Peer Review Information:

Nature Genetics thanks Jozef Gecz and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Editorial Summary:

Whole-genome sequencing in a Canadian cohort of 327 children with cerebral palsy compared to pediatric controls, identifies novel pathogenic SNVs/indels and CNVs. In addition, mitochondrial variants in known disease genes were identified. This highlights the importance of genomic testing for individuals with cerebral palsy.

1. Supplementary Information:

A. PDF Files

Item	Present?	Filename Whole original file name including extension. i.e.: Smith_SI.pdf. The extension must be .pdf	A brief, numerical description of file contents. i.e.: *Supplementary Figures 1-4, Supplementary Discussion, and Supplementary Tables 1-4.*
Supplementary Information	Yes	Supplementary Note_FINAL_20240124.pdf	Supplementary Table Information, Supplementary Note 1, Supplementary Figures 1-3
Reporting Summary	Yes	NG-A61616R2_Scherer_RSf.pdf	
Peer Review Information	Yes	*NG-A61616R2_Scherer_TPR.pdf*	

B. Additional Supplementary Files

Type	Number Each type of file (Table, Video, etc.) should be numbered from 1 onwards. Multiple files of the same type should be listed in sequence, i.e.: Supplementary Video 1,	Filename Whole original file name including extension. i.e.: *Smith_ Supplementary_Video_1.mov*	Legend or Descriptive Caption Describe the contents of the file

	Supplementary Video 2, etc.		
Supplementary Table	1	Supplementary Tables_FINAL.xlsx	Supplementary Tables 1-10 in a single excel file with multiple worksheets
Supplementary Data	1	SupplementaryFig2_Source_Data_FINAL.csv	Source data used to generate Supplementary Table 2

3. Source Data

Parent Figure or Table	Filename Whole original file name including extension. i.e.: *Smith_SourceData_Fig1.xls,* or *Smith_ Unmodified_Gels_Fig1.pdf*	Data description i.e.: Unprocessed western Blots and/or gels, Statistical Source Data, etc.
Source Data Fig. 1	Figure1_Source Data_FINAL.xlsx	Data in excel that was used to generate Figure 1
Source Data Fig. 2	Figure2_Source_Data_FINAL.csv	Data in excel that was used to generate Figure 2

Comprehensive Whole-Genome Sequence Analyses Provide Insights into the Genomic Architecture of Cerebral Palsy

Darcy L. Fehlings[1,2*], Mehdi Zarrei[3,4*], Worrawat Engchuan[3], Neal Sondheimer[2,4,5], Bhooma Thiruvahindrapuram[3], Jeffrey R. MacDonald[3,4], Edward J. Higginbotham[3,6], Ritesh Thapa[1], Tarannum Behlim[7], Sabrina Aimola[1], Lauren Switzer[1], Pamela Ng[7], John Wei[3], Prakroothi S. Danthi[3], Giovanna Pellecchia[3], Sylvia Lamoureux[3], Karen Ho[3], Sergio L. Pereira[3], Jill de Rijke[3], Wilson W.L. Sung[3,4], Alireza Mowjoodi[3,4], Jennifer L. Howe[3], Thomas Nalpathamkalam[3], Roozbeh Manshaei[3,8], Siavash Ghaffari[3,4], Joseph Whitney[3], Rohan V. Patel[3], Omar Hamdan[3], Rulan Shaath[3], Brett Trost[3,4], Shannon Knights[2,9], Dawa Samdup[10], Anna McCormick[11], Carolyn Hunt[9], Adam Kirton[12], Anne Kawamura[1,2], Ronit Mesterman[13], Jan Willem Gorter[13], Nomazulu Dlamini[2,14,15], Daniele Merico[3,16], Murto Hilali[3], Kyle Hirschfeld[3], Kritika Grover[3], Nelson X. Bautista[3], Kara Han[3], Christian R. Marshall[6], Ryan K. C. Yuen[3,4,5], Padmaja Subbarao[2,13,17], Meghan B. Azad[18], Stuart E. Turvey[19], Piush Mandhane[20], Theo J. Moraes[2,21], Elinor Simons[22], George Maxwell[23], Michael Shevell[7,24], Gregory Costain[2,4,6,25], Jacques L. Michaud[26,27], Fadi F. Hamdan[27,28], Julie Gauthier[27,28], Kevin Uguen[27], Dimitri J. Stavropoulos[6,29], Richard F. Wintle[3], Maryam Oskoui[7,24**], Stephen W. Scherer[3,4,30**]

[1]Division of Developmental Paediatrics, Holland Bloorview Kids Rehabilitation Hospital, Toronto, Canada
[2]Department of Paediatrics, Faculty of Medicine, University of Toronto, Toronto, Canada
[3]The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada
[4]Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Canada
[5]Department of Molecular Genetics, University of Toronto, Toronto, Canada
[6]Genome Diagnostics, Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Canada
[7]Centre for Outcomes Research and Evaluation, Research Institute of the McGill University Health Centre, Montreal, Canada
[8]Ted Rogers Centre for Heart Research, Cardiac Genome Clinic, The Hospital for Sick Children, Toronto, Canada
[9]Grandview Children's Centre, Oshawa, Canada
[10]Department of Pediatrics, Queen's University, Kingston, Canada
[11]Children's Hospital of Eastern Ontario and University of Ottawa, Ottawa, Canada
[12]Department of Pediatrics, Department of Clinical Neuroscience, University of Calgary, Calgary, Alberta, Canada
[13]Department of Pediatrics, McMaster University, Hamilton, Canada
[14]Division of Neurology, The Hospital for Sick Children, Toronto, Canada
[15]Neurosciences and Mental Health Program, Research Institute, The Hospital for Sick Children, Toronto, Canada
[16]Deep Genomics Inc., Toronto, Canada; presently at Vevo Therapeutics Inc., San Francisco, USA
[17]Department of Physiology, University of Toronto, Toronto, Canada
[18]Department of Pediatrics and Child Health, University of Manitoba, Winnipeg, Manitoba, Canada
[19]Department of Pediatrics, BC Children's Hospital, University of British Columbia, Vancouver, BC, Canada
[20]Faculty of Medicine & Dentistry – Pediatrics Dept., University of Alberta, Edmonton, Alberta, Canada
[21]Program in Translation Medicine & Division of Respiratory Medicine, The Hospital for Sick Children, Toronto, Canada
[22]Department of Pediatrics and Child Health, Section of Allergy and Clinical Immunology, University of Manitoba; Children's Hospital Research Institute of Manitoba
[23]Women's Health Integrated Research Center, Inova Women's Service Line, Inova Health System, Falls Church, Virginia
[24]Departments of Pediatrics and Department of Neurology and Neurosurgery, McGill University, Montreal, Canada
[25]Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Canada
[26]Departments of Pediatrics and Neurosciences, Université de Montréal, Montreal, Canada
[27]CHU Sainte-Justine Azrieli Research Center, Montreal, Canada
[28]Department of Pediatrics, Université de Montréal, Montreal, Canada
[29]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Canada
[30]Department of Molecular Genetics and McLaughlin Centre, University of Toronto, Toronto, Canada
*These authors contributed equally
**These authors jointly supervised this work

Correspondence to: Dr. Stephen W. Scherer, The Centre for Applied Genomics, The Hospital for Sick Children, 555 University Ave., Toronto, ON, M5G 1X8
Stephen.Scherer@sickkids.ca

Word Count: 4,125

Key Words: cerebral palsy, whole genome sequencing, mitochondrial variants

Abstract

We performed whole-genome sequencing (WGS) in 327 children with cerebral palsy (CP) and their biological parents. We classified 37/327 (11.3%) children as having pathogenic/likely pathogenic (P/LP) variants and 58/327 (17.7%) as having variants of uncertain significance. Multiple classes of P/LP variants included single nucleotide variants (SNV)/indels (6.7%), copy number variations (CNV) (3.4%), and mitochondrial mutations (1.5%). The *COL4A1* gene had the most P/LP SNVs. We also analyzed two pediatric control cohorts (n=203 trios, and n=89 sib-pair families) to provide a baseline for *de novo* mutation rates and genetic burden analyses, the latter of which demonstrated associations between *de novo* deleterious variants and genes related to the nervous system. An enrichment analysis revealed previously undescribed plausible candidate CP genes (*SMOC1, KDM5B, BCL11A, CYP51A1*). A multifactorial CP risk profile and significant presence of P/LP variants combine to support WGS in the diagnostic work-up across all CP and related phenotypes.

Introduction

Cerebral palsy (CP) is the most common childhood-onset physical disability with birth prevalence of 1.5/1000 live births in high and 3.4/1000 in low/middle income countries[1]. An international consensus definition describes CP as "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 brain" [2]. There are often additional developmental impacts including disturbances of cognition and behaviour (e.g., autism spectrum disorder; ASD) [3]. An overlap of CP candidate genes with other neurodevelopmental disorders such as ASD, intellectual disability (ID) and epilepsy has been identified[4]. The CP diagnosis may be associated with significant personal, social and financial burdens[5].

Contributions from genetic variation to CP are increasingly recognized[6]. Phenotypic characteristics associated with higher genetic yield include idiopathic/cryptogenic CP (no identifiable antenatal, neonatal and postnatal etiologic risk factors for CP) [7-11], presence of ID, ASD[12,13], family history of CP[9], term birth[12,14] and normal brain imaging[15]. The most frequently reported monogenic cause of CP is *CTNNB1*[16,17]. The uncovering of rare variants (<1% population frequency) impacting single genes or genomic loci (copy number variations, or CNVs) can result in diagnostic confusion[18] and debate about the validity of a clinical diagnosis of CP in individuals with a "genetic diagnosis" [19]. Resolution of these discussions will be facilitated by the comprehensive understanding of the full range of genetic variation contributing to CP phenotypes in large, prospectively collected cohorts, with concomitant analyses of parental samples to identify inherited and *de novo* variation.

A single genome-wide association study (GWAS) examining common genetic variation in CP identified only one variant, postulated to affect *GRIK4* expression[20]. Although GWAS studies to date are underpowered, this raises the possibility that major genomic contributions to CP susceptibility may alternatively be uncovered by methods detecting rare, highly penetrant variants. Accordingly, most genomic studies (Supplementary Table 1 sheets 1-2) have focused on candidate genes, or detecting rare variants. Genomic CNVs[11,14,21-23], single nucleotide variants (SNVs) [4,7,9,10,13,15,24] and rarely, large cytogenetic changes[25,26] have been observed in CP. These studies have predominantly used exome sequencing and/or chromosomal microarray analysis with limitations in detection of some types of genetic variation. There have been two studies applying whole genome sequencing (WGS) to CP. The first[27] examined 150 Australian patients, and the second[8] studied 120 trio families from China (Supplementary Table 1 sheet 1). Across all studies, there was variability in ascertainment criteria (e.g., focusing only on idiopathic CP cases), the technology employed (e.g., different microarray platforms for CNV analysis), and regional differences in phenotyping and interpretation of pathogenicity[28].

To better capture the full extent of genomic contributions to CP, including all classes of nuclear variation and mitochondrial variants, we present a comprehensive analysis of WGS data in a cohort of over 320 previously undescribed trio families prospectively identified from Canada. We annotated the full spectrum of genomic variation[29], including spontaneous and rare-inherited SNV (1-50 bp), CNV (≥1 kb), structural variation (SV; 50 bp-1 kb), tandem repeat expansions[30], and mitochondrial variants[31]. In parallel, we analyzed two healthy pediatric

control cohorts (n=203 trios and n=89 sib-pair families) to provide a baseline for *de novo* mutation rates and genetic burden analyses. We also compared our CP-relevant findings to three independent clinical genomics cohorts (2,876 samples) and studied the resulting variants for potential functional impact on protein-protein interaction using AlphaFold-Multimer.

Results

Genomic variants of potential clinical relevance

WGS was completed on saliva DNA from 327 children with CP from 323 families including 308 trios (index child and biological parents), four quads (two children with CP and biological parents), ten probands with one parental sample, and one proband where neither parental sample (Supplementary Table 2 sheet 1). A small number of samples (0.7% of the initial cohort) failed sequencing library prep, and 1.2% required top-up sequencing to reach 30x average coverage. Of those sequenced, 4% were excluded post-sequencing due to overall poor quality metrics. The mean age of children with CP was 5.7 years (standard deviation; SD=3.1). The mean maternal and paternal ages at the time of conception were 32.2 (SD=4.9) and 34.1 years (SD=5.9) respectively. One child lost their diagnosis of CP (diagnosed with a dopa-responsive dystonia) by the five-year chart review. Predicted ancestry using WGS data identified 57.8% being of European origin, with other ancestries including African (4.3%), Admixed American (4.9%), East Asian (4.0%), Middle Eastern (5.8%), South Asian (5.8%) and Other/Unidentifiable (17.4%) (Supplementary Table 2 sheet 1).

We classified 37/327 (11.3%) children as having pathogenic/likely pathogenic (P/LP) variants and 58/327 (17.7%) as having variants of uncertain significance (VUS; Tables 1-3, Figure 1, Supplementary Table 3 sheets 1-4). P/LP damaging missense and loss-of-function (LOF) SNVs and indels accounted for 6.7% (22/327), CNVs for 3.4% (11/327), and mitochondrial variants for 1.5% (5/327). One child (case 278746) had both a pathogenic SNV and CNV. No P/LP tandem repeat expansions were identified from a list of 59 known disease-causing loci (Supplementary Table 10). Sixteen of 327 (4.9%) had multiple P/LP/VUS variants identified: one child with four, two children with three and 13 with two variants (Supplementary Table 3 sheet 5).

Of the P/LP variants (37/327) in CP cases, three had aneuploidies: trisomy 21, XYY syndrome, and a probable iso-dicentric chromosome X syndrome. Single-gene, LP deletions were found in *DLG2*, *ARID2*, *NRXN1*, and *EXT2*. P/LP SNVs or sequence-level frameshift variants were identified in *ADCY5*, *COL4A1*, *ERLIN2*, *EXOSC3*, *GATAD2B*, *GNAO1*, *KDM3B*, *MEF2C*, *PDE10A*, *PHKA2*, *PIK3R2*, *SOX2*, *SPAST*, *TH*, *TRIP12*, *TUBA1A*, *ZDHHC9*, and *ZIC2* (Table 2). The *COL4A1* gene had the most P/LP SNVs (four cases). Three P/LP homozygous SNVs were identified in recessive genes (*ERLIN2, EXOSC3, TH*) (Table 2). Of the five mitochondrial variants (Table 3), four were non-coding: one ribosomal gene (*MT-RNR1*) and three tRNA genes for leucine (*MT-TL1*), glutamine (*MT-TQ*), and serine (*MT-TS1*). The fifth was a missense variant in NADH dehydrogenase (*MT-ND5*). No pathogenic mitochondrial variants were identified in the control cohorts.

Clinical genomic cohorts

As an exercise in reverse phenotyping to further the evidence for motor involvement/presence of CP phenotype, we investigated three additional phenotypically heterogeneous cohorts for the presence of the 19 genes with P/LP SNV variants identified in our CP cohort (Table 2). For each individual with a suspected or confirmed molecular genetic diagnosis associated with these 19 genes, we assessed the presence of CP or a neuromotor disability. The first cohort is a database of all variants from The Hospital for Sick Children's Genome Diagnostics laboratory that have been reported clinically through genome-wide sequencing (n=927). The second is a group of 802 children with medical complexity (at least one chronic condition, technology dependence, multiple subspecialist involvement), at The Hospital for Sick Children and Complex Care Satellite Program in Ontario[32]. The third cohort represents 1,147 clinical cases from the CHU Sainte-Justine (Montreal, QC, Canada). In these three cohorts, 13 of 19 genes impacted by damaging SNVs in our CP cohort were also identified in at least one unrelated individual with CP/neuromotor disability (Table 2). These genes include *COL4A1* (6 additional variants), *EXOSC3* (5), *TUBA1A* (4), *GATA2DB* (2), *GNAO1* (2), *TH* (2), *ADCY5* (1), *KDM3B* (1), *MEF2C* (1), *SPAST* (1), *TRIP12* (1), *ZDHHC9* (1) and *ZIC2* (1). There were 42 individuals across the CP cohort and the three clinical genomic cohorts with CP/neuromotor disability and findings in one of the 19 genes. A majority had co-morbid global developmental delay/intellectual disability (76%, 32/42 individuals), and 45% had seizures. Brain malformations were commonly identified, as is expected for variants in genes like *EXOSC3* (cerebellar hypoplasia) and *TUBA1A* (neuronal migration disorder).

We also assessed potential overlap in these cohorts between variants driving the gene burden enrichment results that were identified as having either *de novo* LOF or damaging missense variants, without achieving P/LP/VUS status, and the presence of CP/neuromotor disability to explore the potential contribution of these genes to CP/neuromotor disability. Forty genes with damaging missense variants and 12 with LOF variants fell into these categories (Supplementary Table 4 sheets 6-7). There was no overlap in the first cohort, the second cohort had two cases with variants associated with a neuromotor disability (homozygous *SMOC1* associated with ataxia/motor disability; heterozygous *KDM5B* variants associated with dystonia/motor disability), and the third cohort had three variants (heterozygous *BCL11A* associated with ataxia/motor disability; heterozygous *CYP51A1* associated with spasticity/motor disability; compound heterozygous *KDM5B* associated with ataxia/motor disability).

***De novo* rates in CP and paediatric controls**

De novo variants in CP cases or paediatric controls (Canadian Healthy Infant and Longitudinal Development [CHILD] cohort and Inova Health System Cohort [Inova] cohorts) were assessed. A *de novo* rate of 5.8% (18/312) for CNVs in CP cases was found, compared with 2.9% (11/381) in pediatric controls (odds ratio; OR=2.06, 95% confidence intervals (CI) 0.9-4.99, p=0.08). The rate of *de novo* LOF was 0.13 variants per sample (42/312) in CP and 0.12 (46/381) in pediatric controls (OR=1.13, 95% CI 0.7-1.82, p=0.36). The genome-wide damaging missense *de novo* rate was 0.28 variants per sample (88/312 complete trios) for CP and 0.22 per sample (83/381) in controls (OR=1.41, 95% CI 0.98-2.02, p=0.05). Most notably, no P/LP *de novo* damaging missense or LOF variants were found in controls. In the CP cohort, paternal age significantly

predicted the total number of *de novo* SNVs/indels (beta=0.74, 95% CI 0.20-1.28, *p*=0.008), whereas maternal age did not (beta=0.35, 95% CI -0.11-0.81, *p*=0.14) (Supplementary Table 7). Parental age was not collected in the paediatric controls.

Burden of *de novo* variants

All *de novo* variants in CP cases were compared with singleton variants in the Genome Aggregation Consortium (gnomAD) population samples in CP/neurodevelopmental disorder-associated pathways. Multiple gene-sets were enriched for LOF and missense variants in CP cases. These include pathways associated with development and function of the nervous system, *FMR1* target genes[33], impairments of mental function, and highly brain- expressed genes. However, we found no gene-set to be over-represented for damaging missense variants. No gene-sets had significant enrichment of any variant type in the paediatric control cohorts (Supplementary Table 4). The test was also performed on *de novo* synonymous variants as a negative control in CP cases and the result was not significant (Supplementary Table 4 sheet 1). *De novo* LOF and missense variants in 20 and 109 genes, respectively, contributed to pathway enrichment (Supplementary Table 4). Seven of 20 genes harbouring *de novo* LOF in CP were identified as P/LP/VUS. In addition, 10 are reported as autosomal dominant in Online Mendelian Inheritance in Man (OMIM). Of the variants in the 109 genes contributing to the signal in the *de novo* missense analysis, 43 were classified as damaging and variants in three genes (*COL4A1, GNAO1*, and *PIK3R2*) were identified as P/LP.

Transmission disequilibrium test for tandem repeats

In gene-sets associated with CP and neurodevelopmental disorders (NDD), an over-transmission of large tandem repeats (TRs) detected by ExpansionHunter Denovo in genes associated with axon guidance and postsynaptic density was observed in the CP cohort (Figure 2, Supplementary Table 4). We found a total of 73 genes contributing to the signal from the large TRs (Supplementary Table 4 sheet 5). No gene sets were identified as over-transmitted in the pediatric control cohorts. There was no over-transmission of large TRs within different regions of the genome (e.g., genic, intergenic, intronic) from parents to children in both CP and paediatric control cohorts.

***In silico* functional analysis of missense variants**

To gain preliminary functional insight into potential molecular mechanisms of missense variants detected in the CP cohort, we used *in silico* approaches to investigate two potential mechanisms: disruption of protein-protein interactions[34] and disruption of phosphorylation-mediated signaling[35]. We investigated three selected *de novo* P/LP missense variants using AlphaFold-Multimer[36] to assess if their mechanisms may disrupt protein-protein interactions (p.A334T in *TUBA1A*, p.D132A in *EXOSC3*, and p.E237D in *GNAO1*). These variants were selected based on previous evidence that missense variants in their respective proteins may disrupt protein-protein interactions[37-39]. *TUBA1A* and *EXOSC3* were not predicted by AlphaFold-Multimer to be involved in the interface with their putative interacting protein(s) (Supplementary Table 8 sheet 1). However, the *GNAO1* variant, p.E237D, was predicted by AlphaFold-Multimer to be involved in the interface with its regulator RGS4. Interestingly, a *de novo* variant at the same position, but causing a different amino acid change (p.E237K), has

previously been implicated in CP-like features[40]. We modeled both missense variants in AlphaFold-Multimer to assess the GNAO1-RGS4 interaction. Despite the different amino acid characteristics of the two missense variants (Asp versus Lys), they were predicted to have similar effects on the GNAO1-RGS4 interface (Supplementary Table 8 sheet 2 and Supplementary Figure 3). Further, the p.E237D structure was predicted to be more similar to the p.E237K structure (distance = 0.28) than to the wild-type structure (distance = 0.45). To investigate phosphorylation-mediated signaling, we identified *de novo* missense variants in children with CP that either changed a phosphorylatable residue (serine, threonine, or tyrosine) to a non-phosphorylatable residue, or were predicted to disrupt a protein kinase recognition motif. The genes in which *de novo* missense variants impacted phosphorylation sites were *EPB41L4A*, *EFNB2*, *C19orf53*, *HKDC1*, *DLC1*, *PLEKHG3*, and *SETD6*. Genes in which protein kinase recognition sequences were predicted to be impacted were *ASXL1*, *ENTHD1*, *OTOF*, *PCDH17*, *PLK1*, *WNT2B*, *DLC1*, *LRFN1*, *FOXM1*, *ZNF879*, *H2BC18*, *TTN*, *IGF2R*, *HERC2*, *CYP4F11*, and *PPP1R18.* Using a gene-set enrichment analysis, in CP, we identified involvement in neuronal processes (ephrin receptor signaling, dendritic spine development and axon guidance) and muscle-related processes (titin binding and sarcomere organization) (Supplementary Table 9 sheets 1-2), whereas these processes were not identified in the pediatric control cohorts (Supplementary Table 9 sheets 3-6).

Clinical phenotype of CP cohort with & without P/LP variants

Characteristics of children with CP with P/LP variants compared to those with no variants of potential clinical relevance are presented in Table 4 (with contingency tables in Supplementary Table 5). Significant associations were found for normal brain MRI (p=0.03), communication difficulties (p=0.03), cognitive impairment (p=0.01), term birth (p=0.02) and consanguinity (p=0.01). Assessing antenatal/neonatal/postnatal risk factor profiles (Supplementary Table 3 sheet 6, column O; sheet 7-9, column N), three cases (8%) with P/LP variants had no risk factors, 5 (13.5%) had one and 27 (73%) had two or more risk factors. Thirty-four of 37 (92%) participants with P/LP variants had at least one aspect of their phenotype that was consistent with known features related to that gene (Supplementary Table 3 sheets 6-9).

Discussion

In a large study of meticulously-phenotyped children with CP, ascertained with clinically representative sampling, we identified through WGS and thorough genome annotation, P/LP variants in 11.3% and VUS in 17.7%. This study provides a comprehensive evaluation of the genomic architecture of CP, indicating contributions by multiple classes of rare variants including damaging SNVs and indels (6.7%) CNVs (3.4%) and mitochondrial mutations (1.5%). In parallel, we analyzed two healthy pediatric control cohorts (n=203 trios, and n=89 sib-pair families) to provide a contemporaneous baseline for the proper determination of *de novo* mutation rates and genetic burden analysis, the latter of which demonstrated associations between *de novo* deleterious variants and genes related to neuro-functions, nervous system and *FMR1* target genes. Importantly, we identified overlap in 13 of 19 genes (68%) with P/LP SNV variants in our CP cohort and three independent clinical genomic cohorts. These children also had a CP diagnosis and/or neuromotor disability, thus providing further evidence for motor involvement/presence of CP phenotype for these genes.

An important genetic contribution to etiologic risk for CP is now well established. Rates of P/LP variants in previous studies vary, depending largely on ascertainment criteria[17,41]. A systematic review of the diagnostic yield of exome sequencing and chromosomal microarray CNV in CP had an overall yield of 23% for exome sequencing and 5% for CNV[17]. A subgroup analysis of cryptogenic CP had a higher yield of 35% compared to non-cryptogenic where the yield was 7%. Our rate of 11.3% P/LP is similar to other non-cryptogenic, clinically representative cohorts [13,24].

WGS allows for reliable identification of mitochondrial variants which in this CP cohort were low-penetrance, homoplasmic variants associated with hearing loss (1494T and 7511C) as well as heteroplasmic variants associated with classical mitochondrial phenotypes, including m.3243A>G (mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episodes (MELAS)). For the latter, the heteroplasmy in affected offspring was higher in all three children than in their mothers. Mitochondrial variants represent a vulnerability to neural injury, and may require a synergistic interaction between environmental risk and this genetically determined defect in brain energy metabolism to present with a "CP phenotype". Ascertainment of mitochondrial variants in CP cohorts to date has been very limited[8]. This key finding supports the inclusion of genomic assessment for mitochondrial variants in the diagnostic work-up of CP.

The *COL4A1* gene had the most P/LP SNVs (four cases in our CP cohort and six in the clinical genomic cohorts). *COL4A1* encodes the alpha-1 subunit of collagen type IV and is a monogenic cause of small vessel cerebrovascular disease[42] including ischemic stroke, porencephaly, schizencephaly, and periventricular haemorrhage. *COL4A1* is known to be associated with CP[43-45]. The P/LP variants in our study were similar to those previously identified with the exception of *CTNNB1*, which was absent in our cohort. Beta catenin deficiency associated with *CTNNB1* variation has been identified as the most frequent monogenic cause of CP in a systematic review assessing diagnostic yield of exome sequencing [17]. This finding was largely driven by the high prevalence of *CTNNB1* LOF variants in a cohort referred for laboratory genetic testing [13]. Individuals with *CTNNB1* LOF variants also have ID and hence *CTNNB1* may be over-represented in the Moreno-De-Luca cohort. We note that while our study used the most comprehensive genomic analyses, we have taken a conservative approach in our interpretation using strict diagnostic criteria. Our yield of VUS was 17.7% and with more data to compare against, some of these may be reclassified in the future as pathogenic[46]. In fact, from our analysis of the CP and clinical genomic cohorts, *EXOSC3*, *TUBA1A*, *GATA2DB*, *GNAO1*, *TH*, *ADCY5*, *KDM3B*, *MEF2C*, *SPAST*, *TRIP12*, *ZDHHC9* and *ZIC2* demonstrate multiple lines of data supporting a role in CP-related phenotypes.

Normal brain imaging, communication difficulties, cognitive impairment, term birth and consanguinity were the phenotypic characteristics associated with P/LP variants in our CP cohort. The majority (73%) of children with P/LP variants also had multiple intrinsic and environmental antenatal/perinatal risk factors. This supports multifactorial causal pathways including genetic risk leading to CP. The association with consanguinity is an expected finding, increasing the child's risk of homozygous variants. We observed a higher yield of P/LP variants in term-born children, consistent with our earlier CNV study of hemiplegic CP[14]. Although we identified clinical characteristics associated with increased genomic yield, many children with

P/LP did not exhibit these phenotypic markers. Further study of clinical prediction models may help to define whether a "choosing wisely" strategy based on key phenotype markers or genomic testing for all children with CP, as has recently been proposed for neonatal hypotonia[47], is the preferred route. As WGS becomes mainstream, the cost/benefit ratio of diagnostic testing may shift to favor universal testing of all CP diagnoses or more broadly through population-based newborn screening[48]. Identifying a genetic etiology can inform clinical actionability and have repercussions on medicolegal proceedings, where historically perinatal asphyxia was the presumed cause in many children with CP[49].

A burden analysis of *de novo* variants and transmission tests of large TRs, to identify genetic associations with CP, showed an association between *de novo* LOF and missense variants and CP in genes related to the nervous system, neurological functions, and *FMR1* target genes. These gene-sets are known to be associated with other neurological disorders, including ASD and ID[33,50]. We found 20 genes involved in the signal from LOF variants, out of which eight (*SOX2, SPAST, MEF2C, ADCY5, PDE10A, TRIP12, NDUFV1* and *COL4A1*) have a known link with CP, while five (*R3HDM2, ZIC2, DSCAML1, PDZRN3* and *USP32*) are known to be LOF intolerant (gnomAD v2.1.1 LOEUF <0.35). *ZIC2* has been associated with holoprosencephaly[51]. Ninety-eight genes contributed to the signal in the *de novo* missense variant analysis. Associations between CP and damaging missense variants have been previously reported[4,27]. We found an over-transmission of large TRs in genes related to post-synaptic density and axon guidance pathways in children with CP. TR expansions are associated with over 60 disorders, the majority being neurologic, including ASD and schizophrenia[30,52-55]. This is noteworthy given the higher prevalence of ASD in children with CP[3] as well as an overlap of genes within neurodevelopmental disorders[4]. Four genes (*CACNB1, PLXNA4, CDH4, PTPRD*) were previously reported to be associated with TR expansions in ASD[30] and six (*DAB1, EFNA5, DAPK1, SHANK1, DOCK3* and *CAMK2B*) in schizophrenia[54]. Assessing further potential relationships with CP/neuromotor disabilities, *DOCK3* has been associated with a neurodevelopmental disorder with ataxia and ID (OMIM 618292), and *DAB1* with spinocerebellar ataxia (OMIM 615945). *DAPK1* has also been associated with ischemia-induced overstimulation of the glutamate receptor resulting in neuronal injury in mouse models[56]. By contrast, no enrichments of *de novo* variants nor over-transmission of large TRs were found in the pediatric control cohorts.

The assessment of variants driving the CP gene burden analysis but not reaching P/LP/VUS status and overlap of these variants with cases with CP/neuromotor disability in the clinical genomic cohorts allows an exploration of the potential contributions of these genes (*SMOC1, KDM5B, BCL11A, CYP51A1*) to the CP phenotype. Variants in these genes have not previously been reported to be associated with the CP phenotype/neuromotor disabilities, and our work suggests these as previously unrecognized CP candidate genes. Additional work will need to be undertaken to further investigate these associations.

Understanding the genetic underpinnings of CP is of utmost importance to accurately diagnose etiology, improve family counselling, use targeted interventions where available, and avoid treatments that may be harmful. Genetic testing will enable earlier identification of CP "mimics", allowing earlier targeted treatments and appropriate counseling[18]. In this study, the

CP status of one case was changed to a dopa-responsive dystonia caused by a homozygous missense variant in the tyrosine hydroxylase gene (Table 2) (OMIM:191290) [57], which can be treated effectively with levodopa[58,59]. Our study and others have identified individuals in CP cohorts with hereditary spastic paraplegia (HSP) [60]. HSP is a neurodisability that is distinct from CP, as HSP has more weakness, some forms are neurologically progressive, and its identification can inform rehabilitation strategies. Importantly, an enhanced understanding of the genetic pathways of CP will fuel future innovation of targeted treatments, including neural repair strategies.

This study had certain limitations. First, genomic DNA was prepared from saliva. In general, this provided adequate amounts of DNA, with only 1.2% of samples requiring additional sequencing runs and 4% excluded post-sequencing. We were able to circumvent most challenges using informatics allowing robust variant calling[61]. Second, assessment of the contribution of common variant polygenic risk contributors must await larger cohorts for genome-wide association studies. Moreover, the *de novo* enrichment analysis uses a case control design with modest numbers of controls. There also is incomplete data on paternal age. Given the small sample size/missing data, there is possibility of a type II error.

In conclusion, in this study of CP, which historically has been attributed to multi-risk environmental stressors resulting in brain damage, we identified P/LP variants in 11.3% of a clinically representative cohort, with contributions from multiple classes of damaging rare variants including SNVs/indels, CNVs and mitochondrial variants. In children with CP harbouring P/LP variants, multiple environmental CP risk factors were identified in addition to the genomic variant. This multifactorial risk profile, combined with a significant presence of P/LP variants, supports consideration for the inclusion of WGS with nuclear mitochondrial variant detection, in the diagnostic work-up of CP. The significant overlap of genes in our CP cohort with three clinical genomic cohorts (Table 2) suggests substantial pleiotropy in CP and other related/overlapping phenotypes (e.g. neuromotor disability). This observation adds another example to the developing story in neurodevelopmental/neuropsychiatric conditions where common genes and CNVs are increasingly found[62,63]. Our study provides deep phenotypic and genomic architecture data providing a framework to begin to dissect the multifactorial nature of CP. This type of precision medicine is of utmost importance to accurately diagnose etiology across all CP phenotypes, improve family counselling, and use targeted interventions where available.

Acknowledgements

This research was conducted in part through the Childhood Cerebral Palsy Neuroscience Discovery Network (CP-NET) with the support of the Ontario Brain Institute, an independent non-profit corporation, funded by the Ontario Government. This study was also funded by Project Grants from the Canadian Institutes of Health Research (PJT-153004 and PJT-175329 to RKCY), the Research Foundation of the Cerebral Palsy Alliance (PG4516), Debbie and Don Morrison for support of the CHILD study and the University of Toronto McLaughlin Centre. The Canadian CP Registry was funded by Kids Brain Health Network. RKCY is supported by The Hospital for Sick Children's Research Institute, SickKids Catalyst Scholar in Genetics, Brain

Canada, and The Azrieli Foundation. CHILD was initially funded by CIHR and AllerGen NCE, and receives support from Women and Children's Health Research Institute at the University of Alberta (to PM). Work at The Centre for Applied Genomics (TCAG) at The Hospital for Sick Children is funded by the Canada Foundation for Innovation (CGEn-MSI, Innovation Fund) and Genome Canada through Ontario Genomics. We would like to thank Miranda Lorenti for technical assistance with sample preparation and Vedika Jha, Julia Sangster and Sara McPherson for assistance with variant validation. We acknowledge Illumina, Inc. for providing sequencing reagents for the CHILD and Inova control cohorts. MA, PS and SET hold Canada Research Chairs. MO is a Research Scholar of the Fonds de Recherche du Quebec-Santé. During the work presented in this article, JWG held the Scotiabank Chair in Child Health Research at McMaster University. SWS holds the Northbridge Chair in Paediatric Research, a joint Hospital-University Chair between the University of Toronto, The Hospital for Sick Children, and the SickKids Foundation. We thank all participant families for their dedication and commitment to advancing health research.

Author contributions

Conceptualization: D.L. Fehlings, R.F. Wintle, M. Oskoui, M. Zarrei, W. Engchuan, S.W. Scherer, A. McCormick, A. Kawamura, C. Hunt, D. Samdup, J.W. Gorter, N. Dlamini, R. Mesterman, S. Knights, M. Shevell, A. Kirton. Data Collection: D.L. Fehlings, M. Oskoui, A. McCormick, A. Kawamura, C. Hunt, D. Samdup, J.W. Gorter, N. Dlamini, R. Mesterman, S. Knights, M. Shevell, A. Kirton, R.F. Wintle. Methodology/Data Interpretation: M. Zarrei, B. Thiruvahindrapuram, W. Engchuan, J.R. MacDonald, E.J. Higginbotham, B. Trost, M.B. Azad, G. Pellecchia, J. Wei, R.F. Wintle, R.K.C. Yuen, S.W. Scherer, D.L. Fehlings, M. Oskoui, N. Sondheimer, P. Subbarao, S.E. Turvey, P. Mandhane, T. Moraes, E. Simons, G.L. Maxwell, D. Merico, D.J. Stavropoulos, R. Manshaei, T. Nalpathamkalam, R. Shaath, S. Ghaffari, K. Ho, J. de Rijke, W.W.L. Sung, A. Mowjoodi, G. Costain, J.L. Michaud, C. Marshall, F.F. Hamdan, J. Gauthier, K. Uguen. Validation: S. Lamoureux, J. de Rijke, M. Hilali, K. Hirschfeld, K. Grover, N.X. Bautista, K. Han. Resources: S.W. Scherer, D.L. Fehlings, M. Oskoui. Data Curation: B. Thiruvahindrapuram, J.R. MacDonald, J. Whitney, P. Mandhane, P.S. Danthi, R.V. Patel, O. Hamdan, W.W.L. Sung. Visualization: W. Engchuan, G. Pellecchia. Supervision: S.W. Scherer, D.L. Fehlings, M. Oskoui, R.F. Wintle. Project Administration: R.F. Wintle, S.L. Pereira, S. Aimola, R. Thapa, P. Ng, T. Behlim, L. Switzer, J. Howe. Funding Acquisition: S.W. Scherer, D.L. Fehlings, R.F. Wintle, A. McCormick, A. Kawamura, C. Hunt, D. Samdup, J.W. Gorter, N. Dlamini, R. Mesterman, S. Knights, M. Oskoui, M. Shevell, A. Kirton.

Competing interests

Illumina, Inc. provided sequencing reagents for the CHILD and Inova control cohorts. SWS is on the Scientific Advisory Committee of Population Bio, Inc. and serves as a Highly Cited Academic Advisor for the King Abdulaziz University. RFW discloses consulting activities for Guidepoint Global, GLG, and Bioinformatics LLC, not related to this study. DM was employed by Deep Genomics at time of study, and some annotation tools were provided by Deep Genomics. DJS has equity in PhenoTips. The remaining authors declare no competing interests.

514 Table 1. **Pathogenic and likely pathogenic copy number variations identified in the CP cohort**

Case	Sex	Type	Location	Size (bp)	Genes/variant	inh/zyg	Class
264596	F	DUP	Xp22.33-Xq21.1	77,349,024	570 genes	dn	P
264596	F	DEL	Xq21.1-Xq28	77,903,179	*ZDHHC9* + 586 genes	dn	P
278746	M	DUP	21p13-21q22.3	48,129,895	383 genes (trisomy 21)	dn	P
19_13544	M	DUP	Yp11.32-Yq12	59,373,566	144 genes (XYY syndrome)	dn*	P
19_13560	M	DEL	2q13	1,726,363	*BCL2L11, BUB1* + 8 genes	dn	LP
256962	M	DUP	1q21.1-q21.2	1,366,000	*CHD1L, GJA5* + 22 genes	dn*	P
283751	F	DEL	17p12	1,320,800	*PMP22* + 11 genes	dn	P
308285	M	DEL	20p13-p12.3	673,800	*PCNA* + 9 genes	mat*	LP
21_5356	M	DEL	11q14.1	217,338	*DLG2*	dn	LP
20_5471	M	DEL	12q12	127,200	*ARID2*	dn	LP
314271	F	DEL	2p16.3	59,678	*NRXN1*	pat	LP
20_5423	M	DEL	11p11.2	15,000	*EXT2*	mat	LP

515
516 All variants were inspected using Integrative Genomics Viewer and subsequently confirmed using another method
517 (qPCR or Sanger sequencing), except those marked with an asterisk(*) where adequate DNA was unavailable. F: female;
518 M: male; P: pathogenic; LP: likely pathogenic; DEL: deletion, DUP: duplication; INS: insertion; mat: maternally inherited;
519 pat: paternally inherited; het: heterozygous; dn: de novo.
520

Table 2. **Pathogenic and likely pathogenic single nucleotide variants and indels identified in the CP cohort and overlap of SNVs in the clinical genomic cohorts**

Case	Sex	Type	Location	Size (bp)	Genes/variant	inh/zyg	Class	SK Clinical Cohort[a]	SK Complex Care[b]	SJ Clinical Cohort[c]	Any[d]
264594	F	SNV	3q21.1	1	*ADCY5*:c.2088+2T>G:p.?	dn, het	LP	het, VUS (inheritance unknown)^			YES
266349	M	SNV	13q34	1	*COL4A1*:c.G2317A:p.G773R	dn, het*	P	YES dn, het, P (2 instances)^	dn, het, P^	(3 instances)^: dn, het, LP; mat, het, LP; mat (mosaic), het, P	YES
324623	F	SNV	13q34	1	*COL4A1*:c.G2869A:p.G957R	dn, het	LP	YES dn, het, P (2 instances)^	dn, het, P^	(3 instances)^: dn, het, LP; mat, het, LP; mat (mosaic), het, P	YES
267092	F	SNV	13q34	1	*COL4A1*:c.1537-1G>A:p.?	dn, het	LP	YES dn, het, P (2 instances)^	dn, het, P^	(3 instances)^: dn, het, LP; mat, het, LP; mat (mosaic), het, P	YES
20_14777	F	DEL	13q34	1	*COL4A1*:c.1509delT:p.P503fs	pat, het	P	YES dn, het, P (2 instances)^	dn, het, P^	(3 instances)^: dn, het, LP; mat, het, LP; mat (mosaic), het, P	YES
324576	F	INS	8p11.23	14	*ERLIN2*:c.859_860ins AGGCCATTGCTTCC:p.K287fs	hom	LP				NO
261783	F	SNV	9p13.2	1	*EXOSC3*:c.A395C:p.D132A	hom	P	hom, P^	3 hom P, 1 compound het, P/LP (4 instances)^		YES
302991	F	SNV	1q21.3	1	*GATAD2B*:c.T1321C:p.C441R	dn, het	LP	dn, het, LP^		dn, het, P^	YES
19_1885	M	SNV	16q12.2	1	*GNAO1*:c.A711C:p.E237D	dn, het	P			(2 instances)^: dn, het, P; dn, het, LP	YES
323572	F	SNV	5q31.2	1	*KDM3B*:c.G4879A:p.E1627K	dn, het	LP	pat, het, LP^			YES
285253	M	DEL	5q14.3	3	*MEF2C*:c.51_54del:p.R17fs	dn, het	P			dn, het, LP^	YES
264583	M	DEL	6q27	1	*PDE10A*:c.1783delC:p.H595fs	dn, het	LP				NO
278746	M	INS	6q14.1	4	*PHIP*:c.2517_2518insATGG: p.H840fs	dn, het	P				NO
19_1851	M	SNV	Xp22.13	1	*PHKA2*:c.G557A:p.R186H	dn, hem	P				NO
20_983	F	SNV	19p13.11	1	*PIK3R2*:c.G1117A:p.G373R	dn, het	P				NO
324567	F	DEL	3q26.33	18	*SOX2*:c.59_78del:p.G20fs	dn, het	P				NO
20_13820	M	DEL	2p22.3	6	*SPAST*:c.1413+3_1413+6del:p.?	dn, het	LP		dn, het, P^		YES
324636	F	SNV	11p15.5	1	*TH*:c.G931A:p.G311S	hom	P		hom, P^	compound het (inherited), P^	YES
20_19577	M	SNV	2q36.3	1	*TRIP12*:c.3624+1G>A:p.?	dn, het	P	pat, het, VUS^			YES
260471	M	SNV	12q13.12	1	*TUBA1A*:c.G1000A:p.A334T	dn, het	P		dn, het, P^	(3 instances)^: dn, het, P; unknown inheritance, het, LP; unknown inheritance, het, P	YES
317945	M	SNV	Xq26.1	1	*ZDHHC9*:c.881+1G>T:p.?	mat, hem	P	mat, hem, LP^			YES
19_7945	M	DEL	3q32.3	1	*ZIC2*:c.1330delA:p.S444fs	dn, het	P	pat, het, VUS^			YES

All variants were inspected using Integrative Genomics Viewer and subsequently confirmed using another method (qPCR or Sanger sequencing), except those marked with an asterisk(*) where adequate DNA was unavailable. F: female; M: male; P: pathogenic; LP: likely pathogenic; DEL: deletion, DUP: duplication; SNV: single nucleotide variant; INS: insertion; mat: maternally inherited; pat: paternally inherited; het: heterozygous; hom: homozygous; hem: hemizygous. dn: de novo.

[a] Clinical sequencing cases from The Hospital for Sick Children (SK), Toronto, Canada; [b] Children with medical complexity followed by a structured Complex Care Program over a >10-year period (Haque et al. Arch Dis Child 2023); [c] Clinical sequencing cases from Hôpital Ste-Justine (SJ), Montreal, Quebec, Canada. Six of these have been previously reported (one *COL4A1*, two *GNAO1*, one *MEF2C*) (Hamdan et al., Am J Hum Genet 2017); one *TUBA1A* (Jurgens et al., Eur J Hum Genet 2021); one GATAD2B (Hamdan et al., PLoS Genet 2014); [d] 13/19 genes with SNV/indel variants were also found in at least one of the 3 clinical genomic cohorts. There was no overlap of individuals with variants in the cohorts.

^clinical phenotype encompasses CP/neuromotor disability

Table 3. **Pathogenic mitochondrial variants identified in the CP cohort**

Case	Sex	Gene/variant	inh/zyg	Class	Proband HP	Maternal HP	Mitomaster Phenotype	CP Phenotype (CP Subtype, GMFCS[b], MRI[c])	Identified Risk Factors
18_2847	F	*MT-RNR1*: 1494:C>T	mat, hom	P	0.9908	0.9541	Deafness	Hypotonic (no deafness), GMFCS unavailable, Basal ganglia/thalamic lesion	Intrauterine infection, intrapartum asphyxia, neonatal encephalopathy
19_1933	F	*MT-TL1*: 3243:A>G	dn, het	P	0.1133	0.0033	MELAS[a]	Mixed spastic quadriplegia/dystonic, GMFCS IV, Basal ganglia/thalamic lesion	Maternal hemorrhage/shock (uterine rupture), intrapartum asphyxia, neonatal encephalopathy, placental abnormalities: Multifocal chorangiosis changes and multifocal retroplacental blood clots.
20_1168	F	*MT-TQ*: 4332:G>A	mat, het	P	0.1937	0.0808	Encephalopathy/MELAS	Hemiplegia, GMFCS I, MRI unavailable	Prolonged/obstructed labour, anhydramnios, toxic exposure
21_14069	M	*MT-TS1*: 7511:T>C	mat, hom	P	0.9808	0.9926	Deafness/MELAS	Hemiplegia (no deafness), GMFCS I, Middle Cerebral Artery (MCA) infarct	Intrauterine infection
310799	M	*MT-ND5*: 13513:G>A: p.D393N	dn, het	P	0.4395	0.0006	Leigh Syndrome	Hemiplegia, GMFCS I, Periventricular venous infarction	Maternal illness: Essential hypertension with pregnancy induced hypertension, toxic exposure

All variants were SNV with a 1 bp size and were inspected using Integrative Genomics Viewer and subsequently confirmed using another method (qPCR or Sanger sequencing). HP: heteroplasmy; F: female; M: male; P: pathogenic; mat: maternally inherited; hom: homoplasmy; het: heteroplasmy; dn: de novo.

[a] MELAS, Mitochondrial myopathy, Encephalopathy, Lactic Acidosis, and Stroke-like episodes

[b] GMFCS, Gross Motor Function Classification System

[c] MRI, Magnetic Resonance Imaging

Table 3. **Characteristics of children in the CP cohort with pathogenic/likely pathogenic variants, variants of uncertain significance and no clinically relevant variants**

Participant Characteristics	P/LP Variant n=37 (11%)	VUS n=58 (18%)	No clinically relevant variant n=232 (71%)	Fisher's p-value*
Loss of clinical diagnosis of CP	1 (3%)	0 (0%)	0 (0%)	
Sex (n=327), Male	20 (54%)	39 (67%)	135 (58%)	0.72
CP Subtype (n=326)				0.06[a]
Spastic CP (n=285)	27 (73%)	54 (93%)	204 (88%)	
Unilateral spastic CP (hemiplegia)	12 (32%)	19 (33%)	94 (41%)	
Bilateral spastic CP	15 (41%)	35 (60%)	110 (47%)	
Dyskinetic CP (n=24)	3 (8%)	4 (7%)	17 (7%)	
Hypotonic/Ataxic/Other (n=17)	6 (16%)	0 (0%)	11 (5%)	
GMFCS [b] (n=308)				0.37[c]
Ambulatory (GMFCS Levels I - III) (n=242)	24 (65%)	45 (77%)	173 (75%)	
Non-Ambulatory (GMFCS Levels IV-V) (n=66)	9 (24%)	12 (21%)	45 (19%)	
Brain MRI (n=277) normal	3 (8%)	0 (0%)	2 (1%)	0.02
Brain MRI (n=277) with brain maldevelopments	4 (11%)	6 (10%)	10 (4%)	0.13
Cortical visual impairment (n=313)	3 (8%)	2 (3%)	17 (7%)	0.75
Sensorineural auditory impairment (n=310)	2 (5%)	8 (14%)	21 (9%)	0.75
Communication difficulties[d] (n=311)	21 (57%)	32 (55%)	108 (47%)	0.03
Cognitive impairment (n=134)	18 (49%)	19 (33%)	50 (22%)	0.01
Neonatal seizures [e] (n=290)	3 (8%)	13 (22%)	47 (20%)	0.10
Epilepsy [f] (n=290)	11 (30%)	9 (16%)	43 (18%)	0.30
Major/minor congenital malformation (n=316)	7 (19%)	6 (10%)	55 (24%)	0.54
Gestation (n=312)				
Term (≥37 weeks of gestation)	23 (62%)	29 (50%)	104 (45%)	0.02
Preterm (<37 weeks of gestation)	11 (30%)	26 (45%)	119 (51%)	
No identifiable risk factor (Idiopathic CP) (n=312)	3 (8%)	3 (5%)	17 (7%)	0.73
Small for gestational age[g] (n=299)	2 (5%)	9 (16%)	35 (15%)	0.19
Maternal age at time of delivery (n=322)				
>35 years	10 (27%)	18 (31%)	64 (28%)	1.00
≤35 years	27 (73%)	38 (65%)	165 (71%)	
Paternal age at time of delivery (n=172)				
>35 years	8 (22%)	12 (21%)	44 (19%)	0.60
≤35 years	10 (27%)	19 (33%)	79 (34%)	
History of stillbirth (n=256)	4 (11%)	4 (7%)	15 (6%)	0.51
Consanguinity (n=318)	4 (11%)	5 (9%)	5 (2%)	0.02
Family History of CP [h] (n=315)	2 (5%)	4 (7%)	13 (6%)	1.00

P/LP= pathogenic or likely pathogenic; VUS=variants of unknown significance; n=participants with data available; *association of pathogenic/likely pathogenic variants versus no variant (see Supplementary Table 5 for calculations including missing data). [a] association of spastic CP vs other CP subtypes (dyskinetic, ataxic, hypotonic, other); [b] GMFCS (Gross Motor Function Classification System); [c] association of ambulatory (GMFCS I-III) vs to non-ambulatory (GMFCS IV-V); [d] CP-NET cohort (score of Level II-V on Communication Function Classification System **or** Functional Communication Classification System)/CP Registry cohort (documentation of presence of communication difficulties); [e] history of ≥1 seizures in the first 28 days of life; [f] documented seizure outside of neonatal period; [g] weight for gestational age <10th percentile; [h] presence of CP in the proband's first and/or second-degree relatives.

Figure Legends/Captions

Figure. 1: The number (%) of cases with subtypes of variants of potential clinical relevance in cerebral palsy. Numbers and percentages (%) of cases with each class of variants (P/LP/VUS) identified in the CP cohort.

Figure 2: Gene-set analysis of *de novo* SNVs/indels and large tandem repeats. Three panels represent gene-set level analysis results of *de novo* LOF variants (burden analysis), *de novo* missense variants (burden analysis) and large tandem repeats (transmission test). Colours indicate the different cohorts. Points indicate effect size as estimated by Fisher's exact test as the odds ratio with error bars indicating 95% confidence intervals (n=203, 178, and 314 for INOVA, CHILD and CP, respectively). Significant results are marked with asterisks, i.e., those with BH-FDR<10%.

References

1 McIntyre, S. *et al.* Global prevalence of cerebral palsy: A systematic analysis. *Developmental Medicine & Child Neurology* **64,** 1494-1506 , doi:https://doi.org/10.1111/dmcn.15346 (2022).

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 Christensen, D. *et al.* Prevalence of cerebral palsy, co-occurring autism spectrum disorders, and motor functioning - Autism and Developmental Disabilities Monitoring Network, USA, 2008. *Dev Med Child Neurol* **56**, 59-65, doi:10.1111/dmcn.12268 (2014).

4 Jin, S. C. *et al.* Mutations disrupting neuritogenesis genes confer risk for cerebral palsy. *Nat Genet* **52**, 1046-1056, doi:10.1038/s41588-020-0695-1 (2020).

5 Amankwah, N. *et al.* Cerebral palsy in Canada, 2011-2031: results of a microsimulation modelling study of epidemiological and cost impacts. *Health Promot Chronic Dis Prev Can* **40**, 25-37, doi:10.24095/hpcdp.40.2.01 (2020).

6 Friedman, J. M., van Essen, P. & van Karnebeek, C. D. M. Cerebral palsy and related neuromotor disorders: Overview of genetic and genomic studies. *Mol Genet Metab*, doi:10.1016/j.ymgme.2021.11.001 (2021).

7 Chopra, M. *et al.* Mendelian etiologies identified with whole exome sequencing in cerebral palsy. *Ann Clin Transl Neurol* **9**, 193-205, doi:10.1002/acn3.51506 (2022).

8 Li, N. *et al.* In-depth analysis reveals complex molecular aetiology in a cohort of idiopathic cerebral palsy. *Brain* **145**, 119-141, doi:10.1093/brain/awab209 (2022).

9 Mei, H. *et al.* Genetic Spectrum Identified by Exome Sequencing in a Chinese Pediatric Cerebral Palsy Cohort. *J Pediatr* **242**, 206-212.e206, doi:10.1016/j.jpeds.2021.11.019 (2022).

10 Rosello, M. *et al.* Hidden etiology of cerebral palsy: genetic and clinical heterogeneity and efficient diagnosis by next-generation sequencing. *Pediatr Res* **90**, 284-288, doi:10.1038/s41390-020-01250-3 (2021).

11 Segel, R. *et al.* Copy number variations in cryptogenic cerebral palsy. *Neurology* **84**, 1660-1668, doi:10.1212/wnl.0000000000001494 (2015).

12 May, H. J. *et al.* Genetic testing in individuals with cerebral palsy. *Dev Med Child Neurol* **63**, 1448-1455, doi:10.1111/dmcn.14948 (2021).

13 Moreno-De-Luca, A. *et al.* Molecular Diagnostic Yield of Exome Sequencing in Patients With Cerebral Palsy. *Jama* **325**, 467-475, doi:10.1001/jama.2020.26148 (2021).

14 Zarrei, M. *et al.* De novo and rare inherited copy-number variations in the hemiplegic form of cerebral palsy. *Genet Med* **20**, 172-180, doi:10.1038/gim.2017.83 (2018).

15 Takezawa, Y. *et al.* Genomic analysis identifies masqueraders of full-term cerebral palsy. *Ann Clin Transl Neurol* **5**, 538-551, doi:10.1002/acn3.551 (2018).

16 Kayumi, S. *et al.* Genomic and phenotypic characterization of 404 individuals with neurodevelopmental disorders caused by CTNNB1 variants. *Genet Med* **24**, 2351-2366, doi:10.1016/j.gim.2022.08.006 (2022).
17 Srivastava, S. *et al.* Molecular Diagnostic Yield of Exome Sequencing and Chromosomal Microarray in Cerebral Palsy: A Systematic Review and Meta-analysis. *JAMA Neurology* **79**, 1287-1295, doi:10.1001/jamaneurol.2022.3549 (2022).
18 Pearson, T. S., Pons, R., Ghaoui, R. & Sue, C. M. Genetic mimics of cerebral palsy. *Mov Disord* **34**, 625-636, doi:10.1002/mds.27655 (2019).
19 MacLennan, A. H. *et al.* Genetic or Other Causation Should Not Change the Clinical Diagnosis of Cerebral Palsy. *J Child Neurol* **34**, 472-476, doi:10.1177/0883073819840449 (2019).
20 Hale, A. T. *et al.* Genome-Wide Association Study Identifies Genetic Risk Factors for Spastic Cerebral Palsy. *Neurosurgery* **89**, 435-442, doi:10.1093/neuros/nyab184 (2021).
21 Corbett, M. A. *et al.* Pathogenic copy number variants that affect gene expression contribute to genomic burden in cerebral palsy. *NPJ Genom Med* **3**, 33, doi:10.1038/s41525-018-0073-4 (2018).
22 McMichael, G. *et al.* Rare copy number variation in cerebral palsy. *Eur J Hum Genet* **22**, 40-45, doi:10.1038/ejhg.2013.93 (2014).
23 Oskoui, M. Disentangling racial and ethnic disparities in cerebral palsy. *Developmental Medicine & Child Neurology* **57**, 791-792, doi:https://doi.org/10.1111/dmcn.12772 (2015).
24 McMichael, G. *et al.* Whole-exome sequencing points to considerable genetic heterogeneity of cerebral palsy. *Mol Psychiatry* **20**, 176-182, doi:10.1038/mp.2014.189 (2015).
25 Kadotani T, W. Y., Saito T, Sawano K, Minatozaki K. A Chromosomal Study on 100 Cases of Cerebral Palsy. *International Journal of Human Genetics* **1**, 109-112 (2001).
26 Oskoui, M. *et al.* Clinically relevant copy number variations detected in cerebral palsy. *Nat Commun* **6**, 7949, doi:10.1038/ncomms8949 (2015).
27 van Eyk, C. L. *et al.* Yield of clinically reportable genetic variants in unselected cerebral palsy by whole genome sequencing. *NPJ Genom Med* **6**, 74, doi:10.1038/s41525-021-00238-0 (2021).
28 Pham, R. *et al.* Definition and diagnosis of cerebral palsy in genetic studies: a systematic review. *Dev Med Child Neurol* **62**, 1024-1030, doi:10.1111/dmcn.14585 (2020).
29 Costain, G. *et al.* Genome Sequencing as a Diagnostic Test in Children With Unexplained Medical Complexity. *JAMA Netw Open* **3**, e2018109, doi:10.1001/jamanetworkopen.2020.18109 (2020).
30 Trost, B. *et al.* Genome-wide detection of tandem DNA repeats that are expanded in autism. *Nature* **586**, 80-86, doi:10.1038/s41586-020-2579-z (2020).
31 Trost, B. *et al.* Genomic architecture of autism from comprehensive whole-genome sequence annotation. *Cell* **148**, 4409-4427 .e4418, doi:10.1016/j.cell.2022.10.009 (2022).
32 Haque, B. *et al.* Contemporary aetiologies of medical complexity in children: a cohort study. *Arch Dis Child* **108**, 147-149, doi:10.1136/archdischild-2022-325094 (2023).
33 Darnell, J. C. *et al.* FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146**, 247-261, doi:10.1016/j.cell.2011.06.013 (2011).
34 Cheng, F. *et al.* Comprehensive characterization of protein-protein interactions perturbed by disease mutations. *Nat Genet* **53**, 342-353, doi:10.1038/s41588-020-00774-y (2021).
35 Ramazi, S. & Zahiri, J. Posttranslational modifications in proteins: resources, tools and prediction methods. *Database (Oxford)* **2021**, doi:10.1093/database/baab012 (2021).
36 Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. *bioRxiv*, 2021.2010.2004.463034, doi:10.1101/2021.10.04.463034 (2021).
37 Wan, J. *et al.* Mutations in the RNA exosome component gene EXOSC3 cause pontocerebellar hypoplasia and spinal motor neuron degeneration. *Nat Genet* **44**, 704-708, doi:10.1038/ng.2254 (2012).

38 Fu, Y. *et al.* RGS-insensitive G-protein mutations to study the role of endogenous RGS proteins. *Methods Enzymol* **389**, 229-243, doi:10.1016/s0076-6879(04)89014-1 (2004).

39 Yokoi, S. *et al.* TUBA1A mutation can cause a hydranencephaly-like severe form of cortical dysgenesis. *Sci Rep* **5**, 15165, doi:10.1038/srep15165 (2015).

40 Okumura, A. *et al.* A patient with a GNAO1 mutation with decreased spontaneous movements, hypotonia, and dystonic features. *Brain Dev* **40**, 926-930, doi:10.1016/j.braindev.2018.06.005 (2018).

41 Gonzalez-Mantilla, P. J. *et al.* Diagnostic Yield of Exome Sequencing in Cerebral Palsy and Implications for Genetic Testing Guidelines: A Systematic Review and Meta-analysis. *JAMA Pediatrics* **177**, 472-478, doi:10.1001/jamapediatrics.2023.0008 (2023).

42 Lanfranconi, S. & Markus, H. S. COL4A1 mutations as a monogenic cause of cerebral small vessel disease: a systematic review. *Stroke* **41**, e513-518, doi:10.1161/strokeaha.110.581918 (2010).

43 Meuwissen, M. E. *et al.* The expanding phenotype of COL4A1 and COL4A2 mutations: clinical data on 13 newly identified families and a review of the literature. *Genet Med* **17**, 843-853, doi:10.1038/gim.2014.210 (2015).

44 Yoneda, Y. *et al.* Phenotypic spectrum of COL4A1 mutations: porencephaly to schizencephaly. *Ann Neurol* **73**, 48-57, doi:10.1002/ana.23736 (2013).

45 Khalid, R. *et al.* COL4A1 and fetal vascular origins of schizencephaly. *Neurology* **90**, 232-234, doi:10.1212/wnl.0000000000004890 (2018).

46 Costain, G. *et al.* Periodic reanalysis of whole-genome sequencing data enhances the diagnostic advantage over standard clinical genetic testing. *Eur J Hum Genet* **26**, 740-744, doi:10.1038/s41431-018-0114-6 (2018).

47 Morton, S. U. *et al.* Multicenter Consensus Approach to Evaluation of Neonatal Hypotonia in the Genomic Era: A Review. *JAMA Neurol* **79**, 405-413, doi:10.1001/jamaneurol.2022.0067 (2022).

48 Kingsmore, S. F. *et al.* A genome sequencing system for universal newborn screening, diagnosis, and precision medicine for severe genetic diseases. *Am J Hum Genet* **109**, 1605-1619, doi:10.1016/j.ajhg.2022.08.003 (2022).

49 Donn, S. M., Chiswick, M. L. & Fanaroff, J. M. Medico-legal implications of hypoxic-ischemic birth injury. *Semin Fetal Neonatal Med* **19**, 317-321, doi:10.1016/j.siny.2014.08.005 (2014).

50 Chiurazzi, P. & Pirozzi, F. Advances in understanding - genetic basis of intellectual disability. *F1000Res* **5**, doi:10.12688/f1000research.7134.1 (2016).

51 Brown, S. A. *et al.* Holoprosencephaly due to mutations in ZIC2, a homologue of Drosophila odd-paired. *Nat Genet* **20**, 180-183, doi:10.1038/2484 (1998).

52 Depienne, C. & Mandel, J. L. 30 years of repeat expansion disorders: What have we learned and what are the remaining challenges? *Am J Hum Genet* **108**, 764-785, doi:10.1016/j.ajhg.2021.03.011 (2021).

53 Marshall, C. R. *et al.* Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat Genet* **49**, 27-35, doi:10.1038/ng.3725 (2017).

54 Mojarad, B. A. *et al.* Genome-wide tandem repeat expansions contribute to schizophrenia risk. *Mol Psychiatry* **27**, 3692-3698, doi:10.1038/s41380-022-01575-x (2022).

55 Wen, J. *et al.* Rare tandem repeat expansions associate with genes involved in synaptic and neuronal signaling functions in schizophrenia. *Molecular Psychiatry* **28**, 475-482, doi:10.1038/s41380-022-01857-4 (2023).

56 Tu, W. *et al.* DAPK1 interaction with NMDA receptor NR2B subunits mediates brain damage in stroke. *Cell* **140**, 222-234, doi:10.1016/j.cell.2009.12.055 (2010).

57 Chen, Y. *et al.* Clinical and Genetic Heterogeneity in a Cohort of Chinese Children With Dopa-Responsive Dystonia. *Front Pediatr* **8**, 83, doi:10.3389/fped.2020.00083 (2020).

58 Diaz Heijtz, R., Almeida, R., Eliasson, A. C. & Forssberg, H. Genetic Variation in the Dopamine System Influences Intervention Outcome in Children with Cerebral Palsy. *EBioMedicine* **28**, 162-167, doi:10.1016/j.ebiom.2017.12.028 (2018).
59 Wijemanne, S. & Jankovic, J. Dopa-responsive dystonia--clinical and genetic heterogeneity. *Nat Rev Neurol* **11**, 414-424, doi:10.1038/nrneurol.2015.86 (2015).
60 Fink, J. K. Hereditary spastic paraplegia: clinical principles and genetic advances. *Semin Neurol* **34**, 293-305, doi:10.1055/s-0034-1386767 (2014).
61 Trost, B. *et al.* Impact of DNA source on genetic variant detection from human whole-genome sequencing data. *J Med Genet* **56**, 809-817, doi:10.1136/jmedgenet-2019-106281 (2019).
62 Jacquemont, S. *et al.* Genes To Mental Health (G2MH): A Framework to Map the Combined Effects of Rare and Common Variants on Dimensions of Cognition and Psychopathology. *Am J Psychiatry* **179**, 189-203, doi:10.1176/appi.ajp.2021.21040432 (2022).
63 Zarrei, M. *et al.* Gene copy number variation and pediatric mental health/neurodevelopment in a general population. *Hum Mol Genet* **32**, 2411-2421, doi:10.1093/hmg/ddad074 (2023).

Methods

Ethics Statement

This research complies with all relevant ethical regulations. Research Ethics Board approval was obtained from each participating centre (CP-NET REB #: CTO-1517; Holland Bloorview Kids Rehabilitation Hospital-13-450; Queen's University PAED-353-14; Laurentian University 2014-05-11; Western University 105221; McMaster University 14-332; Children's Hospital of Eastern Ontario 15/01E; Health Sciences North 974. CCPR REB #: MUHC 2017-3286; Alberta Children's Hospital REB18-0601) and all caregivers and participants (if able) provided written informed consent.

Study design and participants

Children with CP and parents were recruited from paediatric rehabilitation hospital clinics associated with nine participating sites of the Childhood Cerebral Palsy Integrated Neuroscience Discovery Network (CP-NET) in Ontario, and the Calgary and Montreal sites of the Canadian Cerebral Palsy Registry (CCPR). Participants were enrolled between August 2014 and September 2020 (CP-NET) and December 2017 and February 2021 (CCPR). Children were eligible if they were born on or after January 1, 2009, were at least 2 years old, and had a physician-confirmed CP diagnosis. The CP diagnosis was confirmed by chart review using the consensus definition[2], and reconfirmed after 5 years of age. Standardized data were collected on CP risk factors, neuroimaging, and neurodevelopmental status. DNA was extracted from saliva samples provided by the index child and biological parents using DNA Genotek Oragene OG-500 & OG-575 collection kits, and sent to The Centre for Applied Genomics (TCAG) at The Hospital for Sick Children for WGS. The study cohort (n=327) was a subgroup of participants where trios (samples for index child and biological parents) were available. Included participants were 59.3% male and had a mean age at enrollment of 5.0 years (standard deviation = 2.5). Participants were reimbursed for travel to a clinical site and were provided with a gift card for compensation of their time.

Pediatric control samples

To generate pediatric control cohorts (Supplementary Table 2), 89 quartet families (parents and two healthy children) from Inova (www.inova.org; Virginia USA) (approved by the Inova Institutional Review Board (IRB) (#15-1804) and the Western IRB (#20120204)), and 203 trios (parents and child) from the CHILD cohort[64] (approved by the Hamilton Integrated Research Ethics Board, the Universities of Manitoba, Alberta, and British Columbia, and The Hospital for Sick Children, Toronto REB#1000060059) were whole-genome-sequenced. These data were used for i) establishing *de novo* rates for LOF, damaging missense, SVs, and CNVs, and ii) gene-set burden (LOF, missense variants, phosphorylation) analysis and transmission disequilibrium tests (tandem repeats). Bioinformatic and annotation processes including variant calling and rare variant definition were identical to procedures used for CP samples. The majority of the CHILD samples were of European ancestry (96.6%), whereas of samples in the Inova cohort represented 56.4% European, 17.3% Admixed American and 4.5% East Asian individuals.

Clinical genomic cohorts

To identify further evidence for involvement in CP of the genes with P/LP SNV variants identified in our CP cohort), we investigated three additional genetically well-characterized,

phenotypically heterogeneous clinical cohorts for reported variants in these genes. We also assessed for reported variants in these cohorts that overlapped with findings driving the gene burden enrichment results that were identified as having either *de novo* LOF or damaging missense variants, without achieving P/LP/VUS status. For the subset of individuals in these additional clinical cohorts with variants of interest, we reviewed the available phenotypic data for the presence of a neuromotor disability and/or CP diagnosis. Further details are found in Supplementary Note 1.

Samples and sequencing

WGS was performed on DNA extracted from saliva (CP Cohort) or blood (CHILD and Inova cohorts) on Illumina HiSeq X (CP and CHILD) or NovaSeq 6000 (Inova) instruments (Illumina Inc., San Diego, CA), following recommended protocols with PCR-free library preparation and 150 nt paired end reads. Prior to library preparation, all samples were tested with 44 markers on a mass spec genotyping system (Agena Inc., San Diego, CA) to verify parent-child relationships. Sample identity and relationships were re-verified from WGS genotypes post-sequencing using PLINK (version v1.9.b3.42) [65]. To assess ancestry, we used GCTA v1.93.2[66] to compute principal components (PCs) on a reference set from the 1000 Genomes Project[67] (n=2,541) and the Human Genome Diversity Panel[68] (n=929). We projected the CP and both pediatric control cohorts onto the reference samples using precomputed SNP loadings. PCs were generated using the PC loading and projection option in GCTA. We used Kinship-based Inference for Genome-wide association studies (KING v2.2.5[69]) to verify family relationships and infer identity by descent (IBD). Further details are found in Supplementary Note 1.

Bioinformatics

FASTQ files were aligned against Human Genome Reference Build GRCh37/hg19 using Burrows-Wheeler Aligner (BWA, version v0.7.12). Population structures were identified using Genome-wide Complex Trait Analysis (GCTA, version v 1.93.2) (see Supplementary Note 1) [66]. SNVs and indels were detected using the Genome Analysis Toolkit (GATK; version v3.7) and variants were annotated using a pipeline based on ANNOVAR[31,70] (further details in Supplementary Note 1). We detected CNVs using two algorithms, CNVnator[71] and ERDS[7270,73]. We retained CNVs that were ≥1kb and identified by both CNVnator and ERDS with >50% reciprocal overlap. A stringent set of SVs was detected using Manta[74] and DELLY[75] (see Supplementary Note 1). Mitochondrial pathogenic variants were annotated using an ANNOVAR-based custom script that determines the most common sequence at each position but also identifies heteroplasmic variants across the genome[31]. Annotations from MitoMaster[76] used the revised Cambridge Reference Sequence (rCRS) as a reference[77]. Following our standard approach[30], ExpansionHunter (version 3.0.2) [78] was used to estimate the size of tandem DNA repeats at 59 known disease-causing repeat loci (Supplementary Table 10). Loci that had a size above the known disease-causing threshold were detected and further analyzed manually for their clinical relevance. Genome-wide tandem repeats were also detected using ExpansionHunter Denovo (EHdn; version 0.7.0) [79].

Rare SNVs and indels were defined as those present in fewer than 1% of individuals in 1000 Genomes Project[80]gnomAD v3.1.2 [81], and MSSNG parental samples[31]. We also required rare variants to be present in <1% of parental samples in the CP group. Rare CNVs were defined as those with <1% frequency in the WGS control CNVs (defined as parents in MSSNG cohort[31]), <1% frequency in 10,851 microarray controls [details in [82]], overlapping with a

region of the genome that is at least 75% copy-number stable, according to the CNV map of the human genome[83], and are less than 70% overlapped by segmental duplications. Rare CNVs in the CP cohort were further defined as those not overlapping with CNV clusters in the Database of Genomic Variants that had frequencies of ≥1%[84].

LOF variants were frameshift insertions, frameshift deletions, those affecting canonical splice-sites and stop-gains. Intronic variants predicted to have effects on splicing using spliceAI[85], SPIDEX[86] or dbscSNV[87] were also classified as LOF (see Supplementary Note 1 for details). Missense variants were defined as damaging if they satisfied cutoffs for at least five of seven criteria as we have described previously (SIFT, PolyPhen, MutationTaster, PhyloP mammalian or vertebrate, CADD, pfam domains, PhastCons elements; see Supplementary Note 1 for details) [31]. Recessive genes with a link to CP were assessed for homozygous, hemizygous, or compound heterozygous mutations within SNVs and indels, CNVs and SVs.

Genomic variants of potential clinical relevance

Pathogenic/likely pathogenic variants (P/LP) were defined as genetic/genomic variation in any region of the human genome, including the mitochondrial genome that have corresponding phenotypic clinical significance. Variants of uncertain significance (VUS) were defined as variants of emerging phenotypic clinical significance with a link to CP and associated disorders, but without sufficient data to classify them as P/LP. P/LP/VUS were evaluated by an expert clinical geneticist (author DJS) following the American College of Medical Genetics (ACMG) and Genomics guidelines[82,88-90]. Criteria for determining P/LP included i) an established strong link in the scientific literature between gene and disease, ii) molecular characteristics of the variant with P/LP falling in the category of loss-of-function or damaging missense versus tolerated missense, and iii) assessment of the transmission pattern of the variant, with greater strength of evidence for pathogenicity given to *de novo* variants. Inherited CNVs that were called as P/LP likely represent instances of incomplete penetrance or variable expressivity of phenotype in the parents. All variants (SNVs, indels, CNVs, mitochondrial variants, and SVs) were identified in the following four classes: aneuploidies, chromosomal abnormalities, known genomic syndromes, and deletions or duplications impacting genes associated with CP or associated neurodevelopmental disorders/comorbidities (e.g., ASD, ID, epilepsy). We annotated variants in all NCBI RefSeq genes, including all genes in OMIM, the NIH Clinical Genomic Database, our own curated list of 1,411 unique genes compiled from the literature (Supplementary Table 1), the GeneDx list of genes implicated in CP/NDD and the CP PanelApp genes (https://panelapp.agha.umccr.org/panels/73/) (Supplementary Table 1 sheet 3). Single nucleotide changes and insertions in the nuclear and mitochondrial genomes (with >5% heteroplasmy) were identified as P/LP when adequate evidence for pathogenicity of the variant was available following ACMG standards using annotations from MitoMaster[91]. Tandem repeats were also investigated for their potential pathogenicity associated with CP. All variants found in a proband were also assessed in parental samples to determine inheritance.

Statistical analysis of *de novo* variants and tandem repeats

A gene-set burden analysis was performed for *de novo* variants in the CP or pediatric control cohorts, compared with singleton variants in CP/NDD associated pathways from gnomAD population samples. The CHILD and Inova cohorts were analyzed both together and

independently. Due to a limited number of trio controls available, as a proxy for *de novo* variants, we used singleton variants from the 71,702 gnomAD samples [8192]. We collected twenty-four gene-sets representing neurological functions, brain expression, neurological phenotypes, and growth, along with seven unrelated (negative control) gene-sets. A Fisher's Exact Test was used to compare rates of *de novo* and singleton variants in a gene-set separately for *de novo* LOF, damaging missense, and missense variants. An odds ratio >1 indicates an enrichment of variants impacting the gene-set (i.e., more variants found than expected in the general population [gnomAD singletons]). Benjamini-Hochberg's false discovery rate (BH-FDR) was used to correct for multiple testing, with BH-FDR <10% considered significant. As a negative control, *de novo* synonymous variants were analyzed in the same manner to validate that this method does not pick up significant association by chance.

Tandem repeats (TR) were detected genome-wide in CP, and control cohorts using EHdn along with the 1000 Genomes Project (1KGP[67]) and the Medical Genome Reference Bank (MGRB[93]). Tandem repeat regions with rare expansions (<0.1% prevalence in 1KGP and MGRB) were identified using density-based spatial clustering of applications with noise (DBSCAN)[94]. Due to the lack of large TR data from a comparable set of control samples, transmission analysis of TR was performed[30]. A one-sided Fisher's Exact Test was applied to test whether large TR (repeat size of ≥99th percentile) of regions with rare expansions were over-transmitted to children. The number of large TR detected in the parents was used as the denominator, while the number transmitted to children was used as the numerator. The over-transmission statistic (odds ratio) is the ratio between the rate of transmitted large TR from the set of interest over the transmission rate of other large TR, where an odds ratio >1 indicates over-transmission. Analyses were done for genic, intergenic, different genic elements (i.e., exons, introns, UTRs, upstream, downstream, splice sites) and the 31 gene-sets. BH-FDR was used for multiple testing correction. Genic regions and gene-sets with BH-FDR <10% were considered significant. The CHILD and Inova cohorts were analyzed independently due to sequencing platform differences potentially influencing a difference in the number of TR detected[30]. Linear regression analysis was used to test if parental age (paternal and maternal) significantly predicted total number of *de novo* SNVs/indels in the CP cohort.

In silico functional analysis of CP-associated missense variants

The mechanisms of loss of function variants and copy number variations likely involve changes in gene dosage; however, the mechanisms of missense variants are more challenging to predict, given the wide variety of mechanisms by which they can be deleterious. To gain preliminary functional insight into potential molecular mechanisms of missense variants in CP-affected children from our cohort, we used *in silico* approaches to investigate two potential mechanisms: phosphorylation-mediated signaling[35] and protein-protein interactions[34] (See Supplementary Note 1 for details).

Assessment of the phenotypic characteristics of children with CP with and without pathogenic/likely pathogenic variants

Phenotypic characteristics evaluated included sex, CP sub-type (spastic CP compared to other subtypes), Gross Motor Function Classification System(GMFCS) [2] as ambulatory (Levels I-III) or non-ambulatory (IV-V), brain magnetic resonance imaging (MRI) (normal brain MRI

compared to abnormal; brain maldevelopments compared to no brain maldevelopments), cortical visual impairment, sensorineural auditory impairment, communication difficulties, cognitive impairment, neonatal seizures, epilepsy, congenital malformations, gestational age (term compared to preterm), no identified etiologic risk (definition of idiopathic CP), small for gestational age, parental age (≤35 years or >35 years), history of still birth, consanguinity and family history of CP. Further details on the range of antenatal, neonatal and postnatal etiologic risk factors assessed are outlined in Supplementary Note 1. When calculating the percentages of each characteristic, missing values were included in the denominator, ultimately to report on the percentage of children from the entire cohort who had that characteristic. Fisher's Exact Test was used to assess whether there was a difference in the characteristics of children with CP with P/LP variants compared to no variants. Individuals with missing data were removed from these analyses.

To identify the association between the clinical phenotype and identified P/LP variant, the affected chromosome, cytoband, chromosomal segment, type of mutation, origin of mutation, and zygosity were identified. The UCSC Genome Browser was used to visualize the affected chromosomal segments and identify genes. With missense variants, the exact base-pair substitutions and the amino acid changes were identified in ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/intro/). If the exact change was not identified either in ClinVar or by literature search, then the clinical features of a substitution by a different nucleotide at that location were explored. Information on known neuromotor, neurodevelopmental, or health co-morbidities associated with each genomic variant (SNV/indel, CNV, SV, mitochondrial) was collected from OMIM and literature searches. This was then compared to the clinical phenotype (neuromotor phenotype (CP subtype, GMFCS), neurodevelopmental (cognitive, communication and behavioral characteristics including any co-existing neurodevelopmental disorders (e.g., autism spectrum disorder (ASD), attention-deficit/hyperactivity disorder (ADHD), global developmental delay and intellectual disability), or health co-morbidities (including epilepsy/seizures, any congenital malformation(s), cortical visual impairment, ocular visual impairment, sensorineural auditory impairment, gastroesophageal reflux disease (GERD), constipation, g-tube feeding, and hip subluxation) of each participant with a P/LP variant. The likelihood that clinical features were associated with the genomic variants was categorized by a paediatrician (authors DF for CP-NET, MO for CCPR) with consensus reached between the two as i) likely or possibly – child's clinical picture is or may be consistent in part with the genomic variant identified, or ii) unlikely – child's clinical picture is not consistent with the genomic variant identified. The percentage agreement of a relationship (likely/possible versus unlikely) between P/LP variants as a group and clinical phenotype (neuromotor, neurodevelopmental, health co-morbidities) was also assessed.

Statistics & Reproducibility

This is a genomic analysis of a subgroup of participants (where trios (samples for index child and biological parents) were available) of the prospective observational CP-NET and CCPR CP cohorts. No statistical method was used to predetermine sample size but this the largest WGS CP cohort study to date (n=327) which is more than double the samples of the two previous studies at n=120[8] and n=150[27] respectively. No data were excluded for the analyses.

This study involved participants in prospective observational cohorts and therefore there was no randomization or blinding. Data met the assumptions of the statistical tests used.

Data availability

WGS data from CP families from the Canadian CP Registry and control data from the CHILD cohort are available at the European Genome-Phenome Archive (Canadian CP Registry, https://ega-archive.org/dacs/EGAC00001003068; CHILD: https://ega-archive.org/dacs/EGAC00001002953). WGS data from CP-NET are stored in the Brain-CODE Neuroinformatics Platform (https://www.braincode.ca/content/controlled-data-releases#dr022Access; doi: 10.60955/fszr-5q79). Access to the EGA data sets is governed by a Data Access Committee with primary contact via author SWS. Access to the Brain-CODE data set is governed by a Data Access Committee managed by the Ontario Brain Institute, with primary contacts via authors RFW and DLF. Control data from subjects enrolled in the Inova cohort were not consented for deposition in public databases and are thus available on request from author GM.

Code availability

This study did not use any custom code or software.

Methods-only References

64 Moraes, T. J. *et al.* The Canadian healthy infant longitudinal development birth cohort study: biological samples and biobanking. *Paediatr Perinat Epidemiol* **29**, 84-92, doi:10.1111/ppe.12161 (2015).

65 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-575, doi:10.1086/519795 (2007).

66 Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82, doi:10.1016/j.ajhg.2010.11.011 (2011).

67 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).

68 Bergström, A. *et al.* Insights into human genetic variation and population history from 929 diverse genomes. *Science* **367**, doi:10.1126/science.aay5012 (2020).

69 Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867-2873, doi:10.1093/bioinformatics/btq559 (2010).

70 Yuen, R. K. C. *et al.* Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat Neurosci* **20**, 602-611, doi:10.1038/nn.4524 (2017).

71 Abyzov, A., Urban, A. E., Snyder, M. & Gerstein, M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res* **21**, 974-984, doi:10.1101/gr.114876.110 (2011).

72 Zhu, M. *et al.* Using ERDS to infer copy-number variants in high-coverage genomes. *Am J Hum Genet* **91**, 408-421, doi:10.1016/j.ajhg.2012.07.004 (2012).

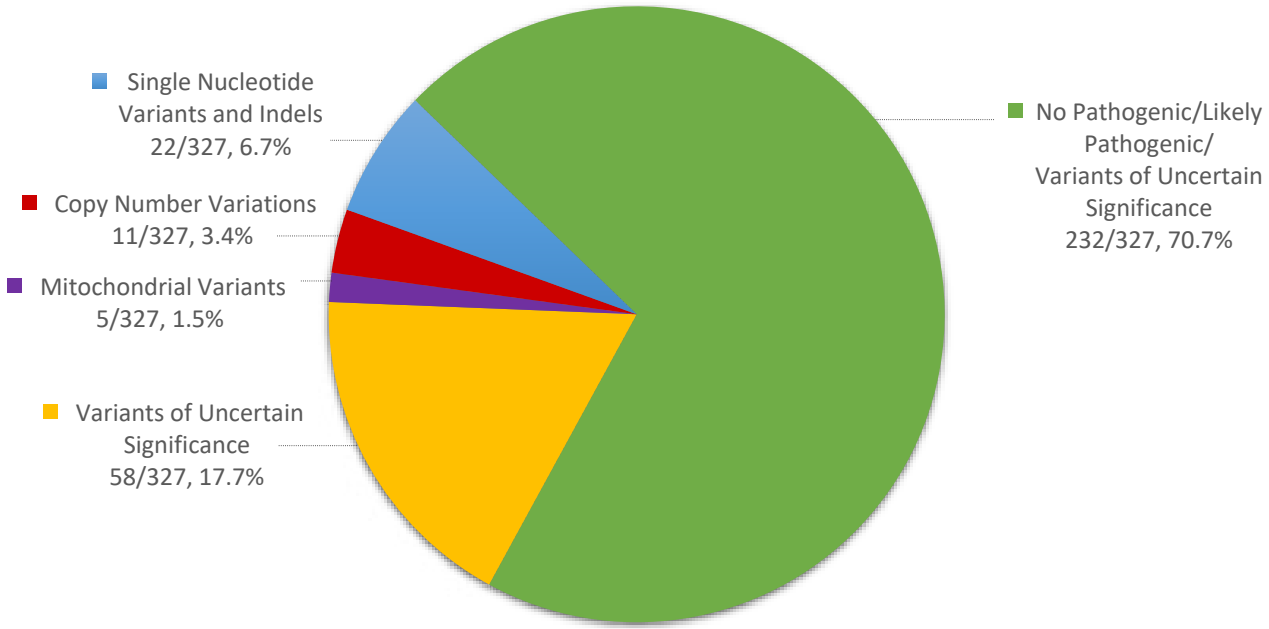
73 Trost, B. *et al.* A Comprehensive Workflow for Read Depth-Based Identification of Copy-Number Variation from Whole-Genome Sequence Data. *Am J Hum Genet* **102**, 142-155, doi:10.1016/j.ajhg.2017.12.007 (2018).

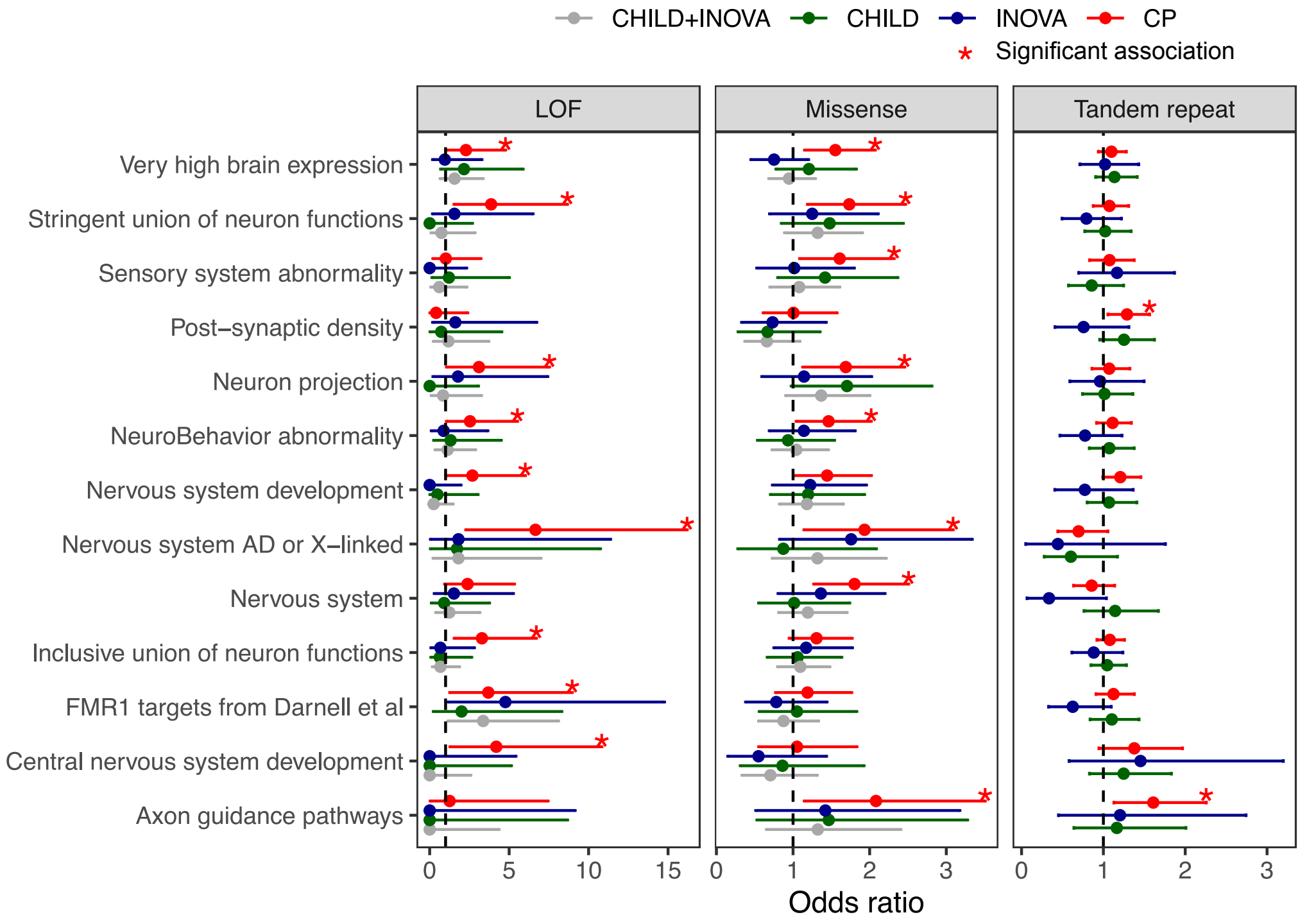
74 Chen, X. *et al.* Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220-1222, doi:10.1093/bioinformatics/btv710 (2016).

75 Rausch, T. *et al.* DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333-i339, doi:10.1093/bioinformatics/bts378 (2012).

76 Lott, M. T. *et al.* mtDNA Variation and Analysis Using Mitomap and Mitomaster. *Curr Protoc Bioinformatics* **44**, 1 23 21-26, doi:10.1002/0471250953.bi0123s44 (2013).

77 Andrews, R. M. *et al.* Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* **23**, 147, doi:10.1038/13779 (1999).
78 Dolzhenko, E. *et al.* ExpansionHunter: a sequence-graph-based tool to analyze variation in short tandem repeat regions. *Bioinformatics* **35**, 4754-4756, doi:10.1093/bioinformatics/btz431 (2019).
79 Dolzhenko, E. *et al.* ExpansionHunter Denovo: a computational method for locating known and novel repeat expansions in short-read sequencing data. *Genome Biol* **21**, 102, doi:10.1186/s13059-020-02017-z (2020).
80 Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).
81 Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443, doi:10.1038/s41586-020-2308-7 (2020).
82 Zarrei, M. *et al.* A large data resource of genomic copy number variation across neurodevelopmental disorders. *NPJ Genom Med* **4**, 26, doi:10.1038/s41525-019-0098-3 (2019).
83 Zarrei, M., MacDonald, J. R., Merico, D. & Scherer, S. W. A copy number variation map of the human genome. *Nat Rev Genet* **16**, 172-183, doi:10.1038/nrg3871 (2015).
84 MacDonald, J. R., Ziman, R., Yuen, R. K., Feuk, L. & Scherer, S. W. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res* **42**, D986-992, doi:10.1093/nar/gkt958 (2014).
85 Jaganathan, K. *et al.* Predicting Splicing from Primary Sequence with Deep Learning. *Cell* **176**, 535-548 e524, doi:10.1016/j.cell.2018.12.015 (2019).
86 Xiong, H. Y. *et al.* RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. *Science* **347**, 1254806, doi:10.1126/science.1254806 (2015).
87 Jian, X., Boerwinkle, E. & Liu, X. In silico prediction of splice-altering single nucleotide variants in the human genome. *Nucleic Acids Res* **42**, 13534-13544, doi:10.1093/nar/gku1206 (2014).
88 Landrum, M. J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* **44**, D862-868, doi:10.1093/nar/gkv1222 (2016).
89 Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**, 405-424, doi:10.1038/gim.2015.30 (2015).
90 Riggs, E. R. *et al.* Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). *Genet Med* **22**, 245-257, doi:10.1038/s41436-019-0686-8 (2020).
91 McCormick, E. M. *et al.* Specifications of the ACMG/AMP standards and guidelines for mitochondrial DNA variant interpretation. *Hum Mutat* **41**, 2028-2057, doi:10.1002/humu.24107 (2020).
92 Manshaei, R. *et al.* Genes and Pathways Implicated in Tetralogy of Fallot Revealed by Ultra-Rare Variant Burden Analysis in 231 Genome Sequences. *Front Genet* **11**, 957, doi:10.3389/fgene.2020.00957 (2020).
93 Pinese, M. *et al.* The Medical Genome Reference Bank contains whole genome and phenotype data of 2570 healthy elderly. *Nat Commun* **11**, 435, doi:10.1038/s41467-019-14079-0 (2020).
94 Schubert, E., Sander, J., Ester, M., Kriegel, H. P. & Xu, X. DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN. *ACM Trans. Database Syst.* **42**, Article 19, doi:10.1145/3068335 (2017).


Single Nucleotide Variants and Indels 22/327, 6.7%
Copy Number Variations 11/327, 3.4%
Mitochondrial Variants 5/327, 1.5%
Variants of Uncertain Significance 58/327, 17.7%
No Pathogenic/Likely Pathogenic/ Variants of Uncertain Significance 232/327, 70.7%


CHILD+INOVA
CHILD
INOVA
CP
Significant association
LOF
Missense
Tandem repeat
Very high brain expression
Stringent union of neuron functions
Sensory system abnormality
Post−synaptic density
Neuron projection
NeuroBehavior abnormality
Nervous system development
Nervous system AD or X−linked
Nervous system
Inclusive union of neuron functions
FMR1 targets from Darnell et al
Central nervous system development
Axon guidance pathways
0
5
10
15
0
1
2
3
0
1
2
3
Odds ratio