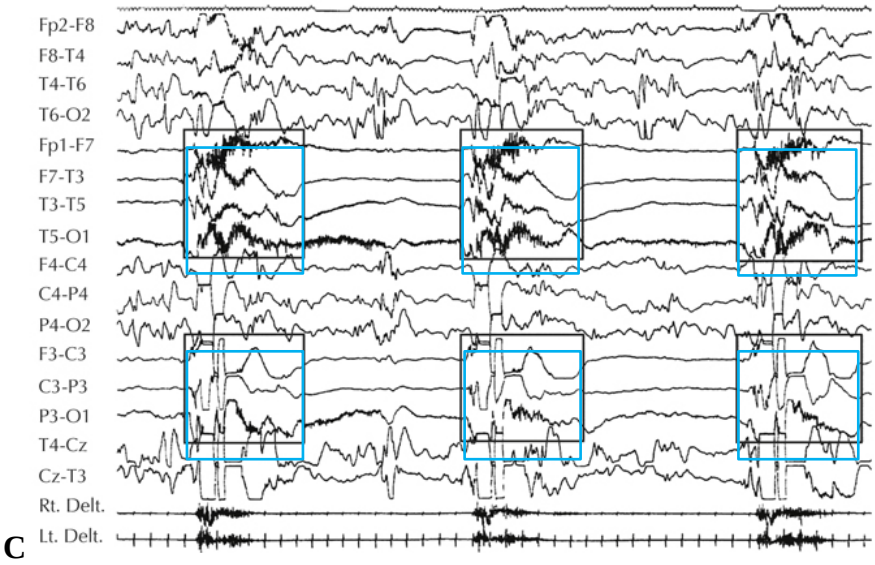
Perturbations at any stage during neurogenesis through to neuronal migration can result in a variety of structural abnormalities of the brain, known as Malformations of cortical development (MCD) (Baffet et al., 2016). The aetiology of MCDs can be of environmental (*in utero* infections or ischaemia) or genetic inception (Leventer et al., 2008). There are genes that regulate a variety of cortical development processes; by both autonomous and extrinsic means (Barkovich, 2005). Thus unsurprisingly, variants in more than 100 genes have been identified as the cause of MCD in various human syndromes (Guerrini and Dobyns, 2014). MCD can lead to a variety of clinical manifestations with the most common being recurrent and intractable seizures; 25-40% of

childhood epilepsy are a consequence of MCDs (Leventer et al., 2008). Many MCD syndromes have been characterised and diagnosed based on clinical, imaging and genetic screening. However, there still remains a subset of unresolved MCD syndromes such as Aicardi Syndrome, whereby the genetic origins have yet to be identified.

1.2 Introduction into Aicardi Syndrome

Aicardi Syndrome (MIM 405060; AIC) is characterized by a classical triad of callosal agenesis, chorioretinal lacunae (CRL) and infantile spasms (**Figure 1.1**) (Aicardi, 2005). There are additional cerebral, ocular and musculoskeletal and systemic defects associated with AIC, which are subject to interindividual variability. The symptoms of AIC were first clinically noted by Krause (1946), who described an infant with intellectual disability, infantile seizures and ocular plaques (reminiscent of CRL). AIC was later recognised as an independent neuropathological entity from other congenital disorders, based on a series of 15 affected individuals (Aicardi et al., 1969). The prevalence of AIC is estimated at 4000 cases worldwide (Kroner et al., 2008). In addition to previous estimates, more than 1000 cases (**Figure 1.2**) have been included in serial studies to date (Aicardi, 2005, Iwamoto et al., 2008).





C

Figure 1.1: Classical Triad of Aicardi Syndrome. A: shows a funduscopic image of typical chorioretinal lacunae (triangles); which usually appear in multiple, rounded and white areas adjacent to the optic disc (circle) (Ospina et al., 2004). **B:** shows a brain magnetic resonance image of complete agenesis of the corpus callosum (red arrow); which is the absence of the structure connecting the two brain hemispheres (Baierl et al., 1988). **C:** shows an interictal electroencephalograph recording of infantile spasms originating from the left hemisphere with high-voltage slow waves (blue squares) (Kobayashi et al., 2011).

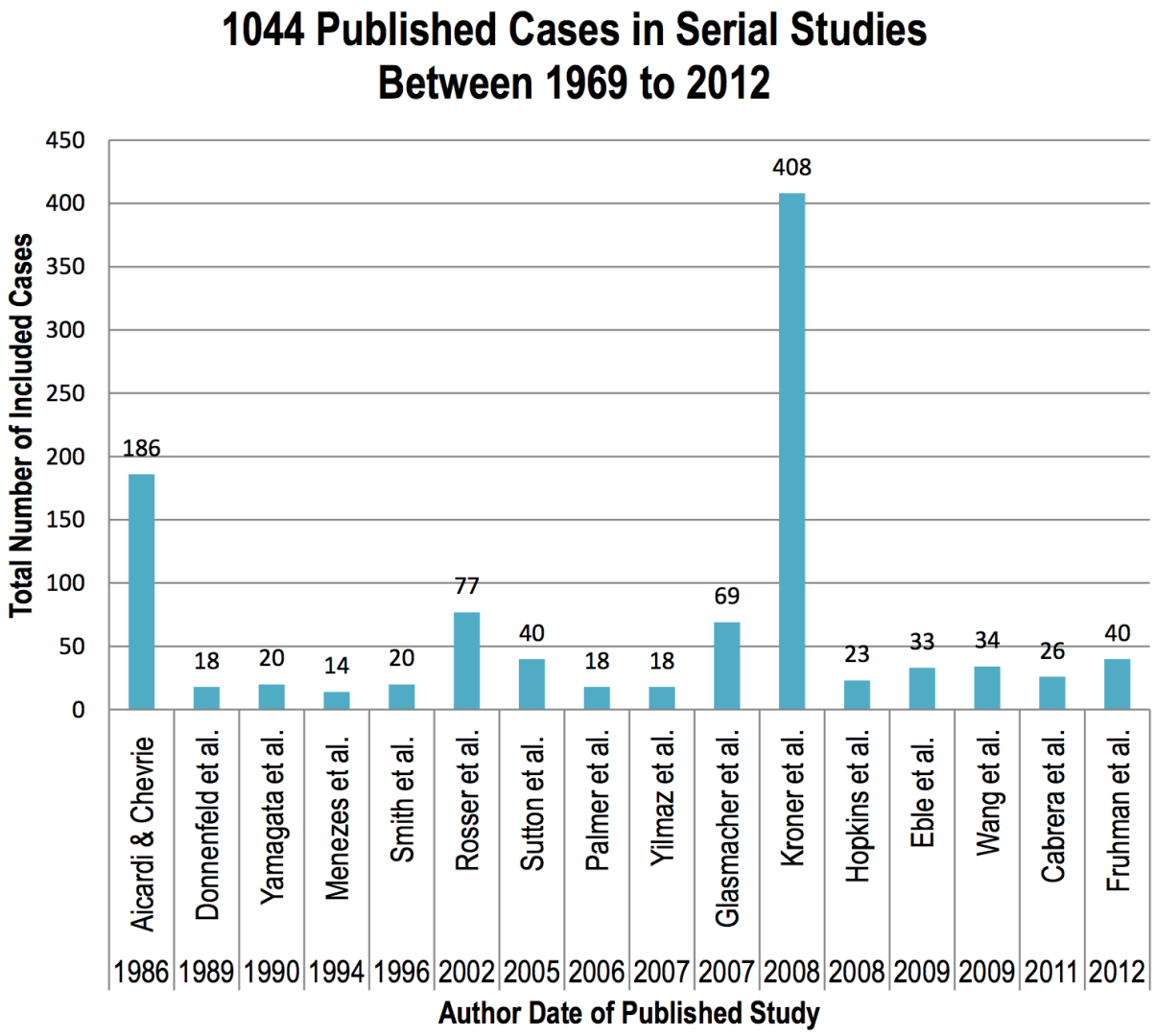


Figure 1.2: Published Cases in Serial Studies between 1969 and 2012. This graph shows the number of Aicardi cases included in serial studies between 1969 and 2012. The author and date of the published studies are shown on the X axis. The numbers of included cases are labelled on top of the corresponding bars. This graph does not show the exact number of all Aicardi cases reported in literature. Individual case reports, females with cytogenetically visible findings and males with Aicardi syndrome are not included in this graph. There may also be replicate cases in studies from 2007 to 2012, where patients are sourced from northern US organisations and institutes.

The aetiology of AIC remains enigmatic (Aicardi, 2005, Van den Veyver, 2002). Environmental factors such as exposure to teratogenic agents or intrauterine disturbances have been refuted in the pathology of AIC, suggesting that the disorder is likely to have a genetic basis Donnenfeld et al. (1989). AIC is reported almost exclusively in females or rarely in Klinefelter males (Aicardi, 2005, Chen et al., 2009, Hopkins et al., 1979, Shetty et al., 2014, Zubairi et al., 2009). The diagnosis of AIC in 46,XY males have either been disputed or disregarded in literature (Chappelow et al., 2008). The gender bias observed in AIC has led many to conclude that an X-linked, male-lethal, *de novo* dominant mutation is responsible for the syndrome (van den Veyver, 2002, Eble et al., 2009, Yilmaz et al., 2007). Approaches used to investigate the pathogenesis of AIC include: Array comparative genome hybridisation, skewed X-inactivation studies, candidate gene sequencing, whole exome and whole genome sequencing and cytogenetic examinations. Despite the investigative efforts thus far, no convincing candidate gene or locus has yet to be identified (Eble et al., 2009, Yilmaz et al., 2007, Wieacker et al., 1985, Anderson et al., 2009, Hoag et al., 1997, Neidich et al., 1990, Nielsen et al., 1991, Wang et al., 2009, Pai et al., 2013, Schrauwen et al., 2015; Lund et al., 2016).

1.3 Clinical Features of Aicardi Syndrome

1.3.1 Epidemiology

Based on a cohort of 408 individuals with confirmed or suspected AIC, the incidence of AIC was estimated at one in every 99,000 to 183,000 live births in European and US populations respectively (Kroner et al., 2008). The emerging reports of individuals with milder forms of AIC suggest the prevalence may be higher than predicted, as a majority of earlier reports described severe or atypical forms of the disorder (Saddichha et al., 2007, King et al., 1998, Menezes et al., 1994, Matlary et al., 2004, Grosso et al., 2007, Yacoub et al., 2003, Lee et al., 2004, Prats Vinas et al., 2005, Glasmacher et al., 2007). It is unknown whether atypical forms of AIC should be considered as distinct syndromes or represent true variant forms allelic to AIC (Bursztejn et al., 2009, Spennato et al., 2013).

The recurrence risk for family members of a proband with AIC is predicted to be low (Sutton and Van den Veyver, 1993). Parents of probands with AIC are mostly asymptomatic and have negative family history for ophthalmologic and/or neurological disorders (Donnenfeld et al., 1989, Glasmacher et al., 2007, Hopkins et al. 2008). There are no known cases of AIC transmission from parent to child There are two known and isolated familial cases of AIC; one pair of sisters from non-consanguineous parents and one pair of monozygotic twins concordant for AIC (Molina et al., 1989, Pons and Garcia, 2008). However, a majority of AIC cases occur sporadically, presumed to be the result of a *de novo* variant, thus siblings are not considered to be at risk (Sutton and Van den Veyver, 1993).

1.3.2 Clinical Diagnosis

The original diagnosis of AIC was based on the classical triad (**Figure 1.1**) and accompanying traits including: subependymal heterotopias, costovertebral defects and various neurological anomalies (Chevrie and Aicardi, 1986). Further AIC-related anomalies were later revealed through advancement in imaging technology (Aicardi, 2005, Hopkins et al., 2008, Chevrie and Aicardi, 1986). These anomalies led to the revision of the diagnostic criteria, to include frequently recurring phenotypes.

The revised diagnosis also allowed the inclusion of cases that did not meet the classical diagnosis; individuals who were missing one or more features of the classical triad (Aicardi, 2005; **Figure 1.1**). The accepted criteria for the diagnosis AIC include two major and two supporting features (**Table 1.1**) (Sutton et al., 2005). This criterion is based on commonly observed neurological, musculoskeletal and ocular defects. There are other common presentations outside of the diagnostic criteria including: intellectual disability, respiratory problems, gastrointestinal upsets and craniofacial abnormalities (Aicardi, 2005, Glasmacher et al., 2007, Sutton et al., 2005). Due to its phenotypic heterogeneity, the diagnosis for AIC is heavily reliant on the accurate assessment of fundus examinations, brain magnetic resonance imaging and radiological finding (Sutton et al., 2005, Sutton and Van den Veyver, 1993).

Type of Findings[1]	Major Features	Supporting Features	Other Features
Classical Triad (Figure 1.1)	Agenesis of the Corpus Callosum - Total/partial Chorioretinal Lacunae Infantile Spasms - Other types of seizures		
Ocular	Optic disc/nerve coloboma	Microphthalmia - Other eye abnormalities	Detached retina Anterior synechae
Neurological	Cortical Malformations - Microgyria Heterotopia - Periventricular - Subcortical Intracranial Cysts - Interhemispheric	Split-brain electroencephalogram Gross cerebral hemispheric asymmetry	Hypotonia
Radiologic		Scoliosis Hemi-vertebrae Absent or abnormal ribs	Hand malformations
Craniofacial		Prominent premaxilla Upturned nasal tip Sparse lateral eyebrows	Cleft lip Cleft palate
Skin			Vascular malformations/ malignancies Pigmentary lesions

Table 1.1: Clinical Presentation of Aicardi Syndrome (AIC).

[1]Clinical features were collated from the following papers: Aicardi, 2005; Glasmacher et al., 2007; Fruhman et al., 2012; Sutton and Van den Veyver, 1993; McMahon et al., 1984; Iwamoto et al., 2008; Sutton et al., 2005

1.3.3 Differential Diagnosis

Chorioretinal Lacunae (CRL) are considered pathognomonic to AIC (**Figure 1.1**) (Aicardi, 2005, Glasmacher et al., 2007, Chevrie and Aicardi, 1986, Fruhman et al., 2012, Sutton et al., 2005). Similar chorioretinal defects have been described in patients with Oculoauricular Syndrome (MIM 142992), Orofaciodigital Syndrome IX (MIM 258865) and Kniest Dysplasia (MIM 120140). The lacunae in AIC can be distinguished from other differential diagnosis and congenital infections based on their distinctive appearance and anatomical position, which can be determined via funduscopic and histological examinations (**Table 1.2**) (Fruhman et al., 2012, Kurugundla et al. 2008). The diagnoses of patients suspected to have AIC are often disputed if the CRL are absent or atypical (Ropers et al., 1982, Burch-Smith et al., 2012, Leng and Moshfeghi, 2011, Willis and Rosman, 1980).

Syndrome	Sex	Infantile Spasms	Agenesis of Corpus Callosum	Chorioretinopathy	Other Ophthalmic Abnormality
Aicardi Syndrome	Female	Always	Always	Bilateral; posterior; little pigment; "lacunar"	Microphthalmia; coloboma; iris synechia; optic atrophy
Congenital Toxoplasmosis	Male or Female	Occasional	No	Uni-/bilateral; posterior of peripheral; often perimacular; pigmentary	Cataracts; microphthalmia; uveitis; optic atrophy; lenticular or corneal opacities
Cytomegalic Inclusion Disease	Male or Female	Occasional	No	Uni-/bilateral; posterior of peripheral; pigmentary	Conjunctivitis
Congenital Rubella	Male or Female	Occasional	No or Rare	Uni-/bilateral; posterior of peripheral; pigmentary; "salt and pepper"	Cataracts; microphthalmia; glaucoma; iris hypoplasia
Congenital Syphilis	Male or Female	Occasional	No	Uni-/bilateral; posterior of peripheral; pigmentary; "salt and pepper"; pleomorphic	Keratitis; uveitis; glaucoma
Congenital Herpes Simplex	Male or Female	Not described	No	Uni-/bilateral; often peripheral; pigmentary	Cataracts; keratitis; conjunctivitis; microphthalmia

Table 1.2: **Aicardi Syndrome versus Congenital Infections.** This table was adapted from O'Driscoll et al. (2010).

In contrast to CRL, Agenesis of the Corpus Callosum (ACC) and Infantile Spasms (IS) are not specific to AIC (**Figure 1.1** & **Table 1.3**) (Paciorkowski et al., 2011, Rosser et al., 2002). For the diagnosis of AIC, the absence of ACC can be substituted by the presence of other neurological defects such as intracranial cysts, heterotopias or cortical malformations (Aicardi, 2005, Sutton and Van den Veyver, 1993). Similarly, IS can be interchanged in the diagnosis of AIC with other early-onset epileptic seizures (Aicardi 2005, Chevrie & Aicardi, 1986). AIC requires brain magnetic resonance imaging and electroencephalograms for the accurate differential diagnosis of syndromes listed in **Table 1.3.**

1.3.4 Development & Prognosis

The prognosis for AIC depends on the extent and severity of the clinical findings (Donnenfeld et al., 1989, Grosso et al., 2007, Glasmacher et al., 2007, Rosser et al. 2001). In addition to the eleven patients noted by Abe et al. (1990), eight more patients were recently reported with favourable outcomes (Saddichha et al., 2007, King et al., 1998, Menezes et al., 1994, Matlary et al., 2005, Grosso et al., 2007, Yacoub et al., 2003, Lee et al., 2004, Vinas et al., 2005, Glasmacher et al., 2007, Kasasbeh et al., 2014). All eight patients had incomplete classical triad (**Figure 1.1**) and relatively mild neurological findings; resulting in less cognitive decline and increased life expectancy (Grosso et al., 2007). The oldest known survivor of AIC is a 49-year-old woman reported by King et al. (1998). The survival estimates for AIC is 62% at 27 years of age, which is higher than previously reported (Kroner et al., 2008, Donnenfeld et al., 1989, Menezes et al., 1994, Glasmacher et al., 2007). The mean and median age of death is estimated at 8.3 and 18.5 years respectively (Sutton and Van den Veyver, 1993).

Syndromes[2]	Genotype	Phenotype
Infantile spasms	ARX (MIM 300350)	Cortical atrophy, reduced WM, abnormal BG, EIEE, variable LGS, dystonia/hyperkinesia common
	CDKL5 (MIM 300203)	Cortical atrophy, abnormal WM, calcifications, LGS, autistic features common
	FOXG1	Variable dysgenesis of CC, autistic features common
	MEF2C	Atrophy, reduced WM, variable subsequent epilepsy, severe cognitive impairment, hyperkinesia common
	SLC25A22	Atrophy, dysgenesis of CC, CBLH, EIEE, severe cognitive impairment, abnormal results of electroretinogram
	SPTAN1	Cortical atrophy and CBLA, hypomyelination, reduced, WM, thin CC
	STXBP1	Atrophy, abnormal WM, EIEE, severe cognitive impairment
X-linked lissencephaly with abnormal genitalia	ARX (MIM 300215)	LIS, neonatal seizures, LGS, severe cognitive impairment, genital hypoplasia
Isolated lissencephaly	DCX (MIM 300067)	LIS, LGS, severe cognitive impairment
	PAFAH1B1 (MIM 607432)	LIS, LGS, severe cognitive impairment
Tuberous sclerosis complex	TSC1 (MIM 191100)	Tubers, Variable LGS, variable cognitive impairment, autistic features common
	TSC2 (MIM 613254)	Tubers, Variable LGS, variable cognitive impairment, autistic features common
	TUBA1A (MIM 611603)	LIS, CBLH, LGS, severe cognitive impairment
Deletion 1p36 syndrome	Deletion in 1p36 (MIM 607872)	Cortical atrophy, abnormal WM, thin CC, EIEE, LGS, moderate to severe cognitive impairment
Williams syndrome plus	Deletion in 7q11.23 (MIM 194050)	Mild to moderate cognitive impairment, characteristic neurobehavioral profile
Pallister-Killian Syndrome	Tetrasomy 12p (MIM 601803)	Cortical atrophy, delayed myelination, severe cognitive impairment
Duplication 15q syndrome	Maternal duplication 15q11q13	Variable LGS and autistic features
Miller-Dieker syndrome	Deletion in 17p13	LIS, mild CBVH, LGS common, severe cognitive impairment

Table 1.3: **Syndromes Associated with Infantile Spasm and Agenesis of the Corpus Callosum.** This table was adapted from Paciorkowski et al. (2011) to show syndromes considered in the differential diagnosis of infantile spasms and agenesis of the corpus callosum of Aicardi Syndrome (Rosser et al., 2002).

[2] Only syndromes with known implicated genes or specific genotypes are included in this table.

Abbreviations: basal ganglia, BG; cerebellar atrophy, CA; diffuse cerebellar hypoplasia, CBLH; corpus callosum, CC; early infantile epileptic encephalopathy, EIEE; Lennox-Gastaut syndrome, LGS; lissencephaly, LIS; white matter, WM.

The increased risk of mortality is attributed to neurological impairments and related defects (Menezes et al., 1994). More than 91% of AIC patients develop severe intellectual disability and intractable epilepsy (Donnenfeld et al., 1989, Rosser et al., 2001). In a study of 77 patients, approximately 80% and 95% of patients were unable to walk and talk respectively (Rosser et al., 2001). Furthermore, 67% of patients had seizures daily, which were refractory to anti-epileptic drugs. These neurological impairments can lead to reduced immobility, gastrointestinal disturbances and respiratory problems (**Table 1.4**) (Glasmacher et al., 2007, Rosser et al., 2001). In a study of 69 patients, approximately 94% were affected by gastrointestinal disturbances; most of these health complications occurred after seizures. In study of 18 patients, more than 72% experienced recurrent pneumonia that was correlated with seizures (Donnenfeld et al., 1989).

Skeletal and visual defects are also common in AIC; the former is the second most common complication following neurological impairments (Aicardi et al., 1969). Individuals with costal and/or vertebral defects are prone to developing scoliosis after 12 months of age (Chevrie and Aicardi, 1986). Scoliosis and kyphosis were reported in 55% and 18% of affected children respectively in study of 69 AIC patients (Glasmacher et al., 2007). The average growth (height and weight) of AIC children are relatively similar to the general population; up to the ages of seven and ten years. After this age, the average growth of AIC patients is lower; partially attributed to scoliosis. Visual impairment in AIC patients is usually associated with optic nerve abnormalities and/or microphthalmia (Aicardi, 2005). More than a quarter of AIC children are reported as legally blind, while half of all cases are blind in one eye (Glasmacher et al., 2007).

Type of Outcome	Common Complications[3]	Rare Complications	References
Cardiovascular		• Mitral valve prolapsed • Benign heart murmur	Glasmacher et al., 2007
Gastrointestinal	• Recurrent constipation • Reflux • Abdominal pain • Diarrhoea • Occasional/abnormal liver valve	• Hiatal hernias • Bowel obstructions	Glasmacher et al., 2007
Ocular	• Legally blind • Blind in one eye • Near or far-sightedness • Nystagmus	• Cataracts	Glasmacher et al., 2007
Developmental	• Intellectual disability • Speech impairment • Walking difficulties		Chevrie and Aicardi, 1986 Modgil et al., 2014
Respiratory	• Chest congestion • Upper airway infections ○ pneumonia • Aspiration problems • Required supplementary oxygen • Apnoea		Glasmacher et al., 2007
Other		• Urinary reflux[a] • Kidney stones[a] • Precocious Puberty	Glasmacher et al., 2007

Table 1.4: Prognosis for Aicardi Syndrome. This table summarises the medical complications experienced by Aicardi patients.

[3] Common complications are those reported in 20% or more of patients in the respective study cohort. Rare complications are those that occur in one to three patients per study cohort.
[a]attribute of anti-epileptic drug, rather than inherent trait of Aicardi Syndrome.

1.4 Treatment

Treatment for AIC is personalized due to the clinical heterogeneity of symptoms and outcomes (Hoyt et al., 1978). Seizures and developmental delay/intellectual disability are the central symptoms targeted for treatment. Seizures are managed through individual or combined treatments of anti-epileptic drugs (AEDs), vagus nerve (VN) stimulation or implantation and/or palliative epilepsy surgery (Figure ***1.6***). The latter two treatments have varying success in managing seizures. Resection of dysplastic brain areas either worsened or alleviated the seizures in some cases; same success rate was observed in VN interventions. No AEDs are universally effective in treating seizures and most AIC-related epilepsy are medically refractory (**Table 1.5**).

Treatment[4]	**Donnenfeld et al., 1989**			**Menezes et al., 1994**			**Rosser et al., 2002**		**Kasabeh et al., 2014**		
	N=18	**IP %**	**IE %**	**N=14**	**IP %**	**IE %**	**N=71**	**IP %**	**N=4**	**IP %**	**IE %**
Barbiturates	17	35	65	8/14	25	75					
ACTH	14	64	36	6/14	50	50					
Valproic acid	12	58	42				32/71				
Phenytoin	11	45	55								
Carbamazepine	9	44	56								
Clonazepam	8	25	75								
Diazepam	2	0	100								
Ethosuximide	2	100	0								
Mephobarbital	1	100	0								
Primidone	1	100	0								
Methsuximide	1	100	0								
Acetozolamide	1	100	0								
Benzodiazepines				10/14	60	40					
Topiramate							20/71				
Corpus Callosotomy							1/71	100	2/4	50	50
Vagus Nerve Stimulation							5/71	40	3/4	25	75
Ketogenic Diet							3/71	30			
Not specified				12/14	36	64					

Table 1.5: Anti-epileptic Treatments. This table summarises the anti-epileptic drugs or surgical interventions used to treat seizures in AIC patients Donnenfeld et al., 1989.

Early treatment of seizures is necessary to prevent further developmental delay and systemic effects; the latter include gastrointestinal disturbances and respiratory problems. To improve the development outcomes in AIC, a number of different therapeutic interventions are needed including: physical, occupational, speech and vision therapy (Sutton and Van den Veyver, 1993). In addition, monitoring of gastrointestinal and skeletal (scoliosis) complications is needed. A systematic plan for the management of AIC symptoms has yet to be established, partly due to the unknown aetiology of the disorder.

1.5 Pathogenesis of Aicardi Syndrome

1.5.1 Prenatal or Intrauterine Disturbances

MCD can be caused by prenatal and intrauterine exposure to toxins, infections or other perturbations (Bertoni et al., 1979). However, infectious origins of AIC were not identified in serial studies or individual case reports, which tested for: toxoplasmosis, rubella, cytomegalovirus, syphilis and others (Donnenfeld et al., 1989, Yamagata et al., 1990, Constad et al., 1985). Furthermore, congenital infections can be differentially diagnosed based on the ophthalmologic and neurologic findings (Figure ***1.3***) (O'Driscoll et al., 2010). Maternal and paternal mean age at conception usually ranges from 26-28 and 30-32 respectively; which is not remarkable in the aetiology of AIC (Donnenfeld et al., 1989, Costa et al., 1997). The proposition of environmental factors and prenatal/intrauterine disturbances as a causative factor is further contradicted by the existence of twins discordant for AIC (Hattori et al., 1984, Taggard and Menezes, 2000, Grayton et al., 2012, Gilissen et al., 2014).

[4]The improvement and ineffectiveness of these treatments are not independent; two or more drugs were administered per individual. The number of each individual per study is given by '**N**'. Improvement (**IP**) and ineffectiveness (**IE**) of each treatment is given as a percentage.

1.5.2 Genetic Predisposition

Genetic Findings

In 1982, Ropers discovered a balanced translocation, 46,X,t(X;3)(p22;q12), in a female patient with an early diagnosis of AIC (Burch-Smith et al., 2012). A second female infant with an Xp22 deletion and partial 3p trisomy was reported by Donnenfeld et al. (1989). Although the diagnosis of this case was disputable based on atypical appearances of the chorioretinal lacunae, early investigations still targeted Xp22.2, which was associated with syndromes listed in **Table 1.6**. These syndromes have various craniofacial, neurological and developmental defects overlapping with AIC (Van den Veyver, 2002). OFD1 and MLS (**Table 1.6**) are X-linked dominant and male lethal disorders as is suspected for AIC. Using southern blotting, Nielsen and Neidich used probes (5 and 8 DNA markers respectively) spanning the Xp22.3 locus to look for microdeletions in AIC patients, but no mutations were found (Neidich et al., 1990, Nielsen et al., 1991). Although recognised as the Aicardi locus, the Xp22 locus has yet to be confirmed, as no other mutations have been found in subsequent cases other than the two atypical cases reported by Ropers et al. (1982) and Donnenfeld et al. (1989).

There are four target genes that have been investigated in AIC patients so far: *CDKL5*, *FLNA*, *OCEL1* and *TEAD1* (Anderson et al., 2009, Pai et al., 2013, Schrauwen et al., 2015). In a male patient, *FLNA* was sequenced due to its association with periventricular heterotopia and patent ductus arteriosus that was presented alongside the classical triad (**Figure 1.1**) (Anderson et al., 2009, Clapham et al., 2012). In another unrelated female patient, *CDKL5* was sequenced based on its association with infantile spasms and early epileptic seizures (Pai et al., 2013, Tao et al., 2004). Targeted gene sequencing for mutations in *CDKL5* and *FLNA* showed negative findings. In a recent WES study, *de novo* variants were found in *TEAD1* and *OCEL1* in two unrelated individuals (Schrauwen et al., 2015). However, subsequent WES studies in unrelated AIC cohorts

found no recurring mutations in either of these genes (Wong et al., 2017, Lund et al., 2015). To date, no recurring mutations have been discovered in any of the genes suspected to be involved in or implicated with AIC. This suggests that future sequencing studies involving AIC cohorts should utilise a disease gene discovery strategy with the expectation of genetic heterogeneity rather than a specific X-linked male-lethal cause.

Syndrome	MIM Number[5]	Gene/Locus	Phenotype
Multiple congenital anomalies-hypotonia-seizures syndrome 2	300868	PIGA PNH1 MCAHS2	Dysmorphic features, neonatal hypotonia, myoclonic seizures, variable congenital anomalies involving central nervous system, cardiac and urinary systems, early infantile epileptic encephalopathy
Opitz syndrome type 1	300000	MID1	Hypertelorism, hypospadias, cleft lip/palate, laryngotracheoesophageal abnormalities, imperforate anus, developmental delay, cardiac defects
Pettigrew syndrome (MRXS5)	304340	AP1S2 MRX59 MRXSF MRXS21 MRXS5 PGS	Choreoathetosis, hydrocephalus, Dandy-Walker malformation, seizures, iron or calcium deposition in the brain
Orofaciodigital syndrome I (OFD1) **Simpson-Golabi-Behmel syndrome**	311200 300209	OFD1	Malformations of face, oral cavity and digits; thickened alveolar ridges, abnormal dentition, absent lateral incisors, polycystic kidney disease
Microphthalmia with linear skin defects syndrome (MLS)	309801	HCCS MCOPS7	Unilateral or bilateral microphthalmia, linear skin defects limited to the face and neck, skin aplasia and hyperpigmentation
Joubert Syndrome 10	300804	OFD1 CXorf5 SGBS2 JBTS10 RP23	Specific hindbrain formation, hypotonia, cerebellar ataxia, dysregulated breathing patterns, developmental delay, ciliary dysfunction
Retinitis Pigmentosa 23	300424	RP23	Progressive retinal degeneration, dychromatopsia, decreased visual acuity, dark pigment clumps in midperiphery and perivenous areas, pigmented vitreous cells, mild hearing loss

Table 1.6: Xp22.2 Syndromes. This table summarises the syndromes associated with the Xp22.2 locus.

[5] Phenotype MIM number, associated genes on the Xp22 locus and clinical features related to each syndrome are given.

There are few published studies with reported genetic abnormalities in AIC patients (**Table 1.7**). A majority of the early studies reported large chromosomal abnormalities, which have yet to be defined as causative or incidental. Furthermore, the diagnosis of AIC in these females range from undisputed to disputed based on the revised diagnostic criteria (Aicardi, 2005). Nonetheless, it can be observed that genetic findings in AIC patients reported thus far are localized across multiple chromosomes. Despite the genetic and clinical heterogeneity seen in **Table 1.8**, the concept of AIC being a polygenic disorder has yet to be considered; particularly for early genetic studies with an ascertainment bias towards the X-linked male lethal hypothesis.

Reference	Genetic Abnormality	Typical AIC traits	Atypical AIC traits	Diagnosis[6]
Ropers et al., 1982	t(X;3)(p22.2;q12)	Scoliosis Rib anomalies Reduced muscle tone Bilateral ptosis Microphthalmia Severe developmental delay	Symblepharon Corneal ulceration Lagophthalmos Clods of retinal pigment	Disputable
Yamagata et al., 1990	t(12;21)(q13.3;q11.2)	Chorioretinal lacunae Coloboma Agenesis of the corpus callosum Epileptic encephalopathy Hypsarrhythmia		Undisputed
Donnenfeld et al., 1990	t(X;3)(p22.3;p23)	Microphthalmia Ventriculomegaly	Sclerocornea Posterior hair whorl Teratoma/lipoma in optic chiasm Sharply demarcated chorioretinal defects No agenesis of the corpus callosum , costovertebral anomalies, infantile spasms or intracranial heterotopias	Disputable
Bursztejn et al., 2009	del(1)(p36.22)	Infantile spasm Bilateral papillary coloboma Agenesis of the corpus callosum Ventricular dilatation delayed psychomotor milestone	Brachydactyly Hypertrichosis Deep set eyes Posterior rotated ears	Phenocopy
Prontera et al., 2013	del(6)(q27) & dup(12)(q24.32q24.33)	Neurodevelopmental delay Epileptic seizures Partial agenesis of the corpus callosum Colpocephaly Ventriculomegaly		Undisputed

[6] The diagnosis and genetic origins of these females are highly heterogenous. The diagnoses of these females are given: disputable cases have associated atypical features, undisputed cases have an accepted diagnosis for AIC.

		Gross cerebral asymmetry Other cortical malformations Coloboma Chorioretinal lacunae Scoliosis Craniofacial abnormalities		
Broomall et al., 2013	del(3)(q21.3q22.1)	Preaxial polydactyly Muscular hypotonia Abnormal corpus callosum Chorioretinal lacunae Partial agenesis of the corpus callosum Ventricular dilatation Mild cortical thickening Cavum septum pellucidum	Retromicrognathia Atrial-septal defect Mild pulmonary valvular stenosis	Disputable
Spennato et al., 2013	45,X0/46,XX	Ventricular dilatation Intraventricular cysts Muscular hypotonia Optic nerve coloboma Agenesis of the corpus callosum Intracranial cysts Multiform epileptic seizures	Short neck Pterygium colli Teletelia Barrel-shaped thorax Lymphedema Hemispondylia Patent foramen ovale Minimal persistent arterial duct Chorioretinal atrophy Hydrocephalus Intracranial hypertension Deafness	Aicardi and Turner Mosaic
Schrauwen et al., 2015	*TEAD1* (Chr11: 12904591G>A; NM_021961.5: c.618G>A; NP_068780.2: p.Trp206Ter)	Chorioretinal lacunae infantile spasm cerebellar cysts periventricular heterotopia		Undisputed
Schrauwen et al., 2015	*OCEL1* (Chr19:17338695G>A; NM_024578.1: c.499G>A; NP_078854.1: p.Ala167Thr)	Partial agenesis of the corpus callosum Chorioretinal lacunae Infantile spasm Postfossa arachnoid cyst		Undisputed
Lund et al., 2016	*NPAS2* (Chr2:101612034C>T; NM_002518: c.2465C>T; NP_002509: p.Pro822Leu)	Chorioretinal lacunae Agenesis of the corpus callosum Infantile spasm		Undisputed

Table 1.7: Genetic Findings in Affected Females. This table summarises individuals who had a reported genetic abnormality in literature to date. SNV are described using HGVS nomenclature and SV are described using ISCN.

Investigation Approaches

Copy number variant (CNV) screening was performed in a cohort study of 18 and 38 girls across the whole chromosome X (ChrX) and genome respectively (Yilmaz et al., 2007, Wang et al.,

2009). CNV screening across ChrX was completed using full coverage array comparative genomic hybridisation (aCGH) with an 82kb resolution and 1875 ChrX clones (Yilmaz et al., 2007). CNV screening across the whole genome was conducted on Agilent 244K DNA 60-mer arrays (Wang et al., 2009). CNV contribute to the pathogenesis of multiple neurodevelopmental disorders including intellectual disability, epilepsy and other neuropsychiatric disorders (Spennato et al., 2013). However, no likely pathogenic CNV were identified in the cohorts of either Yilmaz et al. (2007) or Wang et al. (2009).

The low diagnostic yield for CNV screening in AIC cohort was unexpected as genomic microarrays have contributed to the genetic diagnosis of at least 12% of patients with severe intellectual disability (n=1489), which is the most frequent comorbidity of AIC (**Figure 1.3**) (Gilissen et al. 2014). The low yield may be attributed to sample size: 38 girls in Wang et al. (2009) versus 1489 in Gilissen et al. (2014). Alternatively, it may be due to insufficient resolution of a-CGH used in Yilmaz et al. (2007) and Wang et al. (2009), which could miss the detection of smaller CNV. As shown in **Figure 1.3**, whole exome sequencing (WES) is able to improve the detection of mutations in patients with conclusive genetic diagnosis; among these are eight causative CNV that evaded detection by Sanger sequencing and genomic microarrays. It was further estimated by Gilissen et al. (2014) that whole genome sequencing (WGS) can achieve a conclusive diagnostic yield of 62, of which 21% will be represented by CNV, in a comparably large cohort.

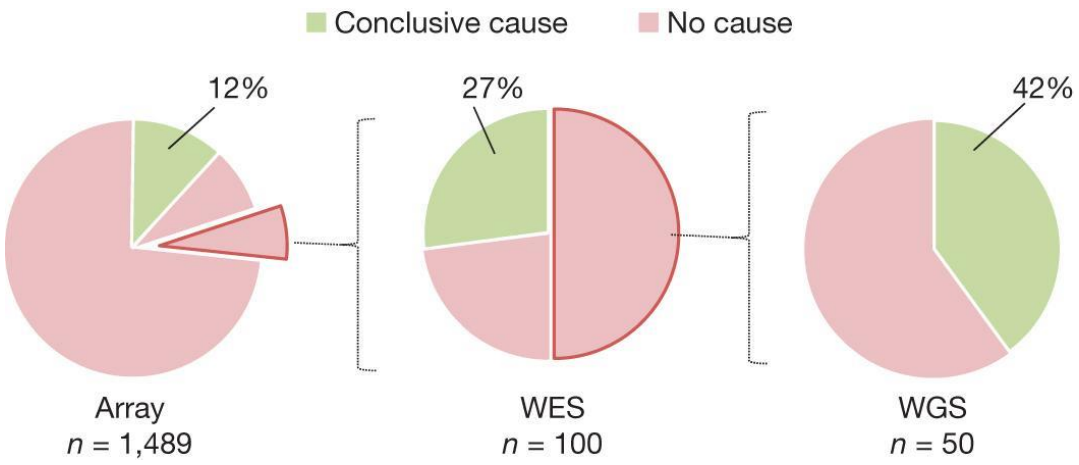


Figure 1.3: Diagnostic Yield versus Technology. This figure obtained from Gilissen et al. (2014) shows the differences in resolution of different technologies in identifying genetic causes of intellectual disability in large patient cohorts. The pie chart represents the cohort; the total number of individuals included in the study is indicated below the pie chart.

WES and WGS are becoming the first-line of approaches for investigating candidate genes in difficult to resolve Mendelian diseases (Majewski et al., 2012). WES in particular has been highly effective at identifying causal mutations (mainly *de novo*) underlying epileptic encephalopathies (EE) (Epi4K Consortium et al., 2013, Veeramah et al., 2013). One of the advantages of WES and WGS is the ability to explore different inheritance models as outlined in **Figure 1.4** (Gilissen et al., 2012). **Figure 1.4** further highlights the power of using parent-proband trios for the identification of candidate disease genes. Only three studies have utilised WES and/or WGS to find candidate AIC genes (Schrauwen et al., 2015, Lund et al., 2016, Wong et al. 2017), however none has reported recurring variants in the same gene.

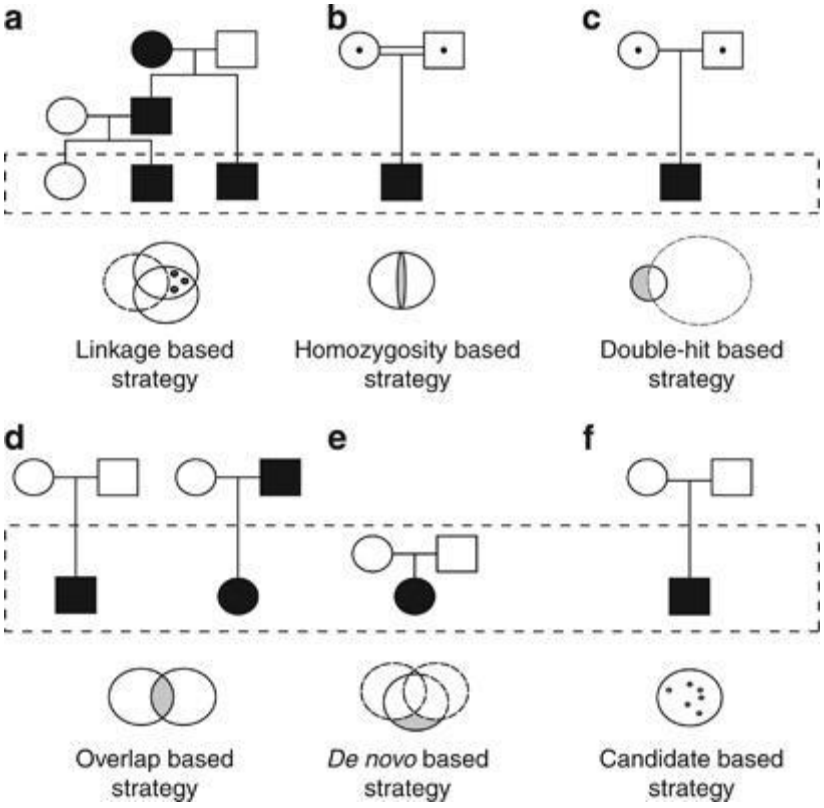


Figure 1.4: Gene Identification Strategies. This diagram extracted from Gilissen et al. (2012) shows different inheritance models that can be studied using whole exome sequencing. Circles and squares represent males and females respectively. Symbols enclosed with a dot represent carriers. Affected individuals are represented by filled symbols, while unaffected symbols represent unaffected individuals. Double lines represent consanguinity.

There is the possibility that causal mutations underlying AIC are not obvious at the genomic level due to a lack of evidence for likely pathogenicity. The utility of matched transcriptome and/or methylome data may increase the resolution of these ambiguous casual mutations. For example, methylome matched with transcriptome data recently revealed the role of brain-specific methylation in the pathogenesis of schizophrenia and bipolar disorder (Xiao et al., 2014). In a different study, methylome and transcriptome data revealed sex differences in the human prefrontal cortex (Xu et al., 2014). The latter could be a contributing factor to the female-biased prevalence of AIC, which remains unexplored in literature.

There is an increasing role of somatic mutations in EE and other neurodevelopmental disorders (Insel et al., 2014, Jamuar et al., 2014). Somatic mutations in *AKT1* and *AKT3* were revealed to predispose the Proteus syndrome and hemi-megalencephaly respectively (Poduri et al., 2012, Yang et al., 2014). Somatic mutations in these genes were not present in peripheral blood; exclusive to affiliated tissues. These studies demonstrate that germinal tissues and whole-brain genomes are less suitable for the detection of somatic (causal) mutations that occur exclusively in the brain (Insel et al., 2014). Investigations into AIC were previously conducted using peripheral blood DNA (Eble et al., 2009, Yilmaz et al., 2007, Hoag et al., 1997, Neidich et al., 1990, Wang et al., 2009). The plausibility of somatic and brain-specific mutations underlying AIC has been postulated, but yet to be explored functionally (Sutton and Van den Veyver, 1993).

Inheritance & Prevalence

There is strong support for an underlying, X-linked, *de novo* dominant and male lethal mutation in AIC (Van den Veyver, 2002). The hypothesis is supported by the almost exclusive female prevalence of AIC and that the only affected males with an undisputed diagnosis of AIC coincidentally have Klinefelter Syndrome (Shetty et al., 2014, Zubairi et al., 2009). Furthermore, the diagnosis of AIC reported in normal 46,XY males have been disputed based on atypical findings (**Table 1.8**) (Chen et al., 2009, Curatolo et al., 1980). However, there are isolated reports of 46,XY males with typical AIC features and 47,XXY males with atypical AIC features (Chen et al., 2009, Chappelow et al., 2008). Despite these observations, there is preferential screening for X-linked pathogenic variants, which is evident from early investigations into the cause of AIC (Wieacker et al., 1985, Neidich et al., 1990, Nielsen et al. 1991, Hoag et al., 1997, Yilmaz et al., 2007, Eble et al., 2009, Anderson et al., 2009, Pai, et al. 2013).

Reference	Karyotype	Classical Triad[7]	Typical AIC Features	Atypical AIC Features
Hopkins et al., 1979	47 XXY	ACC CRL IS	Coloboma Focal seizures Hypotonia Ventriculomegaly Vertebral dysplasia	Narrow palpebral fissure Small right pupil
Curatolo et al., 1980	46 XY	ACC IS	Retinal depigmentation Coloboma Clonic Seizures Hypotonia Cerebellar dysgenesis Cortical heterotopia Vertebral dysplasia	Pseudochoroiditis Tortuous Retinal Vessels
Aggarwal et al., 2002	46 XY	ACC CRL IS	Hypotonia Cortical heterotopia Scoliosis	Lissencephaly Ventricular septal defect
Tateno et al., 2005	46 XY	ACC IS	Coloboma Multifocal Seizure Hypotonia	Lissencephaly
Saddichha et al., 2007	Unknown	ACC IS	Focal Seizure Mild ID Right frontal dysplasia Colpocephaly Kyphoscoliosis	Choroidal calcification
Chappelow et al., 2008	46 XY	ACC CRL IS	Choroidal plexus cyst Colpocephaly Interhemispheric cyst Subependymal heterotopia Bifrontal Polymicrogyria Low muscle tone Craniofacial anomalies Optic Nerve Hypoplasia	
Chen et al., 2009	47 XXY	ACC IS	Generalised tonic-clonic seizure Hypotonia Holoprosencephaly Ventriculomegaly Costovertebral defects Craniofacial anomalies Cleft palate	Salt and pepper pigmentations Lissencephaly Patent ductus arteriosus
Anderson et al., 2009	Unknown	ACC CRL IS	NA	NA
Zubairi et al., 2009	47 XXY	ACC CRL IS	Midline cyst Coloboma Colpocephaly	Lacked vertebral and rib anomalies
Shetty et al., 2014	47 XXY	ACC CRL IS	Hypsarrhythmia Ventriculomegaly Colpocephaly Periventricular heterotopia	No craniofacial abnormalities No skeletal anomalies

Table 1.8: **Males with Aicardi Syndrome.** This table summarises all the males in literature with suspected or confirmed Aicardi Syndrome.

[7] Abbreviations: agenesis of corpus callosum, ACC; chorioretinal lacunae, CRL; infantile spasm, IS; not available, NA.

There are also isolated cases that do not follow the classical X-linked dominant inheritance model, such as the six set of twins described by Chevrie and Aicardi (1986), where only one twin was affected with AIC from each pair. In a 1989 case study, the first familial case of AIC was reported in sisters born five years apart to non-consanguineous parents (Molina et al., 1989). X-linked dominance was contested in these sisters as the parents were phenotypically normal. The twins may be the result of postzygotic mosaicism, while the sisters of germline mosaicism. Although an X-linked cause remains plausible in these isolated cases, the idea of cellular mosaicism has yet to be thoroughly investigated.

The transmission of the disorder will remain ambiguous until the underlying genetic cause is discovered. Progenies from individuals with AIC have not been reported; there are no published reports of transmission from an affected parent to an affected child. It is unknown whether AIC patients have child-bearing potential, considering the mean age of death is estimated at 8.3 years and some girls experienced precocious puberty (Sutton and Van den Veyver, 1993). There are however reports of females with milder forms of AIC, which may be capable of conceiving (Grosso et al., 2007). It is necessary to establish a registry for AIC or AIC-like patients to track for milder forms of the disorder and future offspring.

Hypotheses Underlying Clinical Heterogeneity

X-Inactivation (XIA) was also studied as a potential mechanism underlying the clinical heterogeneity of AIC. Skewed XIA has also been associated with various developmental disorders including Rett syndrome (MIM 312750; *MECP2*), alpha-thalassemia with mental retardation (MIM 301040; *ATRX*), incontinentia pigmenti (MIM 308300; *IKBKG*) and X-linked mental retardation with epilepsy (MIM 300419; *ARX*) (Fruhman et al., 2012). XIA studies in cohorts of AIC patients to date have yielded contradictory results; despite using the same

methodology (Androgen Receptor Assay) and source of DNA (peripheral blood). Hoag et al. (1997) reported XIA is random in affected females, correlating with XIA findings in a 47,XXY male described by Chen et al. (2009).

These results can be interpreted as firstly, XIA is not a mechanism involved in AIC. Secondly, the gene responsible for AIC could be escaping XIA. Thirdly, low expression of the disease allele on chromosome X is sufficient in causing disease; therefore, skewing is not necessary to exert its full disease-potential. Fourthly, in the predicament that mutations underlying AIC occurred somatically, XIA may occur in clonal patterns.

Conversely, Neidich and Eble reported non-random XIA in their cohort of seven and 33 patients respectively (Eble et al., 2009, Neidich et al., 1990). Although these findings support the X-linked hypothesis, it remains controversial whether XIA has a role in the pathogenesis of AIC based on these cohort studies alone. Furthermore, it is unknown whether XIA in lymphocytes is consistent with affiliated tissues in AIC; the brain and eye. Nonetheless, the role can be further elucidated in a large-scale cohort study or upon the identification of a candidate gene/locus. Other mechanisms that could contribute to clinical heterogeneity include pleiotropic genes and genetic heterogeneity, which has yet to be explored.

1.6 Hypothesis & Aims

The diagnosis for AIC thus far has been heavily reliant on accurate clinical diagnosis, which is complicated by the penetrance and expressivity of the major features included in the revised diagnostic criteria (Aicardi, 2005). The pathogenesis of AIC remains largely unknown due to the limited knowledge in the following areas. Firstly, the inheritance of AIC from phenotypically normal parents has been primarily focused on the X chromosome; although an X-linked cause has yet to be identified, other modes of inheritance has been less thoroughly investigated. Secondly, the genetic or molecular mechanisms that contribute to an almost female exclusive prevalence of AIC has yet to be explored. Thirdly, factors contributing to the phenotypic variability among AIC patients are not well understood; those presenting AIC and non-AIC characteristics. To address these disease aspects, we implemented a combined genetic and molecular approach to test the hypothesis with three research aims as outlined below.

1.6.1 Hypothesis

Based on the clinical heterogeneity of reported AIC patients and the chromosomal abnormalities reported thus far, we hypothesized that causes underlying Aicardi Syndrome are genetically heterogeneous.

1.6.2 Research Aims

The overall aim of the project is to identify genetic causes of AIC under the hypothesis previously stated. This will be conducted in the following systematic workflow:

Aim 1: Develop a bioinformatics pipeline that will enrich for causal variants that underlie eye and cortical development. This pipeline will be developed for whole exomes and genomes of Aicardi proband-parent trios.

Aim 2: Prioritise candidate genes by assessing the likely pathogenicity of causal variants using *in silico* and *in vitro* analyses.

Aim 3: Investigate the biological role of causal variants in the pathogenesis of Aicardi Syndrome using *in vivo* models.

The genetics pipeline from Aims 1 and 2 will be first tested on a pedigree that has three generations affected by abnormal eye and brain development (**Chapter 3**). This will test the efficacy of the pipeline as the precedent for assessing Aicardi exomes (**Chapter 4**).

1.7 Expected Outcomes

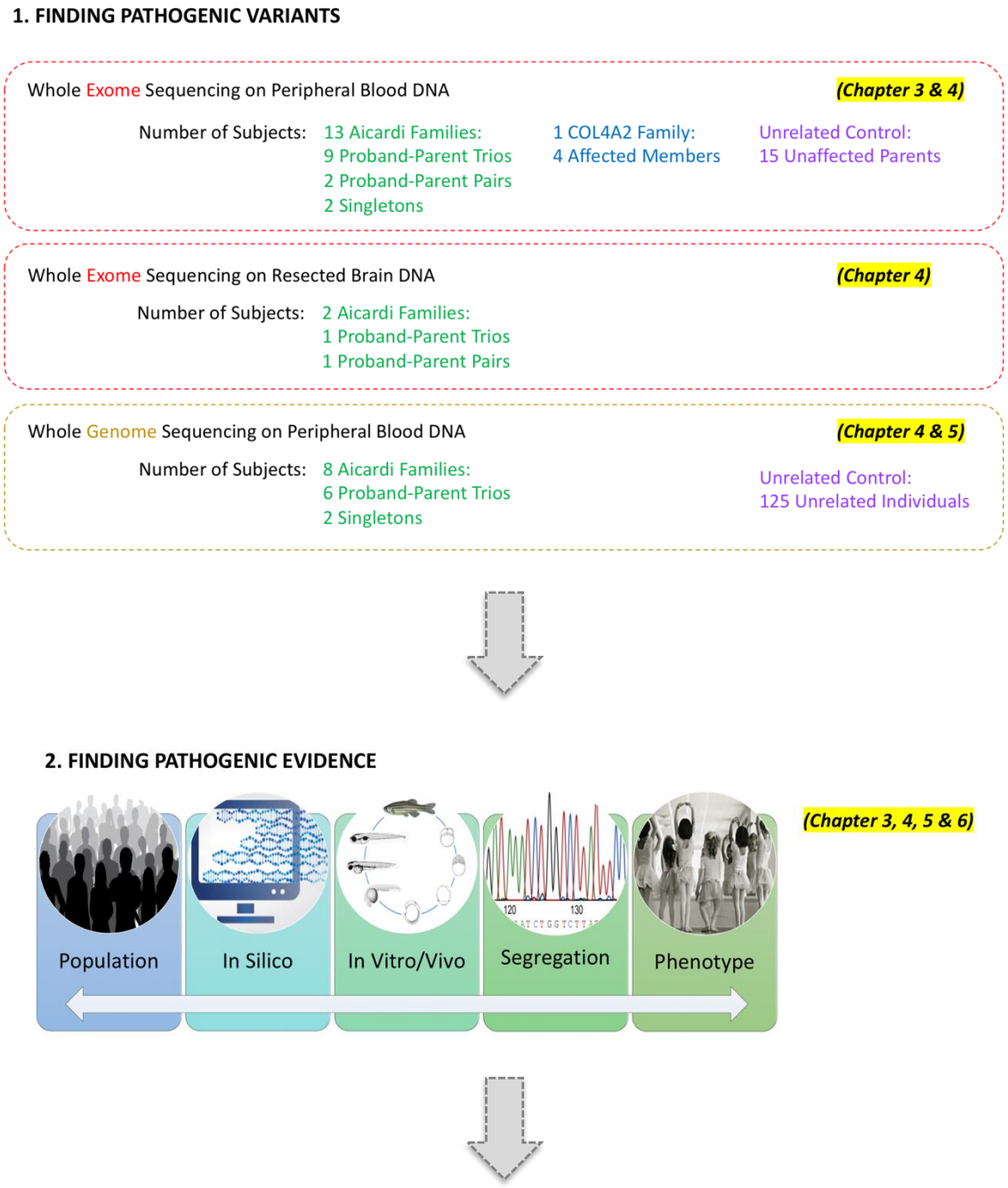
Based on larger sequencing projects involving cohort studies with neurodevelopmental disorders, the following outcomes were expected:

1. Approximately 70% of the cohort will have a likely cause in a protein-coding sequence (Gilissen et al., 2014).
    - This will be examined using whole exome and genome sequencing from all probands in our cohort.
2. *De novo* mutations will account for 15% of yet to be resolved Aicardi cases (EpiK Consortium et al., 2013).
    - This will be investigated using nine parent-proband trio from our cohort and whole exome/genome sequencing.
3. A subset of *de novo* mutations will occur somatically (Jamuar et al., 2014).
    - This will be explored using two paired blood-brain samples from our cohort and whole exome sequencing.

CHAPTER II: Materials & Methods

2.1 Study Design

This PhD Study has three components as represented in **Figure 2.1**, which was designed to achieve the research aims described in **Section 1.6.2**.



(continues onto next page)

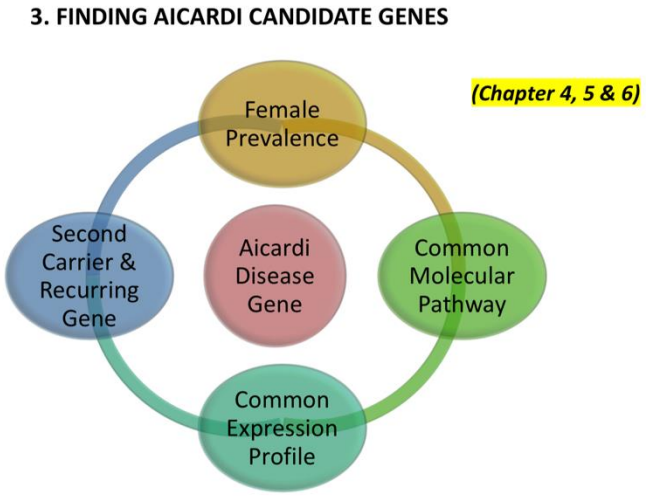


Figure 2.1: Overview of PhD Study. This schematic diagram represents the 3 major components of this PhD study and the respective chapters describing the corresponding results. **[1]** Pathogenic variants identified from WES and WGS using either germline or somatic tissue. **[2]** Finding evidence to support the pathogenicity of variants identified from [1]; in accordance with the American College of Medical Genetics and Genomics, standards and guidelines for the interpretation of sequence variants (Richards et al., 2015). **[3]** Additional evidence that is needed to obtain a definitive genetic diagnosis for Aicardi syndrome is found in the surrounding ovals.

2.1.1 Cohort of Study

All probands recruited in this study had whole exome sequencing (WES). During our first parse, 13 exomes were generated from Aicardi probands. Among the exomes, there were a total of: 9 proband-parent trios, 2 mother-daughter pairs and 2 singletons. Of the 13 probands, 7 were selected for whole genome sequencing (WGS) on the following basis: (1) no candidate genes identified from WES or (2) only variants of unknown significance were identified from WES. Among the genomes, there were 5 proband-parent trios and 2 singletons.

In addition to the Aicardi families, one epileptic encephalopathy family (**Chapter 3**) was included in our PhD study to test our list-based approach for WES analysis. Furthermore, the

family described **in Chapter 3**, who were affected by porencephaly, epilepsy and lens opacities, were the ideal test subjects to screen and enrich for variants involved in eye and brain development. In this family, four affected family members were sent for WES. DNA was available for all four individuals sequenced and one unaffected family member (maternal uncle).

2.1.2 Inheritance-based Strategy

COL4A2 Family

Based on the pedigree of the *COL4A2* family (**Figure 3.1A**), we assumed an autosomal dominant transmission of the disease allele. Using VCFTools 'vcf-contrast' module, we searched for heterozygous variants that were present in the exomes all affected family members and absent from 15 unrelated individuals. Next, we filtered and prioritised for variants that were: (1) harboured in genes within our eye-epilepsy-ID-MCD (EEIM) list and (2) had the highest concordance among multiple prediction algorithms (**Table 3.6**). This list-based approach leads to the discovery of the rare variant in *COL4A2* (**Chapter 3**).

Aicardi Cohort

Based on the pedigree of the Aicardi families (**Figure 4.2**), we assumed three possible inheritance models for the transmission of the disease allele: (1) *de novo* dominant, (2) autosomal recessive or (3) compound heterozygous. We used proband-parent trio analysis (**Figure 1.4**) to filter for variants that complied with the assumed inheritance models. *De novo* variants had to be present as heterozygous (GT: 0/1) in the proband exome, but absent from both parent exomes. Autosomal recessive variants had to be present as homozygous (GT: 1/1) in the proband and present as heterozygous (GT: 0/1) in both unaffected parents. Compound heterozygous variants had to present as heterozygous (GT: 0/1) in the proband and present as heterozygous (GT: 0/1) in one of the unaffected parents, while absent in the other parent. Candidate variants from exomes

and genomes were prioritized by the same list-based and concordant-predictive approach described in the COL4A2 family (**Chapter 3**).

The only difference for trio genomes were the additional investigation of non-coding variants. Non-coding variants were prioritized based on their genomic position in either: (i) evolutionary conserved sequences, (ii) regulatory motifs (iii) sequences absent from control genomes (n=125).

2.1.3 Ethics for human and animals

All sequencing and recruitment to this study was carried out under the Women's and Children's Hospital, Human Research Ethics number REC 2361/3/2014. All zebrafish work in this study was performed under the University of Adelaide, Office of Research Ethics, Compliance and Integrity Research Services approval number S-2017-073.

2.2 Computational Methods

Code and data files used to generate results for **Chapter 4, 5** and **6** are stored in the following GitHub repository: https://github.com/ThuongHa/7.1-Appendix-Data-Code.git (**Appendix 7.1**).

2.2.1 Pre-Processing Raw Reads

Mapping to Reference

WES was performed on a HiSeq2500 (Illumina) by the South Australian Cancer Council Genomics Facility and WGS on the HiSeqXTen (Illumina) by the Australian Genome Research Facility. Raw Illumina reads were processed under the Genome analysis toolkit (GATK) v3.5-0 best practices guidelines (Van der Auwera et al., 2013). First, FASTQ files for WES and WGS generated by Illumina were mapped to the human genome (hg19) using BWA-MEM (Li et al.,

2009). This step generates a SAM file (**Table 2.1**) containing Illumina reads aligned to hg19 reference genome.

Duplicate Marking

Next, duplicate reads introduced by artefacts arising from PCR amplification is marked by Picard (https://broadinstitute.github.io/picard/). This step sorts the hg19-aligned SAM file by coordinate order and marks a duplicated BAM file.

Base Recalibration

Finally, base quality scores are recalibrated using the GATK 'BQSR' function to reduce errors introduced by the sequencing process or experimental preparation. This step generates a recalibrated BAM file with adjusted base quality scores for deletions, insertions and substitutions.

Abbreviations	Definitions	Description
FASTQ		Text-based format that stores biological sequences and its corresponding Phred qualities in a single file.
SAM	Sequence Alignment/Map	Tab-delimited text format that stores biological sequences aligned to a reference sequence.
BAM	Binary Alignment Map	Binary format that stores the same data as SAM files in a compressed binary representation.
BED		Tab-delimited text file that defines a chromosome, start and end position.
VCF	Variant Call Format	Text-based format that stores gene sequence variation and its position in the genome

Table 2.1: Description of common file formats used in PhD study.

2.2.2 Sequencing Coverage

Whole Exome Sequencing

We assessed the sequencing performance of the HiSeq2500 across the target capture plus 50bp flanking sites (SeqCap EZ Exome Enrichment v3.0) using CollectHsMetrics in Picard. We analysed the median target coverage, no base calls and fraction of target bases using the sequencing metrics summarized in **Table 2.2**.

Whole Genome Sequencing

We assessed the sequencing performance of the HiSeqXTen across the whole genome using CollectWgsMetrics and CollectWgsMetricsWithNonZeroCoverage in Picard. Regions with minimum 1bp coverage or no coverage were analysed using the sequencing metrics outlined in **Table 2.2**.

Whole Exome Sequencing	
Field	**Description**
Genome Size	The number of bases in the reference genome used for alignment.
Bait Territory	The number of bases localized to one or more baits.
Target Territory	The unique number of target bases in the target sequence.
On Target Bases	The number of pass-filtered bases mapped to the target region.
Mean Target Coverage	The mean coverage of a target region.
Median Target Coverage	The median coverage of a target region.
Zero CVG Targets PCT	The fraction of targets that did not reach 1bp coverage.
PCT Target Bases	The fraction of all target bases achieved.
Whole Genome Sequencing	
Field	**Description**
Genome Territory	The number of non-N bases in the genome reference[c]
Mean Coverage	The mean coverage in bases of the genome.
SD Coverage	The standard deviation coverage of the genome.
Median Coverage	The median coverage in bases of the genome.
PCT	The fraction of bases attaining sequence coverage post filtering

Table 2.2: Summary of Picard HS metrics used for analysis of WES and WGS sequencing coverage. Descriptions were extracted from Picard HS metrics (https://broadinstitute.github.io/picard/).

Coverage Plots

Coverage data from Picard HS metrics were visualized using the ggplot2 package in a customised R script.

2.2.3 Variant Discovery

Single Nucleotide Variants & Small Insertions and Deletions

Small variants (<50bp) from WES, including single nucleotide variants (SNV) and small insertions and deletions (Indels), were called using GATK HaplotypeCaller, SAMtools Mpileup and DeNovoGear. HaplotypeCaller generated genomic VCF (gVCF) for all WES BAMs encompassing sites with and without variant calls (Van der Auwera et al., 2013). Joint genotyping was then applied to a combined VCF file containing all the WES gVCFs. SAMtools Mpileup was performed on the WES BAMs of just the AIC cohort (n=36) to compute the genotype likelihood across all samples into the VCF format (Li et al., 2009). BCFtools view was then applied to call variants and estimate allele frequencies; the called and genotype variants were then stored in both the VCF and BCF format. BCF files were used as input for DeNovoGear to estimate *de novo* variants in the Aicardi proband; not presents in either unaffected parents (Ramu et al., 2013). Based on the VCF files from HaplotypeCaller and Mpileup, each family/singleton from the AIC cohort was queried using vcf-contrast module in VCFtools. This module pre-filtered variants found in other unrelated individuals; allowing the analysis of recessive and *de novo* variants found in the proband (absent from parents).

For WGS, small variants were called from single-samples using GATK HaplotypeCaller, FreeBayes and Strelka. Similar methods to WES were used to call variants from HaplotypeCaller, except single-sample gVCFs were generated for all WGS BAMs (Van der Auwera et al., 2013). Both FreeBayes and Strelka called variants from WGS BAM files based on a single diploid sample analysis; FreeBayes generated VCF files for SNV and Indels separately, while Strelka generated a single VCF file (Garrison and Marth, 2012, Saunders et al., 2012). A summary of all the SNV and indel callers for WES and WGS is summarized in **Table 2.3**.

	SAMtools	GATK-HC	DenovoGear	Freebayes	Strelka
Language	C	Java	C		C
Model	HMM & MAQ	Bayesian	Likelihood-based error	Bayesian	Empirical variant score
Algorithm	Extended BAQ	Map Reduce	Fragment-based phasing	Haplotype-based	Assembly & Alignment-based haplotyping
Sampling	Single & Multiple	Single & Multiple	Multiple	Single & Multiple	Single & Multiple
Variants	SNP & Indels	SNP & Indels	*De novo* SNP & Indels	SNP & Indels	SNP & Indels
Features	Sorting, indexing, formatting etc.	Realignment, per base recalibration, VQSR			
Version		v3.5.0		v1.1.0	v2.7.1
Reference	Li et al., 2009	Auwera et al., 2013	Ramu et al., 2013	Garrison and Marth, 2012	Saunders et al., 2012

Table 2.3: Summary of SNP and Indels callers. This table was modified from Liu et al. (2013) and summarises the variants callers used to detect SNP and Indels from WES and WGS.

Structural Variants

Larger variants of 1kb or greater were called using Delly2, Lumpy, Manta and Retroseq. The classes of structural variants (SV) that were called included: deletions (DEL), tandem duplications (DUP), insertions (INS), balanced inversions (INV), reciprocal translocations (TRA) and mobile element insertions (MEI). Delly2 operates by using paired-end mapping analysis to calculate read-pair orientation and paired-end insert size; while split-read analysis was used to search for paired-ends where one pair is mapped, while the other is unmapped (Rausch et al., 2012). Different SV classes were then identified based on paired-end evidence (supporting pairs and average mapping quality) and split-read evidence (number of split reads and consensus alignment against reference) (**Figure 2.2**). Delly2 utilised both paired-end and split-read analysis to increase sensitivity and specificity respectively across all SV classes (Rausch et al., 2012).

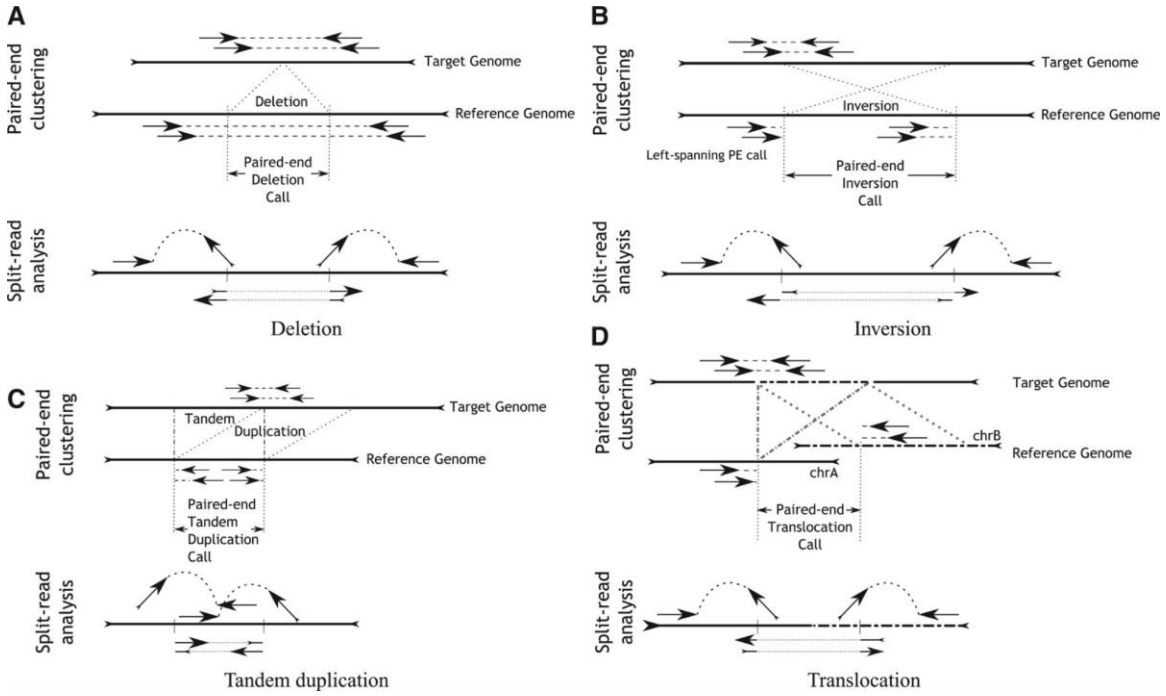


Figure 2.2: Delly2 SV detection based on joint paired-end and split-read analysis.

Lumpy detects candidate breakpoints of genomic SV based on paired-end and split-read alignments; the logistics are similar to Delly2 as described previously (Layer et al., 2014). For DUP and DEL, Lumpy also analyses read-depth alignments, which are based on altered sequence coverage. By integrating all three types of alignment signals, Lumpy computes probability distributions that cover candidate breakpoint regions and then predict the SV subtype based on the multi-signal alignment signatures (**Figure 2.3**) (Layer et al., 2014).

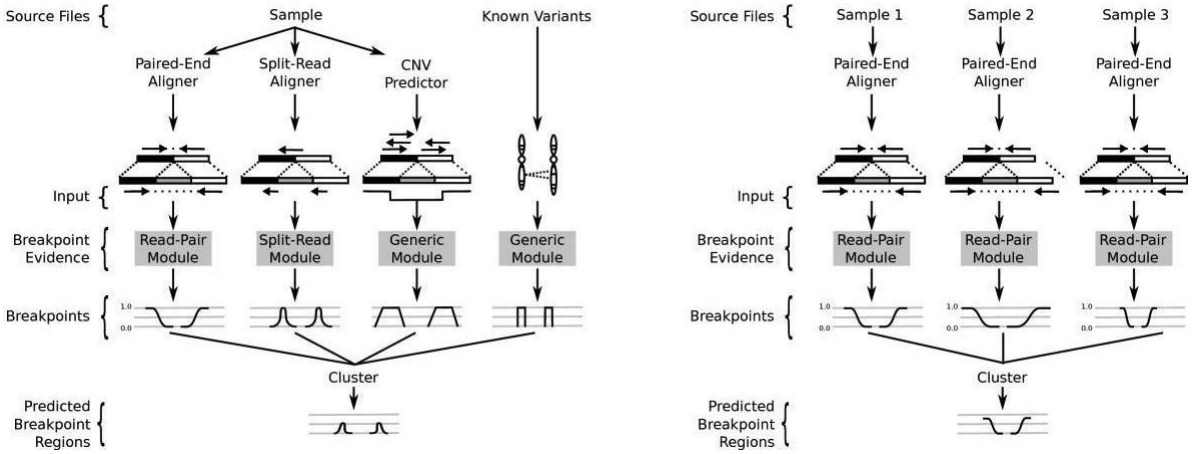


Figure 2.3: Lumpy detection for SV breakpoints based on multiple alignment signals.

Manta's SV detection occurs in three multi-steps: graph construction, local assembly and then variant scoring (Chen et al., 2016). First, a global graph is constructed based on all the breakpoints identified within a genome. Graph edges are then analysed alongside paired-end and split-read evidence to predict candidate SV. Sequencing reads supporting each candidate SV were then assembled locally. Each SV was then scored, genotyped and quality filtered based on a germline diploid model.

Although RetroSeq also utilizes paired-end and split-read analysis, it was solely designed for MEI discovery (Keane et al., 2013). Paired-end and split-read analysis were used to detect discordant paired-ends and single-anchored reads respectively, as described in Delly2. Candidate MEIs were then annotated with a known reference MEI or assigned a MEI class based on a transposable element sequence library. Candidate MEIs were then clustered based on their genomic position and read orientation. The density of the clusters was then used to refine the MEI breakpoints (Keane et al., 2013).

Input files for all SV callers were WGS BAMs, while the output file was in VCF (Lumpy, Manta and Retroseq) or BCF (Delly) format. The sensitivity and specificity for different SV classes were analysed by: (i) running the all the algorithms on the NA12878 reference genome and (ii) comparing to truth sets listed in **Table 2.4** (**Appendix 7.1**).

Truth set Original Name	Bed File Name	Number of Variants	Description	Reference
Complex_1ormoreINDELs_NIST2.19.vcf.gz	complex_indel	184634	Indels within 50bp flanking regions considered 'complex variant regions'	Zook et al., 2014
Complex_noINDELs_NIST2.19.vcf.gz	complex_snp	546712	SNP within 50bp flanking regions considered 'complex variant regions'	Zook et al., 2014
HG001_GRCh37_GIAB_highconf_CG-IllFB-IllGATKHC-Ion-10X-SOLID_CHROM1-X_v.3.3.2_highconf_PGandRTGphasetransfer.vcf.gz	highconf_calls	438100	High-confidence SNP, small Indels, and homozygous reference calls	Zook et al., 2014
NA12878.sorted.vcf.gz	mtsinai_indel	43156	Indels detected by three methods: PBHoney, custom script and Chaisson et al. 2015 (Nature)	Mt. Sinai School of Medicine
Personalis_1000_Genomes_deduplicated_deletions.bed	person_del	2676	Preliminary benchmark deletions from SVclassify	Parikh et al., 2015
Spiral_Genetics_insertions.bed	spiral_ins	68	Preliminary benchmark insertion calls from SVclassify	Parikh et al., 2016
ALL.wgs.mergedSV.v8.20130502.svs.genotypes.vcf.gz		68818	MEI (Alu, Line1,Sva), DEL, DUP, INS, INV, CNV of 2504 individuals from 26 populations	1000 Genomes Project Phase 3
NA12878.wgs.mergedSV.v8.20130502.svs.genotypes.vcf.gz	1000g_sv	3261	SV extracted from 1000g phase 3 project via vcf-subset	1000 Genomes Project Phase 3

Table 2.4: Truth sets with benchmark structural variant calls from published NA12878 datasets.

2.2.4 Annotating Variants

Annovar for SNVs and Indels

SNV and Indels were annotated with five types of gene and function related information by ANNOVAR: (1) gene-based annotations to assign HGNC symbols to variants within or closest to a gene, (2) region-based annotations to identify highly conserved sequences or motifs predicted

to affect gene transcription, (3) filter-based annotations to mark variants previously reported in clinical databases or large sequencing projects, (4) mutation-based annotations to calculate variant effect on protein function or likely contribution to disease and (5) functional-based annotations to associate the candidate gene with known diseases or developmental pathways (Wang et al., 2010). Annotations that were applied to the VCF files of WES and WGS are detailed in Chapter 3 and briefly summarised in **Table 2.5**.

Allele Frequency	Clinical Significance	Gene Annotations	Predicted Pathogenicity
• Unrelated genomes • Public exomes and genomes of healthy individuals	• Gene list associated with: ○ Epilepsy ○ eye development ○ Intellectual disability and ○ malformations of cortical development • Public dataset with reported pathogenic variants	• Hg19 Locus • DNA and amino acid Change • Functional element	• Protein function • Protein structure • Gene Tolerance • Amino Acid Conservation

Table 2.5: Summary of variant annotations for SNV and Indels.

Bedtools for SV and MEI

Annotations for SV and MEI were more limited compared to SNV and Indels, particularly those flanking non-coding regions. Nonetheless, SV and MEI were annotated using the BEDtools 'annotate' function (**Table 2.6 & Appendix 7.1**). The set of benign SV used for filtering were comprised of: the 1000 genomes project phase 3 findings (1000G), SV in healthy controls reported by Coe et al. (2014) (CnvDevDelay), SV from Genomes of the Netherlands consortium (GONL), ISCA curated benign SV (ISCA) and common variants detected from the Deciphering Developmental Delay study (DDD) (Genomes Project, 2015, Genome of the Netherlands, 2014, Riggs et al., 2012, Deciphering Developmental Disorders, 2014). The disease bed files used for annotation were compiled from: top 5% of predicted haploinsufficient genes, known genes causing inherited retinal diseases (Retnet), disease variants from the DDD study, known

Mendelian disease genes (GeneReviews) and SV submitted to ClinVar (Huang et al., 2010, Sun et al., 2015, Deciphering Developmental Disorders, 2015, Adam et al., 2018, Landrum et al., 2016). UCSC Table browser was used to generate a bed file for the canonical transcripts of HGNC gene symbols. The regulatory bed is made of locus provided by the Open Regulatory Annotation database and the Encode project.

Benign.bed	**Disease.bed**	**Gene.bed**	**Regulatory.bed**
1000G CnvDevDelay_control Dgv_merged GONL Isca_benign DDD_common	Huang_hapInsuff RetNet DDD_pathogenic GeneReviews Clinvar	Ucsc hg19	Oreganno wgEncodeDNase wgEncodeTfbs wgEncodeSegment
Description: Variants found in healthy individuals or predicted to be likely benign via curation/functional validation	**Description:** Variants/genes found in individuals with NDD and eye disorders	**Description:** Hg19 position of canonical transcripts with 50 bp flanking 5' and 3' UTR.	**Description:** Functional elements that act at protein or RNA levels and regulatory elements that control gene level activation

Table 2.6: Summary of datasets used for filtering or annotating SV.

2.2.5 Evaluating Variants

Prioritising Likely Causative Variants

Variants were assessed and prioritized in alignment with the standards and guidelines developed by the American College of Medical Genetics and Genomics (Richards et al., 2015). Briefly, evidence for likely pathogenicity was based on: (1) nonsense mutation in a gene associated with a loss of function disease (2) missense mutation predicted to be pathogenic via Polyphen2 and CADD, (3) inherited *de novo* or recessive in the proband, which is confirmed in both parents (4) *in vitro* and/or *in vivo* studies supporting the variants' damaging effect on the gene or protein (5) the prevalence of the variant is rare among a phenotypically normal population.

For variants that lack pathogenic evidence, they were assessed based on the following criteria: (1) resides in known mutational hot spot or functional domain, (2) located in evolutionary conserved sequences, (3) variant alters protein expression or localization, 4) found in a mutation intolerant gene determined by RVIS and previous publications, (5) in a gene underlying eye or brain development but has yet been linked to disease and (6) strong *in silico* evidence for pathogenicity by multiple algorithms as listed in **Table 3.6**. The prediction algorithms listed in **Table 3.6** were used to prioritise variants by: (1) calculating the number of concordant predictions supporting the pathogenicity and (2) ranking the variants from the number of highest to lowest concordant predictions. The rare variant identified in *COL4A2* ranked the highest among the variants identified and listed in **Chapter 3**.

Filtering Benign Variants

Likely benign variants were pre-filtered using in-house dataset by generating a unique key for each called variant comprised of: starting coordinate, reference allele and alternate allele. Next, common variants were filtered out from each proband call set using variant allele frequency thresholds. For the in-house dataset, thresholds were set to consider two scenarios: (1) recurring *de novo* variants present in all Aicardi individuals and absent in unrelated individuals and (2) recurring recessive variants present in proband-parent trios with potentially one or two carriers in other unrelated individuals. For public databases or sequencing studies of phenotypically normal individuals, different thresholds were set based on the number of individuals per reference dataset (**Table 2.7**). Other lines of evidence that excluded likely benign variants included: (i) predicted to be benign by multiple *in silico* prediction programs and (ii) variants that did not segregate in a *de novo* or recessive manner.

Dataset	Source	Number of individuals	Frequency threshold
In-house exomes	This study	~ 20	0.1
In-house genomes	This study	~ 120	0.01

1000 Genomes	1000 Genomes Project	1096	0.001
UK10K	UK10K Consortium	~ 4000	0.001
ExAC	Lek et al. 2016	~ 65000	0.0001
Wellderly	Scripps Institute	~ 400	0.01
CG69	Complete Genomics	69	0.1

Table 2.7. Summary of variant allele frequency thresholds used to exclude likely benign variants.

2.3 Biological Methods

2.3.1 Cell Culture

Human Embryonic Kidney 293T Cells Maintenance & Freezing

Human embryonic kidney 293T (HEK293T) cells were revived from pellets stored at -80^{o}C by transferring frozen cells to a 10mL yellow cap tube and slowly added 5mL of fresh DMEM (Sigma Aldrich, Cat # D6546) growth media (DMEM, 10% foetal calf serum [FCS], 1% L-glutamine [L-glut; 2mM] and no antibiotics). The tubes were spun at 1200rpm for 5 minutes, then supernatant was removed, and the pellet was resuspended with 5mL DMEM growth media. The resuspended pellet was transferred to a 250mL tissue flask with an additional 10mL DMEM growth media (total volume 15mL) and incubated at 37^{o}C with 5% CO_2.

Once the culture reaches 90% confluency, the cells were split by removing all old growth media and the adhered cells were washed with 5mL 1X PBS (Sigma Aldrich, Cat # D8537). PBS was completely removed before adding 1mL of Trypsin EDTA (0.05%; Gibco, Cat # 25300054) and incubating the cells at room temperature for 1 minute to allow the cells to detach from the flask. 10mL of DMEM growth media was added to the old flask and slowly pipetted 3-5 times to achieve single-cell suspension before transferring 0.5mL (1:20) or 0.75mL (1:15) of cell suspension to a new 250mL flask. The new flask is then topped up to 15mL with DMEM growth media and incubated at 37^{o}C with 5% CO_2. Cells from the old flask can be used for enumeration of viable cells by staining with Trypan blue (Invitrogen, Cat # 15250-061) and counting on a haemocytometer.

HEK293T cell lysates were frozen down by transferring the cell culture into a 10mL yellow cap tube, then centrifuging at 1200rpm for 5 minutes. The supernatant is removed, and the cells are resuspended in PBS and transferred to screw-cap tubes; 1mL of resuspended cells per 1.5mL

screw-cap tube. The screw-cap tubes were then centrifuge at 2000rpm for 5 minutes followed by a quick suction and snap freeze (using either LN2 or dry ice). The pellets were then stored in the -80^{o}C freezer for later use.

HEK 293T Cell Transfection for Protein Expression

HEK293T cells were plated onto 12-well plates at a density of 0.5 x 10^6 cells/well with 2mL of growth media (DMEM, 10% FCS, 1% L-glut, no antibiotics) and incubated at room temperature for 2-4 hours. The plates are then incubated at 37^{o}C overnight with 5% CO_2. Prior to adding the transfection mix, 1mL of growth media is removed per well. Plasmids are transfected into HEK293T cells using Lipofectamine 2000 Transfection Reagent (Invitrogen, Cat # 11668027) and under the manufacturer's recommendation with the following exceptions: final lipofectamine2000 reagent per well is reduced to 2µL per well, (2) Opti-MEM medium per well is increased to 250µL, (3) increasing concentration gradient for plasmid DNA (**Table 2.7**) and (4) cells are incubated with the transfection mix overnight at 37^{o}C overnight with 5% CO_2. Cells are then collected and stored at -80^{o}C for further analysis.

500ng empty vector control	500ng empty vector control	50ng plasmid	50ng plasmid
200ng plasmid	200ng plasmid	500ng plasmid	500ng plasmid
1μg plasmid	1μg plasmid	2μg plasmid	2μg plasmid

Table 2.7: Transfection layout on 12-well plate for Plasmid expression.

Lymphoblastoid Cell Line Transformation

Lymphoblastoid cell lines (LCL) are set up from 1-2-day old blood, which was stored in EDTA tubes and kept at room temperature. Lymphocytes are isolated from blood using Ficoll Paque Plus (GE Healthcare, Cat # 17-1440-02) under the manufacturer's protocol with the following adjustments: (1) blood is diluted 1:1 in neat RPMI 1640 (Sigma Aldrich, Cat # R5886) and (2)

after adding the ficoll to the diluted blood, centrifuge for only 20 minutes. After the wash steps, lymphocyte pellets were resuspended in 1mL Epstein-Barr Virus media; 1-part virus and 1-part RPMI growth media (RPMI, 20% FCS, 1% L-glut (2mM) and benzylpenicillin).

To inhibit T-cell activation, 1mL of cyclosporine A (Sigma Aldrich, Cat # C1832) were added to new RPMI growth media per week for the first two weeks of transformation. Cells were incubated at 37^{o}C with 5% CO_2. After two weeks, healthy cultures were transferred to a 50mL tissue culture flask with 3mL of fresh RPMI growth media.

LCL Maintenance & Freezing

Once in flask, growth media was replaced every 3-4 days, gradually increasing the volume up to 10mL. For freezing pellets, 5mL of healthy cell culture were transferred to a 10mL yellow cap tube, while the remaining culture was replaced with RPMI growth media reduced to 10% FCS; L-glut and antibiotics stayed the same. Cells in the yellow cap tube were spun at 1200rpm for 5 minutes, then the supernatant was removed. Pellets were resuspended in storage media (RPMI 1640, 20% FCS, 1% L-glut, 10% filtered DMSO); 1mL of resuspended pellet per cryogenic vial. After the first frozen vial, the cells were maintained with the RPMI growth media with the 10% FCS; media was changed every 3-4 days. Cells for -80^{o}C storage were frozen down using the same method outlined in ‘Human embryonic kidney 293T cell Maintenance & Freezing’. The details of LCLs established for this study are summarised in **Table 2.8**.

DNA #	PK #	Sex	Status	Passage #	Frozen Date
T17262		F	Affected	2	12 SEP 2016
T25388	100234	F	Unaffected parent	2	19 SEP 2016
T25389	100234	M	Unaffected parent	2	19 SEP 2016
T22101		F	Affected	2	21 OCT 2017
T25483	104124	F	Affected	3	21 OCT 2015
T22420	104124	F	Unaffected parent	3	28 OCT 2015
T22419	104124	M	Unaffected parent	3	28 OCT 2015

Table 2.8 Summary of established LCL for Aicardi Cohort

2.3.2 Genomic DNA Sequencing

Primer Design & Storage

For Sanger sequencing validation of variants, primers were designed using NCBI Primer Blast or Primer3Plus under the guidelines outlined in Dieffenbach et al. (1993). Briefly, oligonucleotides were designed to be: (1) between 18-24bp in length, (2) GC content approximately 50%, (4) annealing temperatures between 56-62^{o}C, (5) amplicon length between 250-750bp and (6) avoided tandem repeats. Primer pairs were pre-checked using UCSC I*n-Silico* PCR to exclude for non-specific amplification i.e. binds to sequences other than the target region. Primers were manufactured by GeneWorks or Sigma Aldrich; and re-suspended to a total concentration of 100uM and stored at 4^{o}C. All primers used in this study are listed in **Table 2.9**.

Name	Sequence	Manufacturer	Usage
ANKRD32_SalIF44107	GAGTCGACCGGGAAGATGGAAGATGGTACC	Sigma	cloning
ANKRD32_mutF44108	TGGCCGGTACTTTCATATATCGTCAGCTCTTCT TCACCTGCATCCTCCTT	Sigma	cloning
ANKRD32_NotIR44110	TTCTAGGCGGCCGCTCATCATGAAAACTCCATG ACT	Sigma	cloning
ANKRD32_mutR44109	AAGGAGGATGCAGGTGAAGAAGAGCTGACGA TATATGAAAGTACCGGCCA	Sigma	cloning
SZT2F44119	ACGTGCACTCGTTCAGCTAT	Sigma	sequencing
SZT2R44120	CATGGGGACGTGGTATGTGT	Sigma	sequencing
HCN1F244142	ACTGACATGCTGACATCTCCAA	Sigma	sequencing
HCN1R344143	AGCATGTTTTTCTTCCAGATATGAT	Sigma	sequencing

DOCK7F44144	CTGAGCACTTTTGCATCAGC	Sigma	sequencing
DOCK7R44145	AGTGGAGAATAGGAAAGGGAAT	Sigma	sequencing
KCNAB1F44146	GCATACAAATTACAGCGAAGCC	Sigma	sequencing
KCNAB1R44147	TCTCTGGAATGCATGGTTCTT	Sigma	sequencing
C5orf42F44148	ACACACTCTGCTAATCTGCTTCA	Sigma	sequencing
C5orf42R44149	TGGGAAAAGAATTGTGCTCAT	Sigma	sequencing
ZNF462F44150	GGTGCTCCAATGTCTCCCAA	Sigma	sequencing
ZNF462R44151	CTCAGAGCTGCTTCGGTTCA	Sigma	sequencing
JMJD6F44152	ACTTTCTTCCAGTTCATCTATACCT	Sigma	sequencing
JMJD6R44153	ATTCTGTAGGTGGTTTGTGATGG	Sigma	sequencing
ELAVL4F44154	GGCTGTCATCTTGGTTGTGGA	Sigma	sequencing
ELAVL4R44155	AGCCATCAGCCTTCTGCTCTT	Sigma	sequencing
ANKRD32KpnXho_F	CACACACACAGGTACCCTCGAGAGATGGAAGA TGGTACTCCAAAGCATAT	Sigma	cloning
ANKRD32_NotIR	CACACACACACAGCGGCCGCCTAGCATCATCA TGAAAACTCCATGACTGAC	Sigma	cloning
ANKRD32_1354F	GTATGCGTTCTTCATGCCCTTCTAGAGA	Sigma	cloning
ANKRD32_1630R	CGTCAGGTGGAGCACCTTGGAGCCAATA	Sigma	cloning
ANKRD32_1175F	CTCAGGCTGTCAGATACAACTG	Sigma	cloning
ANKRD32_1354R	AAGGGCATGAAGAACGCATAC	Sigma	cloning
SZT2_679F	CTGGTCAGTATGATTCGTCAGGG	Sigma	cloning
SZT2_915R	AATTCCACATTGGGCACATGG	Sigma	cloning
OPHN1_INT19F	ACCTTCGCTTTCATCCAGGG	Sigma	sequencing
OPHN1_EX20R	TGTCCTACCCCCACCCATAG	Sigma	sequencing
SP6_F	ATTTAGGTGACACTATAG	Sigma	sequencing
T3_F	GCAATTAACCCTCACTAAAGG	Sigma	sequencing
T7_F	TAATACGACTCACTATAGGG	Sigma	sequencing
ANKRD32_480R	ATTTGCCTTTCCAGCCTCCA	Sigma	cloning
TEAD1_298F	CTTGCCAGAAGGAAATCTCG	Sigma	cloning
TEAD1_1050R	GCCATTCTCAAACCTTGCAT	Sigma	cloning
TEAD1_612mF	CTCAGTCCCTGCCTGACAAGGTCGCTCAATTGG CACAACCAAGCTTCG	Sigma	cloning
TEAD1_659mR	CGAAGCTTGGTTGTGCCAATTGAGCGACCTTGT CAGGCAGGGACTGAG	Sigma	cloning

TEAD1_893F	ACTGCAATATTCAAGATGATGC	Sigma	cloning
WNT4_354F	AGGTGTGGCCTTTGCAGTGACG	Sigma	sequencing
WNT4_806F	ACCTGGTGTACTTGGAGCCTAG	Sigma	sequencing
bactin1_F	CGAGCAGGAGATGGGAACC	Sigma	qPCR
bactin1_R	CAACGGAAACGCTCATTGC	Sigma	qPCR
elfa_F	CTTCTCAGGCTGACTGTGC	Sigma	qPCR
elfa_R	CCGCTAGCATTACCCTCC	Sigma	qPCR
SZT2_1542F	CTCAGTGCCTGAGCATTTCACGCTTCCTGACAG CACCAAG	Sigma	cloning
SZT2_1581R	CTTGGTGCTGTCAGGAAGCGTGAAATGCTCAG GCACTGAG	Sigma	cloning
SZT2_4404F	GCTAGACTCTGCCTCGCTGTCAGACGTAGACAC TGTGAATCC	Sigma	cloning
SZT2_4445R	GGATTCACAGTGTCTACGTCTGACAGCGAGGC AGAGTCTAGC	Sigma	cloning
SZT2_6920F	ACATTAGTATTGTGCAGCTGGAGGAGAAACTC CGAGGAGCAGCTCG	Sigma	cloning
SZT2_6965R	CGAGCTGCTCCTCGGAGTTTCTCCTCCAGCTGC ACAATACTAATGT	Sigma	cloning
SZT2_8961F	GTTCAGCTATGACTTCCATCTGCGCCTCGTGCA TCAGCACGTGC	Sigma	cloning
SZT2_9004R	GCACGTGCTGATGCACGAGGCGCAGATGGAAG TCATAGCTGAAC	Sigma	cloning
KMT2B_F1	GAGGACCTGCCATCGGAAAT	Sigma	sequencing
KMT2B_R1	AAGTGGACACCCTCTTGACG	Sigma	sequencing
TCF4_534F	CTATGACCACATGACCAGCAG	Sigma	cloning
TCF4_698R	ATGTCACCTCCAAGGAGACTCTGCTG	Sigma	cloning
TCF4_1560F	CATCCACAGCTATGCCTGGTGGTCAT	Sigma	cloning
TCF4_1750R	CAGACTGGACAGGAAGCTGTGGAA	Sigma	cloning
TCF4_1961F	ATGACGATGAGGACCTGACAC	Sigma	cloning
TCF4_526mutF	CGTCCAGGAACTATGGAGATAGGACTCCCTAT GACCACATGAC	Sigma	cloning
TCF4_526mutR	GTCATGTGGTCATAGGGAGTCCTATCTCCATAG TTCCTGGACG	Sigma	cloning
TCF4_156XhoIF	TGAGAACTGCTACTCGAGCTTCTCC	Sigma	cloning
TCF4_830BstBIR	TTTCGAACTTTCTTTGTCTGTACC	Sigma	cloning
VANGL1_GA_FWD	CTGCATGTTCACTGCCCTTT	Sigma	sequencing

VANGL1_GA_REV	GAAACACAAACACCCGTGGC	Sigma	sequencing
IFT122_CT_FWD	CAAGGATGGAGTGCGGCTT	Sigma	sequencing
IFT122_CT_REV	TGTCCCTCTTCCCAAGGTCT	Sigma	sequencing
CISD2_CT_FWD	GTTGGGATTACAGGCGTAAACTAC	Sigma	sequencing
CISD2_CT_REV	AGTATTAGTGGACCCACATTATCTCC	Sigma	sequencing
SCAF11_GA_FWD	ACCATCATTTGCCCATCTTC	Sigma	sequencing
SCAF11_GA_REV	ATGCCGTCTTCTGATCTTGC	Sigma	sequencing
LARGE_CT_FWD	TGCTGCCAACTACCTGTGT	Sigma	sequencing
LARGE_CT_REV	TGGGATGAAAGGTTAGCAGGA	Sigma	sequencing
ABCD1_TC_FWD	CTGTCTCCAAATAAGGTCACACTG	Sigma	sequencing
ABCD1_TC_REV	TAGCAAGTGTGTGTGGTATTTCCT	Sigma	sequencing
CLPB_GC_FWD	AACTCTTAGGCAGGGCTGGT	Sigma	sequencing
CLPB_GC_REV	GGCCTTCCTAGCATGTTTCCT	Sigma	sequencing
TTN_GA_FWD	ATAGACTGCCGCATCCTCTG	Sigma	sequencing
TTN_GA_REV	TTAACTCGGCACTCTCAGCA	Sigma	sequencing
CYP51A1_AC_FWD	AGCTCCAAATGGCACATAGG	Sigma	sequencing
CYP51A1_AC_REV	GCTGGCTCACTGGAAATCAT	Sigma	sequencing
CYB561_AG_FWD	TTGGCACACTCAGAAAGCTG	Sigma	sequencing
CYB561_AG_REV	AAGGGCTACGCTGACCTGTA	Sigma	sequencing
RNF34_GA_FWD	TGGAGAGAACCAAAGTGCAG	Sigma	sequencing
RNF34_GA_REV	CTTTCCAGTCCCTTCTGTGC	Sigma	sequencing
NAA25_CG_FWD	CATGGTCAACCAGATGTTGC	Sigma	sequencing
NAA25_CG_REV	TGTGTTCTGATGGAGCTTGC	Sigma	sequencing
CLEC18B_GA_FWD	GAAGGATCAGACCCTCAAAGG	Sigma	sequencing
CLEC18B_GA_REV	AGTGACTTCGAGACCAGGAACTT	Sigma	sequencing
WNT8B_ex3_fwd	CTGGTGCCCAGAGTGGTATT	Sigma	qPCR
WNT8B_ex4_rev	CATGCACAAATGCTGTCTCC	Sigma	qPCR
KMT2B_ex23_fwd	GCCTCAGCAACTTCCACTTC	Sigma	qPCR
KMT2B_ex24_rev	ACAATTTCCTTGCCATCCAG	Sigma	qPCR
SZT2_ex11_fwd	TCAGTGCCTGAGCATTTCAC	Sigma	qPCR
SZT2_ex12_rev	GCTTCCAGTAGGCAGCAAAC	Sigma	qPCR
SLF1_ex19_fwd	GAAACAGCCCTGCATAGAGC	Sigma	qPCR

SLF1_ex20_rev	GTCCAGCCAGCATTGTCTTT	Sigma	qPCR
TEAD1_ex3_fwd	AACTCAGGACAGGCAAGACG	Sigma	qPCR
TEAD1_ex6_rev	AGGGCCTTATCCTTTGCAGT	Sigma	qPCR
TCF4_534F	CTATGACCACATGACCAGCAG	Sigma	cloning
TCF4_698R	ATGTCACCTCCAAGGAGACTCTGCTG	Sigma	cloning
TCF4_1560F	CATCCACAGCTATGCCTGGTGGTCAT	Sigma	cloning
TCF4_1750R	CAGACTGGACAGGAAGCTGTGGAA	Sigma	cloning
TCF4_1961F	ATGACGATGAGGACCTGACAC	Sigma	cloning
TCF4_526mutF	CGTCCAGGAACTATGGAGATAGGACTCCCTAT GACCACATGAC	Sigma	cloning
TCF4_526mutR	GTCATGTGGTCATAGGGAGTCCTATCTCCATAG TTCCTGGACG	Sigma	cloning
TCF4_156XhoIF	TGAGAACTGCTACTCGAGCTTCTCC	Sigma	cloning
TCF4_830BstBIR	TTTCGAACTTTCTTTGTCTGTACC	Sigma	cloning
RNF34F43657	GCCTTAGCTCTCTGTTGCCTT	Geneworks	sequencing
RNF34R43658	CCCACACTCCAGTAGGACAC	Geneworks	sequencing
PRDM16F43659	TTCCCAGCACTCATTCTTCC	Geneworks	sequencing
PRDM16R43660	CCAGCTTCTTCTCCTGCATC	Geneworks	sequencing
PCNXL3F43661	CGTGGACCAGGATTGGAAC	Geneworks	sequencing
PCNXL3R43662	ACACCCGAGCACTCACCT	Geneworks	sequencing
GIT1F43663	GAAGGTGCTTGACAATGGAGA	Geneworks	sequencing
GIT1R43664	CTGACCCTTGCCCTCACC	Geneworks	sequencing
PET117F43665	GTGGGGATGTCTAGGAG	Geneworks	sequencing
PET117R43666	ACGTGACACCGACCTGCT	Geneworks	sequencing
SLC2A4RGF43667	GCCTGAGCAGAGTGATGGT	Geneworks	sequencing
SLC2A4RGR43668	GGGGACACTGGAGTCAGG	Geneworks	sequencing
GAL3ST1F43669	TGATGAAGATGGCGTTGG	Geneworks	sequencing
GAL3ST1R43670	GCCGCAATGACTTCGACTAC	Geneworks	sequencing
HCN1F43671	ATCAGCTCGAACACTGG	Geneworks	sequencing
HCN1R43672	ACTTGTTTTAAGTTGTGGGAAG	Geneworks	sequencing
FAM20CF43673	CCAAGGAGGAGTACAAGCTGA	Geneworks	sequencing
FAM20CR43674	TTCCTCTCCACGCAGTCCC	Geneworks	sequencing
GPSM1F43675	GAGGTGCCAGGGTGGTG	Geneworks	sequencing

GPSM1R43676	TCAGCAGGTCCCAGGTCTC	Geneworks	sequencing
CACNA1BF43677	TTCCCAAGCCTTCCTGTTTTCC	Geneworks	sequencing
CACNA1BR43678	GCAGAGCAGGTGCCTTACC	Geneworks	sequencing
FLNAF43691	CTGAGATGTTGATGGGGATG	Geneworks	sequencing
FLNAR43692	CTGGACGCACACTGATGG	Geneworks	sequencing
WNT8BF43693	CTGGTGCCCAGAGTGGTATT	Geneworks	sequencing
WNT8BR43694	GCTGTGGCTGAGGTTTTTCT	Geneworks	sequencing
G2E3F43695	TGTGACCCCTATTTTCACATACTCT	Geneworks	sequencing
G2E3R43696	GAAGTCCACATGGGAAATGA	Geneworks	sequencing
JMJD6F43697	GACAGTCTCTCCTGGTTTTTGTAAG	Geneworks	sequencing
JMJD6R43698	GAGACGAAGGAGGGAACCAG	Geneworks	sequencing
WNT8B2F43741	GCCCTGAGAGAGCCCCGCAGCTGTCCAGCCA	Geneworks	cloning
WMT8B2R43742	TGGCTGGACAGCTGCGGGGCTCTCTCAGGGC	Geneworks	cloning
USP40F43743	CTGGCATATTTACCTGCCTCTC	Geneworks	sequencing
USP40R43744	GCTTGTTTTGCTGAGTCATTCTT	Geneworks	sequencing
COL4A2F43745	CCCACCCATGTTTCCTTTT	Geneworks	sequencing
COL4A2R43746	GATGGGGCACTCACCTCT	Geneworks	sequencing
ANKRD32F43747	GAGGCAGAAAAATCGCTTGA	Geneworks	sequencing
ANKRD32R43748	GGAGACTTCCAAGGAGGATG	Geneworks	sequencing
WNT8B3F43810	AAGCAGAGCTCTCTGGCTAAC	Geneworks	cloning
WNT8B4R43811	CAGGGCCCGCCCGCGCCTTA	Geneworks	cloning
WNT8BF_mut1_43858	CTGGAACTGCCCTGAGAGAGCCCcGCAGCTcTCgAGCCATGGTGGGCTT	Geneworks	cloning
WNT8BR_mut1_43859	AAGCCCACCATGGCTcGAgAGCTGCgGGGCTCTCTCAGGGCAGTTCCAG	Geneworks	cloning
WNT8BR5_43860	ACTTGCAGTTGCAGCTGGACACGGTCT	Geneworks	cloning
attB_Myc_ANKRD32_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCACCATGGAACAGAAACTCATCTCTGAAGAGGATCTGATGGAAGATGGTACtCCAAAGCATATC	Geneworks	cloning
ANKRD32_attB_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTACTAGCATCATCATGAAAACTCCATGACTG	Geneworks	cloning

Table 2.9: Summary of primers used for Sanger sequencing, cloning and qPCR.

PCR Mix & Conditions

We amplified genomic DNA using: (i) 50ng/µL of template DNA, (ii) in-house recombinant Taq DNA polymerase, (iii) 10X PCR buffer (100mM Tris-HCl, 500mM KCl, 15mM $MgCl_2$, pH 8.3 at 20^{o}C, Roche Cat # 11699121001), (iv) 10mM dNTP (2.5mM dATP, dCTP, dGTP and dTTP), (v) 10uM primer (forward and reverse) and (vi) sterilized water up to the volume of 50µL. PCR reaction mixes were placed in 200µL PCR tubes and amplified in a thermal cycler (Eppendorf Mastercycler) with cycling conditions outlined in **Table 2.10**. For each primer pair, we optimised the annealing temperature using a gradient PCR on genomic DNA of an anonymous blood bank donor. For segregation analysis, we performed the PCR on water as a negative control, blood bank DNA as a positive control and LCL-derived DNA from the proband and their parents (if available).

Step	**PCR with in-house Taq**			**PCR with hi-fidelity Taq**		
	Temperature (^{o}C)	**Time (s)**		**Temperature (^{o}C)**	**Time (s)**	
1. Initial Denaturation	94	120		94	120	
2. Denaturation	94	30	**Steps 2-4 x 32**	94	15	**Steps 2-4 x 35**
3. Annealing	60	30		55	30	
4. Elongation	72	30*		68	60	
5. Final Elongation	72	10min		-	-	

Table 2.10: PCR cycling conditions.

Gel Electrophoresis (GE)

PCR amplicons were visualized on 1-1.5% agarose gel (agarose powder from Scientifix, Cat # 9010B; 1X tris-borate-EDTA buffer, 0.5μg/mL of ethidium bromide). Samples were then

homogenized with 6x Orange DNA loading dye (Thermo Scientific, Cat # R0631) and loaded alongside 50ng/µL of 1 kb plus DNA ladder (Invitrogen, Cat # 10787026). Agarose GE was run at 110mV for 30 minutes per 500bp of DNA fragment. Gels were visualized in a UV transilluminator (INGENIUS Syngene Bio Imaging Capture System) and captured using the GeneSnap Image Acquisition Software (v7.05).

PCR & Gel Purification

For single bands on the agarose gel, PCR samples were purified using either the QIAquick PCR Purification Kit (Qiagen, Cat # 28104) or the GenElute PCR Clean-Up Kit (Sigma Aldrich, Cat # NA1020). PCR purification was performed following the respective manufacturer's protocol with one adjustment to both kits: DNA was eluted in 30-50µL of sterilized water instead of the elution buffer described in both protocols. Final DNA concentrations per purified PCR sample was measured on a UV-Vis Spectrophotometer (NanoDrop ND-1000).

For multiple bands on the agarose gel, PCR samples were purified using either the QIAquick Gel Extraction Kit (Qiagen, Cat #28704) or the GenElute Gel Extraction Kit (Sigma Aldrich, Cat # NA1111). Gel purification was performed following each of the manufacturer's protocol with no modifications. Final DNA concentrations were measured using the NanoDrop.

Sanger Sequencing & Purification

For variant validation from WES, 30-50ng of purified PCR amplicons were sequenced using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Cat # 4337455) in a 10µL sequencing reaction (20-25ng PCR template, 2µL Sequencing buffer (5X), 1µL BigDye Reaction Mix, 1µL primer (3.2pmol) and sterilized water for final volume of 10µL). Sequencing reactions were amplified on the Eppendorf MasterCycler (Eppendorf) following the manufacturer's protocol.

DNA from the sequencing mixes was precipitated using 75% isopropanol and washed with 70% ethanol. Residual ethanol was left to vapourise in a Speed Vac (10 minutes) or at room temperature (1 hour) in a dark place. The samples were then processed for Sanger sequencing by the DNA Analysis Facility (Department of Genetic Medicine, SA Pathology at the Women's and Children's Hospital) using the 370xl Sequencer (Applied Biosystems). The sequencer chromatogram files were viewed in the Seqman program (Lasergene DNAStar software package v.8.0.2).

2.3.3 Plasmid cloning

Plasmid Mini & Midi Purification

Plasmids containing the full-length human ORF for *WNT4* (Addgene, Cat # 35929), *WNT8B* (Addgene, Cat # 43819) and *TEAD1* (Addgene, Cat # 33109) were purchased from Addgene, while *SLF1* (Genscript, Cat # OHu04948D) was manufactured by GenScript (**Figure 2.3**). Addgene plasmids arrived in the form of transformed bacteria in a stab culture. Each stab culture was streaked onto LB agar plates with 100μg/mL of ampicillin and incubated at 37^{o}C overnight. GenScript plasmids arrive in 10μg tubes of plasmid DNA, which was first transformed (**Section 2.2.3** 'Ligation and Transformation') into JM109 and then streaked the same as the addgene plasmids.

For transformations where the negative control (water) contained no colonies or less than 1/1000 of the experimental plate, six isolated colonies on the latter were inoculated separately in 5mL of LB with 5µL of ampicillin (100μg/mL). The mini cultures were incubated overnight on a 37^{o}C shaker and 1mL of the culture was purified using the Wizard Plus SV Miniprep DNA Purification System (Promega, Cat #A1330) following the manufacturer's protocol. 200ng of the

Miniprep DNA was used in a diagnostic digest to verify backbone and insert sizes; for those that aligned with predicted sizes, 300ng was used to verify the identity of the plasmids via Sanger sequencing (**Section 2.2.3** 'Plasmid Verification'). The mini cultures were stored at 4^{o}C for 2-3 weeks until validation was complete.

Once the plasmids were sequence validated, 2mL from two of the validated mini cultures were inoculated separately in 200mL of LB with 200µL of ampicillin (100μg/mL). The midi cultures were incubated overnight on a 37^{o}C shaker and the whole 200mL of culture was purified using the QIAGEN Plasmid Midi Kit (Qiagen, Cat # 12243). The midiprep followed the manufacturer's protocol (January 2011) with the following adjustments: replaced step 12 and 13 on page 2/2 with the DNA pellet resuspended in 1mL 100% ethanol, 0.25µL of sodium acetate and 0.5µL of glycogen. The tube was centrifuged, supernatant was removed, and the pellet washed with 70% ethanol. The previous step was repeated without more ethanol and then the pellet was air dried for 0.5-1 hour at room temperature. The DNA pellet was resuspended in 100µL of sterilised water and the final yield was measured using a Nanodrop.

Generation of Mutant Insert

Primers (Geneworks) were designed to amplify candidate variants identified from our WES study using the donor plasmid sequences. The donor plasmid backbone and insert sequence was obtained from the manufacturers site and transferred onto the Seqbuilder program (Lasergene DNAStar software package v.8.0.2). Unique restriction sites in the multiple cloning site were incorporated into the primers flanking the ORF ends. Extra 6 adenine and thymidine pairs were added to the 5' end of the primers to improve restriction digest efficiency. A second pair of primers were designed to amplify the sequence of interest containing the variants listed in **Table**

2.9; the primer sequence contains the variant and a synonymous restriction site for rapid diagnostic digest.

Two successive rounds of PCR were performed: (i) amplification of the ORF ends with the sequence of interest and (ii) amplification of the full-length ORF containing the variant of interest. To amplify the fragments for the first cloning PCR we used Platinum Taq DNA polymerase (Invitrogen, Cat # 10966018) following the manufacturer's' protocol. PCR products were purified using the gel extraction method described in **Section 2.2.2** 'PCR & Gel Purification'. For the second cloning PCR, we combined the purified PCR products to act as cDNA template to amplify the full-length sequence; also using the same amplification and purification methods as the first cloning PCR.

Restriction digest

Restriction enzymes (New England BioLabs) listed in **Table 2.10** were used to digest the mutant insert and the donor plasmid in a 1.5mL eppendorf tube (1μg of DNA, 1.5µL of restriction enzyme, 3µL 10X NEB buffer and sterilized water to 30µL total volume). The tubes were incubated for 3 hours or overnight at their enzyme-specific temperature as recommended by the manufacturer's (New England BioLabs) protocol. For the enzymes that needed to be heat inactivated, the digests were incubated at 65^{o}C or 80^{o}C for 20 minutes. After the digest, 2µL Calf intestinal alkaline phosphatase (New England BioLabs, Cat # M0290S) is added to the digested vector along with 3µL 10X CutSmart Buffer; then incubated at 37^{o}C for 1 hour.

Restriction Endonuclease	Catalogue Number	Reaction Temperature (ºC)	Target Gene
SalI	R0138S	37	*SLF1*
NotI	R0189S	37	*SLF1*
KpnI	R0142S	37	*SLF1*
XhoI	R0146S	37	*SLF1, TCF4, WNT8B*
BstBI	R0519S	65	*TCF4*
HindIII	R0104S	37	*TEAD1*
EcoRII / BstNI	R0168S	60	*TCF4*
SacI	R0156S	37	*WNT8B*
ApaI	R0114S	25	*WNT8B*
BglII	R0144S	37	*TEAD1*
EcoRI	R0101S	37	*TEAD1*

Table 2.10: Restriction endonucleases used for cloning and diagnostic restriction digest.

Insert & Vector Purification

Both digested vector and insert were running on 1% agarose gel, then excised and gel purified using the methods outlined in **Section 2.2.2** 'PCR & Gel Purification'.

Ligation & Transformation

Ligation reactions were set up in parallel containing either: (1) cut vector alone, (2) cut vector and water, (3) cut vector and insert at a ratio of 1:3 and (4) cut vector and insert at a ratio of 1:10. Each ligation reaction also contains: 1µL of 10X ligation buffer (50mM Tris-HCl, 10mM MgCl2, 1mM ATP, 10mM DTT, pH 7.5), 1µL of T4 DNA ligase (New England BioLabs, Cat # M0202S) and sterilized water to a total volume of 10µL. All ligation reactions were incubated overnight at 4ºC. To check for success of ligation, 2µL of each ligation reaction ran on 1% agarose gel (**Section 2.2.2** 'Gel electrophoresis'); reactions observing a ladder of bands on the gel were selected for transformation.

Ligation reactions were purified by adding 1mL of isopropanol (100%), 0.25µL of sodium acetate (3M) and 0.5µL of glycogen (50ng/µL); then incubated at -80^{o}C for 15 minutes. The reactions were then centrifuged for 10 minutes at maximum speed, supernatant removed, and pellet rewashed in 100% ethanol; this step was then repeated with 70% ethanol. After a final centrifuge to remove residual ethanol, the pellet was resuspended in 10µL of sterilized water.

Three 1.5mL Eppendorf tubes were set up for bacterial transformation: (1) negative water control, (2) positive pUC19 vector control and (3) purified vector and insert ligation mix. 1-5µL of these three components was added to separate vials of JM109 cells (40µL) cultured in-house. The cells were then electroporated using the Gene Pulser MXcell Electroporation System (BioRad) following the manufacturer's protocol. Cells were then recovered in 1mL of SOC medium (Sigma Aldrich, Cat # S1797) in a 37^{o}C shaking incubator for 0.5-1 hour. 50-100µL of the transformation mix was plated onto LB agar plates containing the antibiotic used for selection. The plates were then incubated at 37^{o}C overnight. Mini cultures were prepared and purified following the methods outlined in **Section 2.2.3** 'Plasmid Mini & Midi Purification'.

Plasmid Verification & Glycerol Stock

Plasmids containing the vector and mutant insert are screened using a diagnostic restriction digest with the method and enzymes described in **Section 2.2.3** 'Restriction Digest'. To sequence verify the mutant insert and screen for non-synonymous variants in the ORF, Sanger sequencing was performed using 300ng of Mini prep DNA and the methods described in **Section 2.2.2** 'Sanger Sequencing & Purification'.

Glycerol stocks were prepared from mini cultures of plasmid DNA that were sequence validated. 1.5mL of mini culture was transferred to a 1.5mL screw cap tube and centrifuged at low speed for 2 minutes. Supernatant was aspirated, and the pellets were resuspended in 50% glycerol; then stored at -80^{o}C for future use.

2.3.4 RNA, whole exome & Whole Genome Sequencing

RNA Extraction & Sequencing

Two female control and three female Aicardi probands were selected for RNA sequencing based on the availability of LCLs established using the methods described in **Section 2.2.1** 'LCL Maintenance & Freezing'. RNA was first precipitated from LCL pellets using the Trizol Reagent (Invitrogen, Cat # 115596026) and followed the manufacturer's (ThermoFisher Scientific) protocol. Total RNA was then purified using the RNeasy Mini Kit (Qiagen, Cat # 74104) starting from Step 5 on page 27 of the manufacturer's handbook (RNeasy Mini Handbook 06/2012). Final RNA yield was measured using the nanodrop; while the quality was assessed based on visualising the 28S and 28S rRNA ratio on 1% agarose. 3μg of LCL-derived RNA was sent for RNA sequencing to the Australian Genome Research Facility. Preliminary RNA analysis was commenced, however was not comprehensively analysed and completed before the PhD expiration time.

DNA Extraction & Whole Exome/Genome Sequencing

For the main PhD study, 14 unrelated individuals suspected or diagnosed with Aicardi Syndrome (AIC) were selected for whole exome sequencing (WES). In total, there were: ten proband-parent trios, two proband-mother pairs and two singletons that were sent for WES. For 2/10 trios, there were DNA extracted from resected brain tissues; hence two blood-brain pairs were sent for WES. In a parallel study (**Chapter 3**), three affected individuals were also sent for WES. DNA was

extracted from blood using the QIAamp DNA Blood Maxi Kit (Qiagen, Cat # 51192) following the manufacturer's protocol. WES analysis is detailed in **Section 2.1.3** 'Computational Methods'.

1.5μg of genomic DNA were sent for WES to the South Australian Cancer Council Genomics Facility. From the original AIC cohort, five proband-parent trios and two singletons were selected for whole genome sequencing (WGS). 2.5μg of genomic DNA were sent for WGS to the Australian Genome Research Facility. WGS analysis is detailed in **Section 2.1.3** 'Computational Methods'.

2.3.5 TOPFlash Assay

HEK 293T Cell Transfection for TOPFlash Assay

The method described in **Section 2.2.1** 'HEK293T cell transfection for protein expression' was performed with the following exceptions: (1) maximum passage number of cells were 15 and (2) plasmid and final concentrations listed in **Table 2.11** were added to each well along with 50ng of renilla and 500ng of M50 8X TOPFlash plasmid (Addgene, Cat # 12456).

500ng pCDNA3.1 empty vector control	500ng pCDNA3.1 empty vector control	500ng WNT8B wildtype plasmid	500ng WNT8B wildtype plasmid
200ng WNT8B mutant plasmid	200ng WNT8B mutant plasmid	500ng WNT8B mutant plasmid	500ng WNT8B mutant plasmid
500ng WNT8B wildtype plasmid 200ng WNT8B mutant plasmid	500ng WNT8B wildtype plasmid 200ng WNT8B mutant plasmid	500ng WNT8B wildtype plasmid 500ng WNT8B mutant plasmid	500ng WNT8B wildtype plasmid 500ng WNT8B mutant plasmid

Table 2.11: Transfection layout on 12-well plate for TOPFlash assay.

Cell Harvest & Lysis

Reagents from the Dual Luciferase Reporter Assay Kit (Promega, Cat # E1960) were prepared following the manufacturer's recommendation for 12-well plates (Quick Protocol, 2009). Growth

media is removed from 12-well transfections plates and briefly rinsed with 500µL of 1X PBS. PBS is removed and 250µL of 1X Passive Lysis Buffer is added to each well; then gently rocked at room temperature for minimum 1 hour. Cell lysates were then transferred to a new 1.5mL screw cap tube.

Luciferase Reporter Assay

The luciferase reporter assay was then performed according to the manufacturer's protocol (Promega). The luminescence of each cell lysate was measured using the GloMax 20/20 Luminometer (Promega, Cat # E5311). The resulting data, containing the firefly:renilla ratio, was analysed by normalizing the relative luciferase activity (RLA) against the empty vector control measurement. The experiment was replicated three times and provided an indirect measure of canonical WNT signalling activation.

2.3.6 Western Blot

Protein Extraction

Lysis buffer was prepared using 1X stock of Pagano buffer (50mM Tris (pH 7.5), 250mM NaCl, 5mM EDTA and 0.1% Triton X-100 and sterilized water) or RIPA buffer (120mM NaCl, 0.5% Nonidet P-40 (NP-40), 50mM Tris-HCl (pH 8.0) and sterilized water) with 25X protease inhibitor cocktail (Sigma, Cat # P8340), 1mM NaF and 1mM $NaVO_4$. HEK293T cell pellets stored at -80^{o}C were resuspended in 300-500µL lysis buffer, then incubated on ice and after homogenized for 15 seconds (15 x 1 second pulses) via sonification. Samples were then spun at 4^{o}C; the supernatant was transferred to a new 1.5mL eppendorf tube for storage (-80^{o}C) or protein quantification.

Protein Quantification

Protein concentration of whole cell lysates was estimated using the Pierce BSA Protein Assay Kit. (ThermoFisher Scientific, Cat # 23209). From the original bovine serum albumin (BSA; 2 mg/mL) stock, aliquots at the following concentration dilutions were prepared: 1mg/mL, 0.5mg/mL, 0.25mg/mL and 0.10mg/mL. For each sample of unknown protein concentration, the original stock was diluted 1:10 in sterilized water. Unknowns and standards were prepared in triplicates on a 96 well plate following the manufacturer's protocol. The plates were read at 540nm using an automated plate reader (Dynatech MR5000). A concentration for each unknown was determined using the BSA standard curve and averaging the values of the triplicate samples.

Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

Protein samples were denatured in 1X Laemmli buffer (0.0625M Tris-HCl (pH 6.8), 2% sodium dodecyl sulfate, 10% glycerol, 5% beta-mercaptoethanol, 0.001% bromophenol blue and sterilized water) at 95^{o}C for 10-15 minutes in a thermocycler or heating block. 20-30µL of each sample (10-20µg) was then loaded onto a 15-well homemade Bis-Tris gel (30% acrylamide, 0.8% bis-acryl, 0.375M Tris-HCl (pH 8.8), 0.1% SDS, 0.025% Temed and ammonium persulfate and sterilized water); the gel was made up of 7% separating gel (bottom) and 4% stacking gel (top). For molecular weight estimation of the target protein, we loaded the samples next to 2-3µL of the Precision Plus Dual Color Standard (BioRad, Cat #1610374). Proteins were separated on the gel in a Mini-Protean Tetra Cell (BioRad) at 100-140mV powered by the Power Pac 3000 (BioRad) for 1.5-2 hours depending on the target protein size.

Membrane Transfer

Separated proteins were then transferred from the gel to a nitrocellulose membrane (BioTrace NT, Cat # 66485) using the XCell II Blot Module Kit (Invitrogen, Cat # EI9051). Protein transfer

was conducted in a XCell SureLock Mini-Cell (Invitrogen, Cat # EI0002) filled with 1X NuPAGE Transfer Buffer (Invitrogen, Cat # NP0006) at 12mV overnight in a 4ºC fridge. Post transfer, the membranes were treated with Ponceau S solution (Sigma Aldrich, Cat # P7170) to check for even loading and then briefly washed with 1X Tris buffered saline with Tween20 (TBS-T).

Immunoblotting

Transfer membranes were then blocked with 5% blocking milk (95% of 1X TBST and 5% skim milk powder) on an automatic shaker for either 1 hour at room temperature or overnight at 4ºC. Primary antibodies were then diluted in 2.5% blocking milk and added to the membrane in a heat-sealed plastic sleeve for 4-24 hours on an automatic rotator (**Table 2.12**). The membrane was then washed to remove unbound antibodies using 1X TBS-T for 30 minutes (6 x 5 minutes) on the automatic shaker. HRP-conjugated secondary antibodies were then diluted and incubated the same way as primary antibodies except for only 1-2 hours.

Detection

Antibody-bound membranes were then treated with 0.5-1mL of Amersham ECL Western Blotting Detection Reagents (GE LifeSciences, Cat # RPN2109) following the Manufacturer's protocol and exposed to the Amersham Hyperfilm ECL at time intervals of: 10 seconds, 30 seconds, 1 minute and 5 minutes.

Protein Target	Predicted Size (kDa)	Primary Antibody		Secondary Antibody	
		Name (Supplier)	Dilution	Name (Supplier)	Dilution
B-actin	42	Mouse Anti-B-Actin mAb (Sigma Aldrich, Cat # A3854)	1:3000	Goat pAb Anti-Mouse IgG HRP Conjugated (Dako, Cat # P026002-2)	1:3000
B-tubulin	50	Rabbit Anti-beta Tubulin pAb (Abcam, Cat # ab21058)	1:5000	Goat anti-Rabbit IgG HRP Conjugated (Dako, Cat # P044801-2)	1:3000
HSP90	90	Rabbit HSP90 [C45G5] mAb (Cell Signaling, Cat # 79641)	1:2000	Goat anti-Rabbit IgG HRP Conjugated (Dako, Cat # P044801-2)	1:3000
SLF1	116	Rabbit Anti-BRCTD1 [aa166-195) pAB (LifeSpan BioSci,	1:1000	Goat anti-Rabbit IgG HRP Conjugated (Dako, Cat # P044801-2)	1:3000

		Cat # LS-C319479)			
TEAD1	39	Mouse c-Myc [9E10] mAb (Santa Cruz Biotech, Cat # sc-40)	1:3000	Goat pAb Anti-Mouse IgG HRP Conjugated (Dako, Cat # P026002-2)	1:3000
WNT8B	39	Rabbit anti-V5 pAB (Bethyl Lab, Cat # A190-120P)	1:2000	Goat anti-Rabbit IgG HRP Conjugated (Dako, Cat # P044801-2)	1:3000
wnt8b	39	Rabbit anti-WNT8B pAB (Aviva Systems Biology, Cat # ARP33491_T100)	1:1000	Goat anti-Rabbit IgG HRP Conjugated (Dako, Cat # P044801-2)	1:3000
wnt8b	39	Rabbit anti-WNT8B pAB (Aviva Systems Biology, Cat # OAAB22283)	1:2000	Goat anti-Rabbit IgG HRP Conjugated (Dako, Cat # P044801-2)	1:3000

Table 2.12: Detailed summary of primary and secondary antibodies used for immunoblotting.

2.3.7 Morpholino Knockdown in Zebrafish

Breeding & Maintaining Zebrafish Embryo

Tubingen zebrafish was bred and maintained by the Alzheimer's Disease Genetics Laboratory (The University of Adelaide). Briefly, adult zebrafish was kept in filtered and oxygenated fresh water with temperatures maintained at 28.5^{o}C. There were two feeds daily with dry trout pellets in the morning and live adult brine shrimp in the evening. Breeding was controlled by an automated 12-hour light and 12-hour dark cycle; embryos were collected via the marbling technique described in Westerfeld et al. (1995).

Morpholino Antisense Oligonucleotides & Microinjection

Morpholino antisense oligonucleotides (MO) were designed and synthesized by GeneTools (http://www.gene-tools.com). All experimental MO were designed to block translation by targeting sequences upstream of the ATG start codon of genes listed in **Table 2.13**. A control morpholino that shares the same morpholino backbone as the experimental MO, but has no target and minimal biological activity is also included in the experiment as a negative control (**Table 2.13**). Embryos at one or two-cell stage were injected with MO as described by Westerfield et al. (1995); at a concentration gradient of 0.25mM, 0.5mM, 0.75mM and 1mM. The dose of MO that

resulted in less than 50% of early embryonic death were selected for phenotype analysis. We performed double-blinded injections of six different groups: uninjected, control-MO and four experimental-MO. A total of 240 embryos were injected per group for the final analysis.

Target	Morpholino sequences	Reference
control	5'-CCT CTT ACC TCA GTT ACA ATT TAT A 3'	GeneTools
tead1b	5’ - GAGTCGCTCATCCTCTCGATGTC - 3’	N/A
ankrd32/slf1	5’ - GTATTGTCCACCATCCTGAAACCGA - 3’	N/A
hcn1	5’ - TGCCGCTGCATAAACCCGTATTGCC - 3’	N/A
szt2	5’ - TGCCATGCCGTGCTTCTCATTAC - 3’	N/A
wnt8b	5’ - AGGGAGACTTTTCTTCACCTTTCAC - 3’	Hofmeister et al., 2013 Russek et al., 2008 Lee et al., 2006

Table 2.13: Morpholinos used for gene knockdown in zebrafish embryos.

Imaging & Morphant Phenotype Analysis

Embryos were viewed and scored under a dissection microscope (Paul Thomas lab) at 48 and 72 hours post fertilization (hpf). All embryos were classified into four morphological classes (**Figure 2.4**) based on their alignment with or deviation from normal embryonic development (Kimmel et al., 1995). There were three independent scorers for the morphological assessment based on images taken at 72hpf.

Figure 2.4: Morphant Phenotype Classes. Zebrafish embryos were scored according to eye classifications by Miesfeld et al. (2015). Class I: normal phenotype; Class II: normal eye pigmentation with mild body defects; Class III: reduced eye pigmentation with normal to mild body defects; Class IV; reduced eye pigmentation and severe body defects.

Protein Analysis

For wnt8b morphants, 10 72hpf whole embryos were dechorionated, deyolked and lysed in 15µL sample buffer (2% sodium dodecyl sulfate, 5% B-mercaptoethanol, 25% glycerol, 0.0625 M Tris-HCl [pH 6.8] and bromophenol blue). Sample buffer containing the embryos were immediately incubated at 95^{o}C for 20 minutes and stored at -80^{o}C. Protein from the embryo lysates were then separated, transferred, probed and visualized based on methods described in **Section 2.2.6** 'Western Blotting'. Unfortunately, none of the purchased antibodies were specific for the wnt8b protein; there were also no polyclonal zebrafish antibodies available for other gene targets.

CHAPTER III: A mutation in *COL4A2* causes autosomal dominant porencephaly with cataracts.

3.1 Introduction

The central nervous system is mostly recognised for its neural and glial network, however 10-20% of the mammalian adult brain is comprised of the extracellular matrix (ECM) (Lau et al., 2013). There are ECM glycoproteins involved in every facet of early brain development from maintenance and modulation of neural stem cells, to cell differentiation and migration (Barros et al., 2011). Among these glycoproteins, collagens have been shown to facilitate cortical lamination, orchestrate neuronal migration and provide axonal guidance during brain development (Luo et al., 2011, Caglayan et al., 2014, Shinwari et al., 2015, Ackley et al., 2001). Mutations in various collagen subtypes have been linked to cortical malformations such as polymicrogyria (Caglayan et al. 2014).

The ECM is comprised of three compartments: the basal lamina, perineuronal nets and neural interstitial matrix, which dynamically reassembles during different stages of brain development (Lau et al., 2013). Collagen IV (COL IV) is an important component of the basal lamina, which is comprised of six α-polypeptide chains encoded by six differentially expressed genes (Schmidt et al., 1993). These genes are arranged in pairs on chromosomes: 2, 13 and X (Khoshnoodi et al., 2008). Each pair produces two paralogous α-peptides through a bidirectional promoter (Haniel et al., 1995, Khoshnoodi et al., 2008). The COL IV α-peptides assemble into heterotrimers with the following stoichiometry: α1α1α2, α3α4α5 and α5α5α6 (Khoshnoodi et al., 2008).

Aberrations to COL IV heterotrimers can lead to genetic and acquired diseases (**Table 3.1**) (Haniel et al., 1995). For example, autosomal dominant and recessive mutations in COL IV α3

and α4 chains can lead to Alport Syndrome (MIM 104200; MIM 203780) and hematuria (MIM 301050) (Lemmink et al., 1996, Vahedi and Alamowitch, 2011). Meanwhile, X-linked mutations in COL IV α-5 and α-6 can lead to Alport Syndrome (MIM 301050) and deafness (MIM 300914) respectively (Rost et al., 2014, Lemmink et al., 1997, Vahedi and Alamowitch, 2011). COL4A1 and COL4A2 are the most abundantly and ubiquitously expressed COL IV peptides (Favor et al., 2007). Mutations in the corresponding genes can affect multiple organs resulting in various cerebral, ocular, renal and muscular pathologies (Favor et al., 2007, Kuo et al., 2014).

Mutations in *COL4A1* result in well-recognised phenotypes (**Table 3.1**) including: porencephaly (MIM 175780), brain small vessel disease (MIM 607595), cerebral haemorrhages (MIM 614519) and angiopathy with other systemic defects (MIM 607595) (Vahedi and Alamowitch, 2011). The current clinical picture of *COL4A2* is limited to porencephaly type II (MIM 614483) and intracerebral haemorrhage (MIM 614519). While systemic and eye findings have been reported in isolated cases (**Table 3.2**), these clinical features are less well recognised as part of the *COL4A2* spectrum (Gunda et al., 2014, Verbeek et al., 2012).

Cerebral defects are prominent in all individuals with pathogenic *COL4A2* mutations and there is *in vivo* evidence from mice models, that missense mutations in *Col4a2* can also induce ocular, myopathic and systemic defects (Favor et al., 2007, Kuo et al., 2014). Here, we expand the current clinical spectrum of *COL4A2* by identifying a novel missense mutation in a family with autosomal dominant inheritance of porencephaly, focal epilepsy and lens opacities.

GENE[8]	OMIM	DISORDERS	REFERENCE
COL4A1	18000	Tortuosity of retinal arteries	(Gekeler et al., 2006, Plaisier et al., 2010, Zenteno et al., 2015)
	61773	Angiopathy with nephropathy, aneurysms and muscle cramp	(Plaisier et al., 2007)
	607595	Brain small vessel disease	(Gould et al., 2006, Sibon et al., 2007, Lemmens et al., 2013, Coupry et al., 2010, Rodahl et al., 2013, Deml et al., 2014)
	175780	Porencephaly 1	(Gould et al., 2005, Aguglia et al., 2004, Breedveld et al., 2006, Smit et al., 1984, de Vries et al., 2009)
	614519	Susceptibility to intracerebral haemorrhage	(Weng et al., 2012, Shah et al., 2012)
COL4A3	104200	Alport Syndrome (dominant)	(Heidet et al., 2001, van der Loop et al., 2000, Jefferson et al., 1997)
	203780	Alport Syndrome (recessive)	(Mochizuki et al., 1994, Lemmink et al., 1997, Knebelmann et al., 1995, Finielz et al., 1998, Webb et al., 2014)
	141200	Benign hematuria	(Badenas et al., 2002)
COL4A4	203780	Alport Syndrome (recessive)	(Lemmink et al., 1997, Mochizuki et al., 1994, Boye et al., 1998)
	141200	Benign hematuria	(Badenas et al., 2002, Lemmink et al., 1996)
COL4A5	301050	Alport Syndrome	(Turco et al., 1995, Guo et al., 1995, Smeets et al., 1992, Knebelmann et al., 1992, Renieri et al., 1992, Boye et al., 1991, Barker et al., 1990)
COL4A6	300914	X-linked Deafness	(Rost et al., 2014a)

Table 3.1: Collagen IV related disorders. This table summarises the disorders associated with mutations in Collagen IV genes.

[8] Genes are represented by their HUGO Gene Nomenclature Committee (HGNC) symbol The OMIM phenotype number is provided along with the corresponding studies.

	***COL4A2* Family** (Human)	***COL4A2*** *(Human)*		***Col4a2*** *(Mice)*		***COL4A1*** *(Human)*		***Col4a1*** *(Mice)*	
Type of Variant[9]	THR-G	THR-G	NGM	THR-G	NGM	THR-G	NGM	THR-G	NGM
Total Genotypes[10]	1	5	4	1	1	41	7	10	2
Cerebral									
Intracerebral	1[11]	2	3	1	1	21	4	8	2
Haemorrhage		2	1			6	1		
Intracranial	1	3	2	1	1	14	2	8	2
Aneurysms			1_{SEB}			8	1		
MCD	1	4	1	1	1	29	1	8	2
Microbleeds	1	3				26	1		
Porencephaly									
White Matter Change									
Cardiovascular									
Aortic Valve							1		
Replacement						1			
Mitral Valve						5			
Prolapse									
Supraventricular									
Arrhythmia									
Muscular									
Bruising								2	
Muscle		1		1		10	2		
Cramps/Elevated CK				1	1			8	2
Vascular Defects						2			
Other Myopathies									
Neurological									
Developmental	1	2	1			6			
Delay	1	1				9			
Seizures	1	1_{MILD}	1						
Intellectual	1	5	1			9			
Disability			1						
Paresis/Plegia									
ADHD									

[9]Mutations in the corresponding genes are divided into two groups: one that displaces a glycine in the triple helical region (THR-G), while the other displaces a non-glycine residue (NGM).

[10] Total genotypes refer to the total number of genotypes observed per mutation group. The values in the table represent the sum of observed phenotypes per mutation group. Abbreviations: ADHD, Attention Deficit Hyperactive Disorder; BCA, Bowman's capsule abnormality; CK, creatine kinase; IOP, internuclear ophthalmoplegia; MCD, malformation of cortical development; NGM, non-glycine mutation; SEB, subependymal bleeding; THR-G, triple helical region-glycine.

[11] The values in the table represent the sum of observed phenotypes per mutation group.

Abbreviations: ADHD, Attention Deficit Hyperactive Disorder; BCA, Bowman's capsule abnormality; CK, creatine kinase; IOP, internuclear ophthalmoplegia; MCD, malformation of cortical development; NGM, non-glycine mutation; SEB, subependymal bleeding; THR-G, triple helical region-glycine.

Ocular									
Cataracts/Lens	1			1	1	17		10	1
Opacities				1	1	3		10	1
Anterior Segment			1	1		1	2	6	1
Dysgenesis				1		11		6	1
Optic Nerve Hypoplasia	1_{IOP}			1	1	6		10	2
Retinal Artery Tortuosity									
Other Eye Defects									
Renal									
Hematuria				1		3	1	6	
Microalbuminuria				1				6	
Renal Cysts						5			
Renal Insufficiency						4	1		
Urinary Retention						1			
Other nephropathy						2		1_{BCA}	
Other									
Hypothyroidism		1							
Nausea/Vomiting		1							

Table 3.2: Phenotypes of *COL4A1* and COL4A2 variants in humans and mice. This table is modified after Kuo and Meuwissen (Kuo et al., 2012, Meuwissen et al., 2015).

3.2 Materials & Methods

Informed written consent was obtained from all subjects or their parents or legal guardians in the case of minors. The study was approved by the New Zealand Health and Disability Ethics Committee and the Adelaide Women's and Children's Health Network Human Research Ethics Committee.

3.2.1 Exome Sequencing

Whole exome sequencing (Roche Nimblegen SeqCap v3) was performed on a HiSeq2500 (Illumina) by the South Australian Cancer Council Genomics Facility on three affected individuals (**I-2**, **II-2**, **III-1**; **Figure 3.1A**). Reads were mapped to the human genome (hg19) using BWA-MEM (Li and Durbin, 2009) and mapping refined using Genome Analysis Toolkit

(GATK) version 3.2-2 (DePristo et al., 2011). Mapping achieved a minimum median target coverage depth of 46 reads/sample and covered 95% of intended targets with at least 10 reads (Supplementary Fig. S1). Single nucleotide variants and small insertions and deletions were called by the GATK haplotype caller version 3.2-2 (DePristo et al., 2011). Larger copy number and structural variants were analysed by CoNIFER (Krumm et al., 2012); however, no segregating, rare copy number variants were detected. All variants were annotated for allele frequency, clinical significance, locus identity, and likely pathogenicity using ANNOVAR (Wang et al., 2010). Novel genotypes shared between affected individuals, but absent from a control set of 15 exomes, were separated using the vcf-contrast module from VCFtools (Danecek et al., 2011).

3.3 Results

3.3.1 Clinical Findings

The proband (**Figure 3.1A, III-1**) of this family was a 15-year-old male. He was born at term by emergency caesarean section at 41 weeks due to failure to progress. First developmental concerns occurred at 11 weeks due to a lack of head control. By 5 months it was clear he was globally delayed and was developing hydrocephalus. Twelve hours after a VP shunt insertion at 2 years of age, he had his first seizure. Subsequently he has had intractable multifocal seizures including episodes of convulsive status epilepticus. At three years of age he was noted to have bilateral small posterior lens opacities which have not been progressive. He now has spastic quadriplegia and profound Intellectual disability. EEGs show marked asymmetry of the background with suppression of amplitude, increased slowing in the left hemisphere and epileptic activity in the right parietal region. MRI at 3 months shows a large right frontal region of porencephaly that communicates with a very dilated right lateral ventricle and moderate left lateral ventricular dilatation (**Figure 3.1B, i**). Prior intraventricular haemorrhage has resulted in hemosiderin

staining of the ventricular walls with residual blood products evident in the posterior horns. The right caudate nucleus, lentiform nucleus, internal capsule, thalamus and right hippocampus are destroyed. The corpus callosum is extremely thinned (**Figure 3.1B, ii**).

The proband's 11-year-old brother (**Figure 3.1A, III-2**) presented at 3 years of age with focal dyscognitive seizures. He was born at 39 weeks following an uneventful pregnancy by emergency caesarean section due to cord prolapse. His early development was slightly delayed in that he didn't walk until 18 months and had no words until 2 years of age. At 5 years of age he was diagnosed with pervasive developmental disorder with associated mild to moderate intellectual impairment. His focal seizures continue and are all of similar semiology. At 14 months of age he was noted to have bilateral posterior subscapular lens opacities, which were not progressive, but his neurological examination was otherwise in keeping with his developmental level. His MRI revealed a right frontal porencephalic cyst communicating with the right lateral ventricle (**Figure 3.1B, iii**). There is partial destruction of the periventricular white matter and caudate body. A smaller cyst is was present in the right external capsule. The corpus callosum has had a very unusual appearance (**Figure 3.1B, iv**). *COL4A1* sequencing was normal and at the time *COL4A2* was not considered as a candidate gene.

The 42-year-old mother (**Figure 3.1A, II-2**) of the two boys was asymptomatic with no seizures or neurological symptoms and had a normal IQ. Her neurological examination was normal apart from a mild left internuclear ophthalmoplegia. MRI showed a cluster of small cysts in the right frontal periventricular white matter and bilateral anterior caudate nuclei (**Figure 3.1B, v**). There are numerous scattered T2/FLAIR hyperintense lesions in keeping with foci of gliosis, scattered throughout the white matter, more prominent frontally (**Figure 3.1B, vii**).

The 67-year-old maternal grandmother (**Figure 3.1A, I-2**) was noted to have a mild right hemiplegia in the first year of life. At the time, this was felt to be due to a difficult delivery at term, which required high forceps. She was left handed early and delayed in walking. She had learning difficulties at school and although is in the normal range cognitively has never felt her memory was very good. She developed focal dyscognitive seizures in childhood, which were treated with phenobarbital. She did not develop convulsive seizures until the age of 47 years. These continue occurring 4 times per year and she has had episodes of convulsive status. Following the convulsive seizures, she has a right-sided Todd's paresis. Her focal seizures have failed to respond to carbamazepine, valproate, vigabatrin and clobazam. On examination her right limbs are smaller than the left and she has an obvious right hemiplegia. She has refused neuroimaging and has never had an EEG. She developed bilateral cataracts at the age of 65 years.

Informed, written patient consent was obtained from all subjects involved in this study.

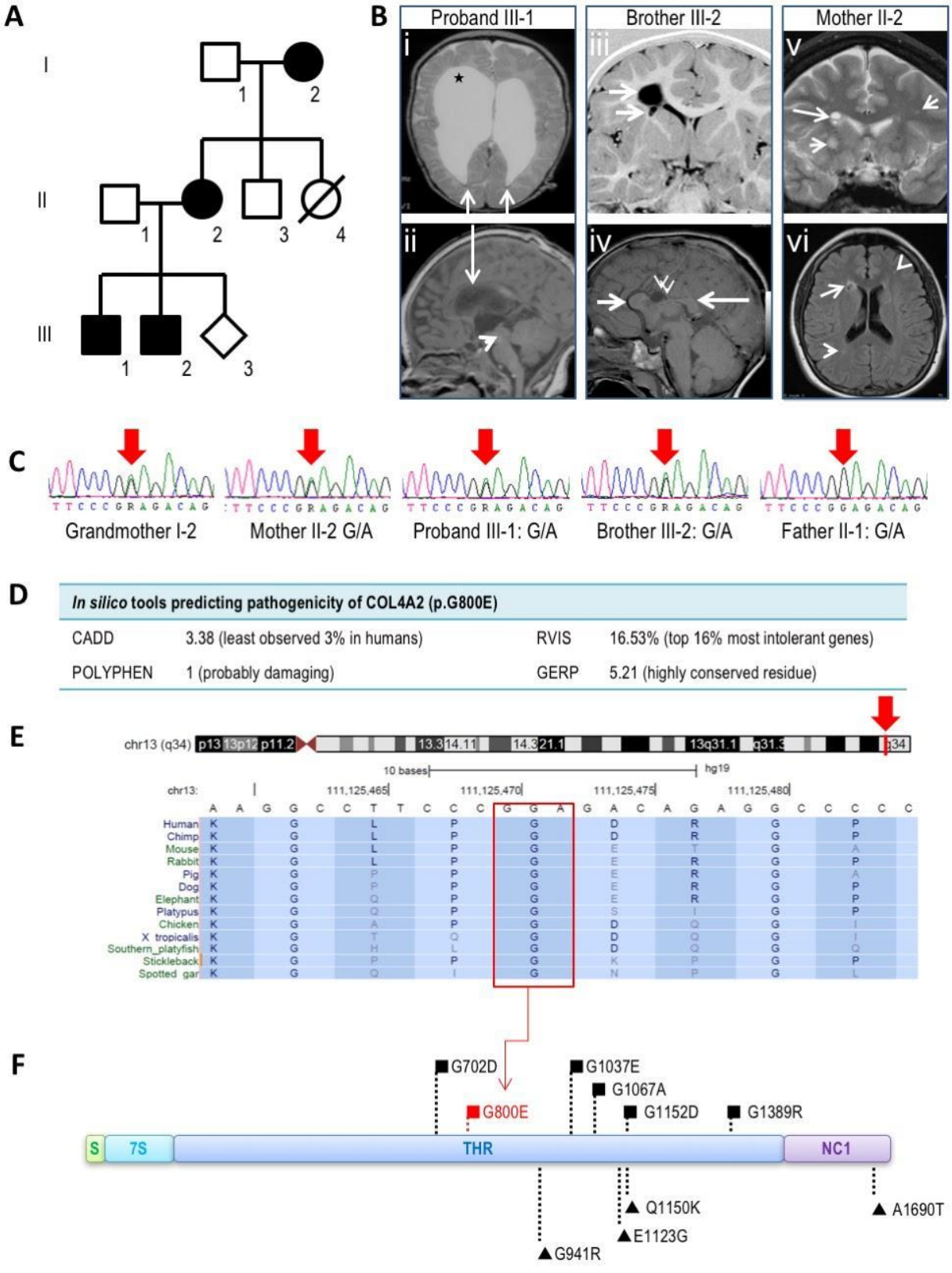


Figure 3.1: Familial porencephaly caused by a COL4A2 mutation. A: Pedigree. Affected individuals shown in black boxes (males) and circles (females). **B:** MRI brain scans. (i) Proband

III-1: Axial T2 demonstrates right frontal porencephalic cyst (star) and hydrocephalus. Hemosiderin staining (T2 hypointensity) of the ependymal lining of the posterior horns of the lateral ventricles (arrows). (ii) Sagittal T1 weighted image demonstrates thinning of the corpus callosum (long arrow) and loss of visualisation of the aqueduct of Silvius (arrowhead). (iii) III-2: Coronal T1 IR sequence shows multiple right frontal porencephalic cysts (arrows), the largest extends into the superior frontal lobe. (iv) Sagittal T1 weighted image demonstrates a dysmorphic corpus callosum the thick rostrum (short arrow) and splenium (long arrow) missing the callosal body (double arrowheads). (v) II-2: Coronal T2 weighted image of the frontal lobes demonstrate multiple small cysts in the right frontal periventricular white matter (long arrow). Numerous scattered foci of T2 hyperintensity are in keeping with gliosis related to small vessel disease (short arrows). (vi) Axial FLAIR demonstrates two small cysts in the right frontal periventricular white matter with surrounding hyperintense gliosis (arrow). Foci of gliosis in the left frontal lobe and right posterior temporal periventricular white matter (arrowheads). **C:** Chromatogram showing the heterozygous c.2339G>A change (arrows), which is present in the affected grandmother (I-2), mother (II-2), proband (III-1) and brother (III-2); and absent in the father (II-1) and maternal uncle (II-3). **D:** Predicted pathogenicity of COL4A2 c.2399G>A, p.G800E determined by four prediction algorithms. **E:** Ideogram of chromosome 13 and the COL4A2 locus (arrow) at 13q34. The mutated amino acid G800 is highly conserved across different orthologs (boxed). **F:** Domain structure and mutations of COL4A2. COL4A2 p.G800E (grey) is located within the triple helical region (THR). Published COL4A2 mutations predisposing to porencephaly (squares) and intracerebral haemorrhage (triangles) are shown.

3.3.2 Overview of filtered variants

Whole exome sequencing (Roche Nimblegen SeqCap v3) was performed on a HiSeq2500 (Illumina) by the South Australian Cancer Council Genomics Facility on three affected individuals (**Figure 3.1A, I-2, II-2, III-1**). Reads were mapped to the human genome (hg19) using BWA-MEM (Li and Durbin, 2009) followed by local realignment around Indels using Genome Analysis Toolkit (GATK) version 3.2-2 (DePristo et al., 2011). Mapping achieved a minimum median target coverage depth of 46 reads per sample and covered 95% of intended targets with at least 10 reads (**Figure 3.2**). Single nucleotide and small insertions and deletions were called by the GATK haplotype caller version 3.2-2 (DePristo et al., 2011). Larger copy number and structural variants were analysed by CoNIFER (Krumm et al., 2012) however, no segregating, rare copy number variants were detected. All variants were annotated for allele frequency, clinical significance, locus identity and likely pathogenicity using ANNOVAR (Wang et al., 2010). Novel genotypes shared between affected individuals, but absent from a control set of 15 exomes, were separated using the vcf-contrast module from VCFtools (Danecek et al., 2011).

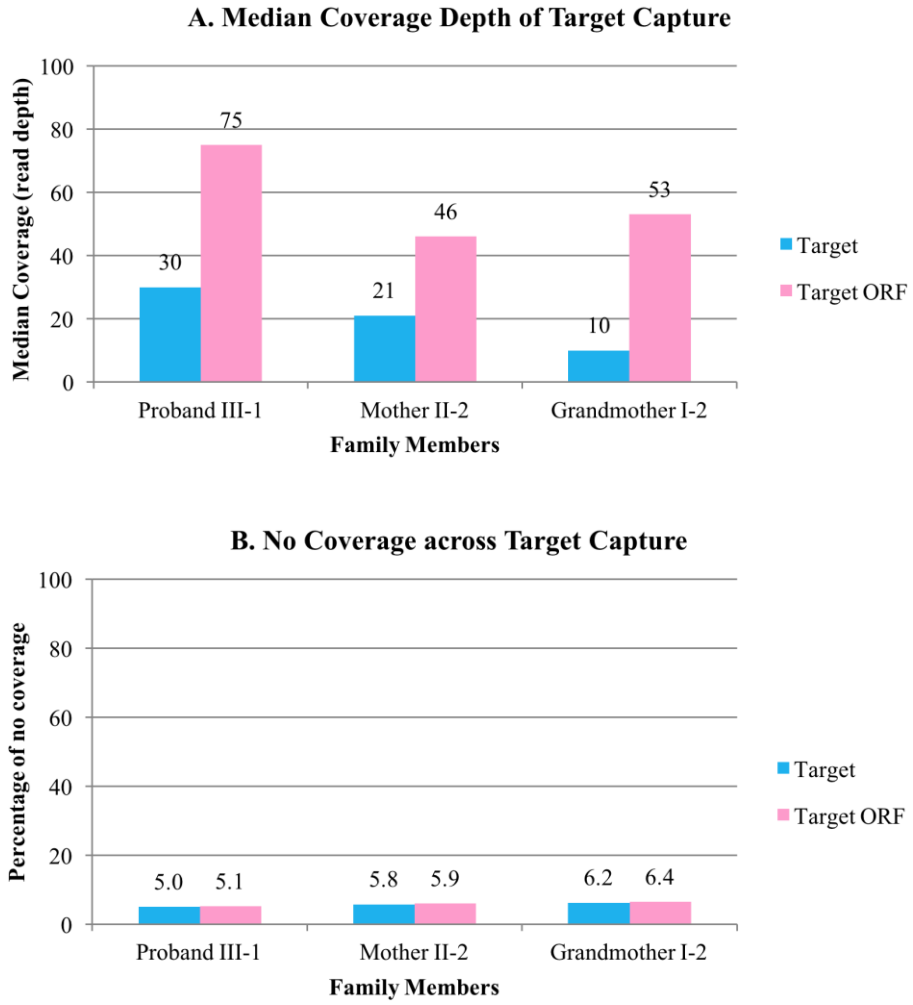


Figure 3.2: Median read depth coverage across capture targets and open reading frames (ORF). A: The median coverage depth across the Roche Nimblegen SeqCap v3 target capture (**blue**) and ORF within the target capture (pink). **B:** The percentage of base pairs with zero coverage across the target capture and ORF within the target capture.

3.2.3 Predicted pathogenicity of COL4A2 p.G800E

There were 8388 variants that were shared between the three affected individuals (**Table 3.3**). Of these, 1049 variants were below the frequency thresholds of 0.01 (EVS, 1000genomes, CG69, Wellderly) and 0.001 (ExAC, UK10K). Then there were 409 variants with a minimum read depth of 10 X that passed GATK variant quality score recalibration. Finally, 69 of these were exonic variants, which were assessed for further likely pathogenicity (**Table 3.3**). Only one variant was

found that passed our filtering criteria and also best fit the clinical description of the patients: CCDS41907.1(*COL4A2*):c.2399G>A, p.(G800E) (**Table 3.4**). Sanger sequencing confirmed the segregation of the variant in the grandmother, mother and grandsons (**Figure 3.1C**). The COL4A2 p.G800E variant was predicted as likely disease causing by multiple variant predictors (**Table 3.5**). The COL4A2 p.G800 is highly conserved across multiple species; primates through to fish (**Figure 3.1E**).

INCLUSION CRITERIA	GRANDMOTHER (I-2)	MOTHER (II-2)	PROBAND (III-1)
All Variants (GATK-HC)	27642	88156	35453
Segregating Variants (VCFtools)		8388	
Rare Variants (MAF <0.01/0.001)[a]		1049	
Coverage Depth ≥ 10 Reads		809	
GATK Quality Pass[b]		409	
Coding		**69**	
Missense		45	
Non-Synonymous		17	
LOF		7	
Highly Conserved (GERP ≥ 4)		25	
Affect Protein Function[c]		6	

Table 3.3: Summary of filtered variants. This table lists the criteria used for filtering common (likely benign) variants from rare (potentially pathogenic) variants. [12]

[12] The number of variants that passes each filtering stage is shown on the right.
[a]The frequency thresholds were set at 0.01 in CG69 (n=69), 0.001 in Wellderly (n=454), 0.0001 in UK10K (n=3781) and 1000 genomes (n=1092); and 0.00001 in ExAC (n>60,000).
[b]Variants flagged as PASS following variant quality score recalibration using GATK. [c]Variants were considered to affect protein function as determined by SIFT and PolyPhen. Abbreviations: GATK-HC, genome analysis toolkit haplotype caller; GERP, genome evolutionary rate profiling; LOF, loss of function; MAF, minor allele frequency; n, individuals.

Variant Information			Pathogenicity Assessment			
HGNC Symbol	**gDNA Change**	**ExAC Frequency**	**Brain**	**Intolerant**	**Affect Protein Function**	**Related Phenotypes**
TTN	chr2:179581972G>A	**AC:** 4 **GT:** 120534	Yes	no (98.04%)	3/6	Cardiomyopathy (MIM 613765; 604145) Muscular dystrophy (MIM 600334; 608807) Hereditary myopathy (MIM 603689; 611705)
USP40	chr2:234465596C>G	**AC:** 2 **GT:** 120736	Yes	no (79.68%)	4/6	Unknown
ST6GALNAC4	chr9:130677116C>T	**AC:** 1 **GT:** 66400	Yes	**yes** (18.59%)	3/6	Unknown
F2	chr11:46744816C>T	**AC:** 4 **GT:** 121260	Yes	**yes** (14.57%)	3/6	Thrombinemia (MIM 613679) Thrombophilia (MIM 188050) Recurrent pregnancy loss (MIM 614390) Ischemic stroke (MIM 601367)
COL4A2	chr13:111125471G>A	0	Yes	**yes** (16.53%)	6/6	Porencephaly (MIM 614483) Intracerebral haemorrhage (MIM 614519)
MIF4GD	chr17:73262833T>A	**AC:** 24 **GT:** 121236	Yes	**yes** (39.95%)	3/6	Unknown

Table 3.4: Summary of potentially pathogenic variants. The likely pathogenicity of variants was determined by: low allele count in normal population, expressed in the brain, gene intolerant to mutations, predicted to affect protein function and related phenotype to our family. [13]

[13] The ExAC frequency provides the observed allele count (AC) among the called genotypes (GT) of individuals without neurodevelopmental disorders. Genes expressed in the brain were determined via Allen Brain Atlas (http://www.brain-map.org/). Six variant prediction tools (SIFT, PolyPhen, LRT, MutationTaster, MutationAssessor and FATHMM) were used to predict the effect on protein function. Related phenotypes were obtained via OMIM (http://www.omim.org/) and PubMed (http://www.ncbi.nlm.nih.gov/pubmed).
Green boxes support causality, **red** boxes indicate insufficient evidence for causality and **yellow** boxes provide inconclusive support for causality.

Abbreviations: gDNA, genomic DNA; HGNC, HUGO gene nomenclature committee; LRT, Likelihood Ratio Test.

ALGORITHM	DESCRIPTION	PREDICTION SCORES
CADD	Scores the deleteriousness of SNVs and INDELs in the human genome	Score: 3.38 Interpretation: amongst the highest 3% of variants least likely to be observed in humans
RVIS	Ranks genes in terms of whether they have more or less common functional genetic variation relative to the genome-wide expectation	Score: -0.64 Interpretation: amongst the 16.53% most intolerant genes
SIFT	Predicts whether an amino acid substitution affects protein function, based on the degree of conservation of amino acid resides in sequence alignments	Score: D Interpretation: Deleterious
PolyPhen-2	Predicts possible impact of amino acid substitution on the structure and function of a human protein	Score: D Interpretation: Damaging
LRT	Likelihood ratio for significantly conserved amino acid positions within the human proteome	Score: D Interpretation: Deleterious
MutationTaster	Evaluates disease causing potential of sequence alterations	Score: D Interpretation: Disease-Causing
MutationAssessor	Predicts functional impact of amino acid substitutions in proteins	Score: H Interpretation: High
FATHMM	Predicts the functional, molecular and phenotypic consequences of protein missense variants using hidden Markov models	Score: D Interpretation: Deleterious
GERP	Identifies constrained elements in multiple alignments of vertebrate species in the human genome	Score: 5.21 Interpretation: Highly conserved
PhyloP	Computers conservation or acceleration p values for multiple alignments of vertebrate species to the human genome	Score: 2.59 Interpretation: Deleterious
SiPhy	Detects bases under selection from a multiple alignment data	Score:18.71 Interpretation: Deleterious

Table 3.5: Predicted pathogenicity of COL4A2 c.G2399A. This table summarises the different algorithms and tools used to assess the likely pathogenicity of variants.[14]

[14] The predicted pathogenicity of COL4A2 c.G2399A (**blue** panel) is given as the raw score, as well as the interpretation of that score in context with the protein or disease likelihood.

3.4 Discussion

3.4.1 Phenotypes in human COL4A2-related disorders

We describe a family that was originally thought to best fit phenotypes associated with *COL4A1* mutations (**Table 3.2**) (Vahedi and Alamowitch, 2011). Lens opacities were observed in the proband and brother (**Table 3.2**). Lens opacities have not been recorded in any previous *COL4A2* cases (**Table 3.2**) but are a well-recognized feature, among other ocular defects, associated with *COL4A1*-related disorders (Coupry et al., 2010). To date, only one *COL4A2* mutation was associated with ocular defects that included myopia, amblyopia and abnormal optic discs (**Table 3.2**) (Verbeek et al., 2012). This suggests that *COL4A2* mutations can manifest dysgenesis of the retina (Verbeek et al., 2012) and lens (present case) in humans.

3.4.2 Phenotypes of COL4A2 and COL4A1 mutations in human and mice

The phenotypes observed in *COL4A2*-related disorders in humans, including the case we present here, have been replicated in and in some cases predicted by mice carrying *Col4a2* mutations (Favor et al., 2007, Kuo et al., 2014). Mice, identified from ethylnitrosurea (ENU) mutagenesis screens that were heterozygous for *Col4a2* mutations showed defects to the eye, including microphthalmia, buphthalmos and lens opacities; brain, such as porencephaly, intracerebral haemorrhage and abnormal neuronal migration; kidney function, indicated by increased urea in blood; vessel system, resulting from reduced vascular development (Favor et al., 2007) and; reduced muscle fibres suggestive of myopathy (Kuo et al., 2014). The renal and myopathic defects observed in mice correlate with the patient phenotype described by Gunda *et al.* (Gunda et al., 2014), which included: hematuria, muscle cramps and elevated creatine levels but these were not observed in the present case (**Table 3.2**). The severity and penetrance of eye, brain and muscular defects varied greatly among a single allele group of 18 mutant *Col4a2* mice on the inbred C57Bl/6J background strain (Kuo et al., 2014). In humans, dominant *COL4A2* mutations

also exhibit incomplete penetrance (**Table 3.6**) (Verbeek et al., 2012, Yoneda et al., 2012). The similarities observed between mice models and patients with *COL4A2* mutations support the notion that the effects of *COL4A2* mutations extend beyond the brain.

3.4.3 Prognosis of THR glycine mutations versus NGMs

A majority of COL4A1/A2 mutations occur at highly conserved glycine residues within the triple helical region (THR), (**Figure 3.1E & 3.1F**) (Favor et al., 2007). It is proposed that glycine substitutions in the THR destabilise COL IV heterotrimer formation leading to accumulation of misfolded protein and subsequent chronic stress on the endoplasmic reticulum (ER) (Kuo et al., 2014, Yoneda et al., 2012). Similar to COL4A1, missense glycine mutations in the THR of COL4A2 are likely to predispose severe cerebral pathologies with early onset (**Table 3.2**). While severe, COL4A2 THR glycine mutations usually lead to milder phenotypes and better prognosis compared to equivalent mutations in COL4A1, in both mice and humans (Favor et al., 2007). It has been hypothesised that the difference in phenotypic severity of COL4A1 versus COL4A2 mutations is due to the stoichiometry of α1: α1: α2 per heterotrimer (Favor et al., 2007, Jeanne et al., 2012, Khoshnoodi et al., 2008). Assuming that heterotrimer formation proceeds as usual, a heterozygous COL4A2 mutation will lead to 50% of COL4A1/A2 trimers with a mutant peptide, whereas a heterozygous COL4A1 mutation will lead to 75% of COL4A1/A2 trimers having a mutant peptide (Favor et al., 2007, Jeanne et al., 2012). Location of the pathogenic variant relative to different functional domains of the COL4A1/A2 peptides also influences phenotype (Des Parkin et al., 2011).

	Present study	Meuwissen *et al.*	Gunda *et al.*	Murray *et al.*	Jeanne *et al.*			Verbeek *et al.*		Yoneda *et al.*	
Variant Inheritance Proband	p.G800E Dom 3 months M	p.E1123G Dom	p.G941R De Novo 5 years M	p.G702D Dom M	p.E1123G De Novo adult	p.Q1150K Dom adult	p.A1690T De Novo adult	p.G1067A Dom 6 months M	p.G1389R Dom 6 months F	p.G1152D Dom 6 months M	p.G1037E De Novo At birth M
Brain MRI Findings	**Proband (III-1)** PC, HC, IVH; atrophy of CC, CN, LN, IC, thalamus and hippocampus **Brother (III-2)** PC, abnormal CC, atrophic CB **Mother (II-2)** Cysts in PV white matter and CN, gliosis	SCHIZ	ICH, intracranial aneurysms, LE	PC	ICH	ICH	ICH	**Proband** cerebral atrophy, PC	**Proband** PC, WMC **Sister** PC, WMC **Mother** Internal carotid artery aneurysm	**Proband** VM, WMC, atrophy of CC and CP **Mother** VM, PC	ICH, VM, atrophy of CC
Neurology	**Proband (III-1)** Profound ID, MF seizures, spastic QP **Brother (III-2)** Focal seizures, pervasive DD, mild ID **Mother (II-2)** Internuclear OP **Grandmother (I-2)** Hemiplegia, LD, focal seizures	DD, Attention deficit hyperactive disorder	Facio-brachial hemiparesis, hemiparesis, HH					**Proband** Dystonic HP, MC, DD	**Proband** HP, mild ID, DD	**Proband** DD, spastic triplegia, spastic paresis **Mother** Focal seizures, monoparesis **Uncle** (Mat) mild MP, HP **Grand Uncle** Congenital HP	Spastic QP
Eye Findings	**Proband (III-1)** Lens opacities **Brother (III-2)** Lens opacities **Grandmother (I-2)** Late onset cataracts		None	None				**Proband** MO, AO, small OD **Mother** tilted OD, MO	None	None	None
Other Anomalies			Hematuria, muscle cramps, elevated CK					**Mother** HT			

Table 3.6: Phenotypes associated with COL4A2 variants in humans. This table provides a summary of human cases with COL4A2 variants. [15]

[15] The variant is described by the predicted protein change identified in the corresponding studies. The age of onset (clinical presentation) and gender of the proband per study is noted. For familial cases, the clinical findings of extended family members are included.

Abbreviations: AO, amblyopia; CB, caudate body; CC, corpus callosum; CK, creatine kinase; CN, caudate nucleus; CP, cerebral peduncle;; DD, developmental disorder; HC, hydrocephalus; HH, hemihypesthesia; HP, hypothyroidism, IC, internal capsule; ICH, intracerebral haemorrhage; IVH, intraventricular haemorrhage; LD, learning difficulties; LE, leukoencephalopathy; LN, lentiform nucleus; Mat, maternal; MC, Microcephaly; MF, multifocal; MO, myopia; MP, monoparesis; OD, optic disks; OP, ophthalmoplegia PC, porencephaly; PV, periventricular; SCHIZ, schizencephaly; QP, quadriplegia; VM, ventriculomegaly; WMC, white matter change.

Non-glycine mutations NGMs within the THR and also the non-collagenous domain are likely low penetrance, risk alleles for late onset ICH with good survivability (**Table 3.1**) (Jeanne et al., 2012). Our review of, all COL4A2 mutations showed NGMs are present in the ExAC database quite frequently while pathogenic THR glycine mutations are absent (**Table 3.7**) (Gunda et al., 2014, Verbeek et al., 2012, Yoneda et al., 2012). Nonetheless, these NGMs were shown to decrease the ratio of extracellular to intracellular COL4A2 relative to wild-type and also two of three variants induced stress in the ER in a similar way to THR glycine mutations, which supports their potential to be causal (Jeanne et al., 2012). It is interesting to note in hindsight that based on their population frequencies, all three NGMs would have been excluded in their original study (Jeanne et al., 2012). This acts as a cautionary tale that filtering of variants using public databases must be tailored to the disease, especially those disorders with late onset.

3.4.4 Concluding remarks

In conclusion, we identified a novel CCDS41907.1(*COL4A2*):c.2399G>A, p.(G800E) mutation that is predicted to be highly pathogenic based on the substitution of a highly conserved, critical THR-glycine residue that segregated with the phenotype in affected individuals. Our findings are supported by the existing genotype to phenotype correlation between previous *COL4A2* THR glycine mutations and porencephaly with focal structural epilepsy due to ischaemic damage. We also extend the *COL4A2* phenotype spectrum, with the observation of lens opacities in some affected individuals; reducing the distinction between *COL4A1* and *COL4A2* mutations. This highlights the importance of simultaneous screening both COL4A1/A2 genes for accurate diagnosis of severe, early onset porencephaly.

REFERENCE	INHERITANCE	DNA CHANGE	AA CHANGE	DBSNP ID	VAF	
					STUDY	EXAC
(Gunda et al., 2014)	De novo dominant	c.2821G>A	p.G941R	--	1	0
(Jeanne et al., 2012)	De novo dominant	c.3368A>G	p.E1123G	rs117412802	2	801
	Autosomal dominant**	c.3448C>A	p.Q1150K	rs62621875	1	131
	De novo dominant	c.5068G>A	p.A1690T	rs201105747	1	65
(Verbeek et al., 2012)	Autosomal dominant	c.3206delC	p.G1067A	--	3	0
	Autosomal dominant	c.4165G>A	p.G1389R	--	3	0
(Yoneda et al., 2012)	Autosomal dominant	c.3455G>A	p.G1152D	rs387906602	4	0
	De novo dominant	c.3110G>A	p.G1037E	rs387906603	1	0
Present	Autosomal dominant	c.2399G>A	p.G800E	--	4	0

Table 3.7: Allele frequency of *COL4A2* mutations in ExAC. This table summarises the variant allele frequency of reported COL4A2 Mutations.[16]

[16] Two allele frequencies (right hand panel) are provided. The first frequency is the observed allele among all participants in the corresponding study (multiple observations are usually from related individuals). The second frequency is the observed allele count among ~60,000 individuals in the ExAC browser.
***COL4A2* c.3448C>A was identified in a male with two affected sisters with intracerebral haemorrhage. The related members were not available for segregation analysis, thus inheritance is unconfirmed but predicted as likely autosomal dominant.

Abbreviations: AA, amino acid; ExAC, Exome Aggregation Consortium; VAF, Variant Allele Frequency.

3.5 Acknowledgement

I would like to firstly thank my co-authors, in particular:

- Simone Mandelstam who comprehensively reviewed and provided the brain MRI scan of the proband, brother and mother.
- Lynette Sadleir, Sarah Paterson and Ingrid Scheffer who meticulously reviewed the patients and provided clinical descriptions.

Secondly, I would like to thank:

- Alison Gardner for the CNV analysis.
- eResearchSA and the SA Cancer Council Genomics Facility for bioinformatics support.

3.6 Publication

The contents of this chapter have been published as a case report in the journal *American Journal of Medical Genetics Part A* under the same title. Details of the publication are provided in **Appendix 7.3.1**.

CHAPTER IV: Genetical heterogeneity and complexity of Aicardi Syndrome revealed through whole exome and whole genome sequencing.

4.1 Introduction

Genetic testing for Aicardi Syndrome is currently unavailable as the underlying cause(s) have yet to be identified. Early investigations into the genetics of AIC were predominantly focused on chromosome X for two reasons (Neidich et al., 1990, Nielsen et al., 1991, Hoag et al., 1997, Yilmaz et al., 2007, Eble et al., 2009, Pai et al., 2013). Firstly, the first chromosomal aberration reported in a suspected AIC female was an X/3 translocation (Ropers et al., 1982). Secondly, an X-linked male-lethal cause would best explain the predominance of the disease in females, which are represented in more than 99% of cases reported in literature. Despite 70 years of genetic and genomic investigations, an X-linked cause has yet to be established. By reviewing the genetic loci of variants reported in AIC thus far, we have yet to observe a recurring gene hit or sub-chromosomal overlap (**Table 1.7**).

Genetic investigations into AIC has yet to be thoroughly explored outside of the X chromosome; except for the study by Schrauwen et al. (2015). In this study, two autosomal variants in *TEAD1* and *OCEL1* were identified in two unrelated individuals respectively. Both of these affected individuals had intact corpus callosum accompanied with chorioretinal lacunae and infantile spasms. Therefore, did not display the classical triad (**Figure 1.1**), but the presence of other MCDs sufficed the AIC diagnosis recommended by Aicardi (2005). No functional validation was performed to assess the molecular consequence of the predicted deleterious variants in *TEAD1* or *OCEL1;* thus, the genotype to phenotype correlation is not proven. In an independent study of a cohort of 38 girls who displayed the hallmark classical triad (**Figure 1.1**), additional variants were not identified in these genes (Wong et al., 2016). Similarly, no further variants were

identified in *TEAD1* or *OCEL1* from two parent-proband trio who were analysed for overlapping autosomal-pathogenic variants (Lund et al., 2016).

In this study, we used a similar sequencing strategy to Schrauwen et al. (2015) by incorporating both WES and WGS for variant discovery. In contrast to previous genetic studies, we searched for small variants (SNP and Indels) using multiple variant callers (**Figure 4.1**). We also explored three potential inheritance models: *de novo*, homozygous recessive and compound heterozygous. Our cohort was comprised of individuals displaying the classical triad (strong diagnosis; **Figure 1.1**), major and minor features (supported diagnosis) and overlapping features with AIC (suspected diagnosis). The clinical heterogeneity of our cohort enabled us to query whether there are: (1) different genetic causes underlying different clinical subsets of AIC or (2) mutations in a recurring gene that is pleiotropic and with variable penetrance or expressivity. We also investigated the molecular consequences of potential candidate genes using protein-specific *in vitro* assays and/or the phenotypic consequences of morpholino-mediated knockdown in zebrafish embryos. Our study aimed to provide insight into the genetic and molecular factors that may contribute to the pathology of AIC.

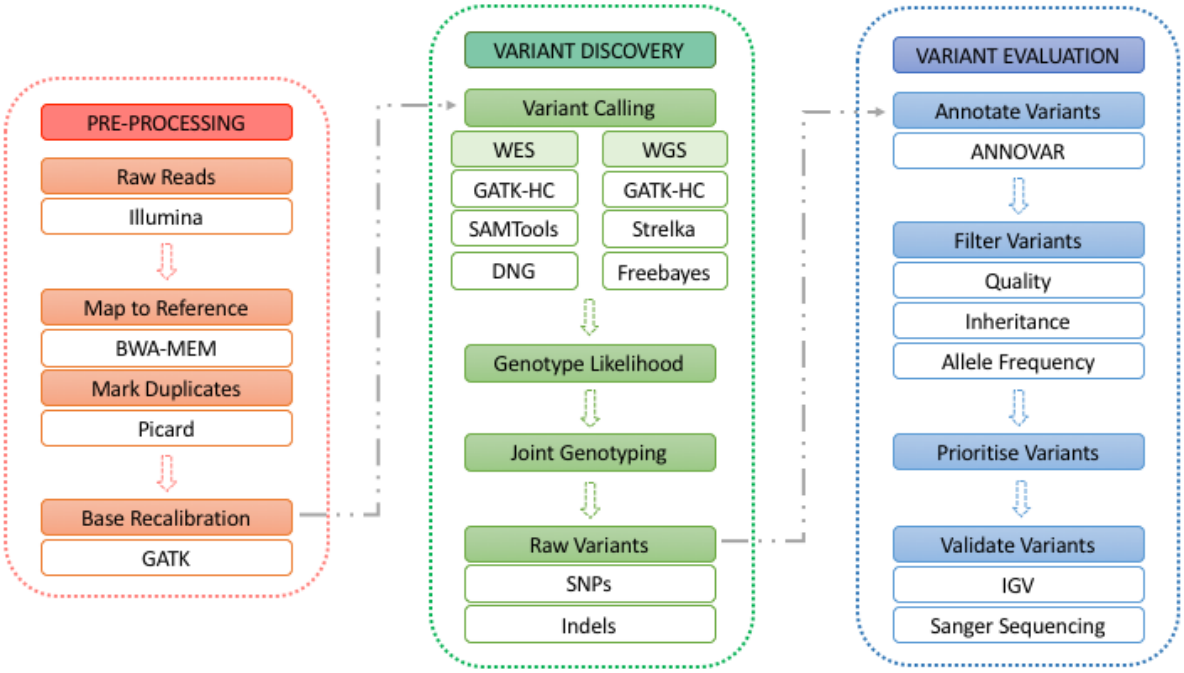


Figure 4.1 Schematic Overview of SNP and Indels processing pipeline. This diagram was adapted from GATK best practice guideline for SNP and Indels discovery. It summarises the processes involved in converting raw sequencing reads to a list of candidate variants, which were then manually assessed for their likely pathogenicity.

4.2 Results

4.2.1 Clinical summary of Aicardi cohort

A majority of the probands were identified and clinically assessed by Professor Ingrid Scheffer of The University of Melbourne and Florey Institute. Other clinicians that recruited single patients for our cohort included: Dr. Nick Smith (WCH Paediatric Neurology), Dr. Chirag Patel (Genetic Health Queensland) and Dr. Vandana Shashi (Duke University Medical Centre). Consent was obtained for all patients enrolled in this study by the ethics committees affiliated with each respective clinician. The clinical characteristics for all probands in our AIC cohort are summarised in **Table 4.1**, while the full phenotypic spectrum is described in **Table 4.2.** Serological tests for congenital viruses and metabolic levels, performed antenatally were negative

for all probands. 9/12 probands had no family history of neurodevelopmental disorders (**Figure 4.2**), while 10/12 had unremarkable pregnancy and delivery. No genomic or genetic abnormalities were detected through prior clinical genetic testing for some of the patients that underwent karyotyping (1/12), array-based CNV screening (5/12) and candidate gene sequencing (1/12).

Clinical Features	**T22842**	**T25217**	**T22101**	**49265**	**T22418**	**T28446**	**T25752**	**43774**	**T22352**	**T17262**	**T25387**	**T25820**	**TOTAL**
Major Features													
Infantile spasms*	1	1	1			1		1	1	1	1	1	8
Chorioretinal lacunae	1	1	1			1		1		1		1	8
Coloboma of the optic disc					1	1					1		3
Agenesis of the corpus callosum			1	1	1			1	1	1		1	7
Cortical malformations (mostly microgyria)		1	1	1	1	1	1		1	1	1	1	10
Periventricular (and subcortical) heterotopia		1	1									1	3
Intracranial cysts			1									1	2
Papilloma of choroid plexuses	1					1					1		3
Supporting Features													
Vertebral and costal abnormalities	1	1	1		1					1			5
Microphthalmia and/or other eye abnormalities			1							1	1		3
'Split brain' EEG		1	1		1	1	1						5
Gross hemispheric asymmetry									1			1	2
Other Features													
Developmental delay	1	1		1	1	1			1				6
Intellectual disability (severe or profound)		1								1			2
Hemiplegia (or diplegia/quadriplegia)		1	1				1						3

Secondary microcephaly													0
Strength of Diagnosis													
Strong			1			1		1		1		1	5
Supported	1	1			1						1		4
Suspected				1			1		1				3

Table 4.1: Clinical summary of patients in study cohort.

HPO ID	Phenotype Description	T228422	T252217	T221201	492665	T224418	T28446	T255752	437744	T223522	T172262	T253887	T258220	Total
HP:0001263	Developmental Delay	1	1	1		1	1	1	1	1	1			9
HP:0007858	Chorioretinal Lacunae	1	1	1			1		1		1		1	7
HP:0001274	Agenesis of the Corpus Callosum			1		1			1		1		1	5
HP:0012469	Infantile Spasm	1	1	1					1		1		1	6
HP:0002126	Polymicrogyria		1	1		1		1					1	5
HP:0002119	Ventriculomegaly	1		1			1				1		1	5
HP:0002282	Heterotopia		1	1									1	3
HP:0200022	Choroid Plexus Papilloma	1					1					1		3
HP:0002521	Hypsarrhythmia		1			1		1						3
HP:0000588	Optic Disc Coloboma					1	1					1		3
HP:0001338	Partial Agenesis of the Corpus Callosum						1			1		1		3
HP:0000729	Autism Spectrum Disorder									1				1
HP:0002539	Cortical dysplasia		1				1							2
HP:0011182	Epileptiform Discharges					1		1						2
HP:0011968	Feeding Problems						1			1				2
HP:0010841	Multifocal Epileptiform Discharges			1			1							2
HP:0002650	Scoliosis			1							1			2
HP:0002069	Tonic Clonic Seizure	1							1					2
HP:0010818	Tonic Seizure						1			1				2
HP:0010576	Intracranial Cystic Lesions												1	1

HP:0005263	Gastritis						1						1	2
HP:0001249	Intellectual Disability		1								1			2
HP:0000610	Abnormality of Choroid											1		1
HP:0002269	Abnormality of Neuronal Migration							1						1
HP:0000493	Abnormality of the Fovea										1			1
HP:0000932	Abnormality of the Posterior Fossa										1			1
HP:0001965	Abnormality of the Scalp												1	1
HP:0002335	Agenesis of Cerebellar Vermis										1			1
HP:0003355	Aminoaciduria												1	1
HP:0100702	Arachnoid Cyst			1										1
HP:0002099	Asthma								1					1
HP:0000483	Astigmatism			1										1
HP:0001272	Cerebellar Atrophy			1										1
HP:0030048	Colpocephaly												1	1
HP:0001305	Dandy Walker Malformation										1			1
HP:0000670	Dental Caries								1					1
HP:0002376	Developmental Regression												1	1
HP:0002307	Drooling									1				1
HP:0001999	Dysmorphic Facial Features										1			1
HP:0000577	Exotropia										1			1
HP:0012427	Femoral Anteversion			1										1
HP:0010734	Fibrous Dysplasia of the Bones			1										1
HP:0007359	Focal Seizures								1					1
HP:0010821	Gelastic Seizures									1				1
HP:0001269	Hemiparesis							1						1
HP:0002937	Hemivertebrae					1								1
HP:0000238	Hydrocephalus											1		1
HP:0000540	Hypermetropia										1			1
HP:0001252	Hypotonia		1											1
HP:0002123	Myoclonic Seizures								1					1
HP:0000121	Nephrocalcinosis							1						1

HP:0001052	Nevus Flammeus												1	1
HP:0001845	Overlapping Toes												1	1
HP:0002445	Quadriplegia			1										1
HP:0010636	Schizencephaly						1							1
HP:0004322	Short Stature										1			1
HP:0010535	Sleep Apnoea										1			1
HP:0000473	Torticollis					1								1
HP:0002948	Vertebral Fusion										1			1
HP:0003422	Vertebral Segmentation Defect										1			1

Table 4.2: Summary of human phenotype ontology (HPO) terms for probands in study cohort.

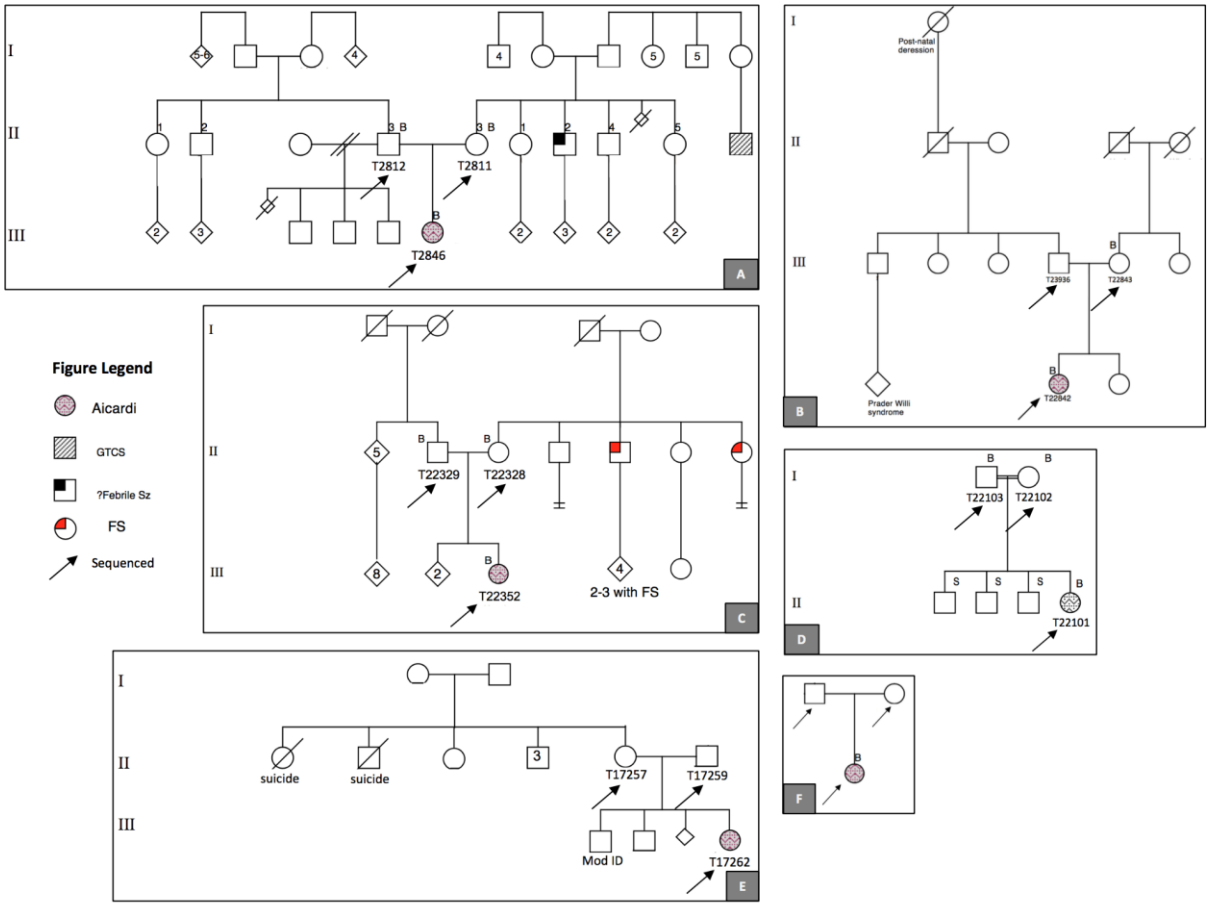


Figure 4.2: Pedigree map of Aicardi cohort study. A-C: The family history of neurodevelopmental disorders in probands T2846, T22842 and T22352. **D:** Proband T22842 is

the offspring of a consanguineous relationship. **E:** Unremarkable family history for Proband T17262. **F:** Schematic diagram representing the remaining probands in the cohort with unremarkable family history. *Abbreviations:* B, blood; S, saliva; Mod ID, moderate intellectual disability; GTCS, generalised tonic-clonic seizure; FS, febrile seizure.

4.2.2 Coverage analysis across whole exome sequencing and whole genome sequencing.

Prior to investigating the variants called using the SNP and indel discovery pipeline (**Figure 4.1**), we assessed SNP and Indels that may have been missed due to inadequate sequencing coverage. The average number of reads aligned to a reference genome, sequencing depth, was assessed across all Aicardi exomes and genomes using Picard (http://broadinstitute.github.io/picard/). Whole exome sequencing (WES) and whole genome sequencing (WGS) achieved a median sequencing depth of 30X and 40X respectively (**Table 4.3**). WGS had greater sequencing coverage and depth across protein-coding regions compared to WES (**Figure 4.3**). This can be observed from 30X sequencing depth, where WES covered 30-70% of target protein-coding sequences, while WES covered 75-95%. Overall, the observed sequencing depth of the Illumina exomes and genomes were comparable to other genetic studies that were successful at identifying disease-related SNP and Indels underlying various neurodevelopmental disorders.

WES_ID	Median Fold Coverage (bp)	Zero Coverage (%)	WGS_ID	Median Fold Coverage (bp)	Zero Coverage (%)
T22842	50	3.5	FR07959033	34	0.26
T25217	28	0.77	-	-	-
T22101	30	1.36	-	-	-
T2846	38	2.73	FR07959023	42	0.26
T25752	23	0.78	FR07958909	48	0.21
43774	30	0.75	FR07958917	35	0.31
T22352	35	3.86	-	-	-
T17262	30	3.76	-	-	-
T22418	-	-	FR07959030	42	0.26
49265	-	-	FR07958956	-	-
SQC0025F8	30	4.89	-	-	-
T25387	22	7.89	FR07958885	38	0.26
T25820	35	0.65	-	-	-
WES Median	**30**	**2.73**	**WGS Median**	**40**	**0.26**

Table 4.3: Median target depth and coverage. This table summarises the sequencing coverage across the open reading frame (ORF) of the SeqCap EZ Exome v3 capture targets (WES) or Hg19 canonical transcripts with 50bp flanking sequences (WGS).[17]

[17] dash (-), samples were not sent for sequencing or failed quality control; ID, identification; WES, whole exome sequencing; WGS, whole genome sequencing.

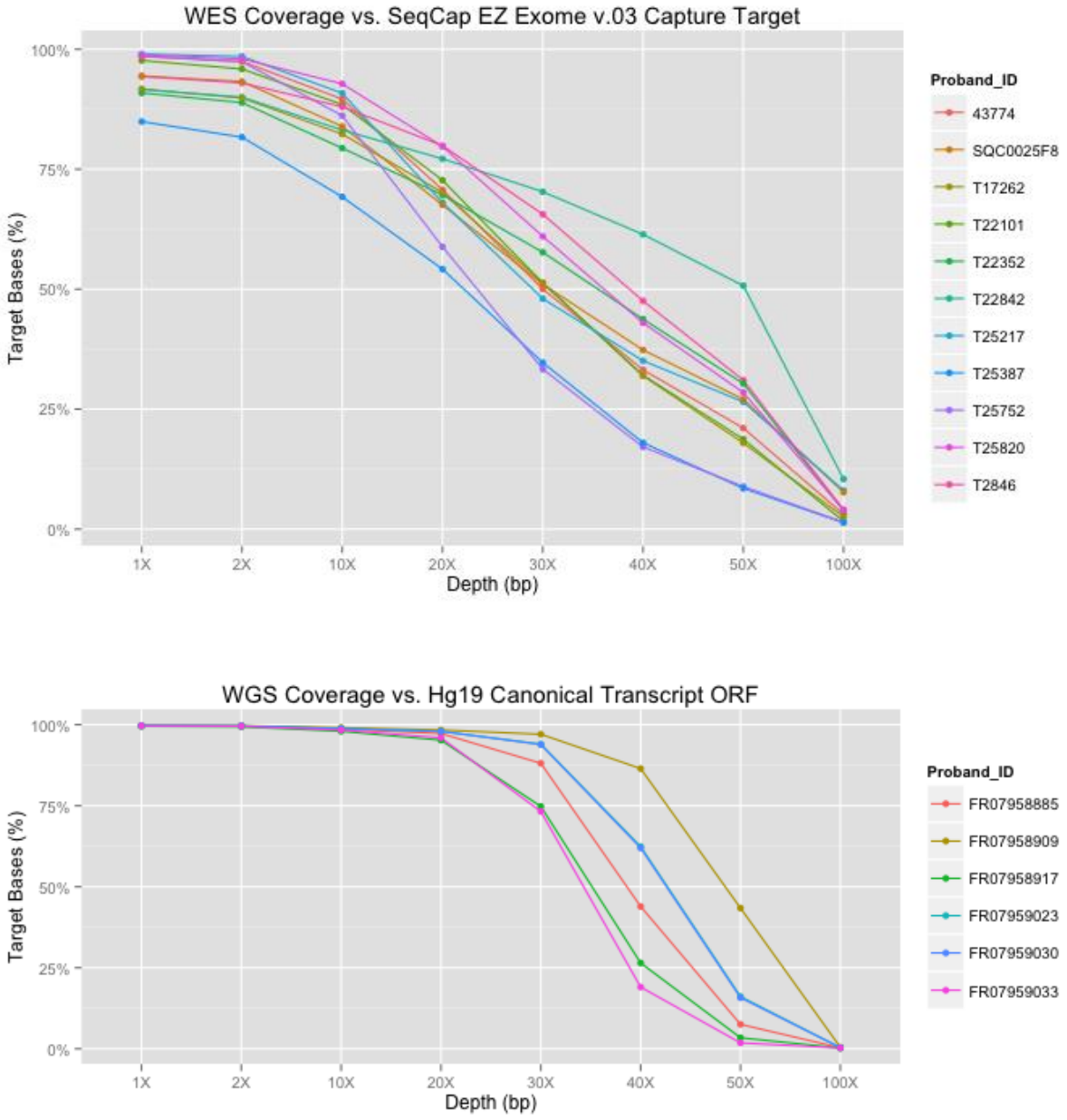


Figure 4.3: Aicardi Cohort WES and WGS coverage and depth across target bases. These line graphs for WES (top) and WGS (bottom) show the percentage of target bases covered (Y-axis) over different coverage depths (X-axis). Hg19, human genome build 19; ID, identification; ORF; v, version; WES, whole exome sequencing; WGS, whole genome sequencing.

We further examined the regions with any base of zero sequencing coverage across exome and genome targets. Approximately 3% (n = ~6.3Mb) of sequence bases were not covered by WES, while less than 0.3% (n = ~73kb) were not covered by WGS (T**able 4.3**). Among the regions with zero sequencing coverage, approximately 74% (n = ~54kb) of WGS and 98% (n = ~6.2Mb) of WES were observed in both Aicardi and unrelated (phenotypically normal) individuals (**Figure 4.4**). The majority of no coverage regions contained interspersed repeats or low complexity DNA sequences; approximately 75% for both WES and WGS (**Figure 4.4**) consistent with the known limitations of Illumina sequencing across repetitive and non-complex DNA sequences (Treangen et al., 2011, Aird et al., 2011).

A majority of the no coverage regions (NCR) resided outside of the open reading frame (ORF). After removing those found in unaffected individuals and flanking repeated sequences, there were less than 0.5% of NCR flanking ORF in WGS (n = ~2700) and WES (n = 37). The higher number of NCR in ORF of WGS is likely due to its increased sequence coverage and depth across ORF compared to WES (**Figure 4.3**). Next, we looked at NCR that flanked putative genes associated with eye, epilepsy, intellectual disability or malformations of cortical development (EEIM). Approximately 14% (WES) and 26% (WGS) of NCR that flanked ORF were in EEIM genes (**Figure 4.4**). Irrespective of being autosomal or X-linked, most of these NCR resided in sequences that were not well conserved, distant from splice junctions or did not overlap with a known regulatory element.

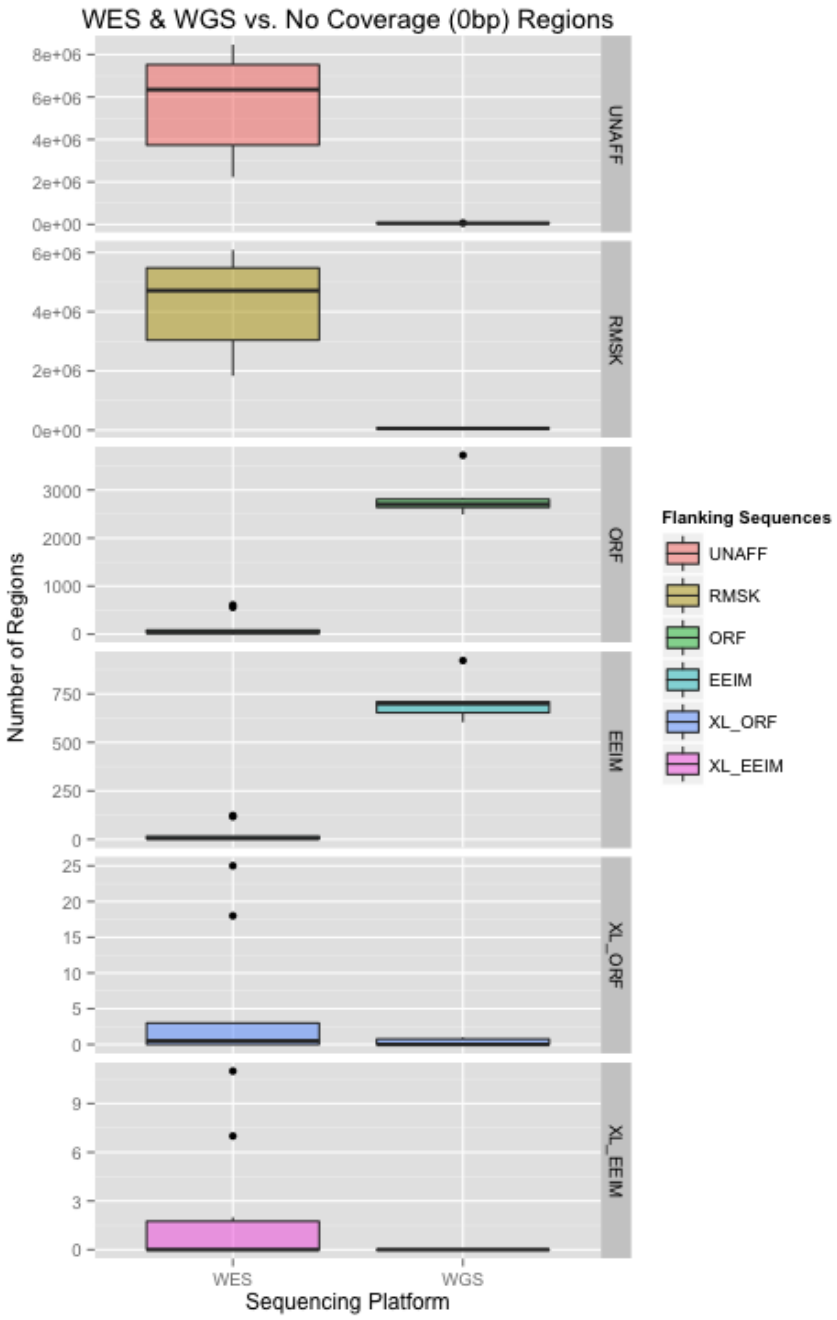


Figure 4.4: WES and WGS with zero sequencing coverage. Faceted boxplot of the different sequencing regions not covered by WES or WGS. UNAFF is the number of no coverage regions (NCR) from Aicardi exome or genome that is also present in unaffected individuals. RMSK is the number of NCR from Aicardi exome or genome that is flanking interspersed repeats and low complexity DNA sequences (Jurka et al., 2000). ORF is the number of NCR remaining after filtering out NCR that overlap with UNAFF and RMSK; and overlaps with human genome build

19 canonical transcripts from UCSC. EEIM are NCR that overlap with ORF and candidate gene list. EEIM, gene list text-mined with eye, epilepsy, intellectual disability and malformation of cortical development; ORF, open reading frame of Hg19 canonical transcripts; WES, whole exome sequencing; WGS, whole genome sequencing; XL, X-linked.

4.2.3 Detecting likely pathogenic SNP and Indels across exomes and genomes.

From our cohort (n=13), we performed whole exome sequencing (WES) on 12 probands and parents (if available) using peripheral blood DNA and the Agilent SureSelect EZ human exome version v3 capture kit (Agilent). We also performed WES using DNA extracted from resected focal cortical dysplasia (Proband T25217) or choroid plexus papilloma (Proband T25387). WES was conducted on the HiSeq2500 (Illumina) platform and the fastq files were processed as described in Figure 4.1. There were approximately 58,000 variants (SNP and Indels) called by GATK Haplotype Caller 3.2-2 per exome and pre-filtered using a control set of 15 exomes using VCFtools 'contrast' module (**Table 4.4**). Four percent of the called variants resided in protein-coding sequences of which ~83% were present in public datasets above a minimum allele frequency threshold for rare variants (**Table 2.5 & 4.4**). Approximately 30 variants per exome were predicted as likely deleterious by CADD20 (=>15) and PolyPhen2 (damaging or possibly damaging). Variants were then viewed on the Integrative Genome Viewer (IGV) using the probands' BAM files and parents (if available) to confirm for likely *de novo* and recessive variants. We applied the same variant filtering and candidate prioritising criteria to the paired blood-brain exomes from probands T25217 and T25387 (**Table 4.1**), but no somatic variants that were likely pathogenic were found in the brain exomes alone.

All variants that validated *in silico* (via IGV) were then manually evaluated and prioritised for sequencing if: (1) there were more than 5 remaining candidate genes and (2) associated with EEIM genes. For individuals with variants in less than 5 genes, all of the variants were sequenced. After manual evaluation and Sanger sequencing, only four variants were confirmed as truly *de novo* (**Table 4.6 & Figure 4.5**). This number of variants that validated is as expected, based on the estimated *de novo* substitution rate of 1.58 (Gilissen et al., 2014).

Proband WES ID	Called Variants	Protein Coding	MAF Threshold	Predicted Pathogenic	*De novo* or Recessive	Validated Variant
T22842	43046[18]	1885	310	38	9	
T25217	24642	2397	530	29	2	*De novo SZT2* (c.9103C>T; NM_015284; p.His3035Tyr)
T22101	86948	2819	406	66	9	*De novo WNT8B* (c.209T>C; NM_003393; p.Leu70Pro)
T2846	64862	2052	476	28	1	
T25752	47480	1671	178	1	0	
43774	69695	2648	207	1	0	
T22352	34009	1894	471	36	14	*De novo HCN1* (c.1625G>T, NM_021072, p.Cys542Phe)
T17262	137573	2262	543	41	2	*De novo SLF1* (c.1409T>C, NM_032290, p.Leu470Ser)
SQC25F8	49698	6921	873	61	1	
T25387	19014	1627	423	31	4	
T25820	58509	1645	188	0	0	
WES Average	57771	2529	419	30	4	

Table 4.4: Summary of filtered SNP and Indels from Whole Exome Sequencing (WES).

[18] This table shows the remaining number of variants per proband following each filtering parameter (first row).

For the remaining cohort with ambiguous or no variant hits, we proceeded with whole genome sequencing (WGS) using the HiSeqXTen (Illumina) platform. Sequence reads were processed and variants called by the Australian Genome Research Facility, using Genome analysis toolkit (GATK) v3.5-0 best practices guidelines (Van den Auwera et al. 2013). Variants within protein-coding regions were annotated and analysed equivalently to WES. On average, approximately 5 million SNP and Indels were called per genome. For our first parse, we analysed variants in protein-coding regions that may have been missed from WES. Here, an average of 42,000 exonic variants remained per proband and a further 82 fulfilled our disease-variant filtering criteria (**Table 4.5**). From our parent-proband trios (n = 5), after we applied inheritance-based filtering and variant prioritisation (**Section 2.1.5**), only one *de novo* variant in *KMT2B* validated via Sanger sequencing (**Table 4.6 & Figure 4.5**). For the remaining two singletons, we compiled a list of potential variant hits that require segregation when parent DNA is available (**Appendix 7.2.1**).

Proband WGS ID	Called Variants	Protein Coding	MAF Threshold	Predicted Pathogenic	Absent In-house	De novo or Recessive	Validated Variant
FR07959033	4937261	41055	4475	56	41	3	*De novo KMT2B (*c.6418C>G; NM_014727; p.Pro214Ala)
FR07959030	5026012	41880	4757	65	43	8	
FR07959023	5029242	41939	4965	58	40	5	
FR07958909	5071754	42832	5551	91	69	36	
FR07958917	5544399	46746	5532	115	102	N/A	
FR07958885	5000206	41386	4731	68	42	1	
FR07958956	5002490	42007	4502	49	40	N/A	
WGS Average	5087338	42549	4930	72	54	11	

Table 4.5: Summary of SNP and Indels filtered from Whole Genome Sequencing. [19]

[19] Abbreviations: WGS, whole genome sequencing; MAF, minimum allele frequency.

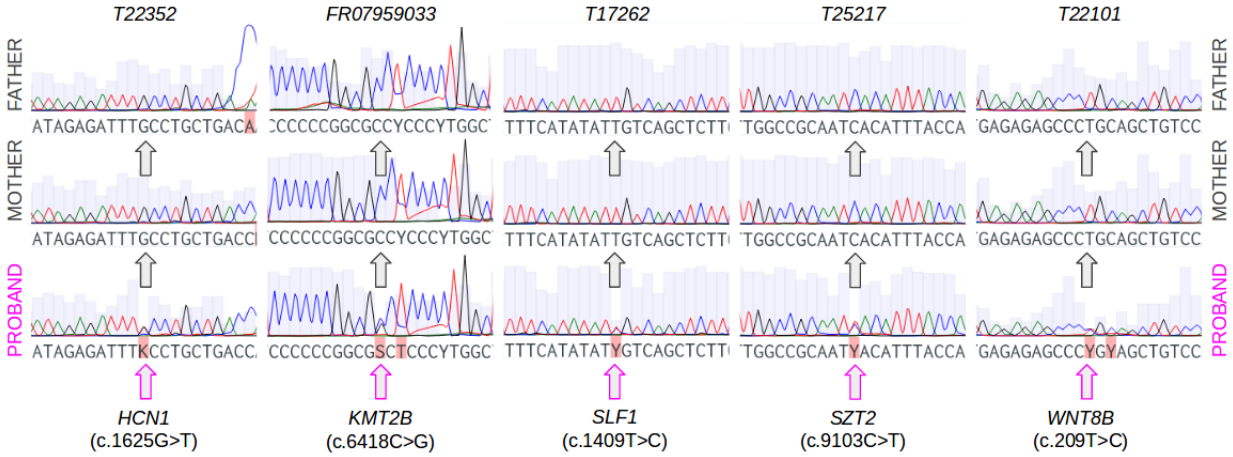


Figure 4.5: Sanger validation of *de novo* variants. Sequence chromatogram of each proband-parent trio (top line) with the locus harbouring the variant indicated by grey arrows (parent sequence) or pink arrows for confirmed *de novo* mutation (proband sequence). Two peaks can be seen in all of the probands, corresponding to the heterozygous variant (bottom row).

	T25217	***T22101***	***T17262***	***T22352***	***T22842***
HGNC Symbol	*SZT2*	*WNT8B*	*SLF1*	*HCN1*	*KMT2B*
Hg19 Position	chr1:43912827	chr10:102239737	chr5:94001606	chr5:45267349	chr19:36223868
cDNA Variant	c.9103C>T	c.209 T>C	c.1409T>C	c.1625G>T	c.6418C>G
Transcript ID	NM_015284	NM_003393	NM_032290	NM_021072	NM_014727
AA Change	p.His303Tyr	p.Leu70Pro	p.Leu470Ser	p.Cys542Phe	p.Pro214Ala
PolyPhen2	0.997 (D)	0.962 (D)	0.946 (D)	0.994 (D)	0.677 (P)
CADD	18.2	21.2	21.1	27.2	15.1
GERP+	5.43	5.7	5.41	5.82	3.99
ExAC	0	0	0	0	0
Zygosity	Heterozygous	Heterozygous	Heterozygous	Heterozygous	Heterozygous
Inheritance	*De novo*	*De novo*	*De novo*	*De novo*	*De novo*
Eye	no	no	no	no	yes
Epilepsy	yes	yes	no	yes	yes
ID	yes	no	no	yes	yes
MCD	no	no	no	no	yes

Table 4.6: Summary of potentially pathogenic variants. This table provides locus information and predicted pathogenic scores for each *de novo* variant found in the respective proband. [20]

[20] Abbreviations: AA, amino acid; D, damaging; ExAC, gDNA, genomic DNA; GERP, genome evolutionary rate profiling; Het, heterozygous; ID, intellectual disability; MCD, malformation of cortical development; P, probably damaging; WGS, whole genome sequencing.

4.2.4 Assessing the HCN1 (p.Cys542Phe) variant via Voltage clamp assay.

Among our implicated genes, only *HCN1* and *WNT8B* had pre-existing *in vitro* assays that can assess the functional consequence of the variants identified in proband T22352 and T22101 respectively. Our collaborators (Dr Melody Li and Dr Christopher Reid) assessed the biophysical properties of the *HCN1* (c.1625G>T, NM_021072, p.Cys542Phe) variant by performing a two-electrode voltage-clamp assay on frog (*Xenopus laevis*) oocytes expressing the wildtype HCN1 or mutant Cys542Phe channel. The mutant Cys542Phe channel displayed a slower activating inward current and a left-shift in the voltage-dependent activation relative to the wildtype channel (**Figure 4.6 A-C**). Oocytes expressing the Cys542Phe channel had a; hyperpolarising shift in the voltage of half activation (mutant = -73.98 ± 0.23 mV vs. wildtype = -6.16 ± 0.25mV, n = 13-17, P = 0.0001), a reduced slope of the activation curves (mutant = 6.35 ± 0.21 vs. wildtype = 7.66 ± 0.23, n = 13-17, $P < 0.05$) and significantly slower activation kinetics (**Figure 4.6 D-F**). The biophysical changes of the mutant Cys542Phe channel compared to the wildtype HCN1 channel were consistent with a loss-of-function.

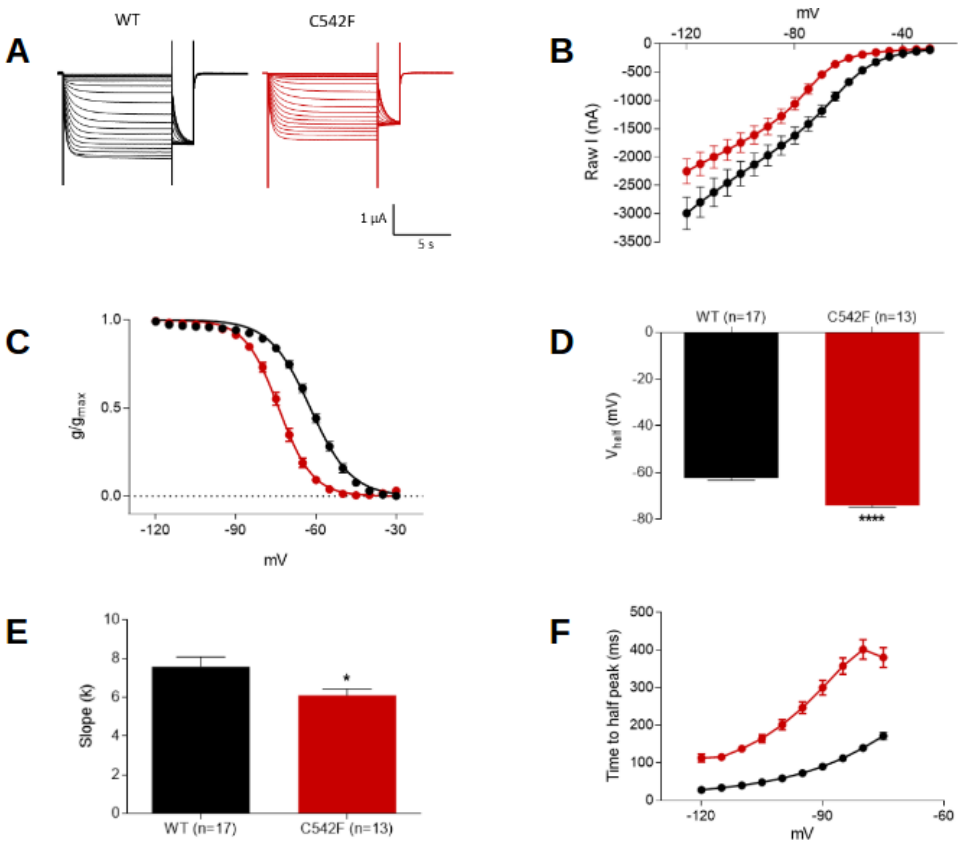
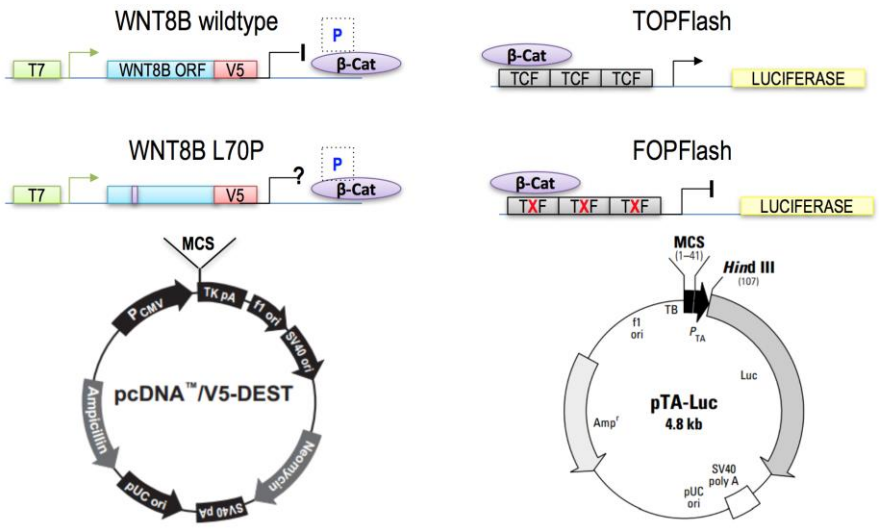


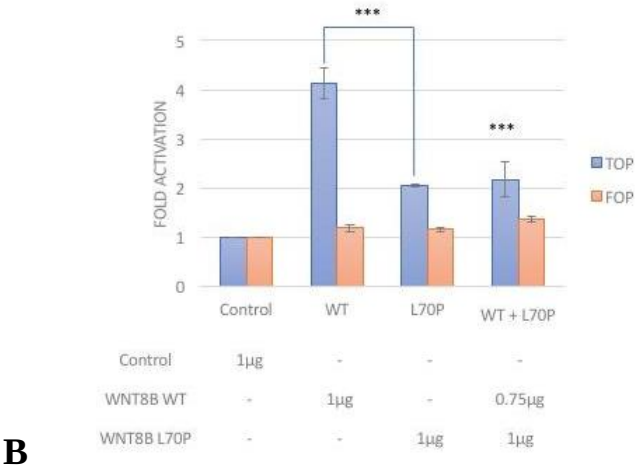
Figure 4.6: Electrophysiological characterisation of HCN1 variant. A: Series of current responses from a single oocyte expressing WT (black) or Cys542Phe (red) plasmids. **B:** Raw current-voltage relationship of WT (n=20) and Cys542Phe (n=14) mutant. Raw current was measured at the end of pre-pulse. C: Activation curve from average normalised tail current of WT (n=17) and Cys542Phe (n=13) mutant with Botlzmann fit. D: Average half activation voltage (V1/2). E: Slope factor derived from Boltzmann fit. F: Time to half maximal activation of WT (n=20) and Cys542Phe (n=14). Asterisks represent Student's t-test P-values (* $P < 0.05$, **** $P = 0.0001$). Error bars represent standard deviations across replicate experiments

4.2.5 Assessing the WNT8B (p.Leu70Pro) variant via TOPFlash assay.

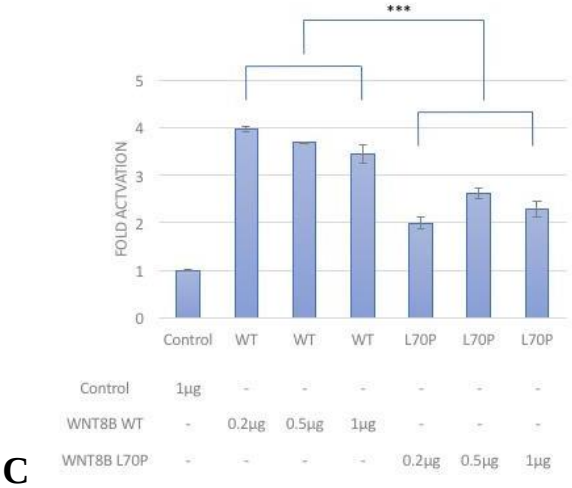
We assessed the effect of the *WNT8B* (c.209T>C; p.Leu70Pro; NM_003393) variant on β-catenin mediated WNT signalling using the TOPFlash assay in human embryonic kidney 293T (HEK 293T) cells (**Appendix 7.2.2**). In HEK 293T cells transfected with either the wildtype WNT8B or mutant Leu70Pro ligand, the relative firefly luciferase activity was 4.13 and 2.04-fold higher ($P < 0.001$), respectively, than the pcDNA 3.1 empty vector control (**Figure 4.7A**). The gene reporter activity was not dose-dependent as increasing plasmid concentrations did not significantly alter the luciferase activity (**Figure 4.7B**). In HEK 293T cells co-expressing the wildtype and Leu70Pro mutant ligand, the gene reporter activity was 2.17-fold higher than the control vector (**Figure 4.7A**), which was comparable to cells expressing the Leu70Pro ligand alone (**Figure 4.7C**). The luciferase activity was not a result of altered protein expression as wildtype, mutant and coexpressed plasmids were stably and equivalently expressed in HEK 293T cells (**Figure 4.7D**). The luciferase activity, resulting from the independent and co-expression of both the wildtype and mutant ligand, suggests that the Leu70Pro ligand has a dominant negative effect on WNT8B-activated beta-catenin mediated WNT signalling pathway.


WNT8B wildtype
T7
WNT8B ORF
V5
P
β-Cat
WNT8B L70P
?
TOPFlash
β-Cat
TCF
TCF
TCF
LUCIFERASE
FOPFlash
TXF
TXF
TXF
MCS
pcDNA™/V5-DEST
Ampicillin
pUC ori
SV40 pA
Neomycin
SV40 ori
f1 ori
TK pA
P CMV
MCS (1–41)
Hind III (107)
TB
f1 ori
Luc
pTA-Luc
4.8 kb
Amp
SV40 poly A
pUC ori

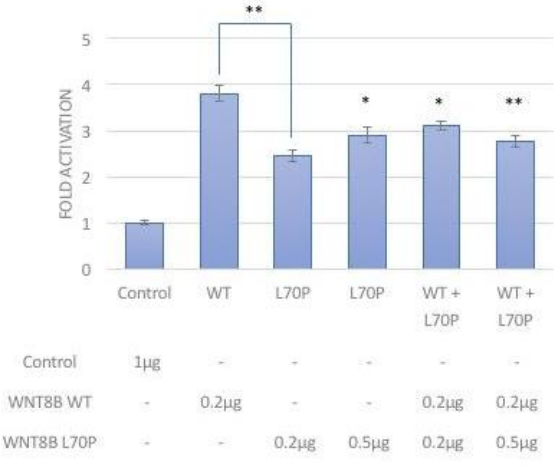
A


FOLD ACTIVATION
TOP
FOP
***
***
Control
WT
L70P
WT + L70P
Control 1µg - - -
WNT8B WT - 1µg - 0.75µg
WNT8B L70P - - 1µg 1µg

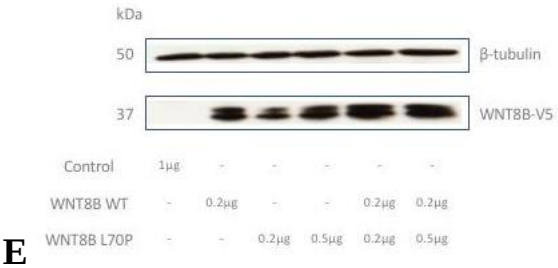
B


FOLD ACTIVATION
***
Control
WT
WT
WT
L70P
L70P
L70P
Control 1µg - - - - - -
WNT8B WT - 0.2µg 0.5µg 1µg - - -
WNT8B L70P - - - - 0.2µg 0.5µg 1µg

C


FOLD ACTIVATION
**
*
*
**
Control
WT
L70P
L70P
WT + L70P
WT + L70P
Control 1µg - - - - -
WNT8B WT - 0.2µg - - 0.2µg 0.2µg
WNT8B L70P - - 0.2µg 0.5µg 0.2µg 0.5µg

D


kDa
50
β-tubulin
37
WNT8B-V5
Control 1µg - - - - -
WNT8B WT - 0.2µg - - 0.2µg 0.2µg
WNT8B L70P - - 0.2µg 0.5µg 0.2µg 0.5µg

E

Figure 4.7: WNT8B Activation of Canonical Wnt Signalling. A: Schematic diagram showing the plasmid maps of the Wnt ligands (left) and reporter constructs (right). **B:** Pilot study showing the fold activation of the luciferase activity from HEK 293T cells expressing the TOPFlash or FOPFlash vector. **C:** Gene reporter activity of the WNT8B wildtype or mutant L70P expressed in HEK 293T cells across a concentration gradient. **D:** Independent and co-transfected luciferase activity of the wildtype and mutant WNT8B ligand in HEK 293T cells across three replicate experiments. **E:** Immunoblot showing stable expression of plasmids across different concentrations. Proteins were extracted from cell lysates from experiment (**D)** of which 20μg were used for western blotting. β-tubulin antibody (Abcam, Cat # ab21058) was used as loading control, while anti-V5 antibody (Bethyl Lab, Cat # A190-120P) was used to probe for WNT8B-tagged (C-terminal V5) protein. No other bands were observed on the blot other than those shown. Asterisks represent Student's t-test P-values (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). Error bars represent standard deviations across replicate experiments (n => 3).

4.2.6 Morpholino-mediated knockdown of Aicardi candidate genes in zebrafish embryos.

For the remaining genes that did not have established molecular or biochemical assays and are yet to be implicated in both eye and brain development, we used morpholino knockdown in zebrafish (*Danio rerio*) embryos to screen for phenotypes that were reminiscent of AIC. We included tead1, which is the only other previously identified gene implicated AIC that is conserved in zebrafish. We designed a study in which the translation-blocking morpholino injections and data analysis were carried out under blinded conditions and morphant classification counts were verified by three independent researchers. More than 90% of uninjected and scrambled morpholino control group displayed normal embryonic development with or without minor body abnormalities such as bent tail (**Figure 4.8**).

In contrast, our experimental morphants displayed a spectrum of developmental abnormalities including: lack of eye pigmentation, body curvatures, pericardial oedema, tail defects and head malformations (**Figure 4.8B**). The most consistent phenotype among the slf1, tead1 and wnt8b morphants was the lack of eye pigmentation, which was patchy in appearance and often unilateral (**Figure 4.8A**). A lack of eye pigmentation was also observed in szt2 morphants, however as 72 hours post fertilization (hpf) szt2 morphants often resembled 48 hpf morphants, this may be attributed to developmental delay. The eye phenotype is similar to the absence of mature melanosomes observed from 48hpf slc24a5 and tyrp1a/b morphants (Lamason et al., 2005, Takamiya et al., 2016). The only exception to the experimental group were hcn1 morphants, whom displayed normal phenotype similar to the uninjected and control group (results not shown). There are known limitations to disease modelling in zebrafish (Kok et al., 2015) and the morpholino knockdown approach models a loss of function whereas our *de novo* variants are likely to be dominant negative. Nonetheless, we showed a unifying morphant phenotype of AIC-like eye and brain defects among at least 3/5 genes tested.

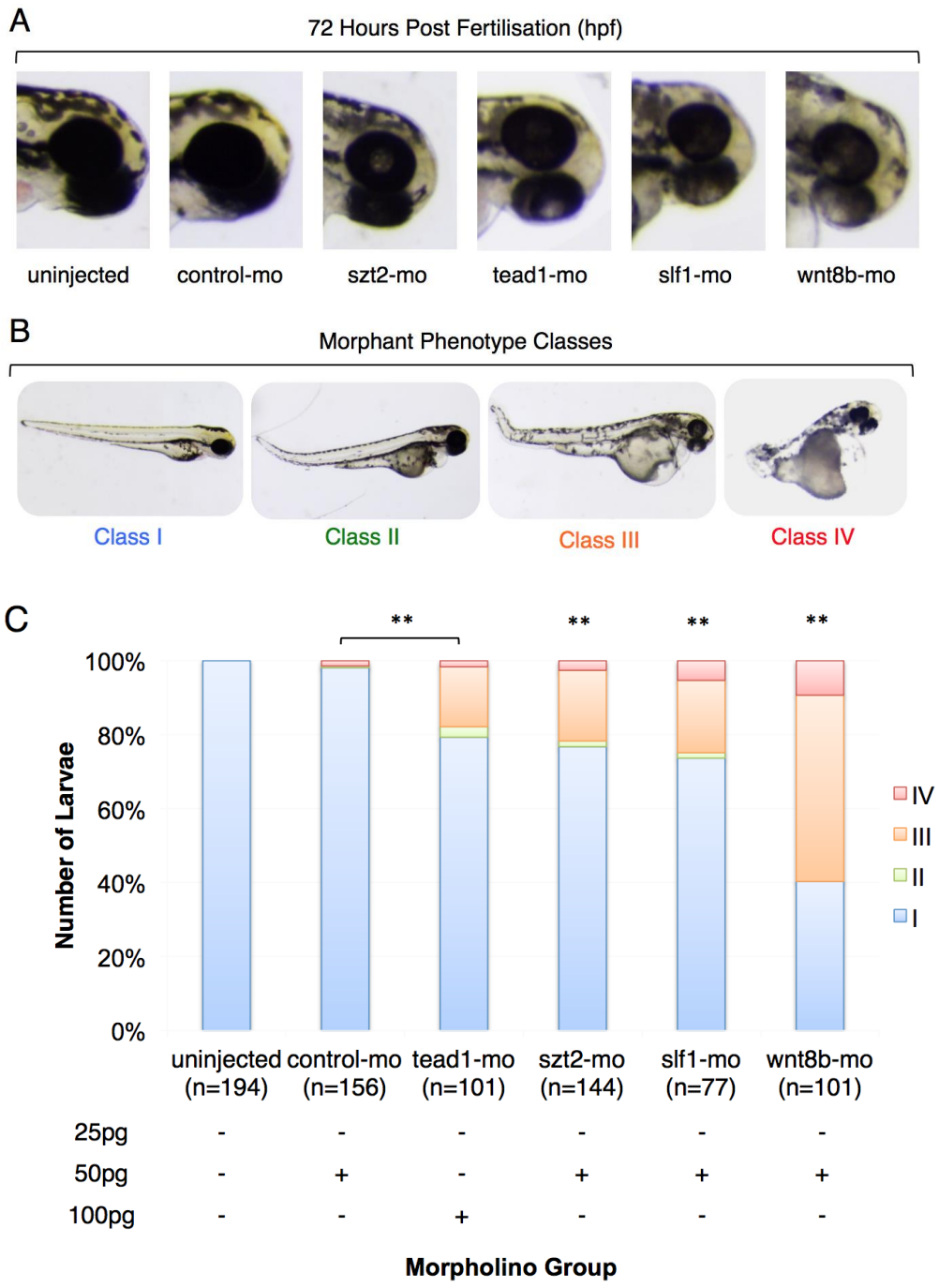


Figure 4.8: Phenotype analysis of morpholino mediated knockdown in zebrafish embryos. A: Ocular morphology of zebrafish embryos at 72hpf either uninjected or injected with a control morpholino or gene-targeted morpholino. B: Zebrafish embryos were scored according to eye classifications by Miesfeld et al. (2015). Class I: normal phenotype; Class II: normal eye pigmentation with mild body defects; Class III: reduced eye pigmentation with normal to mild body defects; Class IV; reduced eye pigmentation and severe body defects. C: Observed

morphant phenotypes after blinded injections of 240 zebrafish embryos per group. The number of viable embryos remaining after 72 hpf are denoted by (n). The amount of morpholino injected per group is indicated by (+).

4.3 Discussion

4.3.1 Lack of X-linked pathogenic variants

We identified likely pathogenic variants in different autosomal genes from five unrelated individuals. There are possible X-linked causes that were not discovered from other and our genetic studies due to: (1) the paired-end sequencing (PES) approach, (2) 40X sequencing coverage, (3) peripheral blood as DNA source and (4) interpretation of variants in non-coding sequences. PES has limited resolution for variants in repetitive DNA sequences, regions with high or low GC bias and non-random DNA fragmentation bias (Treangen et al. 2011; Aird et al. 2011). Terminal, tandem and interspersed repeats have yet to be investigated as a cause of AIC. Our 38X average sequencing coverage was also inadequate for detecting somatic variants with low levels of mosaicism (less than 5% of allelic fraction); which could be disguised as false positives. Furthermore, a majority of our sequencing was based on peripheral blood DNA, thus we were unable to comprehensively screen for brain-specific variants. Finally, due to the lack of paired WGS and RNA sequencing data, we were unable to prioritise and interpret the significance of some non-coding variants; especially in regions lacking functional annotations. For the remaining unresolved individuals in our cohort, an X-linked cause may still be possible, but potentially masked due to the technical limitations listed above.

4.3.2 Variants identified in known neurodevelopmental disorder genes

Six *de novo* mutations in *HCN1*, encoding the hyperpolarization-activated cyclic nucleotide-gated channel 1, that cause early onset epileptic encephalopathy (EIEE; MIM 615871) were described previously by Nava et al. (2014). Proband T22352, in whom we identified the NM_021072.3 (*HCN1*);c1625G>T; p.Cys542Phe variant, has partial agenesis of the corpus callosum, absence of the anterior body, genu and rostrum, hypoplasia of splenium, Lennox Gastaut syndrome and bilateral paroxysmal fast activity. Her MRI brain findings and EEG abnormalities are not present in other individuals with *HCN1* variants. Nonetheless, there were also similar clinical features in proband T22352 to other individuals reported such as autism spectrum disorder and seizures beginning during the first year of life.

Functional differences between HCN1 p.Cys542Phe and previously reported mutations are likely due to the location of the affected amino acid. Mutations reported by Nava et al. were located in the transmembrane domain or the intracellular domain closer to the N-terminus (Nava et al., 2014). Patch clamp recordings in CHO-K1 cells revealed a majority of these human *HCN1* mutations lead to a gain of function or dominant negative effects. Meanwhile, the HCN1 p.Cys542Phe we identified is located in the cyclic nucleotide binding domain, which is responsible for the direct regulation of the channel function (**Figure 4.9**) (Wainger et al., 2001). We conducted voltage clamp recordings in Xenopus oocytes to measure the gating function of the HCN1 p.Cys542Phe mutant channel. We found that the oocytes expressing HCN1 p.Cys542Phe left-shifted the voltage dependence of activation resulting in a loss of function (**Figure 4.5**). In light of our findings and other published cases, the diagnosis for proband T22352 was revised from Aicardi-like syndrome to a HCN1-EIEE.

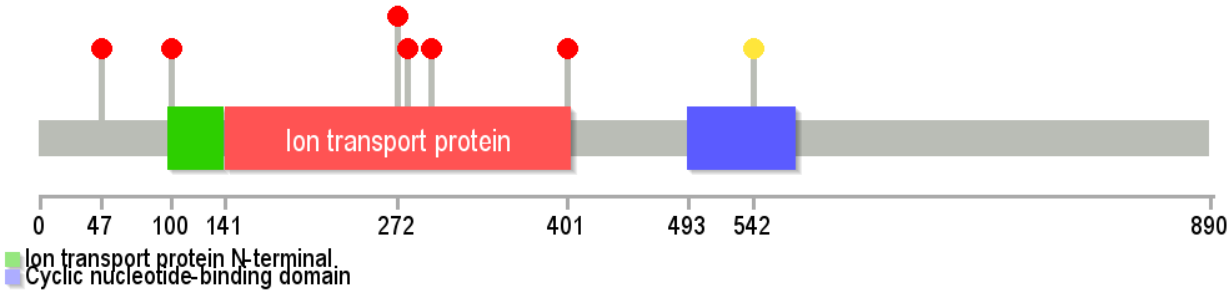


Figure 4.9: HCN1 variants in affected individuals with epileptic encephalopathy. Lollipop plot showing published HCN1 variants from Nava et al. (2014) in red and the Cys542Phe variant from Proband T22352 in yellow.

Mutations in *SZT2*, encoding seizure threshold 2, are associated with autosomal recessive EIEE and non-syndromic intellectual disability (Nakamura et al., 2018, Tsuchida et al., 2018, Venkatesan et al., 2016, Falcone et al., 2013, Basel-Vanagaite et al., 2013) (**Figure 4.10**). We discovered a *de novo* NM_015284 (*SZT2);* c.9103C>T; p.His3035Tyr variant in proband T25217 who displayed: focal cortical dysplasia, polymicrogyria, subcortical heterotopia, modified hypsarrhythmia, developmental delay, ID and hypotonia. The brain MRI findings and neurological anomalies present in proband T25217 share some overlap with the phenotypes described in patients with SZT2. Although the disease mechanism underlying pathologies associated with mutations in *SZT2* has yet to be elucidated, it was recently discovered that SZT2 has a role in GATOR-dependent nutrient sensing and MTORC1 regulation (Peng et al., 2017). SZT2 is part of the KICSTOR complex, which is a negative regulator of MTORC1 signalling. Coincidentally, the knockout of SZT2 in mice had led to an increase in MTORC1 signalling in tissues such as neurons in the brain (Wolfson et al., 2017). Hyperactivity of MTORC1 signalling has many physiological impacts including: neuronal death, axon regeneration, abnormal cell differentiation and morphogenesis (Bockaert and Marin, 2015). These physiological changes are involved in a number of MTORC1-associated brain pathologies listed in **Table 4.5**.

The SZT2 protein is highly conserved across multiple species and has two predicted functional domains: superoxide dismutase and peroxisomal targeting signal (Toutzaris et al., 2010). It is ubiquitously expressed with its highest expression in the CNS (Frankel et al., 2009 & Toutzaris et al., 2010). From an N-ethyl-N-nitrosourea (ENU) mutagenesis screen in mice, a majority of *Szt2* homozygotes died embryonically (Frankel et al., 2009). Those that survived were more susceptible to induced seizures and potentially epileptogenic compared to wildtype mice. Single *Szt2* heterozygotes had less seizures compared to compound *Szt2* heterozygotes, but more compared to wildtype mice; which supports allelism and dominant effects respectively (Frankel et al., 2009). The clinical spectrum associated with *SZT2* mutations are phenotypically heterogeneous, but mostly bi-alleic in cases reported so far. Given that a second variant was not identified in *SZT2* from patient T25217, the likely pathogenicity of the p.His3035Tyr variant remains unknown until further functional validation. The latter may involve measuring the effect of the p.His3035Tyr variants on MTORC1 signalling.

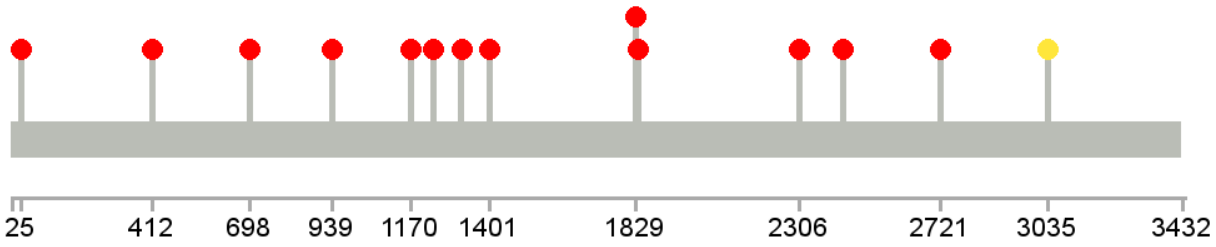


Figure 4.10: SZT2 variants in affected individuals with various neurodevelopmental disorders. Lollipop plot showing the biallelic SZT2 variants reported thus far (red) and the His3035Tyr variant from proband T25217 (yellow). Image was adapted from Figure 1 of the Tsuchida et al. (2017) study.

Brain Disease[21]	Mutated Gene	Phenotype
Tuberous sclerosis	TSC	● SEGA cells ● Benign tumours ● Epilepsies ● Intellectual disability ● ASD
Polyhydramnios-megalencephaly symptomatic epilepsy syndrome	STRADα	● Cranio-facial dysmorphism ● Large brain ● Intractable epilepsy
Familial focal epilepsies with variable foci	DEPDC5	● Intractable epilepsy
Hemimegaloencephaly	AKT3, PI3K and MTOR	● Cerebral asymmetry, laminar cortical abnormalities ● Intractable epilepsy
Ganglioglioma	BRAF	● Giant atypical ganglion cells ● Intractable epilepsy
Cowden syndrome	PTEN	● Multiple haematomas ● Uncontrolled neurogenesis ● Cortical dysplasia
Lhermitte-Duclos disease	PTEN	● Dysplastic gangliocytoma of the cerebellum
Lafora disease	EPM2A	● Accumulation of polyglucosan inclusions ● Myoclonus epilepsy
Brain tumours	PTEN and TSC	● Glial cell proliferation
ASD, Down syndrome, neurofibromatosis	TSC, NF1	● Cognitive deficits ● Decreased autophagy

Table 4.5: Mutations linked to hyperactivation of MTOR signalling in brain pathologies.

[21] Table was modified from Table 2 of the Bockaert and Marin (2015) study.

From patient T22842, a *de novo* (c.6418C>G; p.Pro214Ala; NM_014727) variant was identified in *KMT2B*, encoding the lysine specific methyltransferase 2B (**Table 4.5 & Figure 4.11**) (Zech et al., 2016). In a cohort of 31 patients with dystonia, four missense variants were identified in *KMT2B* in four unrelated probands; three were *de novo* and one was inherited (Zech et al., 2016). In addition to generalized dystonia, other phenotypes observed from these four probands included: microcephaly, short stature, delayed speech, normal to mild cognitive impairment. None of the four probands displayed neurological abnormalities or brain MRI findings. However, the phenotype spectrum of KMT2B-related disorders was later expanded by Meyer et al. (2017) to include for developmental delay, intellectual disability, dysmorphic features and other isolated systemic manifestations. These clinical features were based on 17 unrelated patients with previously unreported missense and nonsense variants in *KMT2B*. Some of the clinical features in patient T22842 were not commonly observed in patients with KMT2B-related disorders, such as: chorioretinal lacunae (**Figure 1.1**), infantile spasm, tonic-clonic seizures and choroid plexus papilloma (**Table 4.2**).

From some of the unrelated probands with nonsense variants in *KMT2B*, there were reduced expression of *KMT2B* in patient fibroblasts, which supports haploinsufficiency as a disease mechanism (Zech et al., 2016, Meyer et al., 2017). However, for the remaining intragenic missense variants in *KMT2B*, including our Pro214Ala variant, the molecular mechanism underlying the pathogenesis of KMT2B-related disorders remains unknown. In the Meyer et al. (2017) study, there were no differences in histone H3K4 methylation between patient and control cell lines. However, in three unrelated probands, there were reduced expression of *THAP1* and *TOR1A*, which suggests some *KMT2B* variants may alter the expression profiles of other known dystonia genes. Given the lack of candidate genes, it is currently unknown whether our

Pro214Ala variant can alter the expression profiles of genes involved in AIC or AIC-related disorders.

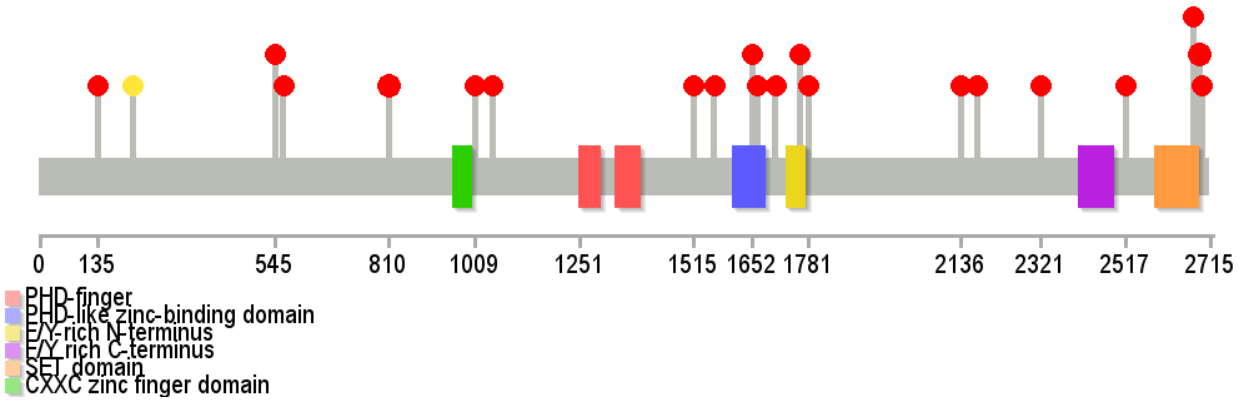


Figure 4.11: KMT2B variants in affected individuals with dystonia. Lollipop plot showing published KMT2B variants from Meyer et al. (2017) and Zech et al. (2016) in red and the Pro214Ala variant from proband T22842 (yellow).

4.3.3 Variants identified in novel disease genes

In *WNT8B,* encoding the wingless-type MMTV integration site family member 8B, we discovered a *de novo* (c.209T>C; p.Leu70Pro; NM_003393) variant in patient T22101 (**Table 4.1 & 4.2**). There are 19 Wnt ligands that can stimulate Wnt signalling pathways including: the canonical β-catenin dependent pathway, the planar cell polarity pathway and the Wnt/Ca^{2+} pathway (Komiya and Habas, 2008). The Wnt signalling pathways play a major role in embryonic development. Disruptions to the Wnt signalling can result in lethal and severe consequences such as cancers, skeletal defects and congenital disorders (Logan and Nusse, 2004). To date, no human pathologies have been associated with *WNT8B* mutations (**Figure 4.12**).

The expression of *WNT8B* appears to be restricted to the developing forebrain, particularly the hippocampus and hypothalamus (Lako et al., 1998). Knockout mouse studies in other Wnt genes show phenotypes ranging from embryonic lethality to complete absence of brain structures (van Amerogen and Berns, 2006). The *Wnt8b* knockout mouse showed no gross morphological differences from the wildtype mouse; but did alter the expression levels of other Wnt genes present in the developing forebrain (Fotaki et al., 2010). There is an overlap between the expression domain of *Wnt8b* and the other Wnt genes, which may provide compensatory functions; thus diminishing the gross effects of *Wnt8b* deficiency (Fischer et al., 2007). Morpholino knockdown of wnt8b in the zebrafish showed commissural defects in structures of the zebrafish brain akin to the corpus callosum and demonstrate the role of Wnt8b in regulating axon guidance (Hofmeister and Key, 2013). In zebrafish and mice, *Wnt8b* has been shown to be a negative regulator of eye development, particularly the specification of retinal progenitor cells (Cavodeassi et al. 2005 & Liu et al. 2010). Although mutations in *WNT8B* have yet to be implicated in epileptic encephalopathy, its role in eye and forebrain development makes it a potential candidate gene for AIC.

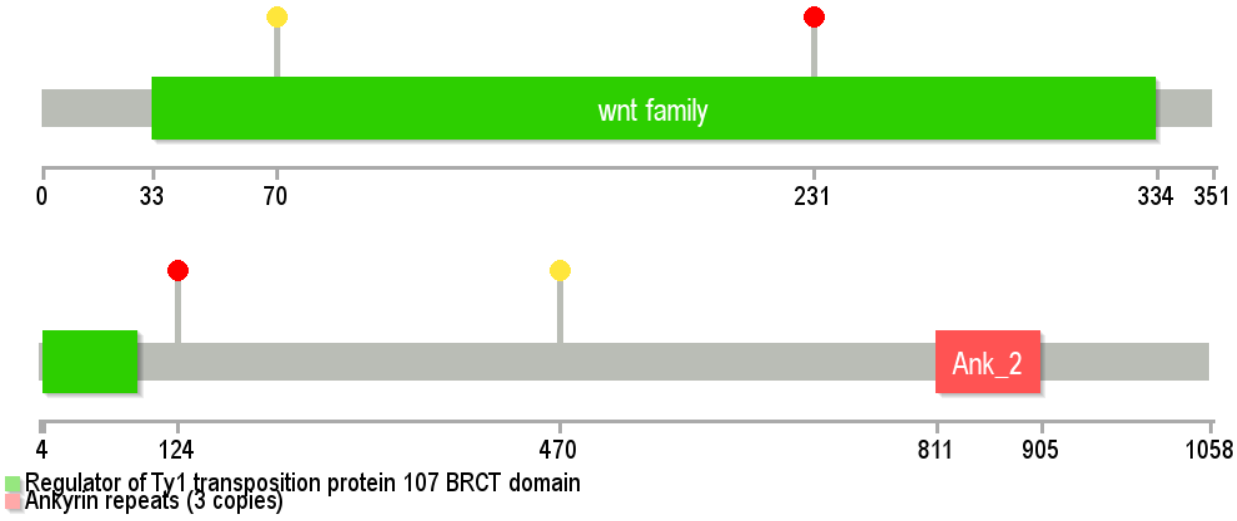


Figure 4.12: *WNT8B* and *SLF1* variants in human disorders. Lollipop plot showing the *WNT8B* (top) and *SLF1* (bottom) variants reported in Lemay et al. (2015) and Sifrim et al. (2016)

respectively (red). For each respective gene, the variants reported in patient T17262 and T22101 are shown in yellow.

In *SLF1*, encoding the SMC5-SMC6 complex localization factor protein 1, we found a (c.1409T>C, p.Leu470Ser; NM_032290) variant in patient T17262 (**Table 4.1 & 4.2**). *SLF1* has primarily been investigated for its potential role in cancer (Liu et al., 2012 & Adams et al., 2005). The *SLF1* gene contains a breast cancer gene 1 carboxyl-terminal (BRCT) domain. There are 23 human genes with a BRCT domain and the knockout of these genes have resulted in embryonic lethality, tumor formation, infertility and growth abnormalities (Liu et al., 2012, Adams et al., 2005, Sharan et al., 1997). However, the *Slf1* knockout mouse showed no histopathological differences from the wildtype mouse (Adams et al., 2005). From a yeast two hybrid screen, SLF1 was shown to interact and co-localise with RAD18, an E3 ligase that is involved in DNA damage signalling pathways. The *Slf1* knockout mouse did not display any DNA repair impairment, genomic instability or DNA growth defects. In a corresponding study, SLF1 and RAD18 interaction was revealed to be important in UV-induced DNA damage repair (Liu et al., 2012). However, SLF1 alone and its role in tumorigenesis has yet to be determined. The SLF1 p.Leu470Ser variant identified in proband T17262 lies outside of the BRCT domain, thus its functional significance has yet to be determined (**Figure 4.12**).

4.3.4 Identifying additional carriers of variants in putative Aicardi genes

To provide additional evidence to support our implicated AIC genes as true disease drivers rather than incidental findings, we sought additional carriers in these genes using: biomedical literature (Pubmed), disease cohorts (Epi4K and DDD) and genotype-to-phenotype matching platforms (Denovo-db, Geno2mp and GeneMatcher) (**Table 4.6**). Variants in *HCN1* and *SZT2* has been reported in individuals with early infantile epileptic encephalopathy (MIM 615871 & MIM

614576). Meanwhile *KMT2B* variants have been implicated in childhood onset dystonia (MIM 617284). While the inheritance mode (*de novo*) of published *HCN1* and *KMT2B* variants overlap with probands T22352 and T22842 respectively, the congenital disorders described in the reported affected individuals share minimal to no phenotypic overlap with our cohort. Conversely, there are some overlap in the neurodevelopmental phenotype of reported individuals with *SZT2* variants and proband T25217, but a majority of reported cases are autosomal recessive versus *de novo* in ours. Variants in *SLF1* and *WNT8B* have only been identified in two different individuals, who both display different clinical characteristics to proband T17262 and T22101 respectively. Finding a second variant in our implicated Aicardi genes from disease cohorts and online patient archives was also unsuccessful; either the inheritance or phenotype was unmatched with individuals in our cohort.

4.3.4 Protein subcellular localisation and interaction of candidate genes

From published immunohistochemistry (IHC) data, we observed that 4/5 protein products of the candidate AIC genes (excluding WNT8B) were localized to the nucleus or nucleoplasm (**Figure 4.13**) (Uhlen et al. 2010). Despite similar subcellular localisation, none of the implicated genes have been predicted to interact on a genetic or protein level (Szklarczyk & Jensen 2015; Chatr-Aryamontri 2016). This may be a reflection of the diverse molecular function and biological processes of the genes harbouring the DNV (**Table 4.7**).

Reference	Cohort/ Disease/ Gene	Molecular Tool	Cohort Number	*HCN1*	*KMT2B*	*SLF1*	*SZT2*	*WNT8B*
EpiK Consortium et al., 2013	EE	WES	264 pros	-	-	-	-	-
Yang et al., 2013	Mendelian Disorders	WES	250 pros		-	-	-	-
Falcone et al,. 2013	*SZT2*	WES	3 sibs	-	-	-	1	-
Basel-Vanagaite et al., 2013	*SZT2*	WES	2 pros	-	-	-	3	-
EuroEpinomics et al., 2014	EE	WES	356 trios	-	1	-	-	-
Wright et al., 2015	Developmental Delay	Array & WES	1133 pros		-	-	-	-
Schrauwen et al., 2015	Aicardi Syndrome	WES & WGS	10 trios	-	-	-	-	-
Lemay et al., 2015	Neural Tube Defects	WES	43 trios		-	1	-	-
Sifrim et al., 2016	Congenital Heart Defects	WES	1891 pros	-	-	-	-	1
Lelieveld et al., 2016	Intellectual Disability	WES	2104 trios	-	1	-	-	-
Venkatesan et al., 2016	*SZT2*	WES	1 pros	-	-	-	1	-
Zech et al., 2016	*KMT2B*	WES	30 pros	-	4	-	-	-
Lund et al., 2016	Aicardi Syndrome	XES	10 trios	-	-	-	-	-
DDD et al., 2017	Developmental Delay	WES	4293 trios	-	-	-	2	
Posey et al., 2017	Multilocus Diseases	WES	7374 pros	-	-	-	-	-
Meyer et al., 2017	*KMT2B*	Array, WES & WGS	27 pros	-	27	-	-	-
Wong et al., 2017	Aicardi Syndrome	WES	38 pros	-	-	-	-	-
Tsuchida et al., 2018	EE	WES	3 pros	-	-		6	-

Table 4.6: Additional carriers of variants in putative Aicardi genes reported in affected individuals. [22]

[22] EE, epileptic encephalopathy; pros, probands; sibs, siblings; trios, parent-proband trios; WES, whole exome sequencing; WGS, whole genome sequencing; XES, chromosome X sequencing.

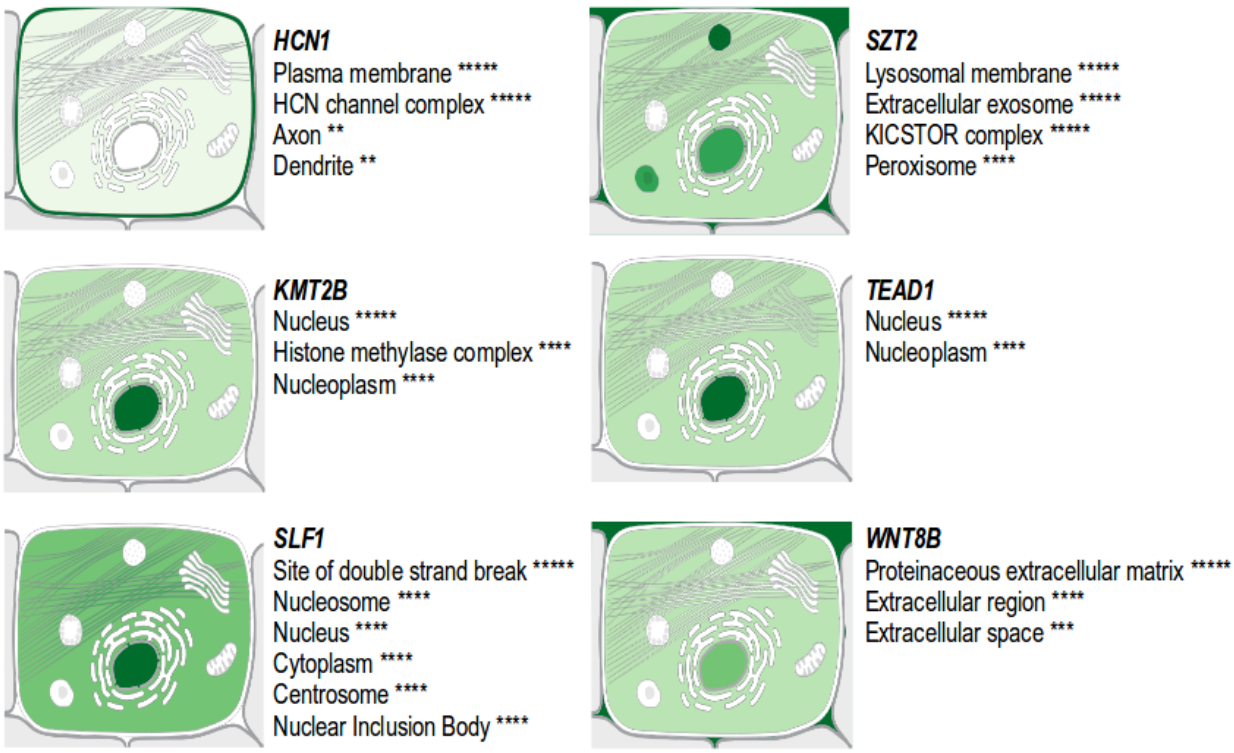


Figure 4.13: Subcellular localisation of Implicated Aicardi genes. This figure shows the localisation of the target genes (bold) derived from database annotations. The confidence of each association is signified by stars, where five stars is the highest confidence and one star is the lowest. Images were obtained from the Jensen Lab 'Compartments' site (https://compartments.jensenlab.org/Search).

A. GO - Molecular Function	***HCN1***	***KMT2B***	***SLF1***	***SZT2***	***TEAD1***	***WNT8B***
cAMP binding	yes					
core promoter proximal region DNA binding					yes	
DNA binding		yes			yes	
frizzled binding						yes
histone-lysine N-methyltransferase activity		yes				
histone methyltransferase activity (H3-K4 specific)		yes				
identical protein binding	yes					
potassium channel activity	yes					
protein complex binding			yes			
protein heterodimerisation activity					yes	
receptor ligand activity						yes
RNA polymerase II transcription factor binding					yes	
sequence-specific DNA binding					yes	
transcription coactivator binding					yes	
transcription factor activity, RNA polymerase II core promoter proximal region					yes	
transcription factor activity, sequence-specific DNA binding		yes			yes	
transcription factor activity, transcription factor recruiting					yes	
ubiquitin protein ligase binding			yes			
voltage-gated cation channel activity	yes					
voltage-gated potassium channel activity	yes					
voltage-gated sodium channel activity	yes					
zinc ion binding		yes				

B. GO - Molecular Function	***HCN1***	***KMT2B***	***SLF1***	***SZT2***	***TEAD1***	***WNT8B***
apical protein localisation	yes					
beta-catenin destruction complex disassembly						yes
canonical Wnt signalling pathway						yes
cell fate commitment						yes
cellular response to amino acid starvation				yes		
cellular response to cAMP	yes					

cellular response to DNA damage stimulus			yes			
cellular response to glucose starvation				yes		
cellular response to retinoic acid						yes
central nervous system development				yes		
chromatin-mediated maintenance of transcription		yes				
corpus callosum morphogenesis				yes		
determination of dorsal identity						yes
DNA repair			yes			
embryonic organ development					yes	
gastrulation						yes
gene silencing		yes				
hippo signalling					yes	
histone H3-K4 methylation		yes				
histone H3-K4 trimethylation		yes				
memory		yes				
negative regulation of TORC1 signalling				yes		
nervous system development						yes
neuron differentiation						yes
oocyte differentiation		yes				
ovarian follicle development		yes				
ovulation		yes				
pigmentation				yes		
positive regulation of double strand break repair			yes			
positive regulation of maintenance of mitotic sister chromatid cohesion			yes			
positive regulation of pri-miRNA transcription from RNA polymerase II promoter					yes	
positive regulation of protein complex assembly			yes			
positive regulation of transcription, DNA-templated					yes	
post embryonic development				yes		
potassium ion transmembrane transport	yes					
protein complex assembly					yes	
protein homotetramerisation	yes					
protein localisation to lysosome				yes		

protein localisation to site of double-strand break			yes			
regulation of histone H3-K4 methylation		yes				
regulation of megakaryocyte differentiation		yes				
regulation of membrane potential	yes					
regulation of superoxide dismutase activity	yes			yes		
response to estradiol						yes
response to retinoic acid						yes
retinal cone cell development	yes					
signal transduction						yes
sodium ion transmembrane transport	yes					
transcription, DNA-templated		yes				
transcription, initiation from RNA polymerase II promoter					yes	
Wnt signalling pathway						yes

Table 4.7: Gene ontology for implicated Aicardi genes. This table summarises the predicted molecular function (**A**) and biological role (**B**) for each gene of interest (http://www.geneontology.org/).

4.3.5 Early developmental expression of candidate genes

Next, we explored whether the implicated genes converge on developmental processes (temporal or spatial) that can contribute to AIC phenotypes. This was observed for nine recognised autism spectrum disorder genes, harbouring DNM identified from WES and WGS studies that converged in the same time point, brain region (mid-foetal layer) and neuronal cell type (cortical projection neurons) Willsey et al. (2013). From human RNA-Seq data across multiple somatic tissues, the expression of *SLF1* and *WNT8B* are sparse; while *KMT2B*, *SZT2* and *TEAD1* are ubiquitously expressed (**Figure 4.14**). In the developing human brain, there is restricted and low expression of *WNT8B* across different regions; while the remaining genes are consistently expressed, but at different magnitudes depending on the tissue type. From the current publicly available transcriptome data, our implicated AIC genes do not appear to converge at a particular

region at whole brain resolution. However, given the heterogeneity and complexity of brain development, our target genes may need to be investigated at single-cell resolution across different tissues and stages to identify potential molecular signatures.

Only some of the target genes have been studied during similar embryonic stages, thus we turned to murine embryonic data to scan for overlapping developmental profiles. We observed that *Kmt2b* and *Tead1* are ubiquitously expressed throughout murine embryonic development; while *Slf1, Szt2* and *Wnt8b* expression is restricted to the brain (Smith et al., 2007). At embryonic day 14.5, the Slf1 and Wnt8b mouse display similar expression in the developing forebrain (**Figure 4.15 & 4.16**) (Diez-Roux et al., 2011, Visel et al., 2004). Both *Wnt8b* and *Tead1* are expressed during early eye development, while *Kmt2b* is expressed later during retinal development (Ang et al., 2004, Blackshaw et al., 2004, Visel et al., 2004). Although the role and expression of *Slf1* and *Szt2* during eye development has yet to be documented, our zebrafish data indicates a potential role of *slf1* in early eye development (**Figure 4.7**). We observed that all genes except for *Szt2* are expressed in the murine brain during 10.5 to 15.5 days of embryonic development (ED) and approximates to the 4th to 7th week of human gestation. This gestational stage coincides with the predicted emergence of eye, brain and skeletal abnormalities observed in AIC; and the divergence of male and female developmental pathways (Rinn and Snyder, 2005, Menezes et al., 1996).

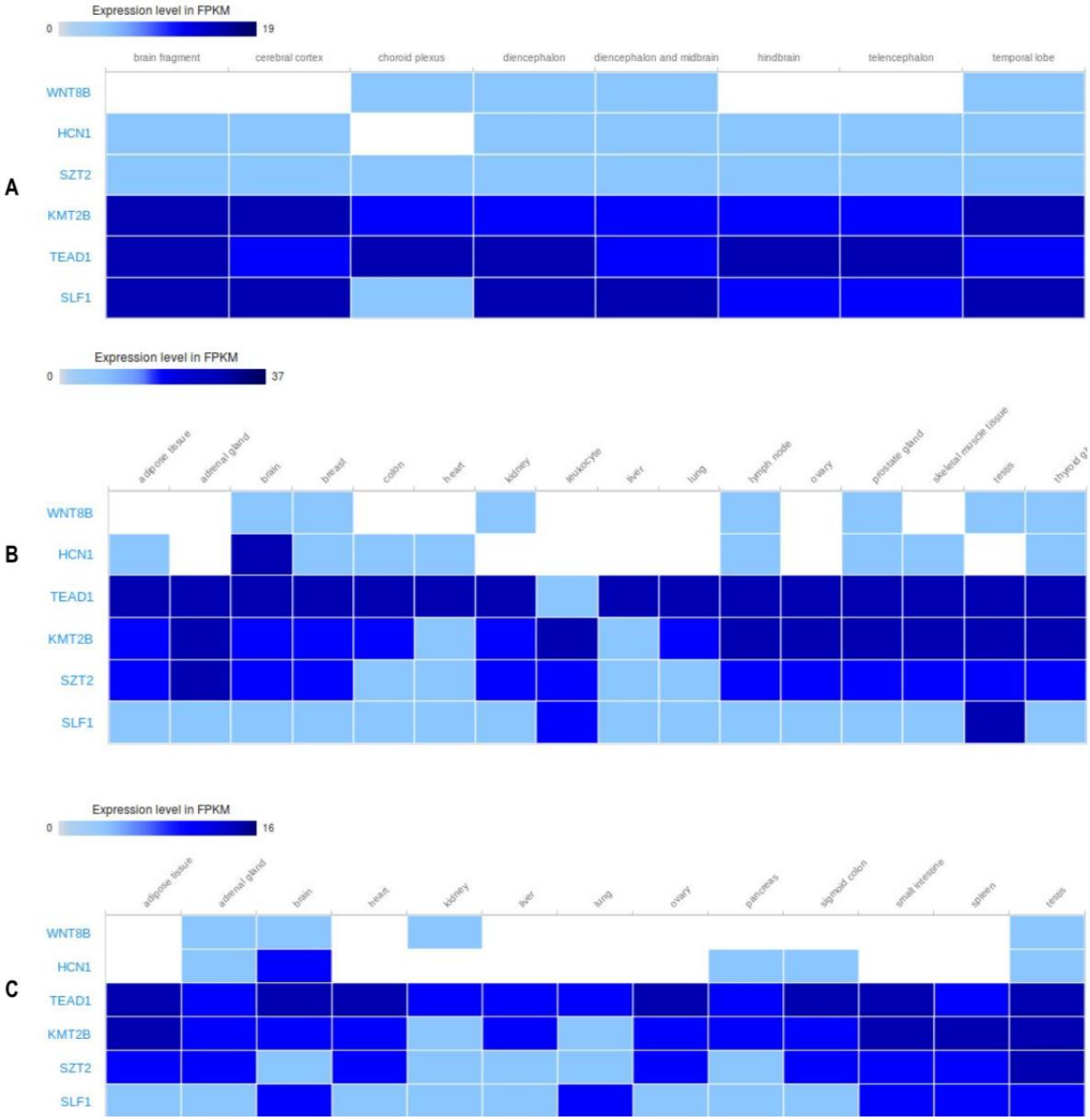


Figure 4.14: RNA-Seq of target Aicardi genes from published datasets. This figure shows the expression level of each target gene (left side, blue writing) across different tissue types. **A:** The expression level of coding RNA across different prenatal human brain regions (source: Human Developmental Biology Resource). **B:** The expression level of the target genes across individual and mixtures of 16 human tissues (source: Illumina Body Map). **C:** The strand-specific expression of the target genes across 16 different human tissue sources: Michael Snyder's lab for the ENCODE project). FPKM, Fragments Per Kilobase of transcript per Million mapped reads.

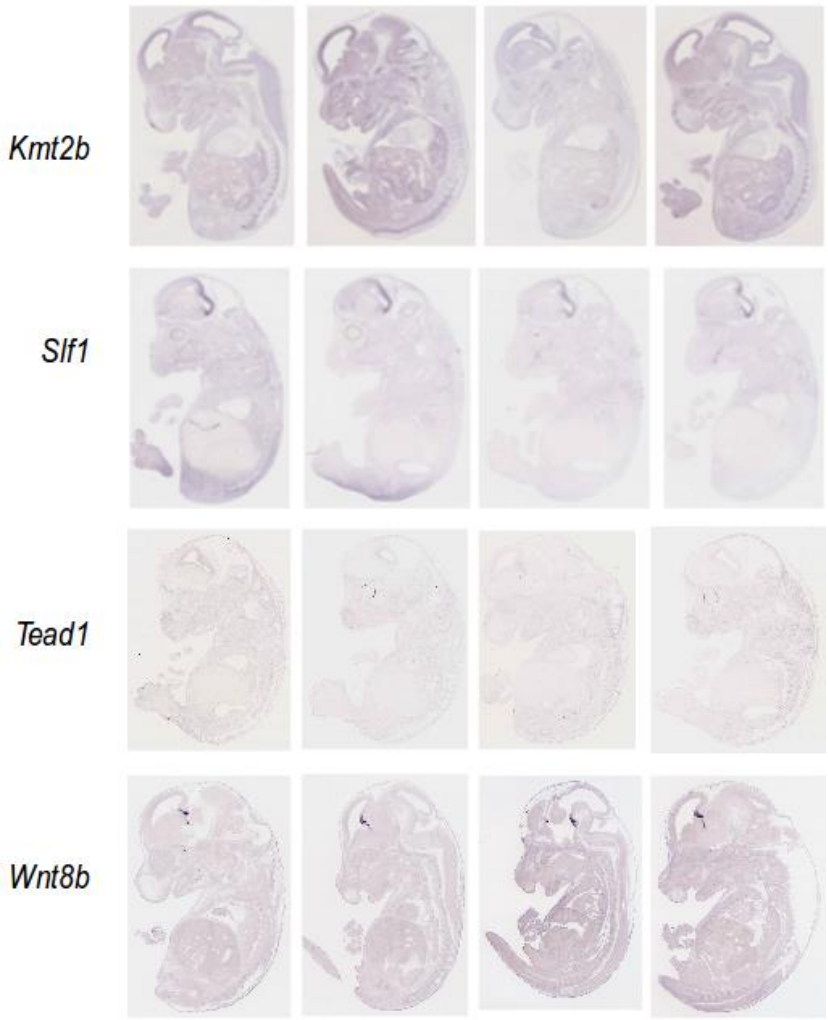


Figure 4.15: RNA in situ hybridisation in mouse embryo at day 14.5 post conception. This figure shows the RNA localisation of Kmt2b, Slf1, Tead1 and Wnt8b across whole mount in situ hybridisation of E14.5 mouse embryos. The images were derived from Visel et al. (2004) and Diez-Rouz et al. (2011).

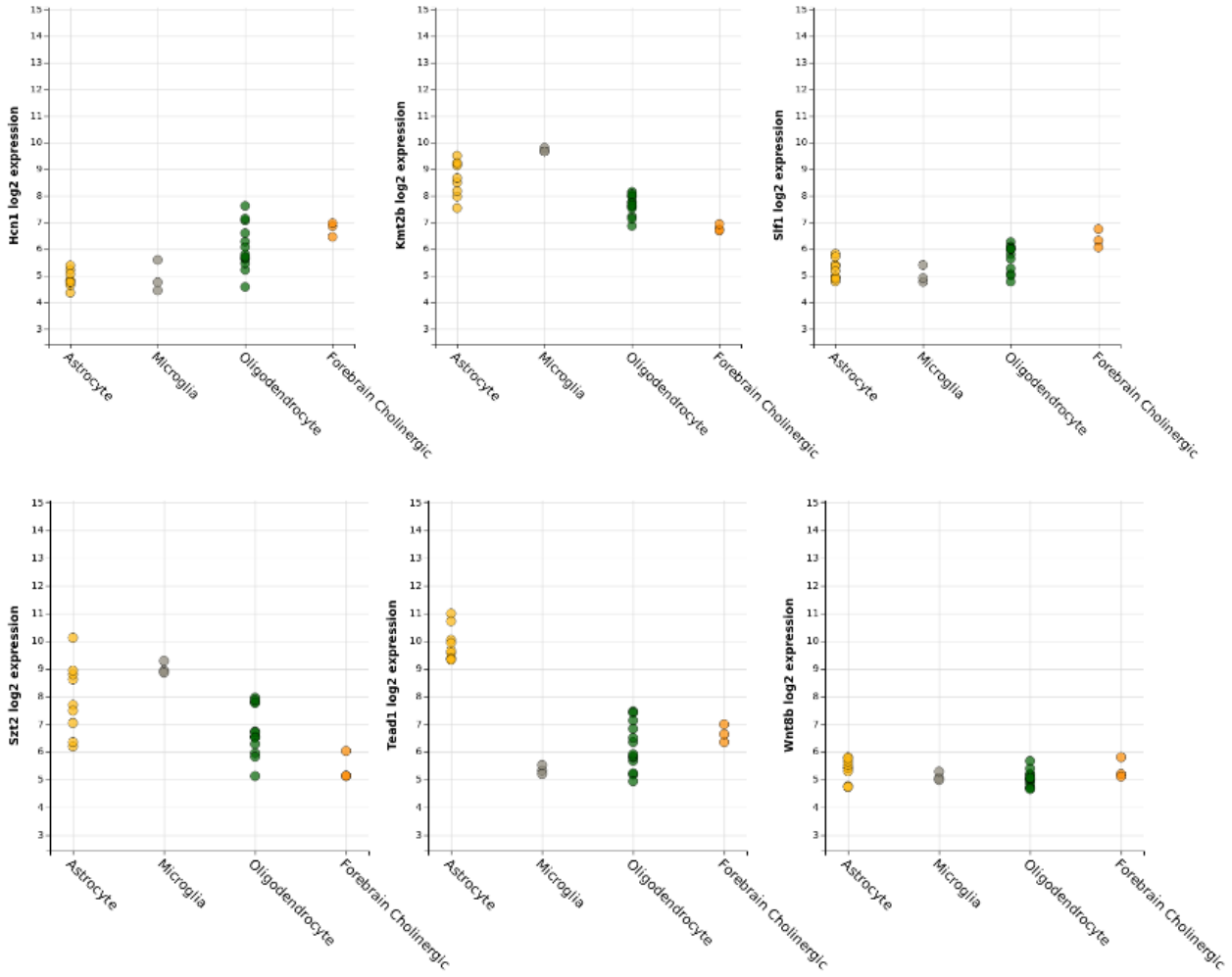


Figure 4.16: Expression of implicated Aicardi genes across different forebrain cell types. This figure shows expression of target genes based on data obtained from Affymetrix GeneChip Mouse Genome 430 2.0 Array (NCBI Geo DataSets platform ID: GPL1261). Each dot graph represents a different target gene and its expression across four different forebrain cell types. Among the six implicated AIC genes, Wnt8b and Slf1 have similar gene expression profiles in the forebrain. Images from NeuroExpresso (https://pavlab.shinyapps.io/neuroexpresso2/).

4.3.6 Concluding Remarks

Our study provided a comprehensive review of existing literature and new genetic findings to shed light on the clinical and genetic heterogeneity of AIC, which for more than 70 years was perceived as a monogenic and likely X-linked disorder. We highlight the importance of utilising genetic studies to guide differential diagnosis of syndromes like AIC that involve multiple complex traits. From our cohort, we identified five likely pathogenic variants in three known disease genes (*HCN1*, *KMT2B* and *SZT2*) and two potentially novel developmental genes (*SLF1* and *WNT8B*). For five of these families, we provided some insight into the potential contributing genetic factors, which were not identified from previous clinical genetic testing. Meanwhile for the remaining unresolved families, we may have missed variants that were: in less annotated regions (non-coding and repeated sequences), transmitted in a non-Mendelian manner (somatic and epigenetic) or undetectable by paired-end sequencing (complex SV, short tandem repeat expansions and trans-regulatory variants).

4.4 Acknowledgement

I would like to thank the following people who contributed to this chapter:

- Miss Sayaka Kayumi for technical assistance in the discovery of the KMT2B variant.
- Ms Alison Gardner analysing the copy number variants of the Aicardi exomes.
- Dr Melody Li for conducting the HCN1 voltage clamp assays and generating the HCN1 result figures.
- Dr Clare van Eyk and Dr Morgan Newman for guiding the zebrafish studies.
- SA Cancer Council Genomics Facility for whole exome sequencing and Dr Mark Corbett for pre-processing of the exome fastq files
- Alzheimer's Disease Genetics Laboratory for providing the facility to conduct the zebrafish studies.
- Australian Genome Research Facility for whole genome sequencing and pre-processing of the genome fastq files
- eResearchSA and Phoenix High Performance Computing Service for providing expertise and increased computational power for data analysis.

4.5 Publication

This chapter has been reformatted into a brief report, which will be submitted to the American Journal of Human Genetics. Please refer to **Appendix 7.3.2** for more details.

CHAPTER V: Contribution of copy number and submicroscopic structural variants to Aicardi Syndrome.

5.1 Introduction

Whole genome sequencing (WGS) has recently emerged as an alternative to array-based platforms, for the detection of structural variants (SV) including: CNV, balanced rearrangements and mobile element insertions (MEI). Paired-end sequencing (PES) has the ability to detect 5000 to 10,000 SV per human genome using the following methods (**Figure 5.1**): read-depth, split-read and sequence assembly (Chiang et al., 2017, Alkan et al., 2011). The advantage of using short PES over older SV discovery methods is the ability to detect multiple SV subtypes across different range of size alterations (**Figure 5.2**) and finding nucleotide resolution breakpoints (Baker et al., 2012, Alkan et al., 2011). The discovery and genotyping of SV in 2504 human genomes, using WGS, revealed that a median of 18.4 megabases of SV are present per diploid human genome (Sudmant et al., 2015). There are molecular models that can explain the formation of SV (**Figure 5.3**), however the challenge remains in interpreting the functional and phenotypic consequences of these SV (Weischenfeldt et al., 2013).

The contribution of SV to human diseases can be monogenic and display positional effect, or multigenic thus relaying gene dosage effects (Weischenfeldt et al., 2013). Large cohort studies have facilitated the discovery of common SV associated with complex traits and autoimmune diseases; as well as rare SV underlying neurodevelopmental and neurocognitive disorders (Girirajan et al., 2011, Conrad et al., 2010, Pinto et al., 2010, Fanciulli et al., 2007). In a cohort study comprised of more than 15,000 children with intellectual disability and other congenital

abnormalities, 14.2% of cases were predicted to have a likely causative CNV that was larger than 400kb (Cooper et al., 2011). In other independent studies, it is estimated that rare CNV contribute to 3-5% of epileptic encephalopathies and an overall 5-10% of all childhood epilepsies (Mefford et al., 2011). More recently, using the WGS platform, putative pathogenic SV were identified in cohorts that share overlapping neurodevelopmental and retinal phenotypes with Aicardi Syndrome (AIC) (Carss et al., 2017, Monlong et al., 2018, Gilissen et al., 2014).

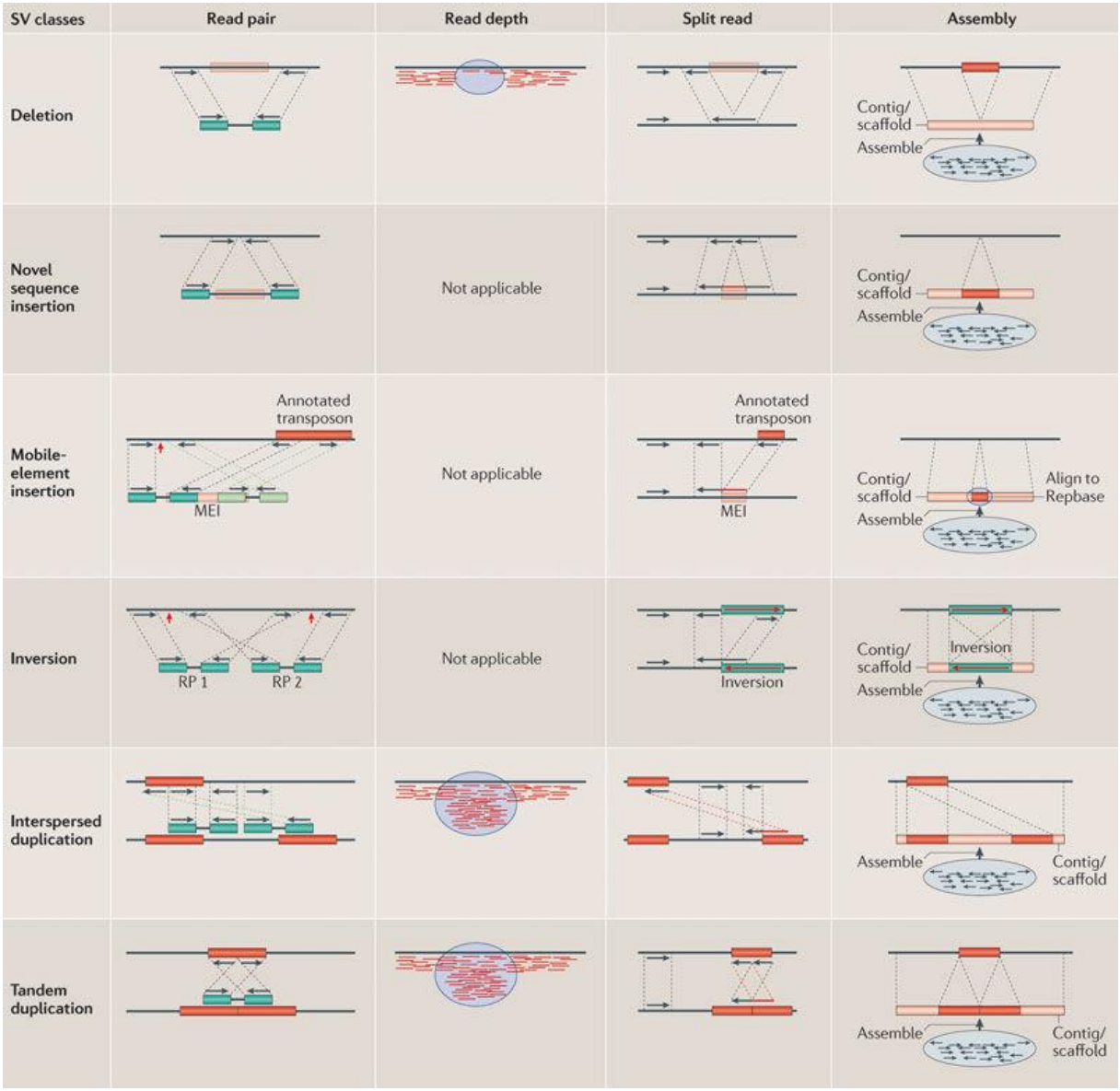


Figure 5.1 Mapping signatures of structural variants. This diagram was extracted from Alkan

et al. (2011) to illustrate the strategies used to identify different classes of structural variants by mapping sequence reads to a reference genome.

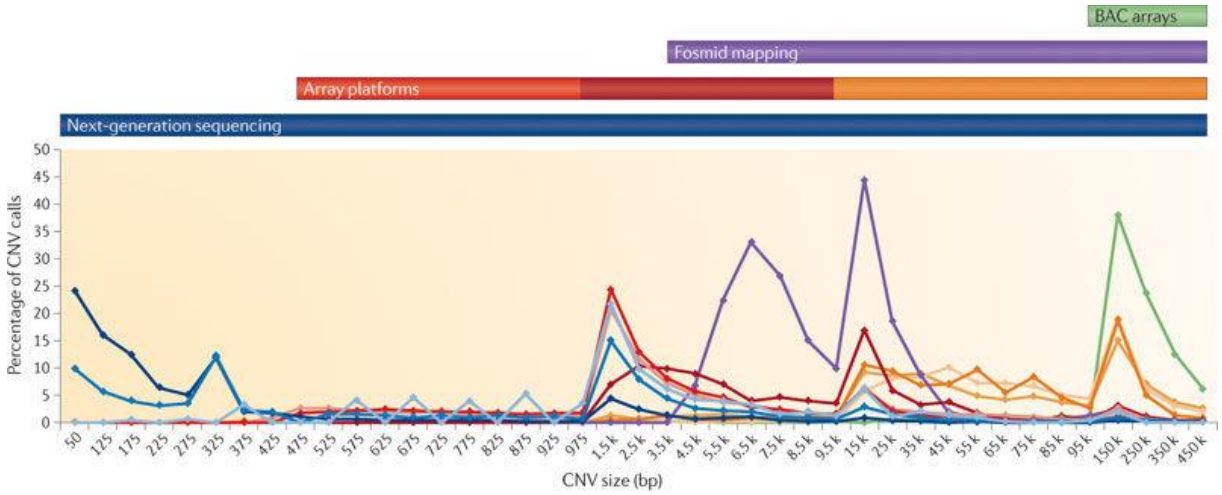


Figure 5.2: Variant size distribution and discovery methods. This graph was extracted from Alkan et al. (2011) to show the differences in variant length detected by four different technologies: BAC arrays, fosmid mapping, array platform and next-generation sequencing.

The genetic causes underlying Aicardi Syndrome have recently been investigated at the single nucleotide level (Lund et al., 2016, Schrauwen et al., 2015). However, earlier studies were focused on finding chromosomal aberrations and submicroscopic abnormalities; using cytogenetic techniques followed by array-based tools that were available at the time. The resolution of these techniques enabled early researchers to detect variants at sizes ranging from hundreds of kilobases to whole chromosomes. Thus, the earliest finding in Aicardi patients were large interchromosomal translocations, deletions and aneuploidies (Ropers et al. 1982, Donnenfeld et al. 1990) (**Figure 5.4**). Although these findings occurred in patients whose AIC diagnosis is disputable, it instigated two decades of further investigations into the X chromosome in hopes of finding an X-linked candidate locus (**Figure 5.5**).

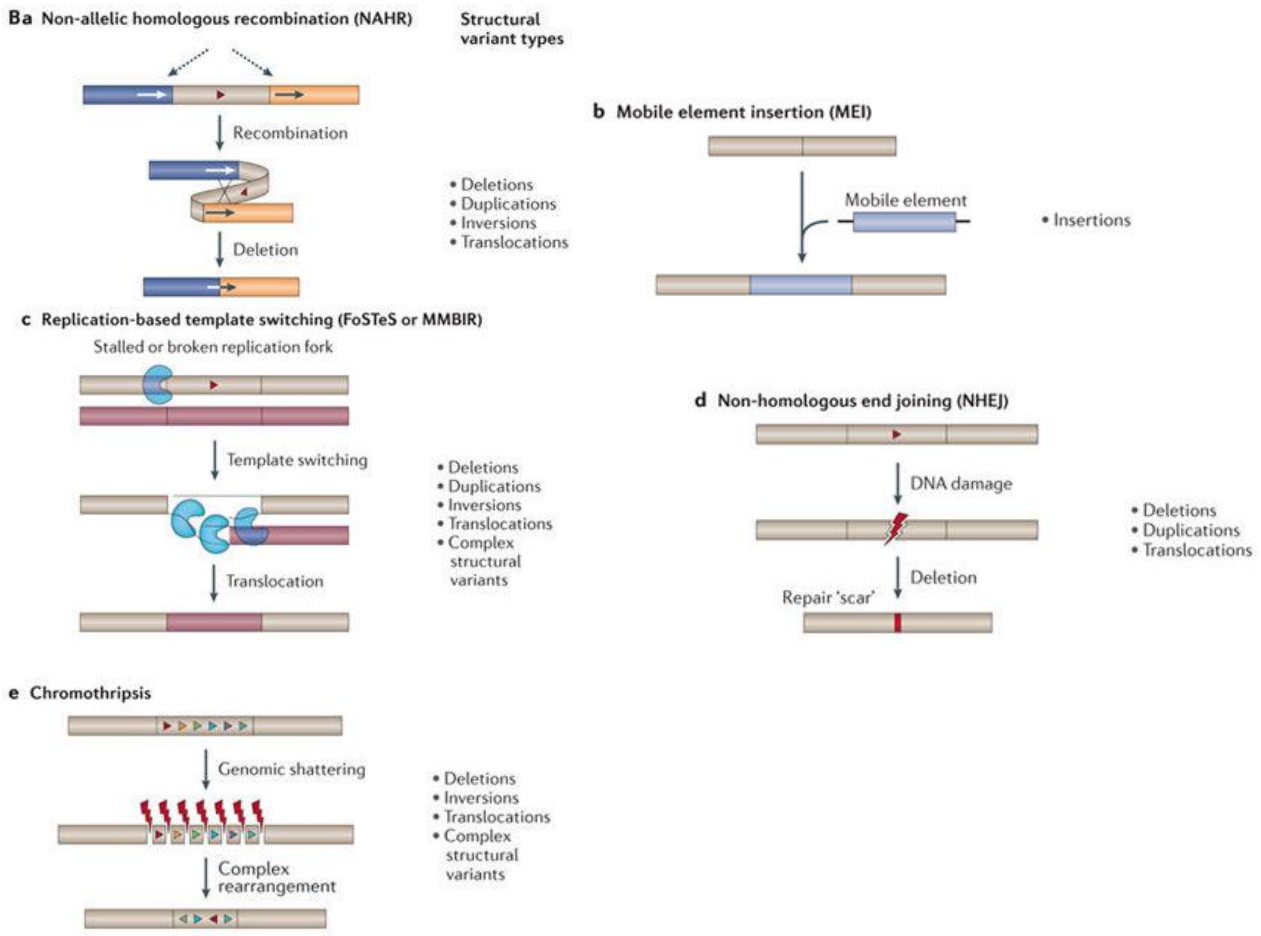


Figure 5.3: Molecular mechanisms underlying structural variant formation. This diagram was extracted from Weischenfeldt et al. (2013) to represent the different mechanisms that lead to the formation of different structural variant subtypes.

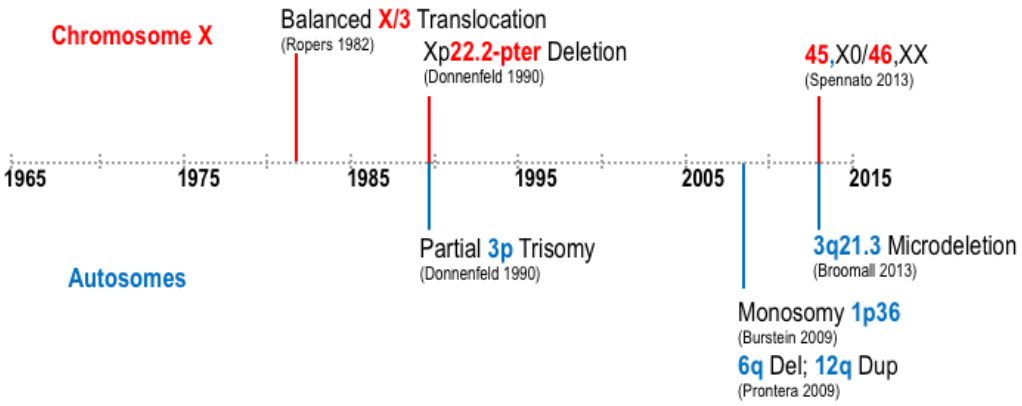


Figure 5.4: Schematic overview of genetic findings. This diagram shows a timeline of the chromosomal abnormalities and aneuploidies reported in Aicardi patients thus far. The chromosomal location and variant type is positioned on top of a reference to the original study.

In 2007, Yilmaz et al. conducted an array based comparative genomic hybridisation (array-CGH) study on a cohort of 18 Aicardi patients; which had a resolution of 82kb and superseded previous X-linked studies. Using BAC clones with a full coverage of the X chromosome, they screened for microdeletions and microduplications, however no deleterious CNV were detected. In a subsequent study, Wang et al. (2009) performed genome-wide screening using a higher resolution array-CGH, which detected 110 SV from a cohort of 38 Aicardi girls. 105 of these SV were either polymorphic or observed in unrelated healthy individuals; four were inherited from an unaffected parent and one *de novo* X-linked CNV was in a region absent of genes. Although the 244k DNA 60-mer array had the spatial resolution of approximately 9kb, the smallest detected CNV was approximately 30Kb; thus CNV smaller than the resolution may potentially be missed or masked by a larger CNV. Although no detectable SV were found in either cohort studies, this does not exclude the potential role of SV to AIC.

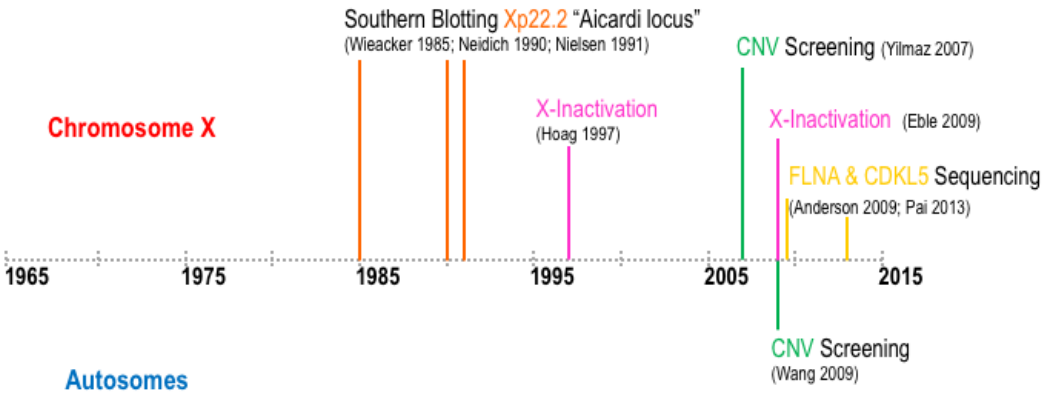


Figure 5.5: Schematic overview of Genetic Investigations. This diagram shows a timeline of the genetic studies conducted on Aicardi patients or cohorts. The molecular technique used for variant discovery is positioned on top of a reference to the original study.

In order to complement our SNP and Indels analysis (**Chapter 4**), we designed an SV discovery pipeline that incorporated multiple SV calling algorithms (**Figure 5.6**). These callers differ in the strategy used to detect different SV signatures (**Chapter 2**) and by combining them, it allowed us to increase the sensitivity and specificity for different SV types from different genomic regions (non-coding and coding) and of different lengths; overall improving breakpoint precision (Tattini et al., 2015). In contrast to the previous array-based CNV studies, we used whole genomes to screen for potential X-linked and autosomal SV at a resolution of 50 base pairs to tens of kilobases (Chiang et al., 2017). This enabled us to search for smaller SV (50bp to 5kb) and submicroscopic SV (10kb to 1mb) that may have been missed by previous methods.

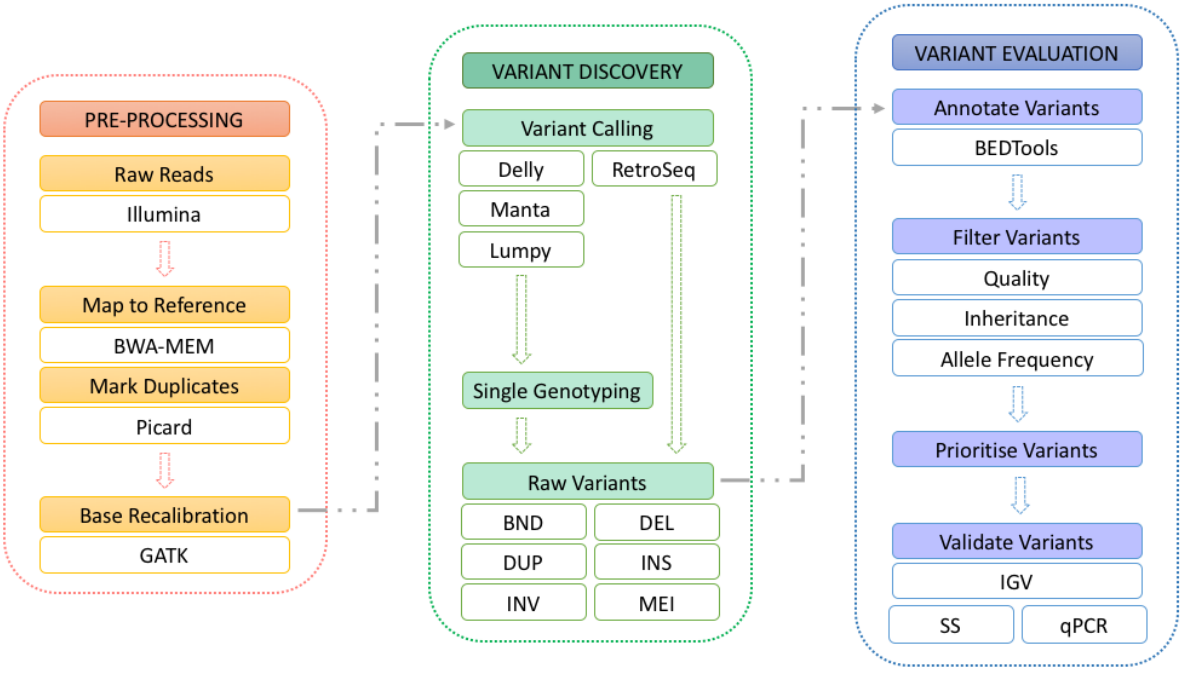


Figure 5.6: Schematic overview of SV detection and filtering pipeline. This flowchart was modified from the SNP and Indels discovery pipeline in Chapter 5 to show the changes made for SV detection, which included: variant callers, variant subtypes and variant annotation methods.

5.2 Results

5.2.1 SV Discovery Pipeline versus NA12878 genome

Our structural variant (SV) pipeline was tested on the GIAB pilot genome (NA12878) that was sequenced on the Illumina HiSeq 2500 v4 platform (Zook et al., 2016). Among the SV callers we selected for our pipeline, Delly called the greatest number of variants (122890) while Manta called the least (9820). In a comparison of the different types of SV, single-anchored breakpoints (BND) represented more than half of the SV called across all SV callers; followed by deletions, duplications, inversions and then insertions (**Figure 5.7**). BND often represent a single breakpoint of an insertion or translocation; the paired-end is either unmapped or located on another chromosome. The low number of insertions called were as expected due to the short-read length (2x125bp) of the WGS platform; the median length (bp) for insertions with resolved break ends were 82 and 24, called by Manta and Delly respectively. In contrast, all SV callers were capable of calling larger deletions, duplications and inversions (**Figure 5.7**).

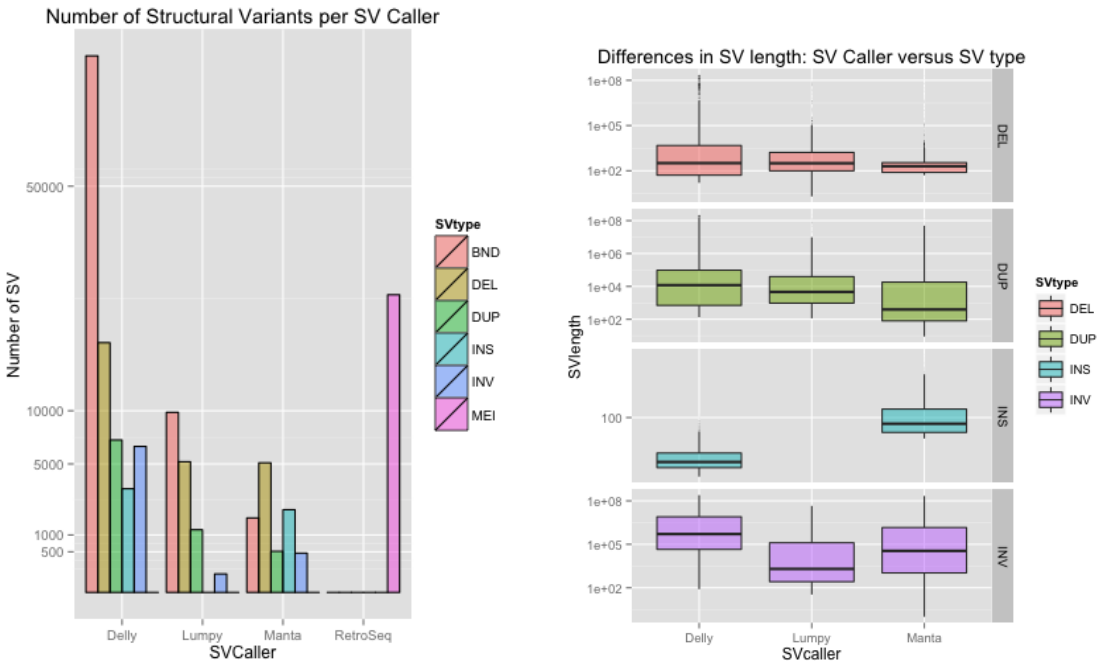


Figure 5.7: NA12878 structural variants with predicted breakpoints detected by different calling algorithms. The bar plot on the left shows the total number of SV types detected across different SV callers. The box plot on the right shows the log10 difference in SV length (base pair) for each SV type across different SV callers. Abbreviations: BND, single break ends; DEL, deletion; DUP, duplication; INS, insertion; INV, inversion; SV, structural variant.

The precision and sensitivity of each SV caller was assessed by comparing each call set with the truth sets from the 1000 Genomes Project phase 3 (1000G) and SVClassify study (Parikh et al., 2016). For the 1000 Genomes Project, different SV types were extracted using a custom script. Overlaps between the call sets and truth sets were obtained via the 'BEDtools intersect' module. For each call set and truth set comparison, three values were calculated: true positive, false positive and false negative (**Figure 5.8**). Based on these values, the precision and sensitivity of each SV caller in detecting each SV subtype was calculated as shown in **Figure 5.9**.

Delly had greater sensitivity but less precision at calling deletions and duplications compared to Lumpy and Manta (**Figure 5.9**). Inversions called by Delly and Manta had similar sensitivity and precision, which were both higher than Lumpy. For insertions, we compared either predicted resolved breakpoints (two-ends) or BND (single-end) (**Figure 5.9**). Lumpy BND detected more insertions that overlap with the 1000G truth set, but less with the SVClassify truth set. Manta's resolved breakpoints had approximately 50% concordance with both insertion truth sets; while its BND had no to minimal overlap. Unsurprisingly, Retroseq superseded all of the other SV callers in detecting true mobile element insertions (**Figure 5.9**).

$$\text{True Positive} = \frac{N_{Intersect}}{N_{Truthset}}$$

$$\text{False Positive} = \frac{N_{Callset} - N_{Intersect}}{N_{Callset}} \qquad Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$\text{False Negative} = \frac{N_{truthset} - N_{Intersect}}{N_{Truthset}} \qquad Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

Figure 5.8: Formulas for calculating concordance. This figure shows all of the formulas used to calculate the overlap between different callsets (SV callers) and truthsets (1000 Genomes or SVClassify). Abbreviation: N, number.

When we combined all of the call sets by their SVtype, we observe a higher concordance between the combined call sets and truth sets (**Figure 5.10**). Based on the NA12878 genome, it can be assumed that almost 100% of putative deletions and inversions will be detected by our SV discovery pipeline; and over 90% for duplications. Combining the BND and resolved INS calls from all the callers increased the detection rate of insertions in the 1000G truth set to 75% and 50% respectively. Taking into account the 75% of the detectable MEI by Retroseq (**Figure 5.9**), we expect our SV discovery pipeline to detect over 75% of any SV subtypes present in a sample

genome that is similar to the NA12878; in terms of read length (2x150bp) and sequence coverage (20-30X).

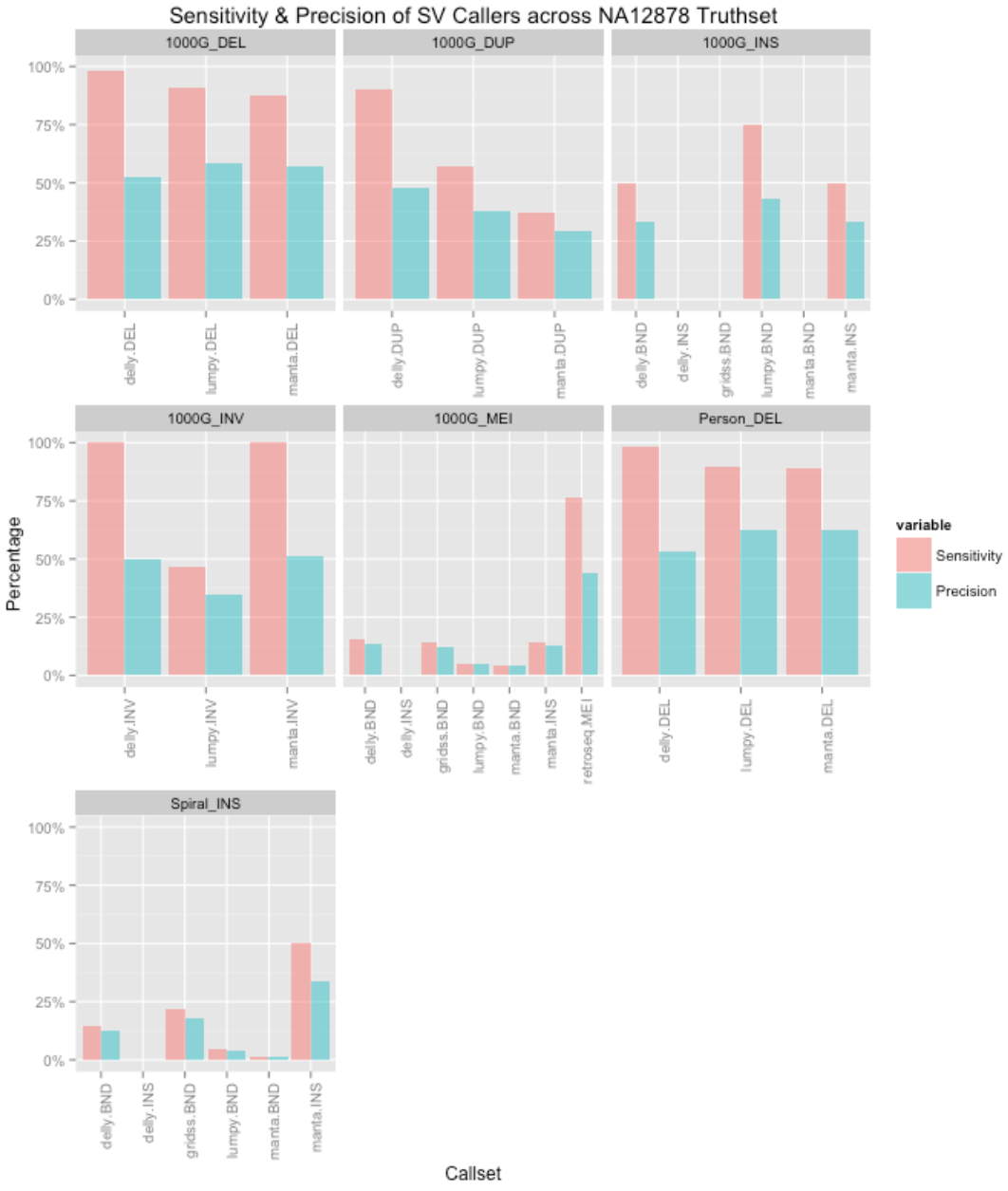


Figure 5.9: Comparing sensitivity and precision of SV callers across NA12878 truthsets. Each bar plot represents an NA12878 truth set for different SV types. Different SV types were extracted from the 1000 Genomes Project Phase 3 to generate "**1000G**" truth sets. Benchmark SV calls from SVclassify (Parikh et al., 2016) were also used as truth sets for deletions (***Person_DEL***) and insertions (***Spiral_INS***). The call sets were based on bed files generated from

the vcf files of each SV caller. Call sets with resolved breakpoints are designated with a SV type suffix; while single break ends are labelled as BND. Abbreviations: BND, single break ends; DEL, deletion; DUP, duplication; INS, insertion; INV, inversion; MEI, mobile element insertions; SV, structural variant.

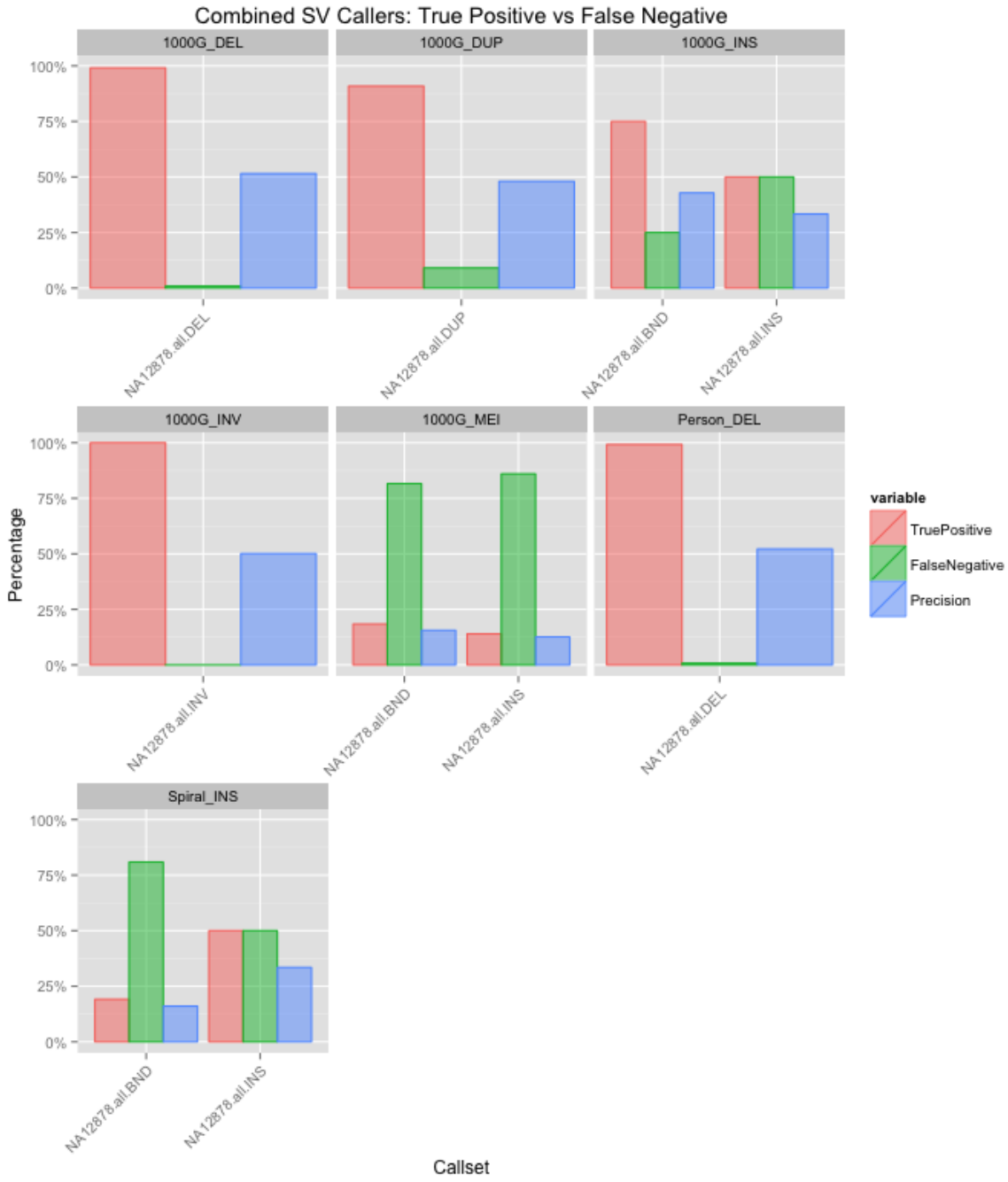


Figure 5.10: Combined SV call sets versus NA12878 truth sets. This faceted boxplot shows the number of real (true positive) and potentially missing (false negative) SV that overlap with

the truth sets, once all of the call sets from the different callers are combined. Abbreviations: BND, single break ends; DEL, deletion; DUP, duplication; INS, insertion; INV, inversion; MEI, mobile element insertions; SV, structural variant.

5.2.2 Unfiltered structural variants in Aicardi cohort and In-house genomes

We ran our SV discovery pipeline across all of the AIC genomes, five proband-parent trios and two singletons. For genomes that were unrelated and unaffected by AIC, we ran the pipeline across the following number of individuals per SV caller: 108 (Delly), 99 (Lumpy), 94 (Manta) and 104 (Retroseq) (**Table 5.1**). We observed a similar trend to NA12878, where BND represented a majority of the SV subtypes called, while INS were the least. Unexpectedly, we observed a 4-fold reduction in the number of MEI present in our set of genomes compared to the NA12878. Correspondingly, there were also slightly less deletions and BND called in our genomes. For the remaining SV subtypes, we observed a slight increase in the median number of called SV. These variations are likely due to different program versions used for pre-processing (**Figure 5.1**) of the Illumina reads; our genomes were sequenced in 2016 while the NA12878 was sequenced in 2015.

Callers	Number of Genomes	BND[23]	DEL	DUP	INS	INV	MEI
Delly	125	66980	18472	7172	3793	5972	-
Lumpy	116	10254	5044	1291	-	2984	-
Manta	111	1708	5248	534	2272	435	-
Retroseq	121	-	-	-	-	-	7506
Total number of unique calls	473	78942	28764	8997	6065	9391	7506
NA12878	1	98746	29164	8737	5328	7029	26846

[23] Abbreviations: BND, single break ends; DEL, deletion; DUP, duplication; INS, insertion; INV, inversion; MEI, mobile element insertions; SV, structural variant.

Table 5.1: Summary of structural variants called by Delly, Lumpy, Manta and Retroseq.

5.2.3 Automated Filtering of Structural Variants in Aicardi Probands

SV in Aicardi genomes were prefiltered to reduce the number of: false positive calls due to sequencing-induced errors and poorly mapped reads; and potentially benign SV that were frequently observed in the control population. From the original number of SV in Aicardi genomes (**Table 5.2,** Unfiltered SV), we were able to filter out approximately 79% of variants using a combined set of SV from unrelated genomes (**Table 5.2,** Absent In-house). Another 6% of original SV were filtered out based on quality parameters recommended by the original authors of each SV caller (**Table 5.2**, Passed Quality) (Chen et al., 2016, Layer et al., 2014, Keane et al., 2013, Rausch et al., 2012). Retroseq calls were filtered based on the number of supporting reads and confident calls that meet breakpoint criteria (GQ tag => 20; FL tag => 5). Lumpy calls were filtered based on the Phred quality that the site is non-reference in any sample in the VCF (QUAL => 90). Manta and Delly calls were filtered based on those that passed all sample filters in the VCF filter fields such as MinQUAL, MinGQ and others (FT = PASS). After filtering variants that were present in a set of published SV that were discovered in an unaffected control population (**Chapter 2; Table 5.2,** Known Benign), we were left with approximately 11% of the original SV.

Sequencing ID	Unfiltered SV	Absent In-house	Passed Quality	Known Benign	Potentially *De novo*	MAF = 0	ORF x Regulatory	IGV Inherited
FR07959033	65169	1820	987	844	177	20	7	0
FR07959030	76860	3400	1121	924	371	38	16	0
FR07959023	76177	3555	962	783	291	34	13	0
FR07958956	142471	13798	1584	1393	1393	146	60	N/A
FR07958917	70610	4738	3212	2403	2403	244	115	N/A
FR07958909	69862	4020	1648	1426	332	31	13	0

FR07958885	72297	3503	1086	908	207	16	5	0
Average	81921	4976	1514	1240	739	76	33	0

Table 5.2: Summary of filtered structural variants from whole genome sequencing.

Next, we filtered for heterozygous SV that were absent from both parents to enrich for *de novo* variants or homozygous SV that were present in both parents for homozygous recessive variants. We also included SV that were heterozygous but present in only one parent for the manual filtering of compound heterozygous variants. The advantage of having a complete proband-parent trio for SV analysis is shown in **Table 5.2**, where inheritance-based filtering reduced the number of filtered SV to a few hundred (177-371); compared to a few thousand (1393-2403) as seen in the singletons FR07958956 and FR07958917. For a first parse analysis, we filtered for variants that flanked the Hg19 canonical transcript of genes; as well as variants that are present in known genes associated with neurodevelopmental or retinal disorders (**Chapter 2**). This left us with approximately 30 SV per genome to manually evaluate (**Table 5.2**).

5.2.4 Manual evaluation of Structural variants in Aicardi probands

Putative SV were manually evaluated by visualising the proband SV alongside the parents' genome (if available) and two unrelated control genomes to assess: (i) the mapping quality of the genomic interval and (ii) heritability of the variant. From manual viewing using IGV and Structural Variant Prediction Viewer (https://github.com/VCCRI/SVPV): 3/35 variants appear to be false positives; 13/35 variants were present in both parents and the remainder were present in one of the parents (**Table 5.3**). The high number of inherited variants that remained after automated filtering is likely due to the break end resolution of the called variants in the proband. For each SV, we buffered the breakpoints by adding 50bp to the left and right outer coordinates, which allows us to eliminate neighbouring and overlapping SV. Furthermore, a minimum 10% of

reciprocal overlap of two SV is needed to be considered overlapping; this is to prevent smaller sized SV from being excluded due to its co-localisation with a larger sized SV. For variants outside of the ORF and flanking regions that lack functional annotations, we were unable to prioritise for potential regulatory variants as RNA sequencing data was unavailable for these probands.

Sequencing ID	Chr	Start	End	SV Type	SV Caller	SV Length	Gene Symbol	Inherited
FR07959033	chr10	101104236	101104286	MEI	Retro	50	CNNM1	mother
FR07959033	chr11	108051282	108051332	MEI	Retro	50	NPAT	father
FR07959033	chr2	118862803	118862823	INS	Delly	20	INSIG2	father
FR07959033	chrX	74728497	74728547	MEI	Retro	50	ZDHHC15	father
FR07959030	chr1	165304109	165304125	INS	Delly	16	LMX1A	mother
FR07959030	chr15	41040027	41040077	MEI	Retro	50	RMDN3	father
FR07959030	chr17	71421936	71421986	MEI	Retro	50	SDK2	mother
FR07959030	chr19	5936661	5936711	MEI	Retro	50	RANBP3	father
FR07959030	chr2	189162673	189162723	MEI	Retro	50	GULP1	father
FR07959030	chr2	189162677	189162727	MEI	Retro	50	GULP1	father
FR07959030	chr3	33225062	33225112	MEI	Retro	50	SUSD5	father
FR07959030	chr4	170658925	170658975	MEI	Retro	50	C4orf27	mother
FR07959030	chr5	168145422	168145472	MEI	Retro	50	SLIT3	father
FR07959030	chr6	136367281	136367331	MEI	Retro	50	PDE7B	father
FR07959030	chr6	136367281	136367331	MEI	Retro	50	AJ606316	father
FR07959030	chr6	155111866	155111916	MEI	Retro	50	SCAF8	father
FR07959030	chr9	111949405	111949455	MEI	Retro	50	EPB41L4B	father
FR07959030	chrX	100357481	100357499	INS	Delly	18	CENPI	father
FR07959023	chr1	172306980	172307030	BND	Manta	50	DNM3	mother
FR07959023	chr1	175920598	175920619	INS	Delly	21	RFWD2	father
FR07959023	chr10	125592193	125592343	DEL	Delly	150	CPXM2	father

FR07959023	chr11	10363248	10363298	MEI	Retro	50	EF537580	mother
FR07959023	chr14	24682592	24682625	INS	Delly	33	TM9SF1	father
FR07959023	chr14	24682592	24682625	INS	Delly	33	NEDD8-MDP1	father
FR07959023	chr2	233857820	233857915	DUP	Manta	95	NGEF	mother
FR07959023	chr5	168244780	168244830	MEI	Retro	50	SLIT3	mother
FR07959023	chr8	99754722	99754772	MEI	Retro	50	STK3	mother
FR07958909	chr1	32584633	32584652	INS	Delly	19	KPNA6	father
FR07958909	chr1	62234144	62234164	INS	Delly	20	INADL	father
FR07958909	chr1	67774469	67774519	MEI	Retro	50	IL12RB2	father
FR07958909	chr15	50592346	50592396	MEI	Retro	50	GABPB1	mother
FR07958909	chr17	73961732	73961782	MEI	Retro	50	ACOX1	mother
FR07958909	chr2	207045494	207045512	INS	Delly	18	GPR1	mother
FR07958909	chr20	37527454	37527469	INS	Delly	15	PPP1R16B	mother
FR07958909	chr8	96697243	96697261	INS	Delly	18	LOC100616530	mother
FR07958885	chr10	61412821	61412871	MEI	Retro	50	SLC16A9	father

Table 5.3: Summary of inherited SV flanking open reading frames and regulatory elements.

5.3 Discussion

We developed a bioinformatics pipeline that was able to detect from 75% to 100% of SV from the benchmark NA12878 genome (**Figure 5.9**). It is known that different SV types have different functional consequences (**Figure 5.11**), thus for the filtering steps we divided SV by their class and length for comparison with unrelated in-house genomes and published datasets (Weischenfeldt et al., 2013). We also ensured that comparisons were made with exact breakpoints (if provided) or a buffered breakpoint of 50 base pair from single break ends. For SV prioritisation, we used multiple published datasets containing putative benign and pathogenic SV, which allowed us to enrich for known SV involved in human NDD and retinal disorders. Although we were able to detect more SV from our cohort of 13 girls compared to the findings in

Yilmaz et al. (2007) and Wang et al. (2009), we did not identify any putative causal SV in our Aicardi trios.

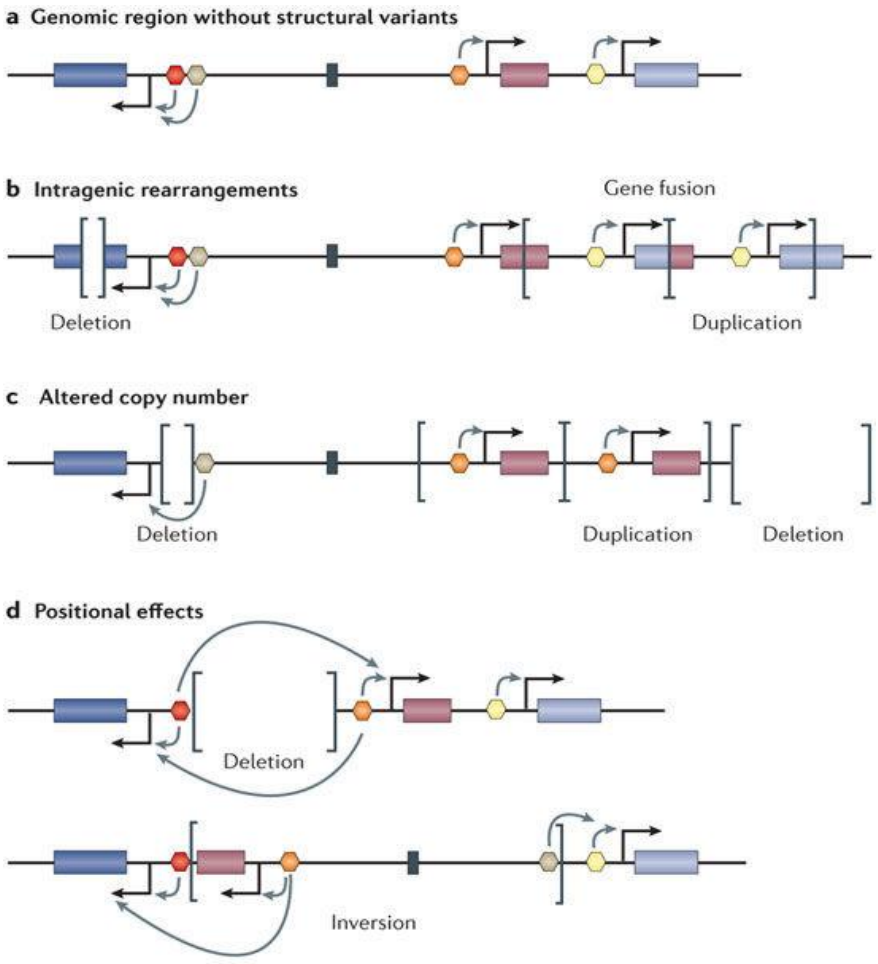


Figure 5.11: Functional consequences of structural variants. This diagram was extracted from Weischenfeldt et al. (2013) to show the different functional impacts of different structural variant subtypes on coding sequences or regulatory elements.

During our first parse analysis, we looked for variants that overlapped protein-coding and disease-associated regions. A majority of these variants were found in either deep intronic sequences free of known expression quantitative trait loci (eQTLs) or regions that were poorly conserved across multiple species. There was insufficient evidence to suggest these SV cis-regulated the expression of the harboured or nearby genes. In a recent deep whole genome

sequencing study, approximately 88% of putative causal SV were non-coding and resided in eQTLs that are predicted to directly alter gene expression; which coincidentally has also been associated with phenotypic traits as inferred by GWAS (Chiang et al., 2017). From our coverage analysis (**Chapter 4**), a majority of the regions not covered by WGS were outside of protein-coding regions or flanking complex (repetitive) sequences. Thus, our SV pipeline may have missed putative causal SV in these non-coding eQTLs.

In contrast our SNP and Indels discovery pipeline, where we incorporated multiple prediction tools to assess the consequences on gene expression and protein function (**Chapter 3 & 4**), our SV pipeline relied on known disease-associated loci or regulatory elements. The lack of *in silico* tools to predict SV impact hindered our ability to prioritise novel-disease SV, especially for our two singleton cases whose parent genomes were unavailable for trio-analysis. There are limited numbers of sequencing-based tools available to assess SV impact due to the fact that different SV types spanning the same genomic region may have: (i) different functional consequences on the sequences they induce or rearrange, (ii) exact SV breakpoints that are difficult to resolve and often non-reciprocal across multiple SV callers and (iii) emphasis on positional effects rather than the potential cis or trans regulatory effects (Weischenfeldt et al., 2013).

For a future second parse analysis, we can incorporate the *in silico* tools (though limited) that are available for SV such as VEP and SVScore. Using Ensembl provided information, VEP (https://github.com/Ensembl/ensembl-vep) can annotate SV that overlap with: (a) transcripts or regulatory regions, (b) known allele frequencies and (c) disease or phenotype association (McLaren et al., 2016). The limitation of VEP is that its predictions can only be applied to balanced SV such as insertions, deletions and duplications. Meanwhile, SVScore (https://github.com/lganel/SVScore) has the ability to aggregate SNP scores across the genomic

region harbouring a putative SV, which generates an SV pathogenic score that predicts the likely deleterious effect (Ganel et al., 2017). However, SVscore is unable to assess the impact of gain of function mutations or cis-regulatory elements. Many of these variant prediction tools are based on a training set of known variants. Thus with the emergence of putative causal SV detected in human disease cohorts and functional datasets with synthetic genetic variants, we can expect better methods for quantitative assessments of SV in the near future (Mahmood et al., 2017).

5.4 Concluding Remarks

In summary, we performed genome-wide screening of submicroscopic SV and CNV in seven Aicardi probands; five proband-parent trio and two singletons. Similar to previous studies, we did not identify any X-linked SV with sufficient evidence to support its pathogenicity (Yilmaz et al., 2007, Wang et al., 2009). Furthermore, we did not detect any recessively inherited variants (homozygous or compound) that were absent from unrelated in-house or public genomes. Our SV analysis also did not reveal likely pathogenic variants in protein-coding or known regulatory regions. Overall our findings reveal that Aicardi Syndrome is unlikely to be caused by SV that were detectable using the WGS platform in our seven-patient cohort.

5.5 Acknowledgement

- I would like to thank Dr Atma Ivancevic for her contribution in generating the Delly and Retroseq datasets for our Aicardi cohort and other unrelated in-house genomes.
- I would also like to thank the Australian Genome Research Facility for providing the whole genome sequencing data and pre-processing of the raw reads from fastq down to vcf file formats.
- I would also like to thank Ms Sayaka Kayumi for providing the NA12878 genome that was assembled using the same pipeline as our other in-house genomes.
- Finally, I would like to thank Dr Mark Corbett for sending DNA samples for whole genome sequencing and his recommended guidance in analysing the structural variants.

5.6 Publication

This chapter have been reformatted into a brief report, which will be submitted to the American Journal of Human Genetics. Please refer to **Appendix 7.3.2** for more details.

CHAPTER VI: Final Discussion & Future Directions

6.1 Significance of Study

Our study is the first to incorporate both computational and molecular approaches to identify candidate genes that may contribute to the pathogenesis of Aicardi Syndrome. To date, *TEAD1* and *OCEL1* are the only genes that have harboured likely pathogenic variants discovered in the exomes of Aicardi probands (Schrauwen et al., 2015). Findings from our WES and WGS study (**Chapter 4 & Chapter 5**), revealed five more genes that are likely to contribute to the clinical manifestations in five of our unrelated probands. Based on the locus heterogeneity of ours and previous findings, we support the hypothesis that causes underlying Aicardi syndrome are genetically heterogeneous. These findings broaden the investigative window for future genetic studies involving individuals diagnosed with Aicardi or Aicardi-like syndromes.

Achieving Aim 1: Develop a bioinformatics pipeline that will enrich for causal variants that underlay eye and cortical development.

We compiled four gene lists to enrich for variants involved in eye and brain development; as well as genes associated with epilepsy and MCDs. For a clinically heterogeneous syndrome like AIC, we screened for variants that can explain for both the pathognomonic CRL and characteristic spectrum of MCDs. This differs from genetic studies like Schrauwen et al. (2015), where the variant identified in the Aicardi proband is likely associated with the eye phenotype, but the connection to the brain MRI findings are not well understood. **Chapter 3** highlighted the effectiveness of our bioinformatics pipeline and variant prioritisation in achieving a genetic diagnosis for a previously unresolved case, wherein the segregating phenotype included both eye and brain abnormalities.

Achieving Aim 2: Prioritise candidate genes by assessing the likely pathogenicity of causal variants using in silico and in vitro assays.

The power of segregation to support a genetic diagnosis in the case of our Aicardi cohort was insufficient due to: (1) lack of additional carriers with overlapping clinical phenotype as our proband and (2) distinguishing the variant as a true disease driver rather than incidental finding. To counteract this issue, we investigated *de novo* variants in *HCN1* and *WNT8B* that were identified from our cohort and had pre-existing *in vitro* assays; tailored to either the gene or its protein family (**Chapter 4.2.4 & 4.2.5**). The *HCN1* variant showed a loss of function in membrane conductivity of the mutant protein compared to its wildtype counterpart; meanwhile the WNT8B mutant protein displayed a dominant negative effect on the intracellular signalling of the wildtype protein. Our *in vitro* findings provided the differential diagnosis for proband T22352, while added further functional evidence to support the pathogenicity of the Leu70Pro variant in proband T22101.

Achieving Aim 3: Investigate the biological role of causal variants in the pathogenesis of Aicardi Syndrome using in vivo models.

To avoid multiple gene testing using independent protein-specific assays, which for a larger cohort study was not practical, we explored the use of zebrafish modelling as a pre-screening method to assess potential candidate genes. Based on the cost and time effectiveness of the morpholino-mediated knockdown approach, we examined four potential Aicardi genes: three from our study and one from Schrauwen et al. (2015). Using the zebrafish assay, we investigated whether the knock down of the implicated genes could lead to an eye pigmentation phenotype; equivalent to the pathognomonic chorioretinal lacunae of AIC (**Chapter 4.2.6**). In all four genes we tested, we saw either a uniform reduction in eye pigmentation (*szt2*) or sparse areas of eye pigmentation often occurring unilaterally (*slf1, tead1* and *wnt8b*). For our implicated Aicardi

genes, we also observed other developmental defects such as skeletal, cerebral and muscular; which correlates with the clinical features observed in the Aicardi patients with likely pathogenic variants in these genes. Although the knockdown in the zebrafish does not model the *de novo* (likely dominant negative) variant observed in the patients, it showed unifying morphant phenotypes in biologically distinct genes; which justifies further functional testing on these genes and their potential role in AIC.

6.2 Limitations

6.2.1 X-linked variants

The most enigmatic aspect of AIC pathology is the excess of affected females compared to males that have been reported in literature thus far. Historical efforts towards answering this enigma were largely focused on finding X-linked candidate genes or loci using investigative tools such as: Southern blotting, skewed X-inactivation studies, X-linked CNV screening and whole X exome sequencing (**Chapter 1.5.2**). Although no X-linked candidates were identified from previous studies, this may be attributed to the resolution of the techniques restricted to either large chromosomal abnormalities or protein-coding regions. In our independent study based on WES and WGS, we extended our window of investigation to include protein-coding and non-coding sequences; as well as regions with no sequencing coverage. Furthermore, our bioinformatics pipeline is able to detect variants at a resolution of single nucleotides to tens of kilobases (**Chapter 4.2.2 & Chapter 5.2.1**). We also investigated regions that may have been missed by sequencing, albeit potentially harbouring pathogenic variants on the X chromosome. Despite all of our considerations, we did not identify any X-linked candidate genes or locus that were enriched in our Aicardi cohort compared to our in-house genomes or public genomes.

The lack of X-linked candidates identified from others and our genetic studies may be attributed to variants residing in sequences inaccessible to the molecular tools utilised thus far. From our coverage analysis (**Chapter 4.2.1**), we show that the X-chromosome was adequately covered across all Aicardi genomes. Combined with segregation analysis *in silico*, we did not find any X-linked denovo or inherited variants. Based on the coverage and segregation analysis, we suggest that X-linked protein-coding variants are unlikely to contribute to the AIC pathology in our cohort. However, for the unresolved individuals, there are potentially X-linked non-coding variants that lie in complex sequence regions that may have been missed by our sequencing approach. For both WES and WGS, we selected paired-end sequencing (PES) for the detection of SNP and genomic rearrangements, which is widely used for large-scale disease studies such as the DDD and Epi4K (Deciphering Developmental Disorders et al., 2015; Epi4K Consortium et al., 2013). Despite its success in disease gene discovery, PES has limited or no coverage across: repetitive DNA sequences, regions with high or low GC bias and regions with non-random DNA fragmentation bias (Treangen et al., 2011, Aird et al., 2011).

6.2.2 Non-coding variants

A number of mutations residing in non-coding sequences have been recently implicated in neurodevelopmental disorders (Kumar et al., 2016, Huang et al., 2012). Approximately 98.5% of the human genome is comprised of non-coding sequences, which were historically considered as 'junk DNA' as they were not functionally transcribed (Estellar et al., 2011). However, the functional importance of non-coding sequences was proposed based on: (1) conserved non-coding sequences were 10x more abundant compared to conserved coding sequences and (2) evolutionary studies have shown that all genomic regions are equally subjective to selective constraints (Attanasio et al., 2008, Thomas et al., 2003, Dermitzakis et al., 2003). In addition, multiple large-scale GWAS studies have presented statistical evidence that show disease-

associated SNP usually reside in non-coding sequences (Herman et al., 2015, Bush et al., 2012). Although the disease-relevance of non-coding sequences have finally been recognised, there are comparatively less resources and tools available to assess the functional effect of non-coding sequences compared to protein-coding sequences (PCS).

One of the limitations of our bioinformatics pipeline is the reliance on existing functional annotations for variant prioritisation outside of (PCS). Our search for likely pathogenic variants flanking exon-intron boundaries, untranslated regions (5' and 3') and known regulatory sequences was unsuccessful. A majority of non-coding variants that passed quality assessment and segregation analysis had: poorly conserved flanking sequences, no prior association to disease trait or unknown regulatory significance. Furthermore, algorithms used to assess protein-coding mutations were not effective at assessing the potential regulatory and structural effects of non-coding sequences (Mather et al., 2016). In addition, we were unable to assess potential trans-regulatory elements as matching RNA sequencing data was unavailable for most of our Aicardi WES and WGS.

6.2.3 Structural variants

The genomic instability of microsatellite and minisatellite repeats, more specifically short tandem repeat (STR) expansions, have been associated with several neurodevelopmental disorders such as Fragile X Syndrome, Myotonic dystrophy and Spinocerebellar ataxia (La Spada et al., 2010). Tandem and dispersed repeats may localise in both coding and non-coding regions; leading to disturbances to DNA processes like replication and recombination or protein consequences such as cellular toxicity (Usdin et al., 2000). Based on the short read-length (2x125bp) of the Illumina platform, our variant discovery pipeline is unable to detect STR that extend into complex sequence regions; especially in non-coding sequences. Fortunately, the systematic biases and

assembly problems presented by PES can be overcome with long-read sequencing (LRS) by Pacific Bioscience or Oxford Nanopore, which provides uniform coverage and *de novo* assembly from average read lengths of 10-15kb (Schatz et al., 2010). Repeated sequences constitute approximately 50% of the human genome (De Koning et al., 2011). Thus utilising LRS technology can expand our investigative window to include potentially pathogenic and previously missed STR on the X chromosome and autosomes.

6.2.4 Somatic variants

A majority of our findings were *de novo* germline variants, as the primary source for our genetic study was peripheral blood, which favours the detection of underlying germline mutations. However, somatic mutations have recently emerged as a genetic driver of various MCD and EE pathologies (Iffland et al., 2016, Moller et al., 2016, Jamuar et al., 2015, Lim et al., 2015). The allelic frequencies of brain-specific somatic mutations can be as high as 6.9-12.6% as reported by Lim et al. (2015) to extremely low mosaic levels of 0.1-3% as detected by Moller et al. (2016). From our cohort, 2/13 proband had paired blood-brain samples available for sequencing; the brain samples were derived from resected focal cortical dysplasia and choroid plexus papilloma. We were unable to detect somatic mutations present in the brain that was also absent from the blood. This could partly be due to our low sequencing coverage of 40x, which would only allow the detection of somatic variants above the allelic fraction of 5%. Variants that are below this frequency threshold were harder to prioritise as germline variants below this proportion were often false positives; thus did not validate with Sanger sequencing.

6.2.5 Non-Mendelian inheritance

Based on the observation of Aicardi individuals born to phenotypically normal parents, we filtered for variants under three potential modes of inheritance: *de novo*, compound heterozygous

and homozygous recessive. However, there are non-Mendelian transmission mechanisms that could also explain the phenotypic variability and lack of heritability of AIC such as: incomplete penetrance, epigenetic factors (including imprinting), repeat expansions and polygenic traits with continuous spectrum. We excluded variants that were present in only one parent and the proband assuming that autosomal dominance was unlikely as family history for a majority of our probands was unremarkable. However, gonadal mosaicism can occur in the parents resulting in affected progenies, which would resemble sporadic cases similar to those observed in bowel and breast cancer (Delon et al., 2013, Hes et al., 2008). This could also potentially explain the underlying cause of monozygotic twins discordant for AIC (Pons et al., 2008, Costa et al., 1997).

6.3 Future Directions

6.3.1 Improving variant screening

X-linked Variants

Other X-linked causes that we may have missed could include genes that escape from X-inactivation, thus protecting females from the dosage effect of losing a functional single allele (Dunford et al, 2017). Alternatively, there could be an X-linked gene that requires the male patient to have a mosaic mutant variant to manifest AIC. This phenomena is observed in the case of PCDH19-related epilepsy, where karyotypically normal females with PCDH19 variants are affected, while males are asymptomatic unless they are carriers of mosaic single nucleotide variants (Terracciano et al., 2016). Variants in genes that are known to escape X-inactivation and those that manifest symptoms under cellular mosaicism need to reassessed in our Aicardi exomes and genomes.

Non-coding Variants

In recent years, there were increases in functional annotations of non-coding sequences from the ENCODE project and other large-scale GWAS studies; as well as improvements in computational methods used to recognise regulatory motifs (Ward et al. 2015; Smedley et al. 2016; Abbas et al., 2016, Zhou et al., 2015). For a comprehensive analysis of non-coding variants, we can reanalyse the exome, genome and RNA sequencing data using updated non-coding annotations and prediction tools.

Structural Variants

There are potentially complex or larger structural variants (SV) that may have been missed by our PES approach such as: dispersed duplications, unbalanced translocations, microinversions and fully assembled large insertions. The detection of some intrachromosomal SV were restricted by the length of our read pair fragment sizes, while interchromosomal SV are reliant on paired-end mapping; thus unbalanced translocations presenting single-anchored breakpoints were difficult to resolve. Furthermore the availability of benchmark SV used to filter false positive or clinically benign SV is scarce compared to SNP or Indels; thus prioritising and interpreting the clinical significance of some SV were inconclusive (**Chapter 5.2.4**). Nonetheless, optical mapping can be an alternative to PES for the detection of novel (disease-relevant) CNV and Indels up to tens of kilobases in length; or combined with haplotype analysis can detect SV up to hundreds of kilobases (Howe et al., 2015, Cao et al., 2014). Unlike the cost associated with most LRS technologies, optical mapping provides a more cost-effective approach for complex SV discovery in smaller cohort studies like ours (Goodwin et al., 2016).

Non Mendelian Inheritance

Epigenetic inheritance of AIC has only been briefly explored in a recent genome-wide methylation study by Piras et al. (2017). Although the study indicated differential methylation in

AIC probands compared to their unaffected parents and controls, there were a few caveats including: small sample size, conducted on peripheral blood and examined 2% of all possible human DNA methylation sites. Nonetheless it is important to explore the epigenome as mutations in epigenetic factors or aberrations to epigenetic processes are involved in many neurodevelopmental disorders (NDD) such as: Rubenstein-Taybi, Coffin-Lowry, ATR-X and Rett Syndromes. Correspondingly, some NDD are caused by genomic imprinting, which is a mechanism that has yet to be explored in AIC studies but has the potential to explain the female-biased incidence. It has been suggested that some imprinted genes are sexually differentiated, independent of the sex chromosome (Faisal et al., 2014, Radford et al. 2012). To comprehensively study whether epigenetics contributes to sexual differentiation or dimorphism in AIC, histone modifications and RNA-associated silencing also needs to be considered in addition to DNA methylation.

Somatic Variants

Nonetheless the lack of paired blood-brain samples for the investigation of brain-specific mutations can be circumvented by: (i) increasing the sequencing coverage of blood-brain DNA to a minimum of 200X or (ii) sequencing cell-free DNA released post seizure (Jamuar et al., 2015, Liimatainen et al., 2013). Both of these options can improve the detection of brain-specific variants, particularly in our cohort for individuals with unresolved causes.

6.3.2 Future experimental plans

SLF1 & SZT2

At the time of our study, the biological role of *SZT2* and *SLF1* was not elucidated and thus were not tested *in vitro*. We now know the role of *SZT2* in the MTOR pathway in both of the gator arms and *SLF1* in the DNA-damage repair pathway with *SLF2* as a duplex. Future experimental

plans for the *SZT2* variant could involve measuring the interaction of the mutant protein with other gator components such as *DEPDC5*, *WDR59* and *SESN2*. Meanwhile, the *SLF1* variant can be assessed using available patient lymphoblastoid cell lines to detect DNA instability; or alternatively measure the mutant response to UV or chemically-induced DNA damage. Although we were unable to perform the necessary assay, we recognise the importance of functional validation for variant prioritisation, especially for cases with multiple complex phenotypes.

In vivo screening

There were a number of practical and theoretical limitations regarding our morpholino knockdown approach. Firstly, we used one translation-blocking morpholino per target gene and the level of knockdown was not measurable due to unavailability of antibodies. Secondly, we were unable to perform rescue experiment of the morphants using wildtype plasmid, which would have consolidated the correlation between the target gene and the suspected chorioretinal lesions. For repeated experiments using the morpholino knockdown approach, we would utilise a cocktail of multiple splice blockers, of which the knockdown effect can be directly measured by quantifying the level of transcript.

Nonetheless, the morpholino knockdown approach enabled us to prioritise genes that may be involved in eye or brain development, however to precisely model the functional consequence of our identified *de novo* variants, we could utilise advance organoid technologies. A transgenic *PAX6* reporter was introduced into mouse embryonic stem cells and differentiated into retinal organoids, which allowed gene expression profiling at different stages of retinogenesis (Volkner et al., 2016). A similar experiment for implicated AIC genes would allow the study of the retinal pigment epithelium, which could give rise to the pathognomonic CRL. Similarly, human brain organoids can be generated to study perturbations to neurodevelopment that may give rise to the signature spectrum of MCDs observed in AIC (Lullo and Kriegstein, 2017).

6.3.3 Explaining female-bias incidence

Sexually dimorphic chorioretinal lacunae

In light of our sequencing findings, which contradicts the X-linked hypothesis and the inability to recruit male AIC patients for our study, we explored other mechanisms that can lead to female-biased traits. As mentioned before, CRL (**Figure 1.1**) are the pathognomonic feature that

distinguishes AIC from other malformation of eye-brain disorders. One possibility that has yet to be explored is that CRL is a female-exclusive phenotype. In proband SQC25F8, we found a paternally inherited variant in *TCF4* in a region free of reported mutations (Sepp et al., 2012). Female, but not male, Tcf4 and Tead1 null mice heterozygotes show abnormal retinal pigmentation and retinal morphology respectively (Koscielny et al., 2014). Remarkably, both of these corresponding genes have been linked to either Pitt Hopkins Syndrome or Sveinsson's chorioretinal atrophy, which do not have a gender-bias aspect. Unfortunately, there is a lack of single-sex studies of female animals compared to males at a ratio of 1 to 5.5 respectively (Beery et al., 2011). In addition, corresponding mouse studies for the remaining AIC candidate genes were unavailable to test whether pigmentation defects are present and exclusive to females. Nonetheless, the observed retinal defects in female mice may be attributed to the relationship between estrogen and retinal pathologies; the role of sex hormones in ocular tissue has been previously established (Bajwa et al., 2012). Notably from immunocytochemistry, the expression of estrogen receptor alpha (ERα) was detected in the retina and retinal pigment epithelium of young female eyes but not in men or postmenopausal women (Ogueta et al., 1999).

Female-sex hormones

This brings us to our next hypothesis that AIC genes are interacting or modulated by female sex hormones. In MCF-7 (breast cancer) cells, it has been shown that KMT2B can epigenetically regulate estradiol (E2) dependent transcription via ERα (Su et al., 2016). Furthermore, the knockdown of KMT2B can lead to the down-regulation of 195 ERα target genes that were upregulated by E2. Another independent study in MCF-7 cells showed that WNT8B expression was significantly upregulated by E2, however it is unknown whether WNT8B is one of 195 genes modulated by KMT2B (Saitoh et al., 2002). E2 can also regulate gene expression via membrane estrogen receptors such as the G-protein-coupled oestrogen receptor 1 (GPER) (Revankar et al.,

2005). In ZR-75-30 (breast cancer) cells, it has been shown that E2 can stimulate GPER and regulate Hippo signalling. Furthermore, GPER-induced transcription requires TAZ and transcriptional factors including TEAD1 (Zhou et al., 2015). E2 can also modulate the DNA damage repair pathway by inducing double stranded DNA breaks and genomic instability (Caldon, 2014). Recently, SLF1 was identified as a DNA repair factor that forms a complex with SLF2 and RAD18 to recruit SMC5/6 to DNA lesions and suppress genome instability (Raschle et al. 2015). Collectively, we have observed the direct (KMT2B, WNT8B) and indirect (TEAD1) or implicated (SLF1) regulation of estrogen in female AIC patients displaying the unifying CRL phenotype.

Sexually dimorphic signalling

Independent of the sex hormones, others and our implicated AIC genes may be involved in sexually dimorphic signalling pathways. It has been previously shown that multiple tissue types, including the brain, exhibit sexual dimorphism as a result of highly specific patterns of gene expression (Yang et al., 2006, Dewing et al., 2003). From the microarray analysis of the F2 progeny of B6.apoE$^{-/-}$ and C3H.apoE$^{-/-}$ mice, there were approximately 600 sexually dimorphic genes identified in the Yang et al. (2006) study; spanning both autosomes and sex chromosomes. In a study based on the meta-analysis of LCL gene expression from the Geuvadis consortium, 587 sex-biased genes were identified and nine transcription factors were enriched near female-biased genes; some of these were involved in the Wnt signalling pathway (Shen et al., 2017). There are known female NDD caused by Wnt-related genes such as *DDX3X*, *USP9X* and *PORCN* (Snijders Blok et al., 2015, Murtaza et al., 2015, Grzeschik et al., 2007). In our study, we identified a *WNT8B* variant that affected canonical wnt signalling; while the Schrauwen et al. (2015) study discovered an upregulation of the wnt receptor *LRP5* and a *TEAD1* variant involved in the hippo pathway that cross-regulates with the wnt/B-catenin pathway (Kim et al., 2014). The

potential role of Wnt signalling or closely related pathways can explain both the clinical and genetic heterogeneity of AIC, as well as the female-bias prevalence; however more collaborative findings from multiple Aicardi cohorts are needed for further investigation.

6.4 Conclusion

Our study provided meta-analysis and new data to shed light on the clinical and genetic heterogeneity of AIC, which for more than 70 years was perceived as a monogenic and likely X-linked disorder. We highlight the importance of utilising genetic studies to guide differential diagnosis of syndromes like AIC that involve multiple complex traits. From our cohort, we identified five likely pathogenic variants in three known disease genes (*HCN1, KMT2B* and *SZT2*) and two potentially novel developmental genes (*SLF1* and *WNT8B*). For five of these families, we provided some insight into the genetic basis of their child's condition; which were unresolved from previous clinical genetic testing. More importantly, there can be immediate health and welfare benefits including: counselling for family planning, closure to parents of affected children by alleviating self-blame and opportunities for improved support through patient outreach services and social media groups.

From a research perspective, there are challenges that remain in establishing genetic causes for unresolved severe disorders of the brain. Firstly, the ability to distinguish between driver and incidental mutations, particularly for AIC where the prevalence of 4000 individuals worldwide reduces the probability of finding a recurring hit in a small cohort study given the unexpected heterogeneity we encountered. Secondly, without cost and timely functional assessment of each of these likely pathogenic variants, there is a lack of acceptance and support for isolated variants identified in one gene in one family. To study and resolve genetically heterogeneous disorders like Aicardi Syndrome, it would be more valuable to concentrate efforts on identifying molecular

signatures, converging pathway or similar developmental expression, which may involve different genes but lead to the same phenotypic consequence. In conclusion, our project established a computational and biological pipeline that can be used to identify potential genetic drivers of unexplained eye and brain disorders. This will lead to better understanding of how these brain disorders arise, accurate genetic diagnosis and potentially gene-tailored treatments.

6.5 Publication

This chapter have been reformatted into a brief report, which will be submitted to the American Journal of Human Genetics. Please refer to **Appendix 7.3.2** for more details.

CHAPTER VII: Appendices

7.1 Appendix Code & Data

Code and data files used to generate results for **Chapter 4, 5** and **6** are stored in the following github repository: https://github.com/ThuongHa/7.1-Appendix-Data-Code.git. Directory and file names are shown below:

```
7.1\ Appendix\ Code\ &\ Data\ /
├── 7.1.1\ Scripts
│   ├── Figure\ 4.3
│   │   ├── WESCoverage.R
│   │   └── WGSCoverage.R
│   ├── Figure\ 4.4
│   │   └── NoCoverage.R
│   ├── Figure\ 5.10
│   │   ├── 01_allintersect.sh
│   │   └── 02_CombinedSV.R
│   ├── Figure\ 5.7
│   │   ├── 01_sv_bed.sh
│   │   ├── 02_SV_Number.R
│   │   └── 03_SVlength.R
│   ├── Figure\ 5.9
│   │   ├── 01_truthset_bed.sh
│   │   ├── 02_intersect.sh
│   │   └── 03_Truthset.R
│   ├── Table\ 4.3
│   │   ├── 01_exomelist.sh
│   │   ├── 02_exomepicard.sh
│   │   ├── 03_genomelist.sh
│   │   └── 04_genomepicard.sh
│   ├── Table\ 4.4\ &\ 4.6
│   │   ├── 01_BWA-GATKHC.pipeline.wrapper.sh
│   │   ├── 02_Illumina-Phred33-PE-BWA-Picard-GATKv3.x.Local.sh
│   │   ├── 03_WES-VCF.wrapper.sh
│   │   ├── 04_vcfFamilyBasedFilterAndAnnotate.sh
│   │   ├── 05_FiltFreqAndInher.sh
│   │   ├── 06_FiltFreqV3.sh
│   │   └── 07_FiltInherV3.sh
│   ├── Table\ 5.1
│   │   ├── 01_allscript.sh
│   │   ├── 02_freebayes_AIC.sh
│   │   ├── 03_lumpy_AIC.sh
│   │   ├── 04_manta_AIC.sh
│   │   ├── 05_strelka_AIC.sh
│   │   ├── 06_rename.sh
│   │   ├── 07_delly_key.sh
│   │   ├── 08_lumpy_key.sh
│   │   ├── 09_manta_key.sh
│   │   └── 10_retro_key.sh
│   └── Table\ 5.2
│       ├── 01_aicardi.sh
│       ├── 02_probed_intersect.sh
│       ├── 03_quality_intersect.sh
│       ├── 04_benign_intersect.sh
│       ├── 05_proanno.sh
│       ├── 06_sinanno.sh
│       ├── 07_rm0bytes.sh
│       ├── 08_inherit.sh
│       ├── 09_inhousecat.sh
│       ├── 10_inhousecat2.sh
│       ├── 11_inhouse_intersect.sh
```

```
│       ├── 12_pass.sh
│       ├── 13_mibinher.sh
│       ├── 14_genotype.sh
│       ├── 15_gene_intersect.sh
│       ├── 16_disease_anno.sh
│       └── 17_catdisease.sh
└── 7.1.2\ Data\ Files
    ├── Figure\ 4.3
    │   ├── 01_WES_targetbasepct.csv
    │   └── 02_WGS_targetbasepct.csv
    ├── Figure\ 4.4
    │   └── WES_WGS_NoCov.csv
    ├── Figure\ 5.10
    │   └── combinedSV.csv
    ├── Figure\ 5.7
    │   └── NA12878_svlength.csv
    ├── Figure\ 5.7\ &\ Table\ 5.1
    │   ├── NA12878.all.BND.key.gz
    │   ├── NA12878.all.DEL.key.gz
    │   ├── NA12878.all.DUP.key.gz
    │   ├── NA12878.all.INS.key.gz
    │   ├── NA12878.all.INV.key.gz
    │   ├── NA12878.delly.BND.key.gz
    │   ├── NA12878.delly.DEL.key.gz
    │   ├── NA12878.delly.DUP.key.gz
    │   ├── NA12878.delly.INS.key.gz
    │   ├── NA12878.delly.INV.key.gz
    │   ├── NA12878.gridss.BND.key.gz
    │   ├── NA12878.lumpy_th.BND.key.gz
    │   ├── NA12878.lumpy_th.DEL.key.gz
    │   ├── NA12878.lumpy_th.DUP.key.gz
    │   ├── NA12878.lumpy_th.INV.key.gz
    │   ├── NA12878.manta.BND.key.gz
    │   ├── NA12878.manta.DEL.key.gz
    │   ├── NA12878.manta.DUP.key.gz
    │   ├── NA12878.manta.INS.key.gz
    │   ├── NA12878.manta.INV.key.gz
    │   └── NA12878.retroseq.MEI.key.gz
    ├── Figure\ 5.9
    │   ├── 01_allntersect.txt
    │   ├── 02_NA12878_intersect.csv
    │   └── 03_NA12878_intersect_4col.csv
    ├── Table\ 2.3
    │   ├── 1000G_ALU.bed.gz
    │   ├── 1000G_CNV.bed.gz
    │   ├── 1000G_DEL.bed.gz
    │   ├── 1000G_DUP.bed.gz
    │   ├── 1000G_INS.bed.gz
    │   ├── 1000G_INV.bed.gz
    │   ├── 1000G_LINE.bed.gz
    │   ├── 1000G_MEI.bed.gz
    │   ├── 1000G_SVA.bed.gz
    │   ├── Complex_INDEL.bed.gz
    │   ├── Complex_SNP.bed.gz
```

```
│   ├── MtSinai_INDEL.bed.gz
│   ├── Personalis_DEL.bed.gz
│   ├── Spiral_INS.bed.gz
│   ├── Zook_INDEL.bed.gz
│   └── Zook_SNP.bed.gz
└── Table\ 5.2
    ├── 01_20130502_1000G_50bp.bed.gz
    ├── 02_20150921_cnvddcontrol.bed.gz
    ├── 03_20160925_dgv.bed.gz
    ├── 04_20161013_GoNL_50bp.bed.gz
    ├── 05_20170619_iscabenign.bed.gz
    ├── 06_20170701_dddcommon_50bp.bed.gz
    ├── 07_20101014_huanghi_top20%.bed.gz
    ├── 08_20170323_retnet_50bp.bed.gz
    ├── 09_20170701_hg19dddgenes_50bp.bed.gz
    ├── 10_20170702_genereviews.bed.gz
    ├── 11_20170615_clinvarSVpatho_50bp.bed.gz
    ├── 12_20170701_hg19genesymbol_50bp.bed.gz
    ├── 13_20160515_oreganno.bed.gz
    ├── 14_20150904_encodeDNASE_50bp.bed.gz
    ├── 15_20130816_encodeTFBS_50bp.bed.gz
    ├── 16_20140325_encodeGM12878_50bp.bed.gz
    ├── benign.bed.gz
    └── pathogenic.bed.gz

19 directories, 112 files
```

7.2 Appendix Tables

7.2.1 List of likely pathogenic variants identified in Aicardi Singletons

The following variants passed our prioritisation criteria for likely pathogenicity, but the parent genomes of FR07958956 and FR07958917 were unavailable in these probands, thus *in silico* segregation analysis was not applicable.

Proband	Chr	Start	End	Ref	Obs	Gene	ExAC	snp142	PP2	CADD	GERP	VAF	DP
FR07958956	chr1	48697786	48697786	T	G	SLC5A9	0.00001	0	D	28.7	6.04	0.51	22
FR07958956	chr1	52823458	52823458	G	A	CC2D1B	0	0	D	25	5.41	0.3	10
FR07958956	chr2	20130279	20130279	T	A	WDR35	0.00006	rs370951527	P	26.3	5.2	0.44	14
FR07958956	chr2	32740484	32740484	C	T	BIRC6	0	rs61757639	D	25.6	3.64	0.5	16
FR07958956	chr2	164467322	164467322	T	G	FIGN	0	0	D	23.1	4.8	0.12	4
FR07958956	chr2	233410292	233410292	C	T	CHRNG	0.00002	rs199767964	D	35	4.53	0.56	14
FR07958956	chr3	148904356	148904356	G	C	CP	0	0	P	23.4	4.11	0.38	11
FR07958956	chr4	187115722	187115722	G	A	CYP4V2	0.00004	rs199476187	D	26.7	5.4	0.68	19
FR07958956	chr5	141034952	141034952	G	A	ARAP3	0.00006	rs141682206	D	35	5.12	0.49	19
FR07958956	chr6	3111076	3111076	T	C	RIPK1	0	0	P	25	5.82	0.36	8
FR07958956	chr6	56357084	56357084	C	A	DST	0	0	D	34	5.54	0.5	23
FR07958956	chr7	102944337	102944337	C	T	PMPCB	0	0	P	23.6	4.3	0.49	17
FR07958956	chr8	2026889	2026889	G	A	MYOM2	0	0	D	25.8	4.71	0.44	11
FR07958956	chr8	37693192	37693192	C	T	ADGRA2	0.00011	0	P	35	5.29	0.53	17
FR07958956	chr8	95523525	95523525	A	G	KIAA1429	0	0	D	25.5	5.5	0.46	13
FR07958956	chr8	145139436	145139436	T	C	GPAA1	0	0	D	25.5	5.01	0.48	14
FR07958956	chr10	70412288	70412288	A	G	TET1	0.00005	0	D	24.2	3.41	0.48	21
FR07958956	chr10	97446261	97446261	C	G	TCTN3	0.00002	rs147928670	P	23.5	4.46	0.49	21
FR07958956	chr11	47469392	47469392	C	A	RAPSN	0	0	P	33	5.14	0.42	16
FR07958956	chr11	56127804	56127804	T	G	OR8J1	0.00062	rs149173319	D	25.9	4.57	0.63	22
FR07958956	chr12	12672907	12672907	C	T	DUSP16	0	0	P	26.7	5.75	0.66	19
FR07958956	chr12	21015427	21015427	T	C	SLCO1B3	0.00001	0	P	23.4	3.9	0.48	22
FR07958956	chr12	133357385	133357385	T	C	GOLGA3	0	0	P	25.4	5.79	0.43	12
FR07958956	chr14	105643306	105643306	G	A	NUDT14	0.00004	0	D	26.7	3.08	0.48	10
FR07958956	chr15	88576096	88576096	G	A	NTRK3	0	rs371590703	D	34	4.91	0.47	14
FR07958956	chr15	89381915	89381915	T	C	ACAN	0	0	D	23.3	4.95	0.13	4

FR07958956	chr15	90764607	90764607	C	A	SEMA4B	0	0	D	26.8	3.59	0.15	4
FR07958956	chr16	5083680	5083680	G	C	NAGPA	0	0	D	25.1	5.23	0.49	17
FR07958956	chr16	84214746	84214746	G	A	TAF1C	0.00002	0	D	25.2	4.56	0.52	13
FR07958956	chr17	8296344	8296344	C	G	RNF222	0	0	P	20.5	4.57	0.59	17
FR07958956	chr17	39671917	39671917	C	G	KRT15	0.00036	rs79344504	D	28.5	4.74	0.55	21
FR07958956	chr17	65163599	65163599	G	A	HELZ	0.00006	rs200790914	P	32	4.99	0.5	16
FR07958956	chr18	48241536	48241536	A	C	MAPK4	0	0	P	25.2	5.35	0.52	17
FR07958956	chr19	17893893	17893893	G	A	FCHO1	0.00004	0	P	26.1	3.84	0.46	11
FR07958956	chr19	18179264	18179264	G	A	IL12RB1	0.0003	rs201548803	P	24.8	4.6	0.49	23
FR07958956	chr19	50305813	50305813	G	A	AP2A1	0.00004	0	P	22	5.11	0.42	11
FR07958956	chrX	9677703	9677703	C	A	TBL1X	0	0	D	24.2	4.19	0.13	3
FR07958956	chrX	47003874	47003874	T	C	NDUFB11	0	0	P	24.6	4.1	0.18	4
FR07958956	chrX	49071927	49071927	A	C	CACNA1F	0	0	D	25.9	3.98	0.09	3
FR07958956	chrX	153595141	153595141	C	T	FLNA	0	0	D	32	5.24	0.45	14
FR07958917	chr1	1563061	1563061	C	T	MIB2	0	0	D	23.4	3.8	0.78	23
FR07958917	chr1	26357016	26357016	G	A	EXTL1	0.00002	rs146384192	D	34	4.61	0.52	33
FR07958917	chr1	50884378	50884378	G	C	DMRTA2	0	0	D	22.2	4.46	0.63	24
FR07958917	chr1	116206387	116206387	G	A	VANGL1	0	0	P	25.1	4.01	0.55	33
FR07958917	chr1	152081033	152081033	C	G	TCHH	0	0	P	16.57	3.78	0.84	19
FR07958917	chr1	155108443	155108443	C	T	SLC50A1	0	0	D	32	4.85	0.69	29
FR07958917	chr1	158724876	158724876	A	T	OR6K6	0	0	D	23.8	5.48	0.44	41
FR07958917	chr1	169279311	169279311	C	A	NME7	0	0	D	28.2	5.58	0.52	27
FR07958917	chr1	178359240	178359240	C	G	RASAL2	0.00001	0	D	22.9	5.66	0.53	36
FR07958917	chr1	227348333	227348333	T	C	CDC42BPA	0	0	D	25.4	5.18	0.56	36
FR07958917	chr2	17692163	17692163	T	C	RAD51AP2	0.00005	0	P	25.3	5.62	0.49	41
FR07958917	chr2	162870961	162870961	T	C	DPP4	0.00013	rs200348396	P	23	5.81	0.59	32
FR07958917	chr2	172379071	172379071	T	C	CYBRD1	0.00003	0	P	24	3.69	0.54	24
FR07958917	chr2	219296640	219296640	C	T	VIL1	0	0	D	33	5.09	0.52	31
FR07958917	chr2	219695023	219695023	G	A	PRKAG3	0.00002	0	D	21.4	4.18	0.66	32
FR07958917	chr2	219875534	219875534	G	A	CFAP65	0.00002	rs370054630	D	27.1	5.41	0.4	30
FR07958917	chr2	234102524	234102524	C	G	INPP5D	0	0	D	24.4	5.28	0.45	33
FR07958917	chr3	46786175	46786175	A	G	PRSS45	0.00002	0	P	23.4	5.83	0.61	38
FR07958917	chr3	48664072	48664072	G	A	SLC26A6	0.00001	0	D	31	3.56	0.52	25
FR07958917	chr3	129195528	129195528	C	T	IFT122	0	0	D	29.7	5.65	0.49	41

FR07958917	chr3	140140107	140140107	G	C	CLSTN2	0.00001	0	P	24.4	5.7	0.41	29
FR07958917	chr3	158364655	158364655	G	A	GFM1	0.00015	rs200727254	D	33	5.86	0.47	43
FR07958917	chr3	169540133	169540133	C	T	LRRIQ4	0.00006	0	D	22.8	4.66	0.33	33
FR07958917	chr3	183774048	183774048	G	C	HTR3C	0	0	D	29.8	4.77	0.52	27
FR07958917	chr3	186940962	186940962	C	T	MASP1	0.00002	0	D	35	5.88	0.49	37
FR07958917	chr4	629697	629697	C	T	PDE6B	0.00005	0	P	24.4	4.25	0.38	21
FR07958917	chr5	1243727	1243727	C	A	SLC6A18	0.00029	rs148220007	D	25.1	4.98	0.3	20
FR07958917	chr5	37293060	37293060	G	A	NUP155	0	0	P	28.8	5.24	0.61	31
FR07958917	chr5	80040326	80040326	C	T	MSH3	0.00005	0	P	28.2	4.91	0.26	34
FR07958917	chr5	80381670	80381670	A	G	RASGRF2	0.00002	rs145244937	D	26.9	5.57	0.53	30
FR07958917	chr5	101583133	101583133	C	A	SLCO4C1	0	0	D	33	6.17	0.7	56
FR07958917	chr5	140256522	140256522	C	G	PCDHA12	0	0	D	20.9	3.43	0.53	40
FR07958917	chr5	151046042	151046042	C	A	SPARC	0.00002	0	D	34	4.81	0.53	32
FR07958917	chr5	176309040	176309040	C	A	HK3	0	0	D	28.4	3.27	0.43	23
FR07958917	chr6	150184795	150184795	G	A	LRP11	0	0	D	33	4.03	0.17	12
FR07958917	chr7	4172034	4172034	G	C	SDK1	0	0	P	21.1	5.31	0.42	26
FR07958917	chr7	30485796	30485796	G	A	NOD1	0.00001	0	D	33	5.38	0.5	34
FR07958917	chr7	44800040	44800040	C	A	ZMIZ2	0	0	P	32	4.4	0.48	23
FR07958917	chr7	80374467	80374467	C	T	SEMA3C	0	0	P	23.2	5.56	0.26	31
FR07958917	chr7	86539236	86539236	T	C	KIAA1324L	0	0	P	16.55	4.88	0.6	45
FR07958917	chr7	100175817	100175817	C	T	LRCH4	0.00006	rs142600009	D	31	4.52	0.59	22
FR07958917	chr7	135329650	135329650	A	G	NUP205	0	0	P	23.3	5.02	0.54	39
FR07958917	chr8	2044236	2044236	G	A	MYOM2	0.00004	0	D	29.6	5.46	0.59	27
FR07958917	chr8	10466998	10466998	G	A	RP1L1	0.00002	rs375128393	D	23.2	4.44	0.48	23
FR07958917	chr8	10480518	10480518	A	G	RP1L1	0.00011	rs377036512	P	25.3	4.78	0.32	28
FR07958917	chr8	87076402	87076402	C	A	PSKH2	0	0	P	28.3	5.1	0.5	36
FR07958917	chr8	92139360	92139360	C	T	LRRC69	0.00004	0	D	25.5	5.93	0.44	41
FR07958917	chr8	145058505	145058505	G	A	PARP10	0.00001	0	D	31	4.21	0.3	20
FR07958917	chr9	100257947	100257947	G	T	TDRD7	0	0	P	19.4	5.19	0.36	33
FR07958917	chr9	125330564	125330564	T	A	OR1L8	0	0	P	24.2	3.16	0.47	36
FR07958917	chr9	131495774	131495774	C	G	ZER1	0	0	D	32	5.77	0.54	26
FR07958917	chr10	5925019	5925019	C	G	ANKRD16	0.00003	0	P	23.4	5.33	0.51	35
FR07958917	chr10	61833470	61833470	C	T	ANK3	0.00001	0	D	26.2	5.7	0.39	28
FR07958917	chr10	71008260	71008260	G	A	HKDC1	0.00002	0	D	28.2	5.1	0.57	23

FR07958917	chr10	75058817	75058817	C	T	CFAP70	0	rs112759937	P	15.58	4.42	0.52	42
FR07958917	chr10	99118300	99118300	T	C	RRP12	0	0	D	24.2	3.37	0.59	17
FR07958917	chr10	104121626	104121626	A	C	GBF1	0	0	D	28.3	6.17	0.48	27
FR07958917	chr10	106025861	106025861	C	T	GSTO1	0.00002	rs369551343	D	33	5.35	0.29	24
FR07958917	chr10	126716025	126716025	G	A	CTBP2	0	0	P	26	3.77	0.44	34
FR07958917	chr11	4615690	4615690	C	T	OR52I1	0.00009	rs202168611	D	25.2	4.04	0.44	34
FR07958917	chr11	6519825	6519825	T	C	DNHD1	0	0	D	23.1	4.28	0.61	31
FR07958917	chr11	6806288	6806288	C	T	OR2AG1	0	0	P	22.7	3.33	0.51	35
FR07958917	chr11	27378978	27378978	C	T	CCDC34	0	0	D	35	5.61	0.4	42
FR07958917	chr11	56143880	56143880	C	A	OR8U1,OR8U8	0	0	D	23	4.78	0.82	62
FR07958917	chr11	59245546	59245546	T	C	OR4D10	0	0	D	23.1	4.58	0.55	33
FR07958917	chr11	73805042	73805042	T	A	C2CD3	0	0	P	21.7	3.43	0.5	34
FR07958917	chr12	30862883	30862883	T	C	CAPRIN2	0	0	D	23.2	5.7	0.64	39
FR07958917	chr12	42499690	42499690	T	C	GXYLT1	0	rs79044728	D	25.5	4.93	0.86	43
FR07958917	chr12	42499694	42499694	A	T	GXYLT1	0	0	D	25	6.07	0.86	43
FR07958917	chr12	46320945	46320945	G	A	SCAF11	0	0	D	31	4.99	0.46	35
FR07958917	chr12	53343084	53343084	G	T	KRT18	0	0	D	27.2	3.69	0.84	19
FR07958917	chr12	54917227	54917227	A	C	NCKAP1L	0	0	P	27	5.13	0.54	39
FR07958917	chr12	56031508	56031508	T	A	OR10P1	0	0	P	24	3.06	0.57	28
FR07958917	chr12	129566452	129566452	C	T	TMEM132D	0.00001	0	P	22.1	3.82	0.54	28
FR07958917	chr13	28004071	28004071	C	T	GTF3A	0	0	P	26.4	3.61	0.45	44
FR07958917	chr13	35782874	35782874	A	T	NBEA	0.00012	rs370712135	P	25.1	5.75	0.46	37
FR07958917	chr15	33261131	33261131	G	A	FMN1	0	0	D	21.6	4.22	0.44	16
FR07958917	chr15	40586540	40586540	C	T	PLCB2	0	0	P	34	4.69	0.62	13
FR07958917	chr15	81610816	81610816	G	A	STARD5	0	0	D	24.9	4.06	0.57	23
FR07958917	chr16	4935248	4935248	T	G	PPL	0	0	P	18.9	3.21	0.4	20
FR07958917	chr16	28495363	28495363	G	A	CLN3	0.00002	0	D	22.8	5.28	0.56	27
FR07958917	chr16	57921802	57921802	G	A	CNGB1	0.00001	0	P	25.9	5.46	0.57	28
FR07958917	chr16	67237752	67237752	G	C	ELMO3	0	0	D	23.2	4.52	0.56	18
FR07958917	chr16	85141552	85141552	G	C	FAM92B	0	0	D	23.3	4.74	0.57	35
FR07958917	chr16	89799814	89799814	G	A	ZNF276	0	0	P	23.2	4.92	0.52	23
FR07958917	chr17	25909962	25909962	C	T	KSR1	0	0	D	27.5	5.7	0.47	15
FR07958917	chr17	27448650	27448650	T	C	MYO18A	0.00001	0	D	22.5	4.61	0.61	33
FR07958917	chr17	35900523	35900523	C	T	SYNRG	0.00004	0	D	26.3	5.52	0.6	30

FR07958917	chr17	40825840	40825840	C	T	PLEKHH3	0.00006	rs200439190	P	24	4.78	0.31	16
FR07958917	chr17	48605541	48605541	G	A	MYCBPAP	0.00001	0	D	31	4.82	0.45	31
FR07958917	chr17	66899494	66899494	C	A	ABCA8	0	0	D	24.1	4.9	0.57	47
FR07958917	chr18	13826638	13826638	C	T	MC5R	0.00004	0	D	23.8	4.58	0.52	29
FR07958917	chr18	21494814	21494814	C	G	LAMA3	0.00003	rs373150935	P	26.5	3.98	0.46	35
FR07958917	chr19	464136	464136	C	T	ODF3L2	0	0	P	25.1	3.8	0.46	13
FR07958917	chr19	36135689	36135689	G	C	ETV2	0.00002	0	D	28.6	4.7	0.62	21
FR07958917	chr19	37488173	37488173	G	A	ZNF568	0	0	P	26.5	3.89	0.4	30
FR07958917	chr20	33264826	33264826	A	C	PIGU	0	0	D	25.7	5.09	0.37	30
FR07958917	chr20	62324281	62324281	G	A	RTEL1	0.00002	rs369357121	D	23.4	4.83	0.54	28
FR07958917	chr22	34157457	34157457	C	T	LARGE	0	0	P	23.5	5.83	0.33	27
FR07958917	chr22	37578725	37578725	C	T	C1QTNF6	0	0	P	20.8	4.05	0.53	19
FR07958917	chr22	50639845	50639845	A	G	SELO	0	0	P	20.2	3.29	0.2	5
FR07958917	chrX	69645241	69645241	T	G	GDPD2	0	0	P	23.2	4.85	0.45	33

7.2.2 TOPFlash experimental conditions and measurements

The raw data that was used to generate the results from **Section 4.2.5** is listed below.

14.05.15					
Plasmids	**ug/uL**	**TOPFlash RLU**	**FOPFLash RLU**	**TOPFlash ABS**	**FOPFlash ABS**
pcDNA 3.1 empty vector	1.0	100.0	100.0	0.0	0.0
WNT8B wildtype	1.0	413.0	118.3	31.2	7.6
WNT8B mutant	1.0	204.5	115.8	3.8	4.5
WNT8B wildtype + mutant	0.75 + 1	217.3	136.5	36.2	5.5
17.05.15					
Plasmids	**ug/uL**	**TOPFlash RLU**	**TOPFlash ABS**		
pcDNA 3.1 empty vector	1.0	100.0	4.9		
WNT8B wildtype	0.2	381.1	16.3		
WNT8B mutant	0.2	245.0	11.7		
WNT8B mutant	0.5	289.7	17.6		
WNT8B wildtype + mutant	0.2 + 0.2	311.6	10.1		
WNT8B wildtype + mutant	0.2 + 0.5	276.3	12.7		
05.07.15					
Plasmids	**ug/uL**	**TOPFlash RLU**	**TOPFlash ABS**		
pcDNA 3.1 empty vector	1.0	100.0	0.1		
WNT8B wildtype	0.2	396.9	5.5		
WNT8B wildtype	0.5	368.6	0.9		
WNT8B wildtype	1.0	344.0	19.9		
WNT8B mutant	0.2	199.4	13.1		
WNT8B mutant	0.5	261.2	11.9		
WNT8B mutant	1.0	228.3	17.6		
14.12.16					

Plasmids	**ug/uL**	**TOPFlash RLU**	**TOPFlash ABS**		
pcDNA 3.1 empty vector	1.0	100.0	0.9		
WNT8B wildtype	0.2	674.1	0.5		
WNT8B mutant	0.2	445.9	0.8		
WNT8B mutant	0.5	458.3	4.7		
WNT8B wildtype + mutant	0.2 + 0.2	452.3	1.5		
WNT8B wildtype + mutant	0.2 + 0.5	469.8	11.0		
15.12.16					
Plasmids	**ug/uL**	**TOPFlash RLU**	**TOPFlash ABS**		
pcDNA 3.1 empty vector	1.0	100.0	0.7		
WNT8B wildtype	0.2	417.3	4.0		
WNT8B mutant	0.2	310.7	8.2		
WNT8B mutant	0.5	351.9	32.7		
WNT8B wildtype + mutant	0.2 + 0.2	380.9	10.9		
WNT8B wildtype + mutant	0.2 + 0.5	390.9	17.6		
Average of N=3 Experiments					
Plasmids	**ug/uL**	**TOPFlash RLU**	**TOPFlash ABS**		
pcDNA 3.1 empty vector	1.0	100.0	2.2		
WNT8B wildtype	0.2	490.8	6.9		
WNT8B mutant	0.2	333.9	6.9		
WNT8B mutant	0.5	366.6	18.3		
WNT8B wildtype + mutant	0.2 + 0.2	381.6	7.5		
WNT8B wildtype + mutant	0.2 + 0.5	379.0	13.7		

7.3 Appendix Publications

7.3.1 A mutation in COL4A2 causes autosomal dominant porencephaly with cataracts.

Chapter 3 has been published as a clinical report in the journal *American Journal of Medical Genetics Part A* under the same title. Contributions of my co-authors are listed in **Section 3.4**.

CLINICAL REPORT

AMERICAN JOURNAL OF medical genetics PART A

A Mutation in *COL4A2* Causes Autosomal Dominant Porencephaly with Cataracts

Thuong T. Ha,[1] Lynette G. Sadleir,[2] Simone A. Mandelstam,[3,4,5] Sarah J. Paterson,[2] Ingrid E. Scheffer,[3,5,6,7] Jozef Gecz,[1,8,9] and Mark A. Corbett[8,9]*

[1]School of Biological Sciences, University of Adelaide, Adelaide, South Australia, Australia

[2]Department of Paediatrics and Child Health, University of Otago Wellington, Wellington South, New Zealand

[3]Department of Paediatrics, University of Melbourne, Royal Children's Hospital, Melbourne, Victoria, Australia

[4]Department of Radiology, University of Melbourne, Royal Children's Hospital Parkville, Melbourne, Victoria, Australia

[5]Florey Institute of Neurosciences and Mental Health, Heidelberg, Victoria, Australia

[6]Department of Neurology, Royal Children's Hospital, Melbourne, Victoria, Australia

[7]Department of Medicine, Austin Health, University of Melbourne, Victoria, Australia

[8]School of Medicine, University of Adelaide, Adelaide, South Australia, Australia

[9]Robinson Research Institute, University of Adelaide, Adelaide, South Australia, Australia

Manuscript Received: 11 September 2015; Manuscript Accepted: 9 December 2015

Mutations in *COL4A1* are well described and result in brain abnormalities manifesting with severe neurological deficits including cerebral palsy, intellectual disability, and focal epilepsy. Families with mutations in *COL4A2* are now emerging with a similar phenotype. We describe a family with an autosomal dominant disorder comprising porencephaly, focal epilepsy, and lens opacities, which was negative for mutations in *COL4A1*. Using whole exome sequencing of three affected individuals from three generations, we identified a rare variant in *COL4A2*. This *COL4A2* (c.2399G>A, p.G800E, CCDS41907.1) variant was predicted to be damaging by multiple bioinformatics tools and affects an invariable glycine residue that is essential for the formation of collagen IV heterotrimers. The cataracts identified in this family expand the phenotypic spectrum associated with mutations in *COL4A2* and highlight the increasing overlap with phenotypes associated with *COL4A1* mutations.

Key words: *COL4A1*; *COL4A2*; porencephaly; epilepsy; cataracts; autosomal dominant

How to Cite this Article:
Ha TT, Sadleir LG, Mandelstam SA, Paterson SJ, Scheffer IE, Gecz J, Corbett MA. 2016. A mutation in *COL4A2* causes autosomal dominant porencephaly with cataracts.
Am J Med Genet Part A 170A:1059–1063.

INTRODUCTION

Collagen IV (COL IV), an important component of the basal lamina, is comprised of six α-polypeptide chains encoded by six differentially expressed genes [Schmidt et al., 1993]. These genes are arranged in pairs on chromosomes 2, 13, and X, and each pair produces two genetically identical paralogous α-peptides through a bidirectional promoter [Haniel et al., 1995; Khoshnoodi et al., 2008]. The COL IV α-peptides assemble into heterotrimers with the following stoichiometries: α1α1α2, α3α4α5, and α5α5α6 [Khoshnoodi et al., 2008]. Aberrations to COL IV heterotrimers can lead to genetic and acquired diseases [Lemmink et al., 1997] (Supplementary Table SI). COL4A1 and COL4A2 are the most abundantly and ubiquitously expressed COL IV peptides [Favor et al., 2007]. Mutations in *COL4A1* and *COL4A2* can affect multiple organs resulting in various cerebral, ocular, renal and muscular pathologies [Favor et al., 2007; Kuo et al., 2014].

Mutations in *COL4A1* result in well-recognized phenotypes (Supplementary Table SI). The current clinical picture of

Grant sponsor: NH&MRC; Grant numbers: 628952, 1041920, 1006110; Grant sponsor: WCH foundation MS McLeod Research Fellowship; Grant sponsor: Health Research Council of New Zealand Project; Grant number: 10/402; Grant sponsor: CureKids New Zealand.

*Correspondence to:
Dr. Mark Corbett, School of Medicine, the University of Adelaide at the Women's and Children's Hospital, 72 King William Rd. North Adelaide, South Australia 5006, Australia.
E-mail: mark.corbett@adelaide.edu.au

Article first published online in Wiley Online Library (wileyonlinelibrary.com): 28 December 2015
DOI 10.1002/ajmg.a.37527

COL4A2 is limited to porencephaly type II (MIM 614483) and intracerebral haemorrhage (MIM 614519). While systemic and eye pathologies have been reported in isolated patients, these clinical features are less well recognized as part of the *COL4A2* spectrum (Supplementary Table SII) [Verbeek et al., 2012; Gunda et al., 2014].

Cerebral defects are prominent in all patients with *COL4A2* mutations and there is in vivo evidence from mice model studies that missense mutations in *Col4a2* can also induce ocular, myopathic, and systemic defects [Favor et al., 2007; Kuo et al., 2014]. We expand the clinical spectrum of *COL4A2* by identifying a novel missense mutation in a family with autosomal dominant inheritance of porencephaly, focal epilepsy and lens opacities.

MATERIALS AND METHODS

Informed written consent was obtained from all subjects or their parents or legal guardians in the case of minors. The study was approved by the New Zealand Health and Disability Ethics Committee and the Adelaide Women's and Children's Health Network Human Research Ethics Committee.

Exome Sequencing

Whole exome sequencing (Roche Nimblegen SeqCap v3) was performed on a HiSeq2500 (Illumina) by the South Australian Cancer Council Genomics Facility on three affected individuals (I-2, II-2, III-1; Fig. 1A). Reads were mapped to the human genome (hg19) using BWA-MEM [Li and Durbin, 2009] and mapping refined using Genome Analysis Toolkit (GATK) version 3.2-2 [DePristo et al., 2011]. Mapping achieved a minimum median target coverage depth of 46 reads/sample and covered 95% of intended targets with at least 10 reads (Supplementary Fig. S1). Single nucleotide variants and small insertions and deletions were called by the GATK haplotype caller version 3.2-2 [DePristo et al., 2011]. Larger copy number and structural variants were analysed by CoNIFER [Krumm et al., 2012]; however, no segregating, rare copy number variants were detected. All variants were annotated for allele frequency, clinical significance, locus identity, and likely pathogenicity using ANNOVAR [Wang et al., 2010]. Novel genotypes shared between affected individuals, but absent from a control set of 15 exomes, were separated using the vcf-contrast module from VCFtools [Danecek et al., 2011].

RESULTS

Clinical Reports

The proband (Fig. 1A, III-1) was a 16-year-old male with spastic quadriplegia and apparent intellectual disability who was non-verbal and fully dependent. First concerns arose at 11 weeks due to poor head control with seizure onset at 2 years. He developed intractable multiple types of focal seizures and status epilepticus. Non-progressive bilateral small posterior lens opacities were found at 3 years. EEGs showed left hemisphere slowing and right parietal epileptiform activity. MRI brain showed right frontal communicating porencephaly with dilated lateral ventricles (Fig. 1B, i) and destruction of the right caudate nucleus, lentiform nucleus, internal capsule, thalamus, and right hippocampus (Fig. 1B, ii).

The proband's 13-year-old brother (III-2) had developmental delay with walking at 18 months and first words at 2 years. Autism spectrum disorder was diagnosed at 5 years. He had apparent intellectual impairment and could read simple sentences and do basic addition. He had focal dyscognitive seizures from 3 to 11 years. At 14 months, he had non-progressive bilateral posterior subscapular lens opacities. His neurological examination was normal. EEG at three years was normal. MRI revealed a right frontal porencephalic cyst communicating with the right lateral ventricle (Fig. 1B, iii). There was partial destruction of the periventricular white matter and caudate body. A smaller cyst was present in the right external capsule. The corpus callosum had an unusual appearance with thickened genu and splenium and absence of the intervening body (Fig. 1B, iv). *COL4A1* sequencing was normal.

The 41-year-old mother (II-2) was asymptomatic and of apparent normal intellect. She had a mild left internuclear ophthalmoplegia. MRI showed a cluster of small cysts in the right frontal periventricular white matter and bilateral anterior caudate nuclei (Fig. 1B, v). There were numerous scattered T2/FLAIR hyperintense lesions in keeping with foci of gliosis, scattered throughout the white matter but prominent frontally (Fig. 1B, vi).

The 69-year-old maternal grandmother (I-2) had a mild right hemiplegia noted in the first year of life. She had learning difficulties at school but was of apparent normal intellect in adult life. She developed focal dyscognitive seizures in childhood and convulsive seizures at 47 years. She continued to have intractable focal seizures with episodes of status epilepticus. On examination, she has an obvious right hemiplegia. She refused neuroimaging and an EEG. She developed bilateral cataracts at 62 years of age.

Exome Analysis

There were 8,388 novel genotypes that were shared between the three affected individuals (Supplementary Table SIII). Of these, 409 variants passed our frequency and read quality criteria (Supplementary Table SIII). Only one out of 69 coding variants passed our filtering criteria and segregated with the clinical phenotype: *COL4A2* (c.2399G>A, p.G800E, CCDS41907.1; Supplementary Table SIV). Sanger sequencing confirmed the segregation of the variant in the proband, his brother, mother and grandmother. (Supplementary Fig. S2F). The COL4A2 p.G800E variant was predicted as likely disease causing by multiple bioinformatics tools (Fig. 1C). The glycine at position p.800 is invariable, from primates through to fish (Fig. 1D).

DISCUSSION

Lens opacities have not been recorded in previous *COL4A2* reports but are well recognized in *COL4A1*-related disorders, together with other ocular defects [Coupry et al., 2010]. To date, only one *COL4A2* mutation has been associated with ocular defects that included myopia, amblyopia, and abnormal optic discs (Supplementary Table SII) [Verbeek et al., 2012]. Lens opacities in humans have not been described and are noted here in young children with the *COL4A2* mutation; in contrast, the grandmother's cataracts may be an unrelated incidental age-related finding.

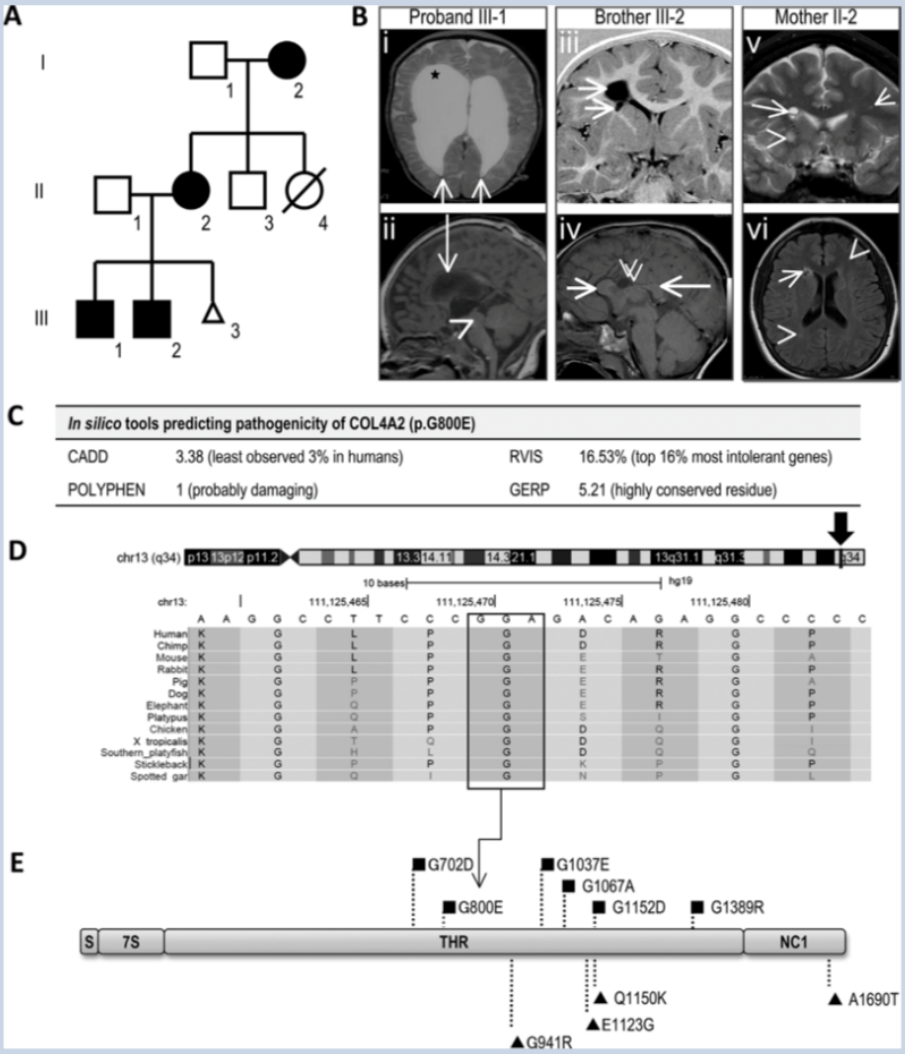


FIG. 1. Familial porencephaly caused by a *COL4A2* mutation. A: Pedigree. Affected individuals shown in black boxes (males) and circles (females). B: MRI brain scans. (i) Proband III-1: Axial T2 demonstrates right frontal porencephalic cyst (star) and hydrocephalus. Hemosiderin staining (T2 hypointensity) of the ependymal lining of the posterior horns of the lateral ventricles (arrows). (ii) Sagittal T1 weighted image demonstrates thinning of the corpus callosum (long arrow) and loss of visualisation of the aqueduct of Sylvius (arrowhead). (iii) III-2: Coronal T1 IR sequence shows multiple right frontal porencephalic cysts (arrows), the largest extends into the superior frontal lobe. (iv) Sagittal T1 weighted image demonstrates a dysmorphic corpus callosum the thick rostrum (short arrow) and splenium (long arrow) missing the callosal body (double arrowheads). (v) II-2: Coronal T2 weighted image of the frontal lobes demonstrate multiple small cysts in the right frontal periventricular white matter (long arrow). Numerous scattered foci of T2 hyperintensity are in keeping with gliosis related to small vessel disease (short arrows). (vi) Axial FLAIR demonstrates two small cysts in the right frontal periventricular white matter with surrounding hyperintense gliosis (arrow). Foci of gliosis in the left frontal lobe and right posterior temporal periventricular white matter (arrowheads). C: Predicted pathogenicity of *COL4A2* c.2399G>A, p.G800E determined by four prediction algorithms. D: Ideogram of chromosome 13 and the *COL4A2* locus (arrow) at 13q34. The mutated amino acid G800 is highly conserved across different orthologs (boxed). E: Domain structure and mutations of COL4A2. COL4A2 p.G800E (gray) is located within the triple helical region (THR). Published COL4A2 mutations predisposing to porencephaly (■) and intracerebral haemorrhage (▲) are shown. 7S, N-terminal propeptide domain; CADD, combined annotation dependent depletion; GERP, genomic evolutionary rate profiling; NC1, non-collagenous domain; RVIS, residual variation intolerance score; S, signal peptide.

The phenotypes observed in human *COL4A2*-related disorders are also found in mice carrying heterozygous *Col4a2* mutations (Supplementary Table SII) [Favor et al., 2007; Kuo et al., 2014]. The severity and penetrance of eye, brain, and muscular defects varies markedly among inbred (C57B1/6J) mice with the same *Col4a2* mutations [Kuo et al., 2014]. In humans, dominant *COL4A2* mutations also exhibit incomplete penetrance (Supplementary Table SII) [Verbeek et al., 2012; Yoneda et al., 2012].

The majority of *COL4A1* and *COL4A2* mutations occur at highly conserved glycine residues within the triple helical region (THR; Fig. 1D) [Favor et al., 2007]. These THR glycine mutations likely destabilize COL IV heterotrimer formation leading to the accumulation of a misfolded protein and subsequent chronic stress on the endoplasmic reticulum [Yoneda et al., 2012; Kuo et al., 2014]. In both mice and humans, *COL4A2* THR glycine mutations usually lead to severe, early onset cerebral pathologies but with a better prognosis compared to similar mutations in *COL4A1* (Supplementary Table SII) [Favor et al., 2007]. It has been hypothesized that the difference in phenotypic severity of *COL4A1* versus *COL4A2* mutations is due to the stoichiometry of $\alpha1{:}\alpha1{:}\alpha2$ per heterotrimer [Favor et al., 2007; Khoshnoodi et al., 2008; Jeanne et al., 2012]. A heterozygous *COL4A2* mutation will lead to 50% of COL4A1/A2 trimers with a mutant peptide, which will be less than a heterozygous *COL4A1* mutation with up to 75% of COL4A1/A2 trimers having a mutant peptide [Favor et al., 2007; Jeanne et al., 2012]. The location of the pathogenic variant relative to different functional domains of the COL4A1/A2 peptides also influences the clinical outcome [Des Parkin et al., 2011]. Non-glycine mutations (NGMs) within the THR and also the non-collagenous domain are likely low penetrance, risk alleles for late onset intracranial haemorrhage with good survival; a distinctly different phenotype to that caused by glycine mutations (Supplementary Table SV) [Jeanne et al., 2012]. Our review of *COL4A2* variations showed that all three previously reported NGMs [Jeanne et al., 2012] are present in the ExAC database while pathogenic THR glycine mutations are absent (Fig. 1E, Supplementary Table SVI). Nonetheless, some NGMs decrease the ratio of extracellular to intracellular COL4A2 relative to wild-type and also two of three variants induced stress in the endoplasmic reticulum in a similar way to THR glycine mutations, which supports their pathogenicity [Jeanne et al., 2012].

In conclusion, we identified a novel *COL4A2* (c.2399G>A, p.G800E, CCDS41907.1) mutation in an autosomal dominant family with porencephaly and ocular abnormalities. The mutation is predicted to be pathogenic based on the substitution of a highly conserved, critical THR-glycine residue and segregation with the phenotype. We extend the *COL4A2* phenotype spectrum, with the observation of lens opacities in some affected individuals, thus reducing the distinction between *COL4A1* and *COL4A2* mutations. This highlights the importance of simultaneous screening of both *COL4A1* and *COL4A2* genes for accurate diagnosis of early onset porencephaly and further ophthalmological management where indicated.

ACKNOWLEDGMENTS

We wish to thank the family for participating in our research, Alison Gardner for support with bioinformatics, eResearchSA, and the SA Cancer Council Genomics Facility. This project was supported by: NH&MRC grant 628952 to IES and JG, Senior Research Fellowship 1041920 (JG) and Practitioner Fellowship 1006110 (IES); WCH foundation MS McLeod research fellowship (MC); Health Research Council of New Zealand project grant 10/402 and CureKids New Zealand (LGS, IES, and SJP).

REFERENCES

Coupry I, Sibon I, Mortemousque B, Rouanet F, Mine M, Goizet C. 2010. Ophthalmological features associated with COL4A1 mutations. Arch Ophthalmol 128:483–489.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, Grp GPA. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491–498.

Des Parkin J, San Antonio JD, Pedchenko V, Hudson B, Jensen ST, Savige J. 2011. Mapping structural landmarks, ligand binding sites, and missense mutations to the collagen IV heterotrimers predicts major functional domains, novel interactions, and variation in phenotypes in inherited diseases affecting basement membranes. Hum Mutat 32:127–143.

Favor J, Gloeckner CJ, Janik D, Klempt M, Neuhauser-Klaus A, Pretsch W, Schmahl W, Quintanilla-Fend L. 2007. Type IV procollagen missense mutations associated with defects of the eye, vascular stability, the brain, kidney function and embryonic or postnatal viability in the mouse, Mus musculus: An extension of the Col4 alpha 1 allelic series and the identification of the first two Col4a2 mutant Alleles. Genetics 175:725–736.

Gunda B, Mine M, Kovacs T, Hornyak C, Bereczki D, Varallyay G, Rudas G, Audrezet MP, Tournier-Lasserve E. 2014. COL4A2 mutation causing adult onset recurrent intracerebral hemorrhage and leukoencephalopathy. J Neurol 261:500–503.

Haniel A, Welgelussen U, Kuhn K, Poschl E. 1995. Identification and characterization of a novel transcriptional silencer in the human collagen type-IV gene Col4a2. J Biol Chem 270:11209–11215.

Jeanne M, Labelle-Dumais C, Jorgensen J, Kauffman WB, Mancini GM, Favor J, Valant V, Greenberg SM, Rosand J, Gould DB. 2012. COL4A2 mutations impair COL4A1 and COL4A2 secretion and cause hemorrhagic stroke. Am J Hum Genet 90:91–101.

Khoshnoodi J, Pedchenko V, Hudson BG. 2008. Mammalian collagen IV. Microsc Res Tech 71:357–370.

Krumm N, Sudmant PH, Ko A, O'Roak BJ, Malig M, Coe BP, Quinlan AR, Nickerson DA, Eichler EE, Project NES. 2012. Copy number variation detection and genotyping from exome sequence data. Genome Res 22:1525–1532.

Kuo DS, Labelle-Dumais C, Mao M, Jeanne M, Kauffman WB, Allen J, Favor J, Gould DB. 2014. Allelic heterogeneity contributes to variability in ocular dysgenesis, myopathy and brain malformations caused by Col4a1 and Col4a2 mutations. Hum Mol Genet 23:1709–1722.

Lemmink HH, Schroder CH, Monnens LA, Smeets HJ. 1997. The clinical spectrum of type IV collagen mutations. Hum Mutat 9:477–499.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760.

Schmidt C, Fischer G, Kadner H, Genersch E, Kuhn K, Poschl E. 1993. Differential-effects of DNA-binding proteins on bidirectional

transcription from the common promoter region of human collagen type-Iv genes Col4a1 and Col4a2. Biochim Biophys Acta 1174:1–10.

Verbeek E, Meuwissen MEC, Verheijen FW, Govaert PP, Licht DJ, Kuo DS, Poulton CJ, Schot R, Lequin MH, Dudink J, Halley DJ, de Coo RIF, den Hollander JC, Oegema R, Gould DB, Mancini GMS. 2012. COL4A2 mutation associated with familial porencephaly and small-vessel disease. Eur J Hum Genet 20:844–851.

Wang K, Li MY, Hakonarson H. 2010. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38:e164.

Yoneda Y, Haginoya K, Arai H, Yamaoka S, Tsurusaki Y, Doi H, Miyake N, Yokochi K, Osaka H, Kato M, Matsumoto N, Saitsu H. 2012. De novo and inherited mutations in COL4A2, encoding the type IV collagen alpha 2 chain cause porencephaly. Am J Hum Genet 90: 86–90.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.

7.3.2 Genetic heterogeneity of Aicardi Syndrome revealed through whole exome and whole genome sequencing.

Chapter 4, **5** and **6** has been reformatted into a brief report, which will be submitted to the journal *American Journal of Human Genetics* under the same title as **Chapter 4**. Contributions of my co-authors are listed in **Section 4.4**.

Genetic heterogeneity of Aicardi Syndrome revealed using whole exome and whole genome sequencing.

Thuong T Ha[1], Rosemary Burgess[2], Morgan Newman[3], Melody Li[4], Alison E. Gardner [5], Atma Ivancevic[5], Duyen Pham[5], Raman Kumar[5], Christopher Reid[6], Nicholas Smith[5,7], Chirag Patel, Katherine Howell[6], Clare van Eyk [10], Michael Lardelli[3], Samuel F. Berkovic[8], Stephen Petrou[4], Ingrid E. Scheffer[6], Jozef Gecz[1,5,11] & Mark A. Corbett [5,10]

[1]School of Biological Sciences, University of Adelaide, Adelaide SA 5005, Australia. [2]Epilepsy Research Centre, University of Melbourne, Heidelberg VIC 3084, Australia. [3]Alzheimer's Disease Genetics Laboratory, University of Adelaide, Adelaide SA 5005, Australia. [4]Florey Institute of Neuroscience and Mental Health, Parkville VIC 3052, Australia. [5]School of Medicine, University of Adelaide, Adelaide SA 5005, Australia. [6]Department of Paediatrics, University of Melbourne, Parkville VIC 5032, Australia. [7]Department of Neurology, Women's and Children's Hospital, North Adelaide SA 5006, Australia. [8]Department of Medicine, Royal Melbourne Hospital, Parkville VIC 3010, Australia. [9]School of Medicine, Duke University, Durham NC 27710, England, [10]Australian Collaborative Cerebral Palsy Research Group, University of Adelaide, Adelaide SA 5005, Australia, [11]South Australian Health and Medical Research Institute, Adelaide SA 5005, Australia

Correspondence: mark.corbett@adelaide.edu.au

Abstract

Aicardi Syndrome (AIC) is a rare neurodevelopmental disorder recognized by a classical triad of chorioretinal lacunae, infantile spasms and agenesis of the corpus callosum. The revised diagnostic criteria of AIC has been broadened to include additional phenotypes outside of the classical triad that are frequently observed. AIC is traditionally thought to be an X-linked and male lethal disorder due to an almost exclusive prevalence in females. There have been numerous genetic and genomic investigations into AIC, however an X-linked cause has yet to be established. In this study, we performed whole exome and whole genome sequencing on a cohort of 13 females diagnosed with or suspected of AIC based on the current criteria. In five unrelated individuals, we identified *de novo* variants in *HCN1*, *KMT2B*, *SLF1*, *SZT2* and *WNT8B* respectively. Notably, a comprehensive genomic analyses of coding, non-coding and structural variation highlighted a distinct lack of X-linked candidates. Next, we assessed the likely pathogenicity of the autosomal variants we identified using a combination of: pre-existing in vitro assays (*HCN1* and *WNT8B*), published expression and phenotype studies in human or mice and morpholino knockdown in zebrafish (*Danio rerio*) embryos. Our findings show that causes underlying AIC are genetically heterogeneous but converge on molecular pathways central to cortical development.

Main Text

Aicardi Syndrome (MIM 304050; AIC) is an early infantile epileptic encephalopathy syndrome estimated to affect 4000 individuals worldwide (Kroner et al. 2008, Aicardi 2005). The diagnostic criteria for the disorder encompasses: a recognizable spectrum of malformations of cortical development including agenesis of the corpus callosum, intracranial cysts and white matter abnormalities; early onset infantile spasms and/or focal seizures within three to four months; and finally chorioretinal lacunae (CRL; **Figure 1.1**), which is the most consistently

observed phenotype among affected individuals to date (Aicardi 2005; Fruhman et al. 2012). In the absence of CRL, alternative eye phenotypes (coloboma and microphthalmia) can satisfy the phenotypic requirements for an AIC diagnosis provided that the accompanying onset and type of seizures and patterns of malformation of cortical development are typical of AIC (Aicardi 2005). In a meta-analysis of 405 cases, 83% of patients without CRL were not clinically diagnosed with AIC, despite sufficing the current diagnostic criteria (Aicard 2005, Kroner et al. 2008). This highlights the difficulty in recognising AIC, which requires combined expertise in neurology and ophthalmology and thus reiterates the need for identification of genetic causes to aid clinical examinations (Fruhman et al. 2012).

Unfortunately, genetic testing for AIC is currently unavailable because the cause(s) underlying AIC have yet to be identified. Early investigations into the genetics of AIC were predominantly focused on chromosome X for two reasons (Neidich et al. 1990; Nielsen et al. 1991; Hoag et al. 1997; Yilmaz et al. 2007; Eble et al. 2009; Pai et al. 2013). Firstly, the first chromosomal aberration reported in a suspected AIC female was an X/3 translocation (Ropers et al. 1982). Secondly, an X-linked male-lethal cause would best explain the predominance of the disease in females, which are represented in more than 99% of cases reported in literature. Despite 70 years of genetic and genomic investigations, an X-linked cause has yet to be established. By reviewing the genetic loci of variants reported in AIC thus far, we have yet to observe a recurring gene hit or subchromosomal overlap (**Table 1.7**).

The clinical presentations and diagnosis of AIC in the cases we reviewed from literature and our cohort are highly variable. There are three observable subgroups of patients in ours and published cases based on the presentation of: (i) the complete classical triad (undisputed diagnosis; **Figure 1.1**), (ii) one or two missing classical features replaced with major features that suffice the

Aicardi (2005) criteria (supported diagnosis) and two or more missing classical features replaced with minor features that do not meet the criteria (suspected diagnosis) (**Table 1.1**). In addition to the variable phenotypes, the efficacy of reported anti-epileptic treatments in Aicardi patients are also highly variable; which suggests diverse origins (post-synaptic processes and neuronal structures) of seizure development in AIC (**Table 1.5**). Based on the genetic and clinical variability observed thus far, we hypothesized that AIC is genetically heterogeneous.

Patient consent was obtained for all patients enrolled in this study. The clinical characteristics of the individuals in our AIC cohort are summarised in Table - Clinical, while the full phenotypic spectrum is described in **Table 4.2**. Serological tests for congenital viruses and metabolic levels, performed antenatally were negative for all probands. 9/12 probands had no family history of neurodevelopmental disorders, while 10/12 had unremarkable pregnancy and delivery. No genomic or genetic abnormalities were detected through prior clinical genetic testing for some of the patients that underwent karyotyping (1/12), array-based CNV screening (5/12) and candidate gene sequencing (1/12).

We performed whole exome sequencing (WES) on 13 patients (and parents if available) using the Agilent SureSelect EZ human exome version v3 capture kit (Agilent) and sequencing on the HiSeq2500 (Illumina) platform. Fastq files were processed, annotated and analysed as previously described (S Figure - Pipeline) (Ha et al. 2015). WES achieved a median coverage depth of 30X across 97.27% of the target capture (**Figure 4.3 & Table 4.3**). On average, approximately 60,000 SNP and Indels were called per exome and then filtered using the following selection criteria: (i) absent from unrelated in-house or published healthy cohorts, (ii) predicted deleterious by CADD and Polyphen2, (iii) transmitted in a *de novo* or recessive manner and (iv) associated with eye or cortical development. Four *de novo* variants remained after manual evaluation and Sanger

validation in the following genes: *HCN1*, *SLF1*, *SZT2* and *WNT8B* (**Table 4.4**) (Kircher et al. 2014; Adzhubei et al. 2010).

For the remaining cohort with ambiguous or no variant hits, we proceeded with whole genome sequencing (WGS) using the HiSeqXTen (Illumina) platform. Sequence reads were processed and variants called by the Australian Genome Research Facility, using Genome analysis toolkit (GATK) v3.5-0 best practices guidelines (Auwera et al. 2013). Variants within protein-coding regions were annotated and analysed equivalently to WES. WGS achieved a median coverage depth of 38X across 99.74% of the Human genome genome build 19 (Hg19) (**Figure 4.3 & Table 4.3**). On average, approximately 5 million SNP and Indels were called per genome. For our first parse, we analysed variants in protein-coding regions that may have been missed from WES. Here, an average of 42,000 exonic variants remained per proband and a further 82 fulfilled our disease-variant filtering criteria (**Table 4.6**). From our parent-proband trios (n=5), after we applied inheritance-based filtering, only one *de novo* variant, in *KMT2B*, validated via Sanger sequencing (**Table 4.5**). For the remaining two singletons, we compiled a list of potential variant hits that require segregation when parent DNA is available (**Appendix 7.2.1**).

We also screened for putative causative structural variants (SV; larger than 50bp) using four widely used programs: Delly, Lumpy, Manta and Retroseq (Rausch et al. 2012; Layer et al. 2014; Chen et al. 2015; Keane et al. 2012). We validated our SV discovery pipeline (**Figure 5.6**) using the Genome in a Bottle benchmark genome 'NA12878' (**Figure 5.7 & Figure 5.9**) (Zook et al. 2016) and applied it to 17 AIC genomes (5 proband-parent trios and two singletons) and approximately 100 unrelated in-house genomes Combining the variants from all the SV callers, there were approximately 82,000 SV per genome (**Table 5.2**). SV were automatically filtered using a custom bash script and the following parameters: unrelated in-house genomes, read

quality scores, benign SV from published datasets and assumed inheritance models (**Table 5.2**). After prioritising for variants that flanked Hg19 canonical transcripts or regulatory elements, none of these variants appeared *de novo in silico* (**Table 5.2**). As RNA sequencing data was unavailable for these probands, we were unable to prioritise regulatory variants.

Of the five putative candidate genes identified, *HCN1* and *WNT8B* had pre-existing in vitro assays that we could employ to test the effect of each variant on protein function. For *HCN1*, encoding the hyperpolarization-activated cyclic nucleotide-gated channel 1, we investigated the effect of the *HCN1* (c1625G>T; p.Cys542Phe; NM_021072) variant on membrane conductivity using voltage-clamp assay in frog (Xenopus laevis) oocytes. We found that oocytes expressing HCN1 (p.Cys542Phe) variant left-shifted the voltage dependence of activation suggesting a loss of channel function (**Figure 4.6**). These findings contrast with the gain of function or dominant negative effects of pathogenic HCN1 variants described by Nava et al. (2014). While most of the published variants reside in the transmembrane or intracellular domain of HCN1, the p.Cys542Phe variant is located in the cyclic nucleotide binding domain that is responsible for the direct regulation of the channel function (Wainger 2001). Previously reported individuals with mutations in *HCN1* have a Dravet-like syndrome with normal brain MRI and seizure onset between 4 and 13 years of age (Nava et al. 2014). The discovery of the HCN1 (p.Cys542Phe) variant also helped redefine the diagnosis for proband V from an AIC-like syndrome to Early Infantile Epileptic Encephalopathy (EIEE) 24 (MIM 615871).

To assess the effect of the *WNT8B* (c.209T>C; p.Leu70Pro; NM_003393) variant that was found in proband III (Table - Clinical; Figure - In vitro) on canonical WNT signalling we performed a TOPFlash assay in human embryonic kidney 293T cells. We observed that in the presence of wild-type WNT8B, the mutant had a dominant negative effect indicated by a significant

reduction in luciferase reporter gene activity (**Figure 4.7**). The mechanism by which WNT8B mutant interferes with the wildtype WNT signalling was not followed up in our study. We observed that the WNT8B mutant was consistently lower in abundance compared to the wildtype at equivalent levels of transfected plasmid, potentially indicating a loss of protein stability (**Figure 4.7**). *WNT8B* has yet to be implicated in any human inherited disorder however, the demonstrated roles for the zebrafish and mouse *WNT8B* orthologues in specifying the anterior neuroectoderm, regulating axon guidance and specification of retinal progenitor cells, suggest this gene as a very plausible candidate for AIC (Kim et al. 2002; Cavodeassi et al. 2005; Hofmeister & Key 2013, Liu 2012).

For the remaining genes that did not have established molecular or biochemical assays and are yet to be implicated in both eye and brain development, we used morpholino knockdown in zebrafish (*Danio rerio*) embryos to screen for phenotypes that were reminiscent of AIC. We included tead1, which is the only other previously identified gene implicated AIC that is conserved in zebrafish. We designed a study in which the morpholino injections and data analysis were carried out under blinded conditions and morphant classification counts were verified by three independent researchers. More than 90% of uninjected and scrambled morpholino control group displayed normal embryonic development with or without minor body abnormalities such as bent tail (**Figure 4.8**). In contrast, our experimental morphants displayed a spectrum of developmental abnormalities including: lack of eye pigmentation, body curvatures, pericardial oedema, tail defects and head malformations (**Figure 4.8**). The most consistent phenotype among the slf1, tead1 and wnt8b morphants was the lack of eye pigmentation, which was patchy in appearance and often unilateral (**Figure 4.8**). A lack of eye pigmentation was also observed in szt2 morphants, however as 72 hours post fertilization (hpf) szt2 morphants often resembled 48 hpf morphants, this may be attributed to developmental delay. There are known limitations to

disease modeling in zebrafish (Kok et al. 2015) and the morpholino knockdown approach models a loss of function whereas our *de novo* variants are likely to be dominant negative. Nonetheless, we showed a unifying morphant phenotype of AIC-like eye and brain defects among at least 3/5 genes tested.

We identified likely pathogenic variants in different autosomal genes from five unrelated individuals. The lack of X-linked candidates identified from others and our genetic studies may be attributed to the paired-end sequencing (PES) approach, which has limited resolution for variants in repetitive DNA sequences, regions with high or low GC bias and non-random DNA fragmentation bias (Treangen et al. 2011; Aird et al. 2011). Structural variant detection from PES is reliant on balanced and paired-end mapping, which may miss larger or complex structural variants such as: dispersed duplications, unbalanced translocations, microinversions, large insertions and repeat expansion mutations. Our 38X average sequencing coverage was also inadequate for detecting somatic variants with low levels of mosaicism (less than 5% of allelic fraction); which could be disguised as false positives. In addition, a majority of our sequencing was based on peripheral blood DNA, thus we were unable to comprehensively screen for brain-specific variants. Finally, due to the lack of paired WGS and RNA sequencing data, we were unable to prioritise and interpret the significance of some non-coding variants; especially in regions lacking functional annotations. For the remaining unresolved individuals in our cohort, an X-linked cause may still be possible.

The most enigmatic aspect and least understood aspect of AIC is the almost exclusive female prevalence. In light of our genetic data that opposes the X-linked-male-lethal hypothesis and lack of males affected by AIC in our study, we explored other potential mechanisms that can lead to female-biased traits. One possibility would be the participation of the implicated AIC genes in a

sexually dimorphic pathway such as Wnt signalling, which has been linked to female neurodevelopmental disorders in X-linked Wnt-related genes such as *DDX3X*, *USP9X* and *PORCN* (Shen et al. 2017; Snijders et al. 2015; Murtaza et al. 2015; Grzeschik et al. 2007). In our study, we identified a WNT8B variant that affected canonical wnt signalling; while the Shrauwen et al. (2015) study showed upregulation of the wnt receptor LRP5 by RNA-Seq and also discovered a pathogenic variant in *TEAD1*. While TEAD1 is involved in the hippo pathway, there is evidence for cross-regulation between the hippo and wnt/β-catenin pathways (Kim & Jho 2014).

Another hypothesis would be the interaction with or modulation by female sex hormones and our candidate AIC genes. In MCF-7 (breast cancer) cells, it has been shown that *KMT2B* can epigenetically regulate estradiol (E2) dependent transcription via estrogen receptor alpha (ERα) (Su et al. 2012). Meanwhile in an independent study in MCF-7 cells, *WNT8B* expression was significantly upregulated by E2 (Saitoh et al. 2002). It is interesting to note that the expression of ERα was detected in the retinal pigment epithelium (RPE) of young female eyes but not in men or postmenopausal women (Ogueta et al. 1999). The role of estrogen in retinal pathologies are well established and it is tempting to speculate that CRL (the pathognomonic feature of AIC) may be a sex-limited trait that appears in an excess of females compared to males (Gupta et al. 2005; Wagner et al. 2008). Nonetheless, more transcriptomic and epigenomic data is needed to assess whether E2 directly or indirectly regulate ours and future implicated AIC genes. Other sex-limited traits that overlap with AIC, such as early infantile epileptic encephalopathy and isolated ACC, have been reported in affected females with autosomal mutations in *FOXG1* (14q12), *YWHAG* (7q11.23), *HCN1* (5p12) and *DCC* (18q21.2) (Mitter et al. 2017; Guella et al. 2017; Marsh et al. 2017; Nava et al. 2014).

Overall, our genetic findings highlight the importance of proband-parent trio analysis for unresolved and clinically heterogeneous disorders like Aicardi Syndrome. We were able to provide a differential diagnosis for proband T22352 and expand the phenotype spectrum of *HCN1* variants. We also identified two new variants in known neurodevelopmental disorder genes, *SZT2* and *KMT2B*. Although these variants have good bioinformatic support for pathogenicity, the inheritance of the *SZT2* variant in proband T25217 and the phenotype of proband T22842 with the *KMT2B* variant do not match with the published cases reported thus far (Table - De novo & S Table - Recurrence). Further functional validation is needed to establish whether these variants are true drivers of the Aicardi phenotypes in the respective individuals or incidental findings. We also discovered likely pathogenic variants in *WNT8B* and *SLF1* in probands T22101 and T17262 respectively. These two genes have yet to be implicated in human neurodevelopmental disorders (**Table - 4.5 & Table 4.6**). Although both of these genes are biologically distinct (**Table 4.7**), they share similarities in their zebrafish morphant phenotype (**Figure 4.8**) and expression in embryonic mouse brain (**Figure 4.14 & Figure 4.15**). Both of the individuals with the *WNT8B* and *SLF1* variant display the classical triad (diagnostic for AIC; **Figure 1.1**) with overlapping accompanying phenotypes including: ventriculomegaly, atrophy of cerebellar hemisphere, developmental delay, scoliosis and skeletal abnormalities (**Table 4.1 & Table 4.2**). Based on our in vivo findings and published expression studies, *WNT8B* and *SLF1* appear to be good candidates for AIC that strongly suggest that this enigmatic disorder is genetically heterogeneous and that not all cases will be X-linked.

Acknowledgements

We wish to thank the families involved in this study. Thank you to Sayaka Kayumi for technical assistance. This work was supported by NHMRC program grant (SB,IS,JG,SP), NHMRC

fellowships (to be listed), MS McLeod Fellowship from the WCH foundation and Cerebral Palsy Alliance Career Development Award (MC).

Web Resources

ANNOVAR	http://annovar.openbioinformatics.org/
CADD	http://cadd.gs.washington.edu/
ExAC Browser	http://exac.broadinstitute.org/
OMIM	http://www.omim.org/
GenBank	https://www.ncbi.nlm.nih.gov/genbank/
GERP	http://mendel.stanford.edu/SidowLab/downloads/gerp/
GATK	https://software.broadinstitute.org/gatk/
DGV	http://dgv.tcag.ca/dgv/app/home
PolyPhen2	http://genetics.bwh.harvard.edu/pph2/
Decipher	https://decipher.sanger.ac.uk/
GONL	http://www.nlgenome.nl/
ISCA	http://dbsearch.clinicalgenome.org/search/
ENCODE	https://www.encodeproject.org/
ORegAnno	http://www.oreganno.org/
1000 Genomes	http://www.internationalgenome.org
CG69	http://www.completegenomics.com/public-data/69-genomes/
UK10K	https://www.uk10k.org/
Wellderly	https://www.stsiweb.org/translational-research/genomic-medicine/wellderly/

References

1. Abbas, Q., Raza, S.M., Biyabani, A.A., and Jaffar, M.A. (2016). A Review of Computational Methods for Finding Non-Coding RNA Genes. Genes (Basel) 7.

2. Ackley, B.D., Crew, J.R., Elamaa, H., Pihlajaniemi, T., Kuo, C.J., and Kramer, J.M. (2001). The NC1/endostatin domain of Caenorhabditis elegans type XVIII collagen affects cell migration and axon guidance. J Cell Biol 152, 1219-1232.

3. Adam, M.P., Ardinger, H.H., and Pagon, R.A. (1993-2018). GeneReviews. In. (Seattle, University of Washington.

4. Adams, D.J., van der Weyden, L., Gergely, F.V., Arends, M.J., Ng, B.L., Tannahill, D., Kanaar, R., Markus, A., Morris, B.J., and Bradley, A. (2005). BRCTx is a novel, highly conserved RAD18-interacting protein. Mol Cell Biol 25, 779-788.

5. Aguglia, U., Gambardella, A., Breedveld, G.J., Oliveri, R.L., Le Piane, E., Messina, D., Quattrone, A., and Heutink, P. (2004). Suggestive evidence for linkage to chromosome 13qter for autosomal dominant type 1 porencephaly. Neurology 62, 1613-1615.

6. Aicardi, J. (2005). Aicardi syndrome. Brain Dev 27, 164-171.

7. Aicardi, J., Chevrie, J.J., and Rousselie, F. (1969). [Spasma-in-flexion syndrome, callosal agenesis, chorioretinal abnormalities]. Arch Fr Pediatr 26, 1103-1120.

8. Aird, D., Ross, M.G., Chen, W.S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D.B., Nusbaum, C., and Gnirke, A. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol 12, R18.

9. Alkan, C., Coe, B.P., and Eichler, E.E. (2011). Genome structural variation discovery and genotyping. Nat Rev Genet 12, 363-376.

10. Anderson, S., Menten, B., Kogelenberg, M., Robertson, S., Waginger, M., Mentzel, H.J., Brandl, U., Skirl, G., and Willems, P. (2009). Aicardi syndrome in a male patient. Neuropediatrics 40, 39-42.

11. Ang, S.J., Stump, R.J., Lovicu, F.J., and McAvoy, J.W. (2004). Spatial and temporal expression of Wnt and Dickkopf genes during murine lens development. Gene Expr Patterns 4, 289-295.

12. Attanasio, C., Reymond, A., Humbert, R., Lyle, R., Kuehn, M.S., Neph, S., Sabo, P.J., Goldy, J., Weaver, M., Haydock, A., et al. (2008). Assaying the regulatory potential of mammalian conserved non-coding sequences in human cells. Genome Biol 9, R168.

13. Badenas, C., Praga, M., Tazon, B., Heidet, L., Arrondel, C., Armengol, A., Andres, A., Morales, E., Camacho, J.A., Lens, X., et al. (2002). Mutations in theCOL4A4 and COL4A3 genes cause familial benign hematuria. J Am Soc Nephrol 13, 1248-1254.

14. Baffet, A.D., Carabalona, A., Dantas, T.J., Doobin, D.D., Hu, D.J., and Vallee, R.B. (2016). Cellular and subcellular imaging of motor protein-based behavior in embryonic rat brain. Methods Cell Biol 131, 349-363.

15. Baierl, P., Markl, A., Thelen, M., and Laub, M.C. (1988). MR imaging in Aicardi syndrome. AJNR Am J Neuroradiol 9, 805-806.

16. Bajwa, S.K., Singh, S., and Bajwa, S.J. (2012). Ocular tissue responses to sex hormones. Indian J Endocrinol Metab 16, 488-489.

17. Baker, M. (2012). Structural variation: the genome's hidden architecture. Nat Methods 9, 133-137.

18. Barker, D.F., Hostikka, S.L., Zhou, J., Chow, L.T., Oliphant, A.R., Gerken, S.C., Gregory, M.C., Skolnick, M.H., Atkin, C.L., and Tryggvason, K. (1990). Identification of mutations in the COL4A5 collagen gene in Alport syndrome. Science 248, 1224-1227.

19. Barkovich, A.J., Kuzniecky, R.I., Jackson, G.D., Guerrini, R., and Dobyns, W.B. (2005). A developmental and genetic classification for malformations of cortical development. Neurology 65, 1873-1887.

20. Barros, C.S., Franco, S.J., and Muller, U. (2011). Extracellular matrix: functions in the nervous system. Cold Spring Harb Perspect Biol 3, a005108.

21. Basel-Vanagaite, L., Hershkovitz, T., Heyman, E., Raspall-Chaure, M., Kakar, N., Smirin-Yosef, P., Vila-Pueyo, M., Kornreich, L., Thiele, H., Bode, H., et al. (2013). Biallelic SZT2 mutations cause infantile encephalopathy with epilepsy and dysmorphic corpus callosum. Am J Hum Genet 93, 524-529.

22. Beery, A.K., and Zucker, I. (2011). Sex bias in neuroscience and biomedical research. Neurosci Biobehav Rev 35, 565-572.

23. Bertoni, J.M., von Loh, S., and Allen, R.J. (1979). The Aicardi syndrome: report of 4 cases and review of the literature. Ann Neurol 5, 475-482.

24. Blackshaw, S., Harpavat, S., Trimarchi, J., Cai, L., Huang, H., Kuo, W.P., Weber, G., Lee, K., Fraioli, R.E., Cho, S.H., et al. (2004). Genomic analysis of mouse retinal development. PLoS Biol 2, E247.

25. Bockaert, J., and Marin, P. (2015). mTOR in Brain Physiology and Pathologies. Physiol Rev 95, 1157-1187.

26. Boye, E., Mollet, G., Forestier, L., Cohen-Solal, L., Heidet, L., Cochat, P., Grunfeld, J.P., Palcoux, J.B., Gubler, M.C., and Antignac, C. (1998). Determination of the genomic structure of the COL4A4 gene and of novel mutations causing autosomal recessive Alport syndrome. Am J Hum Genet 63, 1329-1340.

27. Boye, E., Vetrie, D., Flinter, F., Buckle, B., Pihlajaniemi, T., Hamalainen, E.R., Myers, J.C., Bobrow, M., and Harris, A. (1991). Major rearrangements in the alpha 5(IV) collagen gene in three patients with Alport syndrome. Genomics 11, 1125-1132.

28. Breedveld, G., de Coo, I.F., Lequin, M.H., Arts, W.F., Heutink, P., Gould, D.B., John, S.W., Oostra, B., and Mancini, G.M. (2006). Novel mutations in three families confirm a major role of COL4A1 in hereditary porencephaly. J Med Genet 43, 490-495.

29. Burch-Smith, R., Ordonez, N.G., Ginsberg, L.E., Ater, J.L., and El Naggar, A.K. (2012). Oral extragonadal yolk sac tumor in a patient with Aicardi syndrome: putative origin and differential diagnosis. Hum Pathol 43, 939-942.

30. Bursztejn, A.C., Bronner, M., Peudenier, S., Gregoire, M.J., Jonveaux, P., and Nemos, C. (2009). Molecular characterization of a monosomy 1p36 presenting as an Aicardi syndrome phenocopy. Am J Med Genet A 149A, 2493-2500.

31. Bush, W.S., and Moore, J.H. (2012). Chapter 11: Genome-wide association studies. PLoS Comput Biol 8, e1002822.

32. Cabrera, M.T., Winn, B.J., Porco, T., Strominger, Z., Barkovich, A.J., Hoyt, C.S., Wakahiro, M., and Sherr, E.H. (2011). Laterality of brain and ocular lesions in Aicardi syndrome. Pediatr Neurol 45, 149-154.

33. Caglayan, A.O., Baranoski, J.F., Aktar, F., Han, W., Tuysuz, B., Guzel, A., Guclu, B., Kaymakcalan, H., Aktekin, B., Akgumus, G.T., et al. (2014). Brain malformations associated with Knobloch syndrome--review of literature, expanding clinical spectrum, and identification of novel mutations. Pediatr Neurol 51, 806-813 e808.

34. Caldon, C.E. (2014). Estrogen signalling and the DNA damage response in hormone dependent breast cancers. Front Oncol 4, 106.

35. Cao, H., Hastie, A.R., Cao, D., Lam, E.T., Sun, Y., Huang, H., Liu, X., Lin, L., Andrews, W., Chan, S., et al. (2014). Rapid detection of structural variation in a human genome using nanochannel-based genome mapping technology. Gigascience 3, 34.

36. Carss, K.J., Arno, G., Erwood, M., Stephens, J., Sanchis-Juan, A., Hull, S., Megy, K., Grozeva, D., Dewhurst, E., Malka, S., et al. (2017). Comprehensive Rare Variant Analysis via Whole-Genome Sequencing to Determine the Molecular Pathology of Inherited Retinal Disease. Am J Hum Genet 100, 75-90.

37. Cavodeassi, F., Carreira-Barbosa, F., Young, R.M., Concha, M.L., Allende, M.L., Houart, C., Tada, M., and Wilson, S.W. (2005). Early stages of zebrafish eye formation require the coordinated activity of Wnt11, Fz5, and the Wnt/beta-catenin pathway. Neuron 47, 43-56.

38. Chappelow, A.V., Reid, J., Parikh, S., and Traboulsi, E.I. (2008). Aicardi syndrome in a genotypic male. Ophthalmic Genet 29, 181-183.

39. Chatr-Aryamontri, A., Oughtred, R., Boucher, L., Rust, J., Chang, C., Kolas, N.K., O'Donnell, L., Oster, S., Theesfeld, C., Sellam, A., et al. (2017). The BioGRID interaction database: 2017 update. Nucleic Acids Res 45, D369-D379.

40. Chen, T.H., Chao, M.C., Lin, L.C., Jong, Y.J., Yang, S.N., Lai, Y.H., and Chen, H.L. (2009). Aicardi syndrome in a 47, XXY male neonate with lissencephaly and holoprosencephaly. J Neurol Sci 278, 138-140.

41. Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Kallberg, M., Cox, A.J., Kruglyak, S., and Saunders, C.T. (2016). Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics 32, 1220-1222.

42. Chiang, C., Scott, A.J., Davis, J.R., Tsang, E.K., Li, X., Kim, Y., Hadzic, T., Damani, F.N., Ganel, L., Consortium, G.T., et al. (2017). The impact of structural variation on human gene expression. Nat Genet 49, 692-699.

43. Clapham, K.R., Yu, T.W., Ganesh, V.S., Barry, B., Chan, Y., Mei, D., Parrini, E., Funalot, B., Dupuis, L., Nezarati, M.M., et al. (2012). FLNA genomic rearrangements cause periventricular nodular heterotopia. Neurology 78, 269-278.

44. Coe, B.P., Witherspoon, K., Rosenfeld, J.A., van Bon, B.W., Vulto-van Silfhout, A.T., Bosco, P., Friend, K.L., Baker, C., Buono, S., Vissers, L.E., et al. (2014). Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nat Genet 46, 1063-1071.

45. Conrad, D.F., Bird, C., Blackburne, B., Lindsay, S., Mamanova, L., Lee, C., Turner, D.J., and Hurles, M.E. (2010). Mutation spectrum revealed by breakpoint sequencing of human germline CNVs. Nat Genet 42, 385-391.

46. Consortium, E.P. (2004). The ENCODE (ENCyclopedia Of DNA Elements) Project. Science 306, 636-640.

47. Constad, W.H., Wagner, R.S., and Caputo, A.R. (1985). Aicardi syndrome in one dizygotic twin. Pediatrics 76, 450-453.

48. Cooper, G.M., Coe, B.P., Girirajan, S., Rosenfeld, J.A., Vu, T.H., Baker, C., Williams, C., Stalker, H., Hamid, R., Hannig, V., et al. (2011). A copy number variation morbidity map of developmental delay. Nat Genet 43, 838-846.

49. Corbett, M.A., Turner, S.J., Gardner, A., Silver, J., Stankovich, J., Leventer, R.J., Derry, C.P., Carroll, R., Ha, T., Scheffer, I.E., et al. (2017). Familial epilepsy with anterior polymicrogyria as a presentation of COL18A1 mutations. Eur J Med Genet 60, 437-443.

50. Costa, T., Greer, W., Rysiecki, G., Buncic, J.R., and Ray, P.N. (1997). Monozygotic twins discordant for Aicardi syndrome. J Med Genet 34, 688-691.

51. Coupry, I., Sibon, I., Mortemousque, B., Rouanet, F., Mine, M., and Goizet, C. (2010). Ophthalmological features associated with COL4A1 mutations. Arch Ophthalmol 128, 483-489.

52. Curatolo, P., Libutti, G., and Dallapiccola, B. (1980). Aicardi syndrome in a male infant. J Pediatr 96, 286-287.

53. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics 27, 2156-2158.

54. de Koning, A.P., Gu, W., Castoe, T.A., Batzer, M.A., and Pollock, D.D. (2011). Repetitive elements may comprise over two-thirds of the human genome. PLoS Genet 7, e1002384.

55. de Vries, L.S., Koopman, C., Groenendaal, F., Van Schooneveld, M., Verheijen, F.W., Verbeek, E., Witkamp, T.D., van der Worp, H.B., and Mancini, G. (2009). COL4A1 mutation in two preterm siblings with antenatal onset of parenchymal hemorrhage. Ann Neurol 65, 12-18.

56. Deciphering Developmental Disorders, S. (2015). Large-scale discovery of novel genetic causes of developmental disorders. Nature 519, 223-228.

57. Deciphering Developmental Disorders, S. (2017). Prevalence and architecture of de novo mutations in developmental disorders. Nature 542, 433-438.

58. Delon, I., Taylor, A., Molenda, A., Drummond, J., Oakhill, K., Girling, A., Liu, H., Whittaker, J., Treacy, R., and Tischkowitz, M. (2013). A germline mosaic BRCA1 exon deletion in a woman with bilateral basal-like breast cancer. Clin Genet 84, 297-299.

59. Deml, B., Reis, L.M., Maheshwari, M., Griffis, C., Bick, D., and Semina, E.V. (2014). Whole exome analysis identifies dominant COL4A1 mutations in patients with complex ocular phenotypes involving microphthalmia. Clin Genet 86, 475-481.

60. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43, 491-498.

61. Dermitzakis, E.T., Reymond, A., Scamuffa, N., Ucla, C., Kirkness, E., Rossier, C., and Antonarakis, S.E. (2003). Evolutionary discrimination of mammalian conserved non-genic sequences (CNGs). Science 302, 1033-1035.

62. Dewing, P., Shi, T., Horvath, S., and Vilain, E. (2003). Sexually dimorphic gene expression in mouse brain precedes gonadal differentiation. Brain Res Mol Brain Res 118, 82-90.

63. Di Lullo, E., and Kriegstein, A.R. (2017). The use of brain organoids to investigate neural development and disease. Nat Rev Neurosci 18, 573-584.

64. Dieffenbach, C.W., Lowe, T.M., and Dveksler, G.S. (1993). General concepts for PCR primer design. PCR Methods Appl 3, S30-37.

65. Diez-Roux, G., Banfi, S., Sultan, M., Geffers, L., Anand, S., Rozado, D., Magen, A., Canidio, E., Pagani, M., Peluso, I., et al. (2011). A high-resolution anatomical atlas of the transcriptome in the mouse embryo. PLoS Biol 9, e1000582.

66. Donnenfeld, A.E., Packer, R.J., Zackai, E.H., Chee, C.M., Sellinger, B., and Emanuel, B.S. (1989). Clinical, cytogenetic, and pedigree findings in 18 cases of Aicardi syndrome. Am J Med Genet 32, 461-467.

67. Dunford, A., Weinstock, D.M., Savova, V., Schumacher, S.E., Cleary, J.P., Yoda, A., Sullivan, T.J., Hess, J.M., Gimelbrant, A.A., Beroukhim, R., et al. (2017). Tumor-suppressor genes that escape from X-inactivation contribute to cancer sex bias. Nat Genet 49, 10-16.

68. Eble, T.N., Sutton, V.R., Sangi-Haghpeykar, H., Wang, X., Jin, W., Lewis, R.A., Fang, P., and Van den Veyver, I.B. (2009). Non-random X chromosome inactivation in Aicardi syndrome. Hum Genet 125, 211-216.

69. Epi, K.C., Epilepsy Phenome/Genome, P., Allen, A.S., Berkovic, S.F., Cossette, P., Delanty, N., Dlugos, D., Eichler, E.E., Epstein, M.P., Glauser, T., et al. (2013). De novo mutations in epileptic encephalopathies. Nature 501, 217-221.

70. Esteller, M. (2011). Non-coding RNAs in human disease. Nat Rev Genet 12, 861-874.

71. Euro, E.-R.E.S.C., Epilepsy Phenome/Genome, P., and Epi, K.C. (2014). De novo mutations in synaptic transmission genes including DNM1 cause epileptic encephalopathies. Am J Hum Genet 95, 360-370.

72. Faisal, M., Kim, H., and Kim, J. (2014). Sexual differences of imprinted genes' expression levels. Gene 533, 434-438.

73. Falcone, M., Yariz, K.O., Ross, D.B., Foster, J., 2nd, Menendez, I., and Tekin, M. (2013). An amino acid deletion inSZT2 in a family with non-syndromic intellectual disability. PLoS One 8, e82810.

74. Fanciulli, M., Norsworthy, P.J., Petretto, E., Dong, R., Harper, L., Kamesh, L., Heward, J.M., Gough, S.C., de Smith, A., Blakemore, A.I., et al. (2007). FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. Nat Genet 39, 721-723.

75. Favor, J., Gloeckner, C.J., Janik, D., Klempt, M., Neuhauser-Klaus, A., Pretsch, W., Schmahl, W., and Quintanilla-Fend, L. (2007). Type IV procollagen missense mutations associated with defects of the eye, vascular stability, the brain, kidney function and embryonic or postnatal viability in the mouse, Mus musculus: an extension of the Col4a1 allelic series and the identification of the first two Col4a2 mutant alleles. Genetics 175, 725-736.

76. Finielz, P., Cartault, F., Chuet, C., and Genin, R. (1998). Alport syndrome in Reunion Island: phenotypic heterogeneity of the recessive-autosomal form. Nephron 79, 237.

77. Fischer, T., Guimera, J., Wurst, W., and Prakash, N. (2007). Distinct but redundant expression of the Frizzled Wnt receptor genes at signalling centers of the developing mouse brain. Neuroscience 147, 693-711.

78. Fotaki, V., Larralde, O., Zeng, S., McLaughlin, D., Nichols, J., Price, D.J., Theil, T., and Mason, J.O. (2010). Loss of Wnt8b has no overt effect on hippocampus development but leads to altered Wnt gene expression levels in dorsomedial telencephalon. Dev Dyn 239, 284-296.

79. Frankel, W.N., Yang, Y., Mahaffey, C.L., Beyer, B.J., and O'Brien, T.P. (2009). Szt2, a novel gene for seizure threshold in mice. Genes Brain Behav 8, 568-576.

80. Fruhman, G., Eble, T.N., Gambhir, N., Sutton, V.R., Van den Veyver, I.B., and Lewis, R.A. (2012). Ophthalmologic findings in Aicardi syndrome. J AAPOS 16, 238-241.

81. Ganel, L., Abel, H.J., FinMetSeq, C., and Hall, I.M. (2017). SVScore: an impact prediction tool for structural variation. Bioinformatics 33, 1083-1085.

82. Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. arXiv.

83. Gekeler, F., Shinoda, K., Junger, M., Bartz-Schmidt, K.U., and Gelisken, F. (2006). Familial retinal arterial tortuosity associated with tortuosity in nail bed capillaries. Arch Ophthalmol 124, 1492-1494.

84. Genome of the Netherlands, C. (2014). Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nat Genet 46, 818-825.

85. Genomes Project, C., Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., et al. (2015). A global reference for human genetic variation. Nature 526, 68-74.

86. Gilissen, C., Hehir-Kwa, J.Y., Thung, D.T., van de Vorst, M., van Bon, B.W., Willemsen, M.H., Kwint, M., Janssen, I.M., Hoischen, A., Schenck, A., et al. (2014). Genome sequencing identifies major causes of severe intellectual disability. Nature 511, 344-347.

87. Gilissen, C., Hoischen, A., Brunner, H.G., and Veltman, J.A. (2012). Disease gene identification strategies for exome sequencing. Eur J Hum Genet 20, 490-497.

88. Girirajan, S., Brkanac, Z., Coe, B.P., Baker, C., Vives, L., Vu, T.H., Shafer, N., Bernier, R., Ferrero, G.B., Silengo, M., et al. (2011). Relative burden of large CNVs on a range of neurodevelopmental phenotypes. PLoS Genet 7, e1002334.

89. Glasmacher, M.A., Sutton, V.R., Hopkins, B., Eble, T., Lewis, R.A., Park Parsons, D., and Van den Veyver, I.B. (2007). Phenotype and management of Aicardi syndrome: new findings from a survey of 69 children. J Child Neurol 22, 176-184.

90. Goodwin, S., McPherson, J.D., and McCombie, W.R. (2016). Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 17, 333-351.

91. Gould, D.B., Phalan, F.C., Breedveld, G.J., van Mil, S.E., Smith, R.S., Schimenti, J.C., Aguglia, U., van der Knaap, M.S., Heutink, P., and John, S.W. (2005). Mutations in Col4a1 cause perinatal cerebral hemorrhage and porencephaly. Science 308, 1167-1171.

92. Gould, D.B., Phalan, F.C., van Mil, S.E., Sundberg, J.P., Vahedi, K., Massin, P., Bousser, M.G., Heutink, P., Miner, J.H., Tournier-Lasserve, E., et al. (2006). Role of COL4A1 in small-vessel disease and hemorrhagic stroke. N Engl J Med 354, 1489-1496.

93. Grayton, H.M., Fernandes, C., Rujescu, D., and Collier, D.A. (2012). Copy number variations in neurodevelopmental disorders. Prog Neurobiol 99, 81-91.

94. Grosso, S., Lasorella, G., Russo, A., Galluzzi, P., Morgese, G., and Balestri, P. (2007). Aicardi syndrome with favorable outcome: case report and review. Brain Dev 29, 443-446.

95. Grzeschik, K.H., Bornholdt, D., Oeffner, F., Konig, A., del Carmen Boente, M., Enders, H., Fritz, B., Hertl, M., Grasshoff, U., Hofling, K., et al. (2007). Deficiency of PORCN, a regulator of Wnt signalling, is associated with focal dermal hypoplasia. Nat Genet 39, 833-835.

96. Guerrini, R., and Dobyns, W.B. (2014). Malformations of cortical development: clinical features and genetic causes. Lancet Neurol 13, 710-726.

97. Gunda, B., Mine, M., Kovacs, T., Hornyak, C., Bereczki, D., Varallyay, G., Rudas, G., Audrezet, M.P., and Tournier-Lasserve, E. (2014). COL4A2 mutation causing adult onset recurrent intracerebral hemorrhage and leukoencephalopathy. J Neurol 261, 500-503.

98. Guo, C., Van Damme, B., Vanrenterghem, Y., Devriendt, K., Cassiman, J.J., and Marynen, P. (1995). Severe alport phenotype in a woman with two missense mutations in the same COL4A5 gene and preponderant inactivation of the X chromosome carrying the normal allele. J Clin Invest 95, 1832-1837.

99. Ha, T.T., Sadleir, L.G., Mandelstam, S.A., Paterson, S.J., Scheffer, I.E., Gecz, J., and Corbett, M.A. (2016). A mutation in COL4A2 causes autosomal dominant porencephaly with cataracts. Am J Med Genet A 170A, 1059-1063.

100. Haniel, A., Welge-Lussen, U., Kuhn, K., and Poschl, E. (1995). Identification and characterization of a novel transcriptional silencer in the human collagen type IV gene COL4A2. J Biol Chem 270, 11209-11215.

101. Hattori, H., Yamairi, T., Kim, M., Matsuoka, O., Murata, R., and Isshiki, G. (1984). [Aicardi syndrome in one of dizygotic twins]. No To Hattatsu 16, 317-321.

102. Heidet, L., Arrondel, C., Forestier, L., Cohen-Solal, L., Mollet, G., Gutierrez, B., Stavrou, C., Gubler, M.C., and Antignac, C. (2001). Structure of the human type IV collagen gene COL4A3 and mutations in autosomal Alport syndrome. J Am Soc Nephrol 12, 97-106.

103. Herman, M.A., and Rosen, E.D. (2015). Making Biological Sense of GWAS Data: Lessons from the FTO Locus. Cell Metab 22, 538-539.

104. Hes, F.J., Nielsen, M., Bik, E.C., Konvalinka, D., Wijnen, J.T., Bakker, E., Vasen, H.F., Breuning, M.H., and Tops, C.M. (2008). Somatic APC mosaicism: an underestimated cause of polyposis coli. Gut 57, 71-76.

105. Hoag, H.M., Taylor, S.A., Duncan, A.M., and Khalifa, M.M. (1997). Evidence that skewed X inactivation is not needed for the phenotypic expression of Aicardi syndrome. Hum Genet 100, 459-464.

106. Hofmeister, W., and Key, B. (2013). Frizzled-3a and Wnt-8b genetically interact during forebrain commissural formation in embryonic zebrafish. Brain Res 1506, 25-34.

107. Hopkins, B., Sutton, V.R., Lewis, R.A., Van den Veyver, I., and Clark, G. (2008). Neuroimaging aspects of Aicardi syndrome. Am J Med Genet A 146A, 2871-2878.

108. Hopkins, I.J., Humphrey, I., Keith, C.G., Susman, M., Webb, G.C., and Turner, E.K. (1979). The Aicardi syndrome in a 47, XXY male. Aust Paediatr J 15, 278-280.

109. Howe, K., and Wood, J.M. (2015). Using optical mapping data for the improvement of vertebrate genome assemblies. Gigascience 4, 10.

110. Hoyt, C.S., Billson, F., Ouvrier, R., and Wise, G. (1978). Ocular features of Aicardi's syndrome. Arch Ophthalmol 96, 291-295.

111. Huang, L., Jolly, L.A., Willis-Owen, S., Gardner, A., Kumar, R., Douglas, E., Shoubridge, C., Wieczorek, D., Tzschach, A., Cohen, M., et al. (2012). A noncoding, regulatory mutation implicates HCFC1 in nonsyndromic intellectual disability. Am J Hum Genet 91, 694-702.

112. Huang, N., Lee, I., Marcotte, E.M., and Hurles, M.E. (2010). Characterising and predicting haploinsufficiency in the human genome. PLoS Genet 6, e1001154.

113. Iffland, P.H., 2nd, and Crino, P.B. (2016). Sending Mixed Signals: The Expanding Role of Molecular Cascade Mutations in Malformations of Cortical Development and Epilepsy. Epilepsy Curr 16, 158-163.

114. Insel, T.R. (2014). Brain somatic mutations: the dark matter of psychiatric genetics? Mol Psychiatry 19, 156-158.

115. Iwamoto, T., Fukumoto, S., Yamada, A., Arakaki, M., and Nonaka, K. (2008). Presurgical treatment of cleft lip and palate in Aicardi syndrome: A case report. Pediatric Dental Journal 18, 204-209.

116. Jamuar, S.S., and Walsh, C.A. (2014). Somatic mutations in cerebral cortical malformations. N Engl J Med 371, 2038.

117. Jamuar, S.S., and Walsh, C.A. (2015). Genomic variants and variations in malformations of cortical development. Pediatr Clin North Am 62, 571-585.

118. Jeanne, M., Labelle-Dumais, C., Jorgensen, J., Kauffman, W.B., Mancini, G.M., Favor, J., Valant, V., Greenberg, S.M., Rosand, J., and Gould, D.B. (2012). COL4A2 mutations impair COL4A1 and COL4A2 secretion and cause hemorrhagic stroke. Am J Hum Genet 90, 91-101.

119. Jefferson, J.A., Lemmink, H.H., Hughes, A.E., Hill, C.M., Smeets, H.J., Doherty, C.C., and Maxwell, A.P. (1997). Autosomal dominant Alport syndrome linked to the type IV collage alpha 3 and alpha 4 genes (COL4A3 and COL4A4). Nephrol Dial Transplant 12, 1595-1599.

120. Jurka, J. (2000). Repbase update: a database and an electronic journal of repetitive elements. Trends Genet 16, 418-420.

121. Kasasbeh, A.S., Gurnett, C.A., and Smyth, M.D. (2014). Palliative epilepsy surgery in Aicardi syndrome: a case series and review of literature. Childs Nerv Syst 30, 497-503.

122. Keane, T.M., Wong, K., and Adams, D.J. (2013). RetroSeq: transposable element discovery from next-generation sequencing data. Bioinformatics 29, 389-390.

123. Khoshnoodi, J., Pedchenko, V., and Hudson, B.G. (2008). Mammalian collagen IV. Microsc Res Tech 71, 357-370.

124. Kim, M., and Jho, E.H. (2014). Cross-talk between Wnt/beta-catenin and Hippo signalling pathways: a brief review. BMB Rep 47, 540-545.

125. Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B., and Schilling, T.F. (1995). Stages of embryonic development of the zebrafish. Dev Dyn 203, 253-310.

126. King, A.M., Bowen, D.I., Goulding, P., and Doran, R.M. (1998). Aicardi syndrome. Br J Ophthalmol 82, 457.

127. Knebelmann, B., Deschenes, G., Gros, F., Hors, M.C., Grunfeld, J.P., Zhou, J., Tryggvason, K., Gubler, M.C., and Antignac, C. (1992). Substitution of arginine for glycine 325 in the collagen alpha 5 (IV) chain associated with X-linked Alport syndrome: characterization of the mutation by direct sequencing of PCR-amplified lymphoblast cDNA fragments. Am J Hum Genet 51, 135-142.

128. Knebelmann, B., Forestier, L., Drouot, L., Quinones, S., Chuet, C., Benessy, F., Saus, J., and Antignac, C. (1995). Splice-mediated insertion of an Alu sequence in the COL4A3 mRNA causing autosomal recessive Alport syndrome. Hum Mol Genet 4, 675-679.

129. Kobayashi, K., Watanabe, K., Yoshinaga, H., and Ohtsuka, Y. (2011). Bilaterally independent epileptic spasms in a case of Aicardi syndrome. Epileptic Disord 13, 326-330.

130. Kok, F.O., Shin, M., Ni, C.W., Gupta, A., Grosse, A.S., van Impel, A., Kirchmaier, B.C., Peterson-Maduro, J., Kourkoulis, G., Male, I., et al. (2015). Reverse genetic screening reveals

poor correlation between morpholino-induced and mutant phenotypes in zebrafish. Dev Cell 32, 97-108.

131. Komiya, Y., and Habas, R. (2008). Wnt signal transduction pathways. Organogenesis 4, 68-75.

132. Koscielny, G., Yaikhom, G., Iyer, V., Meehan, T.F., Morgan, H., Atienza-Herrero, J., Blake, A., Chen, C.K., Easty, R., Di Fenza, A., et al. (2014). The International Mouse Phenotyping Consortium Web Portal, a unified point of access for knockout mice and related phenotyping data. Nucleic Acids Res 42, D802-809.

133. Krause, A.C. (1946). Congenital encephalo-ophthalmic dysplasia. Arch Ophthal 36, 387-344.

134. Kroner, B.L., Preiss, L.R., Ardini, M.A., and Gaillard, W.D. (2008). New incidence, prevalence, and survival of Aicardi syndrome from 408 cases. J Child Neurol 23, 531-535.

135. Krumm, N., Sudmant, P.H., Ko, A., O'Roak, B.J., Malig, M., Coe, B.P., Project, N.E.S., Quinlan, A.R., Nickerson, D.A., and Eichler, E.E. (2012). Copy number variation detection and genotyping from exome sequence data. Genome Res 22, 1525-1532.

136. Kumar, R., Ha, T., Pham, D., Shaw, M., Mangelsdorf, M., Friend, K.L., Hobson, L., Turner, G., Boyle, J., Field, M., et al. (2016). A non-coding variant in the 5' UTR of DLG3 attenuates protein translation to cause non-syndromic intellectual disability. Eur J Hum Genet 24, 1612-1616.

137. Kuo, D.S., Labelle-Dumais, C., and Gould, D.B. (2012). COL4A1 and COL4A2 mutations and disease: insights into pathogenic mechanisms and potential therapeutic targets. Hum Mol Genet 21, R97-110.

138. Kuo, D.S., Labelle-Dumais, C., Mao, M., Jeanne, M., Kauffman, W.B., Allen, J., Favor, J., and Gould, D.B. (2014). Allelic heterogeneity contributes to variability in ocular dysgenesis,

myopathy and brain malformations caused by Col4a1 and Col4a2 mutations. Hum Mol Genet 23, 1709-1722.

139. Kurugundla, N., Amchentsev, A., Sinha, R., Romaine, S., and Lombardo, G. (2008). Obstructive Sleep Apnea in a Child with Aicardi Syndrome. The Internet Journal of Pulmonary Medicine 11, 1-3.

140. La Spada, A.R., and Taylor, J.P. (2010). Repeat expansion disease: progress and puzzles in disease pathogenesis. Nat Rev Genet 11, 247-258.

141. Lako, M., Lindsay, S., Bullen, P., Wilson, D.I., Robson, S.C., and Strachan, T. (1998). A novel mammalian wnt gene, WNT8B, shows brain-restricted expression in early development, with sharply delimited expression boundaries in the developing forebrain. Hum Mol Genet 7, 813-822.

142. Lamason, R.L., Mohideen, M.A., Mest, J.R., Wong, A.C., Norton, H.L., Aros, M.C., Jurynec, M.J., Mao, X., Humphreville, V.R., Humbert, J.E., et al. (2005). SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. Science 310, 1782-1786.

143. Landrum, M.J., Lee, J.M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Hoover, J., et al. (2016). ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res 44, D862-868.

144. Lau, L.W., Cua, R., Keough, M.B., Haylock-Jacobs, S., and Yong, V.W. (2013). Pathophysiology of the brain extracellular matrix: a new target for remyelination. Nat Rev Neurosci 14, 722-729.

145. Layer, R.M., Chiang, C., Quinlan, A.R., and Hall, I.M. (2014). LUMPY: a probabilistic framework for structural variant discovery. Genome Biol 15, R84.

146. Lee, S.W., Kim, K.S., Cho, S.M., and Lee, S.J. (2004). An atypical case of Aicardi syndrome with favorable outcome. Korean J Ophthalmol 18, 79-83.

147. Lelieveld, S.H., Reijnders, M.R., Pfundt, R., Yntema, H.G., Kamsteeg, E.J., de Vries, P., de Vries, B.B., Willemsen, M.H., Kleefstra, T., Lohner, K., et al. (2016). Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. Nat Neurosci 19, 1194-1196.

148. Lemay, P., Guyot, M.C., Tremblay, E., Dionne-Laporte, A., Spiegelman, D., Henrion, E., Diallo, O., De Marco, P., Merello, E., Massicotte, C., et al. (2015). Loss-of-function de novo mutations play an important role in severe human neural tube defects. J Med Genet 52, 493-497.

149. Lemmens, R., Maugeri, A., Niessen, H.W., Goris, A., Tousseyn, T., Demaerel, P., Corveleyn, A., Robberecht, W., van der Knaap, M.S., Thijs, V.N., et al. (2013). Novel COL4A1 mutations cause cerebral small vessel disease by haploinsufficiency. Hum Mol Genet 22, 391-397.

150. Lemmink, H.H., Nillesen, W.N., Mochizuki, T., Schroder, C.H., Brunner, H.G., van Oost, B.A., Monnens, L.A., and Smeets, H.J. (1996). Benign familial hematuria due to mutation of the type IV collagen alpha4 gene. J Clin Invest 98, 1114-1118.

151. Leng, T., and Moshfeghi, D.M. (2011). Retinopathy of prematurity in an infant with Aicardi's syndrome. Eye (Lond) 25, 257-258.

152. Leventer, R.J., Guerrini, R., and Dobyns, W.B. (2008). Malformations of cortical development and epilepsy. Dialogues Clin Neurosci 10, 47-62.

153. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754-1760.

154. Liimatainen, S.P., Jylhava, J., Raitanen, J., Peltola, J.T., and Hurme, M.A. (2013). The concentration of cell-free DNA in focal epilepsy. Epilepsy Res 105, 292-298.

155. Lim, J.S., Kim, W.I., Kang, H.C., Kim, S.H., Park, A.H., Park, E.K., Cho, Y.W., Kim, S., Kim, H.M., Kim, J.A., et al. (2015). Brain somatic mutations in MTOR cause focal cortical dysplasia type II leading to intractable epilepsy. Nat Med 21, 395-400.

156. Liu, T., Chen, H., Kim, H., Huen, M.S., Chen, J., and Huang, J. (2012). RAD18-BRCTx interaction is required for efficient repair of UV-induced DNA damage. DNA Repair (Amst) 11, 131-138.

157. Logan, C.Y., and Nusse, R. (2004). The Wnt signalling pathway in development and disease. Annu Rev Cell Dev Biol 20, 781-810.

158. Lund, C., Striano, P., Sorte, H.S., Parisi, P., Iacomino, M., Sheng, Y., Vigeland, M.D., Oye, A.M., Moller, R.S., Selmer, K.K., et al. (2016). Exome Sequencing Fails to Identify the Genetic Cause of Aicardi Syndrome. Mol Syndromol 7, 234-238.

159. Luo, R., Jeong, S.J., Jin, Z., Strokes, N., Li, S., and Piao, X. (2011). G protein-coupled receptor 56 and collagen III, a receptor-ligand pair, regulates cortical development and lamination. Proc Natl Acad Sci U S A 108, 12925-12930.

160. Mahmood, K., Jung, C.H., Philip, G., Georgeson, P., Chung, J., Pope, B.J., and Park, D.J. (2017). Variant effect prediction tools assessed using independent, functional assay-based datasets: implications for discovery and diagnostics. Hum Genomics 11, 10.

161. Majewski, J., and Rosenblatt, D.S. (2012). Exome and whole-genome sequencing for gene discovery: the future is now! Hum Mutat 33, 591-592.

162. Mather, C.A., Mooney, S.D., Salipante, S.J., Scroggins, S., Wu, D., Pritchard, C.C., and Shirts, B.H. (2016). CADD score has limited clinical validity for the identification of pathogenic variants in noncoding regions in a hereditary cancer panel. Genet Med 18, 1269-1275.

163. Matlary, A., Prescott, T., Tvedt, B., Lindberg, K., Server, A., Aicardi, J., and Stromme, P. (2004). Aicardi syndrome in a girl with mild developmental delay, absence of epilepsy and normal EEG. Clin Dysmorphol 13, 257-260.

164. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol 17, 122.

165. McMahon, R.G., Bell, R.A., Moore, G.R., and Ludwin, S.K. (1984). Aicardi's syndrome. A clinicopathologic study. Arch Ophthalmol 102, 250-253.

166. Mefford, H.C., Yendle, S.C., Hsu, C., Cook, J., Geraghty, E., McMahon, J.M., Eeg-Olofsson, O., Sadleir, L.G., Gill, D., Ben-Zeev, B., et al. (2011). Rare copy number variants are an important cause of epileptic encephalopathies. Ann Neurol 70, 974-985.

167. Menezes, A.V., Lewis, T.L., and Buncic, J.R. (1996). Role of ocular involvement in the prediction of visual development and clinical prognosis in Aicardi syndrome. Br J Ophthalmol 80, 805-811.

168. Menezes, A.V., MacGregor, D.L., and Buncic, J.R. (1994). Aicardi syndrome: natural history and possible predictors of severity. Pediatr Neurol 11, 313-318.

169. Meuwissen, M.E., Halley, D.J., Smit, L.S., Lequin, M.H., Cobben, J.M., de Coo, R., van Harssel, J., Sallevelt, S., Woldringh, G., van der Knaap, M.S., et al. (2015). The expanding phenotype of COL4A1 and COL4A2 mutations: clinical data on 13 newly identified families and a review of the literature. Genet Med 17, 843-853.

170. Meyer, E., Carss, K.J., Rankin, J., Nichols, J.M., Grozeva, D., Joseph, A.P., Mencacci, N.E., Papandreou, A., Ng, J., Barral, S., et al. (2017). Mutations in the histone methyltransferase gene KMT2B cause complex early-onset dystonia. Nat Genet 49, 223-237.

171. Miesfeld, J.B., Gestri, G., Clark, B.S., Flinn, M.A., Poole, R.J., Bader, J.R., Besharse, J.C., Wilson, S.W., and Link, B.A. (2015). Yap and Taz regulate retinal pigment epithelial cell fate. Development 142, 3021-3032.

172. Mochizuki, T., Lemmink, H.H., Mariyama, M., Antignac, C., Gubler, M.C., Pirson, Y., Verellen-Dumoulin, C., Chan, B., Schroder, C.H., Smeets, H.J., et al. (1994). Identification of mutations in the alpha 3(IV) and alpha 4(IV) collagen genes in autosomal recessive Alport syndrome. Nat Genet 8, 77-81.

173. Molina, J.A., Mateos, F., Merino, M., Epifanio, J.L., and Gorrono, M. (1989). Aicardi syndrome in two sisters. J Pediatr 115, 282-283.

174. Moller, R.S., Weckhuysen, S., Chipaux, M., Marsan, E., Taly, V., Bebin, E.M., Hiatt, S.M., Prokop, J.W., Bowling, K.M., Mei, D., et al. (2016). Germline and somatic mutations in the MTOR gene in focal cortical dysplasia and epilepsy. Neurol Genet 2, e118.

175. Monlong, J., Girard, S.L., Meloche, C., Cadieux-Dion, M., Andrade, D.M., Lafreniere, R.G., Gravel, M., Spiegelman, D., Dionne-Laporte, A., Boelman, C., et al. (2018). Global characterization of copy number variants in epilepsy patients from whole genome sequencing. PLoS Genet 14, e1007285.

176. Murray, L.S., Lu, Y., Taggart, A., Van Regemorter, N., Vilain, C., Abramowicz, M., Kadler, K.E., and Van Agtmael, T. (2014). Chemical chaperone treatment reduces intracellular accumulation of mutant collagen IV and ameliorates the cellular phenotype of a COL4A2 mutation that causes haemorrhagic stroke. Hum Mol Genet 23, 283-292.

177. Murtaza, M., Jolly, L.A., Gecz, J., and Wood, S.A. (2015). La FAM fatale: USP9X in development and disease. Cell Mol Life Sci 72, 2075-2089.

178. Nakamura, Y., Togawa, Y., Okuno, Y., Muramatsu, H., Nakabayashi, K., Kuroki, Y., Ieda, D., Hori, I., Negishi, Y., Togawa, T., et al. (2018). Biallelic mutations in SZT2 cause a discernible clinical entity with epilepsy, developmental delay, macrocephaly and a dysmorphic corpus callosum. Brain Dev 40, 134-139.

179. Nava, C., Dalle, C., Rastetter, A., Striano, P., de Kovel, C.G., Nabbout, R., Cances, C., Ville, D., Brilstra, E.H., Gobbi, G., et al. (2014). De novo mutations in HCN1 cause early infantile epileptic encephalopathy. Nat Genet 46, 640-645.

180. Neidich, J.A., Nussbaum, R.L., Packer, R.J., Emanuel, B.S., and Puck, J.M. (1990). Heterogeneity of clinical severity and molecular lesions in Aicardi syndrome. J Pediatr 116, 911-917.

181. Nielsen, K.B., Anvret, M., Flodmark, O., Furuskog, P., and Bohman-Valis, K. (1991). Aicardi syndrome: early neuroradiological manifestations and results of DNA studies in one patient. Am J Med Genet 38, 65-68.

182. O'Driscoll, M.C., Black, G.C., Clayton-Smith, J., Sherr, E.H., and Dobyns, W.B. (2010). Identification of genomic loci contributing to agenesis of the corpus callosum. Am J Med Genet A 152A, 2145-2159.

183. Ogueta, S.B., Schwartz, S.D., Yamashita, C.K., and Farber, D.B. (1999). Estrogen receptor in the human eye: influence of gender and age on gene expression. Invest Ophthalmol Vis Sci 40, 1906-1911.

184. Ospina, L.H., Nayak, H., and McCormick, A.Q. (2004). Progressive pigmentation of chorioretinal lesions in aicardi syndrome. Arch Ophthalmol 122, 790.

185. Paciorkowski, A.R., Thio, L.L., Rosenfeld, J.A., Gajecka, M., Gurnett, C.A., Kulkarni, S., Chung, W.K., Marsh, E.D., Gentile, M., Reggin, J.D., et al. (2011). Copy number variants and infantile spasms: evidence for abnormalities in ventral forebrain development and pathways of synaptic function. Eur J Hum Genet 19, 1238-1245.

186. Pai, C.P., Wang, C.Y., Kuo, Y.T., and Liang, J.S. (2013). Aicardi syndrome without CDKL5 gene mutation. Journal of Experimental and Clinical Medicine 5, 81-82.

187. Palmer, L., Zetterlund, B., Hard, A.L., Steneryd, K., and Kyllerman, M. (2006). Aicardi syndrome: presentation at onset in Swedish children born in 1975-2002. Neuropediatrics 37, 154-158.

188. Parikh, H., Mohiyuddin, M., Lam, H.Y., Iyer, H., Chen, D., Pratt, M., Bartha, G., Spies, N., Losert, W., Zook, J.M., et al. (2016). svclassify: a method to establish benchmark structural variant calls. BMC Genomics 17, 64.

189. Parkin, J.D., San Antonio, J.D., Pedchenko, V., Hudson, B., Jensen, S.T., and Savige, J. (2011). Mapping structural landmarks, ligand binding sites, and missense mutations to the

collagen IV heterotrimers predicts major functional domains, novel interactions, and variation in phenotypes in inherited diseases affecting basement membranes. Hum Mutat 32, 127-143.

190. Peng, M., Yin, N., and Li, M.O. (2017). SZT2 dictates GATOR control of mTORC1 signalling. Nature 543, 433-437.

191. Pinto, D., Pagnamenta, A.T., Klei, L., Anney, R., Merico, D., Regan, R., Conroy, J., Magalhaes, T.R., Correia, C., Abrahams, B.S., et al. (2010). Functional impact of global rare copy number variation in autism spectrum disorders. Nature 466, 368-372.

192. Piras, I.S., Mills, G., Llaci, L., Naymik, M., Ramsey, K., Belnap, N., Balak, C.D., Jepsen, W.M., Szelinger, S., Siniard, A.L., et al. (2017). Exploring genome-wide DNA methylation patterns in Aicardi syndrome. Epigenomics 9, 1373-1386.

193. Plaisier, E., Chen, Z., Gekeler, F., Benhassine, S., Dahan, K., Marro, B., Alamowitch, S., Paques, M., and Ronco, P. (2010). Novel COL4A1 mutations associated with HANAC syndrome: a role for the triple helical CB3[IV] domain. Am J Med Genet A 152A, 2550-2555.

194. Plaisier, E., Gribouval, O., Alamowitch, S., Mougenot, B., Prost, C., Verpont, M.C., Marro, B., Desmettre, T., Cohen, S.Y., Roullet, E., et al. (2007). COL4A1 mutations and hereditary angiopathy, nephropathy, aneurysms, and muscle cramps. N Engl J Med 357, 2687-2695.

195. Poduri, A., Evrony, G.D., Cai, X., Elhosary, P.C., Beroukhim, R., Lehtinen, M.K., Hills, L.B., Heinzen, E.L., Hill, A., Hill, R.S., et al. (2012). Somatic activation of AKT3 causes hemispheric developmental brain malformations. Neuron 74, 41-48.

196. Posey, J.E., Harel, T., Liu, P., Rosenfeld, J.A., James, R.A., Coban Akdemir, Z.H., Walkiewicz, M., Bi, W., Xiao, R., Ding, Y., et al. (2017). Resolution of Disease Phenotypes Resulting from Multilocus Genomic Variation. N Engl J Med 376, 21-31.

197. Prats Vinas, J.M., Martinez Gonzalez, M.J., Garcia Ribes, A., Martinez Gonzalez, S., and Martinez Fernandez, R. (2005). Callosal agenesis, chorioretinal lacunae, absence of infantile

spasms, and normal development: Aicardi syndrome without epilepsy? Dev Med Child Neurol 47, 419-420; discussion 364.

198. Radford, E.J., Isganaitis, E., Jimenez-Chillaron, J., Schroeder, J., Molla, M., Andrews, S., Didier, N., Charalambous, M., McEwen, K., Marazzi, G., et al. (2012). An unbiased assessment of the role of imprinted genes in an intergenerational model of developmental programming. PLoS Genet 8, e1002605.

199. Ramu, A., Noordam, M.J., Schwartz, R.S., Wuster, A., Hurles, M.E., Cartwright, R.A., and Conrad, D.F. (2013). DeNovoGear: de novo indel and point mutation discovery and phasing. Nat Methods 10, 985-987.

200. Raschle, M., Smeenk, G., Hansen, R.K., Temu, T., Oka, Y., Hein, M.Y., Nagaraj, N., Long, D.T., Walter, J.C., Hofmann, K., et al. (2015). DNA repair. Proteomics reveals dynamic assembly of repair complexes during bypass of DNA cross-links. Science 348, 1253671.

201. Rausch, T., Zichner, T., Schlattl, A., Stutz, A.M., Benes, V., and Korbel, J.O. (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics 28, i333-i339.

202. Renieri, A., Seri, M., Myers, J.C., Pihlajaniemi, T., Sessa, A., Rizzoni, G., and De Marchi, M. (1992). Alport syndrome caused by a 5' deletion within the COL4A5 gene. Hum Genet 89, 120-121.

203. Revankar, C.M., Cimino, D.F., Sklar, L.A., Arterburn, J.B., and Prossnitz, E.R. (2005). A transmembrane intracellular estrogen receptor mediates rapid cell signalling. Science 307, 1625-1630.

204. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). 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. Genet Med 17, 405-424.

205. Riggs, E.R., Jackson, L., Miller, D.T., and Van Vooren, S. (2012). Phenotypic information in genomic variant databases enhances clinical care and research: the International Standards for Cytogenomic Arrays Consortium experience. Hum Mutat 33, 787-796.

206. Rinn, J.L., and Snyder, M. (2005). Sexual dimorphism in mammalian gene expression. Trends Genet 21, 298-305.

207. Rodahl, E., Knappskog, P.M., Majewski, J., Johansson, S., Telstad, W., Krakenes, J., and Boman, H. (2013). Variants of anterior segment dysgenesis and cerebral involvement in a large family with a novel COL4A1 mutation. Am J Ophthalmol 155, 946-953.

208. Ropers, H.H., Zuffardi, O., Bianchi, E., and Tiepolo, L. (1982). Agenesis of corpus callosum, ocular, and skeletal anomalies (X-linked dominant Aicardi's syndrome) in a girl with balanced X/3 translocation. Hum Genet 61, 364-368.

209. Rosser, T.L., Acosta, M.T., and Packer, R.J. (2002). Aicardi syndrome: spectrum of disease and long-term prognosis in 77 females. Pediatr Neurol 27, 343-346.

210. Rost, S., Bach, E., Neuner, C., Nanda, I., Dysek, S., Bittner, R.E., Keller, A., Bartsch, O., Mlynski, R., Haaf, T., et al. (2014). Novel form of X-linked nonsyndromic hearing loss with cochlear malformation caused by a mutation in the type IV collagen gene COL4A6. Eur J Hum Genet 22, 208-215.

211. Saddichha, S., Manjunatha, N., and Akhtar, S. (2007). Typical Aicardi syndrome in a male. Acta Neuropsychiatrica 19.

212. Saitoh, T., Mine, T., and Katoh, M. (2002). Expression and regulation of WNT8A and WNT8B mRNAs in human tumor cell lines: up-regulation of WNT8B mRNA by beta-estradiol in MCF-7 cells, and down-regulation of WNT8A and WNT8B mRNAs by retinoic acid in NT2 cells. Int J Oncol 20, 999-1003.

213. Saunders, C.T., Wong, W.S., Swamy, S., Becq, J., Murray, L.J., and Cheetham, R.K. (2012). Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics 28, 1811-1817.

214. Schatz, M.C., Delcher, A.L., and Salzberg, S.L. (2010). Assembly of large genomes using second-generation sequencing. Genome Res 20, 1165-1173.

215. Schmidt, C., Fischer, G., Kadner, H., Genersch, E., Kuhn, K., and Poschl, E. (1993). Differential effects of DNA-binding proteins on bidirectional transcription from the common promoter region of human collagen type IV genes COL4A1 and COL4A2. Biochim Biophys Acta 1174, 1-10.

216. Schrauwen, I., Szelinger, S., Siniard, A.L., Corneveaux, J.J., Kurdoglu, A., Richholt, R., De Both, M., Malenica, I., Swaminathan, S., Rangasamy, S., et al. (2015). A De Novo Mutation in TEAD1 Causes Non-X-Linked Aicardi Syndrome. Invest Ophthalmol Vis Sci 56, 3896-3904.

217. Sepp, M., Pruunsild, P., and Timmusk, T. (2012). Pitt-Hopkins syndrome-associated mutations in TCF4 lead to variable impairment of the transcription factor function ranging from hypomorphic to dominant-negative effects. Hum Mol Genet 21, 2873-2888.

218. Shah, S., Ellard, S., Kneen, R., Lim, M., Osborne, N., Rankin, J., Stoodley, N., van der Knaap, M., Whitney, A., and Jardine, P. (2012). Childhood presentation of COL4A1 mutations. Dev Med Child Neurol 54, 569-574.

219. Sharan, S.K., and Bradley, A. (1997). Identification and characterization of a microsatellite marker within murine Brca2 gene. Mamm Genome 8, 79.

220. Shen, J.J., Wang, T.Y., and Yang, W. (2017). Regulatory and evolutionary signatures of sex-biased genes on both the X chromosome and the autosomes. Biol Sex Differ 8, 35.

221. Shetty, J., Fraser, J., Goudie, D., and Kirkpatrick, M. (2014). Aicardi syndrome in a 47 XXY male - a variable developmental phenotype? Eur J Paediatr Neurol 18, 529-531.

222. Shinwari, J.M., Khan, A., Awad, S., Shinwari, Z., Alaiya, A., Alanazi, M., Tahir, A., Poizat, C., and Al Tassan, N. (2015). Recessive mutations in COL25A1 are a cause of congenital cranial dysinnervation disorder. Am J Hum Genet 96, 147-152.

223. Sibon, I., Coupry, I., Menegon, P., Bouchet, J.P., Gorry, P., Burgelin, I., Calvas, P., Orignac, I., Dousset, V., Lacombe, D., et al. (2007). COL4A1 mutation in Axenfeld-Rieger anomaly with leukoencephalopathy and stroke. Ann Neurol 62, 177-184.

224. Sifrim, A., Hitz, M.P., Wilsdon, A., Breckpot, J., Turki, S.H., Thienpont, B., McRae, J., Fitzgerald, T.W., Singh, T., Swaminathan, G.J., et al. (2016). Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. Nat Genet 48, 1060-1065.

225. Smedley, D., Schubach, M., Jacobsen, J.O.B., Kohler, S., Zemojtel, T., Spielmann, M., Jager, M., Hochheiser, H., Washington, N.L., McMurry, J.A., et al. (2016). A Whole-Genome Analysis Framework for Effective Identification of Pathogenic Regulatory Variants in Mendelian Disease. Am J Hum Genet 99, 595-606.

226. Smeets, H.J., Melenhorst, J.J., Lemmink, H.H., Schroder, C.H., Nelen, M.R., Zhou, J., Hostikka, S.L., Tryggvason, K., Ropers, H.H., Jansweijer, M.C., et al. (1992). Different mutations in the COL4A5 collagen gene in two patients with different features of Alport syndrome. Kidney Int 42, 83-88.

227. Smit, L.M., Barth, P.G., Valk, J., and Njiokiktjien, C. (1984). Familial porencephalic white matter disease in two generations. Brain Dev 6, 54-58.

228. Smith, C.D., Ryan, S.J., Hoover, S.L., and Baumann, R.J. (1996). Magnetic Resonance Imaging of the Brain in Aicardi's Syndrome. Journal of Neuroimaging 6.

229. Smith, C.M., Finger, J.H., Hayamizu, T.F., McCright, I.J., Eppig, J.T., Kadin, J.A., Richardson, J.E., and Ringwald, M. (2007). The mouse Gene Expression Database (GXD): 2007 update. Nucleic Acids Res 35, D618-623.

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

231. Spennato, P., La Porta, A., Varone, A., Ruggiero, C., Buono, S., and Cinalli, G. (2013). Aicardi and Turner syndrome in a 45,X0/46,XX female. Clin Neurol Neurosurg 115, 820-822.

232. Su, C.H., Lin, I.H., Tzeng, T.Y., Hsieh, W.T., and Hsu, M.T. (2016). Regulation of IL-20 Expression by Estradiol through KMT2B-Mediated Epigenetic Modification. PLoS One 11, e0166090.

233. Sudmant, P.H., Rausch, T., Gardner, E.J., Handsaker, R.E., Abyzov, A., Huddleston, J., Zhang, Y., Ye, K., Jun, G., Fritz, M.H., et al. (2015). An integrated map of structural variation in 2,504 human genomes. Nature 526, 75-81.

234. Sun, W., Huang, L., Xu, Y., Xiao, X., Li, S., Jia, X., Gao, B., Wang, P., Guo, X., and Zhang, Q. (2015). Exome Sequencing on 298 Probands With Early-Onset High Myopia: Approximately One-Fourth Show Potential Pathogenic Mutations in RetNet Genes. Invest Ophthalmol Vis Sci 56, 8365-8372.

235. Sutton, V.R., Hopkins, B.J., Eble, T.N., Gambhir, N., Lewis, R.A., and Van den Veyver, I.B. (2005). Facial and physical features of Aicardi syndrome: infants to teenagers. Am J Med Genet A 138A, 254-258.

236. Sutton, V.R., and Van den Veyver, I.B. (1993). Aicardi Syndrome. In GeneReviews((R)), M.P. Adam, H.H. Ardinger, R.A. Pagon, S.E. Wallace, L.J.H. Bean, K. Stephens, andA. Amemiya, eds. (Seattle (WA).

237. Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 43, D447-452.

238. Taggard, D.A., and Menezes, A.H. (2000). Three choroid plexus papillomas in a patient with Aicardi syndrome. A case report. Pediatr Neurosurg 33, 219-223.

239. Takamiya, M., Xu, F., Suhonen, H., Gourain, V., Yang, L., Ho, N.Y., Helfen, L., Schrock, A., Etard, C., Grabher, C., et al. (2016). Melanosomes in pigmented epithelia maintain eye lens transparency during zebrafish embryonic development. Sci Rep 6, 25046.

240. Tao, J., Van Esch, H., Hagedorn-Greiwe, M., Hoffmann, K., Moser, B., Raynaud, M., Sperner, J., Fryns, J.P., Schwinger, E., Gecz, J., et al. (2004). Mutations in the X-linked cyclin-dependent kinase-like 5 (CDKL5/STK9) gene are associated with severe neurodevelopmental retardation. Am J Hum Genet 75, 1149-1154.

241. Tattini, L., D'Aurizio, R., and Magi, A. (2015). Detection of Genomic Structural Variants from Next-Generation Sequencing Data. Front Bioeng Biotechnol 3, 92.

242. Terracciano, A., Trivisano, M., Cusmai, R., De Palma, L., Fusco, L., Compagnucci, C., Bertini, E., Vigevano, F., and Specchio, N. (2016). PCDH19-related epilepsy in two mosaic male patients. Epilepsia 57, e51-55.

243. Thomas, J.W., Touchman, J.W., Blakesley, R.W., Bouffard, G.G., Beckstrom-Sternberg, S.M., Margulies, E.H., Blanchette, M., Siepel, A.C., Thomas, P.J., McDowell, J.C., et al. (2003). Comparative analyses of multi-species sequences from targeted genomic regions. Nature 424, 788-793.

244. Toutzaris, D., Lewerenz, J., Albrecht, P., Jensen, L.T., Letz, J., Geerts, A., Golz, S., and Methner, A. (2010). A novel giant peroxisomal superoxide dismutase motif-containing protein. Free Radic Biol Med 48, 811-820.

245. Treangen, T.J., and Salzberg, S.L. (2011). Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet 13, 36-46.

246. Tsuchida, N., Nakashima, M., Miyauchi, A., Yoshitomi, S., Kimizu, T., Ganesan, V., Teik, K.W., Ch'ng, G.S., Kato, M., Mizuguchi, T., et al. (2018). Novel biallelic SZT2 mutations in 3 cases of early-onset epileptic encephalopathy. Clin Genet 93, 266-274.

247. Turco, A.E., Rossetti, S., Biasi, M.O., Rizzoni, G., Massella, L., Saarinen, N.H., Renieri, A., Pignatti, P.F., and De Marchi, M. (1995). A novel missense mutation in exon 3 of the COL4A5 gene associated with late-onset Alport syndrome. Clin Genet 48, 261-263.

248. Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., Zwahlen, M., Kampf, C., Wester, K., Hober, S., et al. (2010). Towards a knowledge-based Human Protein Atlas. Nat Biotechnol 28, 1248-1250.

249. Usdin, K., and Grabczyk, E. (2000). DNA repeat expansions and human disease. Cell Mol Life Sci 57, 914-931.

250. Vahedi, K., and Alamowitch, S. (2011). Clinical spectrum of type IV collagen (COL4A1) mutations: a novel genetic multisystem disease. Curr Opin Neurol 24, 63-68.

251. van Amerongen, R., and Berns, A. (2006). Knockout mouse models to study Wnt signal transduction. Trends Genet 22, 678-689.

252. Van den Veyver, I.B. (2002). Microphthalmia with linear skin defects (MLS), Aicardi, and Goltz syndromes: are they related X-linked dominant male-lethal disorders? Cytogenet Genome Res 99, 289-296.

253. Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 43, 11 10 11-33.

254. van der Loop, F.T., Heidet, L., Timmer, E.D., van den Bosch, B.J., Leinonen, A., Antignac, C., Jefferson, J.A., Maxwell, A.P., Monnens, L.A., Schroder, C.H., et al. (2000). Autosomal dominant Alport syndrome caused by a COL4A3 splice site mutation. Kidney Int 58, 1870-1875.

255. Veeramah, K.R., Johnstone, L., Karafet, T.M., Wolf, D., Sprissler, R., Salogiannis, J., Barth-Maron, A., Greenberg, M.E., Stuhlmann, T., Weinert, S., et al. (2013). Exome sequencing reveals new causal mutations in children with epileptic encephalopathies. Epilepsia 54, 1270-1281.

256. Venkatesan, C., Angle, B., and Millichap, J.J. (2016). Early-life epileptic encephalopathy secondary to SZT2 pathogenic recessive variants. Epileptic Disord 18, 195-200.

257. Verbeek, E., Meuwissen, M.E., Verheijen, F.W., Govaert, P.P., Licht, D.J., Kuo, D.S., Poulton, C.J., Schot, R., Lequin, M.H., Dudink, J., et al. (2012). COL4A2 mutation associated with familial porencephaly and small-vessel disease. Eur J Hum Genet 20, 844-851.

258. Visel, A., Thaller, C., and Eichele, G. (2004). GenePaint.org: an atlas of gene expression patterns in the mouse embryo. Nucleic Acids Res 32, D552-556.

259. Volkner, M., Zschatzsch, M., Rostovskaya, M., Overall, R.W., Busskamp, V., Anastassiadis, K., and Karl, M.O. (2016). Retinal Organoids from Pluripotent Stem Cells Efficiently Recapitulate Retinogenesis. Stem Cell Reports 6, 525-538.

260. Wainger, B.J., DeGennaro, M., Santoro, B., Siegelbaum, S.A., and Tibbs, G.R. (2001). Molecular mechanism of cAMP modulation of HCN pacemaker channels. Nature 411, 805-810.

261. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38, e164.

262. Wang, X., Sutton, V.R., Eble, T.N., Lewis, R.A., Gunaratne, P., Patel, A., and Van den Veyver, I.B. (2009). A genome-wide screen for copy number alterations in Aicardi syndrome. Am J Med Genet A 149A, 2113-2121.

263. Ward, M., McEwan, C., Mills, J.D., and Janitz, M. (2015). Conservation and tissue-specific transcription patterns of long noncoding RNAs. J Hum Transcr 1, 2-9.

264. Webb, B.D., Brandt, T., Liu, L., Jalas, C., Liao, J., Fedick, A., Linderman, M.D., Diaz, G.A., Kornreich, R., Trachtman, H., et al. (2014). A founder mutation in COL4A3 causes

autosomal recessive Alport syndrome in the Ashkenazi Jewish population. Clin Genet 86, 155-160.

265. Weischenfeldt, J., Symmons, O., Spitz, F., and Korbel, J.O. (2013). Phenotypic impact of genomic structural variation: insights from and for human disease. Nat Rev Genet 14, 125-138.

266. Weng, Y.C., Sonni, A., Labelle-Dumais, C., de Leau, M., Kauffman, W.B., Jeanne, M., Biffi, A., Greenberg, S.M., Rosand, J., and Gould, D.B. (2012). COL4A1 mutations in patients with sporadic late-onset intracerebral hemorrhage. Ann Neurol 71, 470-477.

267. Westerfeld, M. (1995). The Zebrafish Book. A Guide for the Laboratory Use of Zebrafish (Danio rerio).(Eugene, OR: University of Oregon Press).

268. Wieacker, P., Zimmer, J., and Ropers, H.H. (1985). X inactivation patterns in two syndromes with probable X-linked dominant, male lethal inheritance. Clin Genet 28, 238-242.

269. Willis, J., and Rosman, N.P. (1980). The Aicardi syndrome versus congenital infection: diagnostic considerations. J Pediatr 96, 235-239.

270. Willsey, A.J., Sanders, S.J., Li, M., Dong, S., Tebbenkamp, A.T., Muhle, R.A., Reilly, S.K., Lin, L., Fertuzinhos, S., Miller, J.A., et al. (2013). Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell 155, 997-1007.

271. Wolfson, R.L., Chantranupong, L., Wyant, G.A., Gu, X., Orozco, J.M., Shen, K., Condon, K.J., Petri, S., Kedir, J., Scaria, S.M., et al. (2017). KICSTOR recruits GATOR1 to the lysosome and is necessary for nutrients to regulate mTORC1. Nature 543, 438-442.

272. Wong, B.K., Sutton, V.R., Lewis, R.A., and Van den Veyver, I.B. (2017). Independent variant analysis of TEAD1 and OCEL1 in 38 Aicardi syndrome patients. Mol Genet Genomic Med 5, 117-121.

273. Wright, C.F., Fitzgerald, T.W., Jones, W.D., Clayton, S., McRae, J.F., van Kogelenberg, M., King, D.A., Ambridge, K., Barrett, D.M., Bayzetinova, T., et al. (2015). Genetic diagnosis of

developmental disorders in the DDD study: a scalable analysis of genome-wide research data. Lancet 385, 1305-1314.

274. Xiao, Y., Camarillo, C., Ping, Y., Arana, T.B., Zhao, H., Thompson, P.M., Xu, C., Su, B.B., Fan, H., Ordonez, J., et al. (2014). The DNA methylome and transcriptome of different brain regions in schizophrenia and bipolar disorder. PLoS One 9, e95875.

275. Xu, H., Wang, F., Liu, Y., Yu, Y., Gelernter, J., and Zhang, H. (2014). Sex-biased methylome and transcriptome in human prefrontal cortex. Hum Mol Genet 23, 1260-1270.

276. Yacoub, M., Missaoui, N., Tabarli, B., Ghorbel, M., Tlili, K., Selmi, H., and Essoussi, A. (2003). [Aicardi syndrome with favorable outcome]. Arch Pediatr 10, 530-532.

277. Yamagata, T., Momoi, M., Miyamoto, S., Kobayashi, S., and Kamoshita, S. (1990). Multi-institutional survey of the Aicardi syndrome in Japan. Brain Dev 12, 760-765.

278. Yang, X., Schadt, E.E., Wang, S., Wang, H., Arnold, A.P., Ingram-Drake, L., Drake, T.A., and Lusis, A.J. (2006). Tissue-specific expression and regulation of sexually dimorphic genes in mice. Genome Res 16, 995-1004.

279. Yang, Y., Muzny, D.M., Reid, J.G., Bainbridge, M.N., Willis, A., Ward, P.A., Braxton, A., Beuten, J., Xia, F., Niu, Z., et al. (2013). Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N Engl J Med 369, 1502-1511.

280. Yang, Z., Xu, Z., Sun, Y.J., and Ma, L. (2014). Heterozygous somatic activating AKT1 mutation in a case of Proteus syndrome with mental retardation. J Dermatol 41, 188-189.

281. Yilmaz, S., Fontaine, H., Brochet, K., Gregoire, M.J., Devignes, M.D., Schaff, J.L., Philippe, C., Nemos, C., McGregor, J.L., and Jonveaux, P. (2007). Screening of subtle copy number changes in Aicardi syndrome patients with a high resolution X chromosome array-CGH. Eur J Med Genet 50, 386-391.

282. Yoneda, Y., Haginoya, K., Arai, H., Yamaoka, S., Tsurusaki, Y., Doi, H., Miyake, N., Yokochi, K., Osaka, H., Kato, M., et al. (2012). De novo and inherited mutations in COL4A2, encoding the type IV collagen alpha2 chain cause porencephaly. Am J Hum Genet 90, 86-90.

283. Zech, M., Boesch, S., Maier, E.M., Borggraefe, I., Vill, K., Laccone, F., Pilshofer, V., Ceballos-Baumann, A., Alhaddad, B., Berutti, R., et al. (2016). Haploinsufficiency of KMT2B, Encoding the Lysine-Specific Histone Methyltransferase 2B, Results in Early-Onset Generalized Dystonia. Am J Hum Genet 99, 1377-1387.

284. Zenteno, J.C., Crespi, J., Buentello-Volante, B., Buil, J.A., Bassaganyas, F., Vela-Segarra, J.I., Diaz-Cascajosa, J., and Marieges, M.T. (2015). Erratum to: Next generation sequencing uncovers a missense mutation in COL4A1 as the cause of familial retinal arteriolar tortuosity. Graefes Arch Clin Exp Ophthalmol 253, 1417.

285. Zhou, M., Wang, X., Li, J., Hao, D., Wang, Z., Shi, H., Han, L., Zhou, H., and Sun, J. (2015). Prioritizing candidate disease-related long non-coding RNAs by walking on the heterogeneous lncRNA and disease network. Mol Biosyst 11, 760-769.

286. Zhou, Z., Hu, T., Xu, Z., Lin, Z., Zhang, Z., Feng, T., Zhu, L., Rong, Y., Shen, H., Luk, J.M., et al. (2015). Targeting Hippo pathway by specific interruption of YAP-TEAD interaction using cyclic YAP-like peptides. FASEB J 29, 724-732.

287. Zook, J.M., Catoe, D., McDaniel, J., Vang, L., Spies, N., Sidow, A., Weng, Z., Liu, Y., Mason, C.E., Alexander, N., et al. (2016). Extensive sequencing of seven human genomes to characterize benchmark reference materials. Sci Data 3, 160025.

288. Zubairi, M.S., Carter, R.F., and Ronen, G.M. (2009). A male phenotype with Aicardi syndrome. J Child Neurol 24, 204-207.