The genetic determinants of cerebral palsy

A thesis submitted for the degree of Doctor of Philosophy (PhD) to the University of Adelaide

By

Gai McMichael

Supervisors: Professors Jozef Gecz and Eric Haan

The University of Adelaide, Robinson Institute

School of Medicine

Faculty of Health Science

May 2016

Statement of Declaration

This work contains no material which has been accepted for the award of any other degree or diploma 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.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.

Gai Lisette McMichael

January 2016

Table of contents

Statement of declaration i

Table of contents ii

Acknowledgements ix

Publications xi

HUGO Gene Nomenclature gene symbol and gene name xiii

Abbreviations xvi

URLs xix

Chapter 1 Introduction 1

1.1 Definition of cerebral palsy 2
1.2 Clinical classification of cerebral palsy 3
1.2.1 Gross motor function classification system 5
1.3 Neuroimaging 7
1.4 Incidence and economic cost of cerebral palsy 8
1.5 Known clinical risk factors for cerebral palsy 9
1.5.1 Preterm birth 9
1.5.2 Low birth weight 9
1.5.3 Multiple birth 10
1.5.4 Male gender 10
1.6 Other known clinical risk factors 11
1.6.1 Birth asphyxia 11
1.7 Other possible risk factors 12
1.8 Evidence for a genetic contribution to cerebral palsy causation 13
1.8.1 Sibling risks and twin studies 13
1.8.2 Effect of consanguinity 14
1.8.3 Associated congenital anomalies 15
1.8.4 Families providing evidence for Mendelian inheritance 16
1.9 Genetic disorders that can be mistaken for cerebral palsy 17
1.10 Approaches to identification of genes involved in cerebral palsy 19
1.10.1 Linkage analysis and homozygosity mapping 19
1.10.2 *GAD1* 20
1.10.3 *KANK1* 20
1.10.4 Chromosome 9p12-q12 21
1.11 Candidate-gene association studies 22
1.12 Copy number variants 23
1.13 Massively parallel sequencing 28
1.14 Whole-exome sequencing 28

1.14.1 Whole-exome capture methods 29
1.15 Whole-exome sequencing identifying single gene mutations 32
1.16 Whole-exome sequencing identifying single gene mutations in cerebral palsy 33
1.16.1 AP-4 complex 33
1.16.2 *ADD3* 34
1.16.3 *KCNC3* 34
1.16.4 *ITPR1* 35
1.16.5 *SPTBN2* 36
1.17 Whole-exome sequencing vs whole genome sequencing 39
1.18 Summary 41
1.19 Hypotheses and aims 42
1.19.1 Hypotheses 42
1.19.2 Aims 42
1.20 References 43

Chapter 2 Clinical cohort 67
2.1 Sporadic cerebral palsy cases (n = 183) 68
2.2 Brain imaging details 69
2.3 Cerebral palsy cohort clinical characteristics 69
2.4 Cerebral palsy subtype 69
2.5 Gestational age 72
2.6 Birth weight distribution 74
2.7 Comorbidities distribution 76
2.8 Gender distribution 77
2.9 Familial cerebral palsy cases 78
2.10 References 80

Chapter 3 Materials and methods 82
3.1 Recruitment of cerebral palsy families 83
3.2 Blood collection 84
3.3 Lymphocyte cell line development 84
3.4 DNA extraction 85
3.5 Whole-exome sequencing 86
3.5.1 Illumina library construction 86
3.5.2 Illumina exome capture 87
3.5.3 Illumina sequencing 88
3.5.4 Read mapping and variant calling 89
3.5.5 Data interrogation and analysis 90
3.6 Validation of candidate variants for cerebral palsy 91
3.6.1 Human Genome Sequencing Centre, Baylor College of Medicine 91
3.6.2 The University of Adelaide 91
3.7 Polymerase Chain Reaction (PCR) 95

3.7.1 PCR product visualization 95
3.8 Sanger sequencing reaction and clean up 95
3.9 DNA variant and gene prioritization 97
3.10 References 98

Chapter 4 Whole exome sequencing points to considerable genetic heterogeneity of cerebral palsy 100
Statement of authorship 101
4.1 Abstract 105
4.2 Introduction 106
4.3 Materials and methods 107
4.31 Study cohort 107
4.3.2 DNA extraction 107
4.3.3 Illumina library construction 108
4.3.4 Illumina exome capture 108
4.3.5 Illumina sequencing 109
4.3.6 Analysis 110
4.3.7 DNA variant and gene prioritization 111
4.4 Results 113
4.4.1 Case-parent trios (n=98) 113
4.4.1.1 *De novo* mutations 113
4.4.1.2 Inherited X-chromosome and recessive variants 114
4.4.2 Whole-exome sequencing 115
4.4.3 Pathway analysis 116
4.4.4 Clinical associations 116
4.5 Discussion 117
4.6 References 125

Chapter 5 A family diagnosed with cerebral palsy and intellectual disability with a X-linked mutation in ZC4H2 130
5.1 Introduction 131
5.2 Clinical reports 131
5.2.1 The family 131
5.2.2 The parents (I-1 and I-2) 132
5.2.3 The son with ‘cerebral palsy’ and intellectual disability (II-1) 132
5.2.4 The son with ‘cerebral palsy’ and intellectual disability (II-5) 135
5.2.5 The daughter with developmental delay (II-7) 136
5.2.6 Remaining siblings (II2, II3, II-4, II-6 and II-8) 136
5.3 Methods 138
5.3.1 DNA isolation 138
5.3.2 Whole-exome sequencing and analysis 138
5.3.3 Sanger sequencing 138

5.4 Results 140
5.5 Discussion 142
5.6 References 144

Chapter 6 NK2 homeobox 1 gene mutations in a family diagnosed with ataxic dyskinetic cerebral palsy 146
Statement of authorship 147
6.1 Abstract 149
6.2 Introduction 150
6.3 Clinical report 151
6.3.1 The family 151
6.3.2 The affected father 151
6.3.3 The affected son 152
6.3.4 The affected daughter 152
6.3.5 Other cases of dyskinetic or ataxic cerebral palsy 153
6.4 Methods 154
6.4.1 DNA isolation 154
6.4.2 Whole-exome sequencing and analysis 154
6.4.3 Sanger sequencing 155
6.5 Results 156
6.6 Discussion 158
6.7 References 160

Chapter 7 A family with autism spectrum disorder and spasticity with novel compound heterozygous mutation in the sodium channel gene SCN2A 162
7.1 Abstract 163
7.2 Introduction 164
7.3 Clinical report 165
7.3.1 The family 165
7.3.2 The parents 165
7.3.3 The 13-year-old son with ASD only 166
7.3.4 The 9-year-old daughter with ASD and spasticity 166
7.3.5 The 7-year-old son with ASD and spasticity 167
7.4 Methods 169
7.4.1 Whole-exome sequencing and analysis 169
7.4.2 Sanger sequencing 170
7.5 Results 171
7.6 Discussion 172
7.7 References 179

Chapter 8 Discussion and future directions 184
8.1 Discussion and future prospects 185
8.2 Sporadic cases 186

8.3 Familial cases 190
8.3.1 Family 1 with *ZC4H2* mutation 190
8.3.2 Family 2 with *NKX2-1* mutation 191
8.3.3 Family 3 with *SCN2A* mutation 192
8.4 Limitations of this study 194
8.5 Future prospects 195
8.5.1 Whole-exome and whole-genome sequencing 196
8.5.2 Custom gene-panels 197
8.5.3 Central data repository 198
8.5.4 Genome-wide expression profiling (RNA sequencing) 198
8.5.5 Epigenetics and cerebral palsy 200
8.6 Conclusion 204
8.7 References 207

Appendix Supplementary data 217
9.1 References 260

List of Tables

Table 1.1 Summary of potentially pathogenic CNVs encompassing genes previously associated with cerebral palsy 27
Table 1.2 Summary of the different solution-based commercial exome capture kits 31
Table 1.3 Summary of novel single gene mutations, in known (OMIM) genes, in cerebral palsy cases 38
Table 2.1 Type of cerebral palsy in 183 sporadic cases 70
Table 2.2 Gestational age distribution of 183 sporadic cases with cerebral palsy compared with South Australian (SA) pregnancy outcome data 73
Table 2.3 Birth weight distribution in 183 sporadic cases with cerebral palsy compared with South Australian (SA) pregnancy outcome data 75
Table 2.4 Distribution of co-morbidities in 183 sporadic cases with cerebral palsy 76
Table 2.5 Gender distributions in 183 sporadic cases with cerebral palsy 77
Table 2.6 Summary of the clinical characteristics for three families With more than one individual with a diagnosis of cerebral Palsy 79
Table 3.1 Primer sequences for validating candidate variants from whole blood (The University of Adelaide) 92
Table 4.1 Clinical characteristics of 183 individuals with cerebral palsy 122
Table 4.2 Novel variants in known (OMIM) genes in CP cases identified in 98 trios: CADD score > 10 and RVIS % < 50 123

Table 4.3	Novel candidate genes in CP cases identified in 98 trios; CADD score > 20 and RVIS % < 50	124
Table 5.1	Multispecies alignment around p.Arg213 of *ZC4H2*	141
Table 7.1	Clinical characteristics, *SCN2A* variants and segregation pattern and pathogenicity predictions and scores	178
Table 7.2	Fragments of sequence chromatograms from 5'-3'	178
Table 9.1	Supplementary Table A	224
Table 9.2	Supplementary Table B	242
Table 9.3	Supplementary Table C	243
Table 9.4	Supplementary Table D	247
Table 9.5	Supplementary Table E	248
Table 9.6	Supplementary Table F	249
Table 9.7	Supplementary Table G	250
Table 9.8	Supplementary Table H	253
Table 9.9	Supplementary Table I	254
Table 9.10	Supplementary Table J	255
Table 9.11	Supplementary Table K	258
Table 9.12	Supplementary Table L	259

List of Figures

Figure 1.1	Different types of cerebral palsy and the regions of the body affected	5
Figure 1.2	Prevalence of cerebral palsy rates per 1000 live births compared with caesarean rates over the past 50 years	12
Figure 1.3	The prevalence and type of co-morbidities seen in individuals diagnosed with cerebral palsy	16
Figure 1.4	Allele frequency correlated with effect size	23
Figure 2.1	Type of cerebral palsy in 183 sporadic cases	71
Figure 2.2	Gestational age distribution of 183 sporadic cases with cerebral palsy compared with South Australian (SA) pregnancy outcome data	73
Figure 2.3	Birth weight distribution in 183 sporadic cases with cerebral palsy compared with South Australian (SA) pregnancy outcome data	75
Figure 2.4	Distribution of co-morbidities in 183 sporadic cases with cerebral palsy	76
Figure 2.5	Gender distributions in 183 sporadic cases with cerebral palsy	77
Figure 2.6-2.8	Overview of the pedigree details for three families with a diagnosis of cerebral palsy	78
Figure 5.1	Two generation pedigree with two brothers (II-1 and II-5) diagnosed with 'cerebral palsy' and intellectual disability and a sister (II-7) significantly behind in language, cognitive and personal/social skills	132

Figure 5.2	Representative sequence chromatograms of *ZC4H2* wildtype, mutations and heterozygous variants identified in nine out of the ten family members	141
Figure 6.1	Three generation pedigree with de novo mutation in II-1 and subsequent autosomal dominant transmission in III-2 and III-3	153
Figure 6.2(A)	Fragments of sequence chromatograms from an affected individual (heterozygous mutation) and an unaffected individual (normal homozygous) from 5'–3' and corresponding amino acid sequences	157
Figure 6.2(B)	Comparison of amino acid sequences of wildtype and mutant NKX2-1 proteins	157
Figure 7.1	Two-generation pedigree	177
Figure 7.2	Fragments of sequence chromatograms from 5'-3'	177
Figure 9.1	Supplementary Figure A	218
Figure 9.2	Supplementary Figure B	219
Figure 9.3	Supplementary Figure C	219
Figure 9.4	Supplementary Figure D	220
Figure 9.5	Supplementary Figure E	220
Figure 9.6	Supplementary Figure F	221
Figure 9.7	Supplementary Figure G	221
Figure 9.8	Supplementary Figure H	222
Figure 9.9	Supplementary Figure I	222
Figure 9.10	Supplementary Figure J	223

Acknowledgments

I extend my thanks to my supervisors, Professor Jozef Gecz and Professor Eric Haan for their support, wisdom and guidance over the duration of this study. Their mentoring and encouragement to me personally and in my career thus far has provided me with many valuable and exciting opportunities for which I am extremely thankful. It has been an absolute pleasure and privilege to work alongside Professor Jozef Gecz and Professor Eric Haan during both my PhD and my career in cerebral palsy research.

I gratefully extend my thanks to Professor Alastair MacLennan for his support for my career and studies in cerebral palsy research, in particular for supervising me during the first two years of my PhD. I very much appreciate the support from the Cerebral Palsy Research Group, University of Adelaide, Robinson Institute. I gratefully extend my sincere thanks to Corinne Reynolds for her support and friendship throughout my PhD. I extend sincere thanks to Professor Richard Gibbs for providing the opportunity to collaborate and work at Baylor College of Medicine, The Human Genome Sequencing Center (HGSC), Houston, Texas, USA.

I am very appreciative and am thankful for the support from the Department of Paediatric Rehabilitation Medicine at Women's and Children's Hospital, Adelaide in particular Dr Ray Russo, Dr James Rice and Dr Andrew Tidemann for allowing us to collect blood samples whilst our participants were under general anaesthetic. I also wish to thank Dr Jane Valentine and Peta Watts for facilitating blood collections at Princess Margaret Hospital, Western Australia.

Thank you to the staff at the University of Adelaide Department of Obstetrics and Gynaecology, both at the Women's and Children's Hospital and the Medical School for supporting me through this candidature. Thank you to Dr Kathie Friend and staff at the Department of Diagnostic Molecular Genetics, SA Pathology, Adelaide and the Australian Genome Research Facility (Adelaide node) for support with DNA extractions.

I sincerely thank my colleagues at the Adelaide Neurogenetics Research Group, The University of Adelaide, Robinson Institute for their support, advice, mentoring and above all their valued friendship. In particular thank you to Dr Raman Sharma, Dr Duyen Pham, Stanley Tan, Dr Mark Corbett and Alison Gardner it was a pleasure to work with you.

This work has been funded in part by the Cerebral Palsy Foundation, Tennix, University of Adelaide, Women's and Children's Research Foundation and National Health and Medical Research Council.

Publications

1. **McMichael, G.L.**, Bainbridge, M., Haan, E., Corbett, M. *et al.* Whole-exome sequencing points to considerable genetic heterogeneity of cerebral palsy. *Molecular Psychiatry*, 1-7 (2015).
2. **McMichael, G.L.**, Hirata, H., Nanda, I., van Riesen, A. *et al.* ZC4H2 mutations are associated with arthrogryposis multiplex congenita and intellectual disability through impairment of central and peripheral synaptic plasticity. *Am J Hum Genet* **92**, 681-95 (2013).
3. **McMichael, G.L.**, Haan, E., Gardner, A., Yap, T.Y. *et al.* NKX2-1 mutation in a family diagnosed with ataxic dyskinetic cerebral palsy. *Eur J Med Genet* **56**, 506-9 (2013).
4. **McMichael, G.L.**, Girirajan, S., Moreno-De-Luca, A., Gecz, J. *et al.* Rare copy number variation in cerebral palsy. *Eur J Hum Genet* (2013).
5. O'Callaghan, M.E., Maclennan, A.H., Gibson, C.S., **McMichael G.L.** *et al.* Genetic and clinical contributions to cerebral palsy: A multi-variable analysis. *J Paediatr Child Health* (2013).
6. O'Callaghan, M.E., MacLennan, A.H., **McMichael, G.L.**, Haan, E.A. & Dekker, G.A. Single-nucleotide polymorphism associations with preterm delivery: A case-control replication study and meta-analysis. *Pediatr Res***74**, 433-8 (2013).
8. O'Callaghan, M.E., MacLennan, A.H., Gibson, C.S., **McMichael G.L.** *et al.* Fetal and maternal candidate single nucleotide polymorphism associations with cerebral palsy: A case-control study. *Pediatrics* **129**, e414-23 (2012).
9. **McMichael, G.L.**, Maclennan, A., Gibson, C., Alvino, E. *et al.* Cytomegalovirus and Epstein-Barr virus may be associated with some cases of cerebral palsy. *J Matern Fetal Neonatal Med* (2012).

10. O'Callaghan, M.E., MacLennan, A., Gibson, C., **McMichael, G.L.** *et al.* Epidemiolgic associations with cerebral palsy. *Obstet Gynecol* **118**, 576-582 (2011).

11. **McMichael, G.L.**, Highet, A.R., Gibson, C.S., Goldwater, P.N. *et al.* Comparison of DNA extraction methods from small samples of newborn screening cards suitable for retrospective perinatal viral research. *J Biomol Tech* **22**, 5-9 (2011).

12. O'Callaghan, M.E., MacLennan, A.H., Gibson, C.S., **McMichael, G.L.** *et al.* The Australian cerebral palsy research study-protocol for a national collaborative study investigating genomic and clinical associations with cerebral palsy. *J Paediatr Child Health* **47**, 99-110 (2011).

13. **McMichael, G.L.**, Gibson, C.S., O'Callaghan, M.E., Goldwater, P.N. *et al.* DNA from buccal swabs suitable for high-throughput snp multiplex analysis. *J Biomol Tech* **20**, 232-5 (2009).

14. Djukic, M., Gibson, C.S., Maclennan, A.H., **McMichael, G.L.** *et al.* Genetic susceptibility to viral exposure may increase the risk of cerebral palsy. *Aust N Z J Obstet Gynaecol* **49**, 247-53 (2009).

15. **McMichael, G.L.**, Gibson, C.S., Goldwater, P.N., Haan, E.A. *et al.* Association between apolipoprotein E genotype and cerebral palsy is not confirmed in a caucasian population. *Hum Genet* **124**, 411-6 (2008).

HUGO Gene Nomenclature gene symbol and gene name

ADD3 – Adducin 3 (Gamma)

AGAP1 – ArfGAP with GTPase Domain, Ankyrin Repeat and PH Domain 1

AHI1 – Abelson Helper Integration Site 1

ANKRD15 – KN Motif And Ankyrin Repeat Domains 1 (KANK1)

AP-4 – AP-4 Complex

AP4B1 – Adaptor-related Protein Complex 4, β1 subunit

AP4E1 – Adaptor-related Protein Complex 4, ε1 subunit

AP4M1 – Adaptor-related Protein Complex 4, μ1 subunit

AP4S1 – Adaptor-related Protein Complex 4, σ1 subunit

CDK17 – Cyclin-Dependent Kinase 17

CD99L2 – CD99 Molecule-Like 2

CEP290 – Centrosomal Protein 290KDa

COPS3 – COP9 Signalosome Subunit 3

CTNND2 – Catenin (Cadherin-Associated Protein), Delta 2

CUL4B – Cullin 4B

ENPP4 – Ectonucleotide Pyrophosphatase/Phosphodiesterase 4 (Putative)

FLNB – Filamin B (Beta)

GAD1 – Glutamate Decarboxylase 1

GCH1 – GTP Cyclohydrolase 1

HSPA4 – Heat Shock 70kDa Protein 4

HSPG2 – Heparan Sulfate Proteoglycan 2

ITPR1 – Inositol 1,4,5-Trisphosphate Receptor, Type 1

JHDM1D – Lysine (K)-Specific Demethylase 7A

KANK1 – KN Motif And Ankyrin Repeat Domains 1 (ANKRD15)

HUGO Gene Nomenclature gene symbol and gene name (continued)

KDM5C – Lysine (K)-Specific Demethylase 5C

KCNC3 – Voltage-Gated Potassium Channel Subunit Kv3.3

L1CAM – L1 Cell Adhesion Molecule

LTN1 – Listerin E3 Ubiquitin Protein Ligase 1

MAOA – Monoamine Oxidase A

MAOB – Monoamine Oxidase B

MAST1 – Microtubule Associated Serine/Threonine Kinase 1

MCPH1 – Microcephalin 1

MED17 – Mediator Complex Subunit 17

MEF2C – Myocyte Enhancer Factor 2C

MIIP – Migration and Invasion Inhibitory Protein

MUM1L1 – Melanoma Associated Antigen (Mutated) 1-Like 1

MYH14 – Myosin, Heavy Chain 14, Non-Muscle

NAA35 – N(Alpha)-Acetyltransferase 35, NatC Auxiliary Subunit

NEMF – Nuclear Export Mediator Factor

NPHP1 – Nephronophthisis 1 (Juvenile)

NKX2-1 – NK2 Homeobox 1

PAK3 – P21 Protein (Cdc42/Rac)-Activated Kinase 3

PACRG – PARK2 Co-Regulated

PARK2 – Parkin RBR E3 Ubiquitin Protein Ligase

PAX5 – Paired Box 5

PLAC4 – Placenta-Specific 4

PLP1 – Proteolipid Protein 1

PCDH11X – Protocadherin 11 X-Linked

HUGO Gene Nomenclature gene symbol and gene name (continued)

RFX2 – Factor X, 2 (Influences HLA Class II Expression)

SCN2A – Sodium Channel, Voltage Gated, Type II Alpha Subunit

SCN8A – Sodium Channel, Voltage Gated, Type VIII Alpha Subunit

SLC11A2 – Solute Carrier Family 11 (Proton-Coupled Divalent Metal Ion Transporter), Member 2

SPAST – Spastin

SPR – Sepiapterin Reductase (7,8-Dihydrobiopterin:NADP+ Oxidoreductase)

SPTBN2 – Spectrin, beta, non-erythrocytic 2

SSPO – SCO-Spondin

SYNGAP1 – Synaptic Ras GTPase Activating Protein 1

TBC1D24 – TBC1 Domain Family, Member 24

TENM1 – Teneurin Transmembrane Protein 1

TGM6 – Transglutaminase 6

TUBA1A – Tubulin, Alpha 1a

UBE3A – Ubiquitin Protein Ligase E3A

UBQLN3 – Ubiquilin 3

WDR45 – WD Repeat Domain 45

WIPI2 – WD Repeat Domain, Phosphoinositide Interacting 2

ZC4H2 – Zinc Finger, C4H2 Domain Containing

ZNF674 – Zinc Finger Protein 674

Abbreviations

ACD – Acid citrate dextrose

AMC – Arthrogryposis multiplex congenita

ASD – Autism spectrum disorder

ATLAS – SNP

BCM – Baylor College of Medicine

BHC – Benign Hereditary Chorea

BWA – Burrows-Wheeler Aligner

CADD – Combined Annotation-Dependent Depletion

CGH – Comparative Genomic Hybridisation

ChIP – Chromatin immunoprecipitation

CNVs – Copy Number Variants

CNS – Central nervous system

CP – Cerebral palsy

DNA – Deoxyribonucleic acid

EDTA – Ethylenediaminetetraacetic acid

EVS – Exome Variant Server

ExAC – Exome Aggregation Consortium

GA – General anaesthetic

GABA – Gamma-aminobutyric acid

GATK – Genome Analysis Toolkit

GERP – Genomic Evolutionary Rate Profiling

GMFCS – Gross Motor Function Classification System

GRA – Genetic Repositories Australia

HGMD – Human Genome Mutation Database

Abbreviations (continued)

HGSC – Human Genome Sequencing Center

HSP – Hereditary Spastic Paraplegia

ID – Intellectual disability

IQ – Intelligence quotient

IUGR – Intrauterine growth restriction

IVF – *In vitro* fertilization

IVH – Intraventricular haemorrhage

LCLs – Lymphoblastoid cell lines

LM-PCR – Ligation mediated-polymerase chain reaction

LOF – Loss of function

MAF – Minor allele frequency

MIM – Mendelian Inheritance in Man

MPS – Massively parallel sequencing

mRNA – Messenger ribonucleic acid

miRNA – Micro ribonucleic acid

MRI – Magnetic resonance imaging

ncRNA –Non-coding ribonucleic acid

NGS – Next generation sequencing

NHLBI – National Heart, Lung, and Blood Institute

OMIM – Online Mendelian Inheritance in Man

OR – Odds ratio

PCR – Polymerase chain reaction

PE – Paired-end

PMD – Pelizaeus-Merzbacher disease

Abbreviations (continued)

POSU – Pregnancy Outcome in South Australia

PVL – Periventricular leukomalacia

RNA – Ribonucleic acid

rRNA – Ribosomal ribonucleic acid

tRNA – Transfer ribonucleic acid

RVIS – Residual Variation Intolerance Score

SHRs – Standardized hospitalisation ratios

SIFT – Scale-invariant feature transform

SNPs – Single nucleotide polymorphisms

SPG4 – Spastic paraplegia 4

UCSC – University of California, Santa Cruz Genome Browser

UTRs – Untranslated regions

WES – Whole-exome sequencing

WGS – Whole-genome sequencing

XLID – X-linked intellectual disability

URLs

Allen Human Brain Atlas – www.brainspan.org/

Australian Bureau of Statistics – www.abs.gov.au

BCM-HGSC protocol – https://hgsc.bcm.edu/ sites/default/files/documents/ Illumina_

Barcoded_PairedEnd_Capture_Library_Preparation.pdf Barcoded Paired-End Capture

Library Preparation.pdf

Cerebral Palsy Research Report – www.cerebralpalsy.org.au/wp-content /uploads

/2013/04/ ACPR/

dbSNP – http://www.ncbi.nlm.nih.gov/projects/SNP/

Ensembl – www.ensembl.org/

EVS – http://evs.gs.washington.edu/ EVS/

ExAC database – exac.broadinstitute.org/

OMIM – www.omim.org/

Partek – http://www.partek.com/

RefSeq – www.ncbi.nlm.nih.gov/refseq/

1000 Genomes – http://browser.1000genomes.org/index.html

UCSC – https://genome.ucsc.edu/

Chapter 1
Introduction

1.1 Definition of cerebral palsy

Cerebral palsy is a non-progressive neurodevelopmental disorder, which presents in early childhood and continues throughout life. The hallmark of cerebral palsy is that it is a disorder of the motor system i.e. a disorder of movement and posture.[1] The extensive clinical heterogeneity of the disorder has led to long time debate over its definition and classification.[2] Little (1843)[3] first described cerebral palsy as 'cerebral paresis', his work focusing on joint contractures and deformities due to spasticity and paralysis; he noted that comorbidities such as epilepsy and behavioural disturbances were infrequent. He suggested that injury to the brain was likely due to a difficult or traumatic birth resulting in cerebral paresis.[3] Freud (1893)[4] was of a different opinion, observing a higher rate of accompanying disturbances in cerebral palsy cases, and proposing that cerebral palsy was the result of an insult to the developing fetal brain. Based on possible causes, Freud assigned three cerebral palsy groups: 1) maternal and idiopathic congenital; 2) perinatal; and 3) postnatal. He also classified all bilateral disorders of central origin as "diplegia".[4] However, despite the opinion of Freud, Little's beliefs left a legacy lasting well over a century, putting undue emphasis on birth asphyxia as the primary cause of cerebral palsy.[5] Several decades later the definitions of cerebral palsy by Bax (1964)[6] *"a disorder of movement and posture due to a defect or lesion of the immature brain",* and Mutch (1991)[7] *"an umbrella term covering a group of non-progressive, but often changing, motor impairment syndromes secondary to lesions or anomalies of the brain arising in the early stages of development"* were the most frequently cited throughout the literature. Both of these classifications failed to encompass the complex heterogeneity of cerebral palsy. A revised definition and reclassification released in 2006 from an International Workshop is as follows: "*Cerebral palsy describes a group of permanent disorders of the development of*

movement and posture, causing activity limitation that are attributed to non-progressive disturbances that occurred in the developing fetal or infant brain. The motor disorders of cerebral palsy are often accompanied by disturbances of sensation, perception, cognition, communication, and behaviour; by epilepsy, and by secondary musculoskeletal problems".[8] This definition more accurately encapsulates the clinical heterogeneity of cerebral palsy.

Some confusion still remains around the accurate diagnosis of cerebral palsy. In summary, there must be a motor impairment which is cerebral in origin, present early in life and non-progressive, but this does not mean non-changing.[9] Clearly its manifestations change as children become older with the wear and tear on joints and other parts of the body with age.[9] Badawi *et al.* (1998)[9] published a report which included a comprehensive list of inclusion and exclusion criteria for cerebral palsy. The aim was to evoke further discussion and create more consistency amongst cerebral palsy registers. This would enable more reliability when measuring trends and prevalence between different populations and assist with collaborative research.[9] Adding to the complexity of a diagnosis of cerebral palsy is the large number of neurogenetic syndromes with clinical features that overlap with cerebral palsy. These include autosomal dominant, autosomal recessive and X-linked disorders. This is an important consideration as several neurogenetic syndromes can be treated or controlled.[10,11]

1.2 Clinical classification of cerebral palsy

Classification of cerebral palsy has largely relied on clinical judgement. For cases in which diagnostic pathways are heterogeneous, diagnosis may depend on which specialist the patient is first referred to, i.e. determining if the patient is diagnosed with

cerebral palsy with intellectual disability or intellectual disability with a movement disorder. Over the years systems have been put in place and revised to assist with more uniform clinical diagnosis. Emphasis is placed on the motor disorder, which affects movement and posture to varying degrees of severity, from mild problems with muscle coordination to severe spasticity of all four limbs, and is often the first reason children present for medical diagnosis.[8] The motor dysfunction of cerebral palsy is divided into four main categories resulting from damage to different parts of the brain: spasticity, resulting from damage to motor "pathways" originating from the motor cortex (i.e. the pyramidal pathways or corticospinal tracts); dyskinesia, resulting from damage to the basal ganglia; ataxia, resulting from damage to the cerebellum; or mixed, in which multiple brain regions are damaged. Spasticity, the most common type, accounts for up to 80% of cases. It is subdivided into hemiplegia, with spasticity affecting one arm and leg on one side of the body; diplegia, with spasticity primarily affecting the lower body; triplegia affecting three limbs; and quadriplegia in which there is spasticity of all limbs (Figure 1.1). Spasticity is characterised by increased muscle tone, i.e. muscles continually contract, making them resistant to passive lengthening in their most relaxed state and resulting in rigid limbs with joints that are resistant to movement. Reflexes can be exaggerated and movements tend to be jerky and awkward. The arms and legs are typically affected, but the tongue, mouth and pharynx can be affected as well, impairing speech, eating, breathing and swallowing. Dyskinesia is found in approximately 6 - 20% of cases (Figure 1.1). Two forms exist: athetosis (~5%), which involves slow, writhing movements that are often repetitive, sinuous and rhythmic, and dystonia (~15%), which involves involuntary movements accompanied by an abnormal, sustained posture, especially in the arms, legs and hands. Ataxia is found in approximately 5 - 10% of cases, affecting coordinated movements, balance and

posture. Walking gait is broad-based and sometimes irregular. Control of eye movements and depth perception can be impaired. Often, fine motor skills requiring coordination of the eyes and hands, such as writing, are difficult.[12,13] In many cases mixed forms of motor dysfunction are diagnosed, such as a combination of spasticity and dystonia.[13]

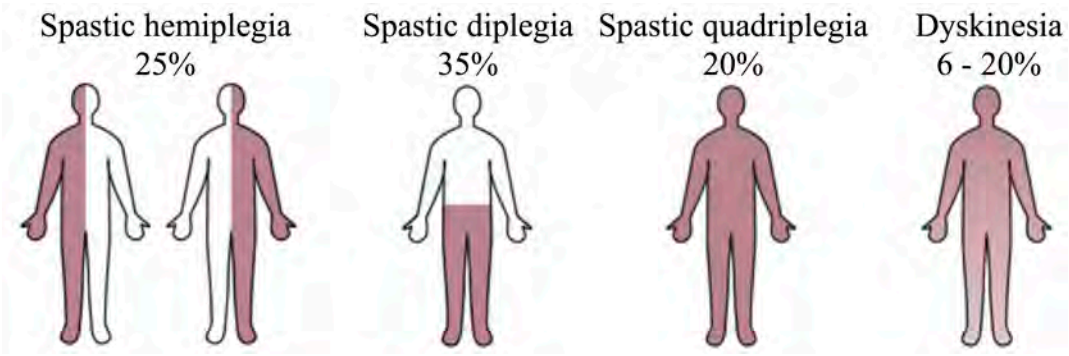


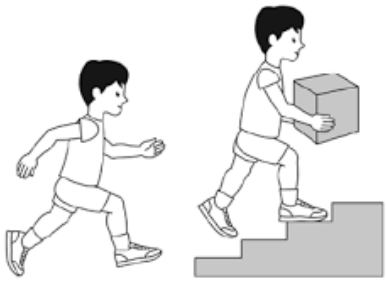
Figure 1.1 Different types of cerebral palsy and the regions of the body affected.

1.2.1 Gross motor function classification system

Severity of cerebral palsy varies greatly, from mild to severe. Motor dysfunction is assessed by the gross motor function classification system (GMFCS), which is a five level clinical classification system for grading gross motor skills, with a focus on sitting (truncal control), walking and wheeled mobility, taking into account the need for assistive devices (i.e. crutches, canes, walkers and wheelchairs).[14]

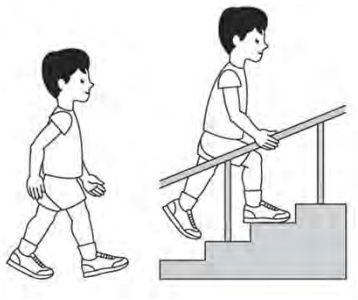
GMFCS Level 1

The individual is able to walk and climb stairs independently, perform gross motor skills such as jumping and running; however co-ordination, balance and speed are affected.[14]



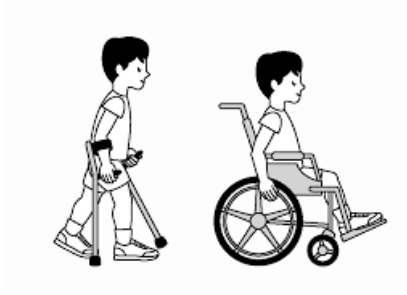
GMFCS Level II

Individuals are able to walk and climb stairs holding on to a rail for support but are limited when walking on undulating surfaces and walking amongst large numbers of people or in confined spaces.[14]



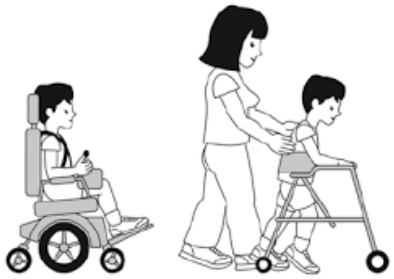
GMFCS Level III

Individuals are able to walk on even surfaces with the assistance of a mobility device, may be able to climb stairs with assistance and manually propel a wheelchair.[14]



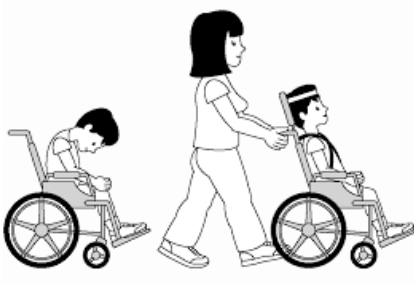
GMFCS Level IV

Individuals may be able to walk short distances with the aid of a walker or be more reliant on wheelchair mobility.[14]



GMFCS Level V

All areas of motor function are affected and individuals are completely wheelchair dependent. Voluntary control of movement and control of head and trunk posture are affected.[14]



Individuals with cerebral palsy may also have a variety of physical and cognitive impairments, including musculoskeletal dysfunction e.g. hip dislocation, scoliosis and

joint contractures; varying degrees of speech impairment, due to the inability to coordinate the muscles of the mouth and tongue required for speech; sensory loss; sleep disturbances; intellectual disability; epilepsy; autism; and behavioural and emotional problems.[8,10]

1.3 Neuroimaging

Neuroimaging, including ultrasound, magnetic resonance imaging (MRI) and computed tomography, have been utilised in some cases to assist with cerebral palsy diagnosis and classification. Approximately 80 - 86% of individuals with cerebral palsy have been reported to have an abnormal MRI brain scan.[15-18] Whilst an abnormal MRI is not required for cerebral palsy diagnosis, associations with cerebral palsy have been described: 1) periventricular leukomalacia with prematurity,[19] 2) basal ganglia injury with some term deliveries,[19] and more recently 3) pancake cerebellum and developmental colpocephaly with extreme prematurity.[20]

A recent review of brain imaging patterns in cerebral palsy cases found brain abnormalities in 86% of scans.[18] The most prevalent imaging patterns were white matter injury (19 – 45%) and grey matter injury (21%), followed by malformations (11%) and focal vascular insults (10%). White matter injury was common in cases diagnosed with spastic cerebral palsy but more so in cases with spastic diplegia (31 - 60%). Grey matter injury was more prevalent in cases with spastic quadriplegia (34%) and dykinesia (21%). Cases diagnosed with ataxic cerebral palsy were more likely to have normal imaging (24% - 57%). White matter injury was more frequent in preterm cases (<37 weeks) compared to term cases, 31 – 79% and 12 – 32%, respectively. In contrast, term cases had a higher rate of grey matter injury (21%) compared to preterm

babies (4 – 20%); a similar trend was seen for malformations (13% compared to 7%) and focal vascular insults (12% compared to 5%). Reported malformations included agenesis of the corpus callosum, polymicrogyria, lissencephaly, schizencephaly, heterotopia, pachygyria, hydrocephalus, holoprosencephaly and hydranencephaly.[18]

A trend towards lower rates of normal imaging was observed for GMFCS levels IV and V compared with levels I and II.[18] In a separate study, normal brain imaging was reported in 10% of cases with a mild degree of spasticity and those with a non-spastic motor type.[21] These results were comparable with earlier findings where 11.7% of cases reported with mild spasticity or a non-spastic motor type had a normal brain MRI.[17] Limitations of contemporary imaging, has been suggested as an explanation for the normal imaging in a minority of cases of cerebral palsy.[18,21]

1.4 Incidence and economic cost of cerebral palsy

Cerebral palsy is the most common physical neurological disability of childhood and is three times more likely to be diagnosed than cancer in a child less than 18 years of age. The prevalence of cerebral palsy throughout the developed world has remained relatively unchanged in the last 50 years and affects 2-3 in every 1,000 live births.[22-28] In contrast, over the same time period, there have been significant improvements to prenatal care with substantial decreases in both perinatal and maternal mortality.[29] In 2007 there were 33,797 estimated cases of Australians with cerebral palsy. This is an immense social and economic burden with an annual financial cost of AU$1.47 billion and an additional AU$2.4 billion in lost wellbeing.[30] In the United States, it is estimated that cerebral palsy affects 2.3 to 3.6 in every 1,000 live births and it is

estimated that 764,000 individuals currently have a cerebral palsy diagnosis. The 2003 estimated lifetime cost per individual was US$921,000.[25,27]

1.5 Known clinical risk factors for cerebral palsy

There are a number of known clinical risk factors for cerebral palsy, which can occur singly or in combination. The Australian Cerebral Palsy Register Report (2013) identified four groups that statistically have a greater risk of cerebral palsy including preterm birth, low birth weight, multiple birth and male gender.[28]

1.5.1 Preterm birth

Preterm birth is the largest risk factor for cerebral palsy and is present in 35% of all cases.[31] Cerebral palsy is the most frequent adverse neurological outcome associated with preterm birth and the risk increases the lower the viable gestation.[31,32] This was evident in a recent large study, which consisted of 1.7 million children delivered in Norway between 1967 and 2001. Gestational age ranged between 23 and 43 weeks and the total prevalence of cerebral palsy in this birth cohort was 1.8 per 1000 births.[33] The absolute risk of cerebral palsy was 8.5% for children born between 23 and 27 weeks gestation; 5.6% for those born at 28 to 30 weeks; 2.0% for those born at 31 to 33 weeks; 0.4% for those born at 34 to 36 weeks and 0.1% for term babies.[33]

1.5.2 Low birth weight

Over a third of all cerebral palsy occurs in children of low birth weight, who include babies born preterm and/or growth restricted. There is a direct correlation with low birth weight and preterm birth, and increased survival rates of preterm and very low birth weight babies (<1500 g) coincides with increased numbers of cerebral palsy

cases.[34] Intrauterine growth restriction (IUGR) is a condition where the fetal size is less than expected for gestation and may be associated with significant fetal or neonatal complications including cerebral palsy. Whilst several factors may cause IUGR, it usually results from placental insufficiency, where the placenta is unable to supply adequate oxygen and nutrients to the developing fetus. IUGR is assessed using customised birth weight percentiles based on the growth potential calculated for each baby.[35] A study by Jacobsson *et al*. used a cut-off for growth restriction of below the 10th percentile and below the 1st percentile for severe growth restriction. Compared to controls matched for gestation, cerebral palsy cases born at term had a significantly lower median birth weight; in contrast, this trend was not seen in cerebral palsy cases born preterm.[36] Another study using customized birth weight centiles found an increased risk of IUGR for cerebral palsy in neonates with birth weights below the 10th percentile and an even greater risk in babies below the 3rd percentile.[37]

1.5.3 Multiple birth

Twins, triplets and other multiple births have a five- to 15-fold increased risk for cerebral palsy.[38,39] The risk increases even further in a surviving twin, whose co-twin died in utero or shortly after birth.[39-42] Approximately 12% of Australian individuals diagnosed with cerebral palsy were from multiple births.[28] In part this is attributed to the higher proportion of preterm births and/or low birth weight in multiple pregnancies compared to singletons.[41]

1.5.4 Male gender

The incidence of cerebral palsy has been reported to be higher in males compared to females (male:female ratio 1.3:1).[43,44] Some of this excess can be explained by

abnormal intrauterine size, either small or large. Male babies are relatively more vulnerable to growth deviation than female babies.[45] The frequency of cerebral palsy has been reported to be higher in male babies with a birth weight 1000 – 1499 g compared to female babies in this category.[34]

1.6 Other known clinical risk factors

Other factors associated with increased risk of cerebral palsy include, maternal and intrauterine infection,[46,47] coagulation disorders,[48] ischaemic stroke in the fetus,[48] placental pathology i.e. chorioamnionitis, funisitis and necrotising funisitis,[49,50] a sibling with cerebral palsy[38] and birth asphyxia, the latter having been considered the leading cause of cerebral palsy in the past.[51,52]

1.6.1 Birth asphyxia

A common non-evidence based belief in the past has been that most cases of cerebral palsy are due to fetal distress in labour, birth asphyxia or birth trauma. However, this belief has had to be revised as epidemiological studies indicate that as few as 10% of cerebral palsy cases show possible signs of intrapartum fetal compromise at birth, and some of these signs may have been a result of chronic pregnancy pathologies.[53-55] Over the last 50 years both elective (prior to labour) and emergency (in labour) caesarean deliveries have increased six-fold (Australian Bureau of Statistics: www. abs.gov.au) without any reduction in the incidence of cerebral palsy [28] (Figure 1.2). This belies the belief that difficult vaginal delivery or “birth asphyxia” commonly contributes to cerebral palsy and suggests that cerebral palsy aetiology is more complex than first thought.

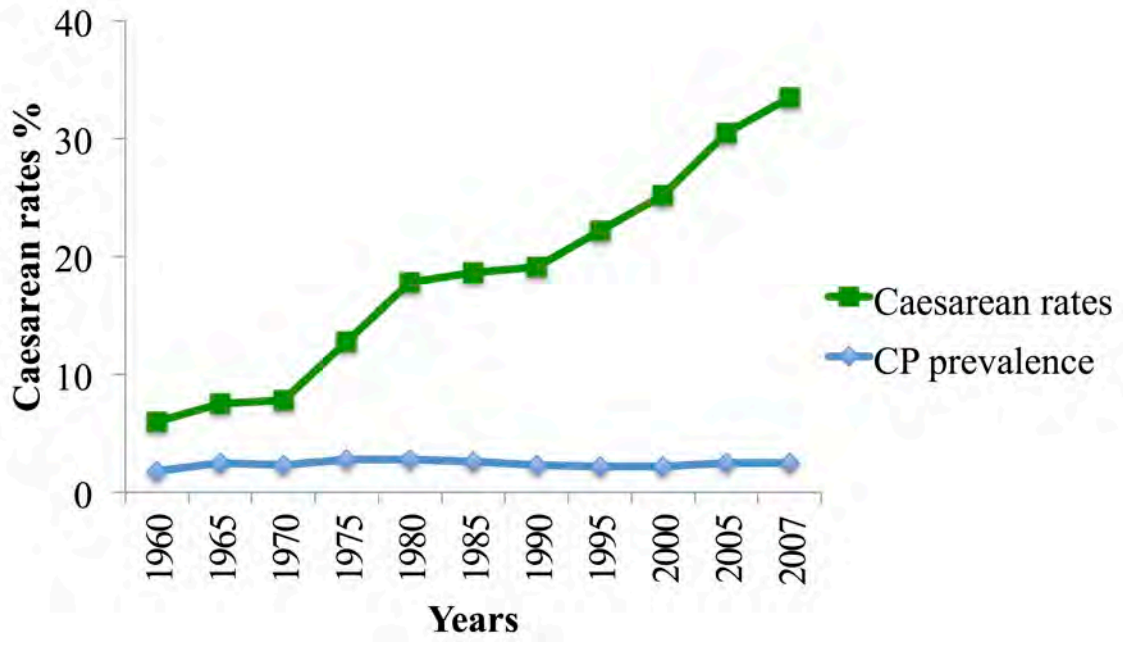


Figure 1.2 Prevalence of cerebral palsy rates per 1000 live births compared with caesarean rates over the past 50 years. Caesarean rates are based on statistics from the Australian Bureau of Statistics: www.abs.gov.au.

1.7 Other possible risk factors

Other possible risk factors that need further consideration include parental age[56] and assisted reproductive technologies.[57] Parental age effect has not been widely studied for cerebral palsy but an effect has been reported in some sub-types, including athetoid/dystonic,[56] ataxia[58] and spastic hemiplegia .[56] In a recent study consisting of four individuals diagnosed with ataxic cerebral palsy, three different *de novo* mutations were found and associated with increased paternal age.[58] Maternal and paternal age have been reported to effect autism and intellectual disability, two disorders comorbid with cerebral palsy.[59,60] An association between assisted reproductive technologies and cerebral palsy has also been investigated. Singleton and multiple births resulting from *in vitro* fertilization (IVF) have a higher prevalence of cerebral palsy compared to naturally conceived children.[61,62] Recently, Davies *et al.* (2012) examined data from 308,974 births, 6,163 the result of assisted conception. Following multivariate

adjustment a two-fold risk for cerebral palsy in singleton IVF cases [OR 2.22 (1.35–3.63)] was observed.[57] Multiple pregnancy and other epidemiological confounding risk factors in assisted conception pregnancies may contribute to the increase in cerebral palsy in this group. Despite these known clinical risk factors, for the majority of cases the exact determinants for injury to the child's developing brain remain unknown.

1.8 Evidence for a genetic contribution to cerebral palsy causation

Cerebral palsy is not considered a genetic disorder, yet there are several lines of evidence suggesting a genetic involvement. The risk of recurrence is higher in families where there is one affected sibling,[38] monozygotic twins have a higher concordance rate than dizygotic twins;[63] cerebral palsy is 2.5 times higher in consanguineous families,[64] and there is a higher rate of congenital anomalies in cerebral palsy individuals compared with the general population.[65] Further, families with more than one individual diagnosed with cerebral palsy are consistent with Mendelian inheritance, however for some of the families polygenic/multifactorial inheritance and/or environmental factors such as illicit drug use, smoking, chorioamnionitis and infection cannot be dismissed as contributing factors.[37,46,47,49,66,67]

1.8.1 Sibling risks and twin studies

Information based on familial clustering of cerebral palsy has been the main avenue to establish recurrence risk for sporadic cases. However, the majority of these studies have been inconclusive due to small sample sizes.[68-70] A much larger study by Hemminki and colleagues analysed the familial risks for cerebral palsy for 3,997 cerebral palsy cases hospitalised in Sweden between 1987 and 2001.[38] The majority of cases were singletons, with twins accounting for 203 (5.1%) and triplets accounting for

eight (0.2%) of all cases. Triplets were not considered further as no multiple affected siblings were identified. The risks were compared between siblings of affected and healthy individuals. Familial cerebral palsy was as low as 1.6% of all cerebral palsy cases (64 siblings). Compared to siblings of healthy individuals the risk of recurrence in another sibling to parents of one affected child was 4.8-fold higher and 29-fold higher if the siblings were twins.[38] A significantly higher concordance rate for cerebral palsy has been reported in monozygotic twins than in dizygotic twins.[63] Two similar versions of the International Classification of Diseases (ICD) versions -9 (1987-96) and -10 (1997-2001) were used for cerebral palsy diagnosis. Standardized hospitalisation ratios (SHRs) were calculated as the ratio of observed to expected number of cases. They reported that congenital diplegia and hemiplegia were amongst the largest specified groups in both IDC classification groups. The SHRs for diplegia were higher for singletons (24.90) compared to hemiplegia (17.20) and quadriplegia (19.47). In contrast, quadriplegia (155.04) and hemiplegia (104.44) were higher than diplegia (37.63) for twins.[38]

1.8.2 Effect of consanguinity

Consanguinity results in a 2-2.5-fold increased risk for cerebral palsy with inheritance likely due to autosomal recessive transmission.[64,71] A study conducted in Saudi Arabia, consisting of 103 children diagnosed with cerebral palsy and 103 matched controls, found consanguinity was the major risk factor for cerebral palsy (OR 2.31).[64] In a separate study, the prevalence of cerebral palsy for non-Asian and Asian populations within ethnic subgroups born in Britain was 3.18 and between 5.48 and 6.42 per 1000 individuals, respectively. First cousin marriages were reported in 51% of the Asian families and nine of these families had another first or second-degree family member

with a similar type of cerebral palsy to the index child. In comparison no consanguinity was reported in the non-Asian families.[71]

1.8.3 Associated congenital anomalies

Compared to the general population, congenital anomalies are reported to be higher in cases diagnosed with cerebral palsy (2 - 4% compared to 12 - 19%, respectively).[65,72-74] A study by Croen *et al*. found congenital anomalies were present in 19% of children diagnosed with cerebral palsy compared to 4% in controls.[74] Two later studies, one by Garne *et al*., found congenital malformations in 12% of the children diagnosed with cerebral palsy[73]; the second, by Rankin *et al*., reported congenital anomalies in 15% of cerebral palsy cases.[75] For each of the above-mentioned studies, microcephaly and congenital hydrocephalus were the most frequent cerebral anomalies.[75] Non-cerebral disturbances included cardiac anomalies (12%), musculoskeletal anomalies (5.4%) and urogenital abnormalities (5.4%) (Figure 1.3).[10,13,65] Other reported comorbidities in cerebral palsy cases include; intellectual disability (30 - 65%),[10,76] epilepsy (22 - 46%),[10,65,76] autism (9%),[10,76] speech deficits (28 - 43%),[10,76] and hearing (5 – 15%)[10,76] and visual impairments (40%)[10] (Figure 1.3). Many of these comorbidities, in particular intellectual disability, epilepsy and autism, have been extensively examined for genetic abnormalities, resulting in the successful identification of numerous risk and causative genes and copy number variants (CNVs).[77-83] In some instances movement disturbances associated with cerebral palsy have been part of the clinical spectrum.

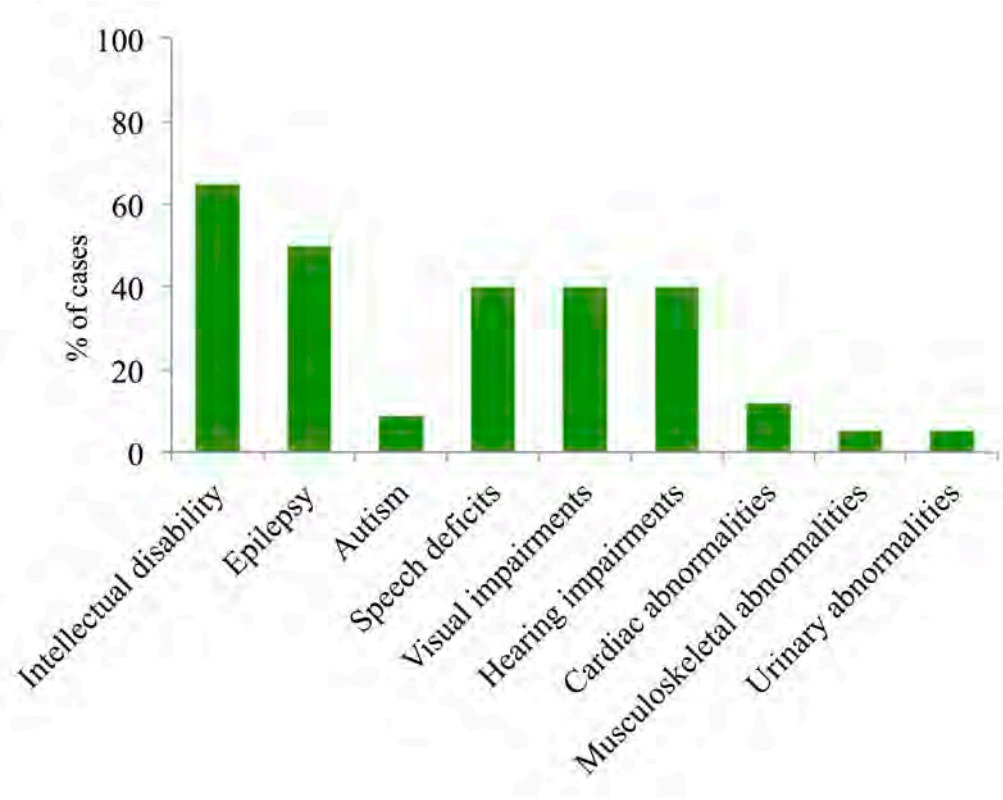


Figure 1.3 The prevalence and type of co-morbidities seen in individuals diagnosed with cerebral palsy.

1.8.4 Families providing evidence for Mendelian inheritance

Families with more than one individual with cerebral palsy, either in the same or different generations, provide evidence for genetic susceptibility to cerebral palsy.

Blumel *et al.* (1957),[66] was among the first to observe a pedigree consistent with an X-linked pattern of transmission in a large non-consanguineous four-generation family. The family consisted of four affected males over two-generations presenting with similar spastic involvement. It was hoped that their preliminary findings would not only encourage investigators to consider the genetic component in cerebral palsy but that X-linked transmission may also explain the higher incidence of males diagnosed with cerebral palsy.[66] A separate study reported on two non-consanguineous families each with two affected individuals. The first family included two sisters diagnosed with spastic diplegia, one with normal intelligence and the other with intellectual disability

and hypothyroidism. Autosomal recessive transmission was considered the most likely pattern of inheritance for this family. The second family consisted of two brothers with spastic quadriplegia and normal intelligence. As there were two affected brothers, both X-linked inheritance and autosomal recessive transmission were considered as a possible mode of inheritance for this family.[84]

Ataxic cerebral palsy, which affects approximately 15% of cases, has also been considered to have genetic origins.[67] There are several reports of families with more than one individual diagnosed with ataxic cerebral palsy, with inheritance patterns consistent with autosomal recessive, dominant and X-linked inheritance.[85-87] One example, consistent with autosomal recessive transmission, is a large consanguineous family with four adults and three children over several generations with identical clinical features including ataxia, signs of spasticity, short stature and normal intelligence.[85] Another example is a non-consanguineous family consisting of an affected mother and daughter. The mother had an unaffected brother and unaffected son. Autosomal dominant inheritance was considered the most likely mode of transmission for this family.[86]

1.9 Genetic disorders that can be mistaken for cerebral palsy

Several reviews have described the shared clinical features of cerebral palsy with other neurodevelopmental disorders, leading to a misdiagnosis of cerebral palsy in some cases, especially early in life.[9-11] One group of genetic disorders, which have been reported to present as cerebral palsy in approximately 24% of cases, are the dopa-responsive dystonic disorders.[88] For example, following whole-exome sequencing (WES), a twin pair initially diagnosed with cerebral palsy was found to have compound

heterozygous mutations in the *SPR* gene.[89] Disruption of this gene leads to a decrease in tetrahydrobiopterin (a cofactor in the synthesis of dopamine and serotonin) and administration of levodopa supplemented with a serotonin precursor (oxitriptan) resulted in immediate improvement of symptoms.[89] In another study, a woman who had lived with a diagnosis of cerebral palsy for 10 years was found to have dopa-responsive dystonia resulting from a mutation in *GCH1*. Again, administration of levodopa dramatically improved her condition.[90]

Another group of genetic disorders that can masquerade as cerebral palsy are disorders associated with spasticity. Several examples exist, including the hereditary spastic paraplegias, which are a heterogeneous group of progressive degenerative disorders, with over 70 genetic types documented to date.[91,92] Hereditary spastic paraplegia is distinguished from cerebral palsy by the worsening of spasticity with age.[93,94] Pelizaeus-Merzbacher disease (PMD), an X-linked neurodegenerative disorder of myelination, is another example and results from mutations in the proteolipid protein (*PLP1*) gene.[95] *PLP1*, a main component of the myelin membrane, plays an essential role in myelin sheath formation.[96] Two known disease genes, *L1CAM,* associated with L1 syndrome, and *TUBA1A,* involved in cortex development and associated with neuronal migration disorders, can be associated with clinical features that meet criteria for a diagnosis of cerebral palsy. Spasticity, poor muscle control and intellectual disability have been reported in patients with mutations in both of these genes.[97,98]

Ataxic/hypotonic disorders including Joubert syndrome and Angelman syndrome, are another group of disorders where affected children can have features that meet the criteria for cerebral palsy. Joubert syndrome is caused by mutations in genes involved in

formation and function of primary non-motile cilia, including *AH11*, *CEP290* and *NPHP1*.[99] Clinical phenotypes for Joubert syndrome include ataxia, intellectual disability and global developmental delay.[99,100] Angelman syndrome is a neurodevelopmental disorder characterised by ataxia, seizures, autism, intellectual disability and global developmental delay.[101] It is an example of genomic imprinting, and results from either a deletion on the maternally inherited chromosome 15, as a result of which the paternal copy is imprinted or silenced, or alternatively, paternal uniparental disomy for the same region of chromosome 15. A sub-set of cases can also result from a mutation in *UBE3A*.[102]

1.10 Approaches to identification of genes involved in cerebral palsy

It needs to be emphasised at the outset that cerebral palsy is highly heterogeneous and not a single disorder. Consequently, the search for 'cerebral palsy genes' should be considered a search for genes and their variants associated with susceptibility to multiple neurodevelopmental disorders that share clinical features meeting the criteria for a clinical diagnosis of cerebral palsy. Conceptually, cerebral palsy can be considered either as a monogenic or a complex genetic disorder. To identify the responsible genetic determinants different methods of investigation, i.e. linkage analysis, candidate and genome wide association studies for complex inheritance and candidate, linkage and whole-exome/genome sequencing for single gene mutation discovery, can be used.

1.10.1 Linkage analysis and homozygosity mapping

Conventional linkage studies have proven to be a useful tool for identifying disease-causing genes segregating within families[104] and homozygosity mapping a powerful tool to identify recessive variants in consanguineous families.[105,106]

1.10.2 GAD1

McHale and colleagues (1999)[107] were the first to report genetic linkage for autosomal recessive spastic cerebral palsy in three out of the eight consanguineous Pakistani families with multiple affected children. Following a genome-wide search for linkage using 290 polymorphic DNA markers, a 5cM region of homozygosity was identified at chromosome 2q24-25. In each case cerebral palsy was described as 'non-progressive' and no perinatal cause was identified. The clinical phenotypes for these families have been previously described.[108] In summary there were eight affected individuals in three families. Clinical features in two out of three families (son and daughter, and two sons and two daughters, respectively) included; spastic diplegia, moderate to severe intellectual disability and microcephaly. The two affected individuals from the remaining family were diagnosed with spastic quadriplegia, severe intellectual disability and microcephaly. Subsequently a novel missense change was found in the *GAD1* gene, which segregated with the affected individuals.[12] *GAD1* encodes glutamate decarboxylase, which is involved in the production of the important inhibitory neurotransmitter gamma-aminobutyric acid (GABA) (Table 1.3).[107] However, this change was not functionally tested and no further *GAD1* changes have been found in individuals with cerebral palsy.

1.10.3 KANK1

Linkage analysis in a large four-generation Israeli family consisting of nine individuals diagnosed with cerebral palsy identified a 225 kb deletion on 9p24.3 involving *KANK1* (previously known as *ANKRD15*) and not found in 210 control individuals. Highly expressed in the fetal brain, the protein encoded by *KANK1* regulates actin polymerisation and cell migration.[109] The affected children were born at term to non-

consanguineous parents and later diagnosed with a congenital neurodegenerative disease resembling cerebral palsy. Congenital hypotonia evolved to spastic quadriplegia within the first year, accompanied by transient nystagmus and moderate to severe intellectual disability with brain atrophy and ventriculomegaly. Pregnancies and deliveries appeared normal for each individual.[109] Only paternal offspring were affected, which, together with gene expression data, led the authors to suggest that this family demonstrated the phenomenon of a deletion creating imprinting-like inheritance.[109]

1.10.4 Chromosome 9p12-q12

A study by McHale and colleagues (2000)[110] identified a region of homozygosity at chromosome 9p12-q12 in a large consanguineous Asian pedigree with four affected children in two sibships. Three of the children were diagnosed with ataxic cerebral palsy and one child with ataxic diplegia. In each case the ataxia was static. All four affected individuals were born at term following an unremarkable pregnancy and normal delivery. In all cases motor milestones were delayed but intelligence was within the normal range. Two positional candidate genes mapped to this region; *FRDA* and *PAX5/BSAP*. Mutations in *FRDA* have been associated with Friedreich's ataxia, a progressive condition that presents later in life with no cases reported with congenital onset. The encoded protein from *PAX5/BSAP*, Pax5, is expressed in the developing mesencephalon, which then forms the cerebellum. Several mouse models map to this region. The mouse model *waddler* (the best candidate but now extinct) maps within 20cM of the murine *Pax5* and presented with a non-progressive ataxia from 14 days of age. *PAX 5/BSAP* was the favoured candidate but as yet no other cases have been identified in cerebral palsy.

1.11 Candidate-gene association studies

Candidate-gene studies focus on a select set of genes with known biological or functional relevance to the disease in question.[111] There have been several hypothesis-driven targeted candidate-gene association studies undertaken to determine a link between genetic risk factors and cerebral palsy. In particular, single nucleotide polymorphisms (SNPs) have been identified in a number of cerebral palsy sub-types in genes involved in inflammation, infection[47] thrombophilia and preterm birth risk.[47,112-120] In a comprehensive review of 22 targeted association studies, following the HuGENet guidelines, O'Callaghan *et al.* (2009) concluded that the majority of these studies were underpowered, resulting in conflicting results.[121] Two recent case-control studies reassessed previously reported associations between cerebral palsy and previously reported candidate SNPs. The first, re-examining 15 SNPs, comprised 138 non-Hispanic white cases ≥36 weeks gestation and 165 randomly selected controls (excluding non-whites).[122] The most recent, consisted of 587 Caucasian mother/child case pairs (including all sub-types and all gestational ages) and 1154 Caucasian mother/child control pairs. Thirty-five candidate SNPs were selected from the literature.[123] Results from earlier association studies were not replicated in either study, with no statistically significant associations remaining after allowing for multiple testing.[122,123] This approach was not only limited by its reliance on existing knowledge but the majority of these studies were underpowered and could not be replicated.[124] Another point of debate for candidate-gene association studies is that most identified SNPs have a relatively small effect size; that is, they are more likely to confer disease risk, in contrast to monogenic disorders where a single gene mutation with complete penetrance can result in disease[125-127] (Figure 1.4).

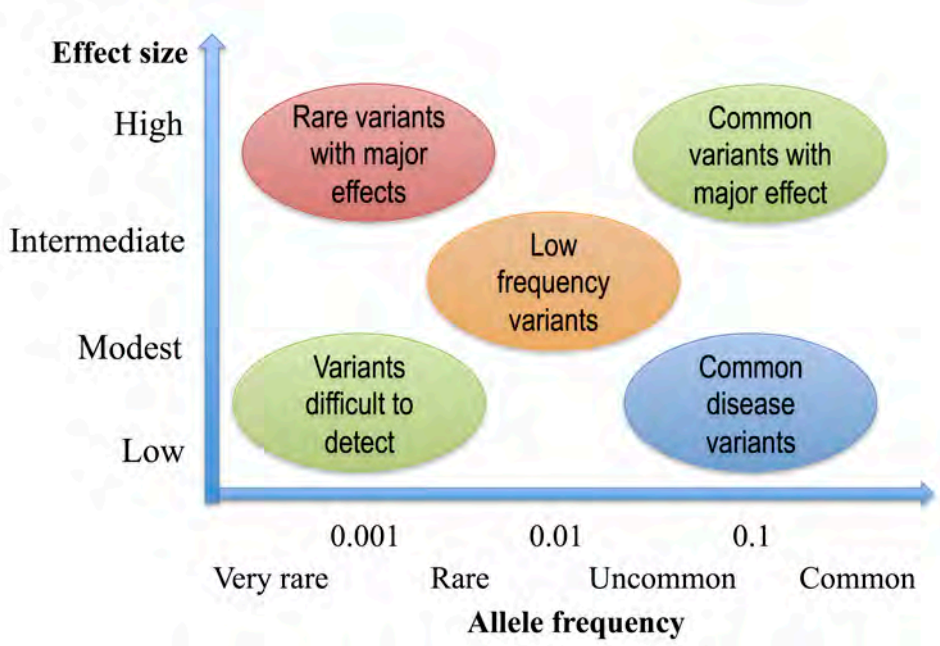


Figure 1.4 Allele frequency correlated with effect size. Adapted from Manolio et al.[127]

1.12 Copy Number Variants

The advances in molecular genetics, in particular comparative genomic hybridisation (CGH) and single nucleotide polymorphism (SNP) microarrays have facilitated the identification of DNA variations involving segments too small to be recognised microscopically but much larger than those previously detectable by conventional sequence analysis.[128,129] These submicroscopic variants (>50 bp), primarily deletions and duplications, alter the diploid status of a particular locus and are referred to as copy number variants (CNVs). CNVs contribute to a substantial amount of the genetic variation in the human genome, accounting for both population diversity and disease.[129,130] Pathogenic CNVs are likely to be rare (0.1 – 1% population frequency) and can occur *de novo* (absent in both parents' genomes) or be inherited (present in one or both parents and demonstrating variable penetrance and expressivity).[131] There is now convincing evidence that rare CNVs have a significant role in the aetiology of both common and more complex neurological disorders such as intellectual disability,[132-136]

autism[137-139] and epilepsy.[140-142] An identical CNV in the same genomic region can be associated with more than one neurological phenotype.[132,137,138,143] Examples include a 1.6 Mbp microdeletion on 1q21.1 and a 1.5 Mbp deletion on 15q13.3, each identified in individuals with developmental delay,[132,144] autism[145,146] and schizophrenia.[143] 15q13.3 deletions have also been associated with an increased risk for epilepsy.[141] Another example is a microdeletion on 3q29 found in individuals with developmental delay[147,148] and schizophrenia[149]. These association between a single CNV and more than one complex neurological phenotype suggest that the phenotypes share genetic susceptibility factors and have overlapping causal pathways.[150] The vast majority of causative CNVs occur spontaneously (*de novo*) with high penetrance due to enormous selection pressure. Causative CNVs can also be inherited from a seemingly unaffected parent.[129] For example, in multiplex autism families, dominant transmission has been reported in affected male offspring from a maternal parent with low or no penetrance.[151] The inherited predisposition may not be sufficient on its own to result in disease and in some families a second gene mutation may be necessary to result in disease.[133]

As yet the contribution of CNVs to cerebral palsy has not been widely investigated. A recent genome-wide search for CNVs in 50 Caucasian families having children with cerebral palsy cases did not yield any *de novo* mutations but identified inherited rare CNVs involving potentially pathogenic genes requiring further investigation.[152] In total, 14 different rare inherited variants (5 deletions and 9 duplications) were seen in 10 out of the 50 cases (20%). Eight were inherited from an unaffected mother, including two different duplications involving clinically relevant genes, *CTNND2* and *MCPH1* (Table 1.1). *CTNND2* has an important role in neuronal functioning, adhesion, and migration and is involved in early embryogenesis.[153-155] Deletions of *CTNND2* have been associated with intellectual disability in Cri-du-chat syndrome and autism.[153,156]

Duplications of *CTNND2* have been associated with schizophrenia.[155] Microcephalin, the product of *MCPH1*, is involved in neurogenesis and regulation of cerebral cortex size. Homozygous loss of function mutations of the *MCPH1* gene cause autosomal recessive disorders including premature chromosome condensation syndrome,[157] intellectual disability,[157] and microcephaly.[158] Heterozygous deletions and duplications of *MCPH1* have been reported in families with autism spectrum disorders, supporting the concept that *MCPH1* is a dosage-sensitive gene, with considerable mutation pleiotropy.[159] Three CNVs were inherited from an unaffected father including a deletion in *COPS3*, involved in several cellular and developmental processes and in signal transduction (Table 1.1).[160] For the remaining three CNVs the inheritance was unknown. *COPS3* maps to chromosome 17p11.2, within the Smith-Magenis syndrome and Potocki-Lupski syndrome critical interval.[161,162] Both syndromes are characterized by congenital anomalies and intellectual disability.[163] Inheritance for the remaining three CNVs was not known.[159]

In contrast, a more recent study identified 39 CNVs in 25 out of 52 individuals (48%) diagnosed with cerebral palsy. Clinically relevant CNVs, 10 pathogenic and six likely pathogenic variants, the majority *de novo*, were found in one-third (31%) of the participants.[164] In total, seven *de novo* CNVs were observed, four containing genes associated with clinical features relevant to cerebral palsy. These included deletions in *SPAST*, a well-known hereditary spastic paraplegia gene; *MEF2C*, which is associated with dystonia; *WDR45*, which is associated with spasticity; and *NKX2-1*, which is associated with benign hereditary chorea. In addition to the *de novo* CNVs, a paternally inherited CNV was identified involving *KANK1*, previously associated with spastic quadriparesis[164] and congenital neurodegenerative disease resembling cerebral palsy

(Table 1.1).[109]

A larger more recent study found 12 *de novo* CNVs (seven deletions and five duplications) in eight out of 115 (7%) cases diagnosed with cerebral palsy.[165] Three cases had one or more CNVs considered pathogenic for cerebral palsy, two cases had a CNV considered likely pathogenic for cerebral palsy and three cases had a CNV of uncertain clinical significance to cerebral palsy. A maternally inherited CNV considered pathogenic for cerebral palsy was also identified. Genes affected by *de novo* and/or inherited CNVs included *KANK1*, recently associated with cerebral palsy in a separate CNV study[164], and three genes involve in endoplasmic reticulum pathways, *PARK2, PACRG* and *HSPA4* (Table 1.1).[165] Previous studies have demonstrated that immature neurons and preoligodendrocytes important to white matter formation are susceptible to apoptosis resulting from endoplasmic reticulum stress.[166] More cases would need to be tested to determine the full extent to which CNVs are involved with cerebral palsy causation.

Table 1.1 Summary of potentially pathogenic CNVs encompassing genes previously associated with cerebral palsy.

Cytoband	Size kb	CNV type	Candidate genes	OMIM	Inheritance	CP subtype	Reference
5p15.2	447	Dup	*CTNND2*	604275	Maternal	Hemiplegia	McMichael et al.[159]
8p23.2-p23.21	219	Dup	*MCPH1*	607117 and 251200	Maternal	Hemiplegia	McMichael et al.[159]
17p11.2	4	Del	*COPS3*	604665	Paternal	Hemiplegia	McMichael et al.[159]
2p23.1-p22.2	5,214	Del	*SPAST*	604277	*de novo*	Diplegia	Segal et al.[164]
5q14.3	3,463	Del	*MEF2C*	600662	*de novo*	Quadriplegia	Segal et al.[164]
Xp11.23-p11.22	4,287	Del	*WDR45*	300526	*de novo*	Quadriplegia	Segal et al.[164]
12q21.2-14q12	11,150	Del	*NKX2-1*	600635	*de novo*	Quadriplegia	Segal et al.[164]
9p24.3	226	Del	*KANK1*	607704	Paternal	Quadriplegia	Segal et al.[164]
9p24.3-p24.1	8,102	Del	*KANK1*	607704	*de novo*	Quadriplegia	Oskoui et al.[165]
6q26	350	Dup	*PARK2* and *PACRG*	602544 and 608427	*de novo*	Quadriplegia	Oskoui et al.[165]

Abbreviations: CNV – Copy number variant; kb – kilobase; Dup – duplication; Del – deletion

1.13 Massively parallel sequencing

The quest to understand the genetic basis of disease has largely been dependent on the available genetic technologies at the time of investigation. The past 10 years have seen several advances in genotyping technology but the development of massively parallel sequencing technologies (MPS) has had the most impact on genomics.[124,167] For more than 30 years traditional Sanger sequencing has been the main DNA sequencing method, ultimately enabling the first human reference sequence in 2001 and completion of the first version of the human genome in 2004.[168] The Human Genome Project required enormous amounts of time and resources and demonstrated the need for faster, higher throughput and cheaper technologies, if sequencing was to make a larger contribution to research and diagnosis.[167] Over the past few years MPS technologies have advanced dramatically, facilitating high throughput, cost-effective genomic DNA sequencing (whole-exome/genome).[167] This new technology has the capacity to read millions of short DNA sequences at the same time.[124]

1.14 Whole-exome sequencing

WES sequences the protein coding regions of the genome, known as the exome. The exome accounts for less than 2% of the entire genome. Of the 100,000 mutations catalogued in the human genome mutation database, 86% are in protein coding regions, which are captured by WES.[168] WES has become an effective and affordable strategy for identifying causative variants in common and complex disease. Several commercial exome-capture kits are available on the market, providing an efficient way to sequence exomes at a high sensitivity.

1.14.1 Whole-exome capture methods

Array-based technology using NimbleGen oligonucleotide arrays was the first method to be used for exome capture.[169] This technology has now been surpassed by solution-based methods. Solution-based methods involve the capture of fragmented genomic DNA with oligonucleotide probes (baits) that collectively cover the majority of the exonic regions.[168] Agilent SureSelect Human All Exon Capture kit was the first commercial sample preparation kit on the market.[170] Shortly after, SeqCap EZ Exome Capture System by Roche NimbleGen was released.[171] The main differences between the two sample preparation methods are the design of the capture probes, bait lengths and bait density. Agilent used long (120 bp) RNA baits targeting 37.6 Mb and NimbleGen used shorter (60 – 90 bp) DNA probes targeting 33.9 Mb (Table 1.2).[170-172]

Since the first 'exome' the concept and definition of an 'exome' has changed dramatically. Earlier capture kits captured approximately 80% of the human consensus-coding sequence regions with a minimum coverage of ≥20X.[173] 'Exome' now involves sequences outside the coding regions including 5' and 3' untranslated regions (UTRs), micro RNAs (miRNAs) and long noncoding transcripts. Capture efficiency of the protein-coding regions is now around 95% at a sequencing depth of ≥20X.[174] Currently there are three major providers of exome capture methods; Agilent, NimbleGen and Illumina.[168,175] Each platform targets specific exomic segments based on databases including UCSC,[176] RefSeq[177] and Ensembl.[178] There are considerable differences in the density of the oligonucleotide probes between the three platforms.[168,175] Agilent uses RNA probes and the corresponding target sequences (exomes or exomes +UTRs) are adjacent to one another as opposed to overlapping.[175] This design has longer baits than the other platforms and can better tolerate mismatches, and is therefore more efficient at

identifying indels (insertions and deletions).[179] Compared to NimbleGen, Agilent has fewer PCR duplicates; however, coverage is not as uniform.[180]

NimbleGen and Illumina use DNA probes and baits for both platforms are shorter compared to the Agilent platform. NimbleGen's Seqcap EZ Exome Library has the highest bait density of all commercial platforms, containing overlapping baits that cover the target bases multiple times. This design requires less sequencing to cover the target region, therefore has improved enrichment efficiency and fewer off-target reads compared to other platforms, however this results in more duplicate reads and a lower alignment rate compared to Agilent.[168,175]

Illumina capture kits (TruSeq and Nextera Rapid) use 95 bp probes designed to leave small gaps between target regions, relying on paired-end reads to fill in the gaps by extending outside the bait sequences. Compared to Agilent and NimbleGen, target efficiency is reduced due to a higher percentage of off-target enrichment.[168,175] Illumina's Nextera Rapid Capture Exome kit is unique as it uses transposomes as opposed to ultrasonication to fragment the genomic DNA. This technique has considerably shortened and simplified the library preparation protocol by optimising and reducing the hybridisation process. Illumina platforms are designed to enrich UTRs, most of which are not targeted by other platforms. Illumina +UTR kits have 22.5Mb of unique targets of which 21.8Mb of these are UTRs.[175] Commercial-based exome capture kits are summarised in Table 1.2.

Table 1.2 Summary of the different solution-based commercial exome capture kits.

Sequencing platforms	Probe type	Probe size (bp)	Target size (Mb)	Coverage design	Fragmentation method	Strengths	Weaknesses
Earlier exome capture kits 2009 - 2010							
Agilent Sure Select Human All Exon Kit	RNA	120	37	Adjacent probes	Ultrasonication	Lower number of duplicate reads	GC content coverage bias
NimbleGen SeqCap EZ Exome Library	DNA	60-90	34	Overlapping probes	Ultrasonication	Higher number of quality reads	GC content coverage bias
More recent exome capture kits as of May 2015							
Agilent Sure Select Human All Exon Kit Basic/+UTR	RNA	114-126	37/64	Adjacent probes	Ultrasonication	Better indel coverage, less duplicate reads, higher alignment, Coverage of UTRs	Less number of high quality reads
NimbleGen SeqCap EZ Exome Library	DNA	55-105	64	Overlapping probes	Ultrasonication	More uniform coverage in difficult regions	Higher number of duplicate reads, lower alignment rate
Illumina TrueSeq Exome Enrichment Kit Basic/+UTR	DNA	95	37/62	Gaps between probes	Ultrasonication	Coverage of UTRs and miRNAs	Increased off-target enrichment
Illumina Nextera Rapid Capture Exome Kit Basic/+UTR	DNA	95	37/62	Gaps between probes	Transposomes	Coverage of UTRs and miRNAs, shorter processing time	Increased off-target enrichment, GC content coverage bias

Abbreviations: bp – base pair, Mb – Megabase, UTRs – untranslated regions, miRNAs – micro ribonucleic acid, RNA – ribonucleic acid, DNA – deoxyribonucleic acid

1.15 Whole-exome sequencing identifying single gene mutations

WES has had an enormous impact on determining disease-causing variants in common and complex diseases and been shown to be a powerful tool for the identification of novel disease genes.[181,182] WES is a suitable platform for identifying potential disease causing mutations in families with a small number of affected individuals. Rare variants common to the affected individuals and not found segregating in unaffected individuals can be filtered out for further interrogation. Corbett *et al.* (2010) identified the first autosomal-recessive syndrome of focal epilepsy, dysarthria and moderate intellectual disability in a large consanguineous family utilising WES.[183] In another study, a homozygous mutation in the Golgi vesicle transport gene was identified in an individual with progressive myoclonus epilepsy with early ataxia.[79] A similar approach identified *TGM6* as a novel causative gene of spinocerebellar ataxias in a single four-generation spinocerebellar ataxia family.[184] The power of WES was clearly demonstrated in a study that identified 50 novel autosomal recessive mental retardation genes.[82] Another applicable strategy involves sequencing child-parent trios to identify rare *de novo* variants in the affected individual. This study design is highly effective in disorders where the majority of cases are sporadic i.e. parents unaffected.[185] The considerable contribution of *de novo* variants to intellectual disability,[77,186-188] autism spectrum disorder,[60,83,150,189,190] schizophrenia[191,192] and more recently epilepsy[193-195] has since been well established due to this novel technology.

WES facilitates the discovery of an unprecedented number of new variants and genes but not all will be pathogenic. A major challenge is to identify which variants and genes are causative for disease. Several published guidelines and recommendations are now

established to assist in determining the pathogenicity of sequence variants and genes prior to experimental validation in animal and *in vitro* models.[196-199]

1.16 Whole-exome sequencing identifying single gene mutations in cerebral palsy

WES technologies have facilitated identification of a plethora of rare and/or novel genes/variants involved in neurodevelopmental disorders that overlap with cerebral palsy, including intellectual disability, autism and epilepsy.[187,188,200] These technologies, whilst in their infancy, are now being applied to the investigation of cerebral palsy causation. Previous estimates have suggested that the contribution of genetic variants to the burden of cerebral palsy is about 2%.[201] Mutations identified in single genes (discussed below and summarised in Table 1.3), which cause or predispose to cerebral palsy, provide 'proof of principle' of genetic involvement.[12,107,110,202,203]

1.16.1 AP-4 complex

The AP-4 complex consists of four sub-units; *AP4E1*, *AP4M1*, *AP4B1*, and *AP4S1* important for vesicular transport, thus necessary for correct intracellular transport, secretion and endocytosis.[204] Mutations in all four subunits of AP-4 *(AP4E1*, *AP4M1*, *AP4B1*, and *AP4S1)* have been reported in individuals with cerebral palsy-like motor disorders (Table 1.3).[205-210] In each case the phenotype included progressive spasticity and therefore does not fit the classical definition of cerebral palsy.[9] Recently, Jameel *et al.*[210] sequenced the exomes of two brothers diagnosed with cerebral palsy and intellectual disability, identifying a novel homozygous *AP4M1* mutation c.194_195delAT, p.Y65Ffs*50 in both brothers. In addition to spastic paraparesis and

intellectual disability, both boys were found to display aggressive behaviour, expanding the clinical spectrum for mutations in the four components of the AP-4 complex.[210]

1.16.2 ADD3

A homozygous mutation in *ADD3*, encoding gamma adducin, was found in a multiplex consanguineous family. The family, a second-cousin marriage, consisted of four affected siblings; both parents were unaffected. Three of the four affected individuals were diagnosed with spastic quadriplegia and one with spastic diplegia. Phenotypes also included borderline microcephaly and intellectual disability (Table 1.3).[203] *In vitro* studies suggested that mutations in *ADD3* disrupt the normal actin-capping function of adducin, resulting in abnormal proliferation and migration. Loss of function studies in *Drosophila* identified a critical role for adducin in locomotion.[203]

Further support for genetic involvement in cerebral palsy can be found in a recent study of 10 unrelated children diagnosed with sporadic ataxic cerebral palsy. They were examined using either targeted capture sequencing incorporating genes known to be associated with ataxia or trio-based exome sequencing. Four out of the 10 cases were found to have four different *de novo* mutations in three genes: *KCNC3* (n=1), *SPTBN2* (n=1) and *ITPR1* (n=2).[58] From a clinical prospective, the cases presented in this study fell within the current definition for cerebral palsy i.e. a group of permanent disorders of movement and posture attributed to non-progressive disturbances.[8]

1.16.3 KCNC3

Case one was the second child to non-consanguineous parents. He was born at term following an unremarkable pregnancy and normal vaginal delivery. He showed early

motor delay affecting sitting, crawling, standing and walking and was later diagnosed with ataxic cerebral palsy and mild intellectual disability. Brain imaging was normal. Targeted capture and sequencing of 57 known ataxic genes identified a novel, *de novo* variant, p.T428I, in *KCNC3,* a voltage-gated potassium channel K_v3.3 gene (Table 1.3).[58] Follow up functional studies showed a direct and indirect effect on channel function, exerting a severe dominant negative loss-of-function phenotype. Overall channel activity was reduced by at least 50% in the mutant p.T428I compared to the wild type. Mutations in *KCNC3* had previously been associated with spinocerebellar ataxia 13, a late onset progressive disorder,[58,211] and two other mutations in the gene, p.R423H and p.P448L, had been associated with a rare early-onset non-progressive autosomal dominant ataxia reminiscent of cerebral palsy.[212]

1.16.4 ITPR1

Analysis of WES identified two different variants, p.N602D and p.S1487D in *ITPR1* in two cases (case two and case three respectively) from two separate non-consanguineous families. Case two, a female, was born following a single miscarriage and conception difficulties. She did not meet her early developmental milestones and at age three was diagnosed with ataxic cerebral palsy and moderate intellectual disability. Brain imaging was normal.[58] Case three had increased nuchal thickness on prenatal ultrasound; karyotype was normal on amniocentesis. He was also behind in his developmental milestones and was diagnosed with ataxic cerebral palsy and moderate intellectual disability. Brain imaging was normal.[58] *ITPR1* encodes an intracellular receptor for inositol 1,4,5-trisphosphate and is associated with spinocerebellar ataxia type 15, a slowly progressive adult-onset disorder, although onset has been reported as early as seven years of age[58,213]; more recently it has been associated with autosomal dominant

non-progressive cerebellar ataxia, an early-onset disorder with features overlapping ataxic cerebral palsy (Table 1.3).[214]

1.16.5 SPTBN2

The fourth child, a female of non-consanguineous parents, was delivered at term by emergency caesarean following fetal distress. Pregnancy and the post-delivery period were normal. At eight months of age she had delayed development and by five she could stand independently and take a few steps; subsequently, she was diagnosed with ataxic cerebral play and moderate intellectual disability. Brian imaging showed mild cerebellar hypoplasia/'atrophy'. A *de novo* variant, p.R480W, in *SPTBN2* was identified. The protein encoded by *SPTBN2* is spectrin beta non-erythrocytic 2 or beta-III spectrin. The protein regulates the glutamate signalling pathway by stabilizing the glutamate transporter EAAT4 at the surface of the plasma membrane. Investigations into the effect of this mutation on voltage-gated sodium channel currents in cultured hippocampal neurons showed that in the presence of R480W, βIII spectrin sodium currents were lower than those for wild-type βIII spectrin, supporting the pathogenicity of this variant.[58] Mutations in this gene are associated with spinocerebellar ataxia 5.[215] Onset of spinocerebellar ataxia 5 is usually later in life; however, age of onset has been as young as 10 years.[216] Spinocerebellar ataxia, autosomal recessive 14, a neurologic disorder with delayed psychomotor development and severe early onset gait ataxia, is also associated with mutations in *SPTBN2* (Table 1.3).[217]

It is well recognised that for slowly progressive neurological disorders, cerebral palsy may be diagnosed at one point in time, with the need to change the diagnosis after a longer period of observation.[9] It is unclear at this time whether genetic susceptibility to

cerebral palsy as classically described, that is “non-progressive”, involves the same genes and genetic variations that are associated with very slowly progressive neurological disorders that cause the motor disorders associated with cerebral palsy.

Table 1.3 Summary of novel single gene mutations, in known (OMIM) genes, in cerebral palsy cases.

	Name	OMIM	Inheritance	Clinical features of cases diagnosed with cerebral palsy
GAD1	Glutamate decarboxylase 1	603513	AR	Spastic diplegia, spastic quadriplegia, microcephaly, ID[107]
AP4M1	Adaptor-related protein complex 4, μ1 subunit	612936	AR	Progressive spasticity, ID, microcephaly[206,210]
AP4E1	Adaptor-related protein complex 4, ε1 subunit	613744	AR	Progressive spasticity, ID, microcephaly[205]
AP4B1	Adaptor-related protein complex 4, β1 subunit	614066	AR	Progressive spasticity, ID, microcephaly[209]
AP4S1	Adaptor-related protein complex 4, σ1 subunit	614067	AR	Progressive spasticity, ID, microcephaly[208]
ADD3	Adducin 3 (Gamma)	601568	AR	Spastic diplegia, spastic quadriplegia, microcephaly, ID[203]
KCNC3	Voltage-Gated Potassium Channel Subunit Kv3.3	605259	AD	Ataxia, mild ID[58]
ITPR1	Inositol 1,4,5-Trisphosphate Receptor, Type 1	606658 117360	AD	Ataxia, moderate ID[58]
SPTBN2	Spectrin, beta, non-erythrocytic 2	600224 615386	AD AR	Ataxia, moderate ID[58]

Abbreviations: OMIM – Online Mendelian Inheritance in Man, AR – autosomal recessive, AD – autosomal dominant, ID – intellectual disability

1.17 Whole-exome sequencing vs whole-genome sequencing

Currently WES is the platform routinely used for common and rare variant detection. Until recently the sequencing requirements and costs of WES, which are largely dependent on capture methods, have been significantly less compared to whole-genome sequencing (WGS). However, as the costs of WGS are reducing, it is becoming an attractive alternative approach and there is considerable debate whether it will become the platform of choice for genome sequencing in both research and diagnostic settings.

The main advantages of WGS include the detection of noncoding variants,[218] more comprehensive coverage of exomes,[168] and more accurate assessment of indels, structural variants and CNVs.[219,220] Preparation of DNA templates for WGS is also much simpler and hybridisation independent.

In a recent comparative study of exome and genome sequencing technologies, WES was comparable to WGS on an exome-wide scale if sequencing coverage was two to three times higher.[174] Another study compared WES and WGS platforms for SNV detection in the exomes of six unrelated individuals. Coverage depth was much more uniform for WGS compared to WES. WES identified 26 variants missed by WGS and WGS identified 655 variants not seen by WES, of which a considerable proportion were predicted to be damaging. In addition, WES failed to adequately cover 380 genes, including 49 genes associated with Mendelian disorders. [221]

WGS offers a more uniform coverage across the genome compared to WES, in particular allowing detection of non-coding variants, the functional impact of which is becoming easier to interpret as a result of the systematic annotation of functional non-coding elements.[221,222] Although WGS sequencing costs are reducing, the increased

requirements and time needed for data storage and analysis significantly add to the costs.[168] WGS identifies approximately 4,000,000 variants per individual compared to 20,000 variants per individual for WES.[223] Filtering to a few potentially causal variants has proven, and remains, a challenge for WES but at present the implications of a missense mutation in the protein-coding regions are better understood. The many variants identified by WGS that are intronic involve uncharted regions of the genome.[196] The vast number of sequence variants that result from WGS compared to WES is likely to create a much greater challenge.

It is predicted that WGS will eventually take the lead. However, WES effectively interrogates the vast majority of the best understood portion of the genome, the protein-coding regions. WES is more cost and time efficient with regards to sequencing and computational requirements, making it more conducive for large-scale sequencing studies where resources are limited.

1.18 Summary

Cerebral palsy, recognised as the most common physical disability in childhood, is a complex disorder that is heterogeneous in its severity, aetiology and comorbidities. Empiric recurrence risk data, sibling (including twin) recurrence data, family studies suggesting Mendelian inheritance, gene mapping studies and the identification of CNVs and single gene mutations provide strong evidence for genetic involvement in cerebral palsy causation; however, for the majority of cases the genetic involvement is unknown. The advent of advanced genotyping technologies, in particular massively parallel sequencing, is a promising way forward in identifying the cause or contribution of genetic mutations in many cases of cerebral palsy. An unbiased WES approach will assist in identifying potential pathogenic variants. The success of genetic research in disorders that overlap with cerebral palsy, including intellectual disability, autism and epilepsy, provide an excellent model for cerebral palsy genetic research. Similar success in cerebral palsy research would dramatically increase our understanding of cerebral palsy causation. This will assist in not only identifying potentially pathogenic genes for cerebral palsy but in identifying cases which may have been misdiagnosed as cerebral palsy or where the diagnosis is not yet clear, resulting in potential for treatment based on a specific diagnosis or prevention through reproductive options.

1.19 Hypotheses and Aims

1.19.1 Hypotheses

1) Cerebral palsy is genetically heterogeneous and associated with many different rare variants either *de novo* or inherited.
2) Massively parallel sequencing provides an effective strategy to identify cerebral palsy mutations.

1.19.2 Aims

1) Identify rare genetic variants in cerebral palsy cases using *WES*.
2) Correlate results with the clinical features in order to understand the specific determinants associated with cerebral palsy.

1.20 References

1. Smithers-Sheedy, H., Badawi, N., Blair, E. *et al.* What constitutes cerebral palsy in the twenty-first century? *Dev Med Child Neurol* **56**, 323-8 (2014).
2. Colver, A., Fairhurst, C. & Pharoah, P.O. Cerebral palsy. *Lancet* **383**, 1240-9 (2014).
3. Little, W.J. Lectures on the deformity of the human frame. *Lancet* **1**, 318-320 (1843).
4. Freud, S. Les diplegies cerebrales infantiles. *1893* **1**, 177-182. (In French) (1893).
5. Naeye, R.L., Peters, E.C., Bartholomew, M. & Landis, J.R. Origins of cerebral palsy. *Am J Dis Child* **143**, 1154-61 (1989).
6. Bax, M.C. Terminology and classification of cerebral palsy. *Dev Med Child Neurol* **6**, 295-7 (1964).
7. Mutch, L., Alberman, E., Hagberg, B., Kodama, K. & Perat, M.V. Cerebral palsy epidemiology: Where are we now and where are we going? *Dev Med Child Neurol* **34**, 547-51 (1992).
8. Rosenbaum, P., Paneth, N., Leviton, A. *et al.* A report: The definition and classification of cerebral palsy april 2006. *Dev Med Child Neurol Suppl* **109**, 8-14 (2007).
9. Badawi, N., Watson, L., Petterson, B. *et al.* What constitutes cerebral palsy? *Dev Med Child Neurol* **40**, 520-527 (1998).
10. Moreno-De-Luca, A., Ledbetter, D.H. & Martin, C.L. Genomic insights into the causes and classification of the cerebral palsies. *Lancet Neurol* **11**, 283-92 (2012).

11. Lee, R.W., Poretti, A., Cohen, J.S. *et al.* A diagnostic approach for cerebral palsy in the genomic era. *Neuromolecular Med* (2014).

12. Lynex, C.N., Carr, I.M., Leek, J.P. *et al.* Homozygosity for a missense mutation in the 67 kda isoform of glutamate decarboxylase in a family with autosomal recessive spastic cerebral palsy: Parallels with stiff-person syndrome and other movement disorders. *BMC Neurol* **4**, 20 (2004).

13. Andersen, G.L., Irgens, L.M., Haagaas, I. *et al.* Cerebral palsy in norway: Prevalence, subtypes and severity. *Eur J Paediatr Neurol* **12**, 4-13 (2008).

14. Palisano, R., Rosenbaum, P., Walter, S. *et al.* Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Dev Med Child Neurol* **39**, 214-23 (1997).

15. Ashwal, S., Russman, B.S., Blasco, P.A. *et al.* Practice parameter: Diagnostic assessment of the child with cerebral palsy: Report of the quality standards subcommittee of the american academy of neurology and the practice committee of the child neurology society. *Neurol* **62**, 851-63 (2004).

16. Korzeniewski, S.J., Birbeck, G., DeLano, M.C., Potchen, M.J. & Paneth, N. A systematic review of neuroimaging for cerebral palsy. *J Child Neurol* **23**, 216-27 (2008).

17. Krageloh-Mann, I. & Horber, V. The role of magnetic resonance imaging in elucidating the pathogenesis of cerebral palsy: A systematic review. *Dev Med Child Neurol* **49**, 144-51 (2007).

18. Reid, S.M., Dagia, C.D., Ditchfield, M.R., Carlin, J.B. & Reddihough, D.S. Population-based studies of brain imaging patterns in cerebral palsy. *Dev Med Child Neurol* **56**, 222-32 (2014).

19. Accardo, P.J. & Hoon, A.H., Jr. The challenge of cerebral palsy classification: The elgan study. *J Pediatr* **153**, 451-2 (2008).

20. Bodensteiner, J.B. & Johnsen, S.D. Magnetic resonance imaging (mri) findings in children surviving extremely premature delivery and extremely low birthweight with cerebral palsy. *J Child Neurol* **21**, 743-7 (2006).

21. Leonard, J.M., Cozens, A.L., Reid, S.M. *et al.* Should children with cerebral palsy and normal imaging undergo testing for inherited metabolic disorders? *Dev Med Child Neurol* **53**, 226-32 (2011).

22. Stanley, F.J. & Watson, L. Trends in perinatal mortality and cerebral palsy in western australia, 1967 to 1985. *BMJ* **304**, 1658-63 (1992).

23. Hagberg, B., Hagberg, G. & Olow, I. The changing panorama of cerebral palsy in sweden. Vi. Prevalence and origin during the birth year period 1983-1986. *Acta Paediatr* **82**, 387-93 (1993).

24. Perlman, J.M. Intrapartum hypoxic-ischemic cerebral injury and subsequent cerebral palsy: Medicolegal issues. *Pediatrics* **99**, 851-9 (1997).

25. Centers for disease control and prevention. Economic costs associated with mental retardation, cerebral palsy, hearing loss and vision impairment - united states. *Morb Mortal Wkly Rep 2004;* **53**, 57-59 (2003).

26. Yeargin-Allsopp, M., Van Naarden Braun, K., Doernberg, N.S. *et al.* Prevalence of cerebral palsy in 8-year-old children in three areas of the united states in 2002: A multisite collaboration. *Pediatrics* **121**, 547-54 (2008).

27. Tosi, L.L., Maher, N., Moore, D.W., Goldstein, M. & Aisen, M.L. Adults with cerebral palsy: A workshop to define the challenges of treating and preventing secondary musculoskeletal and neuromuscular complications in this rapidly growing population. *Dev Med Child Neurol* **51 Suppl 4**, 2-11 (2009).

28. Australian cerebral palsy register report, birth years 1993-2006. (Australian Cerebral Palsy Register, 2013).

29. MacLennan, A. A template for defining a causal relation between acute intrapartum events and cerebral palsy: International consensus statement. *BMJ* **319**, 1054-9 (1999).

30. Access economics.,the economic impact of cerebral palsy in australia in 2007. in *Report for Cerebral Palsy Australia* (2008).

31. Murphy, D.J., Sellers, S., MacKenzie, I.Z., Yudkin, P.L. & Johnson, A.M. Case-control study of antenatal and intrapartum risk factors for cerebral palsy in very preterm singleton babies. *Lancet* **346**, 1449-54 (1995).

32. Stanley F., B.E., Alberman E. Cerebral palsies: Epidemiology and causal pathways. **151**(2000).

33. Tronnes, H., Wilcox, A.J., Lie, R.T., Markestad, T. & Moster, D. Risk of cerebral palsy in relation to pregnancy disorders and preterm birth: A national cohort study. *Dev Med Child Neurol* **56**, 779-85 (2014).

34. Platt, M.J., Cans, C., Johnson, A. *et al.* Trends in cerebral palsy among infants of very low birthweight (<1500 g) or born prematurely (<32 weeks) in 16 european centres: A database study. *Lancet* **369**, 43-50 (2007).

35. Gardosi, J., Mongelli, M., Wilcox, M. & Chang, A. An adjustable fetal weight standard. *Ultrasound Obstet Gynecol* **6**, 168-74 (1995).

36. Jacobsson, B., Ahlin, K., Francis, A. *et al.* Cerebral palsy and restricted growth status at birth: Population-based case-control study. *BJOG* **115**, 1250-5 (2008).

37. O'Callaghan, M.E., MacLennan, A., Gibson, C. *et al.* Epidemiolgic associations with cerebral palsy. *Obstet Gynecol* **118**, 576-582 (2011).

38. Hemminki, K., Li, X., Sundquist, K. & Sundquist, J. High familial risks for cerebral palsy implicate partial heritable aetiology. *Paediatr Perinat Epidemiol* **21**, 235-41 (2007).

39. Pharoah, P.O. Risk of cerebral palsy in multiple pregnancies. *Clin Perinatol* **33**, 301-13 (2006).

40. Petterson, B., Nelson, K.B., Watson, L. & Stanley, F. Twins, triplets, and cerebral palsy in births in western australia in the 1980s. *BMJ* **307**, 1239-43 (1993).

41. Pharoah, P.O., Price, T.S. & Plomin, R. Cerebral palsy in twins: A national study. *Arch Dis Child Fetal Neonatal Ed* **87**, F122-4 (2002).

42. Scher, A.I., Petterson, B., Blair, E. *et al.* The risk of mortality or cerebral palsy in twins: A collaborative population-based study. *Pediatr Res* **52**, 671-81 (2002).

43. Jarvis, S., Glinianaia, S.V., Arnaud, C. *et al.* Case gender and severity in cerebral palsy varies with intrauterine growth. *Arch Dis Child* **90**, 474-9 (2005).

44. Johnston, M.V. & Hagberg, H. Sex and the pathogenesis of cerebral palsy. *Dev Med Child Neurol* **49**, 74-8 (2007).

45. Jarvis, S., Glinianaia, S.V., Arnaud, C. *et al.* Case gender and severity in cerebral palsy varies with intrauterine growth. *Arch Dis Child* **90**, 474-9 (2005).

46. Grether, J.K. & Nelson, K.B. Maternal infection and cerebral palsy in infants of normal birth weight. *JAMA* **278**, 207-11 (1997).

47. McMichael, G., Maclennan, A., Gibson, C. *et al.* Cytomegalovirus and epstein-barr virus may be associated with some cases of cerebral palsy. *J Matern Fetal Neonatal Med* (2012).

48. Nelson, K.B. Causative factors in cerebral palsy. *Clin Obstet Gynecol* **51**, 749-62 (2008).

49. Wu, Y.W. & Colford, J.M., Jr. Chorioamnionitis as a risk factor for cerebral palsy: A meta-analysis. *JAMA* **284**, 1417-24 (2000).

50. Yoon, B.H., Romero, R., Park, J.S. *et al.* The relationship among inflammatory lesions of the umbilical cord (funisitis), umbilical cord plasma interleukin 6 concentration, amniotic fluid infection, and neonatal sepsis. *Am J Obstet Gynecol* **183**, 1124-9 (2000).

51. Windle, W.F. Neurological and psychological deficits from asphyxia neonatorum. *Public Health Rep* **72**, 646-50 (1957).

52. O'Brien, J.R., Usher, R.H. & Maughan, G.B. Causes of birth asphyxia and trauma. *Can Med Assoc J* **94**, 1077-85 (1966).

53. Nelson, K.B. & Ellenberg, J.H. Antecedents of cerebral palsy. Multivariate analysis of risk. *N Engl J Med* **315**, 81-6 (1986).

54. Blair, E. & Stanley, F.J. Intrapartum asphyxia: A rare cause of cerebral palsy. *J Pediatr* **112**, 515-9 (1988).

55. Strijbis, E.M., Oudman, I., van Essen, P. & MacLennan, A.H. Cerebral palsy and the application of the international criteria for acute intrapartum hypoxia. *Obstet Gynecol* **107**, 1357-65 (2006).

56. Fletcher, N.A. & Foley, J. Parental age, genetic mutation, and cerebral palsy. *J Med Genet* **30**, 44-6 (1993).

57. Davies, M.J., Moore, V.M., Willson, K.J. *et al.* Reproductive technologies and the risk of birth defects. *N Engl J Med* **366**, 1803-13 (2012).

58. Parolin Schnekenberg, R., Perkins, E.M., Miller, J.W. *et al.* De novo point mutations in patients diagnosed with ataxic cerebral palsy. *Brain* (2015).

59. Rahbar, M.H., Samms-Vaughan, M., Loveland, K.A. *et al.* Maternal and paternal age are jointly associated with childhood autism in jamaica. *J Autism Dev Disord* (2012).

60. O'Roak, B.J., Vives, L., Girirajan, S. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* **485**, 246-50 (2012).

61. Lidegaard, O., Pinborg, A. & Andersen, A.N. Imprinting diseases and ivf: Danish national ivf cohort study. *Hum Reprod* **20**, 950-4 (2005).

62. Hvidtjorn, D., Grove, J., Schendel, D.E. *et al.* Cerebral palsy among children born after in vitro fertilization: The role of preterm delivery--a population-based, cohort study. *Pediatrics* **118**, 475-82 (2006).

63. Petterson, B., Stanley, F. & Henderson, D. Cerebral palsy in multiple births in western australia: Genetic aspects. *Am J Med Genet* **37**, 346-51 (1990).

64. al-Rajeh, S., Bademosi, O., Awada, A. *et al.* Cerebral palsy in saudi arabia: A case-control study of risk factors. *Dev Med Child Neurol* **33**, 1048-52 (1991).

65. Odding, E., Roebroeck, M.E. & Stam, H.J. The epidemiology of cerebral palsy: Incidence, impairments and risk factors. *Disabil Rehabil* **28**, 183-91 (2006).

66. Blumel, J., Evans, E.B. & Eggers, G.W. Hereditary cerebral palsy; a preliminary report. *J Pediat* **50**, 454-8 (1957).

67. Paneth, N. Birth and the origins of cerebral palsy. *N Engl J Med* **315**, 124-6 (1986).

68. Amor, D.J., Craig, J.E., Delatycki, M.B. & Reddihough, D. Genetic factors in athetoid cerebral palsy. *J Child Neurol* **16**, 793-7 (2001).

69. Palmer, L., Petterson, B., Blair, E. & Burton, P. Family patterns of gestational age at delivery and growth in utero in moderate and severe cerebral palsy. *Dev Med Child Neurol* **36**, 1108-19 (1994).

70. Van Naarden Braun, K., Autry, A. & Boyle, C. A population-based study of the recurrence of developmental disabilities--metropolitan atlanta developmental disabilities surveillance program, 1991-94. *Paediatr Perinat Epidemiol* **19**, 69-79 (2005).

71. Sinha, G., Corry, P., Subesinghe, D., Wild, J. & Levene, M.I. Prevalence and type of cerebral palsy in a british ethnic community: The role of consanguinity. *Dev Med Child Neurol* **39**, 259-62 (1997).

72. Blair, E., Al Asedy, F., Badawi, N. & Bower, C. Is cerebral palsy associated with birth defects other than cerebral defects? *Dev Med Child Neurol* **49**, 252-8 (2007).

73. Garne, E., Dolk, H., Krageloh-Mann, I. *et al.* Cerebral palsy and congenital malformations. *Eur J Paediatr Neurol* **12**, 82-8 (2008).

74. Croen, L.A., Grether, J.K., Curry, C.J. & Nelson, K.B. Congenital abnormalities among children with cerebral palsy: More evidence for prenatal antecedents. *J Pediatr* **138**, 804-10 (2001).

75. Rankin, J., Cans, C., Garne, E. *et al.* Congenital anomalies in children with cerebral palsy: A population-based record linkage study. *Dev Med Child Neurol* **52**, 345-51 (2010).

76. Pakula, A.T., Van Naarden Braun, K. & Yeargin-Allsopp, M. Cerebral palsy: Classification and epidemiology. *Phys Med Rehabil Clin N Am* **20**, 425-52 (2009).

77. Vissers, L.E., de Ligt, J., Gilissen, C. *et al.* A de novo paradigm for mental retardation. *Nat Genet* **42**, 1109-12 (2010).

78. Cooper, G., Coe, B., Girirajan, S. *et al.* A copy number variation morbidity map of develomental delay. *Nat Genet* (2011).

79. Corbett, M.A., Schwake, M., Bahlo, M. *et al.* A mutation in the golgi qb-snare gene gosr2 causes progressive myoclonus epilepsy with early ataxia. *Am J Hum Genet* **88**, 657-63 (2011).

80. Miles, J.H. Autism spectrum disorders--a genetics review. *Genet Med* **13**, 278-94 (2011).

81. Mulley, J.C. & Mefford, H.C. Epilepsy and the new cytogenetics. *Epilepsia* **52**, 423-32 (2011).

82. Najmabadi, H., Hu, H., Garshasbi, M. *et al.* Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature* **478**, 57-63 (2011).

83. O'Roak, B.J., Deriziotis, P., Lee, C. *et al.* Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat Genet* **43**, 585-9 (2011).

84. Wild, N.J. & Rosenbloom, L. Familial cerebral palsy associated with normal intelligence. *Postgrad Med J* **62**, 827-30 (1986).

85. Kvistad, P.H., Dahl, A. & Skre, H. Autosomal recessive non-progressive ataxia with an early childhood debut. *Acta Neurol Scand* **71**, 295-302 (1985).

86. Imamura, S., Tachi, N. & Oya, K. Dominantly inherited early-onset non-progressive cerebellar ataxia syndrome. *Brain Dev* **15**, 372-6 (1993).

87. Young, I.D., Moore, J.R. & Tripp, J.H. Sex-linked recessive congenital ataxia. *J Neurol Neurosurg Psychiatry* **50**, 1230-2 (1987).

88. Neville, B. Congenital dopa-responsive disorders: A diagnostic and therapeutic challenge to the cerebral palsies? *Dev Med Child Neurol* **49**, 85 (2007).

89. Bainbridge, M.N., Wiszniewski, W., Murdock, D.R. *et al.* Whole-genome sequencing for optimized patient management. *Sci Transl Med* **3**, 87re3 (2011).

90. Lee, J.H., Ki, C.S., Kim, D.S. *et al.* Dopa-responsive dystonia with a novel initiation codon mutation in the gch1 gene misdiagnosed as cerebral palsy. *J Korean Med Sci* **26**, 1244-6 (2011).

91. Fink, J.K. Hereditary spastic paraplegia: Clinico-pathologic features and emerging molecular mechanisms. *Acta Neuropathol* **126**, 307-28 (2013).

92. Novarino, G., Fenstermaker, A.G., Zaki, M.S. *et al.* Exome sequencing links corticospinal motor neuron disease to common neurodegenerative disorders. *Science* **343**, 506-11 (2014).

93. Harding, A.E. Hereditary spastic paraplegias. *Semin in Neurol* **13**, 333-6 (1993).

94. Shoukier, M., Neesen, J., Sauter, S.M. *et al.* Expansion of mutation spectrum, determination of mutation cluster regions and predictive structural classification of spast mutations in hereditary spastic paraplegia. *Eur J Hum Genet* **17**, 187-94 (2009).

95. Hubner, C.A., Orth, U., Senning, A. *et al.* Seventeen novel plp1 mutations in patients with pelizaeus-merzbacher disease. *Hum Mutat* **25**, 321-2 (2005).

96. Cailloux, F., Gauthier-Barichard, F., Mimault, C. *et al.* Genotype-phenotype correlation in inherited brain myelination defects due to proteolipid protein gene mutations. Clinical european network on brain dysmyelinating disease. *Eur J Hum Genet* **8**, 837-45 (2000).

97. Yamasaki, M., Thompson, P. & Lemmon, V. Crash syndrome: Mutations in l1cam correlate with severity of the disease. *Neuropediatrics* **28**, 175-8 (1997).

98. Poirier, K., Saillour, Y., Fourniol, F. *et al.* Expanding the spectrum of tuba1a-related cortical dysgenesis to polymicrogyria. *Eur J Hum Genet* **21**, 381-5 (2013).

99. Romani, M., Micalizzi, A. & Valente, E.M. Joubert syndrome: Congenital cerebellar ataxia with the molar tooth. *The Lancet. Neurol* **12**, 894-905 (2013).

100. Karakas, E., Cullu, N., Karakas, O. *et al.* Joubert syndrome: The clinical and radiological findings. *J Pak Med Assoc* **64**, 91-4 (2014).

101. Bird, L.M. Angelman syndrome: Review of clinical and molecular aspects. *Appl Clin Genet* **7**, 93-104 (2014).

102. Sadikovic, B., Fernandes, P., Zhang, V.W. *et al.* Mutation update for ube3a variants in angelman syndrome. *Human Mut* (2014).

103. Rosenbaum, P., Paneth, N., Leviton, A. *et al.* A report: The definition and classification of cerebral palsy april 2006. *Dev Med Child Neurol* **109**, 8-14 (2007).

104. Kerem, B., Rommens, J.M., Buchanan, J.A. *et al.* Identification of the cystic fibrosis gene: Genetic analysis. *Science* **245**, 1073-80 (1989).

105. Harville, H.M., Held, S., Diaz-Font, A. *et al.* Identification of 11 novel mutations in eight bbs genes by high-resolution homozygosity mapping. *J Med Genet* **47**, 262-7 (2010).

106. Lander, E.S. & Botstein, D. Homozygosity mapping: A way to map human recessive traits with the DNA of inbred children. *Science* **236**, 1567-70 (1987).

107. McHale, D.P., Mitchell, S., Bundey, S. *et al.* A gene for autosomal recessive symmetrical spastic cerebral palsy maps to chromosome 2q24-25. *Am J Hum Genet* **64**, 526-32 (1999).

108. Mitchell, S. & Bundey, S. Symmetry of neurological signs in pakistani patients with probable inherited spastic cerebral palsy. *Clin Genet* **51**, 7-14 (1997).

109. Lerer, I., Sagi, M., Meiner, V. *et al.* Deletion of the ankrd15 gene at 9p24.3 causes parent-of-origin-dependent inheritance of familial cerebral palsy. *Hum Mol Genet* **14**, 3911-20 (2005).

110. McHale, D.P., Jackson, A.P., Campbell *et al.* A gene for ataxic cerebral palsy maps to chromosome 9p12-q12. *Eur J Hum Genet* **8**, 267-72 (2000).

111. Zhu, M. & Zhao, S. Candidate gene identification approach: Progress and challenges. *Int J Biol Sci* **3**, 420-7 (2007).

112. Nelson, K.B., Dambrosia, J.M., Grether, J.K. & Phillips, T.M. Neonatal cytokines and coagulation factors in children with cerebral palsy. *Ann Neurol* **44**, 665-75 (1998).

113. Gibson, C.S., MacLennan, A.H., Hague, W.M. *et al.* Associations between inherited thrombophilias, gestational age, and cerebral palsy. *Am J Obstet Gynecol* **193**, 1437 (2005).

114. Nelson, K.B., Dambrosia, J.M., Iovannisci, D.M. *et al.* Genetic polymorphisms and cerebral palsy in very preterm infants. *Pediatr Res* **57**, 494-9 (2005).

115. Gibson, C.S., MacLennan, A.H., Goldwater, P.N. *et al.* The association between inherited cytokine polymorphisms and cerebral palsy. *Am J Obstet Gynecol* **194**, 674 e1-11 (2006).

116. Gibson, C.S., MacLennan, A.H., Janssen, N.G. *et al.* Associations between fetal inherited thrombophilia and adverse pregnancy outcomes. *Am J Obstet Gynecol* **194**, 947 e1-10 (2006).

117. Gibson, C.S., Maclennan, A.H., Dekker, G.A. *et al.* Candidate genes and cerebral palsy: A population-based study. *Pediatrics* **122**, 1079-85 (2008).

118. McMichael, G.L., Gibson, C.S., Goldwater, P.N. *et al.* Association between apolipoprotein e genotype and cerebral palsy is not confirmed in a caucasian population. *Hum Genet* **124**, 411-6 (2008).

119. Blackman, J.A. Apolipoprotein e and functional motor severity in cerebral palsy. *J Pediatr Rehab Med*, 67-74 (2009).

120. Braga, L.W., Borigato, E.V., Speck-Martins, C.E. *et al.* Apolipoprotein e genotype and cerebral palsy. *Dev Med Child Neurol* (2009).

121. O'Callaghan, M.E., MacLennan, A.H., Haan, E.A. & Dekker, G. The genomic basis of cerebral palsy: A huge systematic literature review. *Hum Genet* **126**, 149-72 (2009).

122. Wu, Y.W., Croen, L.A., Vanderwerf, A., Gelfand, A.A. & Torres, A.R. Candidate genes and risk for cp: A population-based study. *Pediatr Res* **70**, 642-6 (2011).

123. O'Callaghan, M.E., Maclennan, A.H., Gibson, C.S. *et al.* Fetal and maternal candidate single nucleotide polymorphism associations with cerebral palsy: A case-control study. *Pediatrics* **129**, e414-23 (2012).

124. Zhang, J., Chiodini, R., Badr, A. & Zhang, G. The impact of next-generation sequencing on genomics. *J Genet Genomics* **38**, 95-109 (2011).

125. Hindorff, L.A., Sethupathy, P., Junkins, H.A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA* **106**, 9362-7 (2009).

126. Bergen, S.E. & Petryshen, T.L. Genome-wide association studies of schizophrenia: Does bigger lead to better results? *Curr Opin Psychiatry* **25**, 76-82 (2012).

127. Manolio, T.A., Collins, F.S., Cox, N.J. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747-53 (2009).

128. Feuk, L., Carson, A.R. & Scherer, S.W. Structural variation in the human genome. *Nat Rev Genet* **7**, 85-97 (2006).

129. Girirajan, S., Campbell, C.D. & Eichler, E.E. Human copy number variation and complex genetic disease. *Annu Rev Genet* **45**, 203-26 (2011).

130. Lupski, J.R. Structural variation in the human genome. *N Engl J Med* **356**, 1169-71 (2007).

131. Itsara, A., Cooper, G.M., Baker, C. *et al.* Population analysis of large copy number variants and hotspots of human genetic disease. *Am J Hum Genet* **84**, 148-61 (2009).

132. Sharp, A.J., Mefford, H.C., Li, K. *et al.* A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. *Nat Genet* **40**, 322-8 (2008).

133. Girirajan, S., Rosenfeld, J.A., Cooper, G.M. *et al.* A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nat Genet* **42**, 203-9 (2010).

134. van Bon, B.W., Balciuniene, J., Fruhman, G. *et al.* The phenotype of recurrent 10q22q23 deletions and duplications. *Eur J Hum Genet* **19**, 400-8 (2011).

135. Coe, B.P., Witherspoon, K., Rosenfeld, J.A. *et al.* Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nat Genet* **46**, 1063-71 (2014).

136. Cooper, G.M., Coe, B.P., Girirajan, S. *et al.* Corrigendum: A copy number variation morbidity map of developmental delay. *Nat Genet* **46**, 1040 (2014).

137. Ben-Shachar, S., Lanpher, B., German, J.R. *et al.* Microdeletion 15q13.3: A locus with incomplete penetrance for autism, mental retardation, and psychiatric disorders. *J Med Genet* **46**, 382-8 (2009).

138. Miller, D.T., Shen, Y., Weiss, L.A. *et al.* Microdeletion/duplication at 15q13.2q13.3 among individuals with features of autism and other neuropsychiatric disorders. *J Med Genet* **46**, 242-8 (2009).

139. Moreno-De-Luca, D., Mulle, J.G., Kaminsky, E.B. *et al.* Deletion 17q12 is a recurrent copy number variant that confers high risk of autism and schizophrenia. *Am J Hum Genet* **87**, 618-30 (2010).

140. Dibbens, L.M., Mullen, S., Helbig, I. *et al.* Familial and sporadic 15q13.3 microdeletions in idiopathic generalized epilepsy: Precedent for disorders with complex inheritance. *Hum Mol Genet* **18**, 3626-31 (2009).

141. Helbig, I., Mefford, H.C., Sharp, A.J. *et al.* 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. *Nat Genet* **41**, 160-2 (2009).

142. de Kovel, C.G., Trucks, H., Helbig, I. *et al.* Recurrent microdeletions at 15q11.2 and 16p13.11 predispose to idiopathic generalized epilepsies. *Brain* **133**, 23-32 (2010).

143. Stefansson, H., Rujescu, D., Cichon, S. *et al.* Large recurrent microdeletions associated with schizophrenia. *Nature* **455**, 232-6 (2008).

144. Mefford, H.C., Sharp, A.J., Baker, C. *et al.* Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N Engl J Med* **359**, 1685-99 (2008).

145. Szatmari, P., Paterson, A.D., Zwaigenbaum, L. *et al.* Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet* **39**, 319-28 (2007).

146. Pagnamenta, A.T., Wing, K., Sadighi Akha, E. *et al.* A 15q13.3 microdeletion segregating with autism. *Eur J Hum Genet* **17**, 687-92 (2009).

147. Ballif, B.C., Theisen, A., Coppinger, J. *et al.* Expanding the clinical phenotype of the 3q29 microdeletion syndrome and characterization of the reciprocal microduplication. *Mol Cytogenet* **1**, 8 (2008).

148. Clayton-Smith, J., Giblin, C., Smith, R.A., Dunn, C. & Willatt, L. Familial 3q29 microdeletion syndrome providing further evidence of involvement of the 3q29 region in bipolar disorder. *Clin Dysmorphol* **19**, 128-32 (2010).

149. Mulle, J.G., Dodd, A.F., McGrath, J.A. *et al.* Microdeletions of 3q29 confer high risk for schizophrenia. *Am J Hum Genet* **87**, 229-36 (2010).

150. Mefford, H.C., Batshaw, M.L. & Hoffman, E.P. Genomics, intellectual disability, and autism. *N Engl J Med* **366**, 733-43 (2012).

151. Zhao, X., Leotta, A., Kustanovich, V. *et al.* A unified genetic theory for sporadic and inherited autism. *Proc Natl Acad Sci U S A* **104**, 12831-6 (2007).

152. McMichael, G., Girirajan, S., Moreno-De-Luca, A. *et al.* Rare copy number variation in cerebral palsy. *Eur J Hum Genet* **22**, 40-5 (2014).

153. Gai, X., Xie, H.M., Perin, J.C. *et al.* Rare structural variation of synapse and neurotransmission genes in autism. *Mol Psychiatry* (2011).

154. Medina, M., Marinescu, R.C., Overhauser, J. & Kosik, K.S. Hemizygosity of delta-catenin (ctnnd2) is associated with severe mental retardation in cri-du-chat syndrome. *Genomics* **63**, 157-64 (2000).

155. Vrijenhoek, T., Buizer-Voskamp, J.E., van der Stelt, I. *et al.* Recurrent cnvs disrupt three candidate genes in schizophrenia patients. *Am J Hum Genet* **83**, 504-10 (2008).

156. Harvard, C., Malenfant, P., Koochek, M. *et al.* A variant cri du chat phenotype and autism spectrum disorder in a subject with de novo cryptic microdeletions involving 5p15.2 and 3p24.3-25 detected using whole genomic array cgh. *Clin Genet* **67**, 341-51 (2005).

157. Trimborn, M., Bell, S.M., Felix, C. *et al.* Mutations in microcephalin cause aberrant regulation of chromosome condensation. *Am J Hum Genet* **75**, 261-6 (2004).

158. Garshasbi, M., Motazacker, M.M., Kahrizi, K. *et al.* Snp array-based homozygosity mapping reveals mcph1 deletion in family with autosomal recessive mental retardation and mild microcephaly. *Hum Genet* **118**, 708-15 (2006).

159. McMichael, G., Girirajan, S., Moreno-De-Luca, A. *et al.* Rare copy number variation in cerebral palsy. *Eur J Hum Genet* (2013).

160. Seeger, M., Thierse, H.J., Lange, H. *et al.* Antigen-independent suppression of the ige immune response to bee venom phospholipase a2 by maternally derived monoclonal igg antibodies. *Eur J Immunol* **28**, 2124-30 (1998).

161. Girirajan, S., Vlangos, C.N., Szomju, B.B. *et al.* Genotype-phenotype correlation in smith-magenis syndrome: Evidence that multiple genes in 17p11.2 contribute to the clinical spectrum. *Genet Med* **8**, 417-27 (2006).

162. Potocki, L., Chen, K.S., Park, S.S. *et al.* Molecular mechanism for duplication 17p11.2- the homologous recombination reciprocal of the smith-magenis microdeletion. *Nat Genet* **24**, 84-7 (2000).

163. Molina, J., Carmona-Mora, P., Chrast, J. *et al.* Abnormal social behaviors and altered gene expression rates in a mouse model for potocki-lupski syndrome. *Hum Mol Genet* **17**, 2486-95 (2008).

164. Segel, R., Ben-Pazi, H., Zeligson, S. *et al.* Copy number variations in cryptogenic cerebral palsy. *Neurology* **84**, 1660-8 (2015).

165. Oskoui, M., Gazzellone, M.J., Thiruvahindrapuram, B. *et al.* Clinically relevant copy number variations detected in cerebral palsy. *Nat Commun* **6**, 7949 (2015).

166. Bueter, W., Dammann, O. & Leviton, A. Endoplasmic reticulum stress, inflammation, and perinatal brain damage. *Pediatr Res* **66**, 487-94 (2009).

167. van Dijk, E.L., Auger, H., Jaszczyszyn, Y. & Thermes, C. Ten years of next-generation sequencing technology. *Trends in Genet : TIG* **30**, 418-26 (2014).

168. Warr, A., Robert, C., Hume, D. *et al.* Exome sequencing: Current and future perspectives. *G3 (Bethesda)* **5**, 1543-50 (2015).

169. Albert, T.J., Molla, M.N., Muzny, D.M. *et al.* Direct selection of human genomic loci by microarray hybridization. *Nat Methods* **4**, 903-5 (2007).

170. Gnirke, A., Melnikov, A., Maguire, J. *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* **27**, 182-9 (2009).

171. Bainbridge, M.N., Wang, M., Burgess, D.L. *et al.* Whole exome capture in solution with 3 gbp of data. *Genome Biol* **11**, R62 (2010).

172. Sulonen, A.M., Ellonen, P., Almusa, H. *et al.* Comparison of solution-based exome capture methods for next generation sequencing. *Genome Biol* **12**, R94 (2011).

173. Parla, J.S., Iossifov, I., Grabill, I. *et al.* A comparative analysis of exome capture. *Genome Biol* **12**, R97 (2011).

174. Lelieveld, S.H., Spielmann, M., Mundlos, S., Veltman, J.A. & Gilissen, C. Comparison of exome and genome sequencing technologies for the complete capture of protein-coding regions. *Hum Mutat* **36**, 815-22 (2015).

175. Clark, M.J., Chen, R., Lam, H.Y. *et al.* Performance comparison of exome DNA sequencing technologies. *Nat Biotechnol* **29**, 908-14 (2011).

176. Hsu, F., Kent, W.J., Clawson, H. *et al.* The ucsc known genes. *Bioinformatics* **22**, 1036-46 (2006).

177. Pruitt, K.D., Tatusova, T., Klimke, W. & Maglott, D.R. Ncbi reference sequences: Current status, policy and new initiatives. *Nucleic Acids Res* **37**, D32-6 (2009).

178. Flicek, P., Amode, M.R., Barrell, D. *et al.* Ensembl 2011. *Nucleic Acids Res* **39**, D800-6 (2011).

179. Chilamakuri, C.S., Lorenz, S., Madoui, M.A. *et al.* Performance comparison of four exome capture systems for deep sequencing. *BMC Genomics* **15**, 449 (2014).

180. Bodi, K., Perera, A.G., Adams, P.S. *et al.* Comparison of commercially available target enrichment methods for next-generation sequencing. *J Biomol Tech* **24**, 73-86 (2013).

181. Bilguvar, K., Ozturk, A.K., Louvi, A. *et al.* Whole-exome sequencing identifies recessive wdr62 mutations in severe brain malformations. *Nature* **467**, 207-10.

182. Choi, M., Scholl, U.I., Ji, W. *et al.* Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci USA* **106**, 19096-101 (2009).

183. Corbett, M.A., Bahlo, M., Jolly, L. *et al.* A focal epilepsy and intellectual disability syndrome is due to a mutation in tbc1d24. *Am J Hum Genet* **87**, 371-5 (2010).

184. Wang, J.L., Yang, X., Xia, K. *et al.* Tgm6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing. *Brain* **133**, 3510-8 (2010).

185. Bamshad, M.J., Ng, S.B., Bigham, A.W. *et al.* Exome sequencing as a tool for mendelian disease gene discovery. *Nat Rev. Genet* **12**, 745-55 (2011).

186. Topper, S., Ober, C. & Das, S. Exome sequencing and the genetics of intellectual disability. *Clin Genet* **80**, 117-26 (2011).

187. de Ligt, J., Willemsen, M.H., van Bon, B.W. *et al.* Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med* **367**, 1921-9 (2012).

188. Rauch, A., Wieczorek, D., Graf, E. *et al.* Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: An exome sequencing study. *Lancet* **380**, 1674-82 (2012).

189. Stephan J. Sanders, Michael T. Murtha, Abha R. Gupta *et al.* De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nat Letter* (2012).

190. Rosti, R.O., Sadek, A.A., Vaux, K.K. & Gleeson, J.G. The genetic landscape of autism spectrum disorders. *Dev Med Child Neurol* **56**, 12-8 (2014).

191. Girard, S.L., Gauthier, J., Noreau, A. *et al.* Increased exonic de novo mutation rate in individuals with schizophrenia. *Nat Genet* **43**, 860-3 (2011).

192. Xu, B., Roos, J.L., Dexheimer, P. *et al.* Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nat Genet* **43**, 864-8 (2011).

193. Allen, A.S., Berkovic, S.F., Cossette, P. *et al.* De novo mutations in epileptic encephalopathies. *Nature* **501**, 217-21 (2013).

194. Veeramah, K.R., Johnstone, L., Karafet, T.M. *et al.* Exome sequencing reveals new causal mutations in children with epileptic encephalopathies. *Epilepsia* **54**, 1270-81 (2013).

195. Dyment, D.A., Tetreault, M., Beaulieu, C.L. *et al.* Whole-exome sequencing broadens the phenotypic spectrum of rare pediatric epilepsy: A retrospective study. *Clinical Genet* (2014).

196. MacArthur, D.G., Manolio, T.A., Dimmock, D.P. *et al.* Guidelines for investigating causality of sequence variants in human disease. *Nature* **508**, 469-76 (2014).

197. Hoischen, A., Krumm, N. & Eichler, E.E. Prioritization of neurodevelopmental disease genes by discovery of new mutations. *Nat Neurosci* **17**, 764-72 (2014).

198. Gilissen, C., Hoischen, A., Brunner, H.G. & Veltman, J.A. Disease gene identification strategies for exome sequencing. *Eur J Hum Genet* **20**, 490-7 (2012).

199. Goldstein, D.B., Allen, A., Keebler, J. *et al.* Sequencing studies in human genetics: Design and interpretation. *Nat Rev Genet* **14**, 460-70 (2013).

200. Sanders, S.J., Murtha, M.T., Gupta, A.R. *et al.* De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237-41 (2012).

201. Rajab, A., Yoo, S.Y., Abdulgalil, A. *et al.* An autosomal recessive form of spastic cerebral palsy (cp) with microcephaly and mental retardation. *Am J Med Genet A* **140**, 1504-10 (2006).

202. Fong, C.Y., Mumford, A.D., Likeman, M.J. & Jardine, P.E. Cerebral palsy in siblings caused by compound heterozygous mutations in the gene encoding protein c. *Dev Med Child Neurol* **52**, 489-93 (2010).

203. Kruer, M.C., Jepperson, T., Dutta, S. *et al.* Mutations in gamma adducin are associated with inherited cerebral palsy. *Annals of Neurol* **74**, 805-14 (2013).

204. Hirst, J., Irving, C. & Borner, G.H. Adaptor protein complexes ap-4 and ap-5: New players in endosomal trafficking and progressive spastic paraplegia. *Traffic* **14**, 153-64 (2013).

205. Moreno-De-Luca, A., Helmers, S.L., Mao, H. *et al.* Adaptor protein complex-4 (ap-4) deficiency causes a novel autosomal recessive cerebral palsy syndrome with microcephaly and intellectual disability. *J Med Genet* **48**, 141-4 (2011).

206. Verkerk, A.J., Schot, R., Dumee, B. *et al.* Mutation in the ap4m1 gene provides a model for neuroaxonal injury in cerebral palsy. *Am J Hum Genet* **85**, 40-52 (2009).

207. Moreno-De-Luca, A., Helmers, S.L., Mao, H. *et al.* Adaptor protein complex-4 (ap-4) deficiency causes a novel autosomal recessive cerebral palsy syndrome with microcephaly and intellectual disability. *J Med Genet* (2010).

208. Abou Jamra, R., Philippe, O., Raas-Rothschild, A. *et al.* Adaptor protein complex 4 deficiency causes severe autosomal-recessive intellectual disability, progressive spastic paraplegia, shy character, and short stature. *Am J Hum Genet* **88**, 788-95 (2011).

209. Bauer, P., Leshinsky-Silver, E., Blumkin, L. *et al.* Mutation in the ap4b1 gene cause hereditary spastic paraplegia type 47 (spg47). *Neurogenetics* **13**, 73-6 (2012).

210. Jameel, M., Klar, J., Tariq, M. *et al.* A novel ap4m1 mutation in autosomal recessive cerebral palsy syndrome and clinical expansion of ap-4 deficiency. *BMC Med Genet* **15**, 133 (2014).

211. Waters, M.F., Minassian, N.A., Stevanin, G. *et al.* Mutations in voltage-gated potassium channel kcnc3 cause degenerative and developmental central nervous system phenotypes. *Nat Genet* **38**, 447-51 (2006).

212. Figueroa, K.P., Waters, M.F., Garibyan, V. *et al.* Frequency of kcnc3 DNA variants as causes of spinocerebellar ataxia 13 (sca13). *PLoS One* **6**, e17811 (2011).

213. van de Leemput, J., Chandran, J., Knight, M.A. *et al.* Deletion at itpr1 underlies ataxia in mice and spinocerebellar ataxia 15 in humans. *PLoS Genet* **3**, e108 (2007).

214. Huang, L., Chardon, J.W., Carter, M.T. *et al.* Missense mutations in itpr1 cause autosomal dominant congenital nonprogressive spinocerebellar ataxia. *Orphanet J Rare Dis* **7**, 67 (2012).

215. Jacob, F.D., Ho, E.S., Martinez-Ojeda, M., Darras, B.T. & Khwaja, O.S. Case of infantile onset spinocerebellar ataxia type 5. *J Child Neurol* **28**, 1292-5 (2013).

216. Bauer, P., Schols, L. & Riess, O. Spectrin mutations in spinocerebellar ataxia (sca). *Bioessays* **28**, 785-7 (2006).

217. Lise, S., Clarkson, Y., Perkins, E. *et al.* Recessive mutations in sptbn2 implicate beta-iii spectrin in both cognitive and motor development. *PLoS Genet* **8**, e1003074 (2012).

218. Spielmann, M. & Klopocki, E. Cnvs of noncoding cis-regulatory elements in human disease. *Curr Opin Genet Dev* **23**, 249-56 (2013).

219. Fang, H., Wu, Y., Narzisi, G. *et al.* Reducing indel calling errors in whole genome and exome sequencing data. *Genome Med* **6**, 89 (2014).

220. Gilissen, C., Hehir-Kwa, J.Y., Thung, D.T. *et al.* Genome sequencing identifies major causes of severe intellectual disability. *Nature* **511**, 344-7 (2014).

221. Belkadi, A., Bolze, A., Itan, Y. *et al.* Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc Natl Acad Sci U S A* **112**, 5473-8 (2015).

222. Consortium, T.E.P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).

223. Shendure, J. Next-generation human genetics. *Genome Biol* **12**, 408 (2011).

Chapter 2
Clinical cohort

2.1 Sporadic cerebral palsy cases (n = 183)

This project was a novel study using whole-exome sequencing, to identify rare, *de novo* and inherited, disease-causing mutations in 183 sporadic cases of cerebral palsy. Only offspring of Caucasian parents (defined for the purpose of this study as 'of European ancestry') were included, in order to have as homogenous a study population as possible. This was done so that the significance of genetic variants identified by the study could be assessed against the best available estimates of the background population frequency of each variant. All participants were Australian-born to allow linkage to birth records. While consanguinity is not common in the Caucasian population (as defined), it was an exclusion criterion to avoid biasing the population towards autosomal recessive inheritance. Parental DNAs were available for more than half of the cases. The sample was a convenience sample of volunteer families recruited throughout Australia between 2010 and 2012 as part of a research study conducted at The University of Adelaide, South Australia, Australia. No case with a confirmed diagnosis of cerebral palsy was excluded. The cohort's overall phenotypic, clinical and demographic characteristics were similar to population distributions described in the 2013 report of the Australian Cerebral Palsy Register www.cerebralpalsy.org.au/wp-content/uploads/2013/04/ACPR-report_Web_2013.pdf. Clinical data including parental age, gender, details of pregnancy and birth, gestational age, Apgar scores, birth weight, birth head circumference and neonatal events were collected through cerebral palsy registers, midwife case-notes and a maternal study questionnaire. Also collected were type and severity of cerebral palsy and any accompanying comorbidities such as intellectual disability (ID), autism and epilepsy. Growth restriction was assessed using a customized birth weight centile program.[2]

2.2 Brain imaging details

Brain imaging reports were available for 112 of the 183 sporadic cases (61%). Results from brain imaging were categorized into seven diagnostic groupings: 1) intraventricular haemorrhage (IVH; 21); 2) white matter damage (periventricular leukomalacia (PVL) and/or porencephalic cysts (n = 46); 3) developmental malformations (n = 20); 4) perinatal cerebral infarction (n = 15); 5) no specific abnormality detected (n = 7); 6) intrauterine infection e.g. cytomegalovirus (n = 2); and 7) intrauterine cerebral tumour (n = 1).

2.3 Cerebral palsy cohort clinical characteristics

Clinical characteristics for this study included:

1. Type of cerebral palsy
2. Gestational age
3. Birth weight
4. Gender
5. Accompanying co-morbidities i.e. intellectual disability, autism and epilepsy

2.4 Cerebral palsy subtype

Four classifications described the different motor impairments of cerebral palsy:[3]

1. Spastic cerebral palsy, classified by the region of the body affected i.e. hemiplegia (one side), diplegia (lower limbs), triplegia (three limbs) and quadriplegia (four limbs). Individuals with spastic cerebral palsy are hypertonic (increased muscle tone).

2. Ataxic cerebral palsy is characterised by hypotonia (low muscle tone) and movements that are characterised by instability and clumsiness and often appear jerky or disorganised resulting in lack of coordination and balance.
3. Dyskinetic cerebral palsy is characterised by varying involuntary movements including athetosis (slow writhing movements), chorea (irregular unpredictable movements) and dystonic (twisting and repetitive movements).
4. Mixed type (hypertonic and hypotonic).

Amongst spastic cerebral palsy cases, hemiplegia is the most prevalent type in term babies and diplegia is the most prevalent type in preterm infants.[4] Table 2.1 and Figure 2.1 illustrate the distribution of cerebral palsy subtypes for the 183 sporadic cases in our cerebral palsy cohort.

Table 2.1 Type of cerebral palsy in 183 sporadic cases.

Type of cerebral palsy	**Number**	**%**
Hemiplegia	76	42
Diplegia	65	36
Triplegia	4	2
Quadriplegia	31	17
Dyskinesia	7	3
Ataxia	0	0
Total	183	100

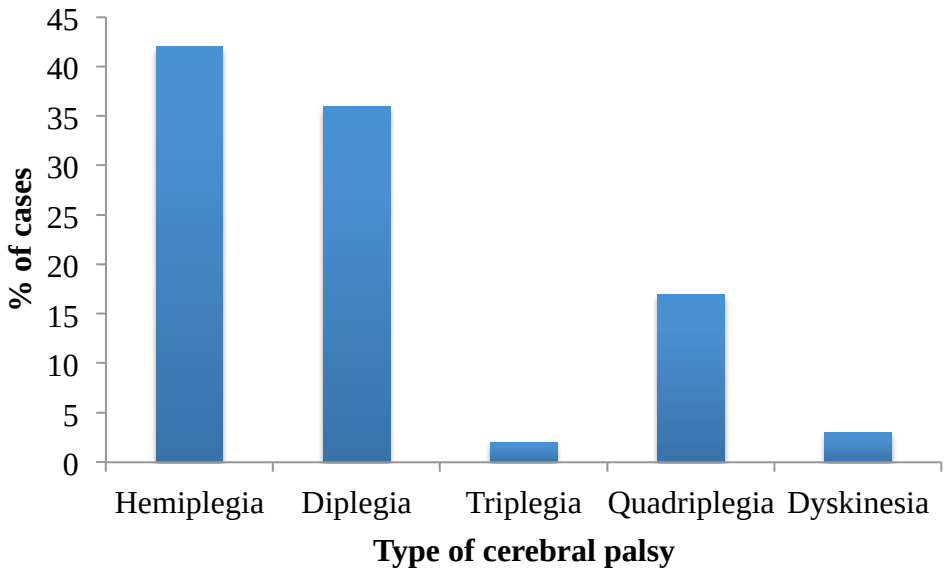


Figure 2.1 Type of cerebral palsy in 183 sporadic cases.

2.5 Gestational age

The risk of cerebral palsy increases with decreasing gestational age. Very preterm infants born 24 – 26 weeks and 27 – 32 weeks represent 20% and 4% of diagnosed cases of cerebral palsy, respectively; with term infants accounting for half of all diagnosed cases.[5] A more recent study based on our Australian cohort found that compared to term infants, babies born between 32 – 36 weeks were at an increased risk for cerebral palsy (OR 5.0) and babies born before 32 weeks were at even greater risk (OR 59.2).[6] There appears to be a direct correlation between increased rates of cerebral palsy and decline in mortality rate amongst preterm babies.[7] Approximately half of the cohort in this thesis (49%) were born at term, 17% were born between 32 and 36 weeks and the remaining 32% fell into the very premature category of less than 32 weeks. In comparison, 90% of all South Australian births in 2009 were term born[8] (Table 2.2 and Figure 2.2).

Table 2.2 Gestational age distribution of 183 sporadic cases with cerebral palsy compared with South Australian (SA) pregnancy outcome data.[7]			
Gestation (weeks)	**Number**	**CP cohort (%)**	**SA (%)**
≥ 37	89	49	91
32 - 36	31	17	7.2
28 - 31	37	20	0.9
< 28	23	12	0.9
Unknown	3	2	0
Total	183	100	100

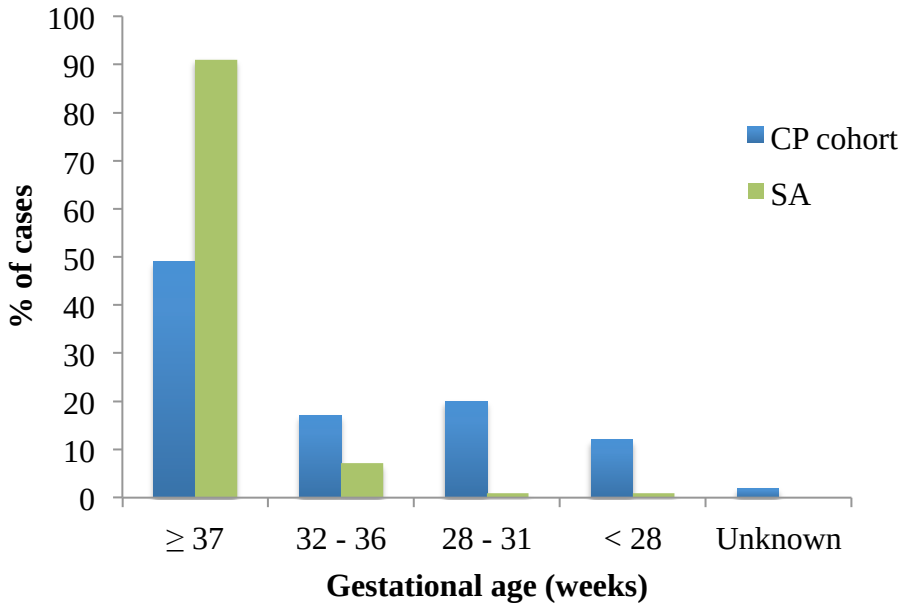


Figure 2.2 Gestational age distributions of 183 sporadic cases with cerebral palsy compared with South Australian (SA) pregnancy outcome data.[7]

2.6 Birth weight distribution

The risk of cerebral palsy increases amongst low birth weight babies compared with normal birth weight babies. Babies delivered less than 2500 grams account for approximately 50% of cerebral palsy cases.[9] Over the last few decades the incidence of low birth weight babies amongst cerebral palsy cases has risen from 33% to 50% and, as with gestational age, this is largely due to increased survival rate.[9] The mean birth weight for our cerebral palsy cohort was 2414 grams (500 – 4999 grams). In comparison the mean weight of all South Australian births in 2009 was 3328 grams (550 – 5930 grams)[8] (Table 2.3 and Figure 2.3).

Table 2.3 Birth weight distribution in 183 sporadic cases with cerebral palsy compared with South Australian (SA) pregnancy outcome data.[7]			
Birth weight (grams)	**Number**	**CP cohort (%)**	**SA (%)**
<1500	52	29	1.6
1500 - 2499	30	16	5.7
2500 - 3499	61	33	52
3500 - 4499	33	18	39.2
≥4500	2	1	1.5
Unknown	5	3	0
Total	183	100	100

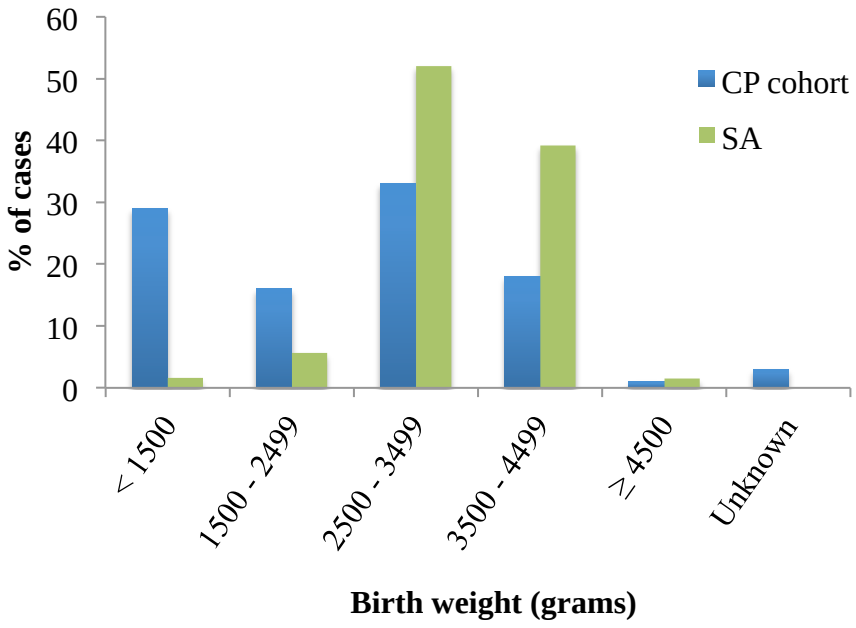


Figure 2.3 Birth weight distribution of 183 sporadic cases with cerebral palsy compared with South Australian (SA) pregnancy outcome data[7].

2.7 Comorbidities distribution

Cerebral palsy is often accompanied by other neurodevelopmental disorders including intellectual disability (ID), autism and epilepsy. Spastic quadriplegic cerebral palsy has the highest reported incidence of additional comorbidities including ID.[10] Epilepsy is reported to be more prevalent in cases with quadriplegic and hemiplegic cerebral palsy.[11] Nearly half of our cohort (51%) had one or more of these comorbidities (Table 2.4 and Figure 2.4).

Table 2.4 Distribution of co-morbidities in 183 sporadic cases with cerebral palsy.

Co-morbidities	**Number**	**%**
Intellectual disability	35	19
Autism	12	7
Epilepsy	46	25
Total	93	51

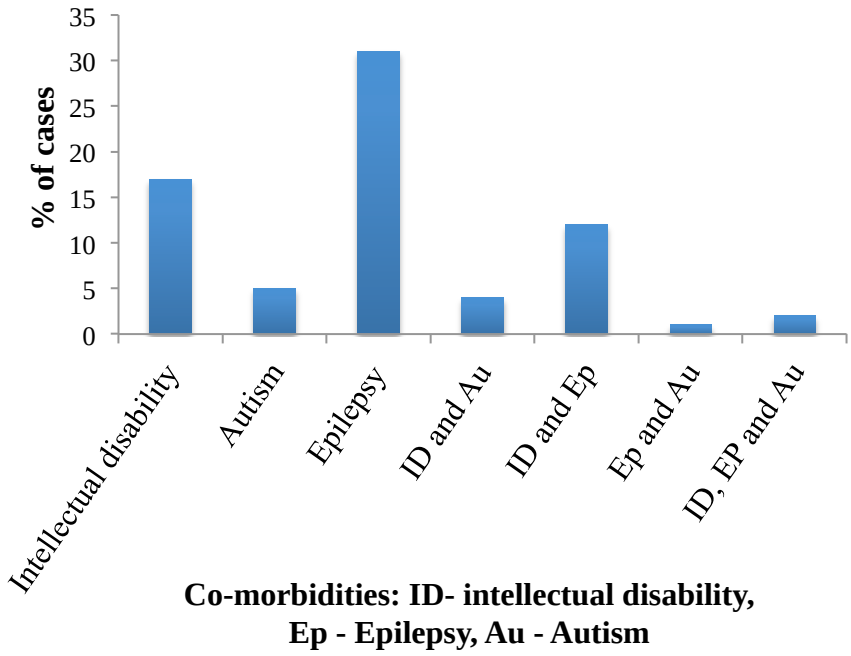


Figure 2.4 Distribution of co-morbidities in 183 sporadic cases with cerebral palsy.

2.8 Gender distribution

There is a consistently higher prevalence for cerebral palsy in the male population, with a male:female ratio of 1.3:1 – 1.5:1.[12] Table 2.5 and Figure 2.5 illustrate the gender distribution of our cerebral palsy cohort, showing a slightly higher prevalence for cerebral palsy in the males compared to females (male:female ratio 1.2:1).

Table 2.5 Gender distributions in 183 sporadic cases with cerebral palsy.

Gender	Number	%
Male	99	54
Female	84	46
Total	183	100

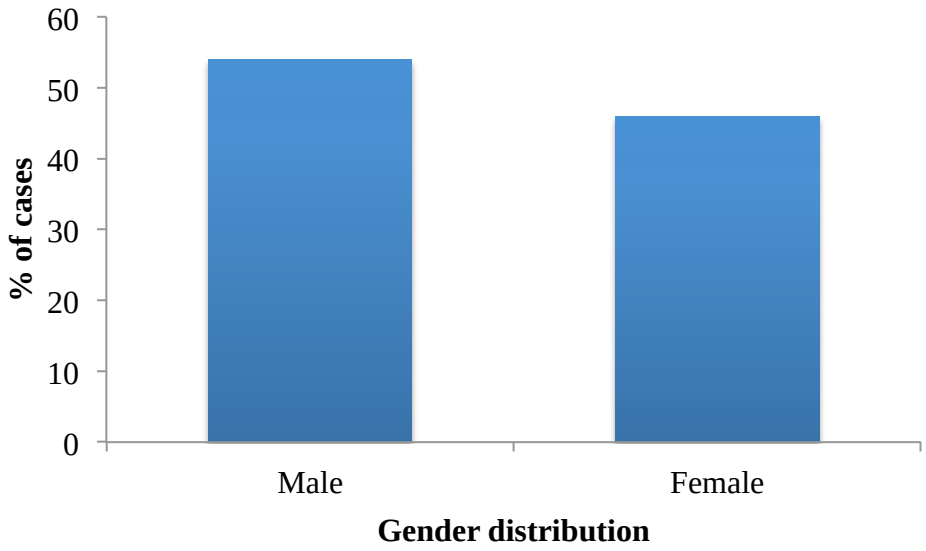


Figure 2.5 Gender distributions in 183 sporadic cases with cerebral palsy.

2.9 Familial cerebral palsy cases

As well as the 183 sporadic cerebral palsy cases we reported on the results from WES in three families with more than one individual with a diagnosis of cerebral palsy.

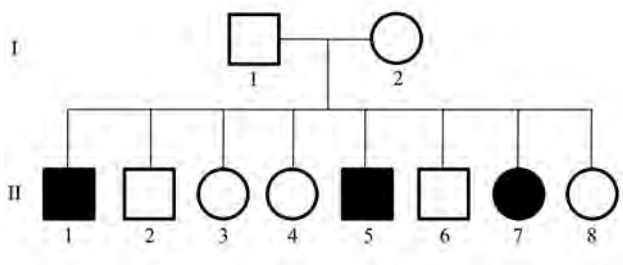


Figure 2.6 Family 1

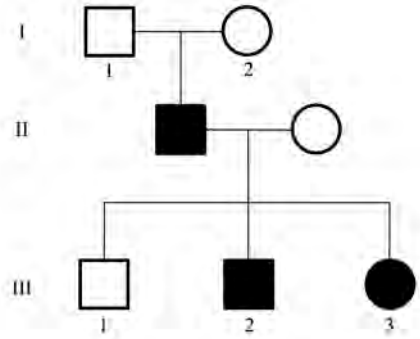


Figure 2.7 Family 2

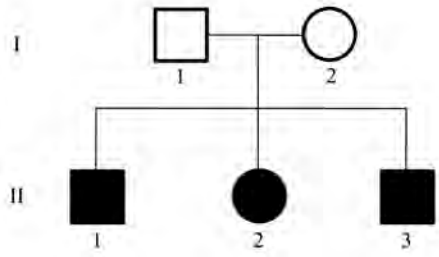


Figure 2.8 Family 3

Figures 2.6, 2.7 and 2.8 Overview of the pedigree details for three families with a diagnosis of cerebral palsy.

Table 2.6 Summary of the clinical characteristics for three families with more than one individual with a diagnosis of cerebral palsy.

Family	Cases	Sex	Gest weeks	BW grams	Clinical characteristics
1 (Fig 2.6)	II-1	M	40	3890	Quadriplegic cerebral palsy, hypotonia, mild dystonia, intellectual disability, seizures, cognitive decline with age and equinovarus feet
	II-5	M	40	4080	Diplegic cerebral palsy, hypotonia, mild dystonia, intellectual disability, cognitive decline with age, epilepsy and equinovarus feet
2 (Fig 2.7)	II-1	M	40	NA	Mild ataxia, mild dysarthria
	III-2	M	39	2970	Developmental delay with ataxia and choreoathetoid movements
	III-3	F	40	3440	Ataxic gait, mild choreoathetoid movements
3 (Fig 2.8)	II-1	M	40	NA	Autism spectrum disorder
	II-2	F	40	3460	Mild spasticity, developmental delay, autism spectrum disorder and cognitive decline
	II-3	M	40	4235	Mild spasticity, autism spectrum disorder

Abbreviations: Gest – gestation, BW – birth weight, M – male, F – female, NA – not available

2.10 References

1. Sinha, G., Corry, P., Subesinghe, D., Wild, J. & Levene, M.I. Prevalence and type of cerebral palsy in a british ethnic community: The role of consanguinity. *Dev Med Child Neurol* **39**, 259-62 (1997).
2. Gardosi, J. Customised assessment of fetal growth potential: Implications for perinatal care. *Arch Dis Child Fetal Neonatal Ed* **97**, F314-7 (2012).
3. Andersen, G.L., Irgens, L.M., Haagaas, I. *et al.* Cerebral palsy in norway: Prevalence, subtypes and severity. *Eur J Paediatr Neurol* **12**, 4-13 (2008).
4. Hagberg, B., Hagberg, G. & Olow, I. The changing panorama of cerebral palsy in sweden. Vi. Prevalence and origin during the birth year period 1983-1986. *Acta Paediatr* **82**, 387-93 (1993).
5. Ancel, P.Y., Livinec, F., Larroque, B. *et al.* Cerebral palsy among very preterm children in relation to gestational age and neonatal ultrasound abnormalities: The epipage cohort study. *Pediatrics* **117**, 828-35 (2006).
6. O'Callaghan, M.E., MacLennan, A., Gibson, C. *et al.* Epidemiolgic associations with cerebral palsy. *Obstet Gynecol* **118**, 576-582 (2011).
7. Vincer, M.J., Allen, A.C., Joseph, K.S. *et al.* Increasing prevalence of cerebral palsy among very preterm infants: A population-based study. *Pediatrics* **118**, e1621-6 (2006).
8. Chan A, S.W., Scott J, Nguyen A-M, Sage L. Pregnancy outcome in south australia. (Pregnancy Outcome Unit, SA Health, Governmento of South Australia, 2011, 2011).
9. Odding, E., Roebroeck, M.E. & Stam, H.J. The epidemiology of cerebral palsy: Incidence, impairments and risk factors. *Disabil Rehabil* **28**, 183-91 (2006).

10. Gabis, L.V., Tsubary, N.M., Leon, O., Ashkenasi, A. & Shefer, S. Assessment of abilities and comorbidities in children with cerebral palsy. *J Child Neurol* **30**, 1640-5 (2015).

11. Pruitt, D.W. & Tsai, T. Common medical comorbidities associated with cerebral palsy. *Physical Med and Rehab Clinics of North America* **20**, 453-67 (2009).

12. Stanley F., B.E., Alberman B. How common are the cerebral palsies? *In: Stanley F, Blair E, Alberman B, editors. Cerebral Palsies: Epidemiology &Causal Pathways. London (UK): Mac Keith Press,* p.22-48 (2000).

Chapter 3
Materials and methods

3.1 Recruitment of cerebral palsy families

Recruitment for this study coincided with the development of a National lymphocyte cell line and blood derived DNA biobank of cerebral palsy family trios (mother, father and affected child). Participating South Australian cerebral palsy families from our existing database (99% agreed to be recontacted for future research) were invited to take part in this current study. New families were also sought through ongoing recruitment and contact with the Paediatric Rehabilitation Department at the Women's and Children's Hospital, Adelaide, South Australia and the Paediatric Rehabilitation Department of the Princess Margaret Hospital, Perth, Western Australia. All participants were of Caucasian decent with a confirmed specialist diagnosis of cerebral palsy, and where possible, confirmation by the South Australian State Cerebral Palsy Register.

Invited families were sent a study kit containing a cover letter informing them about the study, University of Adelaide study information sheet and consent form, Genetic Repositories Australia participant information sheet and consent form, a detailed maternal questionnaire requesting pregnancy, labour and delivery details of their participating child. Participants wishing to be involved in this research were asked to sign and return the enclosed consent forms in the provided pre-paid envelope which allowed the research coordinator to phone to discuss the project further, answer any questions the family may have and arrange a suitable and convenient time for blood collection.

Approval from the Women's and Children's Health Network Human Research Ethics Committee, Adelaide, Australia, and Child and Adolescent Health Service Ethics

Committee at the Princess Margaret Hospital, Perth, Australia and the Internal Review Board at Baylor College of Medicine, Houston, USA was obtained for this project. Written informed consent was obtained from participants or their parents.

3.2 Blood collection

Whole blood (11mL in total: 9mL in an acid citrate dextrose (ACD) collection tube and 2mL in an ethylenediaminetetraacetic acid (EDTA) collection tube) was collected from the affected child. 9mL of whole blood was collected from each parent in an EDTA collection tube. To minimize any inconvenience to the participating families, a number of ways to collect blood were offered. This included a pre-arranged home visit by a registered nurse, collection via a blood pathology centre or collection at SA Pathology at the Women's and Children's Hospital. A qualified phlebotomist experienced in blood collection from children collected all child and parental samples. A number of children with cerebral palsy have Botox therapy under general anaesthetic (GA). In these instances, visits were coordinated with the Paediatric Rehabilitation and Anaesthesiology Departments at the Women's and Children's Hospital, Adelaide and the Princess Margaret Hospital, Perth to collect the child's blood under GA.

3.3 Lymphocyte cell line development

Blood from individuals with cerebral palsy was collected in an ACD blood collection tube and stored at room temperature for no longer than 24 hours before transformation at Genetic Repositories Australia (GRA) a National Health and Medical Research Council supported facility that develops, stores and distributes de-identified DNA samples and lymphoblastoid cell lines (LCLs) to facilitate genetic medical research.

The technique applied by GRA to develop Epstein Barr Viral (EBV) transformed LCLs involves collection of fresh whole blood into 9mL ACD blood collection tubes. Fresh human B lymphocytes are isolated via density gradient separation, washed, counted and inoculated with EBV viral supernatant (B95-8 Monkey Marmoset cell line). Cells are then transferred to a 25 cm^2 tissue culture flask containing transformation cell culture media (RPMI 1640, supplemented with fetal bovine serum, L-glutamine and Phytohaemaglutinin MN) and cultured at 37℃ with 10% CO_2 in a humidity controlled incubator. Developing cultures are routinely checked microscopically for signs of microbial contamination and cell line transformation. Following transformation, cells are washed, counted and cryopreserved in cryovials at 1 x 10^7 in freeze down media containing 10% dimethyl sulfoxide, cooled at a controlled rate of -1℃/minute prior to transferring to a vapour phase liquid nitrogen tank for long term storage.

3.4 DNA extraction

DNA extraction from LCLs was performed at GRA on a fully automated large volume nucleic acid purification system (Qiagen Autopure LS, Stanford, CA, USA), which ensures high quality DNA from EBV transformed LCLs. Subsequently extracted DNA was normalised to 50 ng/µL on a fluorometer using the PicoGreen dsDNA quantification assay (Applied Biosytems, CA, USA).

DNA was extracted from 2mL of the affected child's blood using a QIAamp DNA Blood Mini Kit (Qiagen, Stanford, CA), following the manufacturer's instructions. DNA was eluted in 200 µL of the buffer provided (Buffer AE) and DNA concentration was determined by measuring absorbance at 260 nm using a UV spectrophotometer (CAT # ND-1000, Nanodrop).

Two different DNA extraction methods were used to extract DNA from 9 mL of the parental blood, 1) a fully automated large volume nucleic acid purification system (AGRF, SA, Australia) and 2) a QIAamp DNA Blood Maxi Kit (Qiagen, Stanford, CA, USA), following the manufacturer's instructions. DNA was eluted in buffer (Buffer AE) 1mL and 600 µL respectively and concentration for both methods was determined on a fluorometer using the PicoGreen dsDNA quantification assay (Applied Biosytems, CA, USA). Samples were normalised to 50 ng/µL.

3.5 Whole-exome sequencing

3.5.1 Illumina library construction

After determining DNA concentration and integrity, high molecular weight double strand genomic DNA samples were constructed into Illumina PairEnd precapture libraries according to the manufacturer's protocol (*Illumina Multiplexing SamplePrep Guide 1005361_D*) with modifications as described in the Baylor College of Medicine-Human Genome Sequencing Center (BCM-HGSC) protocol (https://hgsc.bcm.edu/sites/default/files/documents/Illumina_Barcoded_PairedEnd_Capture_Library_Preparation.pdf). 1 µg genomic DNA in 100 µL volume was sheared into fragments of approximately 300 base pairs in a Covaris plate with E210 system (Covaris, Inc. Woburn, MA). The setting was 10% Duty cycle, Intensity of 4,200 Cycles per Burst, for 120 seconds. Fragment size was evaluated using a 2.2 % Flash Gel DNA Cassette (Lonza, Cat. No.57023). Fragmented DNA was end-repaired in 90ul total reaction volume containing sheared DNA, 9 µL 10X buffer, 5 µL END Repair Enzyme Mix and H_2O (NEBNext End-Repair Module; Cat. No. E6050L) and then incubated at 20C for 30 minutes. A-tailing was performed in a total reaction volume of 60 µL containing end-repaired DNA, 6 µL 10X buffer, 3 µL Klenow Fragment (NEBNext dA-Tailing

Module; Cat. No. E6053L) and H_2O followed by incubation at 37C for 30 minutes. Illumina multiplex adapter ligation (NEBNext Quick Ligation Module Cat. No. E6056L) was performed in a total reaction volume of 90 µL containing 18 uL 5X buffer, 5ul ligase, 0.5 µL 100 µM adaptor and H_2O at room temperature for 30 minutes. After Ligation, PCR with Illumina PE 1.0 and modified barcode primers was performed in 170 µL reactions containing 85 2x Phusion High-Fidelity PCR master mix, adaptor ligated DNA, 1.75 µL of 50 µM each primer and H_2O. The standard thermocycling for PCR was 5' at 95°C for the initial denaturation followed by 6-10 cycles of 15 s at 95°C, 15 s at 60°C and 30 s at 72°C and a final extension for 5 min. at 72°C. Agencourt® XP® Beads (Beckman Coulter Genomics, Inc.; Cat. No. A63882) was used to purify DNA after each enzymatic reaction. After bead purification, PCR product quantification and size distribution was determined using the Caliper GX 1K/12K/High Sensitivity Assay Labchip (Hopkinton, MA, Cat. No. 760517).

3.5.2 Illumina exome capture

1 µg of Illumina paired-end pre-capture library DNA samples (six pre-capture libraries pooled together, approximately 166 ng/sample) was hybridised to a custom capture reagent designed at the HGSC (HGSC VCRome 2.1 design; 42 Mb NimbleGen)[1] and constructed at Roche NimbleGen, Madison WI, USA. Briefly, this reagent targets the coding regions of genes from CCDS, ENSEMBL, UCSC, GenCode, VEGA, and RefSeq. It also targets miRNA and snoRNA from UCSC, predicted miRNA binding sites, and 1000 Sanger-predicted miRNA. Hybridisation was conducted according to the manufacturer's protocol with minor revisions. Human COT1 DNA and full-length Illumina adaptor-specific blocking oligonucleotides were added into the hybridisation to block repetitive genomic sequences and the adaptor sequences. Hybridisation enhancing

oligos IHE1, IHE2 and IHE3 replaced oligos HE1.1 and HE2.1 and post-capture LM-PCR was performed using 14 cycles. After the final AMPure XP bead purification, quantity and size of capture libraries were analysed using the Agilent Bioanalyser 2100 DNA Chip 7500 (Agilent, Santa Clara, CA, USA). Efficiency of the capture was evaluated by performing a qPCR-based quality check on the built-in controls (qPCR SYBR Green assays, Applied Biosystems, CA, USA). Four standardized oligo sets, RUNX2, PRKG1, SMG1, and NLK, were employed as internal quality controls. The enrichment of the capture libraries was estimated to range from 7 to 9-fold of ΔCt over background.

3.5.3 Illumina sequencing

Illumina cBot cluster generation system with TruSeq PE Cluster Generation Kits (Cat. no. PE-401-3001) was used to prepare library templates for sequencing. Prior to clustering libraries were denatured with sodium hydroxide and diluted to 3-6 pM in hybridisation buffer in order to achieve a load density of ~800K clusters/mm^2. Each library pool was spiked with 2% phiX control library for run quality control prior to loading in a single lane of a HiSeq flow cell. The sample libraries underwent bridge amplification to form clonal clusters, followed by hybridisation with the sequencing primer. Sequencing runs were performed in paired-end mode using the Illumina HiSeq 2000 platform (Illumina, San Diego, CA, USA). Using the TruSeq SBS Kits (Cat. no. FC-401-3001), sequencing-by-synthesis reactions were extended for 101 cycles from each end, with an additional seven cycles for the index read.

The sequencing method describe above was used for the majority of our study cohort with the exception of two families (Family 1 and Family 2 described in **Chapters 5 and 6** respectively). For these families sequencing was performed at Axeq Technologies,

(Rockville, MD, USA) as a paid service. Samples were prepared using the TruSeq DNA Exome Enrichment Kit +UTRs, following the manufacturers instructions. Briefly 1 µg of genomic DNA was fragmented by nebulization. Fragmented DNA was end-repaired followed by A-tailing and ligation of Illumina adapters (Illumina, San Diego, CA, USA). Size selected products (350-400 bp) were PCR amplified and validated on an Agilent Bioanalyser (Agilent, Santa Clara, CA, USA). Following bridge amplification, sequencing was performed on the HiSeq2000 (Illumina, San Diego, CA, USA). Using TruSeq SBS Kits v3 (Illumina, San Diego, CA, USA), sequencing-by-synthesis reactions were extended for 101 cycles from each end, with an additional seven cycles for the index read. No further changes were made to the methodologies discussed in this chapter.

3.5.4 Read mapping and variant calling

Illumina data was aligned to the human reference genome (HG19) with Burrows Wheeler Aligner (BWA).[2] Variant qualities were recalibrated with Genome Analysis Toolkit (GATK).[3] Variants were called using ATLAS-SNP and the SAMtools program pileup. Reads were locally realigned at presumptive insertion or deletion events. Base qualities were empirically recalibrated and small insertions and deletions were locally realigned using the smith-waterman algorithm. Differences between the human reference and the sequence reads (variants) were identified.

Identified variants were initially annotated with the following information: Population and sub-population frequency from Thousand Genomes, and the Exome Variant Server (EVS) data set (http://evs.gs.washington.edu/EVS/); variant quality metrics; the effect of the mutation on the UCSC and RefSeq gene models; predicted deleteriousness of the

variant by PolyPhen2,[4] SIFT,[5] and Mutation Taster;[6] dbSNP; gene function, gene expression pattern, post translational modification and disease association; Online Mendelian Inheritance in Man (OMIM) and Human Genome Mutation Database (HGMD).

Identified variants were annotated based on their population frequency and predicted effect on the related genes. Variants which were rare (MAF < 0.1%) and had altered translation of the gene in some manner, were prioritized for further biological interpretation of their potential pathogenicity.

3.5.5 Data interrogation and analysis

Samples were analysed to identify *de novo* mutations (case-parent trios), autosomal recessive variants (homozygous or heterozygous) and X-linked variants (case-parent trios and duos), and overlapping variants identified in trios and duos and plausible genes for cerebral palsy (singletons). Variants were considered high-quality *de novo* in the child if there was no equivalent variant in either parent, and both parents had at least 10X coverage at the variant position. Variants were considered low-quality *de novo* if a low quality variant was present in one parent, or low-coverage *de novo* if the coverage in either parent was 10X or less. A gene was considered to contain a compound heterozygous variant if a single parent did not account for all the rare protein-changing variants in a gene. Rare variants (MAF < 0.1%) predicted to alter the gene product were validated by Sanger sequencing.

3.6 Validation of candidate variants for cerebral palsy

Standard Sanger sequencing was performed across two sites, the Human Genome Sequencing Center and The University of Adelaide, to validate potential variants.

3.6.1 Human Genome Sequencing Centre, Baylor College of Medicine

Primers for de novo variant sites were designed by an in-house, automated pipeline (primer sequences not provided). Sanger sequencing was performed using BigDye terminator chemistry 3.1 (Applied Biosytems, Foster City, CA, USA) and sequenced using an ABI 3730xl DNA analyser (Applied Biosytems, Foster City, CA, USA). Sequencing data was analysed using in-house software, SNP-D, and consed.

3.6.2 The University of Adelaide

Maternally inherited X-linked variants and a subset of *de novo* mutations were validated in-house. Primers incorporating the candidate variant were designed using the web resource Primer3 (v.0.4.0) (Table 3.1). Primers were manufactured by GeneWorks (Adelaide, SA, Australia) as a paid service.

Table 3.1 Primer sequences for validating candidate variants from whole blood (The University of Adelaide).

Primer Pair	Forward Sequence (5' - 3')	Reverse Sequence (5' - 3')	Size (bp)
AIM1L	GGTGTATCTGTCAGTTTGTCCCTCTG	CTCCTGTCCTCAAGCAATCCTCCT	796
ALK	GTGTGAATGTGGGTGGGTGT	CAGCCTGTAGCCAGACAATGGA	424
ALPK2	CCTCAGGAATTGACACAGCA	CAAGAGCCCCAGTCCTAGTG	206
ARMCX2	AGGTCATGGTGATTTGCTAAGG	GCAAACTTTTTCCGTTTGCTAT	326
ART1	GTGTGGCAGGGAGGCAAAGTT	AGAGGCTGAGGCAGGAGAATGG	515
ATP1B4	ATTGGCATCATGGGATTTACTC	CATCTGGCTGTGTCCTAATTGA	384
BRWD3	CATGATGCCTGTGAATTATGCT	AGAGAAGTCTGTGGGACTTTGC	371
ASPH	CCAAACTCTTCTCCACCCTACTCTCT	ACACATCTTACATTCACTCGCCTTCAT	474
CARS2	CTCCGCTGACCTGTCCATTGTG	GATTCTCCTGCCTGAGCCTCCT	522
CDC27	ACAGACAGCCTGAGACAGTTCT	ACTAAGAGCTGCTGGTCCTCCT	350
CDH26	GAGCAGATGAGCAGATGAGTGAGTC	GCCAAGTCCAGGAGCCTGAATATC	59
CD99L2	CTTGCAGGGTGGTTTACACTTT	CTGTCTGACCTCACCTGTCCTC	367
CNKSR2	TGCCAGTTACTTTCTTCTACTGT	AGCTCTCTACTGATGGATGGTG	286
CTDSPL	AATGTGCTTTCCTGCTTGTGCTTAT	CCTGTTGCCTGCTAGGAGAGTAG	578
CUL4B	TAACATAAATAGAGCACGTGACC	CACTGATGGTAATACCAGCAC	562
DENND1B	GTAGTCAGACTTTCTCAGGAGGCAAA	GCAGGTGGCAGGGAAGCAAA	689
DYNC1H1	GAGGTTAAGTCACAGAGTTTCCTGAAGA	ACAACATTCTTCACTCTGCCAACACA	616
EIF4E2	GTGGAGGTGGTAAGGATTGACACTG	CATGAACGAACTATCAAGGCACACTC	483
ELFN2	CCACACCTGCCTGACCTTCAC	CTGGTTGACCTTGTCCACCTCC	526
ENOX2	TTACTTGGATTTGGAGGGAAGA	AGCAGAGAGTGTGTGTTGGAAA	475
ENPP4	GGTTCTCTCCTCCTGTGGGCTA	GCCACATAGCAGCAGCACTTGA	529
FEZ1	AGGGTGCCATCCGTGACTTCTC	TGCTGCTCACCACCACTGCT	341
FGB	GCTAAAGATAAGGGAAGAAAGGCAGTT	TTAGGACAGTGGCTGACAGTAAGTG	487
GANC	CATGAACACCAAGCGACGAGAG	CATACTTTACTCACAAAGCAGTAGCAG	369
HUWE1	TCTCCAGCGAAGGTGAGTGATGG	GAGACAAACCCAGCCTTGCTTTCA	578
ICMT	TCTCATGTAAGAGGAGCAACAGGAA	CTTGATTCGGCATAAGGACAGACTG	607
IGSF1	ATACCCCATTGCACAGTCCT	CATCCCAGCAACCCTCTCA	225
IL1RAPL1	GAGCTGGAAACCAGACTTCG	ACGCATTTGTGAATGGTACG	390
IQSEC2	GTTGGACAAGCTAACCCTCAAC	TTATGTATGAGCCACCTTGCAC	332
IRX6	GCTTGGCACTGCTCGCTAG	CCTCCTCCTCCTCCTCTTCCT	329
KIAA0040	GAATCAGTGCCTTCTTCAGCTCTATCT	ACACCTCTGCTTCCTGCCTAACT	314

Table 3.1 Primer sequences for validating candidate variants from whole blood (The University of Adelaide).

KLKB1	CTTTGAGTAGTCAGTCAGTGGGAGAAG	GGACTGTGCTATCCGCTAATGTGAT	668
L1CAM 539	AAGGAGGCCGAAGTCATGAC	CAGGAGGAGAGTGTCAGCC	324
L1CAM 577	CATGACAGTGGGCATCACAG	CTGCTCTCTTCCCTTTGCCT	258
LTN1	GTAATCTATCAGAAGCAAGTCCAGAGG	GATTTGGTCACTGAATGATGCTGGTAG	367
MAOB	GGGATTTCTGTAACAGCTTAGCA	CTATCCAGTCACTCCAGTAAACA	500
MAP4K5	CAGCACAAGCACCACAGTTACC	AATCCACATCAAGAAGAGGACACAGT	553
MNT	GTGCCTCCTCGCCATCTTGTTG	GGTCCCTCGCAGCATGTTAATGAA	645
MTMR1	CCTTGAATTCAGACTTCCAGC	CGTAGTGACTCTCGCATGAC	330
MYH8	GGGTGGGTGTCTCAATCTCTCTAAATC	GCAGGAAGAGTGGATATTCTGGAAGG	330
NCOR2	GTTTATGACCTCTTCTGGGTGGACTC	GGACAGCAGGCTGCCTTACTTG	316
NR1I2	GCCCTGGTCTTCCTTCACTTCC	CTACCTCCAGATGCCACCCTCT	345
OFD1	GCTCCATGCTGGAGCTGAATATC	CAGATGTCTAACTCCATGCTGC	443
OTC	AGAATCGCTTGAACCTGGGAG	GGGACTGGTAGTAATGGAACG	385
PAK3 083	ACACCTTGGACCCTGAAGCTA	AGGAGAATGGCAGCACGATA	245
PAK3 718	GTCTGTGCCAAGCCATGATT	TGCCTCTCTTTGTAATTGTCCT	237
PAK3 772	TGGTTTCATCTCCTCCCACC	ACTGCACACACATGAACCTG	214
PCBP3	GGGACAGGCACTGGAGTGTGA	GCAGTGCCTCCACAGCAACAT	427
PTGFRN	CAGATTCCGTGCTTGTGGTGAAGG	GATTGGAGAGACTGGTTGCTGCTT	396
RAD21	GGTTCAGAACTGGCAGGGTAGATAG	CTGAGGTGGTGGTGGAGGCATA	368
RFX2	GCCACGTCCGAATTGCTCCAT	CGGCTGCCTCTTACTCACCATC	486
SCN2A 911	TGGCATTCTGCATGACATTT	ATTTCGTAGGTTGCCCATGA	241
SCN2A 865	CCTATTCTTAATAGTGGACCT	CAACGTGGTCGTAATGGGCTCA	537
SEMA4	TGTGGGTGGAGTGGATGGAAAGG	ACGCCGACCAATCAGAATGAGAGT	582
SH3PXD2B	GAGTGGCTGGTGGTACATTCAGATT	CCGGGCTGGAGGGATGTGTTTG	513
SLA	GCGTGGTGGTGGATGGTTGTA	GAGGTGTTGTCCTCACTGGTCAG	419
SLC22A14	AGAGGTAACAGAGCAGAGGCAACA	TGACAGCAGGATGGCAGGGTAG	50
SLITRK2	ATTCAGGAAGGTGCCTTTACAA	CCCATGATGTCACAGGTACAGT	377
SPATS1	GCGAGGGCAATTCTAGGGTAAACTAT	AACATATAATGGCTTCCTGGCAGAGG	372
SREK1	GGCCTGCCTGGATTGCCTTCT	CGAGATACCAACCTTCTGCACAAGG	357
STAB2	GAAGGCAGGTCTCACCTATCTCC	CAGTACAGGAAGAGCAACAAGTGG	457
STAB2	GAAGGCAGGTCTCACCTATCTCC	CAGTACAGGAAGAGCAACAAGTGG	455
TENM1	GCTCCATGCTGGAGCTGAATATC	CAGATGTCTAACTCCATGCTGC	513

Table 3.1 Primer sequences for validating candidate variants from whole blood (The University of Adelaide).

TEP1	CTTGGGATTCTAGCAGAGCTGG	GCCATGTGCCAGTCCTACCTTT	439
TIGD	CTAAGATCATTGACGAAGGTGGCTACA	CCAGATCCATCAGAGACATCACTATCC	597
TRIM54	CTAACTCACAGTGTCACTACGGATGG	CAAGCAGGTGCAAGAGACAAGACT	455
TUBA3D	CCTCGCCACGGCAAGTACAT	GCCTGGTTGACAGAGCAAGACT	339
UBXN7	GGGTGACATGCAGTTCGTGAAGG	ATAGTGGCTGCTTTGGGTGGC	595
UPF3	ACCTCGGTGTCTGGCTACATCT	CACACGGAACTCACGCTGCT	590
USP26 343	GTAGCCGTGCCAAGCACATG	ATGCATCATGAACGCCACTG	323
USP26 273	CGGGCACTGTTTGAAAGTCA	ATCCTGGCTCCACACATTGG	292
WIPI2	GCTGTGCGTCTGTGTGAGTAGG	CGATGGAGTGTCGGCTCTAAGAAG	372
ZDHHC9	TACTCTGTGATGACCACGAAGC	GCCTCTGTTTCTTCTCTCGTGT	393
ZMYM3	CCAGAGTCTCTCCCACTGAA	TGGGGCAAATTCCTGTTCAC	209
ZNF628	CGAGTGCGGCAAGTCCTTCAAG	TGGTGCAGCAGCAGGTTGGA	468
ZNF674	CATGTCTAATAAGATGATATGAAC	TTCTAGAAGTCTGGGAAGTTGA	732
ZNF814	CACCAGAGACTGCTCACTAGAGAAGA	TGAACTCGTTGATGGCTCCTAAGATG	680

3.7 Polymerase chain reaction (PCR)

PCRs were amplified in 20 µL or 50 µL reaction volumes using the 10X buffer (Roche NimbleGen, Madison WI, USA) (100 mM Tris-HCl, 15 mM MgCl2, 500 mM KCl (pH8.3 [20°C])) with Taq DNA Polymerase (Roche NimbleGen, Madison WI, USA) as per the manufacturers instructions. The optimum annealing temperature of the primers was 60°C.

94°C	4 min		
94°C	30 sec	}	x 35
60°C	30 sec		
72°C	30 sec		
72°C	2 min		

3.7.1 PCR product visualization

PCR products were run on an agarose gel with ethidium bromide. Either 1 or 2% gels were used for visualisation where required. 1 kb+ DNA ladder (Invitrogen, Carlsbad, CA, USA) for products > 200 bp or pUC19 ladder for products < 200 bp were used as size markers. Gels were photographed using an INGENIUS syngene Bio imaging capture system and GeneSnap image acquisition software version 7.05.

3.8 Sanger sequencing reaction and clean up

The QIAquick PCR purification kit (CAT # 28106, Qiagen, Stanford, CA, USA) was used to clean PCR products. Whole reaction volume was mixed with five times volume buffer PB. Up to 600 µL of the PCR and buffer PB was transferred into the QIAquick spin column, and spun at 10,000 x g to bind PCR products onto the membrane. The

membrane was washed once with 700 µL of buffer PE. Supernatant was discarded and the column was spun again at 15,000 x g for 5 minutes to evaporate all the left over alcohol. PCR bands were eluted in 30 µL of sterile H_2O. Purified PCR product concentration was determined using the UV spectrophotometer (CAT # ND-1000, Nanodrop).

Approximately 100 ng of genomic DNA was subjected to BigDye™ Terminator sequencing reaction. The reactions were set up in 20 µL volume containing appropriate amount of DNA, 1 µL of primer at 10 µM, 1 µL of ABI Big Dye™V3.1 (Perkin Elmer, Waltham, MA, USA), 3 uL of Big Dye™ BufferV3.1 and distilled water. Sequencing conditions were as followed: 15 cycles of denaturation at 96^{o} C for 30 seconds, annealing at 50 C for 15 seconds and extension at 60 C for 4 minutes. To clean up the sequencing reaction, 25 µL of distilled water and 75 µL of 100% isopropanol (Ajax Finechem, Thermo Fisher, Waltham, MA, USA) were added and gently vortexed. The mixture was allowed to sit at room temperature for 15 minutes, followed by 20 minutes of centrifugation at 15,000 x g. After discarding the supernatant 250 µL of 75% isopropanol was added, followed by 10 minutes of centrifugation at 15,000 x g. The supernatant was removed and the tube was allowed to air dry for 10 minutes.

Dried sequencing products were size fractionated by capillary electrophoresis by using an ABI prism 3700 genetic analyser (Applied Biosystems, Foster City, CA, USA).
Sequencing data was analysed using DNASTAR Lasergene 10 Seqman Pro (DNASTAR, Inc. Madison, WI, USA).

3.9 DNA variant and gene prioritization

Stringent filtering parameters following recommended published guidelines[7-10] were used to assist in prioritising potential disease causative variants for cerebral palsy.

1. The type of variant i.e. frameshift, splice, stop-gain and missense. Nonsense and frameshift mutations are predicted to result in loss of protein function making them more obvious candidates; however, not all loss of function variants are deleterious.[11]
2. *In silico* prediction of functional effect at the amino acid level used algorithms such as PolyPhen, Mutational Taster and SIFT.[4,5] These computational tools are designed to discriminate between deleterious and benign amino acid changes with a reported accuracy of 75-80%.[12]
3. Evolutionary conservation (phyloP or GERP scores) – pathogenic mutations tend to affect more highly conserved amino acids compared to benign variants.[13,14]
4. The predicted effect of haploinsufficiency i.e. a single functional copy of one allele may not be sufficient for normal gene function.[15]
5. Brain expression pattern i.e. is the expression pattern of the gene relevant to the disease of interest (Allen Human Brain Atlas).
6. Previous disease association reported in OMIM. Additional support is provided if the variant or gene of interest has previously been associated with the same or similar phenotype.[8]

To further assist in selecting candidate genes we used two recently developed complex tools 1) Combined Annotation-Dependent Depletion (CADD) Score and 2) Residual Variation Intolerance Score (RVIS).[16,17]

3.10 References

1. Bainbridge, M.N. *et al.* Targeted enrichment beyond the consensus coding DNA sequence exome reveals exons with higher variant densities. *Genome Biol* **12**, R68 (2011).
2. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60 (2009).
3. DePristo, M.A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat genetics* **43**, 491-8 (2011).
4. Adzhubei, I., Jordan, D.M. & Sunyaev, S.R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet, Haines et al.* **Chapter 7**, (2013).
5. Kumar, P., Henikoff, S. & Ng, P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4**, 1073-81 (2009).
6. Schwarz, J.M., Rodelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* **7**, 575-6 (2010).
7. Gilissen, C., Hoischen, A., Brunner, H.G. & Veltman, J.A. Disease gene identification strategies for exome sequencing. *Eur J Hum Genet* **20**, 490-7 (2012).
8. Goldstein, D.B. *et al.* Sequencing studies in human genetics: design and interpretation. *Nat Rev Genet* **14**, 460-70 (2013).
9. Hoischen, A., Krumm, N. & Eichler, E.E. Prioritization of neurodevelopmental disease genes by discovery of new mutations. *Nat Neurosci* **17**, 764-72 (2014).

10. MacArthur, D.G. *et al.* Guidelines for investigating causality of sequence variants in human disease. *Nature* **508**, 469-76 (2014).
11. Shendure, J. Next-generation human genetics. *Genome Biol* **12**, 408 (2011).
12. Cooper, G.M. & Shendure, J. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nat Rev Genet* **12**, 628-40 (2011).
13. Cooper, G.M. *et al.* Single-nucleotide evolutionary constraint scores highlight disease-causing mutations. *Nat Methods* **7**, 250-1 (2010).
14. Vissers, L.E. *et al.* A de novo paradigm for mental retardation. *Nat Genet* **42**, 1109-12 (2010).
15. Huang, N., Lee, I., Marcotte, E.M. & Hurles, M.E. Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet* **6**, e1001154 (2010).
16. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310-5 (2014).
17. Petrovski, S., Wang, Q., Heinzen, E.L., Allen, A.S. & Goldstein, D.B. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet* **9**, e1003709 (2013).

Chapter 4

Whole exome sequencing points to considerable genetic heterogeneity of cerebral palsy

Gai McMichael, Matthew N Bainbridge, Eric Haan, Mark Corbett, Alison Gardner, Suzanna Thompson, Bregje W M van Bon, Clare L van Eyk, Jessica Broadbent, Corinne Reynolds, Michael E O'Callaghan, Lam Son Nguyen, David L Adelson, Ray Russo, Shalini Jhangiani, Harsha Doddapaneni, Donna M Muzny, Richard A Gibbs, Jozef Gecz,* Alastair H MacLennan *

*Equal project supervisors

Molecular Psychiatry (2015) 20, 176-182

Statement of Authorship

Title of Paper	Whole-exome sequencing points to considerable genetic heterogeneity of cerebral palsy
Publication Status	Published
Publication Details	Molecular Psychiatry, 2015, 20:176-182

Author Contributions

By signing the Statement of Authorship, each author certifies that their stated contribution to the publication is accurate and that permission is granted for the publication to be included in the candidate's thesis.

Name of Principal Author (Candidate)	Gai McMichael		
Contribution to the Paper	First author and wrote manuscript, study design, coordination and recruitment of families involved in the study, prepared DNA samples for sequencing and validation, performed analysis on all samples, interpreted genetic data and correlated genetic results with clinical data, validated results with Sanger sequencing		
Signature		Date	

Name of Co-Author	Professor Jozef Gecz		
Contribution to the Paper	Study design, genetic data interpretation and manuscript evaluation		
Signature		Date	18/1/2016

Name of Co-Author	Professor Eric Haan		
Contribution to the Paper	Study design, clinical and genetic data interpretation and manuscript evaluation		
Signature		Date	13/1/16

Name of Co-Author	Professor Alastair MacLennan		
Contribution to the Paper	Study design, clinical data interpretation and manuscript evaluation		
Signature		Date	21/1/16

Name of Co-Author	Dr Mark Corbett		
Contribution to the Paper	Bioinformatics support		
Signature		Date	13/1/16

Name of Co-Author	Alison Gardner		
Contribution to the Paper	Bioinformatics support		
Signature		Date	14/1/16

Statement of Authorship

Title of Paper	Whole-exome sequencing points to considerable genetic heterogeneity of cerebral palsy
Publication Status	Published
Publication Details	Molecular Psychiatry, 2015, 20:176-182

Author Contributions

By signing the Statement of Authorship, each author certifies that their stated contribution to the publication is accurate and that permission is granted for the publication to be included in the candidate's thesis.

Name of Co-Author	Dr Ray Russo		
Contribution to the Paper	Provided patient clinical data and support in patient recruitment		
Signature		Date	19/1/16

Name of Co-Author	Dr Suzanna Thompson		
Contribution to the Paper	Evaluation of MRI data		
Signature		Date	14.1.16

Name of Co-Author	Professor David L Adelson		
Contribution to the Paper	Study design		
Signature		Date	18/1/16

Name of Co-Author	Dr Clare L van Eyk		
Contribution to the Paper	Technical support		
Signature		Date	18/1/16

Name of Co-Author	Jessica Broadbent		
Contribution to the Paper	Study recruitment support		
Signature		Date	21/1/16

Statement of Authorship

Title of Paper	Whole-exome sequencing points to considerable genetic heterogeneity of cerebral palsy
Publication Status	Published
Publication Details	Molecular Psychiatry, 2015, 20:176-182

Author Contributions

By signing the Statement of Authorship, each author certifies that their stated contribution to the publication is accurate and that permission is granted for the publication to be included in the candidate's thesis.

Name of Co-Author	Dr Richard Gibbs		
Contribution to the Paper	Study design, data interpretation and manuscript evaluation		
Signature		Date	

Name of Co-Author	Dr Matthew Bainbridge		
Contribution to the Paper	Genetic data analysis and interpretation, algorithm development and manuscript evaluation		
Signature		Date	

Name of Co-Author	Donna M Muzny		
Contribution to the Paper	Baylor College of Medicine-Human Genome Sequencing Centre (BCM-HGSC) Project management		
Signature		Date	

Name of Co-Author	Shalini Jhangiani		
Contribution to the Paper	BCM-HGSC Centre coordinator		
Signature		Date	

Name of Co-Author	Dr Harsha Doddapaneni		
Contribution to the Paper	Sequencing manager		
Signature		Date	

Statement of Authorship

Title of Paper	Whole-exome sequencing points to considerable genetic heterogeneity of cerebral palsy
Publication Status	Published
Publication Details	Molecular Psychiatry, 2015, 20:176-182

Author Contributions

By signing the Statement of Authorship, each author certifies that their stated contribution to the publication is accurate and that permission is granted for the publication to be included in the candidate's thesis.

Name of Co-Author	Dr Bregje W M van Bon		
Contribution to the Paper	Clinical re-examination of cerebral palsy cases with a pathogenic variant		
Signature		Date	10-1-16

Name of Co-Author	Corinne Reynolds		
Contribution to the Paper	Contributed to recruitment of families and clinical data collection		
Signature		Date	06/01/16

Name of Co-Author	Lam Son Nguyen		
Contribution to the Paper	Technical support		
Signature		Date	07.9.16

Name of Co-Author	Dr Michael E O'Callaghan		
Contribution to the Paper	Contributed to recruitment of families		
Signature		Date	5/1/2016

4.1 Abstract

Cerebral palsy (CP) is a common, clinically heterogeneous group of disorders affecting movement and posture. Its prevalence has changed little in 50 years and the causes remain largely unknown. The genetic contribution to CP causation has been predicted to be ~2%. We performed whole-exome sequencing of 183 cases with CP and both parents (98 cases) and one parent (67 cases) and 18 singleton cases (no parental DNA). We identified and validated 61 *de novo* protein-altering variants in 43 out of 98 (44%) case parent trios. Initial prioritization of variants for causality was by mutation type, whether they were known or predicted to be deleterious and if they occurred in known disease genes whose clinical spectrum overlaps CP. Further prioritization used two multidimensional frameworks – the Residual Variation Intolerance Score and the Combined Annotation-dependent Depletion score. Ten *de novo* mutations in three previously identified disease genes (*TUBA1A* (n=2), *SCN8A* (n=1) and *KDM5C* (n=1)) and in six novel candidate CP genes (*AGAP1*, *JHDM1D*, *MAST1*, *NAA35*, *RFX2* and *WIPI2*) were predicted to be potentially pathogenic for CP. Additionally, we identified four predicted pathogenic, hemizygous variants on chromosome X in two known disease genes, *L1CAM* and *PAK3*, and two novel candidate CP genes, *CD99L2* and *TENM1*. In total, 14% of CP cases, by strict criteria, had a potentially disease-causing gene variant. Half were in novel genes. The genetic heterogeneity highlights the complexity of the genetic contribution to CP. Function and pathway studies are required to establish the causative role of these putative pathogenic CP genes.

4.2 Introduction

Cerebral palsy (CP) is the most frequent cause of physical disability in childhood with a prevalence of 2-2.5 per 1000 live births that has changed little in fifty years.[1] It is a clinically heterogeneous group of non-progressive disorders, primarily affecting movement and posture. Comorbidities can include intellectual disability (ID), autism spectrum disorder (ASD), epilepsy, speech and language deficits, and visual and hearing impairments.[2] Known epidemiological risk factors include preterm delivery, intrauterine growth restriction, intrauterine infection and male gender (male: female ratio 1.3:1).[3-5] However, causative pathways are poorly understood. Acute or chronic intrapartum fetal compromise, historically considered the cause of CP, is found in <10% of cases.[6] Substantial empiric recurrence risks for certain types of CP and identification of mutant genes in a small number of familial cases are consistent with a genetic contribution to CP.[7-9] Defining that contribution has been hampered by the largely sporadic nature of the disorder. Candidate CP gene mutation screening and association studies have been inconclusive.[10] A recent genome-wide search for copy number variants (CNVs) in 50 CP cases did not yield any *de novo* or obvious candidate mutations.[11]

High-throughput sequencing of whole-exome captured genomic DNA (WES) is an efficient strategy for finding rare, disease-causing mutations, including *de novo* mutations, and has established that *de novo* mutations are associated with a sizeable proportion of sporadic cases of ID, ASD and schizophrenia.[12-15] To investigate the possible contribution of *de novo* and other rare variants to CP, we sequenced the exomes of 183 sporadic cases and when available, their unaffected parents.

4.3 Materials and Methods

4.3.1 Study cohort

The study cohort comprised 183 Caucasian cases with CP and 263 parents: 98 case-parent trios; 67 case-parent duos (DNA from only one parent) and 18 singletons (no parental DNA). No case with a confirmed diagnosis of CP was excluded. The CP diagnosis was confirmed by a paediatric rehabilitation specialist using standard published criteria relating to non-progressive disorders of movement control and posture.[16] Brain imaging reports were available for 112 of the cases (61%). The cohort's overall phenotypic, clinical and demographic characteristics (Table 4.1) were very similar to population distributions described in the 2013 report of the Australian Cerebral Palsy Register (see Web Resources). Paediatric specialist evaluation of the available medical records for potential known causes of CP was complete before enrollment to this study. Comprehensive details of the cohort are provided in Supplementary Table A.

Approval from the Women's and Children's Health Network Human Research Ethics Committee, Adelaide, Australia, and Child and Adolescent Health Service Ethics Committee at the Princess Margaret Hospital, Perth, Australia and the Internal Review Board at Baylor College of Medicine, Houston, USA was obtained for this project. Written informed consent was obtained from participants or their parents.

4.3.2 DNA extraction

DNA extraction from lymphoblastoid cell lines (LCLs) was performed on a fully automated large volume nucleic acid purification system (Qiagen Autopure LS; Qiagen, Stanford, CA), which ensures high quality DNA from EBV transformed LCLs

(Materials and methods **Chapter 3.3**). The affected child and parent's blood was extracted using a QIAamp DNA Blood Mini Kit and QIAamp DNA Blood Maxi Kit (Qiagen, Stanford, CA) respectively, following the manufacturer's instructions.

4.3.3 Illumina library construction

Genomic DNA samples were constructed into Illumina paired-end pre-capture libraries according to the manufacturer's protocol (*Illumina Multiplexing_SamplePrep_Guide_ 1005361_D*) with modifications as described in the BCM-HGSC protocol (https://hgsc.bcm.edu/sites/default/files/documents/Illumina_Barcoded_Paired_End_ Capture_Library_Preparation.pdf). Briefly, 1 µg of genomic DNA in 100 µl volume was sheared into fragments of approximately 300-400 base pairs in a Covaris plate with E210 system followed by end-repair, A-tailing and ligation of the Illumina multiplexing PE adaptors. Pre-capture Ligation Mediated-PCR (LM-PCR) was performed for seven cycles of amplification using the 2X SOLiD Library High Fidelity Amplification Mix (a custom product manufactured by Invitrogen). Universal primer IMUX-P1.0 and a pre-capture barcoded primer IBC were used in the PCR amplification. In total a set of 12 such barcoded primers were used on these samples. Purification was performed with Agencourt AMPure XP beads after enzymatic reactions. Following the final XP beads purification, quantification and size distribution of the pre-capture LM-PCR product was determined using the LabChip GX electrophoresis system (PerkinElmer).

4.3.4 Illumina exome capture

Six pre-capture libraries were pooled together (approximately 166 ng/sample, 1ug per pool) and hybridized in solution to the HGSC VCRome 2.1 design[17] (42Mb, NimbleGen) according to the manufacturer's protocol *NimbleGen SeqCap EZ Exome*

Library SR User's Guide (Version 2.2) with minor revisions. Human COT1 DNA and full-length Illumina adaptor-specific blocking oligonucleotides were added into the hybridization to block repetitive genomic sequences and the adaptor sequences. Post-capture LM-PCR amplification was performed using the 2X SOLiD Library High Fidelity Amplification Mix with 14 cycles of amplification. After the final AMPure XP bead purification, quantity and size of the capture library was analysed using the Agilent Bioanalyser 2100 DNA Chip 7500. The efficiency of the capture was evaluated by performing a qPCR-based quality check on the four standard NimbleGen internal controls. Successful enrichment of the capture libraries was estimated to range from a 6 to 9 of ΔCt value over the non-enriched samples.

4.3.5 Illumina sequencing

Library templates were prepared for sequencing using Illumina's cBot cluster generation system with TruSeq PE Cluster Generation Kits (Cat. no. PE-401-3001). Briefly, these libraries were denatured with sodium hydroxide and diluted to 3-6 pM in hybridization buffer in order to achieve a load density of ~800K clusters/mm^2. Each library pool was loaded in a single lane of a HiSeq flow cell, and each lane was spiked with 2% phiX control library for run quality control. The sample libraries then underwent bridge amplification to form clonal clusters, followed by hybridization with the sequencing primer. Sequencing runs were performed in paired-end mode using the Illumina HiSeq 2000 platform. Using the TruSeq SBS Kits (Cat. no. FC-401-3001), sequencing-by-synthesis reactions were extended for 101 cycles from each end, with an additional seven cycles for the index read. Sequencing runs generated approximately 300-400 million successful reads on each lane of a flow cell, yielding 5-6 Gb per

sample. With these sequencing yields, samples achieved an average of 92% of the targeted exome bases covered to a depth of 20X or greater.

4.3.6 Analysis

Analysis was conducted as previously described.[18] Briefly, Illumina sequence data was aligned to the human reference genome (HG19) with BWA.[19] Variant qualities were recalibrated with GATK.[20] Variants were called using ATLAS-SNP and the SAMtools program pileup. Reads were locally realigned at presumptive insertion or deletion events. Differences between the human reference and the sequence reads (variants) were identified. Variants were annotated with their minor allele frequency in normal populations, previous association to disease, predicted effect on the human gene models, association of the gene with disease and, when parental data was available, their inheritance. Variants were considered *de novo* if neither parent had the variant, and candidate recessive and inherited X-linked variants were selected by segregation analysis. The availability of unaffected siblings would have provided additional power to address the significance of potential causative variants for CP, however due to the constraints of human research ethics unaffected siblings were not collected for this study.

Sanger sequencing was performed across two sites to test segregation and to validate rare variants predicted to alter the gene product: 1) HGSC - an in-house, automated pipeline designed primers for *de novo* variant sites. Sanger sequencing was performed using BigDye terminator chemistry 3.1 (ABI) and sequenced using an ABI 3730xl DNA analyser (Applied Biosystems, Foster City, CA, USA). Sequencing data was analysed using in-house software, SNP-D, and consed; 2) The University of Adelaide -

primers incorporating the candidate variant for a subset of *de novo* and X-linked variants were designed using Primer3 (v.0.4.0). Sanger sequencing was performed using BigDye terminator chemistry 3.1 (ABI) and sequenced using an ABI prism 3700 genetic analyser (Applied Biosystems, Foster City, CA, USA). Sequencing data was analysed using DNASTAR Lasergene 10 Seqman Pro (DNASTAR, Inc. Madison, WI, USA). All validation was performed using genomic DNA isolated from whole blood.

4.3.7 DNA variant and gene prioritization

Where possible we followed recent guidelines for investigating causality of sequence variants in human disease prior to future experimental validation in animal and *in vitro* models.[21] We identified variants that were in known disease genes whose clinical spectrum overlaps CP or *de novo* in any gene. We checked these against dbSNP, 1000 genomes or Exome Variant Server (EVS). Synonymous and intronic *de novo* variants were not included once they were predicted to be neutral. Unique variants were assessed using multiple criteria. However, to select the best possible candidate genes and variants we used 1) a combination of Residual Variation Intolerance Score (RVIS)[22] and Combined Annotation–Dependent Depletion (CADD)[23] with cut-offs for RVIS<50th percentile for known and novel genes and CADD>10 for known OMIM disease genes and CADD>20 for novel candidate genes; 2) the type of variant (i.e. frameshift, splice, stop-gain, missense); 3) *in silico* prediction of functional effect at the amino acid level by various algorithms;[24,25] 4) evolutionary conservation; 5) brain-expression pattern; 6) the predicted effect of haploinsufficiency[26] and 7) previous disease association reported in Online Mendelian Inheritance in Man (OMIM). The RVIS and CADD tools are complex, multidimensional tools.[22,23] RVIS ranks genes in terms of intolerance to functional genetic variation and CADD integrates several well-known tools, among

these also PolyPhen and SIFT.[22,23] We assessed the validity of RVIS and CADD as prioritization models using two cut-off scores (RVIS <25, CADD >20 and RVIS <50, CADD >10) applied to recent ID[12,13], autism[13,27,28] and schizophrenia[14] WES papers to confirm if the genes identified would be selected under this model. We performed a T-test through Partek (see Web Resources) to compare the haploinsufficiency scores[26] of *de novo* variants (where a score was available) with non-mutated genes.

In total, 2.5 Tbp of sequence data was generated from 446 individuals and aligned to the human reference genome sequence (hg19). This yielded an average redundant coverage of 78.8X with 92% of targeted, coding bases having at least 20X redundant read coverage. Nine out of 14 predicted causative genes for CP had better than 90% of the gene covered with greater than 20X reads (Supplementary Table B).

More comprehensive details of the above methods can be found in the Materials and methods section, **Chapter 3, sections 3.3 – 3.9**.

4.4 Results

4.4.1 Case-parent trios (n = 98)

4.4.1.1 De novo mutations

We identified 61 *de novo* mutations in 43 cases from 98 case-parent trios (44%; one to four *de novo* mutations per individual), 60 autosomal and one X-chromosomal (Supplementary Table C). The *de novo* rate for protein altering mutations (one stopgain, five splice site, three frameshift deletions, two frameshift insertions, two non-frameshift deletions, and 48 missense mutations) was 0.62 per individual. The rate was within the lower end of the range previously reported for ID (0.63 and 1.49 per individual)[12,13] and ASD (0.70 per individual).[14] Assessment of the haploinsufficiency scores[26] for *de novo* mutations (where scores were available; n=38) identified a slight enrichment compared to non-mutated genes but this was not significant (Supplementary Figure A). Based on RVIS and CADD and a set of multidimensional prioritization criteria (see methods in this chapter 4.3.7), we selected 10 *de novo* mutations in 98 case-parent trios as potentially relevant to CP causation. These included one splice site and nine missense *de novo* mutations. Two different *de novo* mutations were identified in *TUBA1A*, in two cases, and the remaining eight *de novo* mutations occurred in different genes (Tables 4.2 and 4.3).

Of the 10 *de novo* mutations predicted to be causative for CP, four occurred in genes associated with neurological disorders (Table 4.2): p.P480L in *KDM5C*, a lysine-specific histone demethylase and known X-linked ID (XLID) gene;[29] p.G1050S in *SCN8A*, associated with cognitive impairment;[30] and p.R123C and p.L152Q in *TUBA1A* associated with neuronal migration disorders[31] (Table 4.2). Six *de novo* mutations occurred in genes not known to be associated with disease (Table 4.3): including a

single mutation (c.957+1G>A) predicted to affect splicing in *AGAP1,* which directly regulates AP-3-dependent trafficking[32] and five missense mutations in *JHDM1D*; *MAST1*; *NAA35*; *WIPI2* and *RFX2*. Additionally, we identified seven loss of function (LOF) variants in non-disease associated genes (*CDK17*, *ENPP4*, *LTN1*, *MIIP*, *NEMF*, *SSPO* and *UBQLN3*), but these genes either had a high RVIS percentile or there were other frequent LOF variants in these genes in EVS.

4.4.1.2 Inherited X-chromosome and recessive variants

In addition to the *de novo* mutations, which included one X-chromosome *de novo* mutation (in *KDM5C*), four maternally inherited X-chromosome variants in four male cases (4%) predicted to be causative for CP were also seen in the 98 case-parent trios (Tables 4.2 and 4.3). Two were in known OMIM disease genes: p.P161A in *L1CAM* associated with L1 syndrome;[33] and p.R493C in *PAK3* associated with XLID. The remaining two were in genes not yet associated with human disease: a nonsense mutation p.K163X in *CD99L2*; and a missense variant p.G2533S in *TENM1* (Tables 4.2 and 4.3 and Supplementary Table D).

No predicted deleterious homozygous autosomal recessive variants were identified. However, we identified a single case with compound heterozygous variants, c.11562+2T>A (paternal allele) and c.7158+5G>A, p.R2682Q and p.A2854D (maternal alleles) in the *HSPG2* gene (Supplementary Table E). HSPG2 mutations are associated with a wide variety of phenotypes, often including chondrodysplasia. However, isolated muscle stiffness without obvious signs of chondrodysplasia has also been noted in individuals with mutations in HSPG2.[34] Although perlecan deficiency has been shown to underlie the chondrodysplasia, severe muscle stiffness may also induce bone

deformities. In individuals with HSPG2 mutations, muscle stiffness or bone deformities are often the first symptoms. Deformities of lower limbs and feet are common.[34] This individual showed joint limitation of his knee, ankle and foot and equinus deformity.

4.4.2 Whole-exome sequencing

In summary 14% of the 98 case-parent trios had a predicted causative variant. Six (6%) had a *de novo* (4) or inherited (2) predicted deleterious mutation in known disease genes. Eight (8%) had an implicated pathogenic variant, *de novo* (6) and inherited (2), in novel candidate CP genes. Multiple prioritization criteria including the assessment of RVIS and CADD scores assisted in implicating the deleteriousness for these mutations (see methods in this chapter 4.3.7). We tested this model in previously published ID, autism and schizophrenia WES data sets (Supplementary Figures B–I and also Supplementary Table F). These conditions have a high level of locus heterogeneity involving hundreds of independent risk loci and this was reflected in the gene distribution pattern. We also looked at the multispecies alignment for each of the 12 predicted causative missense mutations, *de novo* (9) and inherited (3) and found a high level of conservation across species (Supplementary Table G).

Two different variants (p.E304K in *L1CAM* and p.M100I in *PAK3*) in two X-linked genes we predicted to be causative in the trio cases were identified in two of the 67 duo cases (Supplementary Table H). Five singleton cases had a variant in known disease genes, associated with phenotypes overlapping with CP or genes associated with neurological disorders that are frequently comorbid with CP (Supplementary Table I). As inheritance of these variants was unknown we could not confidently associate these with CP.

4.4.3 Pathway analyses

Biological system analysis of the predicted causative CP genes (n=14) was conducted using the Ingenuity Pathway Analysis (IPA) [www.ingenuity.com]. Three genes (*L1CAM*, *PAK3*, and *TUBA1A*) were identified (p=0.006) as being involved in Axonal Guidance Signaling. The top associated network was developmental disorder, hereditary disorder and neurological disease (Supplementary Figure J).

4.4.4 Clinical associations

From 15 cases (13 trio cases and two duo cases) with a predicted pathogenic variant, 14 (93%) had positive findings on brain imaging. Overall we had brain imaging on 112 cases. Two out of 21 cases with intraventricular haemorrhage, four out of 46 cases with white matter damage, five out of 20 cases with developmental brain malformations and three out of 15 cases with unilateral cerebral infarction had a predicted pathogenic variant, respectively. One case out of seven cases with normal imaging had a predicted pathogenic variant. No predicted pathogenic variants were found in two cases with in utero infection or in one case with a brain tumour (Supplementary Table J).

Twenty-two out of 98 trio cases had a diagnosis of ID. Seven of the 22 cases (0.32) had a predicted pathogenic variant. These included one novel protein-truncating mutation (0.05) (*AGAP1*), four missense mutations (0.18) in known ID genes (*KDM5C*, *L1CAM*, *SCN8*A, and *TUBA1A*) and two novel (0.09) missense mutations (*JHDM1D* and *NAA35*) (see also Supplementary Tables K and L). These proportions were slightly lower than previously reported for ID where three out of 22 (0.14)[12] and six out of 45 (0.13)[13] novel truncating mutations and four out of 22 (0.18)[12] and 11 out of 45 (0.24)[13] variants in known ID genes were found.

4.5 Discussion

Cerebral palsy encompasses a large group of childhood movement and posture disorders and can occur as an isolated finding or additional phenotypic features may be present.[5] Previous estimates have suggested that the contribution of genetic variants to the burden of CP is about 2%.[35] We followed, where possible, recent guidelines to resolve which of the many variants may be implicated in human disease[19]. Using strict criteria we found that 14% of 98 case-parent trios had variants that were putatively disease causing in five known disease genes and eight novel candidate genes. *L1CAM* and *PAK3* genes, which we implicate in CP from our trio cases, had one additional predicted deleterious variant each in two of the 67 duo cases. Overall the predicted deleterious variants of this study were in genes other than the currently known CP genes, suggesting considerable genetic heterogeneity underlying CP.[7,8,36] We used multiple criteria to select CP-relevant unique variants, including a combination of two recently developed prioritization algorithms, RVIS[22] and CADD.[23] We tested these algorithms in other genetically heterogeneous neurological disorders[12-15,28] and found a similar distribution pattern between our CP-relevant variants and those found in ID.

In total, five known disease genes had variants predicted to be causative for CP. These included *KDM5C, SCN8A, TUBA1A, L1CAM* and *PAK3*. Weakness, poor muscle control and spasticity have been reported in patients with mutations in these genes.[29,30,33,37,38] In combination with the current sequencing results, these signs of movement and posture disturbances implicate CP as a previously unrecognized diagnosis of the clinical spectrum associated with mutations in these genes.

KDM5C is a well-known XLID gene with varying clinical features including mild-severe ID, microcephaly, spasticity and seizures.[29] Recently the same variant (p.P480L) was reported as a novel mutation in two male siblings diagnosed with ID.[39] Although the majority of *KDM5C* affected individuals are males, carrier females with milder phenotypes including ID and spasticity have been reported.[29,40] This female (26026P) showed spastic diplegia, ID and speech dyspraxia. Two cases, 106115P and 169451P, had different novel *de novo* missense mutations in *TUBA1A*. *De novo* mutations in this gene have been reported in a wide spectrum of neuronal migration disorders.[31] Furthermore, spastic diplegia or quadriplegia, ID and microcephaly are common in *TUBA1A* mutations.[37,41,42] Case 106115P was diagnosed with diplegic CP, optic atrophy, adducted thumbs, seizures and moderate ID. Case 169451P had diplegic CP. Brain MRI of both individuals showed, except for the characteristic sign of pachygyria, *TUBA1A* associated anomalies such as cerebellar and corpus callosum anomalies.[41,42]

One female case (43043P) diagnosed with hemiplegic CP and ID had a missense mutation, p.G1050S, in *SCN8A*. Intellectual disability, epilepsy and varying degrees of motor dysfunction including ataxia and spasticity have been described in children with mutations in *SCN8A*.[13,30,43] In mice with partial- and complete-loss-of-function of *scn8a* ataxic gait, tremor, dystonia, muscle atrophy and loss of hind-limb function have been reported.[43]

Two males (142424P and 156438P) had different maternally inherited X-chromosome variants in *L1CAM*. *L1CAM* is involved in neurite outgrowth and when mutated is associated with L1 syndrome, a disorder with variable features including hydrocephalus, ID, spastic paraplegia, adducted thumbs and agenesis of the corpus

callosum.[33] The phenotype of both cases is clearly compatible with L1 syndrome, although case 156438P was less severely affected showing only subclinical signs of hydrocephalus with ventricular enlargement on brain CT.

We also found two male cases with different maternally inherited missense variants in *PAK3*, a known XLID gene. *PAK3* belongs to a family of serine/threonine p21-activating kinases, critical effectors that link Rho GTPases to cytoskeleton reorganization and nuclear signaling. Two essential domains have been identified, the N-terminal regulatory region with a CRIB domain and a C-terminal catalytic domain that includes a kinase domain.[44] One of the variants affects the CRIB domain (p.M100I in 165447P) and the other the kinase domain (p.R493C in 15015P). Both mutations haven't been reported previously. Case 165447P was diagnosed with a left cerebral artery infarction in the neonatal period and was later diagnosed with dystonic CP. Case 15015P was prematurely born and neonatal brain MRI showed a grade IV intraventricular haemorrhage. He was diagnosed with hemiplegic CP and epilepsy and showed cognitive abilities in the upper limit of the low average range. Different *PAK3* mutations may lead to different clinical outcomes. While isolated loss of *PAK3* kinase activity causes non-syndromic ID, a dual molecular effect is seen in those cases with a syndromic neurocutaneous phenotype.[38] In the latter phenotype also hypotonia, apparent motor delay with inability to walk, hyperreflexia and afinalistic movements have been described.[38] Out of 2400 males in the NHLBI Exome Sequencing Project, only two missense variants are reported, which suggests that indeed missense variants in this gene are not tolerated in males.

In addition to these known disease genes, we identified a splice variant in *MAOB,* which might be of interest for CP. Most duplications and deletions of *MAOB* include nearby genes and have been associated with severe ID.[45] Two missense variants in known XLID genes, *IQSEC2* and *ZDHHC9*, were identified in two cases. Though they met our prioritization criteria, we did not consider them causative of CP mostly because no intellectual disability was reported for either case.

Brain imaging reports were available for 61% of cases. Thirteen per cent of these had a potentially pathogenic variant to explain their CP. It has been well recognized that malformations of cortical development often have a genetic basis.[46] Our findings of putative pathogenic mutations in cases with diffuse white matter damage or a unilateral thrombotic event raise the possibility of a genetic contribution in these pathologies.

The results of this preliminary study suggest that CP is genetically heterogeneous, likely reflecting its clinical heterogeneity. The classification of CP as a non-progressive disorder of movement and posture has been defined by international criteria[4,5]. All individuals with CP in this study met these criteria for CP at time of diagnosis. Physical examination before inclusion was performed by experienced paediatric rehabilitation specialists. The fact that mutations in known syndromic ID genes such as *TUBA1A* and *L1CAM* have been identified within this cohort confirms the clinical variability of these disorders and suggests that movement disturbances are part of their phenotypic spectrum.

It should be emphasized that these data, on their own, do not definitively implicate the identified gene variants in CP causation. Ultimately ongoing cellular, molecular and

animal model functional studies of these candidate CP genes will provide final evidence of pathogenicity. As yet, they cannot be used in diagnostic screening and clinical decision making.

With the exception of *TUBA1A, L1CAM* and *PAK3*, where two different mutations were identified for each gene, only single mutations were observed for the remaining genes. Much larger sample sizes are required to confirm the involvement of any single gene and this can be addressed through targeted re-sequencing of candidate genes.[47] Whole-exome sequencing and whole-genome sequencing of new CP cohorts will undoubtedly reveal other candidate mutations and potential novel syndromes associated with movement disturbances. In time, we may gain a better understanding of the interactions between heterogeneous genetic susceptibility, environment and chance that appear to underlie this relatively common, complex and burdensome neurodevelopmental disorder.

Table 4.1 Clinical characteristics of 183 individuals with cerebral palsy.		
Characteristics	**Number (%)**	**Mean**
Cerebral Palsy Type		
Quadriplegia	31 (17)	
Diplegia	65 (36)	
Hemiplegia	76 (42)	
Triplegia	4 (2)	
Dyskinesia	7 (3)	
Other Demographics		
Gestational age (weeks)		34.7
Birth weight (grams)		2414
Maternal age (years)		30
Paternal age (years)		33
Male	99 (54)	
Plurality (n>1)	25 (14)	
Co-morbidities		
Intellectual disability	35 (19)	
Epilepsy	46 (25)	
Autism	12 (7)	

Table 4.2 Novel variants in known (OMIM) genes in CP cases identified in 98 trios: CADD score > 10 and RVIS % < 50.

Case number, CP type, Co-mobidities	Sex	Gene	Type	Protein change	Brain expression	PPH2 score	GERP score	HI	RVIS %	CADD Phred Score	Function	Disease association
15015P Hemi Ep	M	*PAK3*	X	p.R493C	Mod	0.910	5.54	U	26.23	21.2	Synapse formation and plasticity in the hippocampus	Intellectual disability, X-linked 30/47 (300558)
26026P Di ID	F	*KDM5C*	DN	p.P480L	Mod	1.000	5.42	0.431	3.50	22.5	Histone demethylase	Intellectual disability, X-linked, syndromic, Claes-Jensen type (300534)
43043P Hemi ID	F	*SCN8A*	DN	p.G1050S	Low	0.638	4.46	0.632	2.34	13.21	Mediates the voltage-dependent sodium ion permeability of excitable membranes	Cognitive impairment with or without cerebellar ataxia (614306); epileptic encephalopathy, early infantile, 13 (614558)
106115P Di ID, Ep	M	*TUBA1A*	DN	p.R123C	High	0.997	3.18	U	35.42	12.3	Major constituent of microtubules	Lissencephaly 3 (602529);
156438P Hemi ID	M	*L1CAM*	X	p.P161A	High	0.929	5.12	U	4.92	11.11	Neuronal migration and differentiation	Corpus callosum (304100); CRASH syndrome (303350); Hydrocephalus (307000); MASA syndrome (303350)
169451P Di	F	*TUBA1A*	DN	p.L152Q	High	0.785	5.15	U	35.42	12.3	See above	See above

Abbreviations: RVIS – Residual Variation Intolerance Score; CADD – Combined Annotation-Dependent Depletion; HI – Haploinsufficiency Index; PPH2 – Polyphen2; Dip – diplegia, Hem – hemiplegia; M-male, F – female; DN – de novo, X – X-chromosome; U - unknown

Table 4.3 Novel candidate genes in CP cases identified in 98 trios; CADD score > 20 and RVIS % < 50.

Case number, CP type, Co-mobidities	Sex	Gene	Type	Protein change	Brain expression	PPH2 score	GERP score	HI	RVIS %	CADD Phred Score	Function
19019P Dip	M	*TENM1*	X	p.G2533S	Low	0.993	5.83	0.726	2.61	22.8	May function as a cellular signal transducer
33033P Quad ID, EP	M	*AGAP1*	DN	Splice	Mod	U	5.08	U	1.10	29.7	Regulates the adapter protein 3 (AP-3)-dependent trafficking of proteins in the endosomal-lysosomal system
39039P Di	M	*CD99L2*	X	Stopgain	High	U	5.06	0.15	34.31	32	May function as a homophilic adhesion molecule
44044P Hem	F	*WIPI2*	DN	p.Y246C	Mod	0.999	5.93	0.181	7.43	25.7	Early component of the autophagy machinery
90099P Dip	M	*MAST1*	DN	p.P500L	Mod	1.000	4.39	0.571	1.94	25.8	Links the dystrophin/utrophin network with microtubule filaments via the syntrophins
96105P Quad ID, Ep	F	*JHDM1D*	DN	p.S727W	Low	1.000	5.85	U	40.56	21.1	Histone demethylase required for brain development
108117P Hemi ID, Ep	M	*NAA35*	DN	p.W532C	Low	1.000	5.40	U	43.77	23.1	Component of the N-terminal acetyltansferase C (NatC) complex
161443P Hemi	F	*RFX2*	DN	p.Y91C	Low	0.997	5.36	0.820	10.87	20.2	Encodes transcription factors that contain a highly-conserved winged helix DNA binding domain factors

Abbreviations: RVIS – Residual Variation Intolerance Score; CADD – Combined Annotation-Dependent Depletion; HI – Haploinsufficiency Index; PPH2 – Polyphen2; Dip – diplegia, Hem – hemiplegia, Quad - quadriplegia; M-male, F – female; DN – de novo, X – X-chromosome; U - unknown

4.6 References

1. Stanley F., B.E., Alberman B. How common are the cerebral palsies? *In: Stanley F, Blair E, Alberman B, editors. Cerebral Palsies: Epidemiology &Causal Pathways. London (UK): Mac Keith Press:* , p.22-48 (2000).
2. Rosenbaum, P. *et al.* A report: the definition and classification of cerebral palsy April 2006. *Dev Med Child Neurol Suppl* **109**, 8-14 (2007).
3. O'Callaghan, M.E. *et al.* Epidemiolgic Associations With Cerebral Palsy. *Obstet Gynaecol* **118**, 576-582 (2011).
4. Colver, A., Fairhurst, C. & Pharoah, P.O. Cerebral palsy. *Lancet* **383**, 1240-9 (2014).
5. Moreno-De-Luca, A., Ledbetter, D.H. & Martin, C.L. Genomic insights into the causes and classification of the cerebral palsies. *Lancet Neurol* **11**, 283-92 (2012).
6. Nelson, K.B. & Ellenberg, J.H. Antecedents of cerebral palsy. Multivariate analysis of risk. *N Engl J Med* **315**, 81-6 (1986).
7. Hirata, H. *et al.* ZC4H2 mutations are associated with arthrogryposis multiplex congenita and intellectual disability through impairment of central and peripheral synaptic plasticity. *Am J Hum Gen* **92**, 681-95 (2013).
8. Moreno-De-Luca, A. *et al.* Adaptor protein complex-4 (AP-4) deficiency causes a novel autosomal recessive cerebral palsy syndrome with microcephaly and intellectual disability. *J Med Genet* **48**, 141-4 (2011).
9. Tollanes, M.C., Wilcox, A.J., Lie, R.T. & Moster, D. Familial risk of cerebral palsy: population based cohort study. *BMJ* **349**, g4294 (2014).

10. O'Callaghan, M.E. *et al.* Fetal and maternal candidate single nucleotide polymorphism associations with cerebral palsy: a case-control study. *Paediatrics* **129**, e414-23 (2012).
11. McMichael, G. *et al.* Rare copy number variation in cerebral palsy. *Eur J Hum Genet* (2013).
12. de Ligt, J. *et al.* Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med* **367**, 1921-9 (2012).
13. Rauch, A. *et al.* Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* **380**, 1674-82 (2012).
14. Sanders, S.J. *et al.* De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237-41 (2012).
15. Xu, B. *et al.* De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nat Genet* **44**, 1365-9 (2012).
16. Badawi, N. *et al.* What constitutes cerebral palsy? *Dev Med Child Neurol* **40**, 520-7 (1998).
17. Bainbridge, M.N. *et al.* Targeted enrichment beyond the consensus coding DNA sequence exome reveals exons with higher variant densities. *Genome Biol* **12**, R68 (2011).
18. Bainbridge, M.N. *et al.* De novo truncating mutations in ASXL3 are associated with a novel clinical phenotype with similarities to Bohring-Opitz syndrome. *Genome Med* **5**, 11 (2013).
19. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60 (2009).

20. DePristo, M.A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**, 491-8 (2011).

21. MacArthur, D.G. *et al.* Guidelines for investigating causality of sequence variants in human disease. *Nature* **508**, 469-76 (2014).

22. Petrovski, S., Wang, Q., Heinzen, E.L., Allen, A.S. & Goldstein, D.B. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet* **9**, e1003709 (2013).

23. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310-5 (2014).

24. Adzhubei, I., Jordan, D.M. & Sunyaev, S.R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* **Chapter 7**, (2013).

25. Schwarz, J.M., Rodelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* **7**, 575-6 (2010).

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

27. Neale, B.M. *et al.* Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* **485**, 242-5 (2012).

28. O'Roak, B.J. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* **485**, 246-50 (2012).

29. Rujirabanjerd, S. *et al.* Identification and characterization of two novel JARID1C mutations: suggestion of an emerging genotype-phenotype correlation. *Eur J Hum Genet* **18**, 330-5 (2010).

30. Trudeau, M.M., Dalton, J.C., Day, J.W., Ranum, L.P. & Meisler, M.H. Heterozygosity for a protein truncation mutation of sodium channel SCN8A in a patient with cerebellar atrophy, ataxia, and mental retardation. *J Med Genet* **43**, 527-30 (2006).

31. Poirier, K. *et al.* Expanding the spectrum of TUBA1A-related cortical dysgenesis to Polymicrogyria. *Eur J Med Genet* **21**, 381-5 (2013).

32. Nie, Z. *et al.* Specific regulation of the adaptor protein complex AP-3 by the Arf GAP AGAP1. *Dev Cell* **5**, 513-21 (2003).

33. Yamasaki, M., Thompson, P. & Lemmon, V. CRASH syndrome: mutations in L1CAM correlate with severity of the disease. *Neuropaediatrics* **28**, 175-8 (1997).

34. Stum, M. *et al.* Spectrum of HSPG2 (Perlecan) mutations in patients with Schwartz-Jampel syndrome. *Hum Mutat* **27**, 1082-91 (2006).

35. Rajab, A. *et al.* An autosomal recessive form of spastic cerebral palsy (CP) with microcephaly and mental retardation. *Am J Med Genet A* **140**, 1504-10 (2006).

36. Kruer, M.C. *et al.* Mutations in gamma adducin are associated with inherited cerebral palsy. *Ann Neurol* **74**, 805-14 (2013).

37. Bahi-Buisson, N. *et al.* The wide spectrum of tubulinopathies: what are the key features for the diagnosis? *Brain* **137**, 1676-700 (2014).

38. Magini, P. *et al.* A mutation in PAK3 with a dual molecular effect deregulates the RAS/MAPK pathway and drives an X-linked syndromic phenotype. *Hum Mol Genet* (2014).

39. Grafodatskaya, D. *et al.* Multilocus loss of DNA methylation in individuals with mutations in the histone H3 lysine 4 demethylase KDM5C. *BMC Med Genomics* **6**, 1 (2013).

40. Jensen, L.R. *et al.* Mutations in the JARID1C gene, which is involved in transcriptional regulation and chromatin remodeling, cause X-linked mental retardation. *Am J Hum Genet***76**, 227-36 (2005).

41. Bahi-Buisson, N. *et al.* Refinement of cortical dysgeneses spectrum associated with TUBA1A mutations. *J Med Genet* **45**, 647-53 (2008).

42. Poirier, K. *et al.* Large spectrum of lissencephaly and pachygyria phenotypes resulting from de novo missense mutations in tubulin alpha 1A (TUBA1A). *Hum Mutat* **28**, 1055-64 (2007).

43. Veeramah, K.R. *et al.* De novo pathogenic SCN8A mutation identified by whole-genome sequencing of a family quartet affected by infantile epileptic encephalopathy and SUDEP. *Am J Hum Genet* **90**, 502-10 (2012).

44. Rejeb, I. *et al.* A novel splice mutation in PAK3 gene underlying mental retardation with neuropsychiatric features. *Eur J Hum Genet* **16**, 1358-63 (2008).

45. Klitten, L.L., Moller, R.S., Ravn, K., Hjalgrim, H. & Tommerup, N. Duplication of MAOA, MAOB, and NDP in a patient with mental retardation and epilepsy. *Eur J Hum Genet* **19**, 1-2 (2011).

46. Mochida, G.H. & Walsh, C.A. Genetic basis of developmental malformations of the cerebral cortex. *Arch Neurol* **61**, 637-40 (2004).

47. O'Roak, B.J. *et al.* Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* **338**, 1619-22 (2012).

Chapter 5

A family diagnosed with cerebral palsy and intellectual disability with a X-linked mutation in ZC4H2

5.1 Introduction

Cerebral palsy is the most common physical neurological disability of childhood. The rate of cerebral palsy throughout the developed world has remained largely unchanged in the last 50 years and affects 2-3 in every 1,000 live births.[1-6] The vast majority of cases are singletons, however cerebral palsy occurs more than expected among twins[7,8] and siblings[9-11] suggesting a heritable component, likely genetic.[12]

ZC4H2 is a member of the zinc finger domain-containing protein family. Mutations are extremely rare causing neurodevelopmental disorders of the central and peripheral nervous systems.[13] We report on a family with two male siblings diagnosed with 'cerebral palsy' and intellectual disability. An initial diagnosis of cerebral palsy was made when the first affected boy was young. However, over the course of time the progressive nature of the disorder became apparent taking it out of the realm of cerebral palsy. Further a female sibling was significantly behind in her language, cognitive and personal/social skills. Analysis of whole-exome sequencing (WES) data of five out of 10 family members (affected and unaffected individuals) identified a maternally inherited X-chromosome variant in the *ZC4H2* gene segregating throughout this family.

5.2 Clinical reports

5.2.1 The family

We recruited nine out of the 10 family members (at the time of the study the youngest sibling was too young to participate) of a non-consanguineous family. The mother was born in Uruguay to parents of Spanish decent and the father was born in Chile to parents of Chilean and German decent. The family consisted of eight siblings and was part of a larger study aimed at identifying novel genes for cerebral palsy. The inheritance pattern

of ‘cerebral palsy’ in the family was not unequivocally X-linked. We sequenced five individuals (affected and unaffected) by WES (I-1, I-2, II-1, II-2 and II-5, Figure 5.1).

Approval from the Women’s and Children’s Health Network Human Research Ethics Committee, Adelaide, Australia was obtained for this project. Written informed parental consent was obtained.

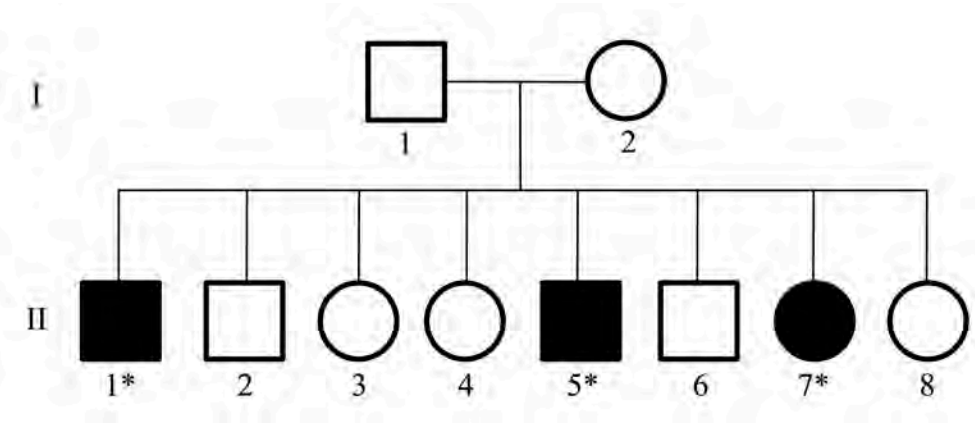


Figure 5.1 Two generation pedigree with two brothers (II-1 and II-5) diagnosed with ‘cerebral palsy’ and intellectual disability and a sister (II-7) significantly behind in language, cognitive and personal/social skills.

5.2.2 The parents (I-1 and 2, Figure 5.1)

Both parents were intellectually normal, however the father had been diagnosed with schizophrenia and it had been suggested that he may have had attention deficit hyperactivity disorder as a child. The mother described herself as having dyslexia. There was no family history of intellectual disability, neurological disorder or psychiatric disorder among the parents’ siblings, parents or more distant relatives.

5.2.3 The son with ‘cerebral palsy’ and intellectual disability (II-1, Figure 5.1)

The pregnancy for this individual was complicated by what was considered to be a maternal ‘liver infection’ at 35 weeks gestation. He was born at 40 weeks gestation after a normal labour and delivery with birth weight 3890 grams, birth length 51 cm and birth

head circumference 37.5 cm. Brief resuscitation with a bag and mask was required. There were no feeding difficulties but gastro-oesophageal reflux was a problem from around six weeks of age and he responded to treatment.

At around 6 months of age he was not rolling and was floppy. Brain magnetic resonance imaging (MRI) was performed at that stage and showed only prominent ventricles and subarachnoid spaces. He did not sit until 12 months of age. A diagnosis of cerebral palsy was made at 14-18 months of age. Examination by a neurologist at 18 months of age showed him to be essentially hypotonic. Knee reflexes were brisk but other reflexes were normal, plantar responses were extensor and there was no ankle clonus.

He crawled at 24 months of age. Examination of the eyes by an ophthalmologist at 2.5 years of age was normal. Examination by a neurologist at 34 months of age showed him to be hypertonic with minor increase in tone in his upper limbs with brisk reflexes, and substantial increase in tone in the lower limbs with hyperreflexia, reflex spread, absence of clonus and extensor plantar responses. He exhibited a diplegic crawl, he was able to pull himself up against furniture and his grasp was clumsy and slow with a tendency to tuck the thumb under the fingers. He walked with a frame at 4-5 years of age. At around 6 years of age, there was a change in his motor findings, with the development of dystonia. He could walk with crutches at 7 years of age.

He had a seizure during intercurrent illnesses at 3, 6 and 7 years of age, which was considered the result of hypoglycaemia (lowest blood glucose 1.1 mmol/L).

At 11.5 years of age his clinical picture was described as consistent with mixed spastic/dystonic quadriplegic cerebral palsy at level III on the Gross Motor Function

Classification System. The spasticity involved predominantly the lower limbs. The dystonia worsened with effort and emotional change.

At 12 years of age he had a few single words and mainly used sign language, he was able to write his name and draw a little, and he could feed himself, although chewing was slow and dyscoordinate. Drooling was a significant problem. He was unable to walk independently but could walk holding on to furniture or with elbow crutches but preferred to crawl when at home. For long distances he used a wheelchair. Glasses had been prescribed for refractive error. He had cognitive impairment and attended a special unit in a normal school and received physiotherapy, occupational therapy and speech therapy at school. His mother considered that his understanding was at a 6-7 year old level.

On examination at 12 years of age head circumference was 58.3 cm (2 cm above the 98th centile), he was considered short for his age and weight was 46.5 kg (50th centile). He was not dysmorphic. Cranial nerve examination was normal. To communicate he used a communication device. Finger-nose-finger testing was precise but slow and there was no intention tremor. With arms outstretched, his fingers distally became slightly spooned and asymmetrically postured, suggesting a degree of dystonia. There was no tremor. Subtle dystonia was evident in most positions. Tone varied markedly in his arms, depending on his degree of excitement. Upper limb reflexes were normal but in his legs, the knee jerks were pathologically brisk and ankle jerks were easy to obtain in spite of the fact that he had very tight Achilles tendons. Plantar responses could not be assessed.

5.2.4 The son with 'cerebral palsy' and intellectual disability (II-5, Figure 5.1)

This individual was born at term after a normal pregnancy, labour and delivery with Apgar score of 9 at 5 minutes, birth weight 4080 grams, birth length 50 cm and birth head circumference 36 cm. There were no feeding difficulties. As with his brother, gastro-oesophageal reflux was a problem from birth and was controlled by treatment.

The family was concerned about his poor head control at 4-5 months of age and suspected that he had the same disorder as his older brother. At 8 months of age he tended to choke on some solid foods, was not rolling, tended to fist his hands, and had increased lower limb tone, delayed head control and truncal hypotonia. Adductor and hamstring tone was increased and lower limb reflexes were increased. He was considered to have cerebral palsy, although it was recognised that this was a diagnosis of convenience, given his similarly affected brother. He sat at 9 months, crawled at 18 months and walked with a frame at 3 years.

There were documented episodes of hypoglycaemia at 4 and 4.5 years of age, the first associated with a grand mal seizure. He had a speech disorder, considered a mixed dyspraxia and dysarthria. His physical disability was rated level III on the Gross Motor Function Classification System. A Griffiths developmental assessment (GMDS-ER) at 53 months showed an average mental age of 22 months; Locomotor scale 16.8 months; Personal-Social scale 21.9-24 months; Hearing and Speech scale 16.8 months; Eye and Hand Coordination scale 24 months and Performance scale 22.7 months.

At 6 years of age, he was vocalising but had no words, he was able to feed himself with a fork and spoon but had a tendency to choke on food, and drooling was a significant problem. He scribbled when given a pencil, was not toilet trained, tended to crawl about

the house when at home, used a walking frame at school and had a wheelchair for longer distances. He attended a special school and had physiotherapy, occupational therapy and speech therapy at school. Glasses had been prescribed because of a refractive error.

On examination at 6 years, head circumference was 49.6 cm (2-50th centile). He was not dysmorphic. A trial of L-dopa was without effect. On examination at 6 years 9 months, he could walk using a walker. His gait was diplegic and he was up on his toes with legs internally rotated. Cranial nerve function was normal. He had subtle dystonia. In the arms, his tone was normal, although it increased with excitement and contralateral tapping. In his legs (after recent Botox treatment), his knee jerks were brisk and there was no clonus at the ankles.

5.2.5 The daughter with developmental delay (II-7, Figure 5.1)

Development was assessed using the Australian Developmental Screening Test. She showed age appropriate abilities in her fine motor and gross motor skills but was significantly behind in her language, cognitive and personal/social skills. Hearing was normal. She did not have features of cerebral palsy.

5.2.6 Remaining siblings (II2, II3, II-4, II-6 and II-8, Figure 5.1)

Brother II-2 did not speak well until after 3 years of age. There were no learning problems at school. Sister II-3 had minor problems with literacy skills at school but was not receiving extra help. Sister II-4 was unable to speak well until after 3 years of age. There were no reported learning problems at school. Brother II-6 was assessed at 5 years of age because of concerns about development and behaviour. His teacher

observed that he had difficulty forming friendships with children of his own age, concentrating in the classroom and sitting still. He was unable to work independently, struggled to maintain concentration and was not achieving at the same level as his peers. Assessment using the Behavioural Rating Scales showed severe inattention, hyperactivity and impulsivity consistent with a diagnosis of Attention Deficit Hyperactivity Disorder. A Wechsler Preschool and Primary Scale of Intelligence-III assessment showed a low average full-scale IQ, with strengths in performance skills, where he performed in the average range. The youngest sister II-8 was 2 years of age at the time of this study and there were no apparent concerns.

5.3 Methods

5.3.1 DNA isolation

For each individual in the two-generation pedigree, with the exception of individual (II-8), DNA was isolated from whole blood using a QIAmp DNA Midi kit (Qiagen, Stanford, CA, USA) following the manufacturer's instructions.

5.3.2 Whole-exome sequencing and analysis

For five individuals (I-1, I-2, II-1, II-2, and II-5, Figure 5.1), WES was performed with Illumina TrueSeq WES protocols (Materials and methods **Chapter 3.5.3**). The enriched DNA (350–400 bp products) was sequenced on the Illumina HiSeq2000 platform (Axeq Technologies), which returned on average 48.3 Mb of 100 bp paired-end reads per individual. Reads were quality trimmed with the FASTX toolkit and aligned to UCSC Genome Browser HG19 with the Burrows-Wheeler Aligner with default parameters, except for -d 5 and -l 35. SAMtools was used for generating BAM files. Sequence variants were realigned, recalibrated, and reported with the Genome Analysis Toolkit and categorized with Annovar. Variants were filtered based on dbSNP132, 1000 genomes (sited November, 2012) and Exome variant server (http://evs.gs.washington.edu/EVS/) (Materials and methods **Chapter 3.5.4**). We applied different inheritance models for this family: autosomal recessive (homozygous or compound heterozygous), X-linked, and *de novo* (postulated germline mosaicism in one of the parents).

5.3.3 Sanger sequencing

Sanger sequencing was performed in all family members (with the exception of II-8) for the purpose of validation and to assess segregation within the family of rare variants

predicted to alter the gene product. Unique primers incorporating the variants of interest were designed using Primer3 (v.04.0).

5.4 Results

We identified unique sequence variants in seven genes. There were three X chromosome variants: c.637C>T (p.Arg213Trp) in *ZC4H2*; c.20T>A (p.Leu7Gln) in *MUM1L1*; and c.3140G>A (p.Gly1047Glu) in *PCDH11X* validated by Sanger sequencing. All three variants segregated in the extended family according to postulated X chromosome inheritance. It was difficult to determine which variant was causative, however during data analysis, the latter two variants (in *MUM1L1* and *PCDH11X*) appeared as SNPs in the latest release of dbSNP135 and were thus considered to be rare, non-causative variants. There were two autosomal variants, both in *MYH14* and both on the same allele (i.e. not compound heterozygous). These *MYH14* variants did not segregate, given that the healthy mother and her healthy daughter (II-3, Figure 5.1) had them as well. There were *de novo* heterozygous variants in *FLNB* and *PLAC4* and potentially compound-heterozygous changes in *SLC11A2*. These were not confirmed by Sanger sequencing.

We considered the variant c.637C>T (p.Arg213Trp) in *ZC4H2* to be deleterious in this family. Multispecies alignment for p.Arg213Trp found a high level of conservation across species (Table 5.1). PolyPhen2 and MutationTaster predicted this amino acid change to be damaging to ZC4H2 function.[14] Sanger sequencing confirmed the segregation pattern and validated this variant within the family (Figure 5.2). The mother (I-2) and two daughters (II-4 and II-7) each carried one copy of the normal variant and one copy of the mutated variant. The two affected boys (II-1 and II-5) both carried the mutated variant. The remaining family members tested (I-1, II-2, II-3, II-6) carried the wild type allele/alleles. *ZC4H2* has been shown to be subject to X inactivation,[13] however preferential inactivation of the mutated X chromosome in the carrier mother (I-

2) and two carrier female siblings (II-4 and II-7) was not observed.

Table 5.1. Multispecies alignment around p.Arg213 of *ZC4H2*.									
Mutant	K	S	R	S	W	N	P	K	K
Human	K	S	R	S	R	N	P	K	K
Mouse	K	S	R	S	R	N	P	K	K
Rat	K	S	R	S	R	N	P	K	K
Frog	K	S	R	S	R	N	P	K	K
Zebrafish	K	S	R	S	R	N	P	K	K

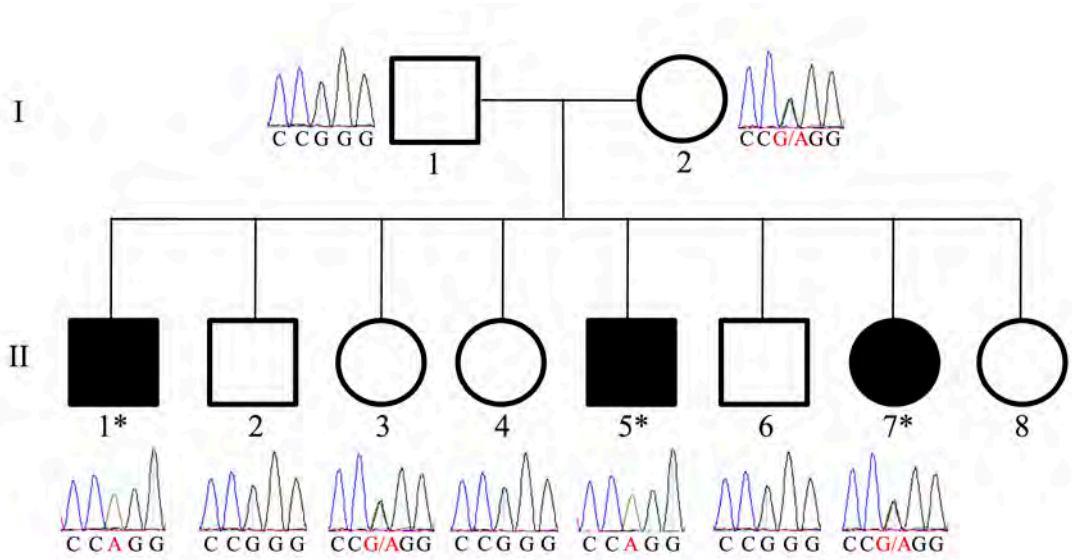


Figure 5.2 Representative sequence chromatograms of *ZC4H2* wildtype, mutations and heterozygous variants identified in nine out of the ten family members. Wildtype – I-1, II-2, II-4, II-6; mutation – II-1 and II-5; heterozygous variant – I-2, II-3 and II-7. II-8 was too young to be tested at the time of this study.

5.5 Discussion

We performed WES for five out of 10 individuals in a two-generation family and Sanger sequencing for the remaining family members (with the exception of II-8 as she was too young to partake at the time of the study). We identified an association between a mutation in the X-linked zinc-finger gene *ZC4H2* and cerebral palsy and intellectual disability. As well as the two affected male individuals hemizygous for the *ZC4H2* gene mutation, one of the three carrier females (II-7) was significantly behind in her language, cognitive and personal/social skills.

Subsequently, we were asked by another research group to collaborate in *ZC4H2* functional studies.[13] They had identified four familial and three simplex cases with four different mutations in *ZC4H2*. In one of their families they identified exactly the same *ZC4H2* mutation as found in our family. Each of the four families in their study had a diagnosis of arthrogryposis multiplex congenita (AMC) characterised by an onset of muscle weakness in utero, severe contractures, dysmorphic facial features, skeletal anomalies including scoliosis, hip dislocation and pes equinovarus and intellectual disability.[13,15] Affected individuals in our family and the other family shared a number of features including spasticity, intellectual disability, seizures, dysarthria, deficit in expressive language and drooling. Some of these clinical features were also present in some of the other families.[13] Other overlapping features with remaining families included; equinovarus feet, contracture of Achilles tendon, dystonia and hypoglycemia.

Functional studies of the different variants in the combined study of five families and three singletons (including the variant in our family) using zebrafish models (I did not carry out this work and it is not part of my PhD thesis) showed loss of ZC4H2 function.

This resulted in abnormal swimming and impaired alpha-motoneuron development. In primary mouse hippocampal neurons, transiently expressed zc4h2 protein localized to the postsynaptic compartment of excitatory synapses and loss of zc4h2 function led to reduced dendritic spine density and impaired central and peripheral synaptic plasticity.[13]

The functional studies confirmed that the mutation in *ZC4H2* identified in our family was the primary cause of the clinical features in the affected boys, including 'cerebral palsy' and intellectual disability. Cerebral palsy is classified as a group of 'non-progressive' disorders, primarily affecting movement and posture, presenting early in childhood, and continuing throughout life.[16,17] A diagnosis of cerebral palsy was made when the first affected boy was young. Whilst at the time a diagnosis of cerebral palsy was appropriate, by early adolescence the progressive nature of the disorder became apparent, no longer supporting the initial diagnosis. Severe spasticity and intellectual disability were the prominent features in both affected boys, features also prominent in another family with the same mutation.[13] The variable clinical manifestations resulting from different mutations in *ZC4H2* demonstrate considerable mutation pleiotropy and fits with the concept of the "expanded phenotype," where mutations in one gene may be associated with a broad spectrum of clinical features rather than a single narrow phenotype.[18] Our study has highlighted the utility of WES as a tool for genetic diagnosis and in particular, the improved efficiency and likelihood of diagnostic success associated with sequencing multiple family members at the outset.

5.6 References

1. Centers for Disease control and Prevention. Economic costs associated with mental retardation, cerebral palsy, hearing loss and vision impairment - United States. *Morb Mortal Wkly Rep 2004;* **53**, 57-59 (2003).
2. Australian Cerebral Palsy Register Report, Birth Years 1993-2006. (Australian Cerebral Palsy Register, 2013).
3. Hagberg, B., Hagberg, G., Beckung, E. & Uvebrant, P. Changing panorama of cerebral palsy in Sweden. VIII. Prevalence and origin in the birth year period 1991-94. *Acta Paediatr* **90**, 271-7 (2001).
4. Hagberg, B., Hagberg, G. & Olow, I. The changing panorama of cerebral palsy in Sweden. VI. Prevalence and origin during the birth year period 1983-1986. *Acta Paediatr* **82**, 387-93 (1993).
5. Stanley, F.J. & Watson, L. Trends in perinatal mortality and cerebral palsy in Western Australia, 1967 to 1985. *BMJ* **304**, 1658-63 (1992).
6. Yeargin-Allsopp, M., Van Naarden Braun, K., Doernberg, N.S. *et al.* Prevalence of cerebral palsy in 8-year-old children in three areas of the United States in 2002: a multisite collaboration. *Paediatrics* **121**, 547-54 (2008).
7. Grether, J.K., Nelson, K.B. & Cummins, S.K. Twinning and cerebral palsy: experience in four northern California counties, births 1983 through 1985. *Paediatrics* **92**, 854-8 (1993).
8. Petterson, B., Nelson, K.B., Watson, L. & Stanley, F. Twins, triplets, and cerebral palsy in births in Western Australia in the 1980s. *BMJ* **307**, 1239-43 (1993).
9. Gustavson, K.H., Hagberg, B. & Sanner, G. Identical syndromes of cerebral palsy in the same family. *Acta Paediatr Scand* **58**, 330-40 (1969).

10. Hemminki, K., Li, X., Sundquist, K. & Sundquist, J. High familial risks for cerebral palsy implicate partial heritable aetiology. *Paediatr Perinat Epidemiol* **21**, 235-41 (2007).

11. O'Callaghan, M.E., MacLennan, A., Gibson, C. *et al.* Epidemiolgic associations with cerebral palsy. *Obstet Gynaecol***118**, 576-582 (2011).

12. Tollanes, M.C., Wilcox, A.J., Lie, R.T. & Moster, D. Familial risk of cerebral palsy: population based cohort study. *BMJ* **349**, g4294 (2014).

13. Hirata, H., Nanda, I., van Riesen, A. *et al.* ZC4H2 mutations are associated with arthrogryposis multiplex congenita and intellectual disability through impairment of central and peripheral synaptic plasticity. *Am J Hum Genet* **92**, 681-95 (2013).

14. Robinson, P.N., Kohler, S., Bauer, S. *et al.* The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. *Am J Hum Genet* **83**, 610-5 (2008).

15. Reimann, J., Jacobson, L., Vincent, A. & Kornblum, C. Endplate destruction due to maternal antibodies in arthrogryposis multiplex congenita. *Neurology* **73**, 1806-8 (2009).

16. Badawi, N., Watson, L., Petterson, B. *et al.* What constitutes cerebral palsy? *Dev Med Child Neurol* **40**, 520-7 (1998).

17. Rosenbaum, P., Paneth, N., Leviton, A. *et al.* A report: the definition and classification of cerebral palsy April 2006. *Dev Med Child Neurol Suppl* **109**, 8-14 (2007).

18. Schaefer, G.B. Genetics considerations in cerebral palsy. *Semin Paediatr Neurol* **15**, 21-6 (2008).

Chapter 6

NK2 homeobox 1 gene mutations in a family diagnosed with ataxic dyskinetic cerebral palsy

Gai McMichael, Eric Haan, Alison Gardner, Tzu Ying Yap, Suzanna Thompson, Robert Ouvrier, Russell Dale, Jozef Gecz and Alastair H MacLennan

European Journal of Medical Genetics (2013) 56, 506-9

Statement of Authorship

Title of Paper	NK2 homeobox 1 gene mutations in a family diagnosed with ataxic dyskinetic cerebral palsy
Publication Status	Published
Publication Details	European Journal of Medical Genetics (2013) 56, 506-9

Author Contributions

By signing the Statement of Authorship, each author certifies that their stated contribution to the publication is accurate and that permission is granted for the publication to be included in the candidate's thesis.

Name of Principal Author (Candidate)	Gai McMichael		
Contribution to the Paper	First author and wrote manuscript, involved the design of the study, coordination and recruitment of the family involved in the study, prepared DNA samples for sequencing and validation, performed analysis on all samples, interpreted genetic data and correlated genetic results with clinical data, validated results with Sanger sequencing		
Signature		Date	

Name of Co-Author	Professor Jozef Gecz		
Contribution to the Paper	Study design, genetic data interpretation and manuscript evaluation		
Signature		Date	18/1/2016

Name of Co-Author	Professor Eric Haan		
Contribution to the Paper	Study design, clinical assessment, genetic data interpretation and manuscript evaluation		
Signature		Date	6/1/16

Name of Co-Author	Professor Alastair MacLennan		
Contribution to the Paper	Study design and manuscript evaluation		
Signature		Date	21/1/16

Name of Co-Author	Alison Gardner		
Contribution to the Paper	Bioinformatics support		
Signature		Date	11/1/16

Name of Co-Author	Dr SuzannThompson		
Contribution to the Paper	Assessment of clinical data		
Signature		Date	14.1.16

Statement of Authorship

Title of Paper	NK2 homeobox 1 gene mutations in a family diagnosed with ataxic dyskinetic cerebral palsy
Publication Status	Published
Publication Details	European Journal of Medical Genetics (2013) 56, 506-9

Author Contributions

By signing the Statement of Authorship, each author certifies that their stated contribution to the publication is accurate and that permission is granted for the publication to be included in the candidate's thesis.

Name of Co-Author	Tzu Ying Yap		
Contribution to the Paper	Technical support		
Signature		Date	08 Jan 2016

Name of Co-Author	Professor Russell Dale		
Contribution to the Paper	Neurological assessment		
Signature		Date	06/01/16

Name of Co-Author	Professor Robert Ouvrier		
Contribution to the Paper	Neurological assessment		
Signature		Date	06/01/16

6.1 Abstract

Benign hereditary chorea caused by mutations in the NK2 homeobox 1 gene (*NKX2-1*), shares clinical features with ataxic and dyskinetic cerebral palsy (CP), resulting in the possibility of misdiagnosis. A father and his two children were considered to have ataxic CP until a possible diagnosis of benign familial chorea was made in the children in early teenage. The father's neurological condition had not been appreciated prior to examination of the affected son. WES of blood derived DNA and bioinformatics analysis were performed.

A 7 bp deletion in exon 1 of *NKX2-1*, resulting in a frame shift and creation of a premature termination codon, was identified in all affected individuals. Screening of 60 unrelated individuals with a diagnosis of dyskinetic or ataxic CP did not identify *NKX2-1* mutations. BHC can be confused with ataxic and dyskinetic CP. Occasionally these children have a mutation in *NKX2-1*.

6.2 Introduction

Cerebral palsy (CP) describes a group of permanent disorders of the development of movement and posture, that are attributed to non-progressive disturbances that occurred in the developing fetal or infant brain.[1] The motor dysfunction of CP is divided into three categories: spastic, ataxic and dyskinetic. Spastic CP is characterised by increased muscle tone. Ataxic CP is characterised by low muscle tone and cerebellar features including intention tremor, poor balance and poor coordination.[2] Dyskinetic CP is characterised by involuntary movements, either athetosis or dystonia, and mixed muscle tone.[1]

Benign Hereditary Chorea (BHC) is an autosomal dominant movement disorder characterised by non-progressive or very slowly progressive chorea with normal cognitive function. Choreiform movements may involve the limbs, face, neck, trunk and tongue. While chorea is the characteristic movement disorder, atypical features including gait disturbance, dystonia, ataxia, intention tremor, dysarthria and pyramidal signs may accompany BHC. In a minority of affected individuals, the disorder may involve the thyroid, presenting as congenital hypothyroidism, and/or lung, with respiratory distress syndrome or recurrent lung infection.[3,4] Benign Hereditary Chorea results from mutations of the NK2 homeobox 1 gene (*NKX2-1*), located at chromosome 14q13.3, which encodes a transcription factor important for the development of the lung, thyroid and brain.[5] *NKX2-1* consists of three coding exons, which are transcribed into two major *NKX2-1* mRNA isoforms encoding different proteins of 371 and 401 amino acids in length.[4] Deletions, missense mutations and nonsense mutations have been described in *NKX2-1* and result in haploinsufficiency.[6,7] The gene is highly expressed in the fetal brain and is involved in neuronal migration and development of

the basal ganglia.[8] We report a father and his two children who were diagnosed with ataxic CP in early childhood. The father's neurological condition was not appreciated until examination of his son. The clinical diagnosis was revised to BHC, with identification of a pathogenic mutation in *NKX2-1* by whole-exome sequencing (WES) of the family during a research study into the genetic basis of CP. To determine other BHC cases with similar phenotypes, we sequenced *NKX2-1* in 60 unrelated individuals diagnosed with ataxic, dyskinetic and/or athetoid CP.

6.3 Clinical report

6.3.1 The Family

The affected family members are a father, his son and his daughter (II-1, III-2 and III-3, in Figure 6.1) with a previously unreported mutation of the *NKX2.1* gene. The father and his wife are Caucasian, non-consanguineous and have an unaffected son.

6.3.2 The affected father

He was born at term after a normal pregnancy, labour and delivery but had delayed onset of respiration. There were no problems in the neonatal period. He walked late and is described as having had genu valgum, an awkward gait and a tendency to fall easily; coordination was generally poor through childhood. He was an above average student at school and completed tertiary studies. Neurological examination at the time his son was diagnosed revealed definite though mild ataxia; he was able to stand on one leg for approximately 10 seconds but tandem gait was ataxic. He did not have nystagmus or chorea. When seen at 49 years of age he described mild dysarthria when tired, and mild functional difficulties due to his ataxia. Walking was not usually a problem but he avoided running; when he ran, his legs "locked" and he overbalanced. He avoided

carrying things in both hands, as this made him less stable. He found it difficult to use both hands simultaneously to perform separate tasks. Examination revealed an inability to heel-toe walk and minor left intention tremor on finger-nose testing.

6.3.3 The affected son

Following a normal pregnancy, he was born at 39 weeks gestation by caesarean delivery after failure of labour to progress with Apgar scores of 9 and 9 at one and five minutes, birth weight 2970 g (10^{th}-50^{th} percentile), birth length 51.5 cm (50^{th}-90^{th} percentile) and birth head circumference 34.0 cm (10^{th}-50^{th} percentile). There was no respiratory distress or abnormalities on newborn screening. A paediatrician documented developmental delay with ataxia at 16 months of age and a diagnosis of ataxic CP was made. He sat at 18 months, walked and had about 40 words at 20 months and used phrases by three years. When seen by a neurologist at 20 months he had an unsteady gait, fell frequently and there were some choreiform movements. At 24 months he had choreoathetoid movements and ataxia persisted. He had speech therapy between two and three years. A diagnosis of BHC was considered at 12 years of age. At 15 years he had choreiform movements and no ataxia; the chorea was exacerbated by emotion. Handwriting was poor and he ran awkwardly, fell frequently and had had a number of fractures. Intelligence was above average, his speech was clear and he was a gifted swimmer in spite of his neurological disorder.

6.3.4 The affected daughter

Following a normal pregnancy, she was born at 40 weeks gestation by elective caesarean delivery with Apgar scores of 6 and 9 at one and five minutes, birth weight 3440 g (50^{th} percentile), birth length 53 cm (90^{th} percentile) and birth head

circumference 34.0 cm (10^{th}-50^{th} percentile). There was mild respiratory distress attributed to meconium aspiration. Newborn screening was normal. Motor milestones were delayed: she sat at 6-8 months, crawled at 12 months and walked at 17 months, with frequent falls but without the jerky movements seen in her brother. Speech development was normal. A neurologist assessed her at 20 months, at which time she was considered to have an ataxic gait; upper limb movement was normal. At 3.5 years, speech and cognitive function were normal and there were mild choreoathetoid movements and an ataxic gait. At 11 years, there was chorea affecting her face, trunk, arms and legs; there was no ataxia. She was a gifted student, with good handwriting.

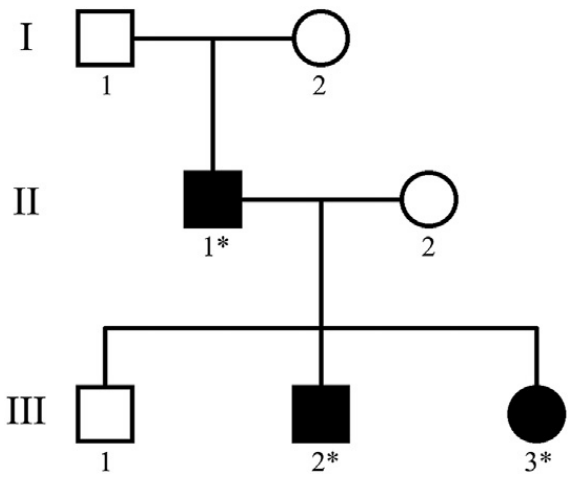


Figure 6.1 Three generation pedigree with de novo mutation in II-1 and subsequent autosomal dominant transmission in III-2 and III-3. All family individuals were sequenced with Sanger sequencing. *Represents individuals carrying the mutation, remaining family members were unaffected. See Fig. 2 for Sanger sequencing trace.

6.3.5 Other cases of dyskinetic or ataxic CP

We selected 60 unrelated Caucasian individuals with a diagnosis of dyskinetic or ataxic CP from the Australian CP Research Study[9] – 29 (48%) had ataxic CP, 16 (27%) had dyskinetic athetoid CP, 10 (17%) had dyskinetic dystonic CP and 5 (8%) had dyskinetic CP without known subtype. There were no reported family histories of CP.

Research ethics approval was obtained from Women's and Children's Health Network Human Research Ethics Committee (Approval No. REC 1946/4/10) and signed parental consent was obtained from the family and all participants involved in the mutational screening.

6.4 Methods

6.4.1 DNA isolation

For each individual in the three generation pedigree DNA was isolated from whole blood using a QIamp DNA Midi kit (Qiagen, Stanford, CA) following the manufacturer's instructions. DNA had previously been extracted from buccal swabs obtained from a convenience cohort of volunteer cerebral palsy cases selected for follow up mutational screening.[10]

6.4.2 Whole-exome sequencing and analysis

Initially WES was performed for the three affected individuals (II-1, III-2 and III-3, Figure 6.1) and the two unaffected individuals in the second and third generation (II-2 and III-1, Figure 6.1). Following TruSeq DNA Exome Enrichment Capture, enriched genomic DNA was massively parallel sequenced on the Illumina HiSeq 2000 platform (Axeq Technologies, Rockville, MD, USA), which returned on average 48.3 million 100 bp pair end reads per individual. These reads were quality trimmed using the FASTX toolkit. Quality reads were mapped to the hg19 build of the human genome by using Burrows-Wheeler Alignment tool. Samtools were used to generate BAM files. Sequence variants were realigned, recalibrated and reported with Genome Analysis ToolKit and categorised with Annovar. We filtered the variants based on dbSNP132 and 1000 genomes, exonic/splice sites, nonsynonymous and then further filtered for

inheritance models including, autosomal dominant, autosomal recessive both homozygous and compound heterozygous and *de novo* variants.

6.4.3 Sanger sequencing

To confirm the mutation in the family unique primers incorporating the 7 bp deletion were designed using Primer3 (v.0.4.0), (forward 5'-CTGTTCCTCATGGTGTCCTGGT-3', reverse 5'-GAATCATGTCGATGAGTCCA AAG-3'). For the 60 unrelated individuals diagnosed with dyskinetic or ataxic CP, the entire protein coding region of the *NKX2-1* gene was amplified in four fragments. Four primer pairs were designed using Primer3 (v.0.4.0). Primer set 1: (forward: 5'-CAGTCGATCCCCTACTCAGC-3', reverse: 5'GTAACAGAGGAGGAGAGATGGT TG-3'), primer set 2: (forward: 5'-AATGCTTTGGGTCTCGTCTCT-3', reverse: 5'-CACT TTCTTGTAGCTTTCCTCCAG-3'), primer set 3: (forward: 5'-TACGTG TACATCCAACAAGATCG-3', reverse: 5'-TACCAAGTCCCTGTTCTTGGAC-3'), primer set 4: (forward: 5'-TCGAAAGAGGGAACTGAGACTGAG-3', reverse: 5'-GCCAGGTTGTTAAGAAAA GTCGA -3'). Sanger sequencing was performed using BigDye terminator chemistry 3.1 (ABI) and sequenced using an ABI prism 3700 genetic analyser (Applied Biosystems, Foster City, CA, USA). Sequencing data was analysed using DNASTAR Lasergene 10 Seqman Pro (DNASTAR, Inc. Madison, WI, USA).

6.5 Results

Analysis of WES data for all individuals in generations II and III identified a 7 bp deletion within exon 1 of *NKX2-1* in all three affected individuals. Both unaffected individuals do not carry this deletion (Figure 6.1). Results were confirmed by Sanger sequencing (Figure 6.2A). Neither of the father's unaffected parents had the deletion, suggesting that it occurred *de novo* in the father (II-1) and was transmitted in an autosomal dominant manner to his affected children (III-2 and III-3). The deletion in *NKX2-1* (NM_003317: exon1: c.84_90del) is predicted to result in a premature termination codon (PTC) and a truncated NKX2-1 protein (p.M59fs*39) (Figure 6.2B). It is plausible that the PTC containing *NKX2-1* mRNA is degraded by non-sense mediated mRNA decay (NMD) resulting in the absence of one copy of the NKX2-1 protein. The effect of NMD on the c.84_90del *NKX2-1* mRNA could not be tested as *NKX2-1* is not expressed in routinely available tissue (i.e. blood, skin or saliva). The c.84_90del mutation is predicted to affect both *NKX2-1* mRNA isoforms.

Sanger sequencing of 60 unrelated participants from the Australian CP Research Study who had been diagnosed with dyskinetic or ataxic CP harboured no *NKX2-1* mutations.

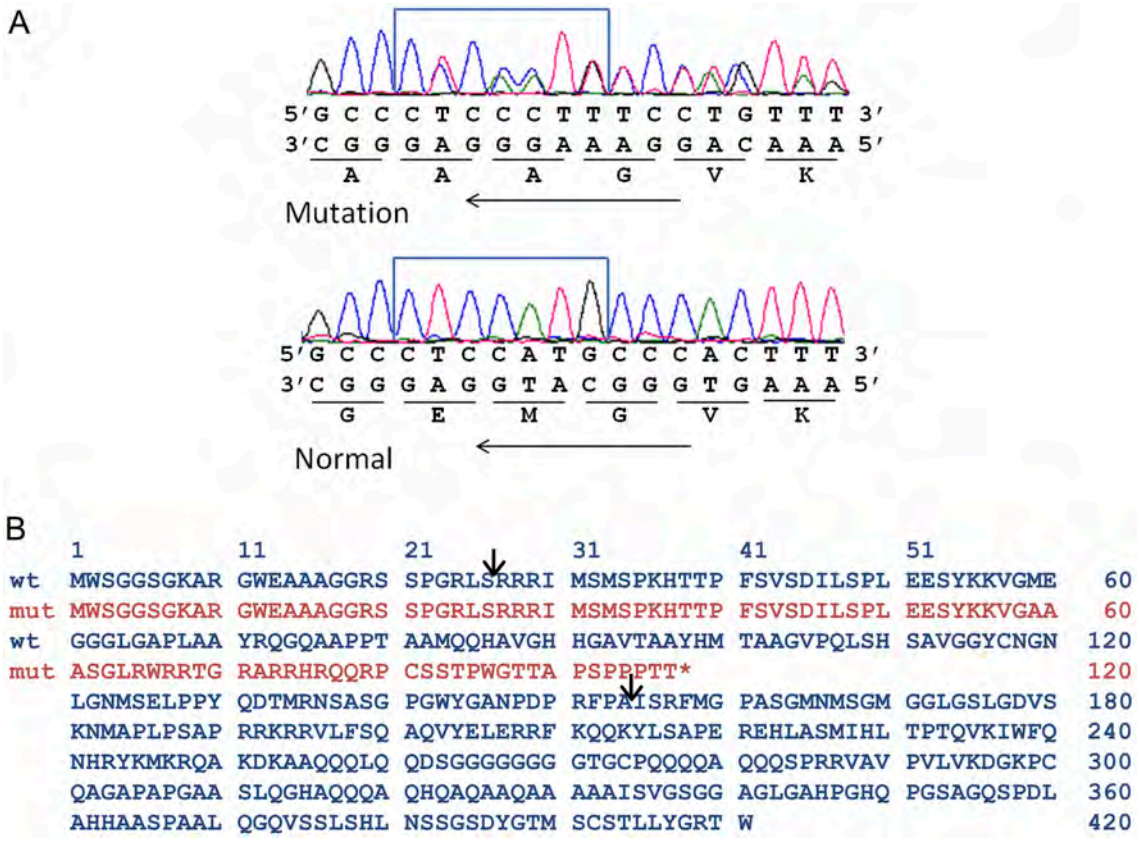


Figure 6.2 (A) Fragments of sequence chromatograms from an affected individual (heterozygous mutation) and an unaffected individual (normal homozygous) from 5'–3' and corresponding amino acid sequences. (B) Comparison of amino acid sequences of wildtype and mutant NKX2-1 proteins. The change was at position 29 introducing a premature termination codon at position 68 resulting in a truncated protein. ↓ represents the exon/exon boundary.

6.6 Discussion

We describe two siblings who were considered to have ataxic CP in early childhood. Their father had a mild cerebellar syndrome, not appreciated at the time of their presentation. Diagnosis was based on the presence of delayed motor milestones and a non-progressive movement disorder, ataxic in nature, associated with gait disturbance and frequent falls. As the children grew older, choreoathetosis became the predominant neurological feature, although ataxia persisted through the first decade. It was recognised that the diagnosis was BHC when reviewed by a neurologists in their early teenage years. Coincidentally, a mutation was identified by WES in *NKX2-1* at around the same time because the family was participating in the Australian CP Research Study. The affected individuals in this family shared a 7 bp deletion of exon 1 of *NKX2-1* resulting in a frame shift, with creation of a premature termination codon. This is predicted to cause haploinsufficiency, either as a result of a non-functional, truncated NKX2-1 protein or 50% reduction of the protein levels due to NMD degradation of the premature termination codon containing *NKX2*-1 mRNA allele.[6,7]

While BHC, with its characteristic choreiform movements, would not commonly be confused with ataxic CP, similar cases have been reported[11,12]. Doyle et al.[12] reported siblings with an *NKX2-1* mutation who had congenital hypothyroidism, global developmental delay and later ataxia, choreoathetosis and dysarthria. Their mother had been diagnosed with CP during childhood; she was described as having ataxia as an adult. Carre et al[12] described a child with an *NKX2-1* mutation who had congenital hypothyroidism and respiratory distress syndrome and in whom ataxic movements and psychomotor delay presented in the first year.

BHC may be difficult to diagnose in early childhood before the characteristic choreiform movements are apparent, especially in the absence of a family history of the disorder and when the child has an atypical movement disorder at the time of first diagnostic evaluation. It is possible that there are a small number of BHC cases among patients diagnosed to have dyskinetic or ataxic CP. However, we did not find any individuals with an *NKX2-1* mutation among 60 unrelated cases of dyskinetic or ataxic CP.

The family reported here exemplifies the problem that clinicians face in distinguishing between CP and genetic neurological disorders that include the motor components of CP among their features, especially when the diagnostic evaluation is being done early in life. Other examples are disorders caused by mutations in *GAD1*,[13,14] *KANK1*,[15] and the adaptor protein complex-4 (*AP4E1*, *AP4M1*, *AP4B1*, and *AP4S1*).[16-19]

This study highlights the importance of genetic investigation of individuals with CP, because a proportion, yet to be defined, will have an underlying genetic disorder, with clinical features that meet the currently accepted criteria for diagnosis of CP.

6.7 References

1. Rosenbaum, P. *et al.* A report: the definition and classification of cerebral palsy April 2006. *Dev Med Child Neurol Suppl* **109**, 8-14 (2007).
2. Rajab, A. *et al.* An autosomal recessive form of spastic cerebral palsy (CP) with microcephaly and mental retardation. *Am J Med Genet A* **140**, 1504-10 (2006).
3. Guazzi, S. *et al.* Thyroid nuclear factor 1 (TTF-1) contains a homeodomain and displays a novel DNA binding specificity. *EMBO J* **9**, 3631-9 (1990).
4. Inzelberg, R., Weinberger, M. & Gak, E. Benign hereditary chorea: an update. *Parkinsonism Relat Disord* **17**, 301-7 (2011).
5. Breedveld, G.J. *et al.* Mutations in TITF-1 are associated with benign hereditary chorea. *Hum Mol Genet* **11**, 971-9 (2002).
6. Asmus, F. *et al.* Clinical differentiation of genetically proven benign hereditary chorea and myoclonus-dystonia. *Mov Disord* **22**, 2104-9 (2007).
7. Krude, H. *et al.* Choreoathetosis, hypothyroidism, and pulmonary alterations due to human NKX2-1 haploinsufficiency. *J Clin Invest* **109**, 475-80 (2002).
8. Kleiner-Fisman, G. *et al.* Alterations of striatal neurons in benign hereditary chorea. *Mov Disord* **20**, 1353-7 (2005).
9. O'Callaghan, M.E. *et al.* The Australian cerebral palsy research study--protocol for a national collaborative study investigating genomic and clinical associations with cerebral palsy. *J Paediatr Child Health* **47**, 99-110 (2011).
10. McMichael, G.L. *et al.* DNA from buccal swabs suitable for high-throughput SNP multiplex analysis. *J Biomol Tech* **20**, 232-5 (2009).
11. Doyle, D.A., Gonzalez, I., Thomas, B. & Scavina, M. Autosomal dominant transmission of congenital hypothyroidism, neonatal respiratory distress, and ataxia caused by a mutation of NKX2-1. *J Paediatr* **145**, 190-3 (2004).

12. Carre, A. *et al.* Five new TTF1/NKX2.1 mutations in brain-lung-thyroid syndrome: rescue by PAX8 synergism in one case. *Hum Mol Genet* **18**, 2266-76 (2009).

13. Lynex, C.N. *et al.* Homozygosity for a missense mutation in the 67 kDa isoform of glutamate decarboxylase in a family with autosomal recessive spastic cerebral palsy: parallels with Stiff-Person Syndrome and other movement disorders. *BMC Neurol* **4**, 20 (2004).

14. McHale, D.P. *et al.* A gene for ataxic cerebral palsy maps to chromosome 9p12-q12. *Eur J Hum Genet* **8**, 267-72 (2000).

15. Lerer, I. *et al.* Deletion of the ANKRD15 gene at 9p24.3 causes parent-of-origin-dependent inheritance of familial cerebral palsy. *Hum Mol Genet* **14**, 3911-20 (2005).

16. Verkerk, A.J. *et al.* Mutation in the AP4M1 gene provides a model for neuroaxonal injury in cerebral palsy. *Am J Hum Genet* **85**, 40-52 (2009).

17. Abou Jamra, R. *et al.* Adaptor protein complex 4 deficiency causes severe autosomal-recessive intellectual disability, progressive spastic paraplegia, shy character, and short stature. *Am J Hum Genet* **88**, 788-95 (2011).

18. Moreno-De-Luca, A. *et al.* Adaptor protein complex-4 (AP-4) deficiency causes a novel autosomal recessive cerebral palsy syndrome with microcephaly and intellectual disability. *J Med Genet* **48**, 141-4 (2011).

19. Bauer, P. *et al.* Mutation in the AP4B1 gene cause hereditary spastic paraplegia type 47 (SPG47). *Neurogenetics* **13**, 73-6 (2012).

Chapter 7

A family with autism spectrum disorder and spasticity with novel compound heterozygous mutation in the sodium channel gene SCN2A

Presented as prepared manuscript for publication

7.1 Abstract

SCN2A is one of four α subunits of voltage-gated sodium channels and is abundantly expressed in the brain. Mutations in the gene have been associated with epilepsy and more recently with autism spectrum disorders. We describe three siblings with autism spectrum disorder, one of whom also had spasticity and intellectual disability and one of whom also had spasticity. Analysis of whole-exome sequencing data revealed a maternally inherited variant in *SCN2A* in all three siblings and a different paternally inherited *SCN2A* variant in the two siblings with spasticity. We hypothesize that the maternal variant contributed susceptibility to autism spectrum disorder in this family, while the paternally inherited variant contributes to spasticity in its own right or via compound heterozygosity for both maternal and paternal *SCN2A* variants. The intellectual disability in one sibling may be related to the ASD and hence to the maternal variant, or may be a variably expressed component of a 'spasticity/intellectual disability' phenotype related to the paternal variant, or may be a variably expressed feature related to the combined effect of both parental variants.

7.2 Introduction

The sodium channel, voltage-gated, type II, alpha subunit (*SCN2A*) gene encodes $Na_v1.2$, one of four α subunits of voltage-gated sodium channels responsible for the generation and propagation of action potentials in neurons and muscle. It is an important arbiter of excitability in the CNS and is critical for the upstroke of the action potential.[1] *SCN2A* is highly expressed in the brain, specifically in the cerebellum and hippocampus.[2,3] Mutations in *SCN2A* have been associated with two seizure disorders, benign familial seizures, 3 (MIM 607745)[4,5] and epileptic encephalopathy, early infantile, 11 (MIM 613721) with and without severe intellectual disability and/or motor impairment.[6-8] Recently *SCN2A* has emerged as a major gene implicated in autism spectrum disorder (ASD).[9-16] So far only single allele *SCN2A* mutations, i.e. autosomal dominant and primarily *de novo*, have been reported.[4,6-8,11,17]

ASD is a complex and highly variable neurodevelopmental disorder characterized by repetitive behaviour, impaired social skills and verbal and non-verbal communication impairments. The global prevalence for ASD is estimated to be 6.2 per 1000 individuals[18] with an excess of males to females as high as 4:1.[19] ASD is highly genetically heterogeneous, with *de novo* mutations and copy number variants accounting for 20% of cases.[13]

We report on a family with three siblings diagnosed with ASD. Two of the siblings also had spasticity, raising the possibility of cerebral palsy or hereditary spastic paraplegia (HSP) as a co-morbidity of the ASD or as an additional diagnosis. In the absence of a clear diagnosis in this family we performed whole-exome sequencing and identified a maternally inherited *SCN2A* variant in all three siblings with ASD and a different,

paternally inherited, *SCN2A* variant in the two siblings with spasticity. We postulate that the maternal variant confers susceptibility to ASD in this family, while either the paternal variant or the combination of both maternal and paternal variants contributes to spasticity. The intellectual disability in one sibling may be related to the ASD (50% of children with ASD have ID)[20] and hence to the maternal variant, or may be a variably expressed component of a 'spasticity/intellectual disability' phenotype related to the paternal variant, or may be a variably expressed feature related to the combined effect of both parental variants.

7.3 Clinical report

7.3.1 The family

We recruited five members of a non-consanguineous Caucasian family in which three siblings had ASD (II-1, II-2 and II-3, Figure 7.1). In addition to ASD, one of the three siblings (II-2, Figure 7.1) had mild spasticity and ID and one had mild spasticity (II-3, Figure 7.1). Neither parent had evidence of ASD, ID or spasticity (I-1 and I-2, Figure 7.1) and no family member had had seizures.

Approval from the Women's and Children's Health Network Human Research Ethics Committee, Adelaide, Australia and the Internal Review Board at Baylor College of Medicine, Houston, USA was obtained for this project. Written informed parental consent was obtained.

7.3.2 The parents (I-1 and 2, Figure 7.1)

The 39-year-old father and 38-year-old mother are healthy, intellectually normal, non-consanguineous Caucasians with no history of mental illness. The mother walked at

nine months and was a toe-walker until five years of age. The father was a ‘clumsy child’ but there were no other reported problems.

7.3.3 The 13-year-old son with ASD only (II-1, Figure 7.1)

The boy was born at 40+5 weeks gestation and is the eldest sibling in the family. His ASD is accompanied by a sensory processing disorder, dyslexia and attention deficit hyperactivity disorder; he does not have spasticity. Comprehensive clinical information about his early development, including birth growth parameters was not available at the time of this study.

7.3.4 The 9-year-old daughter with ASD, ID and spasticity (II-2, Figure 7.1)

The girl was born at 40+3 weeks gestation following an uncomplicated pregnancy. Birth growth parameters were weight 3460 gm, length 49 cm and head circumference 34 cm. Other than some concerns about eye contact, there were no reported medical issues in the first year of life. However, she had motor developmental delay, sitting for the first time at 11 months and walking at 2.5 years. Increased tone, limited to lower limbs, was first noticed at 18 months of age. At 2 years, Griffiths assessment indicated delays, with results on the 1st percentile for most areas of development, brain MRI was normal with no evidence of structural abnormality and multiple investigations did not provide an explanation for her neurological problems. She was subsequently diagnosed with ASD at age 5 following assessment by a multidisciplinary child development team with expertise in ASD. The girl had increased muscle tone in both lower limbs and was clumsy. She was unable to walk more than 500 m without needing to stop. She struggled with climbing stairs and could not manage larger steps. She was fully toilet trained at night well before successful toilet training during the day, with both day and

night training successful by age 5. She also lacked arm strength, particularly for turning door handles and removing lids from jars. She made little progress in cognitive development over two years and was diagnosed with a significant intellectual deficit at 9 years of age. She was in a special class at school. She had also been seen several times by a Paediatric Neurologist. There was evidence of pathologically brisk deep tendon reflexes with catch at the ankles, crossed adductor reflexes, slightly brisk finger jerks and bilateral plantar reflexes that were flexor. The differential diagnoses tendered by the Paediatric Neurologist were cerebral palsy related diplegia and hereditary spastic paraplegia (HSP).

7.3.5 The 7-year-old son with ASD and spasticity (II-3, Figure 7.1)

The boy was born at 40+5 weeks gestation following an uncomplicated pregnancy. His birth growth parameters were weight 4235gm, length 52 cm and head circumference 36 cm. His neonatal examination was normal. He sat for the first time at 8 months and did not walk until 2.5 years. He was noted at 2 years to have increased tone in his lower limbs with increased reflexes. He was formally diagnosed with ASD at age 2 years following assessment by a multidisciplinary child development team with expertise in ASD. He was in a special needs class at school. Brain MRI at 3.5 years showed a slight thickening of the corpus callosum but was otherwise normal. At age 7 years, he was less severely affected than his sister but had significant issues with climbing and was not able to run as quickly as his peers. He fell over regularly. He was described as clumsy and like his sister he also lacked arm strength, particularly for opening doors. He was not fully toilet-trained until age 5 but there was no ongoing evidence of bowel or bladder dysfunction. The boy had been seen several times by a Paediatric Neurologist. There was evidence of marginally brisk reflexes in the upper and lower limbs, without

definite evidence of overflow. Plantar reflexes were flexor. As with his sister, the differential diagnoses tendered by the Paediatric Neurologist were cerebral palsy related diplegia and HSP.

7.4 Methods

7.4.1 Whole-exome sequencing and analysis

Whole-exome sequencing was performed on DNA isolated from whole blood (Qiagen, Stanford, CA). Genomic DNA samples were constructed into Illumina paired-end pre-capture libraries according to the manufacturer's protocol (*Illumina Multiplexing_SamplePrep_Guide_ 1005361_D*) with modifications as described in the BCM-HGSC protocol (https://hgsc.bcm.edu/sites/default/files/documents/Illumina_ Barcoded _Paired-End_Capture_Library_Preparation.pdf).

Illumina sequence data was aligned to the human reference genome (HG19) with Burrows-Wheeler Aligner.[20] Variant qualities were recalibrated with GATK, variants called using ATLAS-SNP and SAMtools were used to generate BAM files[21,22] and base qualities were empirically recalibrated and realigned using the Smith-Waterman algorithm. Variants were selected by segregation analysis applying two different inheritance models; autosomal recessive (homozygous and compound heterozygous) and inherited X-linked. Variants with a minor allele frequency (MAF) of <0.1% were considered for further evaluation.

We used multiple prioritization criteria to identify potential pathogenic variants. These included: 1) type of variant (i.e. missense, frameshift and splice site) 2) MAF <0.1% in dbSNP137, 1000 genomes, exome variant server (EVS: http://evs.gs.washington.edu/EVS/) and Exome Aggregation Consortium (ExAC: http://exac.broadinstitute.org) databases, 3) in silico prediction of functional effect using PolyPhen2[23] and MutationTaster;[24] 4) high conservation score predicted by Genomic Evolutionary Rate Profiling (GERP); 5) the predicted effect of haploinsufficiency,[25] 6)

brain expression pattern (Allen Human Brain Atlas; www.brainspan.org/) and 7) previous disease association as reported in Online Mendelian Inheritance in Man (OMIM). Variants were further filtered based on a gene-based intolerance ranking system, Residual Variation Intolerance Score (RVIS)[26] and Combined Annotation-Dependent Depletion (CADD);[27] with cut-offs for RVIS<25th percentile and CADD<10.

7.4.2 Sanger sequencing

To confirm the mutation in this family two primer pairs incorporating the contributing mutation from each parent were designed using Primer3 (v.0.4.0) (HG19). Primer set 1: (forward 5'-TGGCATTCTGCATGACATTT-3' forward, reverse 5'- ATTTCG TAGGTTGCCCATGA-3') and primer set 2: (forward 5'-CCTATTCTTAATA GTGGACCT-3', reverse 5'- CAACGTGGTCGTAATGGGC TCA-3'). Sanger sequencing was performed using BigDye terminator chemistry 3.1 (ABI) and sequenced using an ABI prism 3100 genetic analyser (Applied Biosystems, Foster City, CA, USA). Sequencing data was analysed using DNASTAR Lasergene 10 Seqman Pro (DNASTAR, Inc. Madison, Wi, USA).

7.5 Results

Analyses of whole-exome sequencing data for all five family members identified two different novel missense variants in *SCN2A* segregating within the family (Figure 7.1). The maternal variant (NM_001040143: exon 26: c.5549A>G: HG19 chr2:166245865) predicts a p.Asp1850Gly change and the paternal variant (NM_ 001040143: exon 6: c.776T>C: HG19 chr2:166166911) predicts a p.Leu259Pro change. The maternal variant was present in all three siblings with ASD. The paternal variant was present in only the two siblings (II-2 and II-3, Figure 7.1) with both ASD and spasticity, one of whom also had ID (II-2, Figure 7.1). Both SCN2A amino acid residues, p.Leu259 and p.Asp1850, are predicted to be intolerant to functional variation, with a RVIS below the 25th percentile, and have a CADD score above the 10th percentile. Both were also predicted to be damaging by PolyPhen2 and Mutation Taster and have been highly conserved (invariable) throughout evolution (Tables 7.1 and 7.2). Both variants were novel when checked against dbSNP, 1000 Genomes, EVS and the ExAC database (see URLs). Sanger sequencing confirmed the results and segregation of variants in this family (Figures 7.1 and 7.2). No other variants that would explain the phenotype in the family, including variants in currently known HSP genes, or genes associated with cerebral palsy, were identified.

7.6 Discussion

We report three siblings with ASD, one of whom had mild spasticity and ID and one of whom had mild spasticity. Whole exome sequencing identified an inherited maternal *SCN2A* variant in all three siblings with ASD, and an inherited paternal *SCN2A* variant in the two siblings with spasticity. Both variants were predicted to be damaging to SCN2A protein function.

Previously reported *SCN2A* mutations have followed an autosomal dominant mode of inheritance and have been associated with seizures and, more recently, ASD. In this family, neither parent had features of ASD and none of the family members had seizures. The maternal variant was present in all three siblings with ASD but it is difficult to quantify the role of the *SCN2A* variant in ASD causation in this family given the multifactorial nature of ASD. It is possible that this variant is one factor related to ASD, due to variable expressivity and incomplete penetrance. The phenomenon of incomplete penetrance has been described previously for many monogenic disorders including disorders caused by sodium channel genes, eg. *SCN1A*.[28] In our family we have in addition to incomplete penetrance, another phenotype, spasticity, in the two siblings (II-2 and II-3, Figure 7.1) with the paternally inherited *SCN2A* variant. We postulate that their spasticity is either a result of compound heterozygosity for the *SCN2A* variants or secondary to the single paternally inherited *SCN2A* variant, invoking incomplete penetrance given the father's normal phenotype. To our knowledge compound heterozygosity for *SCN2A* mutations has not yet been described. The story is further complicated with the diagnosis of ID in the daughter, which as mentioned earlier may be related to the ASD (50% of children with ASD have ID)[20] and hence to the maternal variant, or may be a variably expressed component of a 'spasticity/intellectual

disability' phenotype related to the paternal variant, or may be a variably expressed feature related to the combined effect of both parental variants. Another consideration is additional unidentified variants predisposing to ID not present in the other two siblings and not detected as they were outside the exome-capture design used in this study.

The p.Leu259 residue is invariant across species and located within the 5th transmembrane domain of SCN2A flanked by two mutations (p.Met252Val and p.Val261Met) reported to cause benign familial neonatal infantile seizures.[29] The p.Asp1850 residue, located within the cytoplasmic domain of SCN2A is also invariant across species. This domain of SCN2A has no specific function assigned to it as yet; recently however, two *de novo* mutations within this region, p.Arg1882Gln and p.Arg1882Leu have been reported in epileptic encephalopathy.[14,30] Taken together we propose that both *SCN2A* variants found in the family are deleterious and disease-causing.

SCN2A demonstrates considerable mutation pleiotropy and is best known for its association with epilepsy syndromes.[4-8] However, in several recent studies, ASD with and without ID and without any history of seizures has been associated with *SCN2A* mutations.[12-16] In one such study, a young boy with ASD and mild dysmorphic features was also diagnosed with abnormal motor abnormalities, including mild hypotonia and abnormal waddling gait, but no spasticity was reported.[13]

The $Na_V1.2$ channel is found in the axon initial segment of excitatory neurons where it is critical for action potential generation.[31] More recent data also suggests that Nav1.2 is also found in the axon initial segment of a subclass of GABAergic neurons suggesting a

more complex neurophysiological role than originally thought.[32] Because of this expression profile, both gain and loss of function could impact neuronal network excitability and contribute to disease phenotypes. In the study of Li *et al*[38] the authors found that reductions in $Na_V1.2$ mediated current could increase recurrent network activity and could give rise to hyperexcitability. Conversely gain of function mutations could directly increase the activity of excitatory pyramidal neurons and produce similar hyperexcitability phenotypes. Functional analysis of the variants identified using *in vitro* and *in vivo* models will be needed to fully describe pathomechanisms for spasticity.

Both cerebral palsy and HSP had been considered in the clinical diagnosis for this family. Cerebral palsy is an umbrella term applied to a clinically heterogeneous group of non-progressive disorders, primarily affecting the movement and posture.[33] It is the most frequent cause of physical disability in childhood with a prevalence of 2-2.5 per 1000 live births.[34] Co-morbidities can include ID, ASD spectrum disorder, epilepsy, speech and language deficits, and visual and hearing impairments.[33]

The hereditary spastic paraplegias are monogenic disorders primarily resulting in progressive lower limb spasticity and pyramidal weakness.[38,39] Depending on the clinical manifestations, HSP is classified as either pure or complex. HSP is considered uncomplicated (pure) if limited to lower limb spastic weakness; other features can include pes cavus, urinary bladder disturbance and mild cognitive decline.[40] HSP is considered complex when spastic paraplegia is accompanied by clinical features such as ID, epilepsy, amyotrophy, ataxia, extrapyramidal disturbance and peripheral neuropathy.[40,41] Onset is from early childhood through to 70 years of age.[40] HSP is

genetically heterogeneous with over 70 genetic types documented to date. The majority are classified by their genetic loci ([SPG] 1-71) in order of their discovery.[41,42] Several genetic types are associated with both pure and complicated HSP and the disorder can be inherited either as an autosomal dominant, recessive or X-linked trait.[38,41] The majority of genes associated with HSP are involved in cellular transport, synapse and axon development, and nucleotide metabolism pathways.[42] Age of onset, progression of the disease, other clinical features and presence of family history are important clues in determining between a diagnosis of cerebral palsy and HSP. Spasticity is the primary overlapping clinical feature of cerebral palsy and HSP but the significant difference between these disorders is whether the spasticity is non-progressive or progressive.

For the children in this study, their young age, the mild degree of spasticity and no known family history of either disorder made an accurate diagnosis at this stage difficult. They are still too young to know whether the spasticity will be progressive but the more recent diagnosis of ID in the 9-year-old girl raises the possibility of a slowly progressive disorder in this family. Only further investigation of the children as they increase in age will determine whether the spasticity is static or progressive. The genetic heterogeneity of HSP is well documented with over 70 genes to date.[41,42] There is also increasing evidence that cerebral palsy is genetically heterogeneous, likely reflecting its clinical heterogeneity as highlighted in a recent study where potentially pathogenic variants were identified in 14% of cases in both novel and previously known disease genes.[43] Mutations in *SCN2A* have not been previously reported with either cerebral palsy or HSP and in the absence of a clear diagnosis from the clinical phenotype we could not confidently associate *SCN2A* with either disorder.

The genetics within this family may be more complex than just the interaction of the two *SCN2A* variants. However, the idea that the maternal variant is a susceptibility factor for ASD and that the paternal variant is a susceptibility factor for spasticity is plausible. The maternal variant may also be a susceptibility factor for spasticity, but the spasticity phenotype may only be seen in the presence of homozygous or compound heterozygous variants in *SCN2A*. Alternatively, the paternal variant could be considered recessive, being associated with a phenotype only when accompanied by a pathogenic variant in the second allele. The story is further complicated with the diagnosis of ID in the daughter, which as mentioned earlier may be related to the ASD[20] and hence related to the maternal variant or may be a component of a 'spasticity/ID' phenotype related to the paternal variant. Another consideration is additional unidentified variant/s predisposing to ID not present in the other two siblings and not detected as they were outside the exome-capture design used in this study.

It may be that movement disorders are a yet unappreciated phenotype for *SCN2A* mutations. As far as we are aware this is the first study raising the possibility that *SCN2A* mutation(s) may be associated with movement disorders; if confirmed, this would further expanding the phenotypes associated with *SCN2A* mutations.

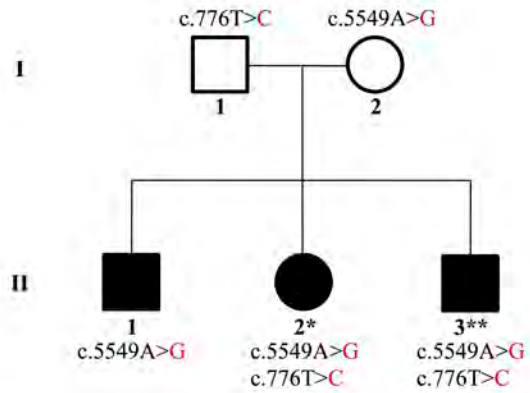


Figure 7.1 Two-generation pedigree. The maternal variant (c.5549A>G) was common to the three siblings diagnosed with ASD. *Represents the individual diagnosed with spasticity and ID and **represents the individual diagnosed with spasticity, both carrying the paternal variant (c.776T>C). Both parents are unaffected.

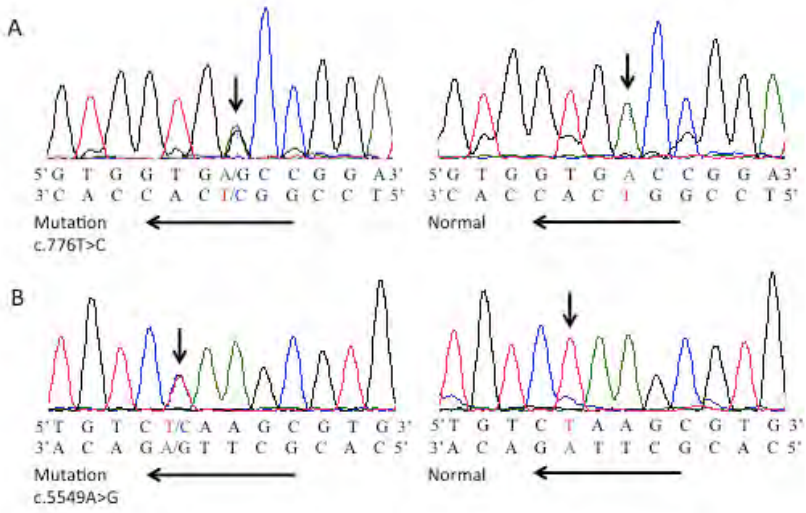


Figure 7.2 Fragments of sequence chromatograms from 5'-3'. (A) Paternally inherited variant (c.776T>C) and normal variant and (B) Maternally inherited variant (c.5549A>G) and normal variant.

Table 7.1 Clinical details, *SCN2A* variants and segregation pattern and pathogenicity predictions and scores.

Sample	Clinical details	cDNA	Protein change	PPH2	MT	GERP	RVIS %	CADD
Father I-1	Unaffected	c.776T>C	p.Leu259Pro	0.999	1	4.09	1.77	16.10
Mother I-2	Unaffected	c5549A>G	p. Asp1850Gly	1	0.5	5.53	1.77	16.97
Male sibling II-1	ASD	c5549A>G	p. Asp1850Gly	1	0.5	5.53	1.77 1.77	16.97
Female sibling II-2	ASD, ID, spasticity	c5549A>G c.776T>C	p. Asp1850Gly p.Leu259Pro	1 0.999	0.5 1	5.53 4.09	1.77 1.77	16.97 16.10
Male sibling II-3	ASD, spasticity, sensory processing disorder, dyslexia, ADHD	c5549A>G c.776T>C	p. Asp1850Gly p.Leu259Pro	1 0.999	0.5 1	5.53 4.09	1.77 1.77	16.97 16.10

Abbreviations: ASD – autism spectrum disorder, HSP – hereditary spastic paraplegia, ADHD – attention deficit hyperactivity disorder, ID – intellectual disability, PPH2- polyphen2, MT – Mutation Taster, RVIS – residual variance intolerance score, CADD - combined annotation-dependent depletion

Table 7.2 Multispecies alignment across nine different species for maternal and paternal variants.

Species	Maternal: p.Asp1850Gly							Paternal: p.Leu257Pro						
Mutant	V	S	G	**G**	R	I	H	V	F	C	**P**	S	V	F
Human	V	S	G	D	R	I	H	V	F	C	L	S	V	F
Rhesus	V	S	G	D	R	I	H	V	F	C	L	S	V	F
Mouse	V	S	G	D	R	I	H	V	F	C	L	S	V	F
Dog	V	S	G	D	R	I	H	V	F	C	L	S	V	F
Elephant	V	S	G	D	R	I	H	V	F	C	L	S	V	F
Chicken	V	S	G	D	R	I	H	V	F	C	L	S	V	F
X-tropicalis	V	S	G	D	R	I	H	V	F	C	L	S	V	F
Zebrafish	V	S	G	D	R	I	H	V	F	C	L	S	V	F
Lamprey	V	I	N	D	R	I	H	V	F	C	L	S	V	F

7.7 References

1. Shi, X., Yasumoto, S., Kurahashi, H. *et al.* Clinical spectrum of scn2a mutations. *Brain Dev* **34**, 541-5 (2012).
2. Schaller, K.L. & Caldwell, J.H. Expression and distribution of voltage-gated sodium channels in the cerebellum. *Cerebellum* **2**, 2-9 (2003).
3. Schaller, K.L. & Caldwell, J.H. Developmental and regional expression of sodium channel isoform nach6 in the rat central nervous system. *J Comp Neurol* **420**, 84-97 (2000).
4. Berkovic, S.F., Heron, S.E., Giordano, L. *et al.* Benign familial neonatal-infantile seizures: Characterization of a new sodium channelopathy. *Ann Neurol* **55**, 550-7 (2004).
5. Herlenius, E., Heron, S.E., Grinton, B.E. *et al.* Scn2a mutations and benign familial neonatal-infantile seizures: The phenotypic spectrum. *Epilepsia* **48**, 1138-42 (2007).
6. Kamiya, K., Kaneda, M., Sugawara, T. *et al.* A nonsense mutation of the sodium channel gene scn2a in a patient with intractable epilepsy and mental decline. *J Neurosci* **24**, 2690-8 (2004).
7. Liao, Y., Anttonen, A.K., Liukkonen, E. *et al.* Scn2a mutation associated with neonatal epilepsy, late-onset episodic ataxia, myoclonus, and pain. *Neurol* **75**, 1454-8 (2010).
8. Ogiwara, I., Ito, K., Sawaishi, Y. *et al.* De novo mutations of voltage-gated sodium channel alphaii gene scn2a in intractable epilepsies. *Neurol* **73**, 1046-53 (2009).

9. Samocha, K.E., Robinson, E.B., Sanders, S.J. *et al.* A framework for the interpretation of de novo mutation in human disease. *Nat Genet* **46**, 944-50 (2014).

10. Weiss, L.A., Escayg, A., Kearney, J.A. *et al.* Sodium channels scn1a, scn2a and scn3a in familial autism. *Mol Psychiatry* **8**, 186-94 (2003).

11. Krumm, N., O'Roak, B.J., Shendure, J. & Eichler, E.E. A de novo convergence of autism genetics and molecular neuroscience. *Trends Neurosci* **37**, 95-105 (2014).

12. Sanders, S.J., Murtha, M.T., Gupta, A.R. *et al.* De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237-41 (2012).

13. Tavassoli, T., Kolevzon, A., Wang, A.T. *et al.* De novo scn2a splice site mutation in a boy with autism spectrum disorder. *BMC Med Genet* **15**, 35 (2014).

14. Baasch, A.L., Huning, I., Gilissen, C. *et al.* Exome sequencing identifies a de novo scn2a mutation in a patient with intractable seizures, severe intellectual disability, optic atrophy, muscular hypotonia, and brain abnormalities. *Epilepsia* **55**, e25-9 (2014).

15. Celle, M.E., Cuoco, C., Porta, S., Gimelli, G. & Tassano, E. Interstitial 2q24.3 deletion including scn2a and scn3a genes in a patient with autistic features, psychomotor delay, microcephaly and no history of seizures. *Gene* **532**, 294-6 (2013).

16. Jiang, Y.H., Yuen, R.K., Jin, X. *et al.* Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. *Am J Hum Genet* **93**, 249-63 (2013).

17. Samocha, K.E., Robinson, E.B., Sanders, S.J. *et al.* A framework for the interpretation of de novo mutation in human disease. *Nat Genet* **46**, 944-50 (2014).

18. Elsabbagh, M., Divan, G., Koh, Y.J. *et al.* Global prevalence of autism and other pervasive developmental disorders. *Autism Res* **5**, 160-79 (2012).

19. Newschaffer, C.J., Croen, L.A., Daniels, J. *et al.* The epidemiology of autism spectrum disorders. *Ann Rev Public Health* **28**, 235-58 (2007).

20. Li, H. & Durbin, R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* **25**, 1754-60 (2009).

21. DePristo, M.A., Banks, E., Poplin, R. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**, 491-8 (2011).

22. Li, H., Handsaker, B., Wysoker, A. *et al.* The sequence alignment/map format and samtools. *Bioinformatics* **25**, 2078-9 (2009).

23. Adzhubei, I., Jordan D. M. & R., S.S. Predicting functional effect of human missense mutations using polyphen-2. *Curr Protoc Hum Genet* **Chapter 7**, 7-20 (2013).

24. Schwarz, J.M., Rodelsperger, C., Schuelke, M. & Seelow, D. Mutationtaster evaluates disease-causing potential of sequence alterations. *Nat Methods* **7**, 575-6 (2010).

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

26. Petrovski, S., Wang, Q., Heinzen, E.L., Allen, A.S. & Goldstein, D.B. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet* **9**, e1003709 (2013).

27. Kircher, M., Witten, D.M., Jain, P. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310-5 (2014).

28. Mhanni, A.A., Hartley, J.N., Sanger, W.G., Chudley, A.E. & Spriggs, E.L. Variable expressivity of a novel mutation in the scn1a gene leading to an autosomal dominant seizure disorder. *Seizure* **20**, 711-2 (2011).

29. Liao, Y., Deprez, L., Maljevic, S. *et al.* Molecular correlates of age-dependent seizures in an inherited neonatal-infantile epilepsy. *Brain* **133**, 1403-14 (2010).

30. Carvill, G.L., Heavin, S.B., Yendle, S.C. *et al.* Targeted resequencing in epileptic encephalopathies identifies de novo mutations in chd2 and syngap1. *Nat Genet* **45**, 825-30 (2013).

31. Hu, W., Tian, C., Li, T. *et al.* Distinct contributions of na(v)1.6 and na(v)1.2 in action potential initiation and backpropagation. *Nat Neurosci* **12**, 996-1002 (2009).

32. Li, T., Tian, C., Scalmani, P. *et al.* Action potential initiation in neocortical inhibitory interneurons. *PLoS Biol* **12**, e1001944 (2014).

33. Rosenbaum, P., Paneth, N., Leviton, A. *et al.* A report: The definition and classification of cerebral palsy. *Dev Med Child Neurol Suppl* **109**, 8-14 (2007).

34. Stanley F., B.E., Alberman B. How common are the cerebral palsies? *In: Stanley F, Blair E, Alberman B, editors. Cerebral Palsies: Epidemiology &Causal Pathways. London (UK): Mac Keith Press,* p.22-48 (2000).

35. O'Callaghan, M.E., MacLennan, A., Gibson, C. *et al.* Epidemiolgic associations with cerebral palsy. *Obstet Gynaecol* **118**, 576-582 (2011).

36. Moreno-De-Luca, A., Ledbetter, D.H. & Martin, C.L. Genomic insights into the causes and classification of the cerebral palsies. *The Lancet Neurol* **11**, 283-92 (2012).

37. Colver, A., Fairhurst, C. & Pharoah, P.O. Cerebral palsy. *Lancet* **383**, 1240-9 (2014).

38. Harding, A.E. Hereditary spastic paraplegias. *Semin Neurol* **13**, 333-6 (1993).

39. Shoukier, M., Neesen, J., Sauter, S.M. *et al.* Expansion of mutation spectrum, determination of mutation cluster regions and predictive structural classification of spast mutations in hereditary spastic paraplegia. *Eur J Hum Genet* **17**, 187-94 (2009).

40. Salinas, S., Proukakis, C., Crosby, A. & Warner, T.T. Hereditary spastic paraplegia: Clinical features and pathogenetic mechanisms. *Lancet Neurol* **7**, 1127-38 (2008).

41. Fink, J.K. Hereditary spastic paraplegia: Clinico-pathologic features and emerging molecular mechanisms. *Acta Neuropathologica* **126**, 307-28 (2013).

42. Novarino, G., Fenstermaker, A.G., Zaki, M.S. *et al.* Exome sequencing links corticospinal motor neuron disease to common neurodegenerative disorders. *Science* **343**, 506-11 (2014).

43. McMichael, G., Bainbridge, M., Haan, E., Corbett, M., Gardner, A., Thompson, S., van Bon, BWM., van Eyk, CL., Broadbent, J., Reynolds, C., O'Callaghan, ME., Nguyen, LSl, Adelson, DL., Russo, R., Jhangiani, S., Doddapaneni, H., Muzny, DM., Gibbs, RA., Gecz, J., MacLennan, AH. Whole-exome sequencing points to considerable genetic heterogeneity of cerebral palsy. *Mol Psychiatry*, 1-7 (2015).

Chapter 8
Discussion and future directions

8.1 Discussion and future prospects

The main goal of genetic evaluation of individuals living with cerebral palsy is to understand the underlying origin of the disorder, so that those affected, their families and the wider community can benefit from further research leading to prevention, better management and ultimately, treatment. The advent of next generation sequencing technologies (NGS), in particular massively parallel sequencing, is a promising way forward in identifying the causes of, or contribution of genetic variation to, cerebral palsy. These technologies have been successful in increasing understanding of the causes of neurodevelopmental disorders whose phenotypes overlap with cerebral palsy (e.g. intellectual disability,[1,2] autism[3-5] and epilepsy[6-8]), and provide an excellent way forward for cerebral palsy genetic research. However, they also bring with them new challenges, including determining which variants/genes are potentially pathogenic for cerebral palsy.[9]

We took an unbiased whole-exome sequencing approach to identify potential pathogenic variants for cerebral palsy in two groups of individuals with cerebral palsy: sporadic cases (no previous family history of cerebral palsy) and families with more than one individual with a confirmed diagnosis of cerebral palsy. Until our work, WES had only been performed in a small number of familial cases[10-12]. Our overall findings suggest that cerebral palsy is genetically heterogeneous, involving mutations, mainly *de novo*, in previously known neurodevelopmental genes and novel candidate cerebral palsy genes. The *de novo* origin of these mutations may explain the typically sporadic nature of cerebral palsy, with most affected individuals having no apparent family history. *De novo* mutations are also associated with a sizable proportion of sporadic cases of ID and ASD and represent an important cause for both disorders.[1,13,14]

8.2 Sporadic cases

Chapter 4 reports on WES of 183 sporadic cases comprising: 98 cases and both parents (trios), 67 cases and one parent (duos), and 18 singleton cases (no parental DNA). High quality variants were annotated with their status in parental samples, when available. Variants were considered *de novo* in the child if the variant was not present in either parent. A gene was considered to contain compound heterozygous variants if two variants in the gene could not be accounted for by a single parent.

We identified and validated 61 *de novo* protein-altering variants in 43 out of 98 (44%) case-parent trios. Following a strict set of multidimensional prioritization criteria (Materials and methods, **Chapter 3.9**), we identified 14 out of the 98 trio cases (14%) with a variant considered to be potentially causative for cerebral palsy. Ten per cent of cases had a *de novo* mutation and 4% had an X-chromosome variant; half of these variants were in novel genes. This is the first time these genes have been associated with cerebral palsy. Of the ten *de novo* mutations predicted to be causative for cerebral palsy, four occurred in genes associated with neurological disorders with phenotypic characteristics that overlap with cerebral palsy (*TUBA1A* (n=2), *SCN8A* (n=1) and *KDM5C* (n=1)) and six occurred in novel candidate cerebral palsy genes (*AGAP1*, *JHDM1D*, *MAST1*, *NAA35*, *RFX2* and *WIPI2*). In addition to the *de novo* mutations, four maternally inherited potentially pathogenic X-linked variants, including two in known disease genes (*PAK3* and *L1CAM*) and two in novel candidate genes (*CD99L2* and *TENM1*), were found in four male cases. *PAK3* and *L1CAM*, implicated in cerebral palsy from our trio cases, had one additional potentially pathogenic variant each in two of the 67 duo cases.

Mutations in *TUBA1A* and *L1CAM* have been previously reported in known genetic disorders with clinical features said to mimic cerebral palsy. *De novo* mutations in *TUBA1A* have been reported in a wide spectrum of neuronal migration disorders.[15] Spastic diplegia or quadriplegia and intellectual disability are common features in *TUBA1A* mutations.[15-17] Two cases in the 98 trio cases, one female and one male, had different novel missense mutations in *TUBA1A*. Both individuals were diagnosed with diplegic cerebral palsy; the male also had intellectual disability, seizures, adducted thumbs and optic atrophy. Brain imaging for both individuals showed tubulin associated anomalies such as cerebellar and corpus callosum anomalies.

L1CAM is involved in neurite outgrowth and when mutated, is associated with L1 syndrome, a disorder with variable features including hydrocephalus, spastic paraplegia, adducted thumbs and agenesis of the corpus callosum.[18] In both individuals with potentially pathogenic *L1CAM* variants, the clinical phenotype and brain imaging results were compatible with L1 syndrome, but L1 syndrome was not recognised in either case in the neonatal period.

There has been considerable discussion in the literature on genetic disorders that masquerade as cerebral palsy and among these are disorders associated with mutations in *TUBA1A* and *L1CAM.*[19] Had these cases been sequenced prior to a diagnosis of cerebral palsy, they may have been diagnosed with neuronal migration disorder and L1 syndrome respectively; that being said, these results do not negate a diagnosis of cerebral palsy, as different mutations within the same gene can result in variable phenotypes. It may be that cerebral palsy is a yet unrecognised phenotype of mutations in *TUBA1A* and *L1CAM*.

For 44% of the 98 trio cases the candidate variant was unresolved. These included several loss of function (LOF) variants in genes not yet known to cause a disease when mutated, including one stop gain p.W113X in *ENPP4,*the function of which is currently unknown, and a splice site at c.114-1G>A in *MIIP,* whose protein product inhibits cell migration.[20] Three frame shift deletions were also identified, each predicted to result in a premature stop codon: p.E1289fs* in *LTN1,* a ring finger protein which functions as an E3 ubiquitin ligase;[21] p.Q189fs* in *UBQLN3* which encodes an ubiquitin-like protein specifically expressed in the testis and thought to regulate cell-cycle progression during spermatogenesis[22]; and p.D324fs* in *NEMF,* which plays a role in nuclear export. These genes either had a high RVIS percentile or there were other frequent LOF variants in these genes in EVS. Given the nature of the EVS resource we could not determine if any of these LOF variants were associated with a movement phenotype.

In addition to several unresolved LOF variants, we found two males with maternally inherited X-linked variants in known intellectual disability genes: p.L107V in *CUL4B,* which encodes a member of the cullin-RING E3 ubiquitin ligase complex[23]; and a splice site mutation c.1137-1G>A in *MAOB.* Deletions involving *MAOA* and *MAOB* have been reported in males with severe intellectual disability.[24] We also identified a missense mutation, p.T167A, in *ZNF674,* mutations in which had previously been associated with intellectual disability. However, it has recently become apparent that stop gains, deletions and other variants in the *ZNF674* gene are not only found in intellectual disability cases but also in controls, and is considered benign.[25]

Further, several variants of unknown inheritance were identified in the 68 duo and 18 singleton cases that may be of particular interest for cerebral palsy. These included: two

compound heterozygous variants in two of the 68 duo cases, p.M109L and p.L618V in *MED17*, which is associated with primary microcephaly[26]; and p.S44L and p.E531K in *SPAST*, a known gene for hereditary spastic paraplegia (HSP)[27]. Two additional cases with variants with high pathogenicity and conservation scores were identified in *SPAST* (p.D581N, p.E432V) in two of the 18 singleton cases. *SPAST* codes for spastin, an ATP-dependent microtubule severing protein.[27] Mutations in *SPAST* cause SPG4, the most common form of HSP. The characteristic clinical feature of HSP is progressive bilateral weakness and spasticity of the lower limbs[27], whereas historically cerebral palsy has been defined as a non-progressive disorder.[28] These findings raise the possibility that these cases were misdiagnosed as cerebral palsy or, alternatively, that *SPAST* mutations can be associated with non-progressive early-onset spastic paraplegia. In another singleton case, an as yet unreported variant p.G119R in *SYNGAP1* was detected. *SYNGAP1* encodes a protein involved in synaptic plasticity and has previously been reported to be mutated in intellectual disability, autism and schizophrenia.[29]

If we conservatively consider the 14% of cases with predicted pathogenic variants, we ought to postulate a considerably heterogeneous genetic contribution to cerebral palsy causation, involving at least five known disease genes and eight novel candidate cerebral palsy genes. This is in addition to the currently known list of cerebral palsy genes.[30] From the clinical perspective we saw a trend towards a higher number of potentially pathogenic *de novo* variants in cases born at term compared to preterm cases. No other correlations including severity of cerebral palsy, type of cerebral palsy and/or cerebral palsy with other comorbidities were identified.

8.3 Familial cases

Chapters 5 and **6** report on two non-consanguineous families with more than one individual with an initial diagnosis of cerebral palsy. **Chapter 7** reports on a non-consanguineous family where the diagnosis of cerebral palsy is not as clear. Familial cases with identical or similar clinical features increase the power for gene discovery.[9] Sequencing multiple family members and including both affected and unaffected individuals can reduce the number of potential variants to a few.[31]

*8.3.1 Family 1 with ZC4H2 mutation (reported in **Chapter 5**)*

We sequenced the exomes for five out of 10 individuals in a two-generation non-consanguineous family with two male siblings, (II-1 and II-5) with an initial diagnosis of quadriplegic cerebral palsy and diplegic cerebral palsy, respectively. The mother was born in Uruguay to parents of Spanish decent and the father was born in Chile to parents of Chilean and German decent. We used Sanger sequencing for the remaining family members (with the exception of youngest child (II-8) in **Figure 5.1**). The initial diagnosis of cerebral palsy in this family was made when the first affected boy was young but as the boys increased in age, the progressive nature of the disorder became apparent and the diagnosis of cerebral palsy became inappropriate. In addition, both boys had intellectual disability, hypotonia, mild dystonia, equinovarus feet and a history of seizures. We identified a maternally inherited X-chromosome variant in the X-linked zinc-finger gene *ZC4H2* segregating within this family. The mother (I-2 in **Figure 5.2**) and two daughters (II-4 and II-7 in **Figure 5.2**) each carried one wild type copy of the gene and one copy of the variant. The two affected boys (II-1 and II-5 in **Figure 5.2**) were hemizygous for the variant. One of the three carrier females (II-7 in **Figure 5.21**) was significantly behind in her language, cognitive and personal/social skills. The

remaining family members tested (I-1, II-2, II-3, II-6 in **Figure 5.2**) carried wild type allele/alleles.

Subsequently, another group who had identified other families with mutations in *ZC4H2* contacted us to partake in functional studies.[32] Four different mutations in *ZC4H2* were identified in the five families, which included the family in our study. Motor dysfunction, muscle weakness, intellectual disability and seizures were common features amongst these families. *In vitro* and *in vivo* functional studies confirmed the pathogenicity of these mutations.[32]

8.3.2 *Family 2 with NKX2-1 mutation* (reported in **Chapter 6**)

The affected family members of a non-consanguineous Caucasian family are a father, his son and his daughter (II-1, III-2 and III-3 in **Figure 6.1**). A diagnosis of ataxic cerebral palsy was made in infancy in the son and daughter based on the presence of delayed motor milestones and a non-progressive movement disorder, ataxic in nature, associated with gait disturbance and frequent falls. However, as the children grew older, choreoathetosis became the predominant neurological feature, although ataxia persisted through the first decade. It was recognised that the diagnosis was benign hereditary chorea when the children were reviewed by a neurologist in their early teenage years, after our studies were initiated. The father's neurological condition had not been appreciated prior to examination of the affected son. Whole-exome sequencing was performed for the three affected individuals (II-1, III-2 and III-3 in **Figure 6.1**) and the two unaffected individuals in the second and third generation (II-2 and III-1 in **Figure 6.1**). A 7 bp deletion in exon 1 of *NKX2-1*, resulting in a frame shift and creation of a premature termination codon, was identified in all affected individuals (**Figure 6.2A**).

Benign hereditary chorea, caused by mutations in the NK2 homeobox 1 gene (*NKX2-1*), shares clinical features with ataxic and dyskinetic cerebral palsy, resulting in the possibility of misdiagnosis.[33,34]

*8.3.3 Family 3 with SCN2A mutation (reported in **Chapter 7**)*

We sequenced the exomes for five members of a non-consanguineous Caucasian family with three siblings diagnosed with ASD (II-1, II-2 and II-3 in **Figure 7.1**), one of whom also had spasticity and intellectual disability. (II- 2 in **Figure 7.1**) and one of whom also had spasticity (II-3 in **Figure 7.1**). Due to the young age of the children it was not clear whether the spasticity was progressive or not, making it difficult to distinguish between cerebral palsy and hereditary spastic paraplegia (HSP) in this family. However, the more recent diagnosis of ID in the 9-year-old girl raises the possibility of a slowly progressive disorder in this family. Neither parent had been diagnosed with ASD, spasticity or ID. Analysis of WES data revealed a maternally inherited variant in *SCN2A* in all three siblings and a different paternally inherited *SCN2A* variant segregating in the siblings with spasticity. Both variants were predicted to be damaging. The presence of the *SCN2A* maternal variant could potentially confer susceptibility to ASD in this family, while the presence of either the paternal variant alone or the combination of both maternal and paternal *SCN2A* variants, could potentially contribute to the spasticity. The ID in one sibling may be related to the ASD (50% of children with ASD have ID)[35] and hence to the maternal variant, or may be a variably expressed component of a 'spasticity/intellectual disability' phenotype related to the paternal variant, or may be a variably expressed feature related to the combined effect of both parental variants.

SCN2A demonstrates considerable mutation pleiotropy and is best known for its association with epilepsy syndromes.[33-37] In several recent studies, ASD without any history of seizures has been associated with *SCN2A* mutations.[38-42] However, to the best of our knowledge this is the first study raising the possibility that *SCN2A* mutation(s) may be associated with movement disorders. If confirmed, these findings would expand the phenotypes associated with *SCN2A* mutations. Two potentially pathogenic mutations for *SCN2A* were found segregating within this family. However, as mutations in *SCN2A* have not previously been associated with either cerebral palsy or HSP, the diagnosis for the spasticity in this family remains uncertain. Further investigation of the children as they increase in age will determine whether the spasticity is progressive or non-progressive. Identification of other cases with a similar phenotype with mutations in *SCN2A* is also likely to assist in the diagnostic process. It may be that movement disorders are a yet unappreciated phenotype for *SCN2A* mutations. One example where different mutations in the same gene have resulted in autosomal dominant as well as autosomal recessive disease (homozygous or compound heterozygous) is *TBC1D24*.[36] Mutations in *TBC1D24* are associated with infantile myoclonic epilepsy,[37] infantile epileptic encephalopathy,[38,39] DOORS (deafness, onychodystrophy, osteodystrophy, intellectual disability and seizures)[40] and NSHL (non-syndromic hearing loss).[36] For NSHL dominant and recessive mutations have been identified.[41,42]

The diagnosis of cerebral palsy/spasticity for all three families was made in the first few years of life. Cerebral palsy is sometimes a 'working diagnosis' and the affected child needs to be reviewed to determine whether or not the spasticity is progressive. Also,

assessment of young children by a neurologist may sometimes identify features about the child's movements that result in a specific alternative diagnosis.

8.4 Limitations of this study

Since 2007 there has been enormous progress in the development of exome capture technologies. Earlier capture kits achieved 80% coverage of the targeted bases and today ≥95% of targeted bases are captured. The capture design (HGSC VCRome 2.1 design; 42Mb, NimbleGen, Madison, WI, USA) used in our study achieved an average of 92% of the targeted exome bases covered.[43] There is no single exome platform providing 100% coverage and all exome capture designs, including the one used in this study, have limitations: 1) Capture probes can only target the exons that are known to date; 2) exons buried in stretches of repeats out towards the chromosome tips are not part of the capture design; 3) sequences with a high GC or AT content may not be captured efficiently;[44] 4) capture efficiency is not uniform across the exons[44]; and 5) not all sequences can be aligned back to the reference genome.

The capture design used in our study (HGSC VCRome 2.1 design; 42Mb, NimbleGen, Madison, WI, USA) included 8 Mbp of randomly selected UTR exons,[45] and therefore the majority of UTRs and non-coding regions were not assessed for potential disease-causing variants. Copy number variants are also likely to be important in cerebral palsy causation but were not assessed using the sequencing data at the time of my PhD project but this work is currently being done. Finally, our study was not large enough to examine the complexity of the contribution of oligogenic, polygenic and/or environmental factors such as illicit drug use,[46] smoking,[46] chorioamnionitis,[47] fever

during labour[48] and infection[49] for cerebral palsy. At least a proportion of cerebral palsy cases will be explained by more complex genetics and not just single major gene effect.

8.5 Future prospects

This study highlighted the genetic heterogeneity and complexity of cerebral palsy and demonstrated the utility of NGS as a gene discovery tool for cerebral palsy. Compared to other neurodevelopmental disorders such as intellectual disability, autism and epilepsy, genetic research for cerebral palsy is very much in its infancy and what we know so far is only the tip of the iceberg. Independent replication studies using WES or WGS in cerebral palsy cases will be crucial, not only to confirm the candidate genes identified in this project but also for understanding the extent of the genetic involvement and which genes are the main contributors. Such studies are already happening. The cost of WES has dramatically dropped since its début in the commercial market in 2008-2009 and over the course of this study, making it more accessible than ever. Until recently the cost of WES has been significantly less than WGS, making WES a more favoured platform for routine use for the detection of common and rare variants. As the cost of WGS continues to fall (currently, WGS costs $1,700 and WES costs $1,100 the cost of WGS is predicted to be comparable to that of WES in the near future), it is becoming an attractive alternative approach to WES.[50] Another approach is the development of targeted gene-panels designed to interrogate multiple genes of interest for a specific disease. Whilst not within the scope of my PhD, I was instrumental in the design of a gene-panel for screening cerebral palsy individuals. Utilising the Haloplex Target Enrichment System (Agilent Technologies, Santa Clara, CA, USA) we created our own custom panel comprising 100 genes including all validated cerebral palsy genes from this study together with several genes

already published in the literature to be associated with cerebral palsy and cerebral palsy-like disorders. To date, in excess of 700 cerebral palsy cases have been tested and the analysis for this work is currently underway. Each of these sequencing strategies will contribute to our further understanding of the extent of the genetic involvement in cerebral palsy.

8.5.1 Whole-exome and whole-genome sequencing

Whole-exome sequencing is a proven successful strategy to identify rare and extremely rare (MAF <0.01%) alleles in the protein coding genes in Mendelian disorders and complex traits, including several neurodevelopmental disorders[1,13,14,51] and more recently cerebral palsy.[10,12] WES effectively interrogates the vast majority of the protein-coding regions of the genome where the majority of disease-causing mutations are found, and is therefore a good place to start.[52] However, not all variants of interest fall within the protein-coding regions of the genome. Whole-genome sequencing, regarded as the most comprehensive genetic screening thus far, extends the search for causative variants to the non-coding regions of the genome[53] as well as more accurately detecting indels, structural variants and CNVs compared to WES.[53,54] Comparative studies have shown that high-coverage WES and WGS are able to successfully generate adequate coverage for reliable variant calling of 95% of the coding regions, however sequencing biases are more pronounced in WES data.[55,56] The power of WGS was demonstrated in a recent study of 50 cases diagnosed with intellectual disability, for whom previous microarray and exome studies had failed to identify causative variants/genes. Following WGS, a genetic cause, including CNVs and de novo SNVs affecting the coding regions, was found in 21 out of the 50 cases (42%). Mutations in

the non-coding regions are yet to be fully examined, therefore the true genetic contribution may be underestimated.

Whole genome sequencing, as the result of rapidly falling costs and better technical characteristics, is predicted to take the lead role for future variant discovery. This does not mean there will not be a place for WES; on the contrary, for large-scale sequencing studies, WES is likely to be more accessible for many research organisations where resources are limited.[56] This is an important consideration for cerebral palsy genetic research, as it is very much in its infancy and large-scale sequencing studies will be fundamental to determining causative disease variants/genes and the extent of genetic involvement.

8.5.2 Custom gene-panels

Targeted gene-panels provide sequence data for several known genes of interest with higher coverage compared to WES and WGS, thus providing more confidence that no variants within these genes will be missed.[57] They are a cost-effective way of screening thousands to tens of thousands of samples and require less DNA than whole-exome and whole-genome sequencing.[58] Gene panels have been successfully developed for, and applied to, several neurodevelopmental disorders that overlap with cerebral palsy, including ataxia,[59] epilepsy[60] and ID[58,61]. However, targeted gene panels are inherently biased; you will only discover variants in genes you have targeted. As genetic research for cerebral palsy is still in its infancy, with likely many more relevant genes yet to be discovered, it is very possible you may miss something of importance.

8.5.3 Central data repository

Collaborations between research groups, both nationally and internationally, would allow the genetic contribution to cerebral palsy causation to be defined as quickly and cost-effectively as possible and is already happening. Collaboration will facilitate the screening of larger cohort sizes in multiple unrelated families, an essential strategy in the identification of variants/genes likely causative for cerebral palsy.[9] One future tool could be the establishment of a central biospecimens biobank and phenotype-genotype database specifically for cerebral palsy linked to existing disease databases, to better collect and understand the rare variants and their clinical association. Access to such a resource would aid clinicians and researchers in the interpretation of specific variants in particular with relation to the patient's clinical details.

8.5.4 Genome-wide expression profiling (RNA sequencing)

The transcriptome comprises a set of all RNA molecules, including messenger RNAs (mRNA), ribosomal RNAs (rRNA), transfer RNAs (tRNA) micro RNAs (miRNA) and other non-coding RNAs (ncRNA), transcribed in a cell or a population of cells. The majority of cells have the same genetic constitution but not every gene is transcriptionally active in every cell; therefore different cells exhibit diverse gene expression patterns.[62] These different gene expression patterns are essential for development, differentiation and homeostasis and their dysregulation can lead to disease.[63] RNA sequencing has provided a high-throughput approach for quantifying transcriptomes. Genome-wide expression profiling (RNA sequencing) is a powerful tool that has facilitated identification of dysregulated genes, i.e. those whose expression is silenced, up-regulated or down-regulated in disease populations compared to controls. Sequencing the transcriptome provides a quantitative estimate of gene expression that

can be used in conjunction with WGS or WES data to identify the functional impact of identified rare variants.[64]

We identified several potentially pathogenic variants associated with cerebral palsy in our recent WES study (**Chapter 4**) but the functional impact on downstream RNA expression is yet to be examined. The work done as part of my PhD has led to a large-scale RNA-sequencing analysis on all 183 cerebral palsy cases that underwent WES and will be completed in the near future. Combining gene expression data with our recent WES data can provide supporting evidence for the pathogenicity of variants in known or candidate genes and can identify variants in new genes which, based on our strict prioritization criteria, were not predicted as likely pathogenic.[64] Functional prediction of these dysregulated genes using *in silico* programs such as Ingenuity Pathway Analysis or the Database of Array Visualisation and Integrated Discovery will identify functional pathways that are altered in cerebral palsy cases.[64]

While brain or neuronal tissue would be optimum biological material but is not accessible for cerebral palsy cases. An alternative source of biological material, although not ideal, but that has been extensively used in literature, are patient-derived lymphoblastoid cell lines (LCLs). It is well recognized that LCLs may not express all disease-relevant genes that are expressed in the brain, making it impossible to study those genes using LCLs.[64] However, LCLs are easily generated by immortalizing patient B cells, making them suitable for high-throughput transcriptome analysis.

Ultimately cellular, molecular and animal model functional studies of candidate genes, in particular genes not previously associated with disease, will be necessary to provide

conclusive evidence of pathogenicity. Experimental data will confirm if the effect of the mutation is consistent with the motor disturbances associated with cerebral palsy.

8.5.5 Epigenetics and cerebral palsy

Epigenetics is another area of growing importance for cerebral palsy research. Several epidemiological studies have provided evidence of prenatal environmental factors associated with cerebral palsy. These factors include intrauterine growth restriction, maternal infection, intrauterine infection, illicit drug use and tobacco smoking.[30,46,65] How these environmental factors impact the epigenome of cerebral palsy affected individuals during *in utero* development is yet to be elucidated.

The recent revolution in human genome sequencing has had a major impact in our understanding of genetic factors contributing to disease, but the interaction of genes with their environment is also of paramount importance when examining disease susceptibility and causation. Several studies have shown that environmental influences i.e. chemical and/or nutritional factors, during the prenatal and early postnatal period, can cause epigenetic changes that can have latent effects on the health of an individual.[66,67] Epigenetic studies provide a platform for a greater understanding of gene-environment interactions.[66,68] Epigenetics refers to mitotically or meiotically heritable changes in gene expression that do not alter the DNA sequence.[66,69] Epigenetic marks are modifiers of DNA and histones that include DNA methylation, histone acetylation and methylation, and chromatin remodelling. DNA methylation, which takes place at the carbon-5 position of cytosine in CpG dinucleotides is the most common covalent modification of DNA in eukaryotes.[70]

Genomic imprinting, predominantly driven by DNA methylation, histone modification and RNA silencing, is one of the most comprehensively studied epigenetically-regulated phenomena in mammals.[66,67] Most human genes are biallelically expressed (concomitant expression of both maternal and paternal alleles) but a small subset of genes (~1%) are monoallelically expressed and can play an important role in fetal development.[67]

Several imprinted genes have been reported to be critical in embryonic growth and development, including placental development.[67,71] Previous studies have shown that epigenetic marks in placentas from pregnancies with fetal growth restriction were different from normal full-term pregnancies.[67,72] This is an extremely important consideration for cerebral palsy. Over a third of all cerebral palsy occurs in children with low birth weight and includes babies born preterm and/or growth restricted.[73]

Tobacco smoking has been associated with an increased risk for cerebral palsy.[46] It has been known for some considerable time that maternal smoking during pregnancy has a major impact on the short and long term health outcomes in children. The most documented short-term outcome is low birth weight.[74] Long-term increased risks include neurobehavioral defects, type 2 diabetes and elevated blood pressure.[68,75] Many studies have shown that tobacco smoking results in epigenetic alterations, not only in adults but also in the developing fetus, by altering the patterns of DNA methylation.[74,76] A recent study consisting of 10 mother-female newborn duos, where the mother had been a regular daily smoker prior to and during pregnancy, and 10 mother-female newborn duos, where the mother was a non-smoker, observed that methylation levels were significantly lower in newborns exposed to tobacco smoke compared to non-

exposed newborns. Significant differences in methylation between the exposed and non-exposed group were identified across 31 CpG sites in 25 different genes.[68]

Epigenetic marks are known to play an important role in the regulation of candidate genes critical for brain development.[77] The role of epigenetics in neurodevelopmental disorders is now a growing area of research.[78] Epigenetic dysregulation has been reported to play an important role in several neurodevelopmental disorders, including Angelman syndrome, Kabuki syndrome, Rett syndrome, Prader-Willi syndrome and more recently autism spectrum disorders.[77]

Understanding the interaction of epigenetic and genetic events is proving to be a successful strategy for understanding causation of neurodevelopmental disorders that can be applied to cerebral palsy research.[77] Placental tissue, umbilical cord blood and/or neonatal dried blood spots provide excellent biological resources for retrospectively examining the epigenome in cerebral palsy individuals at different stages of development compared to healthy individuals. Several methods have been developed to assess various epigenetic processes, particularly DNA methylation and chromatin modifications.[79] Two methods, bisulfite sequencing for DNA methylation and chromatin immunoprecipitation (ChIP) for DNA-protein interactions and chromatin modifications, have been considered the gold standard for epigenetic studies.[79]

The role of epigenetic regulation in neurodevelopment is complex.[80] It has been postulated that epigenetic modifications may offer another explanation, besides variable expressivity or reduced penetrance, as to why the same mutation can be benign in one individual and seemingly pathogenic in another or perhaps even result in a different

clinical features.[80,81] This may be relevant to cerebral palsy as several pathogenic variants identified in our recent WES study (**Chapter 4**) were in genes associated with neurological disorders, including intellectual disability.

8.6 Conclusion

Cerebral palsy is a clinical diagnosis encompassing a heterogeneous group of non-progressive disorders, primarily affecting movement and posture.[30] Applying WES, this study found multiple variants, *de novo* and/or inherited, of known and novel genes, suggesting considerable genetic heterogeneity for cerebral palsy. The level of genetic heterogeneity may in part explain the well-established clinical variability of cerebral palsy. These results suggest that different variants of previously reported genes implicated in known neurological disorders could give rise to cerebral palsy as another phenotype. Since the introduction of sequencing studies it has become increasingly apparent that the range of phenotypes in a specific gene can be broad.[82] This would suggest at least some clinical bias, i.e. selection of specific gene screening in specific clinical phenotypes. Systematic whole exome or whole genome sequencing has the power to address such biases and can redefine many existing gene-phenotype associations.

In the past birth asphyxia has been considered the leading cause of cerebral palsy and this belief has not only hindered cerebral palsy research but also led to an increase in caesarean section rates. The results from this study challenge the long held belief that cerebral palsy is largely due to birth asphyxia. Litigation related to cerebral palsy has been costly in Australia and elsewhere in the developed world. It is now evident that cerebral palsy is more complex than first thought and clinical investigations coupled with genetic screening are likely to provide greater understanding of cerebral palsy causation in the future.[83]

Previous estimates have suggested that the contribution of genetic variants to the burden of cerebral palsy is about 2%.[84] Our landmark study found 14% of the 98 trio cases had a potential pathogenic variant for cerebral palsy. These results highlighted for the first time the genetic heterogeneity and complexity of cerebral palsy. Follow-up studies to confirm or refute pathogenicity of cerebral palsy candidate genes are essential. Further WES, and eventually WGS and targeted gene panels, in new cerebral palsy cohorts are likely to uncover more candidate cerebral palsy genes which are individually rare, heterogeneous, as a group common and together increase the genetic burden in this neurodevelopmental disorder.

WES kits originally designed as a research tool are now becoming part of the diagnostic tool kit.[85,86] Considerable effort is being placed into designing clinical grade "medical exome" kits. The current approach is to build upon the research exome through enriched coverage of medical relevant genes with regular updating as new disease genes are discovered, and this can include cerebral palsy genes.[85] There is continuing support for WES as an effective diagnostic test for Mendelian and common neurodevelopmental disorders known to have a genetic basis. Important to this approach is the collection of detailed clinical data as it provides the basis for genotype-phenotype correlation.[86] Revealing gene-environment interactions will also be important for our understanding of the pathogenicity of cerebral palsy. DNA sequencing alone is likely to be inadequate. In the future a combined approach of DNA sequencing, genome-wide expression profiling (RNA sequencing) and epigenetic studies i.e. DNA methylome profiling, and functional studies in biological systems will be essential if we are to better understand cerebral palsy causation.

Cerebral palsy is of major importance to the affected individuals, their families and society. Identifying genetic causes of cerebral palsy is important to families for many reasons: it provides parents with an explanation for their child's disorder and knowing the cause of cerebral palsy relieves uncertainty, allows improved recurrence risk counselling, can inform prognostic counselling, increases reproductive options for some families and opens the way for treatment based on an improved understanding of the individual's, or a group of individuals', specific genetic diagnosis.

The identification of causal mutations in cerebral palsy genes can assist in providing clinicians with a conclusive genetic diagnosis for cerebral palsy or in distinguishing between cerebral palsy and genetic neurological disorders that include the motor components of cerebral palsy among their features, especially when the diagnostic evaluation is being done early in life. A proportion, yet to be defined, will have an underlying genetic disorder, with clinical features that meet the currently accepted criteria for diagnosis of cerebral palsy.

This study highlights the importance of genetic investigation of individuals with cerebral palsy. Gaining a more comprehensive understanding of cerebral palsy causation will provide better recognition of the needs for management and intervention, and ultimately lead to treatment for some cerebral palsy cases, thereby lessening the burden for some affected individuals and their families.

8.7 References

1. Rauch, A., Wieczorek, D., Graf, E. *et al.* Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: An exome sequencing study. *Lancet* **380**, 1674-82 (2012).
2. Vissers, L.E., de Ligt, J., Gilissen, C. *et al.* A de novo paradigm for mental retardation. *Nat Genet* **42**, 1109-12 (2010).
3. Mefford, H.C., Batshaw, M.L. & Hoffman, E.P. Genomics, intellectual disability, and autism. *N Engl J Med* **366**, 733-43 (2012).
4. O'Roak, B.J., Deriziotis, P., Lee, C. *et al.* Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat Genet* **43**, 585-9 (2011).
5. O'Roak, B.J., Vives, L., Girirajan, S. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* **485**, 246-50 (2012).
6. Allen, A.S., Berkovic, S.F., Cossette, P. *et al.* De novo mutations in epileptic encephalopathies. *Nature* **501**, 217-21 (2013).
7. Dyment, D.A., Tetreault, M., Beaulieu, C.L. *et al.* Whole-exome sequencing broadens the phenotypic spectrum of rare pediatric epilepsy: A retrospective study. *Clin Genet* (2014).
8. Veeramah, K.R., Johnstone, L., Karafet, T.M. *et al.* Exome sequencing reveals new causal mutations in children with epileptic encephalopathies. *Epilepsia* **54**, 1270-81 (2013).
9. MacArthur, D.G., Manolio, T.A., Dimmock, D.P. *et al.* Guidelines for investigating causality of sequence variants in human disease. *Nature* **508**, 469-76 (2014).

10. Kruer, M.C., Jepperson, T., Dutta, S. *et al.* Mutations in gamma adducin are associated with inherited cerebral palsy. *Annals of neurology* **74**, 805-14 (2013).
11. McHale, D.P., Mitchell, S., Bundey, S. *et al.* A gene for autosomal recessive symmetrical spastic cerebral palsy maps to chromosome 2q24-25. *Am J Hum Genet* **64**, 526-32 (1999).
12. Parolin Schnekenberg, R., Perkins, E.M., Miller, J.W. *et al.* De novo point mutations in patients diagnosed with ataxic cerebral palsy. *Brain* (2015).
13. de Ligt, J., Willemsen, M.H., van Bon, B.W. *et al.* Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med* **367**, 1921-9 (2012).
14. Sanders, S.J., Murtha, M.T., Gupta, A.R. *et al.* De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237-41 (2012).
15. Poirier, K., Saillour, Y., Fourniol, F. *et al.* Expanding the spectrum of tuba1a-related cortical dysgenesis to polymicrogyria. *European journal of human genetics : EJHG* **21**, 381-5 (2013).
16. Bahi-Buisson, N., Poirier, K., Boddaert, N. *et al.* Refinement of cortical dysgeneses spectrum associated with tuba1a mutations. *Journal of medical genetics* **45**, 647-53 (2008).
17. Poirier, K., Keays, D.A., Francis, F. *et al.* Large spectrum of lissencephaly and pachygyria phenotypes resulting from de novo missense mutations in tubulin alpha 1a (tuba1a). *Human mutation* **28**, 1055-64 (2007).
18. Yamasaki, M., Thompson, P. & Lemmon, V. Crash syndrome: Mutations in l1cam correlate with severity of the disease. *Neuropediatrics* **28**, 175-8 (1997).

19. Lee, R.W., Poretti, A., Cohen, J.S. *et al.* A diagnostic approach for cerebral palsy in the genomic era. *Neuromolecular Med* (2014).

20. Wu, Y., Song, S.W., Sun, J. *et al.* Iip45 inhibits cell migration through inhibition of hdac6. *J Biol Chem* **285**, 3554-60 (2010).

21. Rujirabanjerd, S., Nelson, J., Tarpey, P.S. *et al.* Identification and characterization of two novel jarid1c mutations: Suggestion of an emerging genotype-phenotype correlation. *Eur J Hum Genet* **18**, 330-5 (2010).

22. Conklin, D., Holderman, S., Whitmore, T.E., Maurer, M. & Feldhaus, A.L. Molecular cloning, chromosome mapping and characterization of ubqln3 a testis-specific gene that contains an ubiquitin-like domain. *Gene* **249**, 91-8 (2000).

23. Zou, Y., Liu, Q., Chen, B. *et al.* Mutation in cul4b, which encodes a member of cullin-ring ubiquitin ligase complex, causes x-linked mental retardation. *Am J Hum Genet* **80**, 561-6 (2007).

24. O'Leary, R.E., Shih, J.C., Hyland, K. *et al.* De novo microdeletion of xp11.3 exclusively encompassing the monoamine oxidase a and b genes in a male infant with episodic hypotonia: A genomics approach to personalized medicine. *European journal of medical genetics* **55**, 349-53 (2012).

25. Delphin, N., Hanein, S., Taie, L.F. *et al.* Intellectual disability associated with retinal dystrophy in the xp11.3 deletion syndrome: Znf674 on trial. Guilty or innocent? *Eur J Hum Genet* **20**, 352-6 (2012).

26. Kaufmann, R., Straussberg, R., Mandel, H. *et al.* Infantile cerebral and cerebellar atrophy is associated with a mutation in the med17 subunit of the transcription preinitiation mediator complex. *Am J Hum Genet* **87**, 667-70 (2010).

27. Shoukier, M., Neesen, J., Sauter, S.M. *et al.* Expansion of mutation spectrum, determination of mutation cluster regions and predictive structural classification of spast mutations in hereditary spastic paraplegia. *Eur J Hum Genet* **17**, 187-94 (2009).

28. Wolf, S.I., Braatz, F., Metaxiotis, D. *et al.* Gait analysis may help to distinguish hereditary spastic paraplegia from cerebral palsy. *Gait & Posture* **33**, 556-61 (2011).

29. Veltman, J.A. & Brunner, H.G. De novo mutations in human genetic disease. *Nat Rev Genet* **13**, 565-75 (2012).

30. Moreno-De-Luca, A., Ledbetter, D.H. & Martin, C.L. Genomic insights into the causes and classification of the cerebral palsies. *Lancet Neurol* **11**, 283-92 (2012).

31. Manolio, T.A., Collins, F.S., Cox, N.J. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747-53 (2009).

32. Hirata, H., Nanda, I., van Riesen, A. *et al.* Zc4h2 mutations are associated with arthrogryposis multiplex congenita and intellectual disability through impairment of central and peripheral synaptic plasticity. *Am J Hum Genet* **92**, 681-95 (2013).

33. Carre, A., Szinnai, G., Castanet, M. *et al.* Five new ttf1/nkx2.1 mutations in brain-lung-thyroid syndrome: Rescue by pax8 synergism in one case. *Hum Mol Genet* **18**, 2266-76 (2009).

34. Doyle, D.A., Gonzalez, I., Thomas, B. & Scavina, M. Autosomal dominant transmission of congenital hypothyroidism, neonatal respiratory distress, and ataxia caused by a mutation of nkx2-1. *J Pediatr* **145**, 190-3 (2004).

35. Yeargin-Allsopp, M., Rice, C., Karapurkar, T. *et al.* Prevalence of autism in a us metropolitan area. *JAMA* **289**, 49-55 (2003).

36. Bakhchane, A., Charif, M., Salime, S. *et al.* Recessive tbc1d24 mutations are frequent in moroccan non-syndromic hearing loss pedigrees. *PLoS One* **10**, e0138072 (2015).

37. Corbett, M.A., Bahlo, M., Jolly, L. *et al.* A focal epilepsy and intellectual disability syndrome is due to a mutation in tbc1d24. *Am J Hum Genet* **87**, 371-5 (2010).

38. Duru, N., Iseri, S.A., Selcuk, N. & Tolun, A. Early-onset progressive myoclonic epilepsy with dystonia mapping to 16pter-p13.3. *J Neurogenet* **24**, 207-15 (2010).

39. Milh, M., Falace, A., Villeneuve, N. *et al.* Novel compound heterozygous mutations in tbc1d24 cause familial malignant migrating partial seizures of infancy. *Hum Mutat* **34**, 869-72 (2013).

40. Campeau, P.M., Kasperaviciute, D., Lu, J.T. *et al.* The genetic basis of doors syndrome: An exome-sequencing study. *Lancet Neurol* **13**, 44-58 (2014).

41. Azaiez, H., Booth, K.T., Bu, F. *et al.* Tbc1d24 mutation causes autosomal-dominant nonsyndromic hearing loss. *Hum Mutat* **35**, 819-23 (2014).

42. Rehman, A.U., Santos-Cortez, R.L., Morell, R.J. *et al.* Mutations in tbc1d24, a gene associated with epilepsy, also cause nonsyndromic deafness dfnb86. *Am J Hum Genet* **94**, 144-52 (2014).

43. Clark, M.J., Chen, R., Lam, H.Y. *et al.* Performance comparison of exome DNA sequencing technologies. *Nat Biotechnol* **29**, 908-14 (2011).

44. Sims, D., Sudbery, I., Ilott, N.E., Heger, A. & Ponting, C.P. Sequencing depth and coverage: Key considerations in genomic analyses. *Nat Rev Genet* **15**, 121-32 (2014).

45. Bainbridge, M.N., Wang, M., Wu, Y. *et al.* Targeted enrichment beyond the consensus coding DNA sequence exome reveals exons with higher variant densities. *Genome Biol* **12**, R68 (2011).

46. O'Callaghan, M.E., MacLennan, A., Gibson, C. *et al.* Epidemiolgic associations with cerebral palsy. *Obstet Gynecol* **118**, 576-582 (2011).

47. Wu, Y.W. & Colford, J.M., Jr. Chorioamnionitis as a risk factor for cerebral palsy: A meta-analysis. *JAMA* **284**, 1417-24 (2000).

48. Grether, J.K. & Nelson, K.B. Maternal infection and cerebral palsy in infants of normal birth weight. *JAMA* **278**, 207-11 (1997).

49. McMichael, G., Maclennan, A., Gibson, C. *et al.* Cytomegalovirus and epstein-barr virus may be associated with some cases of cerebral palsy. *J Matern Fetal Neonatal Med* (2012).

50. Biesecker, L.G., Shianna, K.V. & Mullikin, J.C. Exome sequencing: The expert view. *Genome Biol* **12**, 128 (2011).

51. Bamshad, M.J., Ng, S.B., Bigham, A.W. *et al.* Exome sequencing as a tool for mendelian disease gene discovery. *Nat Rev Genet* **12**, 745-55 (2011).

52. Stenson, P.D., Ball, E.V., Howells, K. *et al.* The human gene mutation database: Providing a comprehensive central mutation database for molecular diagnostics and personalized genomics. *Hum Genomics* **4**, 69-72 (2009).

53. Belkadi, A., Bolze, A., Itan, Y. *et al.* Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc Natl Acad Sci U S A* **112**, 5473-8 (2015).

54. Fang, H., Wu, Y., Narzisi, G. *et al.* Reducing indel calling errors in whole genome and exome sequencing data. *Genome Med* **6**, 89 (2014).

55. Lelieveld, S.H., Spielmann, M., Mundlos, S., Veltman, J.A. & Gilissen, C. Comparison of exome and genome sequencing technologies for the complete capture of protein-coding regions. *Hum Mutat* **36**, 815-22 (2015).

56. Warr, A., Robert, C., Hume, D. *et al.* Exome sequencing: Current and future perspectives. *G3 (Bethesda)* **5**, 1543-50 (2015).

57. Sun, Y., Ruivenkamp, C.A., Hoffer, M.J. *et al.* Next-generation diagnostics: Gene panel, exome, or whole genome? *Hum Mutat* **36**, 648-55 (2015).

58. Lohmann, K. & Klein, C. Next generation sequencing and the future of genetic diagnosis. *Neurotherapeutics* **11**, 699-707 (2014).

59. Nemeth, A.H., Kwasniewska, A.C., Lise, S. *et al.* Next generation sequencing for molecular diagnosis of neurological disorders using ataxias as a model. *Brain* **136**, 3106-18 (2013).

60. Lemke, J.R., Riesch, E., Scheurenbrand, T. *et al.* Targeted next generation sequencing as a diagnostic tool in epileptic disorders. *Epilepsia* **53**, 1387-98 (2012).

61. Saudi Mendeliome, G. Comprehensive gene panels provide advantages over clinical exome sequencing for mendelian diseases. *Genome Biol* **16**, 134 (2015).

62. Wang, Z., Gerstein, M. & Snyder, M. Rna-seq: A revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57-63 (2009).

63. Handel, A.E., Disanto, G. & Ramagopalan, S.V. Next-generation sequencing in understanding complex neurological disease. *Expert Rev Neurother* **13**, 215-27 (2013).

64. Luo, R., Sanders, S.J., Tian, Y. *et al.* Genome-wide transcriptome profiling reveals the functional impact of rare de novo and recurrent cnvs in autism spectrum disorders. *Am J Hum Genet* **91**, 38-55 (2012).

65. Colver, A., Fairhurst, C. & Pharoah, P.O. Cerebral palsy. *Lancet* **383**, 1240-9 (2014).

66. Jirtle, R.L. & Skinner, M.K. Environmental epigenomics and disease susceptibility. *Nat Rev Genet* **8**, 253-62 (2007).

67. Pozharny, Y., Lambertini, L., Ma, Y. *et al.* Genomic loss of imprinting in first-trimester human placenta. *Am J Obstet Gynecol* **202**, 391 e1-8 (2010).

68. Ivorra, C., Fraga, M.F., Bayon, G.F. *et al.* DNA methylation patterns in newborns exposed to tobacco in utero. *J Transl Med* **13**, 25 (2015).

69. Meaburn, E. & Schulz, R. Next generation sequencing in epigenetics: Insights and challenges. *Semin Cell Dev Biol* **23**, 192-9 (2012).

70. Law, J.A. & Jacobsen, S.E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* **11**, 204-20 (2010).

71. Isles, A.R. & Holland, A.J. Imprinted genes and mother-offspring interactions. *Early Hum Dev* **81**, 73-7 (2005).

72. Diplas, A.I., Lambertini, L., Lee, M.J. *et al.* Differential expression of imprinted genes in normal and iugr human placentas. *Epigenetics* **4**, 235-40 (2009).

73. Jarvis, S., Glinianaia, S.V., Torrioli, M.G. *et al.* Cerebral palsy and intrauterine growth in single births: European collaborative study. *Lancet* **362**, 1106-11 (2003).

74. Aagaard-Tillery, K.M., Porter, T.F., Lane, R.H., Varner, M.W. & Lacoursiere, D.Y. In utero tobacco exposure is associated with modified effects of maternal factors on fetal growth. *Am J Obstet Gynecol* **198**, 66 e1-6 (2008).

75. Brion, M.J., Leary, S.D., Lawlor, D.A., Smith, G.D. & Ness, A.R. Modifiable maternal exposures and offspring blood pressure: A review of epidemiological studies of maternal age, diet, and smoking. *Pediatr Res* **63**, 593-8 (2008).

76. Wilhelm-Benartzi, C.S., Houseman, E.A., Maccani, M.A. *et al.* In utero exposures, infant growth, and DNA methylation of repetitive elements and developmentally related genes in human placenta. *Environ Health Perspect* **120**, 296-302 (2012).

77. Rangasamy, S., D'Mello, S.R. & Narayanan, V. Epigenetics, autism spectrum, and neurodevelopmental disorders. *Neurotherapeutics* **10**, 742-56 (2013).

78. Fagiolini, M., Jensen, C.L. & Champagne, F.A. Epigenetic influences on brain development and plasticity. *Curr Opin Neurobiol* **19**, 207-12 (2009).

79. Zhang, Y. & Jeltsch, A. The application of next generation sequencing in DNA methylation analysis. *Genes (Basel)* **1**, 85-101 (2010).

80. Zahir, F.R. & Brown, C.J. Epigenetic impacts on neurodevelopment: Pathophysiological mechanisms and genetic modes of action. *Pediatr Res* **69**, 92R-100R (2011).

81. Van Winkel, R., Esquivel, G., Kenis, G. *et al.* Review: Genome-wide findings in schizophrenia and the role of gene-environment interplay. *CNS Neurosci Ther* **16**, e185-92 (2010).

82. Schaefer, G.B. Genetics considerations in cerebral palsy. *Semin Pediatr Neurol* **15**, 21-6 (2008).

83. Hankins, G.D., MacLennan, A.H., Speer, M.E., Strunk, A. & Nelson, K. Obstetric litigation is asphyxiating our maternity services. *Obstet Gynecol* **107**, 1382-5 (2006).

84. Rajab, A., Yoo, S.Y., Abdulgalil, A. *et al.* An autosomal recessive form of spastic cerebral palsy (cp) with microcephaly and mental retardation. *Am J Med Genet A* **140**, 1504-10 (2006).

85. van Zelst-Stams, W.A., Scheffer, H. & Veltman, J.A. Clinical exome sequencing in daily practice: 1,000 patients and beyond. *Genome Med* **6**, 2 (2014).

86. Shen, Y., Lopez, R., Milos, P.M. The rapidly emerging role for whole exome sequencing in clinical genetics. *Curr Genet Med Rep* **2**, 103-112 (2014).

Appendix
Supplementary data

Whole exome sequencing points to considerable genetic heterogeneity of cerebral palsy

Gai McMichael, Matthew N Bainbridge, Eric Haan, Mark Corbett, Alison Gardner, Suzanna Thompson, Bregje W M van Bon, Clare L van Eyk, Jessica Broadbent, Corinne Reynolds, Michael E O'Callaghan, Lam Son Nguyen, David L Adelson, Ray Russo, Shalini Jhangiani, Harsha Doddapaneni, Donna M Muzny, Richard A Gibbs, Jozef Gecz,* Alastair H MacLennan *

*Equal project supervisors

Molecular Psychiatry (2015) 20, 176-182

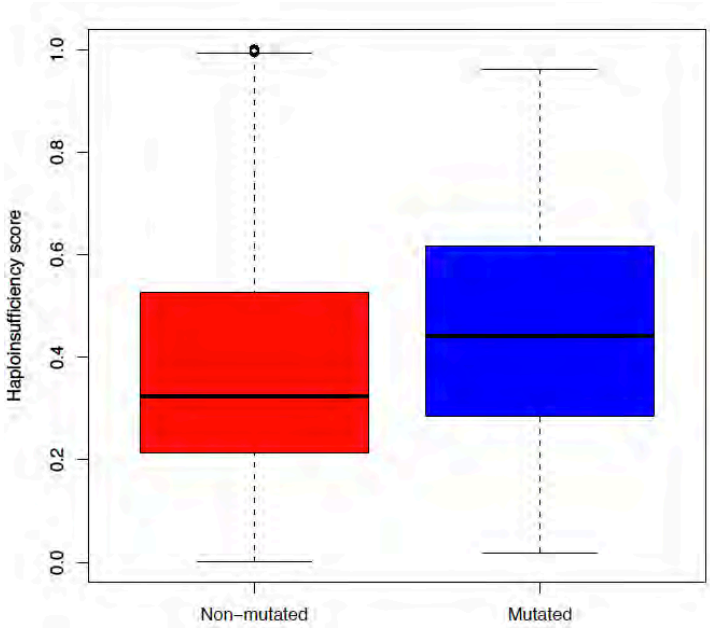


Figure 9.1 Supplementary Figure A. Results from t-test analysis through Partek on the haploinsufficiency scores for non-mutated genes (n=12,218) and genes with a *de novo* mutation (where scores were available; n = 38) identified a slight enrichment compared to non-mutated genes but this was not significant (P = 0.08).

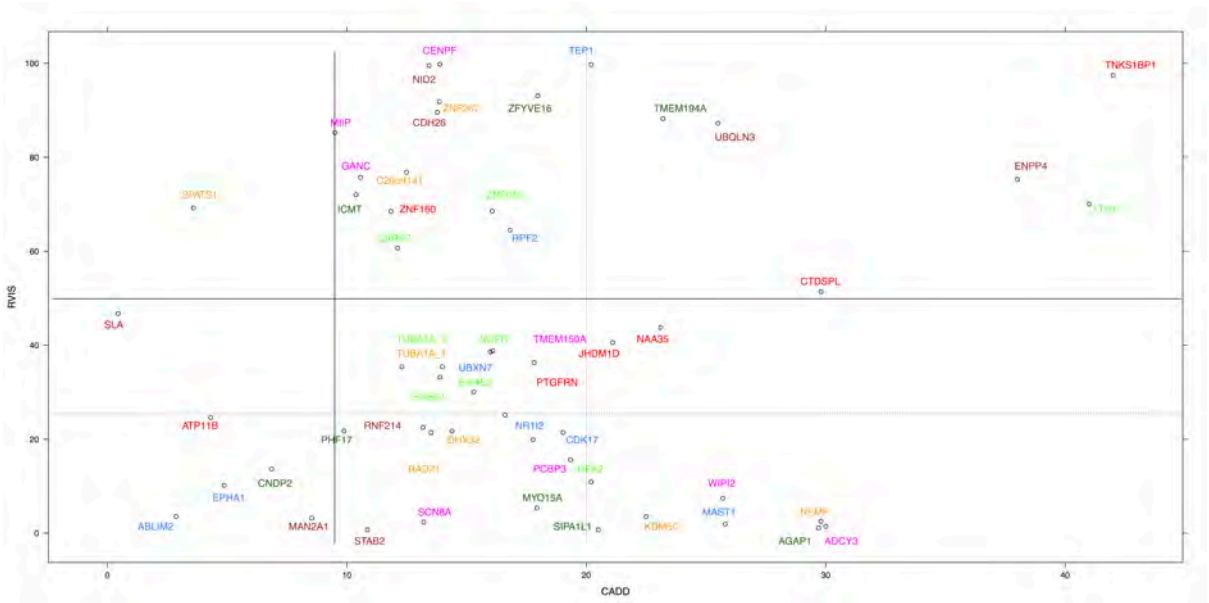


Figure 9.2 Supplementary Figure B. Cerebral palsy *de novo* variants where both RVIS and CADD scores were available (n = 55). Two different variants identified in *TUBA1A* are labeled *TUBA1A*_1 (p.R123C) and *TUBA1A*_2 (p.L152Q). Two different cut-off criteria were applied: solid lines RVIS <25 and CADD >20; dotted lines RVIS <50 and CADD >10.

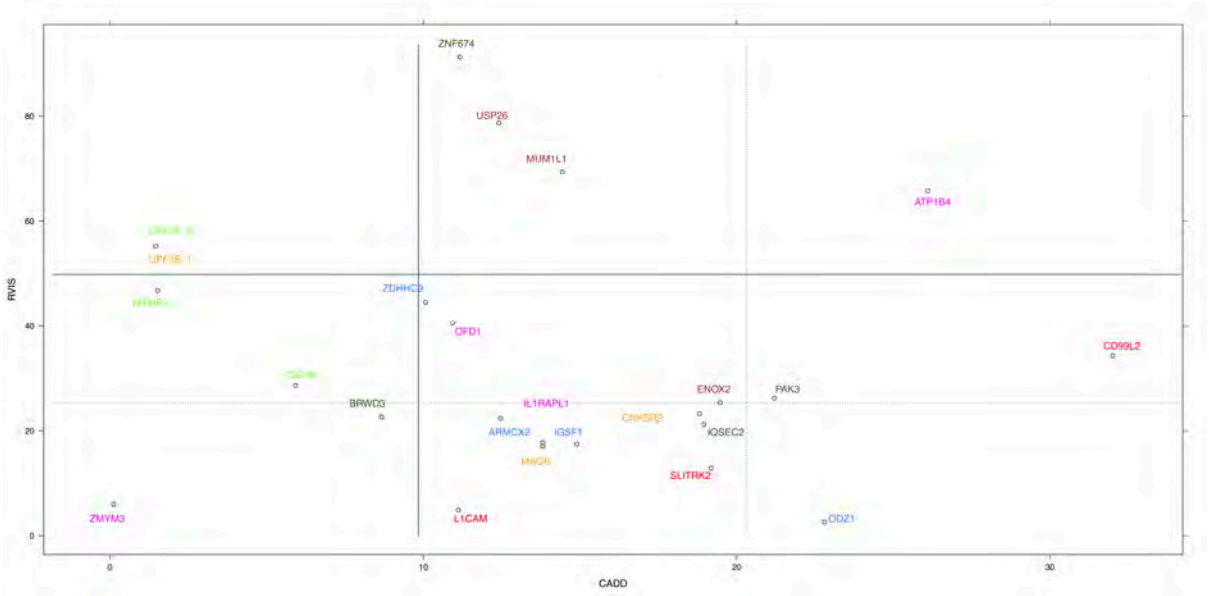


Figure 9.3 Supplementary Figure C. Cerebral palsy inherited X-chromosome variants where both RVIS and CADD scores were available (n = 24). These variants were inherited from an unaffected mother to affected son. Two different cut-off criteria were applied: solid lines RVIS <25 and CADD >20; dotted lines RVIS <50 and CADD >10.

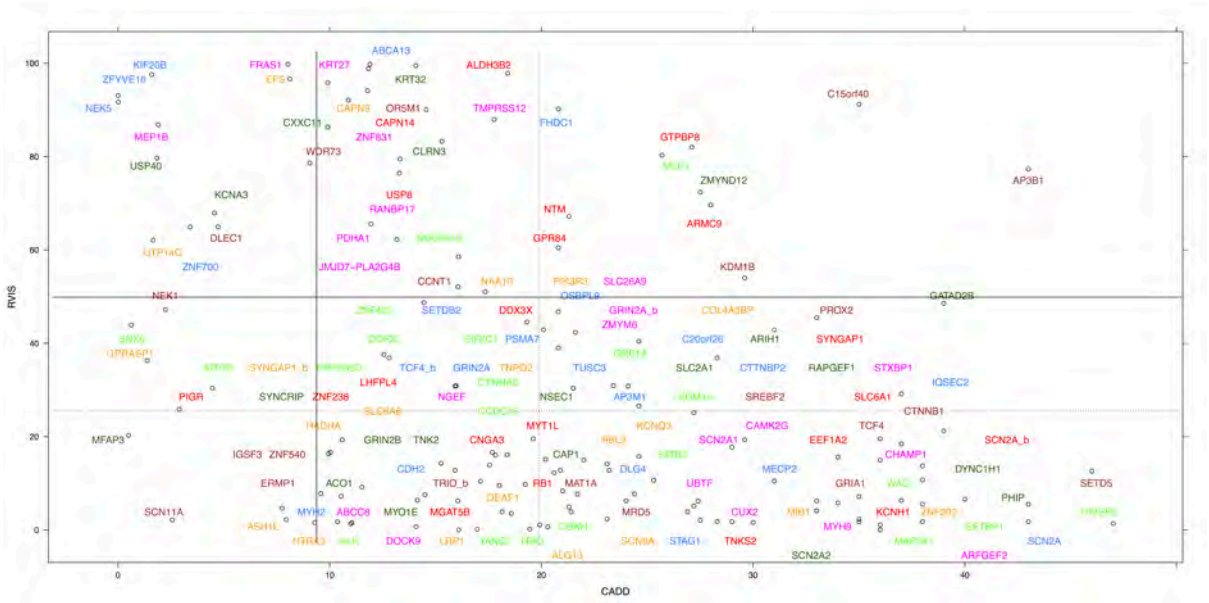


Figure 9.4 Supplementary Figure D. *De novo* mutations in individuals with severe intellectual disability (n = 149).[1,2] Two different cut-off criteria were applied: solid lines RVIS <25 and CADD >20; dotted lines RVIS <50 and CADD >10.

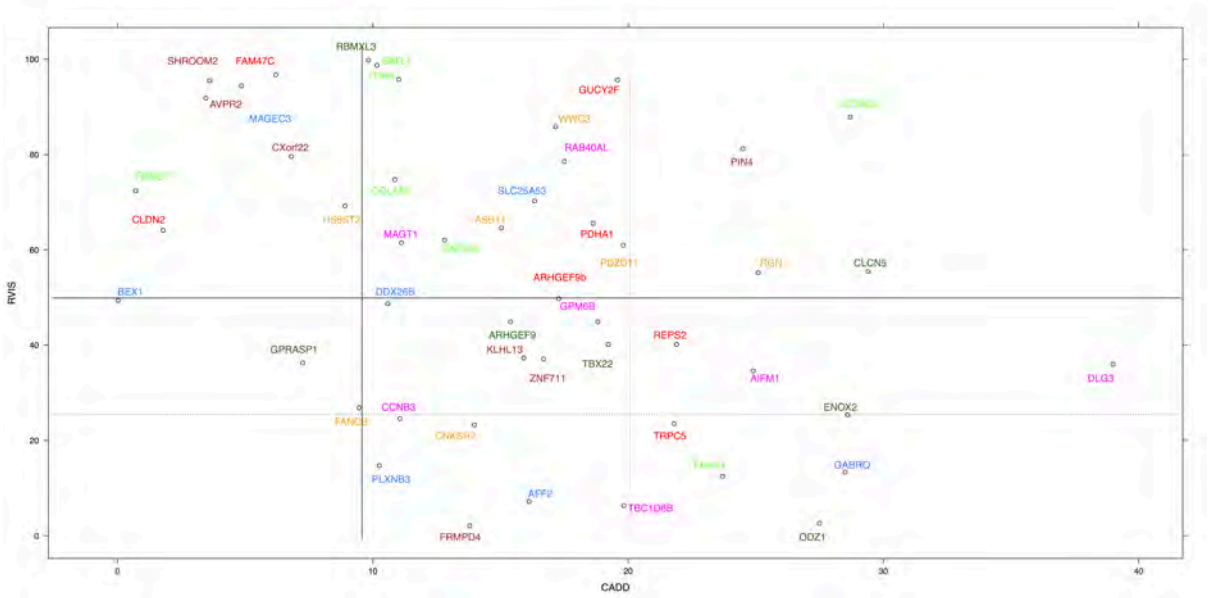


Figure 9.5 Supplementary Figure E. X-chromosome mutations associated with severe non-syndromic sporadic intellectual disability (n = 49).[1,2] Two different cut-off criteria were applied: solid lines RVIS <25 and CADD >20; dotted lines RVIS <50 and CADD >10.

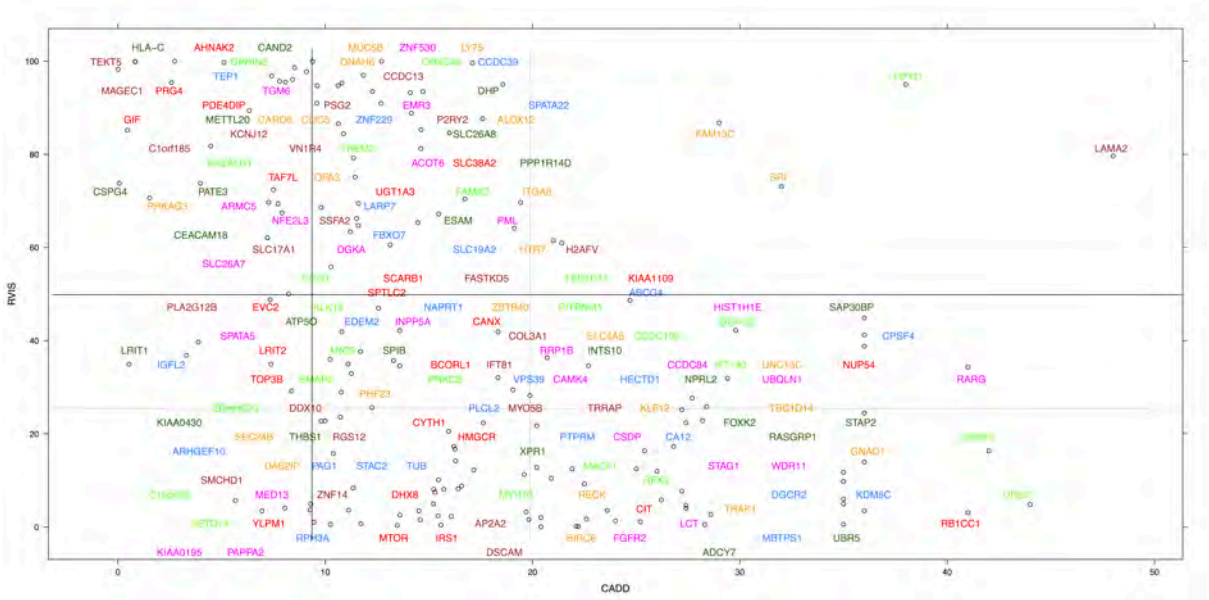


Figure 9.6 Supplementary Figure F. *De novo* mutations in schizophrenia (n = 170). Schizophrenia is widely held to stem from the combined effects of multiple common polymorphisms, each with a small impact on disease risk and reflected in the above scatterplot.[3] Two different cut-off criteria were applied: solid lines RVIS <25 and CADD >20; dotted lines RVIS <50 and CADD >10.

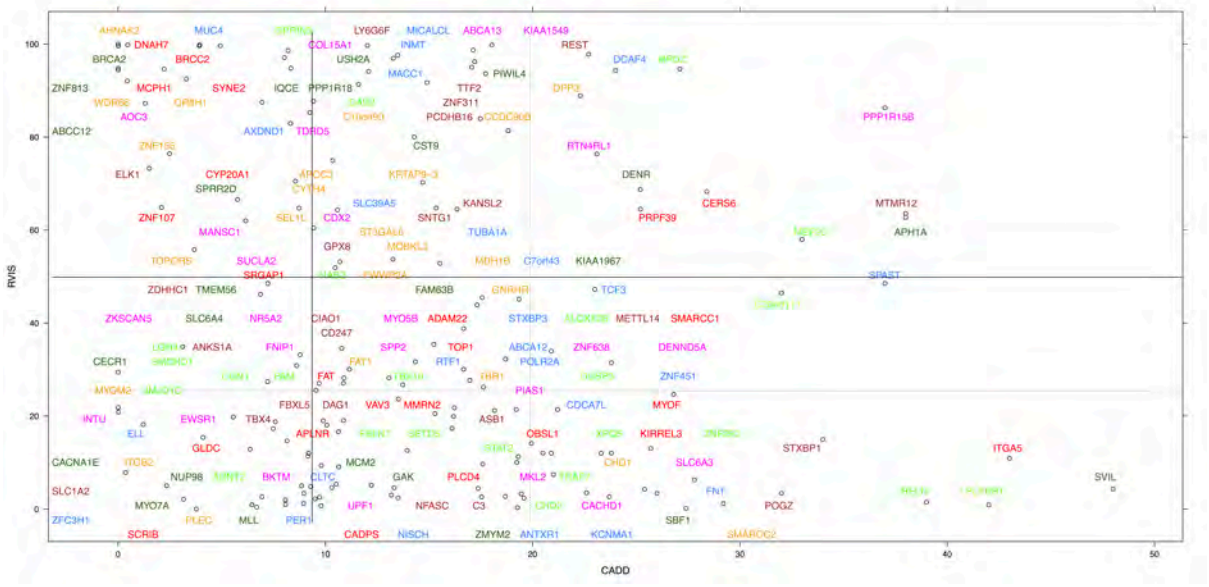


Figure 9.7 Supplementary Figure G. *De novo* mutations in autism spectrum disorders (n = 161).[4] Two different cut-off criteria were applied: solid lines RVIS <25 and CADD >20; dotted lines RVIS <50 and CADD >10.

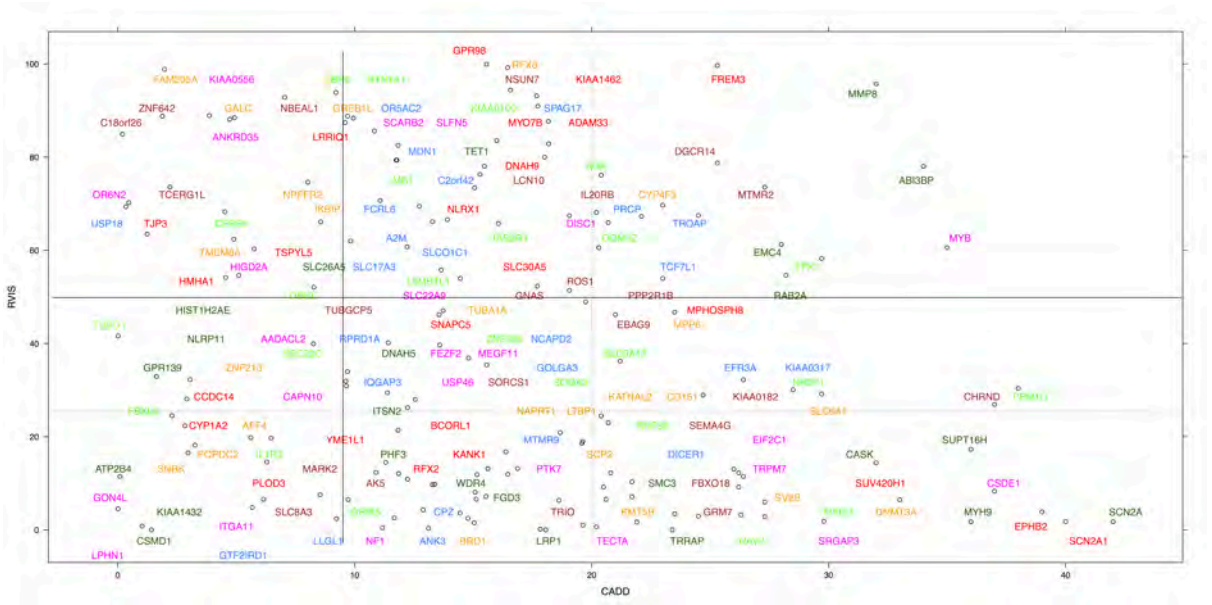


Figure 9.8 Supplementary Figure H. *De novo* mutations in autism spectrum disorders (n = 168).[5] Two different cut-off criteria were applied: solid lines RVIS <25 and CADD >20; dotted lines RVIS <50 and CADD >10.

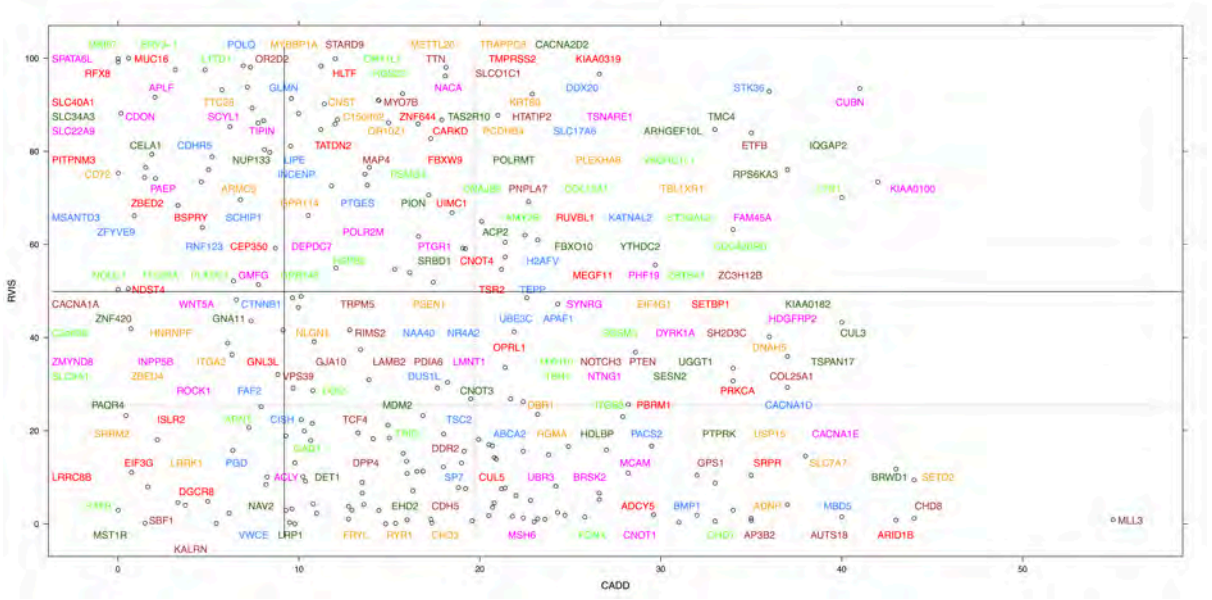


Figure 9.9 Supplementary Figure I. *De novo* mutations in sporadic cases with autism spectrum disorders (n = 228).[6] Two different cut-off criteria were applied: solid lines RVIS <25 and CADD >20; dotted lines RVIS <50 and CADD >10.

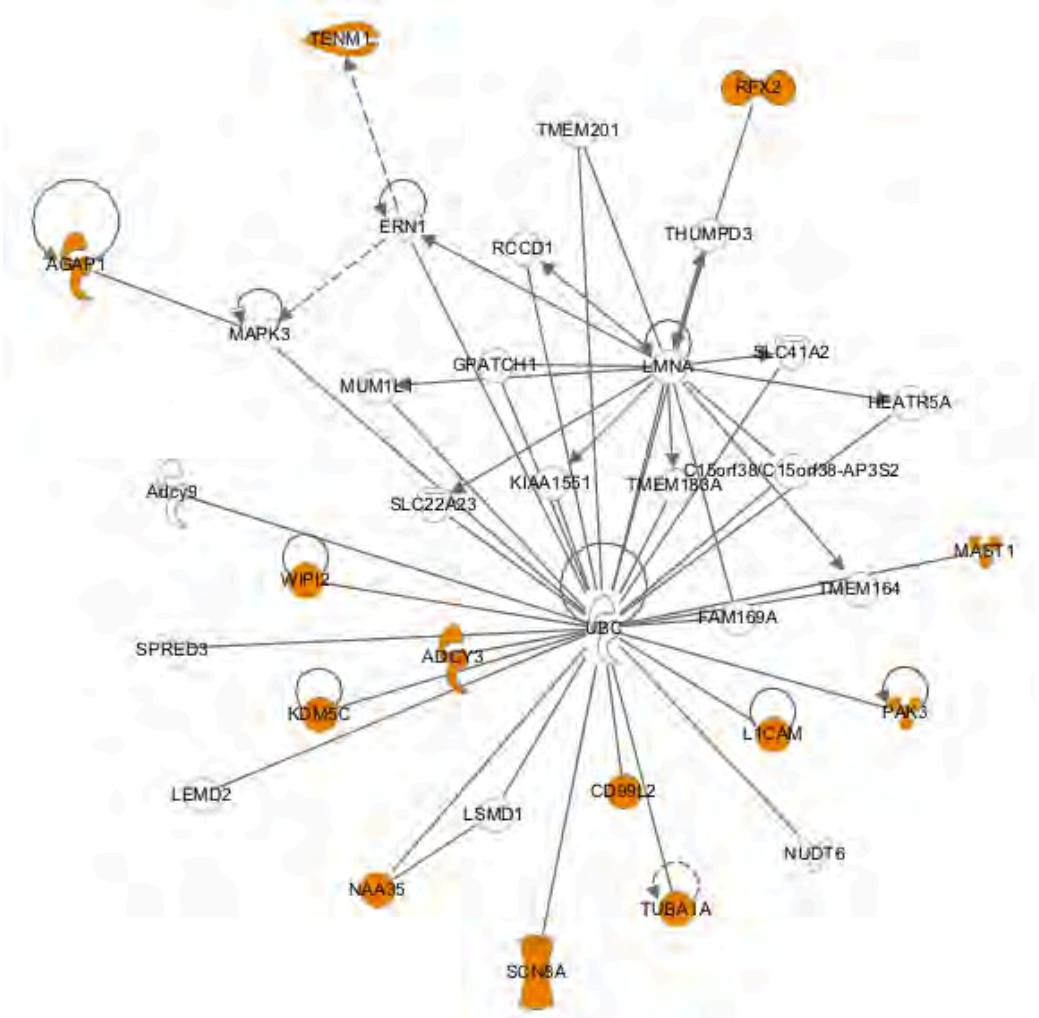


Figure 9.10 Supplementary Figure J. IPA associated network functions. The top interaction network from IPA analysis of 14 predicted causative genes for CP was developmental disorder, hereditary disorder and neurological disease and involved 13 of the 14 predicted causative genes. Solid lines indicate direct interactions while dashed lines represent indirect connection.

Table 9.1 Supplementary Table A. Clinical characteristics of 183 cases with CP.

Cases	Gest weeks	Sex	CP type	Mat age years	Pat age years	BW grams	BW centile	HC	Pregnancy complications	Plurity	GMFCS	Apgar scores 1 - 5 min	Neuro-developmental comorbidities	Clinical details and neuroimaging
1001P	26	F	Q	NA	NA	910	9			S	4	NA		US day 2 – Grade IV left IVH, MRI at 11 mths: PVL
2002P	29	F	D	42	NA	1510	89	29.5	Hypothyroidism	T	1	8-9		US at day 1: bilateral IVH
3003P	32	F	D	38	37	1835	36	30	Emergency caesarean	T	4	9-9	Epilepsy	US: IVH, MRI 7 mths & 6 yrs: PVL: Cortical visual impairment
4004P	40	F	H	27	27	3175	30	32.5		S	2	NA		MRI 16 mths: left porencephalic cyst
5005P	32	F	H	27	34	1559	5			NA	NA	NA		US day 1 to day 8: Grade I IVH
7007P	32	F	D	31	31	1145	0		Assisted conception, breech, emergency caesarean, fetal distress	T	2	9-10		No reports
8008P	34	M	T	35	34	3120	100	34	Maternal diabetes, emergency caesarean, fetal distress	S	2	9-9		No reports
9009P	34	M	H	29	34	2790	88		Fetal distress	S	2	6-8		MRI at 10 yrs: PVL, cystic encephalomalacia
10010P	40	M	D	27	30	3710	47			S	5	8-9	Severe ID, autism	MRI at 1 yr and 11 yrs: aqueductal stenosis, hydrocephalus, Cortical visual impairment
11011P	30	F	D	NA	NA	1480	NA	28		S	3	NA		MRI at 2 yrs: PVL
12012P	29	F	D	34	32	1390	64	29.5	Assisted conception, TTTS, emergency caesarean	T	4	NA	Mild ID, autism	Cranial US day 3 onwards: bilateral Grade II IVH, MRI at 2 yrs: PVL

13013P	40	F	H	35	34	3520	60		Emergency caesarean, fetal distress	S	2	3-9	Mild ID, autism, epilepsy	US day 2 to 5 mths: Grade IV IVH and L MCA territory infarct, then PVL, MRI day 6 & 18 mths: L MCA and ACA territory infarct
14014P	31	M	Q	30	32	1500	8		Emergency caesarean	T	NA	NA		No reports
15015P	29	M	H	36	38	960	1		Emergency caesarean	S	1	NA	Epilepsy, cognitive ability lies at the upper limit of the low average range	US day 2 onwards: right Grade IV IVH, MRI 10 yrs: marked right hemisphere leukomalacia
16016P	41	M	Q	41	40	2700	2		Antenatal infection, fetal distress	S	5	1-4	Epilepsy	No reports
17017P	39	M	Q	27	23	3685	45			S	NA	NA		MRI at 7 mths: diffuse polymicrogyria
18018P	29	F	D	38	NA	1165	18	26.5		S	2	6-9		US at day 1 and 1 mth: Right PVL. Strabismus, bilateral, operated; development normal at 19 mths
19019P	37	M	D	23	28	3190	86	36	Maternal diabetes, breech, emergency caesarean	S	2	9-10		MRI 1 yr to 14 yrs: Asymmetry of lateral ventricles, likely physiological
20020P	40	F	H	29	41	3500	67		Planned caesarean	S	2	8-9	Epilepsy	MRI porencephaly, PVL
21021P	39	M	H	30	29	2940	7	34		S	1	NA	Mild ID	MRI day 5: acute left MCA territory infarct. Hypoxic-ischemic encephalopathy with seizures day 2
22022P	38	M	Q	24	32	2610	6		Emergency caesarean, fetal distress	S	4	NA		MRI day 2, 2 mths & 5 yrs: ischemic changes of basal ganglia bilaterally,

														consistent with severe hypoxic injury. Mild IUGR. Severe neonatal encephalopathy
23023P	39	F	Q	38	NA	3200	29	33.5	Cholestasis of pregnancy, maternal depression	S	5	9-9	Moderate ID, epilepsy	MRI at 7 mths: diffuse polymicrogyria, microcephaly. CGH array dup 2q37.3
24024P	41	F	H	24	29	3540	43		Fetal distress	S	1	4-7	Mild ID, epilepsy	No reports
25025P	40	M	H	32	33	3870	72	35		S	1	9-9		MRI at 7 mths: left PVL with porencephaly
26026P	39	F	D	36	36	2650	4	34		S	NA	9-9	Mild ID	MRI at 14 yrs: PVL Apneic episodes, migraines, low IgA, speech dyspraxia
27027P	29	M	Q	26	26	820	0		Maternal hypertension, Emergency caesarean	S	5	9-9	Mild ID	No reports Impaired hearing
28028P	41	F	H	32	29	3185	19			S	1	6-8	Epilepsy	MRI 10 yrs, 11 yrs, 16 yrs: left MCA territory infarct.
29029P	34	M	H	29	27	2120	13		Placental abruption, emergency caesarean	S	4	4-8	Severe ID	US day 3: normal. MRI at 4 yrs: PVL. Mosaic ring chromosome 22
30030P	38	M	Q	36	44	2052	0		Emergency caesarean, fetal distress	S	NA	3-6	Epilepsy	MRI at 8 yrs: PVL, probably subependymal grey matter heterotopias, scaphocephaly. Severe speech delay (no speech), cortical visual impairment, cataracts
31031P	28	M	H	24	25	710	0	25	Emergency caesarean	S	NA	-	Borderline ID, epilepsy	US day 2 to 3 mths: Grade IV IVH, with secondary hydrocephalus (required shunt). MRI at 16 mths and

													3 yrs: pontocerebellar hypoplasia in addition to changes related to previous haemorrhage
32032P	27	M	T	33	32	1030	32		Emergency caesarean	S	1	6-8		US day 1 to 6 wks: Grade III right IVH. MRI at 1 mth: PVL
33033P	38	M	Q	32	28	2870	15	33	Emergency caesarean	S	5	1-7	Severe ID, epilepsy	MRI at 4 mths and 4 yrs: PVL (L>R). Dystonia, de novo cytogenetically visible chromosome 8p abnormality, interpreted as interstitial duplication involving bands p23.1-p23.2 – "not necessarily the cause of his problems"
34034P	41	M	H	32	39	3700	33	36		S	1	NA	Epilepsy	MRI at 2 yrs: PVL
35035P	27	F	H	27	NA	680	1	25	Antenatal infection, maternal hypertension, breech, emergency caesarean	S	3	5-8	Moderate ID, epilepsy	US day 1 to 3 mths: left PVL. CT at 2 yrs and 5 yrs: mild ex-vacuo ventricular dilation and prominence extra-axial CSF spaces. Impaired hearing
36036P	41	F	H	31	28	2438	0	33	Emergency caesarean, fetal distress	S	1	9-9		MRI at 7 yrs and 8 yrs: PVL. Congenital hypothyroidism, IUGR
37037P	23	F	H	28	29	830	100		Breech	S	NA	NA		Craniofacial CT with 3D at 11 yrs: metopic and sagittal synostosis treated earlier in life, right porencephaly
38038P	31	M	H	24	27	1814	59		Breech	T	NA	NA		Medically controlled hypertension
39039P	38	M	D	34	37	3290	72	34.5		S	1	7-9		MRI at 2 yrs: delayed myelination

40040P	38	M	Q	33	42	3740	88		Antenatal infection, abrupt placenta, fetal distress	S	5	2-7	Severe ID, epilepsy	CT scan at 18 mths: global cerebral atrophy
41041P	40	F	H	31	31	3120	NA			S	NA	9-9	Epilepsy	CT at 8 yrs: extensive left hemispheric encephalomalacia
42042P	31	F	H	29	47	835	0		Emergency caesarean, fetal distress	S	NA	NA	Epilepsy	CT at 21 yrs: left cerebral hemisphere atrophy Heart murmur
43043P	41	F	H	33	35	3600	38	33	Emergency caesarean	S	1	5-8	Mild ID	CT at 3 yrs: old left MCA territory infarct
44044P	30	F	H	35	35	1400	31		Emergency caesarean, fetal distress	S	NA	NA		US 10 wks and 3 mths: hydrocephalus, right porencephaly. MRI at 17 mths: right PVL, bilateral ventriculomegaly, old bilateral IVH. Motor vehicle accident at 29 wks, resulting in head injury in utero
45045P	40	F	H	33	34	3470	45	35	Forceps birth, fetal distress	S	2	7-8		MRI at 7 mths: old right thalamic haemorrhage
46046P	38	F	D	24	NA	2975	48	34	Maternal hypertension, emergency caesarean, fetal distress	S	3	6-7		US day 2: global cerebral edema. MRI at 8 day: bilateral infarction of frontal and parietal lobes
47047P	29	M	Q	33	38	1000	3		Emergency caesarean	T	NA	NA		Cranial US day 3 to 6 wks: PVL (R>L)
49049P	35	M	H	32	32	2155	8	31.5		S	1	9-9		Mild right internal foot progression and decreased selective control at the ankle. Mild deviation in the saggital plane with equinus deformity and abnormal

														knee motion. MRI at 3 yrs: two subependymal nodules on the lateral wall of the right lateral ventricle
50050P	36	M	D	33	37	2980	65	34.5	Maternal epilepsy, maternal hypertension, forceps birth, fetal distress	S	NA	4-6		US day 2 and CT day 3: left extradural haemorrhage with bilateral cortical attenuation (frontal & perisylvian). MRI day 9 and 4 wks: left frontal subdural haemorrhage, PVL
51051P	38	M	H	38	46	3460	76			S	1	9-9		US at 18 mths: right porencephalic cyst
52052P	36	F	Q	38	39	3660	87		Breech, planned caesarean	S	5	10-10	Mild ID, epilepsy	CT MRI at 2 yrs: PVL with porencephalic cysts. Bilateral frontal pachygyria. PVL, porencephaly, haemorrhage in utero
53053P	36	M	Q	26	28	3830	99	35	Maternal diabetes, emergency caesarean, fetal distress	S	4	9-9	Mild ID, autism	US day 15: right IVH. CT at 1 yr: right periventricular calcification. Cortical visual loss, anxiety
55055P	30	F	H	28	28	1450	49		Assisted conception	T	2	NA		US day 2, day 5: normal. MRI at 3 yrs: PVL. Impaired hearing
56056P	37	M	D	34	NA	1956	1		Emergency caesarean	S	3	NA		CT at 13 mths: ex-vacuo ventricular dilation, with widening of subarachnoid spaces
57057P	29	M	H	22	27	1230	13			S	NA	NA		US day 3 to 6 wks: Grade III IVH then hydrocephalus (required shunt). MRI at 6 wks and 12 yrs: PVL, extensive cystic encephalomalacia (L>R),

														shunted hydrocephalus with obstructed aqueduct
58058P	38	M	Q	34	40	3170	34	33.5	Antenatal infection, planned caesarean	S	4	NA	Mild ID	MRI at 6 mths: diffuse poly-microgyria, foci of increased T1 signal through periventricular white matter. Described as 'likely congenital TORCH infection'
59059P	25	M	D	28	NA	770	46		Assisted breech delivery	S	2	4-6	Mild ID, autism	CT at 9 yrs: PVL
60060P	40	F	H	19	32	3025	11			S	NA	NA	Epilepsy	MRI at 4 yrs: bilateral frontal and perisylvian polymicrogyria, macrocephaly, ventriculomegaly
61061P	28	M	D	39	45	1450	98		Emergency caesarean	S	NA	NA		US day 1 to 2 mths: Grade IV left IVH. CT 3 mths, 1 yr, 2 yrs, 5 yrs: hydrocephalus (required shunt), left cerebral hemiatrophy, hypoplastic right cerebellar hemisphere
62062P	29	F	D	27	NA	1400	NA			S	4	NA		US day 5 to 2 mths: evolving PVL MRI at 2 mths: PVL
63063P	42	F	T	30	NA	3845	52			S	2	NA		MRI at 17 mths: likely PVL
64064P	27	M	T	36	37	1012	66			S	2	NA	Moderate ID, epilepsy	MRI at 7 yrs: PVL, post-ischemic bilateral encephalopathic cortical defects with right parietal schizencephaly. CT brain 13 yrs, 15 yrs: hydrocephalus (required shunt) and schizencephaly

65065P	33	F	H	34	29	1820	9	30	Antenatal infection	S	1	7-9		US day 4 to 6 wks: Grade II left IVH
66066P	25	M	D	25	24	720	24		Antenatal infection	S	NA	6-8	Autism	No reports
67067P	28	F	D	32	35	1300	74	26	Emergency caesarean, fetal distress	S	3	5-8		US day 2 to 3 mths: bilateral Grade II IVH, then PVL MRI at 3 yrs: PVL
68068P	39	M	H	31	30	3920	83	35		S	1	9-9		US at 6 mths: ex-vacuo dilatation of right ventricle. MRI 7 mths: right PVL and cortical involvement in region of right pre-central gyrus
69069P	38	F	H	28	29	3203	75		Emergency caesarean	S	1	NA		MRI at 8 mths: old left MCA territory infarct
70070P	34	M	H	29	28	1770	0		Emergency caesarean	T	1	NA		MRI: IVH, hydrocephalus
71071P	40	M	H	40	41	3640	35	35		S	2	7-9	Mild ID, epilepsy	MRI at 2 yrs: right PVL
72072P	40	M	Q	20	NA	3005	23		Fetal distress	S	4	8-10		US at 1 yr: left porencephaly with ex-vacuo left ventricular dilation
73073P	40	M	Dy	30	46	3520	30	35	Fetal distress	S	2	3-6		MRI at 23 mths: PVL
74074P	38	F	D	26	27	3884	93			S	2	NA		CT at 4 mths: normal. MRI at 6 mths and 4 yrs: normal
75075P	35	F	Q	32	33	1710	1	37.8	Emergency caesarean, fetal distress	S	NA	NA		US day 1, 5 and 13, and CT day 2: right frontoparietal haemorrhage. MRI at 4 wks: right frontoparietal cortical haemorrhage, bilateral PVL
76076P	37	F	D	28	31	2535	16	30.5	Planned caesarean	S	2	6-7	Autism	MRI at 14 mths and 8 yrs: normal. Microcephaly
77077P	29	F	H	14	NA	1425	75			S	1	NA		Absent septum pellucidum
78078P	29	M	D	28	29	1500	73		Emergency caesarean, fetal	S	3	NA		US day 2 to 6 wks: Asymmetry of lateral

									distress					ventricle (right larger than left). MRI at 22 mths: bilateral PVL. Sensory processing disorder
79079P	40	M	H	18	NA	2760	4		Instrumental delivery	S	1	NA		MRI at 18 mths: right PVL
80080P	26	M	Q	27	30	880	42			S	NA	NA		No Reports
81081P	36	F	H	24	18	2875	71	32.5	Breech, emergency caesarean	S	3	NA		MRI at 9 mths: right frontoparietal open schizencephaly with associated cortical dysplasia
82082P	39	F	Q	29	30	2380	1	32.5	Breech, emergency caesarean	S	4	NA	Moderate ID	US day 8 and 4 mths: calcification of basal ganglia, cystic areas bilaterally in occipital white matter. MRI at 5 mths: diffuse polymicrogyria. Consistent with congenital infection
83083P	25	M	D	36	33	770	28		Breech, emergency caesarean, fetal distress	S	3	NA		US day 2 to 5 mths: normal, then PVL. MRI at 11 mths: PVL
84084P	40	F	D	32	29	3742	77		Fetal distress	S	3	NA	Epilepsy	US day 2 to 4 mths: mild cerebral edema, then increased echogenicity of white matter. MRI at 9 yrs: changes of gliosis in thalami bilaterally, posterior limb of the left external capsule and high subcortical white matter of the parietal lobes/corticospinal tract
87096P	31	M	D	37	44	1390	6		Emergency caesarean	S	NA	NA		US day 2 to 5 wks: normal. MRI at 9 mths: PVL
88097P	30	M	D	33	38	1595	47		Emergency caesarean	S	NA	NA		US day 5 to 6 wks: normal. MRI at 4 yrs: PVL

89098P	28	F	D	31	40	1040	13		Emergency caesarean	S	2	NA		US day 2 to 17: increased echogenicity within periventricular region (day 2), else normal. MRI at 23 mths: PVL
90099P	32	M	D	34	36	1900	63	31.2		T	3	NA		US day 2 and day 8: normal
91100P	40	F	D	24	NA	2980	4		Fetal distress	S	2	NA		MRI at 4 yrs: normal. Severe dystonia
92101P	37	M	Q	32	33	3290	7		Planned caesarean, fetal distress	S	4	NA	ID	Brain MRI: cortical dysplasia. Severe dystonia
93102P	40	M	H	38	NA	3200	38			S	1	NA		MRI at 4 yrs and 15 yrs: left frontal and parietal cortical dysplasia
94103P	36	M	H	23	28	2892	67			S	NA	NA	Epilepsy, autism	No Reports
95104P	39	M	H	24	39	2490	1	35	Emergency caesarean, fetal distress	S	NA	NA		US day 10: bilateral Grade III IVH, ventriculomegaly. MRI at day 12 (following severe hypoglycemia and sepsis, severe thrombocytopenia): extensive bilateral basal ganglia, intraventricular and basal cistern haemorrhage with evolving moderate hydrocephalus, subacute spinal cord infarct.
96105P	38	F	Q	30	36	2490	4	35	Emergency caesarean	S	3	NA	ID, epilepsy	US day 8: echogenicity of periventricular white matter and left thalamus. MRI at 6 wks and 14 mths PVL. Neonatal seizures
97106P	40	M	D	31	30	3234	30			S	NA	NA		No Reports
98107P	23	F	H	33	NA	500	8		Fetal distress	T	NA	NA	Epilepsy	US day 2 to 4 mths: left Grade IV IVH,

													porencephaly, then hydrocephalus (required shunt). MRI at 2 yrs and 5 yrs: marked left hemiatrophy and encephalomalacia
99108P	23	M	H	21	24	500	4		Emergency caesarean, fetal distress	S	NA	NA		US day 2 to 3 mths: PVL
100109P	36	M	H	38	43	2110	1	32.5	Maternal hypertension, emergency caesarean, fetal distress	S	1	9-9	Epilepsy	CT at 6 mths and 8 yrs: large right MCA territory infarct
101110P	29	M	H	40	38	1690	91	30	Maternal diabetes, emergency caesarean, fetal distress	T	2	NA		US at day 2 to 3 mths: bilateral Grade IV IVH. PVL
102111P	40	M	Q	31	28	3440	53	33.5	Emergency caesarean, fetal distress	S	3	NA		US day 3: increased periventricular echogenicity. MRI at day 6: restricted diffusion with parasagittal and basal ganglia involvement (suggesting perinatal hypoxic ischemic injury) and PVL. Brain MRI at 2 mths: residual changes of PVL only
103112P	41	F	H	31	32	3815	71	35	Planned caesarean	S	4	NA	Epilepsy	MRI at 13 mths: Very large supratentorial teratoma Bilateral sensorineural hearing loss. Shunted hydrocephalus
104113P	40	M	Dy	27	30	3540	49		Emergency caesarean, fetal distress	S	4	3-5	ID	US day 4: normal. MRI day 11: bilateral symmetrical signal changes in the

														lentiform nuclei and perirolandic cortex consistent with changes of ischemia. Neonatal seizures, bilateral undescended testes/operated; No dysmorphic features
105114P	38	F	H	39	29	3420	72	34		S	NA	NA		MRI at 9 mths: Right MCA territory encephalomalacia, consistent with previous infarct
106115P	41	M	D	32	30	3725	34		Planned caesarean, fetal distress.	S	NA	NA	Moderate ID, epilepsy	US day 4: hydrocephalus. MRI at day 8, 2 yrs and 5 yrs: hydrocephalus communicating with Blake's pouch cyst, cerebellar folial disorganization, incomplete bilateral hippocampal inversion, optic nerve hypoplasia, partial agenesis of corpus callosum. Adducted thumbs, seizures, severe speech delay
107116P	26	F	Q	28	30	1020	93	25		S	4	6-9	ID	US day 2 to 2 mths: IVH (Grade II left, Grade II/III right), then ventriculomegaly, porencephaly
108117P	40	M	H	34	36	3540	67	34.5	Fetal distress	S	NA	8-9	Borderline ID, epilepsy	MRI day 3: right MCA territory infarction, with occlusion of origin of right MCA on MRA
109118P	40	M	H	30	31	3690	60		Emergency caesarean, fetal distress	S	1	NA	Autism	US 4mths: normal
110119P	39	F	H	34	38	2750	12	34.5		S	1	7-9	Epilepsy	US 6 mths: Mature left

														MCA territory infarct. MRI at 6 yrs: as above
111128P	36	F	Q	31	33	2585	53		Breech, planned caesarean	S	NA	3-8		Chronic renal failure
112129P	27	M	H	24	NA	920	35	23.5	Fetal distress	S	2	NA	Mild ID	US day 3 to 6 wks: prominent left periventricular flare, then extensive left cystic encephalomalacia. MRI 5 yrs: encephalomalacia of left cerebral hemisphere, basal ganglia and brainstem
119401P	35	F	H	34	37	2268	29		Breech, emergency caesarean	T	NA	NA		No Reports
120402P	26	F	D	34	31	825	26		Forceps delivery	T	NA	NA	Epilepsy	No Reports
121403P	35	M	D	39	NA	1945	8			S	1	7-9		Strabismus
122404P	29	F	D	29	NA	1105	13		Emergency caesarean	Trip	1	-		No Reports
123405P	30	M	D	23	28	1465	24			S	1	9-9		No Reports
124406P	30	F	H	NA	NA	NA	NA			NA	NA			No Reports
125407P	29	M	H	33	NA	1405	48			T	NA	NA		No Reports
126408P	36	M	H	33	NA	2500	18		Emergency caesarean, tight cord around neck, intrapartum fever	S	NA	NA		No Reports
127409P	37	M	H	22	NA	3515	81		Maternal diabetes, emergency caesarean, fetal distress	S	2	NA	Autism	Attention deficit hyperactivity disorder
128410P	40	F	H	31	31	2176	0		Emergency caesarean	S	1	NA	Epilepsy	No Reports
129411P	27	F	Q	40	NA	808	8		Breech, emergency caesarean	S	5	NA		No Reports
130412P	30	M	Q	36	NA	1134	2		Breech, emergency	S	NA	NA		No Reports

								caesarean					
131413P	37	M	D	29	43	4999	100	Emergency caesarean	NA	NA			No Reports
132414P	40	M	H	34	35	4210	91	Emergency caesarean, fetal distress	S	1	2-4	Epilepsy	Von Willebrand disease
133415P	39	F	D	40	46	3810	94	Breech, planned caesarean	S	2	9-10	Mild ID	Microcephaly
134416P	NA	M	D	NA	NA	NA	NA		NA	4	4-5		MRI: normal
135417P	39	M	H	31	35	4536	97	Maternal diabetes, forceps delivery, fetal distress	S	NA	NA		Temporal lobe infarct, PVL
136418P	36	F	H	23	NA	3374	97	Planned caesarean	S	NA	NA		No Reports
137419P	40	F	D	33	36	2790	5		S	3	9-9		Strabismus
138420P	38	M	H	31	31	2835	**33**		S	NA	5-9	Epilepsy	MRI: in utero 3rd trimester
139421P	26	F	Q	31	NA	790	28	IVF assisted conception, emergency caesarean	Trip	NA	NA		Non-verbal, ITB pump in situ
140422P	28	F	D	43	NA	1200	60		S	NA	NA		Dysplastic kidney, ureter in incorrect position
141423P	29	F	D	40	NA	1275	22		T	NA	NA		No Reports
142424P	38	M	D	28	30	3420	56	36	S	NA	9-9	Moderate ID	Hydrocephalus
143425P	38	M	H	36	NA	2885	16	Fetal distress	S	NA	9-9	Epilepsy	Stroke in utero
144426P	33	F	H	28	NA	1814	15	Emergency caesarean, fetal distress	T	2	NA	Epilepsy	Genital anomaly
145427P	38	M	H	35	40	3090	19	Breech, planned caesarean	S	2	NA		Stroke in utero
146428P	29	M	D	29	30	1245	NA	Emergency caesarean	Trip	2	4-7	Mild ID	No Reports
147429P	38	M	H	35	NA	3459	74		T	NA	NA	Epilepsy	No Reports
148430P	37	F	D	21	NA	3062	54		S	3	NA	Epilepsy	Non verbal

149431P	40	F	H	34	39	2665	4	Fetal distress	S	NA	NA		No Reports
150432P	41	F	D	33	NA	3487	22		S	2	NA	Autism	No Reports
151433P	38	F	Q	24	NA	2980	19	Tight cord around neck	S	5	NA	Epilepsy with seizure onset as neonate	Neonatal seizure
152434P	24	F	H	32	NA	600	11	Maternal fever, emergency caesarean, fetal distress.	S	NA	NA		PVL
153435P	39	M	D	32	NA	2800	3	Antenatal infection	S	1	U-9		Dyscalculia, attention deficit hyperactivity disorder, dyslexia
154436P	30	F	D	32	NA	1200	3	IVF, maternal diabetes, breech, emergency caesarean	T	3	NA		No Reports
155437P	28	M	Dy	30	NA	1180	38	Placental abruption, breech, emergency caesarean, fetal distress	S	5	2-2	Epilepsy	Bilateral deafness
156438P	33	M	H	33	NA	2948	100	Fetal distress	S	1	NA	Mild ID	Brain MRI: ventriculomegaly, degeneration of thalamus and the corpus cullosum
157439P	29	M	D	29	NA	665	0	Breech, planned caesarean, fetal distress	T	NA	NA	Epilepsy	Retinopathy of prematurity, cleft palate
158440P	35	M	H	20	NA	2515	57		S	NA	NA		No Reports
159441P	NA	M	H	NA	29	NA	NA		NA	1	NA		No Reports
160442P	24	M	Dy	30	NA	710	44		S	NA	NA		No Reports
161443P	39	F	H	27	36	3023	30	Emergency caesarean, fetal distress	S	1	NA		No Reports

162444P	37	F	H	23	NA	2695	35	Breech, emergency caesarean	S	NA	NA		No Reports
163445P	29	M	D	34	39	1050	3		S	3	NA		No Reports
164446P	27	F	D	NA	NA	965	29	Antenatal infection	S	2	NA		Hyperbilirubinemia, hypertonia
165447P	39	M	Dy	28	NA	3300	46		S	NA	NA		Left cerebral artery infarction
166448P	36	M	D	NA	NA	NA	NA		NA	2			No Reports
167449P	39	F	Dy	28	NA	3005	30	Instrumental delivery, fetal distress	S	NA	NA		No Reports
168450P	34	M	D	41	NA	2325	34	Emergency caesarean, fetal distress	S	NA	NA		No Reports
169451P	40	F	D	27	NA	3657	58		S	NA	NA		MRI at 2 yrs: hypoplasia of inferior cerebellar vermis, PVL Facial glabella hemangioma
170452P	27	M	Q	35	NA	1500	100	Emergency caesarean	S	4	NA		No Reports
171453P	41	F	D	25	NA	4220	90		S	2	NA	Epilepsy	No Reports
172454P	41	M	D	28	36	4337	79		S	1	NA		No Reports
173455P	27	M	D	26	NA	1000	70		S	4	NA		No Reports
174456P	27	M	Q	30	NA	1245	75	Abruptio placenta, emergency caesarean	S	1	NA	Epilepsy	Hydrocephalus
175457P	40	M	H	31	NA	3770	51	Fetal distress	S	NA	5-9	Mild ID, autism, epilepsy	Factor V Leiden mutation
176458P	42	F	H	31	NA	4025	43	Emergency caesarean	S	1	NA		No Reports
177459P	31	F	D	22	23	1700	45		S	2	NA		PVL
178460P	41	M	H	30	30	3785	37		S	1	9-10	Epilepsy	No Reports

179461P	28	M	D	31	NA	1145	34		Emergency caesarean	S	2	7-8		No Reports
180462P	40	F	D	34	NA	3240	22			S	4	NA	Epilepsy	Holoprosencephaly, diabetes insipidus, GERD, bilateral cleft lip and palate
181463P	38	F	Q	28	37	2565	5		Emergency caesarean, fetal distress	S	4	NA	Turner Syndrome	No Reports
182464P	33	M	D	23	NA	1900	34	30		S	2	NA	Severe ID	Periventricular leukomalacia, dystonia
183465P	34	F	D	36	NA	2268	32		Placental abruption, emergency caesarean	T	3	NA		Clotting disorder, vanishing twin
184466P	37	M	H	36	36	2600		33		S	1	9-10		MRI and CT both at 5 yrs: multiple foci altered signal intensity in deep white matter bilaterally, right frontal gliosis with slight dilation of right lateral ventricle. FVL in child and mother
185467P	35	F	D	28	NA	2580	49			S	3	NA	Epilepsy	No Reports
186468P	39	F	D	22	NA	2920	21	34		S	3	8-9	Mod-severe ID	No Reports
P187469	37	F	Dy	32	NA	2550	15		Maternal diabetes, emergency caesarean	S	5	NA		No Reports
188470P	39	M	H	33	NA	3590	46		Maternal liver infection	S	3	NA		PVL
189471P	27	F	D	29	NA	1080	78		Placental abruption, emergency caesarean	S	1	NA		No Reports
190472P	NA	M	D	19	NA	NA	NA		Emergency caesarean, fetal	S	2	NA	Wolff-Parkinson-	Unilateral renal dysfunction/failure,

								distress				White syndrome	spontaneously resolved, supraventricular tachycardia at birth, ataxia
191474P	40	M	H	29	NA	3969	65		S	5	NA		No Reports
192475P	36	M	H	35	36	3250	69		S	3	NA	Velo-Cardio-Facial syndrome	Brain MRI: polymicrogyria, heart: abnormal arch and branching, C22q11 deletion, T-cell immune-deficiency, mandibular osteomyelitis, laryngomalacia
193476P	39	F	Q	35	36	3290	50	IVF, TTTS (other T died), forceps delivery	T	4	NA		Microcephaly, bilateral closed lip schizencephaly
194477P	40	F	D	28	NA	3490	39	Forceps delivery	S	4	NA		No Reports

Abbreviations: GMFCS – Gross Motor Function Classification Score; S – singleton, T – twin, Tri - triplet; M – male, F- female; H – hemiplegia, D – diplegia, Q – quadriplegia, Dy – dyskinetic; ID – intellectual disability; FVL-Factor V Leiden gene, GERD-Gastro Esophageal Reflux Disease, IVF – in vitro fertilization, IUGR-Intrauterine growth restriction, TTTS – twin twin transfusion syndrome; MCA-middle cerebral artery, PVL- periventricular leukomalacia; SVT- supraventricular tachycardia; CT-Computed Tomography scan, MRI-Magnetic Resonance Imaging, US-Ultrasound; NA – not available.

Table 9.2 Supplementary Table B. WES read depth and coverage of genes containing the 16 predicted pathogenic variants.		
Gene	**Average Coverage**	**% < 20x**
AGAP1	80.6	35.5%
ATP11B	104.3	4.7%
CD99L2	134.1	11.3
JHDM1D	146.7	5.4%
KDM5C	70.9	6.1%
L1CAM	62.7	7.1%
MAST1	59.9	28.9%
NAA35	111.4	0.0%
PAK3	44.5	14.8%
RFX2	66.9	15.1%
SCN8A	144.5	7.1%
TENM1	68.8	3.7%
TUBA1A	319.7	0.0%
WIPI2	126	9.0%

Table 9.3 Supplementary Table C. Total number of *de novo* variants (n = 61) identified in 43 out of 98 case-parent trios (43%).

Cases	Gene	OMIM	Mut type	Genomic position	Accession number	cDNA	Protein change	RVIS	CADD	HI %	Brain Exp	PPH2 Score	PPH2 Pred	MT Score	MT Pred	GERP
7007P Dip Female	*CARS2*		Mis	Chr13:g.111353828C>G	NM_024537	c.350G>C	p.G117A	60.68	12.12	0.171	Mod	0.468	Pos D	0.015	N	4.45
	CDK17		Fs ins	Chr12:g.96717851C>CTG	NM_002595	c.158_159 insTG	p.S53fs	21.41	19.03	U	Mod	U	U	U	U	5.71
	RNF214		Mis	Chr11:g.117152637C>T	NM_001077239	c.1363C>T	p.P455S	22.51	13.18	U	Low	0.996	Prob D	0.186	N	4.76
8008P Trip Male	*ICMT*		Mis	Chr1:g.6285291C>T	NM_012405	c.704G>A	p.S235N	72.00	10.38	0.651	Mod	0	B	0	N	0
9009P Hem Male	*MAST1*		Mis	Chr19:g.12975755C>T	NM_014975	c.1499C>T	p.P500L	1.94	25.8	0.571	Mod	1	Prob D	0.785	U	4.39
12012P Dip ID, Au Female	*DHX32*		Mis	Chr10:g.127548338T>C	NM_018180	c.683A>G	p.Y228C	21.73	14.39	U	Low	0.988	Prob D	0.997	D	3.84
20020P Hem Ep Female	*TMEM194A*		Mis	Chr12:g.57457943A>T	NM_015257	c.446T>A	p.V149E	88.23	23.2	U	Low	1	Prob D	0.852	D	4.57
26026P Dip ID Female	*CENPF*		Mis	Chr1:g.214822118A>G	NM_016343	c.7931A>G	p.N2644S	99.81	13.88	0.844	Mod	0.131	B	0.071	N	1.38
	KDM5C	300534	Mis	ChrX:g.53240002G>A	NM_004187	c.1439C>T	p.P480L	3.50	22.5	U	Mod	1	Prob D	0.980	D	5.42
29029P Hem ID Male	*SSPO*		Fs ins	Chr7:g.149477410A >AGG	NM_198455	C1483_1484 insGG	p.R495fs	U	12.41	U	No	U	U	U	U	4.55
30030P Quad Ep Male	*TNKS1BP1*		Nfs del	Chr11:g.57075991 CTCCCTGGCCTCC AAGGGG>C	NM_033396	c.4176_4193 delTCCCTG GCCTCCA AGGGG	p.1392_1398 del	97.43	U	0.433	Mod	0.996	Prob D	U	U	U
	CTDSPL		Mis	Chr3:g.38009345C>T	NM_005808	c.365C>T	p.P122L	51.40	29.8	0.349	Mod	0.996	Prob D	1	D	5.25
32032P Trip Male	*LOC388946*		Mis	Chr2:g.46711491T>C	NM_001145051	c.586T>C	p.F196L	U	23.1	U		U	U	U	U	U
33033P Quad ID, Ep Male	*AGAP1*		Splice	Chr2:g.236708167G>A	NM_014914	c.957+1G>A		1.10	29.7	U	Mod	U	U	U	U	5.08
37037P Hem Female	*LENG8*		Splice	Chr19:g.54969489C>T	NM_052925	c.2032-4C>T		7.28	U	0.260	Mod	U	U	U	U	-10.4

38038P Hem Male	*SLA*		Mis	Chr8:g.134050969G>A	NM_001045556	c.631C>T	p.P211S	46.74	0.443	0.668	High	0	B	0.003	N	1.72
39039P Dip Male	*CNDP2*		Mis	Chr18:g.72168642G>A	NM_018235	c.139G>A	p.V47I	13.67	6.85	0.393	Mod	0.034	B	0.427	N	3.99
43043P Hem ID Female	*ATP11B* *SCN8A* *SIPA1L1*	 614306 614558	Mis Mis Mis	Chr3:g.182605492C>T Chr12:g.52162895G>A Chr14:g.72137984C>T	NM_014616 NM_014191 NM_015556	c.2834C>T c.3148G>A c.2404C>T	p.T945I p.G1050S p.R802W	24.60 2.34 0.70	4.3 13.21 20.5	0.331 0.632 0.556	Low Low Mod	0.025 0.638 1	B Pos D Prob D	0.213 U 0.995	N U D	4.50 4.46 4.21
44044P Hem Female	*WIPI2*		Mis	Chr7:g.5266876A>G	NM_001033520	c.737A>G	p.Y246C	7.43	25.7	0.181	Mod	0.999	Prob D	0.999	D	5.93
45045P Hem Female	*ABLIM2*		Mis	Chr4:g.8010820C>T	NM_001130083	c.1433G>A	p.R478Q	3.54	2.87	U	Low	0.988	Prob D	U	U	1.95
52052P Quad ID, Ep Female	*RPF2* *C20orf141*		Mis Mis	Chr6:g.111312985C>T Chr20:g.2795907G>A	NM_032194 NM_080739	c.230C>T c.77G>A	p.S77L p.G26E	64.51 76.81	U 12.49	U 0.019	Low No	0.629 0.940	Pos D Pos D	0 0.768	N U	5.54 2.85
58058P Quad ID Male	*PTPN14* *INHBB*	613611	Splice Mis	Chr1:g.214546051T>G Chr2:g.121106698C>T	NM_005401 NM_002193	c.3036+3 A>C c.472C>T	 p.R158C	32.28 U	U 16.25	0.344 0.410	Low Mod	U 0.983	U Prob D	U 0.442	U N	4.27 5.09
67067P Dip Female	*UBXN7*		Mis	Chr3:g196120470C>T	NM_015562	c.310G>A	p.A104T	38.58	15.99	U	Mod	0.007	B	0.999	D	4.47
71071P Hem ID, Ep Male	*LTN1*		Fs del	Chr:21:g.30318165CCACT>C	NM_015565	c.3866_3869 delCACT	p.1289_1290 del	70.06	41	U	Low	U	U	U	U	4.8
82082P Quad ID Female	*NEMF*		Fs del	Chr14:g.50296116TTATC>T	NM_004713	c.970_973 delTATC	p.324_325del	2.54	29.8	U	Low	U	U	U	U	5.34
83083P Dip Male	*SPATS1*		Mis	Chr6:g.44328209T>C	NM_145026	c.314T>C	p.L105P	69.21	3.6	0.146	No	0	B	0	N	2.87
89098P Dip Female	*TEP1*		Mis	Chr14:g.20869194G>C	NM_007110	c.1498C>G	p.R500G	99.71	20.2	0.336	No	0.999	Prob D	0.613	D	4.80
92101P Quad ID Male	*NGFR*		Mis	Chr17:g.47589394C>T	NM_002507	c.962C>T	p.T321M	38.82	16.09	0.555	Low	0.223	B	0.688	D	5.18

94103P Hem Ep, Au Male	*STAB2*		Nfs del	Chr12:103981262 TACA>T	NM_01764	c.9_11del ACA	p.3_4del	0.72	10.86	0.156	No	U	U	U	U	4.81
96105P Quad ID, Ep Female	*PHF17* *JHDM1D*		Mis Mis	Chr4:g.129770233G>A Chr7:g.139796549G>C	NM_199320 NM_030647	c.395G>A c.2180C>G	p.R132Q p.S727W	21.73 40.56	9.88 21.1	0.796 U	Mod Low	0.125 1	B Prob D	0.858 0.987	D D	4.69 5.85
97106P Dip Male	*ZNF160*		Mis	Chr19:g.53577434C>A	NM_198893	c.230G>T	p.R77I	68.49	11.84	U	Low	0.847	Pos D	0.004	N	0.769
102111P Quad Male	*EIF4E2*		Mis	Chr2:g.233422672C>G	NM_004846	c.214C>G	p.P72A	30.07	15.3	0.285	Mod	0.002	B	0.542	D	5.27
105114P Hem Female	*EPHA1*		Mis	Chr7:g.143095852G>A	NM_005232	c.1178C>T	p.S393L	10.17	4.87	0.672	Low	0.994	Prob D	0.906	D	2.37
106115P	*C10orf131*		Mis	Chr10:g.97686995T>C	NM_001130446	c.218T>C	p.L73S	U	U	U	No	0.144	B	U	U	4.08
Dip	*TUBA1A*	611603	Mis	Chr12:g.49580101G>A	NM_006009	c.367C>T	p.R123C	35.42	12.3	U	High	0.997	Prob D	0.995	D	3.18
ID, Ep	*MYO15A*	600316	Mis	Chr17:g.18051510T>C	NM_016239	c.6677T>C	p.V2226A	5.35	17.94	0.464	Low	0.996	Prob D	0.921	D	4.1
Male	*ADCY3*		Mis	Chr2:g25062743C>G	NM_004036	c.1354G>C	p.G452R	1.42	30	0.477	Low	1	Prob D	0.999	D	5.05
108117P	*NAA35*		Mis	Chr9:g.88631481G>T	NM_024635	c.1596G>T	p.W532C	43.77	23.1	U	Low	1	Prob D	0.999	D	5.4
Hem	*NID2*		Mis	Chr14:g.52474638G>A	NM_007361	c.3770C>T	p.T1257M	99.53	13.43	0.815	Low	0.997	Prob D	0.055	N	-0.615
ID, Ep	*PCBP3*		Mis	Chr21:g.47355203T>C	NM_001130141	c.815T>C	p.L272P	15.62	19.34	0.450	Low	0.933	Pos D	0.999	D	3.72
Male																
121403P Dip Male	*TMEM150A*		Mis	Chr2:g.85826777G>A	NM_001031738	c.445C>T	p.P149S	25.15	16.61	U	Low	0.871	Pos D	0.984	D	-2.05
123405P Dip Male	*CDH26*		Mis	Chr20:g.58571005G>T	NM_177980	c.1784G>T	p.C595F	89.58	13.78	U	No	0.999	Prob D	0.628	D	4.56
132414P Hem VWD, Ep Male	*MAN2A1*		Mis	Chr5:g.109159498T>A	NM_002372	c.2526T>A	p.F842L	3.20	8.53	0.516	Low	0	B	0.135	N	4.37
146428P Dip ID	*UBQLN3*		Fs del	Chr11:g.5530218CCAG CTGG>C	NM_017481	c.564_570 delCAGCT GG	p.188_190del	87.24	25.5	0.267	No	U	U	U	U	4.63
Male	*ZNF646*		Mis	Chr16:g.31089186A>G	NM_014699	c.1541A>G	p.H514R	68.55	16.07	0.350	Low	0.999	Prob D	0.311	N	5.77
153435P Dip Male	*ZNF267*		Mis	Chr16:g.31926517A>G	NM_003414	c.947A>G	p.H316R	91.89	13.86	0.527	Low	0.994	Prob D	0.004	N	0.458
154436P	*GANC*		Mis	Chr15:g.42632088G>C	NM_198141	c.2065G>C	p.V689L	75.71	10.57	0.166	Low	0	B	0.061	N	4.15
Dip	*SREK1*		Mis	Chr5:g.65449298C>T	NM_001077199	c.169C>T	p.P57S	33.20	13.89	U	Low	0.017	B	0.670	D	5.61

Female															
161443P Hem Female	*NR1I2* *RFX2*		Splice Mis	Chr3:g.119534572G>T Ch19:g.6040241T>C	NM_0033013 NM_000635	c.943-1G>T c.272A>G	 p.Y91C	19.86 10.89	17.78 U	0.590 0.820	No Low	U 0.971	U Pos D	U 0.999	U D	5.04 5.36
168450P Dip Male	*ENPP4* *RAD21*	 614701	Stopgain Mis	Chr6:g.46107658G>A Chr8:g.117864308C>T	NM_014936 NM_006265	c.338G>A c.1349G>A	p.W113X p.R450H	75.29 21.41	38 13.52	0.139 0.962	Low High	0.736 0.140	Pos D B	1 1	D D	5.97 4.65
169451P Dip Female	*TUBA1A*	611603	Mis	Chr12:g.49579694A>T	NM_006009	c.455T>A	p.L152Q	35.42	13.99	U	High	0.785	Pos D	0.996	D	5.15
172454P Dip Male	*MIIP*		Splice	Chr1:g.12082151G>A	NM_021933	c.114-1G>A		85.24	9.5	U	Mod	U	U	U	U	3.47
178460P Hem Ep Male	*ZFYVE16*		Mis	Chr5:g.79743924A>G	NM_001105251	c.2804A>G	p.E935G	93.08	16.07	0.617	Low	0.992	Prob D	0.021	N	5.43
183465P Dip Female	*PTGFRN*		Mis	Chr1:g.117509692A>G	NM_020440	c.1799A>G	p.N600S	36.28	17.82	0.379	Mod	0.991	Prob D	0.844	D	5.56
Abbreviations: OMIM – Online Mendelian Inheritance in Man; RVIS – Residual Variation Intolerance Score; CADD – Combined Annotation-Dependent Depletion; HI – Haploinsufficiency Index; PPH2 Pred – Polyphen2 prediction, MT – Mutational Taster; Dip – diplegia, Hem – hemiplegia, Quad – quadriplegia, Trip – triplegia; ID – intellectual disability, Ep – epilepsy, Au – autism; Mis – missense, Fs del – frameshift deletion, Nfs del – non frameshift deletion; Prob D – probably damaging, Pos D – possibly damaging, B – benign, D – disease causing, N – polymorphism, U – unknown.															

Table 9.4 Supplementary Table D. Total number of X-chromosome variants (n = 19) inherited from an unaffected mother to affected son identified in 14 out of 98 case- parent trios (14%).

Cases CP type	Gene	OMIM	Mut type	Genomic position	NM number	cDNA	Protein change	RVIS %	CADD	HI %	Brain Exp	PPH2 Score	PPH2 Pred	MT Score	MT Pred	GERP
15015P Hem Ep Male	*PAK3*	300558	Mis	ChrX:g.110459718C>T	NM_001128166	c.1477C>T	p.R493C	26.23	21.2	U	Mod	0.91	Pos D	0.999	D	5.54
17017P Quad Male	*UPF3B*	300676	Splice	ChrX:g.118986991C>T		U		55.22	1.46	0.354	High	U	U	U	U	-1.47
19019P	*TENM1*		Mis	ChrX:g.123514967C>T	NM_014253	c.7597G>A	p.G2533S	2.62	22.8	0.726	Low	0.993	Pos D	0.891	D	5.83
Dip	*IL1RAPL1*		Mis	ChrX:g.29973540A>C	NM_014271	c.1694A>C	p.E565A	17.17	13.81	0.759	Low	0.881	Pos D	0.977	D	5.25
Male	*ATP1B4*		Mis	ChrX:g.119513352G>T	NM_001142447	c.937G>T	p.A313S	65.75	26.1	0.278	No	0.991	Prob D	0.999	D	5.72
25025P Hem Male	*MAOB*		Splice	ChrX:g.436334520C>T		c.1137-1G>A		17.03	13.81	0.585	Low	U	U	U	U	6.17
38038P	*ZDHHC9*	300799	Mis	ChrX:g.128975851C>T	NM_016032	c.71G>A	p.C24Y	44.53	10.07	0.629	Mod	0.255	B	0.998	D	5.51
Hem	*OFD1*	300804	Mis	ChrX:g.13774751A>G	NM_003611	c.1276A>G	p.M426V	40.56		0.217	Low	0.947	Pos D	0	N	-8.12
Male		311200														
39039P Dip Male	*CD99L2*		Stopgain	ChrX:g.149944668T>A	NM_134446	c.487T>A	p.K163X	34.31	32	0.15	Mod	0.735	Pos D	1	D	5.06
40040P Quad ID, Ep Male	*MTMR1*		Splice	ChrX:g149905063G>A	NM_03828	c.1056-4G>A		46.74	1.52	0.449	Low	U	U	U	U	1.84
51051P Hem Male	*IGSF1*		Mis	ChrX:g130413296T>A	NM_001170962	c.1639A>T	p.T547S	17.47	14.89	0.679	Low	0.897	Pos D	0.105	D	5.04
61061P Dip Male	*BRWD3*	300659	Mis	ChrX:g.79965045A>G	NM_153252	c.2357T>C	p.L786S	22.64	8.66	0.725	Low	0	B	0.93	D	5.47
83083P	*IQSEC2*	309530	Mis	ChrX:g.53267455T>C	NM_001111125	c.3149T>C	p.V1050A	21.2	18.95	U	Low	0.654	B	0.877	D	4.81
Dip	*SLITRK2*		Mis	ChrX:g.144905474G>A	NM_001144006	c.1531G>A	p.V511M	12.88		0.239	Mod	0.865	Pos D	0	N	5.84
Male																
106115P Dip ID, Ep Male	*CNKSR2*		Mis	ChrX:g.21508607A>G	NM_001168649	c.592A>G	p.I198V	23.25	18.81	0.748	Low	0.978	Pos D	0.994	D	5.83
108117P	*CUL4B*	300354	Mis	ChrX:g.119694175G>C	NM_001079872	c.319C>G	p.L107V	28.63	5.92	0.840	Mod	0	B	0.505	D	1.95
Hem	*USP26*	300676	Mis	ChrX:g.132161343G>C	NM_031907	c.906C>G	p.N302K	78.69	12.41	U	High	0.999	Prob D	0.662	D	0.143
ID, Ep Male																

123405P Dip Male	*ENOX2*		Splice	ChrX:g.129822993C>G	NM_182314	c.97-1G>C		25.36	19.47	0.215	Low	U	U	U	U	5.1
156438P Hem ID Male	*L1CAM*	304100 303350 307000	Mis	ChrXg:153136539G>C	NM_001143963	c.481C>G	p.P161A	4.92	11.11	U	High	0.929	Pos D	0.896	D	5.12

Abbreviations: OMIM – Online Mendelian Inheritance in Man; RVIS – Residual Variation Intolerance Score; CADD – Combined Annotation-Dependent Depletion; HI – Haploinsufficiency Index; PPH2 Pred – Polyphen2 prediction, MT – Mutational Taster; Dip – diplegia, Hem – hemiplegia, Quad – quadriplegia; ID – intellectual disability, Ep – epilepsy, Mis – missense; Prob D – probably damaging, Pos D – possibly damaging, B – benign, D – disease causing, N – polymorphism, U – unknown.

Table 9.5 Supplementary Table E. Compound heterozygous variants in a known disease gene with overlapping features with cerebral palsy.

Cases CP type	**Inheritance**	**Gene**	**OMIM**	**Mut type**	**Genomic position**	**NM number**	**cDNA**	**Protein change**	**RVIS %**	**CADD**	**HI %**	**Brain Exp**	**PPH2 Score**	**PPH2 Pred**	**MT Score**	**MT Pred**	**GERP**
49049P	Paternal	*HSPG2*	224410	Splice	Chr1:g.22157706A>T	NM_005529	c.11562+2T>A		32.07	21.3	0.869	Low	U	U	U	U	5.46
Hem	Maternal		255800	Mis	Chr1:g.22170696G>T		c.8561C>A	p.A2854D		22.6			0.911	Pos D	0.738	D	4.43
Male	Maternal			Mis	Chr1:g.22173966C>T		c.8045G>A	p.R2682Q		22.3			1	Prob D	0.855	D	5.5
	Maternal			Splice	Chr1:g.22178034C>T		c.7158+5G>A			8.7			U	U	U	U	-2.73

Abbreviations: OMIM – Online Mendelian Inheritance in Man; RVIS – Residual Variation Intolerance Score; CADD – Combined Annotation-Dependent Depletion; HI – Haploinsufficiency Index; PPH2 Pred – Polyphen2 prediction, MT – Mutational Taster; Hem – hemiplegia, Mis – missense, Prob D – probably damaging, Pos D – possibly damaging, B – benign, D – disease causing, N – polymorphism, U – unknown.

Table 9.6 Supplementary Table F. Percentage (%) of variants in nine quadrants defined by the various RVIS and CADD threshold values (Supplementary Figures B – I). Red represents the two cut-off threshold values for RVIS (<25 and <50) and CADD (>20 and >10) considered potentially pathogenic.

RVIS:CADD Threshold values	**CP *De novo* % Figure B**	**CP X-chrom % Figure C**	**ID *De novo* % Figure D**	**ID X-chrom % Figure E**	**SZ *De novo* % Figure F**	**Autism *De novo* % Figure G**	**Autism *De novo* % Figure H**	**Autism *De novo* % Figure I**
<25:>20	14	4	33	6	21	12	17	21
<25:>10<20	20	34	18	4	17	17	17	17
<50>25:>20	4	8	10	8	7	2	5	6
<50>25:>10<20	13	13	4	15	9	9	6	4
<25:<10	9	8	5	10	5	17	10	9
<50>25:<10	2	8	3	6	3	5	5	5
>50:>20	13	4	7	10	3	7	10	9
>50:>10<20	23	13	11	33	18	14	16	13
>50:<10	2	8	9	8	17	17	14	16

Abbreviations: RVIS – Residual Variation Intolerance Score, CADD – combined annotation detection depletion, Chrom - chromosome, SZ – schizophrenia

Table 9.7 Supplementary Table G. Multispecies alignment across nine different species for 12 missense mutations predicted to be causative.

MAST1 p.P500L

Mutant	D	L	K	L	D	N	G
Human	D	L	K	P	D	N	G
Rhesus	D	L	K	P	D	N	G
Mouse	D	L	K	P	D	N	G
Dog	D	L	K	P	D	N	G
Elephant	D	L	K	P	D	N	G
Chicken	D	L	K	P	D	N	G
X-tropicalis	D	L	K	P	D	N	G
Zebrafish	D	L	K	P	D	N	G
Lamprey	D	L	K	P	D	N	G

WIPI2 p.Y246C

Mutant	S	T	S	C	L	P	S
Human	S	T	S	Y	L	P	S
Rhesus	S	T	S	Y	L	P	S
Mouse	S	T	S	Y	L	P	S
Dog	S	T	S	Y	L	P	S
Elephant	S	T	S	Y	L	P	S
Chicken	S	T	S	Y	L	P	S
X-tropicalis	A	S	N	Y	L	P	S
Zebrafish	S	T	T	Y	L	P	S

TENM1 p.G2533S

Mutant	T	V	I	S	D	K	I
Human	T	V	I	G	D	K	I
Rhesus	T	V	I	G	D	K	I
Mouse	T	V	I	G	E	K	I
Dog	T	V	I	G	D	K	I
Elephant	T	V	I	G	D	K	I
Chicken	T	V	I	G	D	K	I
X-tropicalis	T	V	=	G	E	K	I
Zebrafish	A	V	G	G	D	R	I
Lamprey	Y	T	R	G	H	N	N

JHDM1D p.S727W

Mutant	T	S	T	W	T	P	C
Human	T	S	T	S	T	P	C
Rhesus	T	S	T	S	T	P	C
Mouse	T	S	T	S	T	P	C
Dog	T	S	T	S	T	P	C
Elephant	T	S	T	S	T	P	C
Chicken	T	S	T	S	T	P	C
X-tropicalis	M	S	T	S	T	P	C

NAA35 p.W532C

Mutant	L	Y	A	C	L	M	S
Human	L	Y	A	W	L	M	S
Rhesus	L	Y	A	W	L	M	S
Mouse	L	Y	A	W	L	M	S
Dog	L	Y	A	W	L	M	S
Elephant	L	Y	A	W	L	M	S
Chicken	L	Y	A	W	L	M	S
X-tropicalis	L	Y	A	W	L	M	S
Zebrafish	L	Y	A	W	L	M	S
Lamprey	L	Y	G	W	L	I	S

RFX2 p.Y91C

Mutant	N	Y	T	C	A	T	R
Human	N	Y	T	Y	A	T	R
Rhesus	N	Y	T	Y	A	T	R
Mouse	N	Y	A	Y	A	A	R
Dog	N	Y	A	Y	A	T	R
Elephant	N	Y	T	Y	A	T	R
Chicken	N	Y	S	Y	A	T	R
X-tropicalis	N	Y	S	Y	A	T	R
Zebrafish	N	Y	S	Y	A	A	R

KDM5C p.P480L

Mutant	E	L	V	L	M	V	N
Human	E	L	V	P	M	V	N
Rhesus	E	L	V	P	M	V	N
Mouse	E	L	V	P	M	V	N
Dog	E	L	V	P	M	V	N
Elephant	E	L	V	P	M	V	N
Chicken	E	L	I	P	M	N	N
X-tropicalis	E	L	V	P	M	V	N
Zebrafish	E	L	V	P	M	V	N
Lamprey	P	L	V	P	M	N	N

SCN8A p.G1050S

Mutant	N	H	T	S	A	D	I
Human	N	H	T	G	A	D	I
Rhesus	N	H	T	G	A	D	I
Mouse	N	H	T	G	V	D	I
Dog	N	H	T	G	A	D	I
Elephant	N	H	T	G	A	D	I
Chicken	N	H	T	G	A	D	I
X-tropicalis	N	N	T	G	A	D	I
Zebrafish	N	H	T	G	V	D	I
Lamprey	N	H	T	G	V	Q	V

TUBA1A p.R123C

Mutant	C	L	K	C	I	R	D
Human	C	L	K	R	I	R	D
Rhesus	C	L	K	R	I	R	D
Mouse	C	L	K	R	I	R	D
Dog	C	L	K	R	I	R	D
Elephant	C	L	K	R	I	R	D
Chicken	C	L	K	R	I	R	D
X-tropicalis	C	L	K	R	I	R	D
Zebrafish	C	L	K	R	I	R	D
Lamprey	C	L	K	R	I	R	D

TUBA1A p.L152Q

Mutant	E	M	L	Q	S	T	F
Human	E	M	L	L	S	T	F
Rhesus	E	M	L	L	S	T	F
Mouse	E	M	L	L	S	T	F
Dog	E	M	L	L	S	T	F
Elephant	E	M	L	L	S	T	F
Chicken	E	M	L	L	S	T	F
X-tropicalis	E	M	L	L	S	T	F
Zebrafish	E	M	L	L	S	T	F
Lamprey	E	M	L	L	A	A	F

PAK3 P.R493C

Mutant	A	V	F	C	D	F	L
Human	A	V	F	R	D	F	L
Rhesus	A	V	F	R	D	F	L
Mouse	A	V	F	R	D	F	L
Dog	A	V	F	R	D	F	L
Elephant	T	V	F	R	D	F	L
Chicken	A	V	F	R	D	F	L
X-tropicalis	A	V	F	R	D	F	L
Zebrafish	P	I	F	R	D	F	L

L1CAM p.P161A

Mutant	I	R	L	A	E	A	S
Human	I	R	L	P	E	A	S
Rhesus	I	R	L	P	E	A	S
Mouse	I	R	L	P	A	A	S
Dog	I	R	L	P	E	A	S
Elephant	I	R	P	P	E	A	S
Chicken	=	=	=	=	=	=	=
X-tropicalis	V	R	P	P	V	A	S
Zebrafish	I	K	P	S	L	V	R
Lamprey	I	V	P	L	P	I	G

Table 9.8 Supplementary Table H. Total number of X-chromosome variants (n = 7) inherited from an unaffected mother to affected son identified in 7 out of 67 duo cases (7%). The 67 duo cases included 13 paternal and 54 maternal DNA samples**.**

Cases	Gene	OMIM	Mut Type	Genomic position	NM number	cDNA	Protein change	RVIS%	CADD	HI%	Brain Exp	PPH2 score	PPH2 Pred	MT Score	MT Pred	GERP
31031P Hem ID, Ep Male	*ARMCX2*		Mis	ChrX:g.100910891C>T	NM_177949	c.1684G>A	p.V562M	22.35	22.35	0.296	Mod	0.785	Pos D	0.1	N	4.1
56056P Dip Male	*UPF3B*	300676	Splice	ChrX:g.118986991				55.22	1.46	0.354	High	U	U	U	U	-1.47
78078P Dip Male	*ZNF674*		Mis	ChrX:g.70472607C>A	NM_001171162	c.499A>G	p.T167A	91.26	4.57	U	Low	0.778	Pos D	0.005	N	-0.96
93102P Hem Male	*ZMYM3*		Mis	ChrX:g.70472607T>C	NM_001171162	c.499A>G	p.T167A	5.97	0.116	0.666	High	0.778	Pos D	0.005	N	-0.96
125407P Hem Male	*MUM1L1*		Mis	ChrX:g.105449828C>T	NM_152432	c.403C>T	p.R135W	69.36	69.36	0.162	No	U	U	U	U	1.82
142424P Dip ID Male	*L1CAM*	304100 303350 307000	Mis	ChrX:g.153135577C>T	NM_001143963	c.910G>A	p.E304K	4.92	19.32	U	High	0.891	Pos D	0.997	D	4.58
165447P Dy Male	*PAK3*	300558	Mis	ChrX:g.110389772	NM_001128173	c.300G>A	p.M100I	26.23	12.37	U	Mod	U	U	U	U	5.35

Abbreviations: OMIM – Online Mendelian Inheritance in Man; RVIS – Residual Variation Intolerance Score; CADD – Combined Annotation-Dependent Depletion; HI – Haploinsufficiency Index; PPH2 Pred – Polyphen2 prediction, MT – Mutational Taster; Dip – diplegia, Hem – hemiplegia, Quad – quadriplegia, Dy – dyskinetic; ID – Intellectual disability; Mis – missense; Prob D – probably damaging, Pos D – possibly damaging, B – benign, D – disease causing, N – polymorphism, U – unknown.

Table 9.9 Supplementary Table I. Variants in known disease genes (n = 7) identified in 5 out of 18 (27%) singleton cases, which met our prioritization criteria including RVIS<50th percentile and CADD>20.

Cases	Gene	OMIM	Mut Type	Genomic position	NM number	cDNA	Protein change	RVIS %	CADD	HI%	Brain Exp	PPH2 Score	PPH2 Pred	MT Score	MT Pred	GERP
53053P Quad ID, Au Male	*HUWE1* *HUWE1* *KIF1A*	300706 300705 614255 614213 610357	Mis Nfs ins Splice	ChrX:g.53607865T>C ChrX:g.53574640C>CAGT Chr2:g.241715346G>A	NM_031407 NM_031407 NM_004321	c.5642A>G c.10630_10631insACT c.882-3C>T	p.N1881S p.V3544delinsTV	0.39 0.39 0.26	19.39 17.41 12.53	0.84 0.84 U	High High High	0.756 U U	Pos D U U	0.961 U U	D U U	5.84 4.33 2.12
60060P Hem Ep Female	*KANK1*	612900	Mis	Chr9:g.711483C>G	NM_153186	c.243C>G	p.I81M	9.78	14.64	0.653	Low	0.954	Pos D	0.549	U	2.81
152434P Hem Female	*HLCS*	253270	Mis	Chr21:g.38137354C>T	NM_000411	c.1639G>A	p.V547M	18.25	12.67	0.147	Low	0.075	B	0.002	N	0.566
177459P Dip Female	*DMD*	310200 300376 302045	Mis	ChrX:g.32382808T>G	NM_004011	c.1022A>C	p.E341A	11.28	15.95	0.493	Low	0.644	Pos D	0	N	5.38
186468P Dip ID Female	*SPAST*	182601	Mis	Chr2:g.32362015A>T	NM_199436	c.1295A>T	p.E432V	48.54	25.3	U	Mod	0.999	Prob D	0.999	D	5.63

Abbreviations: OMIM – Online Mendelian Inheritance in Man; RVIS – Residual Variation Intolerance Score; CADD – Combined Annotation-Dependent Depletion; HI – Haploinsufficiency Index; PPH2 Pred – Polyphen2 prediction, MT – Mutational Taster; Dip – diplegia, Hem – hemiplegia, Quad – quadriplegia, Dy – dyskinetic; au – autism; Nfs ins – non-frameshift insertion, Mis – missense; Prob D – probably damaging, Pos D – possibly damaging, B – benign, D – disease causing, N – polymorphism, U – unknown.

Table 9.10 Supplementary Table J. Brain imaging data (112 out of 183 cases) correlated with genetic results.

Cases	Gestation (weeks)	Gender	CP type	Birth Weight centile	Potentially causative genes
Intraventricular Haemorrhage (IVH) (n = 21)					
1001P	26	Female	Quadriplegia	9	
101110P	29	Male	Hemiplegia	91	
107116P	26	Female	Quadriplegia	93	
12012P	29	Female	Diplegia	64	
13013P	40	Female	Hemiplegia	60	
15015P	29	Male	Hemiplegia	1	*PAK3*
2002P	29	Female	Diplegia	89	
3003P	32	Female	Diplegia	36	
31031P	28	Male	Hemiplegia	0	
44044P	30	Female	Hemiplegia	31	*WIPI2*
5005P	32	Female	Hemiplegia	5	
53053P	36	Male	Quadriplegia	99	
57057P	29	Male	Hemiplegia	13	
59059P	25	Male	Diplegia	46	
61061P	28	Male	Diplegia	98	
64064P	27	Male	Triplegia	66	
65065P	33	Female	Hemiplegia	9	
67067P	28	Female	Diplegia	74	
70070P	34	Male	Hemiplegia	0	
95104P	39	Male	Hemiplegia	1	
98107P	23	Female	Hemiplegia	8	
White mater damage (periventricular leukomalacia) (n = 46)					
102111P	40	Male	Quadriplegia	53	
104113P	40	Male	Dykinetic	49	
11011P	30	Female	Diplegia	NA	
112129P	27	Male	Hemiplegia	35	
152434P	24	Female	Hemiplegia	11	
177459P	31	Female	Diplegia	45	
18018P	29	Female	Diplegia	18	
182464P	33	Male	Diplegia	34	
184466P	37	Male	Hemiplegia	NA	
188470P	39	Male	Hemiplegia	46	
20020P	40	Female	Hemiplegia	67	
22022P	38	Male	Quadriplegia	6	
25025P	40	Male	Hemiplegia	72	
26026P	39	Female	Diplegia	4	*KDM5C*
29029P	34	Male	Hemiplegia	13	
30030P	38	Male	Quadriplegia	0	
32032P	27	Male	Triplegia	32	
33033P	38	Male	Quadriplegia	15	*AGAP1*
34034P	41	Male	Hemiplegia	33	
35035P	27	Female	Hemiplegia	1	

36036P	41	Female	Hemiplegia	0	
37037P	23	Female	Hemiplegia	100	
4004P	40	Female	Hemiplegia	30	
40040P	38	Male	Quadriplegia	88	
41041P	40	Female	Hemiplegia	NA	
47047P	29	Male	Quadriplegia	3	
50050P	36	Male	Diplegia	65	
51051P	38	Male	Hemiplegia	76	
55055P	30	Female	Hemiplegia	49	
56056P	37	Male	Diplegia	1	
62062P	29	Female	Diplegia	NA	
63063P	42	Female	Triplegia	52	
68068P	39	Male	Hemiplegia	83	
72072P	40	Male	Quadriplegia	23	
73073P	40	Male	Dykinetic	30	
78078P	29	Male	Diplegia	73	
79079P	40	Male	Hemiplegia	4	
83083P	25	Male	Diplegia	28	
84084P	40	Female	Diplegia	77	
87096P	31	Male	Diplegia	6	
88097P	30	Male	Diplegia	47	
89098P	28	Female	Diplegia	13	
9009P	34	Male	Hemiplegia	88	*MAST1*
92101P	37	Male	Quadriplegia	7	
96105P	38	Female	Quadriplegia	4	*JHDM1D*
99108P	23	Male	Hemiplegia	4	
Developmental malformations (n = 20)					
10010P	40	Male	Diplegia	47	
106115P	41	Male	Diplegia	34	*TUBA1A*
133415P	39	Female	Diplegia	94	
142424P	38	Male	Diplegia	56	*L1CAM*
156438P	33	Female	Hemiplegia	100	*L1CAM*
169451P	40	Female	Diplegia	58	*TUBA1A*
17017P	39	Male	Quadriplegia	45	
174456P	27	Male	Quadriplegia	75	
180462P	40	Female	Diplegia	22	
192475P	36	Male	Hemiplegia	69	
193476P	39	Female	Quadriplegia	50	
23023P	39	Female	Quadriplegia	29	
39039P	38	Male	Diplegia	72	*CD99L2*
49049P	35	Male	Hemiplegia	8	
52052P	36	Female	Quadriplegia	87	
58058P	38	Male	Quadriplegia	34	
60060P	40	Female	Hemiplegia	11	
77077P	29	Female	Hemiplegia	75	
81081P	36	Female	Hemiplegia	71	
93102P	40	Male	Hemiplegia	38	

Cerebral infarction (n = 15)					
100109P	36	Male	Hemiplegia	1	
105114P	38	Female	Hemiplegia	72	
108117P	40	Male	Hemiplegia	67	*NAA35*
110119P	39	Female	Hemiplegia	12	
135417P	39	Male	Hemiplegia	97	
145427P	38	Male	Hemiplegia	19	
165447P	39	Male	Dyskinetic	46	*PAK3*
21021P	39	Male	Hemiplegia	7	
28028P	41	Female	Hemiplegia	19	
42042P	31	Female	Hemiplegia	0	
43043P	41	Female	Hemiplegia	38	*SCN8A*
45045P	40	Female	Hemiplegia	45	
46046P	38	Female	Diplegia	48	
69069P	38	Female	Hemiplegia	75	
71071P	40	Male	Hemiplegia	35	
MRI but no abnormality seen (n = 7)					
19019P	37	Male	Diplegia	86	*TENM1*
109118P	40	Male	Hemiplegia	60	
134416P	NA	Male	Diplegia	NA	
P74074	38	Female	Diplegia	93	
P76076	37	Female	Diplegia	16	
P90099	32	Male	Diplegia	63	
P91100	40	Female	Diplegia	4	
Inutero infection (e.g. CMV) (n = 2)					
75075P	35	Female	Quadriplegia	1	
82082P	39	Female	Quadriplegia	1	
Inutero cerebral tumor (n = 1)					
103112P	41	Female	Hemiplegia	71	

Table 9.11 Supplementary Table K. Novel candidate genes in CP cases identified in 98 case-parent trios, clinical features and mouse phenotypes (where known)

Study design	Gene	Type	Protein change	Sex	Birth weeks	Clinical features	Mouse phenotype
Trio	*AGAP1*	DN	Splice	M	38	Quadriplegia, ID, Ep; MRI: PVL	MGI:2653690; Conditional ready, Null/knockout reporter mice have hearing/vestibular/ear, and vision/eye defects
Trio	*CD99L2*	X	Stopgain	M	38	Diplegia; MRI: delayed myelination	No reports
Trio	*JHDM1D*	DN	S727W	F	38	Quadriplegia, ID, Ep; MRI: PVL	MGI:2443388; Conditional ready, Null/knockout reporter mice have a abnormal limbs/digits/tails, mortality/aging and skeleton defects
Trio	*MAST1*	DN	P500L	M	32	Diplegia; MRI: No reports	No reports
Trio	*NAA35*	DN	W532C	M	40	Hemiplegia, ID, Ep; MRI: MCA territory infarction	No reports
Trio	*RFX2*	DN	Y91C	F	39	Hemiplegia; MRI: no reports	No reports
Trio	*TENM1*	X	G2533S	M	37	Diplegia; MRI: Asymmetry of lateral ventricles	No reports
Trio	*WIPI2*	DN	Y246C	F	30	Hemiplegia, hydrocephalus; MRI: PVL	No reports

Abbreviations: DN – de novo, X – X-chromosome, M – male, F – female, ID – intellectual disability, Ep – epilepsy, MRI – magnetic resonance imaging, MCA- middle cerebral artery, PVL – periventricular leukomalacia

Table 9.12 Supplementary Table L. Novel variants in known (OMIM) genes in CP cases identified in 98 case-parent trios, clinical features and known disease association

Study design	Gene	Type	Protein change	Sex	Birth weeks	Clinical features	Disease association (OMIM)
Trio	*KDM5C*	DN	P480L	F	39	Diplegia, ID, speech dyspraxia; MRI: PVL	ID (300534)
Trio	*L1CAM*	X	P161A	M	33	Hemiplegia, ID; MRI: agenesis of the corpus callosum	Corpus callosum (304100), CRASH syndrome (303350), Hydrocephalus (307000)
Duo	*L1CAM*	X	E304K	M	38	Diplegia, ID; MRI; h Hydrocephalus	Corpus callosum (304100), CRASH syndrome (303350), Hydrocephalus (307000)
Trio	*PAK3*	X	R493C	M	29	Hemiplegia, ID, Ep; MRI: PVL	ID (300558)
Duo	*PAK3*	X	M100I	M	39	Dyskinetic; MRI: Left cerebral artery infarction	ID (300558)
Trio	*SCN8A*	DN	G1050S	F	41	Hemiplegia, ID; MRI: Left MCA territory infarction	Cognitive impairment (614306), epileptic encephalopathy (614558)
Trio	*TUBA1A*	DN	R123C	M	41	Diplegia, ID, Ep; MRI; cerebellum and corpus callosum anomalies	Lissencephaly 3 (602529)
Trio	*TUBA1A*	DN	L152Q	F	40	Diplegia; MRI: cerebellum and corpus callosum anomalies	As above

Abbreviations: DN – de novo, X – X-chromosome, M – male, F – female, ID – intellectual disability, Ep – epilepsy, MRI – magnetic resonance imaging, MCA – middle cerebral artery, PVL – periventricular leukomalacia

9.1 References

1. de Ligt, J., Willemsen, M.H., van Bon, B.W., Kleefstra, T., Yntema, H.G., Kroes, T. et al. Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med* **367,** 1921-9 (2012).
2. Rauch, A., Wieczorek, D., Graf, E., Wieland, T., Endele, S., Schwarzmayr, T. et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* **380,** 1674-82 (2012).
3. Xu, B., Ionita-Laza, I., Roos, J.L., Boone, B., Woodrick, S., Sun, Y. et al. De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nat Genet* **44**, 1365-9 (2012).
4. Neale, B.M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K.E., Sabo, A. et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* **485**, 242-5 (2012).
5. Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J. et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485,** 237-41 (2012).
6. O'Roak, B.J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B.P. et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* **485**, 246-50 (2012).