

## IMMEDIATE COMMUNICATION

## Whole-exome sequencing points to considerable genetic heterogeneity of cerebral palsy

G McMichael<sup>1</sup>, MN Bainbridge<sup>2</sup>, E Haan<sup>3,4</sup>, M Corbett<sup>1,4</sup>, A Gardner<sup>1,4</sup>, S Thompson<sup>4,5</sup>, BWM van Bon<sup>3,6</sup>, CL van Eyk<sup>1</sup>, J Broadbent<sup>1</sup>, C Reynolds<sup>1</sup>, ME O'Callaghan<sup>1</sup>, LS Nguyen<sup>4</sup>, DL Adelson<sup>7</sup>, R Russo<sup>8</sup>, S Jhangiani<sup>2</sup>, H Doddapaneni<sup>2</sup>, DM Muzny<sup>2</sup>, RA Gibbs<sup>2</sup>, J Gecz<sup>1,4,9,10</sup> and AH MacLennan<sup>1,9,10</sup>

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 including both parents (98 cases) or 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 whether 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. In addition, we identified four predicted pathogenic, hemizygous variants on chromosome X in two known disease genes, *L1CAM* and *PAK3*, and in 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.

*Molecular Psychiatry* (2015) **20**, 176–182; doi:10.1038/mp.2014.189; published online 10 February 2015

## 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 50 years.<sup>1</sup> It is a clinically heterogeneous group of non-progressive disorders, primarily affecting the movement and posture. Co-morbidities can include intellectual disability (ID), autism spectrum disorder, epilepsy, speech and language deficits, and visual and hearing impairments.<sup>2</sup> Known epidemiological risk factors include preterm delivery, intrauterine growth restriction, intrauterine infection and male gender (male:female ratio 1.3:1).<sup>3–5</sup> However, causative pathways are poorly understood. Acute or chronic intrapartum fetal compromise, historically considered the cause of CP, is found in < 10% of cases.<sup>6</sup> 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.<sup>7–9</sup> 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.<sup>10</sup> A recent genome-wide search for copy number variants in 50 CP cases did not yield any *de novo* or obvious candidate mutations.<sup>11</sup>

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, autism spectrum disorder and schizophrenia.<sup>12–15</sup> To investigate the possible contribution of *de novo* and other rare variants to CP, we sequenced the exomes of 183 sporadic cases and where available, their unaffected parents.

## MATERIALS AND METHODS

## 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 pediatric rehabilitation specialist using standard published criteria relating to non-progressive disorders of movement control and posture.<sup>16</sup> Brain imaging reports were available for 112 of the cases (61%). The cohort's overall phenotypic, clinical and demographic characteristics (Table 1) were very similar to population distributions described in the 2013 report of the Australian Cerebral Palsy Register (see Web Resources). Pediatric specialist evaluation

<sup>1</sup>Robinson Research Institute, The University of Adelaide, Adelaide, SA, Australia; <sup>2</sup>Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA; <sup>3</sup>South Australian Clinical Genetics Service, SA Pathology (at Women's and Children's Hospital), North Adelaide, SA, Australia; <sup>4</sup>School of Pediatrics and Reproductive Health, The University of Adelaide, Adelaide, SA, Australia; <sup>5</sup>Department of Pediatric Neurology, Women's and Children's Hospital, North Adelaide, SA, Australia; <sup>6</sup>Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands; <sup>7</sup>School of Molecular and Biomedical Science, The University of Adelaide, Adelaide, SA, Australia and <sup>8</sup>Department of Pediatric Rehabilitation, Women's and Children's Hospital, North Adelaide, SA, Australia. Correspondence: Professor J Gecz, School of Pediatrics and Reproductive Health, University of Adelaide, Robinson Institute, Adelaide, SA 5006, Australia.  
E-mail: jozef.gecz@adelaide.edu.au

<sup>9</sup>These two authors contributed equally towards project supervision.

<sup>10</sup>These two authors contributed equally to this work.

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.

#### DNA extraction

DNA extraction from lymphoblastoid cell lines was performed on a fully automated large volume nucleic acid purification system (Qiagen Autopure LS; Qiagen, Stanford, CA, USA), which ensures high quality DNA from Epstein–Barr virus-transformed lymphoblastoid cell lines. 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, USA), respectively, following the manufacturer's instructions.

#### DNA 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](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 ~300–400 base pairs in a Covaris plate with E210 system followed by end repair, A-tailing and ligation of the Illumina multiplexing PE adapters. Pre-capture ligation-mediated PCR was performed for seven cycles of amplification using the 2× SOLiD Library High Fidelity Amplification Mix (a custom product manufactured by Invitrogen, Grand Island, NY, USA). 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 (Beckman Coulter, Inc., Brea, CA, USA) after enzymatic reactions. Following the final XP beads purification, quantification and size distribution of the pre-capture ligation-mediated PCR product was determined using the LabChip GX electrophoresis system (PerkinElmer).

#### Whole-exome capture

Six pre-capture libraries were pooled together (~166 ng per sample and 1 µg per pool) and hybridized in solution to the HGSC VCRome 2.1 design<sup>17</sup> (42 Mb, 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 adapter-specific blocking oligonucleotides were added into the hybridization to block repetitive genomic sequences and the adapter sequences. Post-capture ligation-mediated PCR amplification was performed using the 2× 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 analyzed using the Agilent Bioanalyzer 2100 DNA Chip 7500 (Agilent Technologies, Santa Clara, CA, USA). The efficiency of the capture was evaluated by performing a quantitative PCR-based quality check on the four standard NimbleGen (Roche NimbleGen, Madison, WI, USA) internal controls. Successful enrichment of the capture libraries was estimated to range from 6 to 9 of ΔCt value over the non-enriched samples.

#### DNA sequencing

Library templates were prepared for sequencing using Illumina's cBot cluster generation system with TruSeq PE Cluster Generation Kits (Illumina, San Diego, CA, USA; Cat. no. PE-401-3001). Briefly, these libraries were denatured with sodium hydroxide and diluted to 3–6 pM in hybridization buffer to achieve a load density of ~800 K clusters per mm<sup>2</sup>. Each library pool was loaded in a single lane of a HiSeq (Illumina) 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

**Table 1.** Clinical characteristics of 183 individuals with CP

Characteristic	Number (%)	Mean
<i>Cerebral palsy type</i>		
Quadriplegia	31 (17)	
Diplegia	65 (36)	
Hemiplegia	76 (42)	
Triplegia	4 (2)	
Dyskinesia	7 (3)	
<i>Other demographics</i>		
Gestational age (weeks)		34.7
Birth weight (grams)		2414
Maternal age (years)		30
Paternal age (years)		33
Male	99 (54)	
Plurality ( <i>n</i> > 1)	25 (14)	
<i>Co-morbidities</i>		
Intellectual disability	35 (19)	
Epilepsy	46 (25)	
Autism	12 (7)	

runs generated ~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 20× or greater.

#### Analysis

Analysis was conducted as previously described.<sup>18</sup> Briefly, Illumina sequence data were aligned to the human reference genome (HG19) with BWA.<sup>19</sup> Variant qualities were recalibrated with GATK.<sup>20</sup> 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 were 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 (Applied Biosystems, Foster City, CA, USA) and sequenced using an ABI 3730xl DNA analyzer (Applied Biosystems). Sequencing data were analyzed 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; [www.primer3plus.com/web\\_0.4.0/input.htm](http://www.primer3plus.com/web_0.4.0/input.htm)). Sanger sequencing was performed using BigDye terminator chemistry 3.1 (Applied Biosystems) and sequenced using an ABI prism 3700 genetic analyzer (Applied Biosystems). Sequencing data were analyzed using DNASTAR Lasergene 10 Seqman Pro (DNASTAR, Inc. Madison, WI, USA). All validation was performed using genomic DNA isolated from whole blood.

#### DNA variant and gene prioritization

Where possible, we followed recent guidelines for investigating causality of sequence variants in human disease before future experimental validation in animal and *in vitro* models.<sup>21</sup> We identified variants that were in known disease genes whose clinical spectrum overlaps CP or *de novo* in any gene. We checked these against single-nucleotide polymorphism database, 1000 genomes or Exome Variant Server. 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)<sup>22</sup> and Combined Annotation-Dependent Depletion (CADD)<sup>23</sup> with cut offs for RVIS < 50th percentile for known and novel genes and CADD > 10 for known Online Mendelian Inheritance in Man disease genes and CADD > 20 for novel candidate genes; (2) the type of variant (that is, frameshift, splice, stopgain and missense); (3) *in silico* prediction of functional effect at the amino-acid level by various algorithms;<sup>24,25</sup> (4) evolutionary conservation; (5) brain expression pattern; (6) the predicted effect of haploinsufficiency<sup>26</sup> and (7) previous disease association reported in Online Mendelian Inheritance in Man. The RVIS and CADD tools are complex, multidimensional tools.<sup>22,23</sup> RVIS ranks genes in terms of intolerance to functional genetic variation and CADD integrates several well-known tools, among these also PolyPhen and SIFT.<sup>22,23</sup> 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,<sup>12,13</sup> autism<sup>13,27,28</sup> and schizophrenia<sup>14</sup> WES papers to confirm whether the genes identified would be selected under this model. We performed a *t*-test through Partek (see Web Resources) to compare the haploinsufficiency scores<sup>26</sup> of *de novo* variants (where a score was available) with non-mutated genes.

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

## RESULTS

### Case-parent trios (*n* = 98)

**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)<sup>12,13</sup> and autism spectrum disorder (0.70 per individual).<sup>14</sup> Assessment of the haploinsufficiency scores<sup>26</sup> for *de novo* mutations (where scores were available; *n* = 38) identified a slight enrichment compared with non-mutated genes but this was not significant (Supplementary Figure A). On the basis of RVIS and CADD and a set of multidimensional prioritization criteria (see methods), 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 2 and 3).

Of the 10 *de novo* mutations predicted to be causative for CP, four occurred in genes associated with neurological disorders (Table 2): p.P480L in *KDM5C*, a lysine-specific histone demethylase and known X-linked ID (*XLID*) gene,<sup>29</sup> p.G1050S in *SCN8A*, associated with cognitive impairment,<sup>30</sup> and p.R123C and p.L152Q in *TUBA1A* associated with neuronal migration disorders<sup>31</sup> (Table 2). Six *de novo* mutations occurred in genes not known to be associated with disease (Table 3): including a single mutation (c.957+1G>A) predicted to affect splicing in *AGAP1*, which directly regulates AP-3-dependent trafficking<sup>32</sup> and five missense mutations in *JHDM1D*, *MAST1*, *NAA35*, *WIPI2* and *RFX2*. In addition, we identified seven loss-of-function 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 loss-of-function variants in these genes in Exome Variant Server.

### 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

**Table 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-morbidities	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: CADD, Combined Annotation-Dependent Depletion; CP, cerebral palsy; Dip, diplegia; DN, *de novo*; F, female; GERP, Genomic Evolutionary Rate Profiling; HI, haploinsufficiency index; ID, intellectual disability; M, male; OMIM, Online Mendelian Inheritance in Man; PPH2, Polyphen2; RVIS, Residual Variation Intolerance Score; U, unknown; X, X-chromosome.

**Table 3.** Novel candidate genes in CP cases identified in 98 trios; CADD score > 20 and RVIS % < 50

Case number, CP type, co-morbidities	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	WIP12	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	NA435	DN	p.W532C	Low	1.000	5.40	U	43.77	23.1	Component of the N-terminal acetyltransferase 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: CADD, Combined Annotation-Dependent Depletion; CP, cerebral palsy; Dip, diplegia; DN, *de novo*; F, female; GERP, Genomic Evolutionary Rate Profiling; Hem, hemiplegia; HI, haploinsufficiency index; ID, intellectual disability; M, male; PPH2, Polyphen2; RVIS, Residual Variation Intolerance Score; Quad, quadriplegia; U, unknown; X, X-chromosome.

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 2 and 3). Two were in known Online Mendelian Inheritance in Man disease genes: p.P161A in *L1CAM* associated with L1 syndrome,<sup>33</sup> 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 2 and 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+2 T > A (paternal allele) and c.7158+5 G > 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.<sup>34</sup> 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.<sup>34</sup> This individual showed joint limitation of his knee, ankle and foot and equinus deformity.

#### 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). 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.

#### Pathway analyses

Biological system analysis of the predicted causative CP genes ( $n=14$ ) was conducted using the Ingenuity Pathway Analysis ([www.ingenuity.com](http://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).

#### 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 hemorrhage, 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 tumor (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*, *SCN8A* 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)<sup>12</sup> and six out of 45 (0.13)<sup>13</sup> novel truncating mutations and four out of 22 (0.18)<sup>12</sup> and 11 out of 45 (0.24)<sup>13</sup> variants in known ID genes were found.

## DISCUSSION

CP encompasses a large group of non-progressive childhood movement and posture disorders and can occur as an isolated finding or together with additional phenotypic features.<sup>5</sup> Previous estimates have suggested that the contribution of genetic variants to the burden of CP is about 2%.<sup>35</sup> We followed, where possible, recent guidelines to resolve which of the many variants may be implicated in human disease.<sup>19</sup> 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 a considerable genetic heterogeneity underlying CP.<sup>7,8,36</sup> We used multiple criteria to select CP-relevant unique variants, including a combination of two recently developed prioritization algorithms, *RVIS*<sup>22</sup> and *CADD*.<sup>23</sup> We tested these algorithms in other genetically heterogeneous neurological disorders<sup>12–15,28</sup> 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.<sup>29,30,33,37,38</sup> 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.<sup>29</sup> Recently, the same variant (p.P480L) was reported as a novel mutation in two male siblings diagnosed with ID.<sup>39</sup> Although the majority of *KDM5C* affected individuals are males, carrier females with milder phenotypes including ID and spasticity have been reported.<sup>29,40</sup> 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.<sup>31</sup> Furthermore, spastic diplegia or quadriplegia, ID and microcephaly are common in *TUBA1A* mutations.<sup>37,41,42</sup> Case 106115P was diagnosed with diplegic CP, optic atrophy, adducted thumbs, seizures and moderate ID. Case 169451P had diplegic CP. Brain magnetic resonance imaging of both the individuals showed, except for the characteristic sign of pachygyria, *TUBA1A* associated anomalies, such as cerebellar and corpus callosum anomalies.<sup>41,42</sup>

One female case (43043P) diagnosed with hemiplegic CP and ID had a missense mutation, p.G1050S, in *SCN8A*. ID, epilepsy and

varying degrees of motor dysfunction including ataxia and spasticity have been described in children with mutations in *SCN8A*.<sup>13,30,43</sup> 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.<sup>43</sup>

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.<sup>33</sup> The phenotype of both the 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 computed tomography.

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.<sup>44</sup> One of the variants affects the CRIB domain (p.M100I in 165447P) and the other the kinase domain (p.R493C in 15015P). Both mutations have not 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 magnetic resonance imaging showed a grade IV intraventricular hemorrhage. 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. Although isolated loss of *PAK3* kinase activity causes non-syndromic ID, a dual molecular effect is seen in those cases with a syndromic neurocutaneous phenotype.<sup>38</sup> In the latter phenotype also hypotonia, apparent motor delay with inability to walk, hyperreflexia and afinalistic movements have been described.<sup>38</sup> Out of 2400 males in the NHLBI Exome Sequencing Project (EPS), 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.<sup>45</sup> Two missense variants in known *XLID* genes, *IQSEC2* and *ZDHH9*, were identified in two cases. Although they met our prioritization criteria, we did not consider them causative of CP mostly because no ID 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.<sup>46</sup> 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.<sup>4,5</sup> All individuals with CP in this study met these criteria for CP at time of diagnosis. Physical examination before inclusion was performed by experienced pediatric 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.<sup>47</sup> 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.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGMENTS

We wish to thank participating CP families, Cerebral Palsy Alliance; Dr James Lupski for critical comments; Dr Jane Valentine and Peta Watts for facilitating blood collections at Princess Margaret Hospital, Western Australia; Dr James Rice and Dr Andrew Tidemann for facilitating blood collections in South Australia; Associate Professor Christopher Barnett for clinical reports; staff of the South Australian Cerebral Palsy Register; Kelly Harper for correlation of neuroimaging reports; Josh Woenig for the technical support; and Dr Kathie Friend and staff at the Department of Genetic Medicine, Women's and Children's Hospital, Adelaide and AGRF (Adelaide node) for the support with DNA extractions. DNA/cell lines were established by Genetic Repositories Australia, an Enabling Facility supported by the Australian National Health and Medical Research Council (Grant No. 401184). This work was funded by the Australian National Health and Medical Research Council (Grant No. 1041920 and Grant No. 1019928). JG is supported by NHMRC research fellowship 1041920, The Cerebral Palsy Foundation, The Tenix Foundation, The Robinson Institute, The University of Adelaide, Women's & Children's Research Foundation, MC is supported by MS McLeod research fellowship, and the Ter Meulen Fund (stipend to BWM/vanB). Supported in part by the National Human Genome Research Institute U54 HG003273 (RAG).

### Web Resources

The URLs for data presented herein include:

Cerebral Palsy Research Report; [www.cerebralpalsy.org.au/wp-content/uploads/2013/04/ACPR/](http://www.cerebralpalsy.org.au/wp-content/uploads/2013/04/ACPR/)

BCM-HGSC protocol; ([https://hgsc.bcm.edu/sites/default/files/documents/Illumina\\_Barcode\\_Paired-End\\_Capture\\_Library\\_Preparation.pdf](https://hgsc.bcm.edu/sites/default/files/documents/Illumina_Barcode_Paired-End_Capture_Library_Preparation.pdf)).

Partek; <http://www.partek.com/>

1000 Genomes; <http://browser.1000genomes.org/index.html>

dbSNP; <http://www.ncbi.nlm.nih.gov/projects/SNP/>

Online Mendelian Inheritance in Man (OMIM); <http://www.omim.org/>

## REFERENCES

- Stanley F, Alberman B. How common are the cerebral palsies? In: Stanley F, Blair E, Alberman B (eds). *Cerebral Palsies: Epidemiology and Causal Pathways*. Mac Keith Press: London, UK, 2000, pp 22–48.
- Rosenbaum P, Paneth N, Leviton A, Goldstein M, Bax M, Damiano D et al. A report: the definition and classification of cerebral palsy April 2006. *Dev Med Child Neurol Suppl* 2007; **109**: 8–14.
- O'Callaghan ME, MacLennan AH, Gibson C, McMichael G, Haan E, Broadbent J et al. Epidemiologic associations with cerebral palsy. *Obstet Gynaecol* 2011; **118**: 576–582.
- Colver A, Fairhurst C, Pharoah PO. Cerebral palsy. *Lancet* 2014; **383**: 1240–1249.
- Moreno-De-Luca A, Ledbetter DH, Martin CL. Genomic insights into the causes and classification of the cerebral palsies. *Lancet Neurol* 2012; **11**: 283–292.
- Nelson KB, Ellenberg JH. Antecedents of cerebral palsy. Multivariate analysis of risk. *N Engl J Med* 1986; **315**: 81–86.
- Hirata H, Nanda I, van Riesen A, McMichael G, Hu H, Hambrock M 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* 2013; **92**: 681–695.
- Moreno-De-Luca A, Helmers SL, Mao H, Burns TG, Melton AM, Schmidt KR 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* 2011; **48**: 141–144.
- Tollanes MC, Wilcox AJ, Lie RT, Moster D. Familial risk of cerebral palsy: population based cohort study. *BMJ* 2014; **349**: g4294.
- O'Callaghan ME, MacLennan AH, Gibson CS, McMichael GL, Haan EA, Broadbent JL et al. Fetal and maternal candidate single nucleotide polymorphism associations with cerebral palsy: a case-control study. *Pediatrics* 2012; **129**: e414–e423.
- McMichael G, Girirajan S, Moreno-De-Luca A, Gecz J, Shard C, Nguyen LS et al. Rare copy number variation in cerebral palsy. *Eur J Hum Genet* 2013; **22**: 40–45.
- de Ligt J, Willemsen MH, van Bon BW, Kleefstra T, Yntema HG, Kroes T et al. Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med* 2012; **367**: 1921–1929.
- Rauch A, Wiczorek D, Graf E, Wieland T, Ende S, Schwarzmayr T et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* 2012; **380**: 1674–1682.
- Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 2012; **485**: 237–241.
- Xu B, Ionita-Laza I, Roos JL, Boone B, Woodrick S, Sun Y et al. De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nat Genet* 2012; **44**: 1365–1369.
- Badawi N, Watson L, Petterson B, Blair E, Slee J, Haan E et al. What constitutes cerebral palsy? *Dev Med Child Neurol* 1998; **40**: 520–527.
- Bainbridge MN, Wang M, Wu Y, Newsham I, Muzny DM, Jefferies JL et al. Targeted enrichment beyond the consensus coding DNA sequence exome reveals exons with higher variant densities. *Genome Biol* 2011; **12**: R68.
- Bainbridge MN, Hu H, Muzny DM, Musante L, Lupski JR, Graham BH et al. De novo truncating mutations in ASXL3 are associated with a novel clinical phenotype with similarities to Bohring-Opitz syndrome. *Genome Med* 2013; **5**: 11.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25**: 1754–1760.
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genet* 2011; **43**: 491–498.
- MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR et al. Guidelines for investigating causality of sequence variants in human disease. *Nature* 2014; **508**: 469–476.
- Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet* 2013; **9**: e1003709.
- Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genet* 2014; **46**: 310–315.
- Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* 2013; **Chapter 7**: Unit7 20.
- Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nature Methods* 2010; **7**: 575–576.
- Huang N, Lee I, Marcotte EM, Hurler ME. Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet* 2010; **6**: e1001154.
- Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 2012; **485**: 242–245.
- O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 2012; **485**: 246–250.
- Rujirabanjerd S, Nelson J, Tarpey PS, Hackett A, Edkins S, Raymond FL et al. Identification and characterization of two novel JARID1C mutations: suggestion of an emerging genotype-phenotype correlation. *Eur J Hum Genet* 2010; **18**: 330–335.
- Trudeau MM, Dalton JC, Day JW, Ranum LP, Meisler MH. Heterozygosity for a protein truncation mutation of sodium channel SCN8A in a patient with cerebellar atrophy, ataxia, and mental retardation. *J Med Genet* 2006; **43**: 527–530.
- Poirier K, Saillour Y, Fourniol F, Francis F, Souville I, Valence S et al. Expanding the spectrum of TUBA1A-related cortical dysgenesis to Polymicrogyria. *Eur J Hum Genet* 2013; **21**: 381–385.
- Nie Z, Boehm M, Boja ES, Vass WC, Bonifacino JS, Fales HM et al. Specific regulation of the adaptor protein complex AP-3 by the Arf GAP AGAP1. *Dev Cell* 2003; **5**: 513–521.

- 33 Yamasaki M, Thompson P, Lemmon V. CRASH syndrome: mutations in L1CAM correlate with severity of the disease. *Neuropediatrics* 1997; **28**: 175–178.
- 34 Stum M, Davoine CS, Vicart S, Guillot-Noel L, Topaloglu H, Carod-Artal FJ *et al*. Spectrum of HSPG2 (Perlecan) mutations in patients with Schwartz-Jampel syndrome. *Human Mutat* 2006; **27**: 1082–1091.
- 35 Rajab A, Yoo SY, Abdulgaliil A, Kathiri S, Ahmed R, Mochida GH *et al*. An autosomal recessive form of spastic cerebral palsy (CP) with microcephaly and mental retardation. *Am J Med Genet A* 2006; **140**: 1504–1510.
- 36 Krueger MC, Jepperson T, Dutta S, Steiner RD, Cottenie E, Sanford L *et al*. Mutations in gamma adducin are associated with inherited cerebral palsy. *Ann Neurol* 2013; **74**: 805–814.
- 37 Bahi-Buisson N, Poirier K, Fourniol F, Saillour Y, Valence S, Lebrun N *et al*. The wide spectrum of tubulinopathies: what are the key features for the diagnosis? *Brain* 2014; **137**: 1676–1700.
- 38 Magini P, Pippucci T, Tsai IC, Coppola S, Stellacci E, Bartoletti-Stella A *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; **23**: 3607–3617.
- 39 Grafodatskaya D, Chung BH, Butcher DT, Turinsky AL, Goodman SJ, Choufani S *et al*. Multilocus loss of DNA methylation in individuals with mutations in the histone H3 lysine 4 demethylase KDM5C. *BMC Med Genomics* 2013; **6**: 1.
- 40 Jensen LR, Amende M, Gurok U, Moser B, Gimmel V, Tzschach A *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* 2005; **76**: 227–236.
- 41 Bahi-Buisson N, Poirier K, Boddaert N, Saillour Y, Castelnau L, Philip N *et al*. Refinement of cortical dysgeneses spectrum associated with TUBA1A mutations. *J Med Genet* 2008; **45**: 647–653.
- 42 Poirier K, Keays DA, Francis F, Saillour Y, Bahi N, Manouvrier S *et al*. Large spectrum of lissencephaly and pachygyria phenotypes resulting from de novo missense mutations in tubulin alpha 1A (TUBA1A). *Human Mutat* 2007; **28**: 1055–1064.
- 43 Veeramah KR, O'Brien JE, Meisler MH, Cheng X, Dib-Hajj SD, Waxman SG *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* 2012; **90**: 502–510.
- 44 Rejeb I, Saillour Y, Castelnau L, Julien C, Bienvenu T, Taga P *et al*. A novel splice mutation in PAK3 gene underlying mental retardation with neuropsychiatric features. *Eur J Hum Genet* 2008; **16**: 1358–1363.
- 45 Klitten LL, Moller RS, Ravn K, Hjalgrim H, Tommerup N. Duplication of MAOA, MAOB, and NDP in a patient with mental retardation and epilepsy. *Eur J Hum Genet* 2011; **19**: 1–2.
- 46 Mochida GH, Walsh CA. Genetic basis of developmental malformations of the cerebral cortex. *Arch Neurol* 2004; **61**: 637–640.
- 47 O'Roak BJ, Vives L, Fu W, Egerton JD, Stanaway IB, Phelps IG *et al*. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* 2012; **338**: 1619–1622.

Supplementary Information accompanies the paper on the Molecular Psychiatry website (<http://www.nature.com/mp>)