HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2022 May 29.

Published in final edited form as:

Nat Genet. 2020 October ; 52(10): 1046–1056. doi:10.1038/s41588-020-0695-1.

Mutations disrupting neuritogenesis genes confer risk for cerebral palsy

A full list of authors and affiliations appears at the end of the article.

Abstract

In addition to commonly associated environmental factors, genomic factors may cause cerebral palsy. We performed whole-exome sequencing of 250 parent–offspring trios, and observed enrichment of damaging de novo mutations in cerebral palsy cases. Eight genes had multiple damaging de novo mutations; of these, two (*TUBA1A* and *CTNNB1*) met genome-wide significance. We identified two novel monogenic etiologies, *FBXO31* and *RHOB*, and showed that the *RHOB* mutation enhances active-state Rho effector binding while the *FBXO31* mutation diminishes cyclin D levels. Candidate cerebral palsy risk genes overlapped with neurodevelopmental disorder genes. Network analyses identified enrichment of Rho GTPase, extracellular matrix, focal adhesion and cytoskeleton pathways. Cerebral palsy risk genes in enriched pathways were shown to regulate neuromotor function in a *Drosophila* reverse genetics screen. We estimate that 14% of cases could be attributed to an excess of damaging de novo or recessive variants. These findings provide evidence for genetically mediated dysregulation of early neuronal connectivity in cerebral palsy.

Cerebral palsy (CP) is the cardinal neurodevelopmental disorder (NDD) impacting motor function, affecting ~2–3 per 1,000 children worldwide[1,2]. Movement disorder (spasticity, dystonia, choreoathetosis and/or ataxia) onset occurs within the first few years of life as a manifestation of disrupted brain development[3]. Historically, although Little and Osler considered CP to occur largely as a result of perinatal anoxia[4], Freud disputed this claim[5].

Correspondence and requests for materials should be addressed to M.C.K. kruerm@email.arizona.edu.

Author contributions

K.B., S.P.-L., Q.X., C. Zhu, R.P.L., A.H.M., J.G. and M.C.K. contributed to study design, data interpretation and oversight. B.Y.N., J.G.B., K.H., C. Zhou, D.Z., B.Z., B.K., S.W., J.B., S.P., J.B.V., J.B.-H., A.P., M.C.F., L.X., Y.X., M.C., K.R., F.M., Y.W., J.L.W., L.R., J.S.C., A.F., A.E.L., J.P.P., T.F., S.J.M., K.E.C., S.M.R., D.S.R., Q.S., C.G., Y.A.W., N.B., I.N., S.C.M., X.W., D.J.A., J.H. and M.C.K. provided cohort ascertainment, recruitment and phenotypic characterization. K.B., C.C., A.E., J.L., C.L.v.E., H.M., S.M.M., I.R.T., F.L.-G., Y.A.W., B.S.G., J.Z., D.L.W., M.S.B.F., C. Zhou and M.A.C. performed exome sequencing production and validation. S.B., S.C.J., M.A.C., M.C.S., X.Z., J.R.K. and A.H.S. performed WES analysis. A.E., H.M., J.L., B.S.G. and S.P.-L. performed RHOB validation. S.M.N., S.P.-L., S.P., J.B.V., D.D. and S.A.L. performed FBXO31 validation. S.A.L., S.V. and D.C.Z. performed *Drosophila* locomotor experiments. S.C.J., S.A.L., S.B., S.S., B.L., Q.L., M.C.S. and X.Z. conducted statistical analysis. S.H. performed biophysical simulation for RHOB and FBXO31. S.C.J., S.A.L., J.G., Q.L., S.P.-L., R.P.L., A.H.M., S.M., B.Y.N., M.C.S., X.Z., C.L.v.E., X.W., Q.X., C. Zhu and M.C.K. wrote and reviewed the manuscript. K.B., R.P.L., Q.X., C. Zhu, A.H.M., J.G., S.P.-L. and M.C.K. acquired funding and supervised the project and were considered co-senior authors. All authors have read and approved the final manuscript.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41588-020-0695-1.

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-020-0695-1.

Competing interests

The authors declare no competing interests.

To this day, debate about the origin of CP continues, particularly in individual cases, with widespread medical and legal implications[6,7].

As for other NDDs, such as autism spectrum disorders (ASDs) and intellectual disability (ID), no single causative factor has been implicated in CP, although several environmental factors, including prematurity, infection, hypoxia–ischemia, and pre- and perinatal stroke, are major contributors to CP risk[8]. However, as many as ~40% of CP cases may not have a readily identifiable etiology[9], defined as cryptogenic or idiopathic CP[10]. Registry-based data have shown that 21–40% of CP cases have an associated congenital anomaly, implicating genomic alterations in many of these cases[11]. A heritability of 40% has been estimated in CP[12], supported by probabilistic modeling of CP etiology in a western Swedish cohort[13], comparable to the heritability of 38–58% estimated for ASD[14,15].

To date, five studies have analyzed genomic copy number variations (CNVs) in CP cases[10,16–19], identifying predicted deleterious CNVs in 10–31% of cases. Three previous whole-exome sequencing (WES) studies have been performed in CP cases[20–22]. The largest study to date reported putatively deleterious variants in ~14% of 98 parent–offspring trios with unselected forms of CP[22]. These studies indicate potentially important genetic risks in CP, but insufficient availability of controls limited the statistical inferences that could be made, and functional validation of novel candidate gene variants was not performed. We sought to address these limitations in the current study.

Results

CP cohort characteristics and WES.

We performed WES of 250 CP trios, including 91 previously reported[22] and 159 ascertained from centers in the United States, China and Australia after written informed consent was obtained according to local ethical requirements (Methods). Cases were diagnosed by clinical specialists using international consensus criteria[23] (Supplementary Table 1 and Supplementary Dataset 1); CP was thus defined as a non-progressive developmental disorder of movement and/or posture impairing motor function. Cases experienced symptom onset by age two. This operational definition thus excluded progressive neurological disorders such as neurodegenerative diseases. No cases had known chromosomal anomalies or aneuploidies, clinically or molecularly diagnosed syndromes (that is, Rett syndrome, Angelman syndrome and so on), pathogenic microdeletion or microduplication syndromes, mitochondrial disorders or traumatic brain injuries.

Detailed patient phenotypes are available in the Supplementary Note. Representative neuroimaging findings are presented in Extended Data Fig. 1, and videos highlighting movement disorder phenotypes in representative individuals can be found in Supplementary Videos (43 videos available via https://figshare.com/s/a4f914ab77958ab3e4b6) and in Supplementary Photos (https://figshare.com/s/0f200402e51de5875390). Within our 250 family cohort, 157 trios (62.8%) were classified as idiopathic (no known cause), 84 cases (33.6%) had a known environmental insult associated with CP (including prematurity defined as <32 weeks of gestation, perinatal hypoxia–ischemia (as defined by treating

clinicians), ischemic/hemorrhagic stroke and/or infection) and the remaining 9 trios (3.6%) were not able to be assigned to either category ('unclassified'; Supplementary Table 1).

WES was performed as previously described[24] (see Supplementary Table 2 for exome metrics). Control trios consisting of 1,789 unaffected siblings of autism cases and their unaffected parents from the Simons Simplex Collection were analyzed in parallel[25]. BWA-MEM was used to align the sequencing reads, and GATK Best Practices was used to call variants[26,27]. MetaSVM[28] and Combined Annotation Dependent Depletion (CADD v1.3)[29] algorithms were used to predict deleteriousness of missense variants (D-Mis, defined as Meta-SVM-deleterious or CADD ≥ 20). Inferred loss-of-function (LoF) variants consist of stop-gain, stop-loss, frameshift insertions/ deletions, canonical splice sites and start-loss. LoF and D-Mis mutations were considered 'damaging'. De novo mutations (DNMs) were called by the TrioDeNovo program[30]. Sanger sequencing was conducted to validate mutations in genes of interest.

Damaging DNMs are significantly enriched in the CP cohort.

We began by assessing the contribution of DNMs to CP at a cohort level. The number of observed DNMs in cases and controls closely approximates the Poisson distribution (Extended Data Fig. 2), indicating that DNMs are independent probabilistic events. We found an enrichment of damaging DNMs in CP cases, which became more apparent when focusing the analysis on genes intolerant to LoF variation (pLI score ≥ 0.9 in gnomAD v2.1.1 (ref. [31])) (enrichment = 1.78; $P = 1.2 \times 10^{-5}$ for damaging DNMs; Table 1). No significant enrichment of any mutation category was found in controls (Table 1). When we considered the ascertainment differential (the observed number of damaging DNMs versus the expected number of damaging DNMs, divided by the number of trios in the cohort), 11.9% of CP cases in our cohort could be attributed to an excess of damaging DNMs. When stratifying cases by CP subtype, we found greater enrichment of damaging DNMs in idiopathic (enrichment = 1.98; $P = 2.1 \times 10^{-5}$) compared to environmental cases (enrichment = 1.28; $P = 0.19$; Supplementary Table 3), suggesting that idiopathic cases harbor a higher burden of damaging DNMs.

Recurrent damaging DNMs implicate both known and novel CP genes.

We next considered individual genes recurrently implicated in our CP cohort via a de novo mechanism (Supplementary Dataset 2). We identified eight genes harboring ≥2 damaging DNMs, with *TUBA1A* ($P = 4.8 \times 10^{-8}$) and *CTNNB1* ($P = 9.8 \times 10^{-10}$) surpassing Bonferroni correction cutoffs for genome-wide significance (Table 2 and Supplementary Table 4). The gene-level enrichment of protein-damaging DNMs in these genes we observed strongly implicates these genes as bona fide CP-associated genes (Supplementary Table 5). Among these eight genes, *ATL1*, *CTNNB1*, *SPAST* and *TUBA1A* have previously been associated with human CP phenotypes[20,22,32]. We also identified identical but independently arising damaging DNMs in two genes, *RHOB* and *FBXO31*.

Identical gain-of-function DNMs in *RHOB* and *FBXO31*.

RHOB, encoding a Rho GTPase, harbored two identical DNMs (encoding p.Ser73Phe; Fig. 1a and Supplementary Table 4) in two unrelated spastic–dystonic CP cases, representing

an unlikely chance event ($P = 1.6 \times 10^{-3}$; Supplementary Note). Ser73 is predicted to be phosphorylated (0.997 by NetPhos 3.1)[33] and located in a conserved position in the Switch II domain, where Rho protein kinases associate with Rho- and Rac-related proteins (Fig. 1b). Comparing structural models of RHOB wild type and p.Ser73Phe suggests an alteration of both the shape of the binding site and the surface charge of the protein (Fig. 1b). Both patients have a remarkably concordant phenotype, including a hyperintense T2 white matter signal (periventricular leukomalacia) on magnetic resonance imaging (MRI), spastic–dystonic diplegia, expressive language disorder and aortic arch abnormalities (Fig. 1c, Supplementary Table 4 and Supplementary Videos F064 and F244). RHOB is known to control dendritic spine outgrowth[34] but has not previously been associated with a human disease. Biochemical analyses indicated that this variant shows accentuated responses to both GTPase-activating proteins (GAPs) and GDP exchange factors (GEFs; Fig. 1d,e), ultimately leading to enhanced binding in the active state to the Rho effector rhotekin (Fig. 1f).

We also identified two unrelated cases with an identical DNM (encoding p.Asp334Asn; Fig. 2a and Supplementary Table 4) in *FBXO31*, which encodes the F-box only protein 31. An FBXO31/SKP1/Cullin1 complex ubiquitinates targets such as cyclin D to control protein abundance by tagging them for proteasomal degradation[35]. Asp334 is a conserved residue within the binding pocket on FBXO31 (Fig. 2b), where it is thought to mediate hydrogen bonding to cyclin D1 (ref. [36]). FBXO31 is known to control axonal outgrowth and is essential for dendrite growth and neuronal migration in the developing brain[37]. FBXO31 p.Asp334Asn affects the cyclin D interaction site[36] (Fig. 2b), leading to an apparent gain of function of cyclin D degradation (Fig. 2c). A homozygous truncating mutation in *FBXO31* has previously been reported in association with ID (MIM 615979)[38]. Both patients in our cohort exhibited spastic diplegic CP (Supplementary Table 4 and Supplementary Videos F218 and F699), ID, expressive language disorder and attention-deficit/hyperactivity disorder. F218 had gut malrotation and constipation, cleft palate, strabismus and normal brain morphology on MRI, while F699 had strabismus, severe constipation and ventricular dilation with thin corpus callosum on MRI. Therefore, this DNM in *FBXO31* leads to a phenotype distinct from the previously described autosomal recessive truncating mutation-associated non-syndromic ID phenotype[38].

DNMs in previously implicated genes *TUBA1A*, *CTNNB1*, *ATL1* and *SPAST*.

TUBA1A, encoding the microtubule-related protein α-tubulin, harbors three damaging DNMs (encoding p.Arg123Cys, p.Leu152Gln and p.Tyr408Asp; Supplementary Table 4) in three unrelated probands, two of whom have previously been reported[22]. Both p.Arg123Cys and p.Leu152Gln map to the tubulin nucleotide-binding domain-like domain, and p.Tyr408Asp maps to the carboxy-terminal stabilization domain[39] (Extended Data Fig. 3). *TUBA1A* heterozygous mutations have been described as being associated with a spectrum of cortical malformations[40] (MIM 611603), and our patients exhibit MRI findings within this spectrum (Extended Data Fig. 3). Clinically, our cases demonstrate spasticity in their lower limbs, and two out of three exhibit concurrent ID.

CTNNB1, encoding β-catenin, harbors three LoF DNMs (encoding p.Glu54*, p.Phe99PhefsTer5 and p.Arg449GlnfsTer24; Supplementary Table 4) in three unrelated probands, one of whom was previously reported[21]. p.Glu54* and p.Phe99fs are located in the amino-terminal domain and predicted to lead to nonsense-mediated decay, while p.Arg449fs is located in the central armadillo repeat domain, which is essential for the phosphorylation of β-catenin by protein kinase CK2 (ref. [41]) (Extended Data Fig. 4). Autosomal dominant germline inactivating mutations in *CTNNB1* have been implicated in exudative vitreoretinopathy 7 (ref. [42]) (MIM 617572) and NDD with spastic diplegia and visual defects[43–45] (MIM 615075). All of our patients exhibited spasticity, ID, behavior problems and language disorders. We also found dystonia and microcephaly in two out of three patients. While one patient had possible bilateral frontal pachygyria, brain findings were notably absent from the other patients (Extended Data Fig. 4). We found strabismus in two out of three patients, but no other visual defects.

ATL1 encodes atlastin-1, which is critical for the formation of the tubular endoplasmic reticulum network and axon elongation in neurons[46–48]. *ATL1* harbors two damaging DNMs in our cohort (encoding p.Ala350Val and p.Lys406Gln; Supplementary Table 4) located in the GBP domain (Extended Data Fig. 5). Autosomal dominant germline mutations have been associated with neuropathy type 1D[49] (MIM 613708) and spastic paraplegia type 3A[50] (MIM 182600). Our patients exhibited spasticity and dystonia with brain findings of T2 hyperintensities and bihemispheric periventricular leukomalacia (Extended Data Fig. 5). There was no evidence of phenotypic progression at the time of last follow-up (patient ages 10 years and 29 months).

SPAST, encoding spastin, harbors two damaging DNMs (encoding p.Asp441Gly and p.Ala495Pro; Supplementary Table 4). Both mutations occur at conserved positions in the AAA domain, which is essential for the regulation of ATPase activity (Extended Data Fig. 6). Autosomal dominant germline mutations in *SPAST* have been linked to spastic paraplegia 4 (ref. [51]; MIM 182601). p.Asp441Gly has been reported in association with hereditary spastic paraplegia (HSP)[52,53]. Our patients exhibited spasticity with one also exhibiting dystonia, with scattered subcortical T2 hyperintensities present in one patient and no brain findings in the other (Extended Data Fig. 6). There was no evidence of phenotypic progression (patient ages 21 years and 40 months, respectively).

DNMs in *DHX32* and *ALK*.

DHX32, encoding putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX32, harbored two damaging DNMs (encoding p.Tyr228Cys and p.Ile266Met; Supplementary Table 4). p.Tyr228Cys falls within the helicase ATP-binding domain, which is required for ATP binding, hydrolysis and nucleic acid substrate binding[54] (Extended Data Fig. 7). Mutations in *DHX32* have not previously been associated with human diseases. Both of our patients exhibited ID, and one demonstrated spastic diplegia, with the other characterized as a generalized dystonia. Brain findings included periventricular leukomalacia and mildly diminished cerebral volume (Extended Data Fig. 7).

ALK, encoding ALK receptor tyrosine kinase, harbored one damaging DNM (encoding p.Ser1081Arg) and one stop-gain DNM (encoding p.Trp1320*; Supplementary Table 4).

p.Trp1320* is located in the tyrosine kinase domain[55] and p.Ser1081Arg is located just upstream in the juxtamembrane domain (Extended Data Fig. 8). Germline and somatic activating mutations in *ALK* have previously been associated with neuroblastoma[56,57] (MIM 613014). One patient exhibited spastic diplegia with mild tremor, scattered subcortical hyperintensities (Extended Data Fig. 8) and an atrial septal defect. The other patient had spastic–dystonic diplegia, white matter abnormalities and epilepsy. There was no evidence of neuroblastoma in either patient.

Enriched recessive genotypes in genes associated with HSP.

We performed a one-tailed binomial test coupled with a polynomial model[24] to evaluate the burden of recessive genotypes (RGs) for each gene in our CP cohort (Supplementary Dataset 3). We did not observe enrichment of damaging RGs in the cohort meeting genome-wide significance (Supplementary Table 6). However, we noted biallelic damaging variants in several genes previously associated with HSP. HSP is clinically distinguished from CP by its progressive, neurodegenerative nature and later (often adult) onset in many cases.

We carefully reassessed the clinical phenotypes of these cases and found no evidence of progression from the time of ascertainment. Interestingly, early onset with protracted clinical stability has previously been identified as an endophenotype in a subset of patients with mutations in HSP-associated genes[58]. For example, patients with *SPAST* missense mutations (as our cases had) may have onset in toddlerhood with extended clinical stability[59] consistent with a CP phenotype. In contrast, truncating *SPAST* mutations are often translated and accumulate over time, putatively leading to later onset and a neurodegenerative course[60]. In addition, important roles for SPAST[61] and ATL1 (ref. [62]) in developmental neuritogenesis have been shown, indicating their importance in neuronal development.

We observed six damaging RGs (in *AMPD2*, *AP4M1*, *AP5Z1, FARS2*, *NT5C2* and *SPG11*; Supplementary Table 7) among genes previously associated with recessive HSP (Supplementary Dataset 4; enrichment = 7.74; one-tailed binomial $P = 1.5 \times 10^{-4}$; Table 3). By ascertainment differential, ~2.1% of the CP cases in our cohort could thus be accounted for by an excess of RGs. The enrichment of RGs in known HSP-associated genes was predominantly driven by idiopathic cases (idiopathic enrichment = 9.22; one-tailed binomial $P = 2.4 \times 10^{-4}$ versus environmental enrichment = 4.48; one-tailed binomial $P = 0.20$; Table 3).

No gene was enriched for rare X-linked hemizygous variants.

Male sex is a risk factor for developing CP[63]. Therefore, we compared rare hemizygous variants (minor allele frequency (MAF) $\leq 5.0 \times 10^{-5}$) in 154 male CP probands to male controls in gnomAD. No gene surpassed the Bonferroni correction cutoff (Supplementary Table 8), suggesting that the current study is statistically underpowered to assess hemizygous burden.

Clinical and genetic overlap of CP with other NDDs.

Clinically, NDDs frequently co-occur. In the case of CP, ~45% of individuals with CP have concurrent ID[64], ~40% also have epilepsy, and ~7% have ASD in addition to CP[1].

Accordingly, we sought to determine the degree of overlap between genes harboring rare damaging variants with de novo, X-linked recessive or autosomal recessive segregation (putative CP risk genes; $n = 439$, Supplementary Datasets 6–15) from our CP cohort with known NDD risk genes. The analysis was performed using the disease–gene network tool DisGeNET, which identifies associations between genes and diseases curated from the literature and databases including ClinVar, ClinGen and UniProt[65]. We found substantial genetic overlap between our CP candidate gene list and the major NDDs (CP versus ID, enrichment = 2.0, $P = 2.56 \times 10^{-16}$; CP versus epilepsy, enrichment = 1.7, $P = 1.6 \times 10^{-4}$; CP versus ASD, enrichment = 2.0, $P = 1.2 \times 10^{-5}$; hypergeometric two-tailed test; Fig. 3a). In contrast, when we examined overlap with a neurodegenerative disorder, Alzheimer's disease, there was no enrichment (Fig. 3b). A total of 28.9% of CP risk genes overlapped with genes linked to ID, 11.1% for epilepsy and 6.3% for ASD. Our data suggest that CP has significant genetic overlap with other genetic NDDs, indicating potential genetic pleiotropy and common etiologies of co-occurring NDDs.

Extracellular matrix, cell–matrix focal adhesions, the cytoskeletal network and Rho GTPase genes are highly associated with CP.

We identified a large number of individual genes harboring predicted damaging variants and employed a suite of tools for unbiased discovery of conserved pathways and biological functions relevant to CP. STRING-based clustering[66] of the 439 putative CP risk genes (Supplementary Datasets 6–15) showed greater connectivity than predicted by chance (enrichment = 1.2, $P = 1.51 \times 10^{-4}$), indicating a functional network encompassing damaging variants. We then performed gene over-representation analysis[67,68] of these genes using DAVID[69], MSigDB[70] and PANTHER[71] for functional annotation and pathway characterization. This approach indicated statistical over-representation of candidate genes stratified by Gene Ontology (GO) and pathways (KEGG/Reactome), and curated functional and expression data to identify meaningful relationships. Consistent with the STRING findings, this approach identified multiple gene sets representing enriched pathways (false discovery rate (FDR) < 0.05) and conserved functions (Supplementary Datasets 6–15).

We noted functionally related findings supported by multiple tools, including non-integrin membrane–extracellular matrix (ECM) interactions and laminin interaction pathways identified by all three algorithms. We then inferred hierarchical associations among ontological terms using dcGO[72] (Table 4). Taken together, these findings indicate an over-representation of genes involved in ECM biology, cell–matrix interactions (focal adhesions), cytoskeletal dynamics and Rho GTPase function.

Genes from Rho GTPase, cytoskeleton and cell projection pathways govern neuromotor development in *Drosophila*.

Subsequently, we independently assessed the role for over-represented pathway members in normal locomotor development by conducting a reverse genetic screen in *Drosophila*. A similar approach has been applied previously in studies of ASD and HSP using *Drosophila* and zebrafish, respectively[73,74]. We focused on genes with damaging variants from our cohort of patients with CP with GTPase, cytoskeleton and cell projection GO terms. We

hypothesized that our screen could newly indicate a key role for these genes in neuromotor development.

We selected genes with conserved *Drosophila* orthologs (DIOPT ≥ 5) that had available molecularly characterized alleles (complete results and genotypes in Supplementary Table 9). We utilized hypomorphic/LoF alleles in a biallelic state to help map phenotypes to the gene of interest in *Drosophila* assays. We excluded genes that would cause confounding phenotypes such as lethality or had a previously described locomotor phenotype, except for *ATL1*, which was included as a positive control. Genes with known roles in brain development or NDDs were prioritized. Two genes with variants that did not meet the filtering criteria for deleteriousness were included as negative controls. Altogether, we screened 22 genes for locomotor ability using turning assays in larvae[75] and negative geotaxis/positive phototaxis assays in adults[76,77].

We found locomotor phenotypes in mutants of gene orthologs encoding regulators of GTPase signal transduction (*AGAP1*, *DOCK11*, *RABEP1*, *SYNGAP1* and *TBC1D17*), the cytoskeleton (*MKL1* and *MPP1*) and cell projection (*PTK2B*, *SEMA4A* and *TENM1*) pathways (Fig. 4). When assays were conducted in both larvae and adults, we often found locomotor phenotypes at both time points, suggesting that defects arose in the developmental period and persisted throughout the lifespan (Supplementary Table 9). Of potential interest, we found evidence for sexual dimorphism, as male flies with mutations in orthologs of *AKT3*, *RABEP1* or *PRICKLE1*/*2* exhibited locomotor deficits while females did not.

In total, we found 71% (10/14) of the genes from our enriched pathways exhibited a locomotor phenotype in *Drosophila* (Fig. 4 and Extended Data Fig. 9). In comparison, genome wide, only 3.1% of annotated *Drosophila* genes are known to lead to a locomotor phenotype[78] (enrichment = 23.4, $P = 2.2 \times 10^{-16}$; Fig. 4). Overall, our *Drosophila* studies supported a role for candidate CP genes in the cytoskeletal, Rho GTPase and cell projection pathways in motor development.

Discussion

In the past, damaging genomic variants have not been considered to be a major contributor to CP, but our findings and those of others challenge this dogma. Previous studies suggested that both CNVs and single-nucleotide variants contribute to CP[10,16–22]. Here we expand on those earlier findings and provide robust statistical evidence at a cohort level that rare, damaging single-nucleotide variants represent an independent risk factor for CP. The cohort-wide enrichment of DNMs we detected is consistent with the observation that most cases of CP occur sporadically[79]. Using the distribution of LoF-intolerant genes with multiple damaging DNMs in this cohort, we estimated the number of genes that contribute to CP through a de novo mechanism to be 75 (95% confidence interval = 26.5–123.5; Extended Data Fig. 10a and Supplementary Note). Saturation analysis estimates that WES of 2,500 and 7,500 CP trios will yield 65.3% and 91.8% saturation, respectively, for CP risk genes with DNMs, suggesting a high yield for CP gene discovery as additional samples are sequenced (Extended Data Fig. 10b). Accordingly, the International Cerebral Palsy

Genomics Consortium (ICPGC; https://www.icpgc.org) was recently founded to address the need for international data sharing and collaboration to advance the pace of discovery[80]. Conservatively, we estimate that 14% of the cases in our cohort can be accounted for by damaging genomic variants (based on ascertainment differentials of 11.9% for DNMs and ~2% for RGs). In comparison, recent estimates indicate that acute intrapartum hypoxia–ischemia is seen in ~6% of CP cases[81], indicating that genomic mutations represent an important, independent contributor to CP etiology that historically has been overlooked.

We found evidence for both known disease-associated genes and genes not previously associated with human phenotypes in our cohort. The identification of independently arising yet identical DNMs in *RHOB* and *FBXO31* indicates that monogenic contributions to CP exist but may be under-recognized. Our parallel identification of genetic correlation of CP with other NDDs implicates shared susceptibility, as suggested previously[82]. In some cases, this may reflect ascertainment bias, as motor phenotypes may have been under-reported in previous studies of other NDDs. In other cases, typified by *FBXO31*, our findings likely represent phenotypic expansions. Finally, in some contexts, NDD manifestations may prove pleiotropic, with a genetic disruption of early neurodevelopment manifesting variably, as is increasingly being recognized[83]. As for other NDDs, individual CP cases may prove to be environmental in origin, genetic, or some combination thereof. However, uniquely among the NDDs, environmental contributions to CP are relatively well characterized, and CP may represent a model disorder within which to study gene–environment interactions in a developmental context.

Altered motor circuit connectivity is thought to be part of CP pathophysiology[84]. By integrating orthogonal lines of evidence, including recurrent gene analyses, in vitro and in vivo functional assays, cohort-wide network biology approaches and *Drosophila* locomotor studies, we found converging evidence supporting a role for ECM components, cell–matrix focal adhesions, cytoskeletal organization and Rho GTPases in CP etiology. These processes are known to drive the conserved process of cell projection extensions during nervous system development[85]. On the basis of known disease and developmental biology, we therefore predict that disruption of genes involved in neurodevelopmental patterning may alter early neuritogenesis and neuronal functional network connectivity in CP. Further studies will be needed to determine more specifically how variants identified in patients with CP affect neuronal circuit development.

Our findings have important clinical implications. Specific genetic findings may provide closure for families and guide preventative healthcare as well as family planning, such as counseling for recurrence risk (often quoted as ~1% for CP but potentially much higher for inherited mutations). In some cases, identification of specific variants in individuals in our cohort led to recommendations for changes in management, including personalized treatments that would not otherwise have been initiated (that is, ethosuximide for *GNB*[86] (F068), levodopa for *CTNNB1* (*ref.* [87]) (F066, GRA8913, F428) and 5-aminoimidazole-4-carboxamide riboside (AICAr) for *AMPD*[88] (F623) (Supplementary Note).

In the near future, studies will be able to overcome our limitations of small sample size and further utilize available clinical data to expand on genotype–phenotype correlations.

Additionally, as more information about CP genetic etiology becomes available, it will become possible to assign likely genetic causation to more individual cases. Future studies of well-characterized unselected CP cohorts will be instrumental in determining the true contributions of genetic and environmental factors side by side to clarify the epi demiology of CP.

Overall, our data indicate that genomic variants should be considered alongside environmental insults when assessing the etiology of an individual's CP. Such considerations will have important clinical, research and medico-legal implications. In the near future, genomic data may help stratify patients and identify likely responders to currently available medical and/or surgical therapies. Finally, over time, mechanistic insights derived from the identification of core pathways via genomic studies of CP may help guide therapeutic development efforts in a field that has not seen a novel therapy introduced for decades.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-020-0695-1.

Methods

Case cohorts, enrollment, phenotyping and exclusion criteria.

A total of 159 CP cases (132 idiopathic, 24 environmental and 3 unclassified) and their unaffected parents were recruited via Phoenix Children's Hospital (PCH), the University of Adelaide and Zhengzhou City Children's Hospital. Six of these were recently published as part of a gene panel study[89]. Exclusion criteria and detailed descriptions about these cohorts are provided separately below. Further, 91 previously published[22] trios (25 idiopathic, 60 environmental, 6 unknown) were included to allow for comparison of idiopathic and environmental subtypes of CP.

CP classification.

CP cases were subdivided into idiopathic, environmental and unclassified groups on the basis of data available at the time of ascertainment. This designation was revised as appropriate as additional data became available. Cases were designated 'environmental' if any idiopathic exclusion criteria were met.

Exclusion criteria for idiopathic status.

Potential participants were excluded from an 'idiopathic' designation if any of the following were present: prematurity (estimated gestational age <32 weeks), stroke, intraventricular hemorrhage, major brain malformation (that is, lissencephaly, pachygyria, polymicrogyria, schizencephaly, simplified gyri, brainstem dysgenesis, cerebellar hypoplasia and so on), hypoxic–ischemic injury (as defined by treating physicians), in utero infection, hydrocephalus, traumatic brain injury, respiratory arrest, cardiac arrest or brain calcifications. The following did not automatically indicate environmental status even if

parents believed this was the cause of the child's CP: history of prematurity (but delivery at greater than or equal to 32 weeks of gestational age), nuchal cord, difficult delivery, fetal decelerations, urgent C-section, preterm bleeding or maternal infection. In equivocal cases, additional data were sought until a decision regarding group assignment could be made by the corresponding author. Periventricular leukomalacia was not considered universally indicative of environmental status[90].

Movement disorder, pattern of involvement and functional status.

Spasticity, dystonia, chorea/athetosis, ballism, hypotonia and/or ataxia were assessed by the treating specialist, who also assigned Gross Motor Functional Classification System scores as well as the pattern of involvement.

PCH (n = 52).—Patients with CP diagnosed according to international consensus criteria[23] were recruited from CP subspecialty clinics (pediatric movement disorders neurology, pediatric orthopedics, pediatric neurosurgery and pediatric physiatry) at PCH or the clinics of collaborators at outside institutions using a local ethics-approved protocol or a PCH-approved central IRB protocol (no. 15–080). Written informed consent was obtained for parents and assent was obtained for children as appropriate for families wishing to participate. Blood, buccal swab and/or saliva samples were collected from the affected child and both parents. DNA was extracted with the support of the PCH Biorepository using a Kingfisher Automated Extraction System, and quality control metrics, including yield, 260/280 and 260/230 ratio, were recorded.

University of Adelaide Robinson Research Institute (n = 63).—Ethics permission was obtained in each state and overall from the Adelaide Women's and Children's Health Network Human Research Ethics Committee South Australia. Families were enrolled from among children attending major children's hospitals in South Australia, New South Wales and Queensland where a diagnosis of CP had been confirmed by a specialist in pediatric rehabilitation according to international consensus criteria[23]. Blood for DNA from cases was collected under general anesthesia during procedures such as Botox injections or orthopedic surgery and parental blood was collected whenever possible. Lymphoblastoid cell lines were generated for each case at Genetic Repositories Australia.

Zhengzhou City Children's Hospital (n = 44).—This study was approved after review by the ethics committee of Zhengzhou City Children's Hospital. Parent–offspring trios were recruited from children with CP without apparent cause at Zhengzhou City Children's Hospital. Cases were additionally excluded if intrauterine growth retardation, threatened preterm birth, premature rupture of membranes, pregnancy-induced hypertension or multiple births was present. All participants and their guardians provided written informed consent under the auspices of the local ethics board. DNA was extracted from blood samples using standard methods.

Control cohorts.

The controls consisted of 1,789 previously sequenced families that included one child with autism, one unaffected sibling and the unaffected parents[25]. For use in this study, only the

unaffected sibling and parents were analyzed. Controls were designated as unaffected by the Simons Simplex Collection. Permission to access the genomic data in the Simons Simplex Collection via the National Institute of Mental Health Data Repository was obtained. Written informed consent for all participants was provided by the Simons Foundation Autism Research Initiative.

Exome sequencing.

Most trios were sequenced at the Yale Center for Genome Analysis following an identical protocol (Supplementary Table 2). Briefly, genomic DNA from venous blood, buccal swabs, saliva or lymphoblastoid cell lines (Adelaide) was captured using the Nimblegen SeqxCap EZ MedExome Target Enrichment Kit (Roche) or the xGEN Exome Research Panel v1.0 (IDT) followed by Illumina DNA sequencing as previously described[24]. Trio samples from Zhengzhou were prepared using Exome Library Prep kits (Illumina), followed by Illumina sequencing. Eight trios from Adelaide sequenced at the University of Washington were prepared using the SureSelect Human All Exon V5 (Agilent) and underwent Illumina sequencing. One trio sequenced by GeneDx was captured using the Agilent SureSelect Human All Exon V4 while one trio sequenced by the Hôpital Pitié-Salpêtrière used the Roche MedExome capture kit, in both cases followed by Illumina sequencing. Ninety-one previously published trios from Adelaide were captured using the VCRome 2.1 kit (HGSC), followed by Illumina sequencing as described previously[22] (Supplementary Dataset 1). Sequencing metrics suggest that, regardless of the exome capture reagent used, all samples had sufficient sequencing coverage to make confident variant calls with a mean coverage of ≥46× at each targeted base and more than 90% of targeted bases with ≥8 independent reads.

Mapping and variant calling.

WES data were processed using two independent pipelines at the Yale School of Medicine and PCH. At each site, sequence reads were independently mapped to the reference genome (GRCh37) with BWA-MEM and further processed using GATK Best Practice workflows, which include duplication marking, indel realignment and base quality recalibration, as previously described[26,27,91]. Single-nucleotide variants and small indels were called with GATK HaplotypeCaller and annotated using ANNOVAR[92], dbSNP (v138), 1000 Genomes (August 2015), NHLBI Exome Variant Server (EVS) and the Exome Aggregation Consortium v3 (ExAC)[93]. MetaSVM and CADD (v1.3) algorithms were used to predict deleteriousness of missense variants (D-Mis, defined as MetaSVM-deleterious or CADD ≥ 20)[28,29]. Inferred LoF variants consist of stop-gain, stop-loss, frameshift insertions/deletions, canonical splice sites and start-loss. LoF + D-Mis mutations were considered ‘damaging’. Variant calls were reconciled between Yale and PCH before downstream statistical analyses. Variants were considered by mode of inheritance, including DNMs, RGs and X-linked variants. Protein annotations in Extended Data Figs. 3–8 were obtained using Geneious Prime 2020.0.5 (https://www.geneious.com).

Variant filtering.

DNMs were called using the TrioDenovo[30] program by Yale and PCH separately as described previously[24], and filtered using stringent hard cutoffs. These hard filters include: MAF ≤ 4×10^{-4} in ExAC; a minimum of 10 total reads, 5 alternate allele reads, and a

minimum 20% alternate allele ratio in the proband if alternate allele reads ≥10 or, if alternate allele reads were <10, a minimum 28% alternate ratio; a minimum depth of 10 reference reads and alternate allele ratio <3.5% in parents; and exonic or canonical splice-site variants.

For the X-linked hemizygous variants, we filtered for rarity (MAF ≤ 5×10^{-5} across all samples in 1000 Genomes, EVS and ExAC) and high-quality heterozygotes (pass GATK variant score quality recalibration, a minimum of 8 total reads, genotype quality score ≥20, mapping quality score ≥40, and a minimum 20% alternate allele ratio in the proband if alternate allele reads ≥10 or, if alternate allele reads were <10, a minimum 28% alternate ratio)[93,94]. Additionally, variants located in segmental duplication regions (as annotated by ANNOVAR[28]), RGs and DNMs were excluded. Finally, in silico visualization was performed on variants that appear at least twice and variants in the top 20 significant genes from the analysis.

We filtered RGs for rare (MAF ≤ 10^{-3} across all samples in 1000 Genomes, EVS and ExAC) homozygous and compound heterozygous variants that exhibited high-quality sequence reads (pass GATK variant score quality recalibration) and had a minimum of 8 total reads for the proband. Only LoF variants (stop-gain, stop-loss, canonical splice-site, frameshift indels and start-loss), D-Mis (MetaSVM = D or CADD ≥ 20) and non-frameshift indels were considered potentially damaging to protein function.

Estimation of expected number of RGs.

We implemented a multivariate regression model to quantify the enrichment of damaging RGs in a specific gene or gene set in cases, independent of controls. Additional details about the modeling of the distribution of RG counts are described in our recent study[24].

Statistical analysis.

De novo enrichment analysis.—The R package denovolyzeR was used for the analysis of DNMs based on a mutation model developed previously[95]. The probability of observing a DNM in each gene was derived as described previously[96], except that the coverage adjustment factor was based on the full set of 250 case trios or 1,789 control trios (separate probability tables for each cohort). The overall enrichment was calculated by comparing the observed number of DNMs across each functional class to that expected under the null mutation model. The expected number of DNMs was calculated by taking the sum of each functional class specific probability multiplied by the number of probands in the study, multiplied by two (diploid genomes). The Poisson test was then used to test for enrichment of observed DNMs versus expected as implemented in denovolyzeR[95]. For gene-set enrichment, the expected probability was calculated from the probabilities corresponding to the gene set alone.

To estimate the number of genes with >1 DNM, 1 million permutations were performed to derive the empirical distribution of the number of genes with multiple DNMs. For each permutation, the number of DNMs observed in each functional class was randomly distributed across the genome adjusting for gene mutability[24]. The empirical *P* value

was calculated as the proportion of times that the number of recurrent genes from the permutation is greater than or equal to the observed number of recurrent genes.

To examine whether any individual gene contains more DNMs than expected, the expected number of DNMs for each functional class was calculated from the corresponding probability adjusting for cohort size. A one-tailed Poisson test was then used to compare the observed DNMs for each gene versus the expected. As separate tests were performed for damaging DNMs and LoF DNMs, the Bonferroni multiple-testing threshold is, therefore, equal to 1.3×10^{-6} (0.05/(19,347 genes × 2 tests)). The most significant P value of the two tests was reported.

Gene-set enrichment analysis.—To test for over-representation of damaging RGs in a gene set without controls and correct for consanguinity, a one-sided binomial test coupled with the polynomial regression model was conducted by comparing the observed number of variants to the expected count estimated as described before[24]. Assuming that our exome capture reagent captures N genes and the testing gene set contains M genes, then the P value of finding k variants in this gene set out of a total of x variants in the entire exome is given by

$$P = \sum_{i=k}^{x} \binom{x}{i} (p)^i (1-p)^{n-i}$$

where

$$P = \left(\sum_{\text{gene set}} \text{expected value}_i\right) / \left(\sum_{\text{all genes}} \text{expected value}_j\right)$$

Enrichment was calculated as the observed number of genotypes/variants divided by the expected number of genotypes/variants.

Gene-based binomial test.—A one-tailed binomial test was used to compare the observed number of damaging RGs within each gene to the expected number estimated using the approach detailed above. Enrichment was calculated as the number of observed damaging RGs divided by the expected number of damaging RGs.

Genetic overlap across NDDs.—We compared the list of 439 putative CP risk genes (Supplementary Datasets 6–15) with genes identified in other major NDDs using DisGeNET (updated May 2019)[65]. We first extracted all of the genes from DisGeNET that were associated with ASD (CUI: C1510586, 571 genes), ID (CUI: C3714756, 2,502 genes) and epilepsy (CUI: C0014544, 1,176 genes). We used the hypergeometric probability to calculate the overlap significance. The hypergeometric distribution formula is given by:

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

where K represents the number of genes in DisGeNET associated with the disease, k represents the number of genes in the overlapping set with that disease, N represents the total number of genes in DisGeNET and n represents the total number of genes in the observed set.

A Venn diagram representing the gene number appearing in more than one list was created in R using the VennDiagram package.

Pathway analysis.

STRING protein–protein interaction enrichment.—We used the list of 439 genes (Supplementary Datasets 6–15) to conduct a protein–protein interaction enrichment analysis for gene networks. We used STRINGv11 to further study protein interaction networks in our set of 439 putative CP risk genes with de novo, X-linked recessive or autosomal recessive damaging variants. We used a 0.70 (high confidence) cutoff to derive these interaction networks as described previously[66]. The network visualization can be accessed at https://version-11-0.string-db.org/cgi/network.pl?networkId=sKvp4sjmxO4O.

Gene-set over-representation analysis.—We used the list of 439 genes (Supplementary Datasets 6–15) for further downstream gene-set over-representation analysis using DAVID v6.8 (refs. [69,97]) (updated October 2016), PANTHER v15.0 (ref. [98]) (updated 14 February 2020) and MSigDB v7.0 (ref. [99]) (updated August 2019). The background gene list for all three tools was their respective pool of all human genes. To measure statistical over-representation of gene sets in the client set, PANTHER uses a Fisher's exact two-tailed test, DAVID uses a modified Fisher's test and MSigDB uses the hypergeometric distribution two-tailed test.

The DcGO[72] algorithm identifies parent and child nesting GO terms to determine hierarchical relationships. We started from the most specific GO terms (fewest genes) to identify first-level parents. These terms were used with DcGO to identify terms where parent, middle and child terms were all represented on our list with significant FDR. These nested terms were manually curated for Table 4.

RHOB functional assays.

GAP assay.—Human reference or S73F purified RHOB protein (13 μg, Origene) was incubated with 20 μM GTP with or without 5 μg or 13 μg of p50 RhoGAP for 30 min at 37 °C, and then incubated with CytoPhos reagent for 15 min at room temperature (Cytoskeleton). Hydrolyzed GTP was detected at 650 nm on a SpectraMax paradigm microplate reader as per the manufacturer's instructions. Data are from three independent biological replicates.

Guanine exchange factor (GEF) assay.—Human reference or S73F purified RHOB protein (2 μM, Origene) was incubated with or without a 2 μM concentration of the GEF domain of the human Dbs protein for 30 min at 20 °C. The fluorescence of *N*-methylantraniloyl GTP-analog binding was measured every 30 s at 360 nm with the

SpectraMax as per the manufacturer's instructions (Cytoskeleton). Data are from five independent biological replicates.

Rhotekin assay.—Agarose beads (50 μg) were coated with the Rho-GTP binding domain (residues 7–89) of the human rhotekin protein (Cytoskeleton) and were incubated with 500 μg of lysate from yeast expressing human RHOB–V5 or the S73F variant under gentle agitation for 1 h at 4 °C. Beads were pelleted by centrifugation at 2,400*g* (5,000 r.p.m.) for 4 min at 4 °C and washed three times in wash buffer (25 mM Tris pH 7.5, 30 mM $MgCl_2$, 40 mM NaCl). Beads were resuspended in Laemmli blue 2× and 40 μg of lysate was used for western blotting. RHOB was identified with a primary monoclonal anti-V5 antibody (Thermo Fisher) 1:5,000 in BSA and a secondary goat anti–mouse HRP (GE Healthcare) 1:5,000. Data are from five independent biological replicates.

***FBXO31* cyclin D abundance assay.—**Three independent, passage-matched control fibroblast lines (GMO8398, GMO2987 and GMO8399 from the Corriell Institute) and two patient primary fibroblasts obtained from each patient via punch biopsies were used. The total sample consisted of $n = 7$ controls and $n = 6$ patient measurements. Plates were seeded at 600,000 cells per well and cultured in DMEM supplemented with 1 mM sodium pyruvate, 1 mM glutamine (Gibco) and 10% FBS. Fibroblasts were collected at confluence with RIPA buffer (Thermo Fisher) supplemented with protease cocktail (Fischer Scientific) on ice and centrifuged. Western blotting was conducted using 10 μg protein per lane with antibodies against cyclin D (rabbit polyclonal; ab134175) 1:1,000, β-tubulin (rabbit polyclonal, ab6046) 1:5,000 in 5% BSA and detected with anti-rabbit HRP (GE Health Sciences) 1:5,000. Signal was quantified using Image Studio Lite and the ratio of cyclin D/β-tubulin was normalized to the within-experiment control GMO8398. The difference in cyclin D abundance was determined using an unpaired *t*-test.

Drosophila locomotor experiments.

Fly rearing and genetics.—*Drosophila* were reared on a standard cornmeal, yeast, sucrose food from the BIO5 media facility, University of Arizona. Stocks for experiments were reared at 25 °C, 60–80% relative humidity with a 12:12 light/dark cycle. Cultures for controls and mutants were maintained with the same growth conditions, with attention to the density of animals within the vial. Descriptions of alleles used for each CP candidate gene can be found in Supplementary Table 9 and include 5' insertional hypomorphs, missense mutations, targeted excision and deficiency chromosomes. Fly stocks were obtained from the Bloomington Drosophila Stock Center (NIH P40OD018537) and other investigators. We performed crosses of background markers for genetic controls.

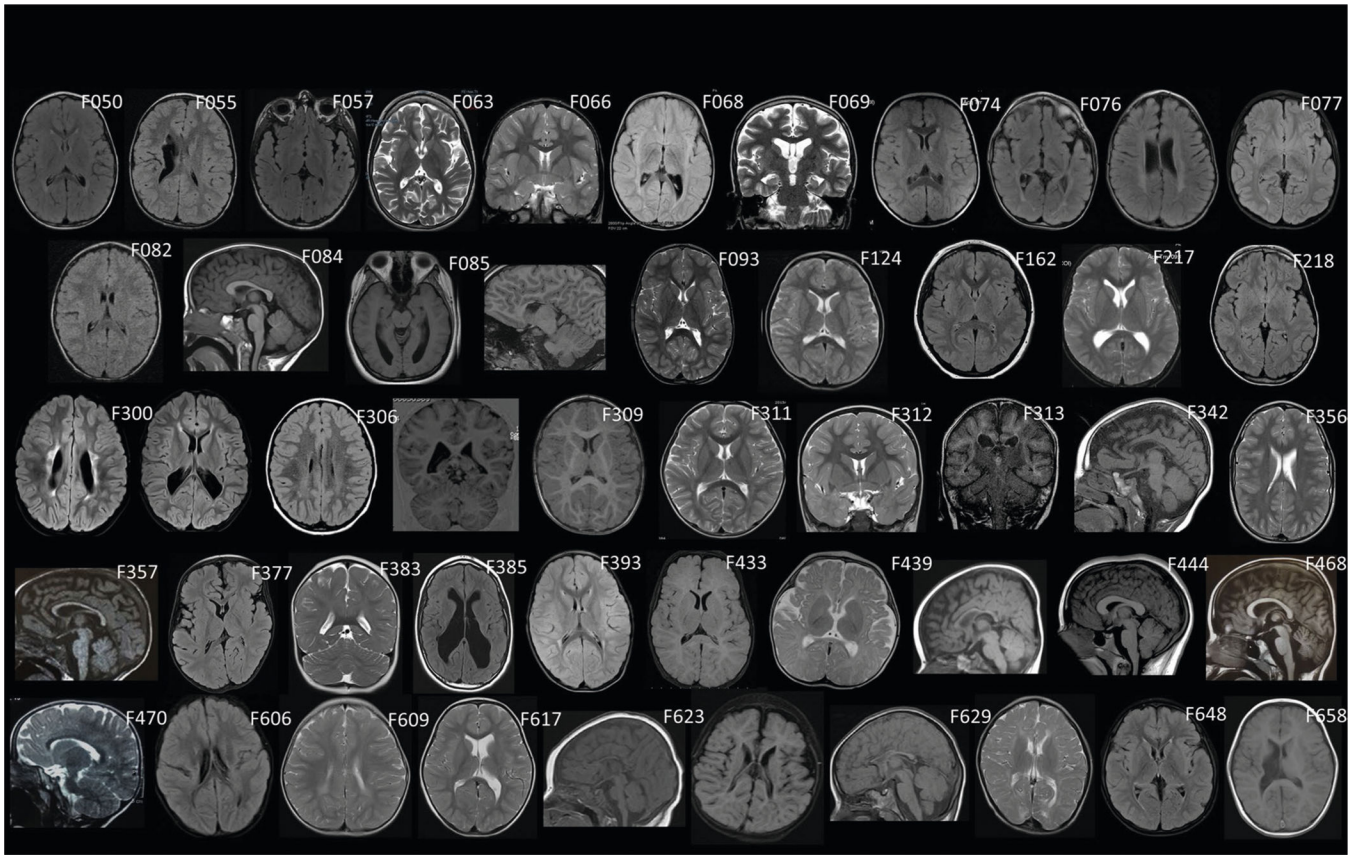
Locomotor assays.—We used naive, unmated flies collected as pharate adults. To minimize variables, we used no anesthesia, and humidity, temperature and time of day were controlled (30–60% RH, 21–23.5 °C, 9:00–12:00). Flies were adapted to room conditions for 1 h before running in groups of 3–20 in a 250-ml graduated cylinder for 2 min (ref. [76]). If <50% crossed the 250 ml (22.5 cm) mark, flies were re-assayed immediately up to three iterations. Flies crossing the 250 ml mark (22.5 cm) were manually scored from coded videos in 10-s bins for 10–21 trials per genotype. The number of falls, defined as downward

movement while detached from the cylinder wall, was manually counted and normalized to the number of flies in the recording window per 10-s bin for 10–21 trials per genotype. A significant difference of locomotor performance between mutants and controls required $P < 0.05$ for both a Kolmogorov–Smirnov test for the whole curve and a Mann–Whitney rank sum test for at least one time bin between 10 and 30 s. The distance traveled assay was performed using paired, coded vials of control and mutant flies[77]. The distance was measured from a still image from a video at 3 s post-tapping using the ImageJ measure distance function from the middle of the fly to the bottom of the vial for 10–11 trials. Larval turning time was defined as the amount of time required to turn onto the ventral surface and initiate forward movement after rotation onto the dorsal surface and measured for 50 larvae per genotype[75]. Significance for vial and larval turning assays was determined using a *t*-test. Graphs were prepared and statistical analysis was performed in R. Enrichment in the number of genes with locomotor defects from our screen compared to the frequency of reports of locomotor defects in the entire *Drosophila* genome was performed as described previously[78]. We used www.MARRVEL.org and https://www.flybase.org to identify the *Drosophila* ortholog and compared to genome-wide number of genes identified by the terms locomotor/ locomotion, flight and taxis (photo- or geo-). The significance of the enrichment was determined using the Fisher exact two-tailed test. Assay validation and additional genetics information is provided in the Supplementary Note.

Reporting Summary.

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

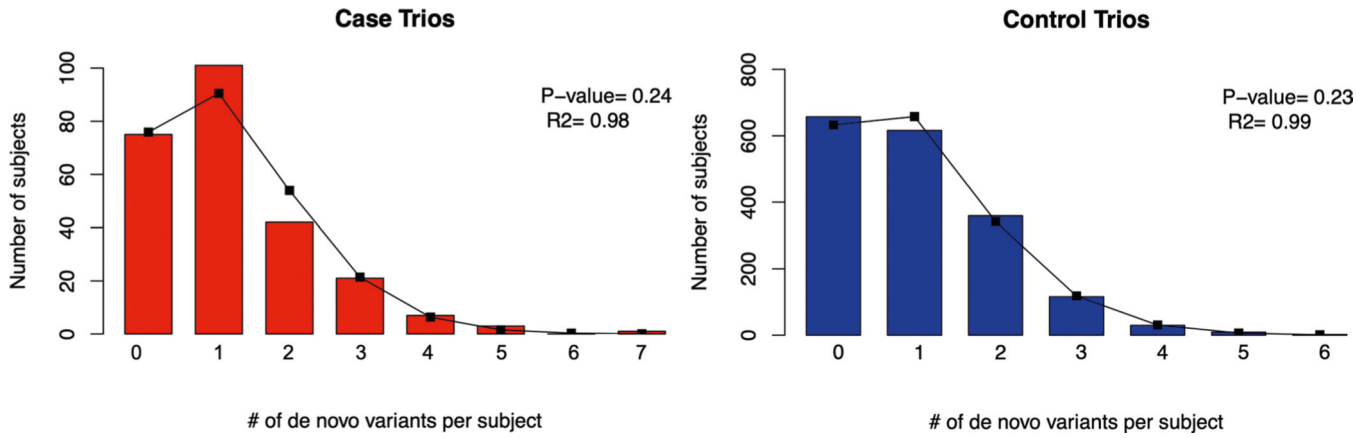
Extended Data



Extended Data Fig. 1 |. Brain MRI features of idiopathic cerebral palsy.

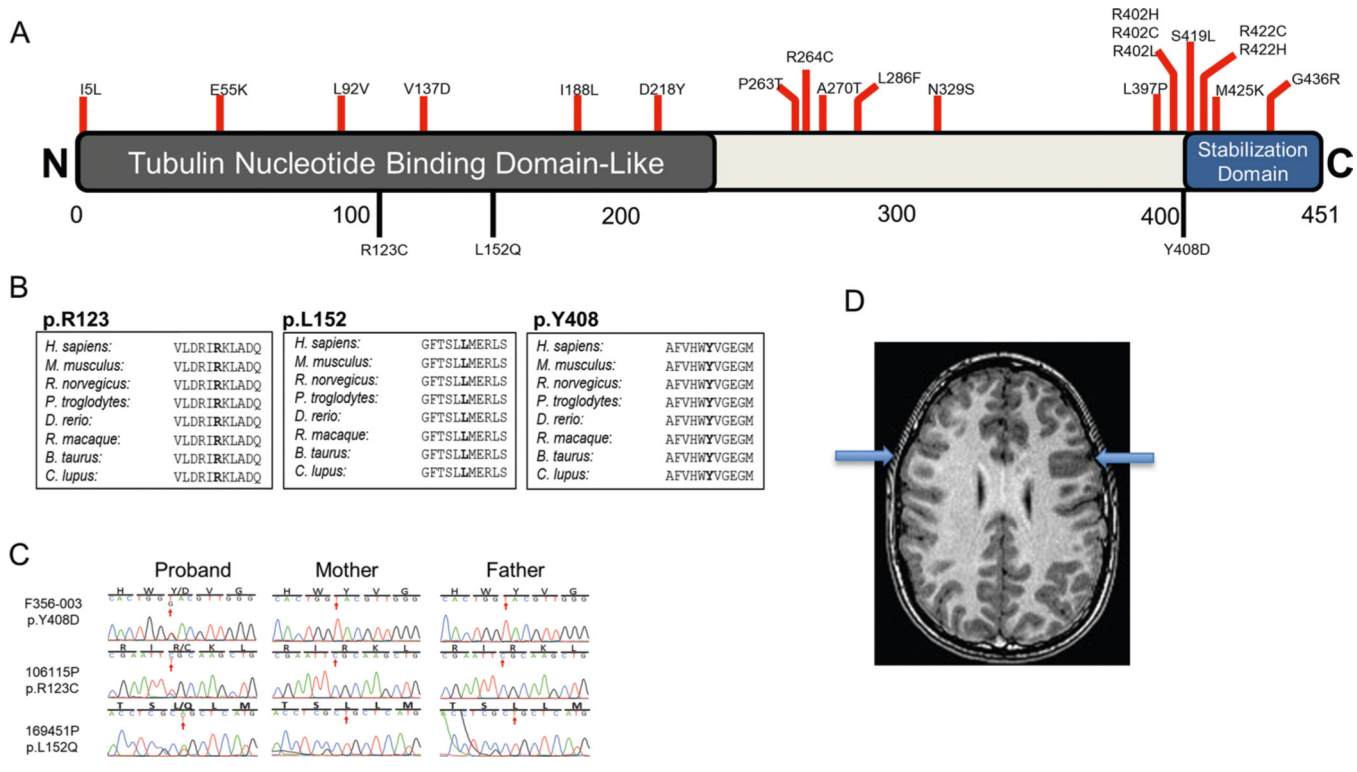
F050: bilateral periventricular leukomalacia; F055: right sided porencephaly; F057: normal (equivocal putaminal rim hyperintensity); F063: mildly globally diminished cerebral volume; F066: normal; F068, bilateral mild periventricular leukomalacia, white matter thinning and colpocephaly; F069: diminished cortical more than cerebellar volumes; F074: normal; F076: ex vacuo ventriculomegaly; bilateral periventricular leukomalacia, and bilateral perisylvian pachygyria; F077: mild periventricular leukomalacia; F082: scattered subcortical T2 hyperintensities; F084: normal; F085: colpocephaly, thinning of periventricular white matter, hypoplastic corpus callosum, diminished left cerebellar hemispheric volume; F093: normal; F124: normal; F162: normal; F217: equivocal ex vacuo ventriculomegaly; F218: normal; F300: bilateral periventricular leukomalacia with thin corpus callosum; F306: scattered bilateral subcortical punctate t2/FLAIr hyperintensities; F309: simplified gyral pattern; F311: normal; F312: normal; F313: normal; F342: diminished cortical volume, thinning and t2/FLAIr signal hyperintensity of periventricular white matter, thin corpus callosum; F356: bilateral perisylvian polymicrogyria; F357: thin corpus callosum; F377: equivocally simplified gyri with ‘open opercula’; F383: bilateral occipital horn heterotopias; F385: hydrocephalus and periventricular leukomalacia; F393: periventricular leukomalacia; F433: normal; F439: increased frontotemporal extra-axial fluid spaces and thin corpus callosum; F444: normal (equivocally thickened corpus callosum); F468: slight ex vacuo ventriculomegaly; F470: equivocally diminished cortical

volume; F606: bilateral perislyvian pachygyria; F609: bi hemispheric periventricular leukomalacia; F617: ex vacuo ventriculomegaly; F623: dysplastic corpus callosum, bitemporal diminished cortical volumes; F629: thin corpus callosum, colpocephaly, with periventricular leukomalacia; F648: periventricular leukomalacia; F658: right sided encephalomalacia affecting putamen and thalamus.



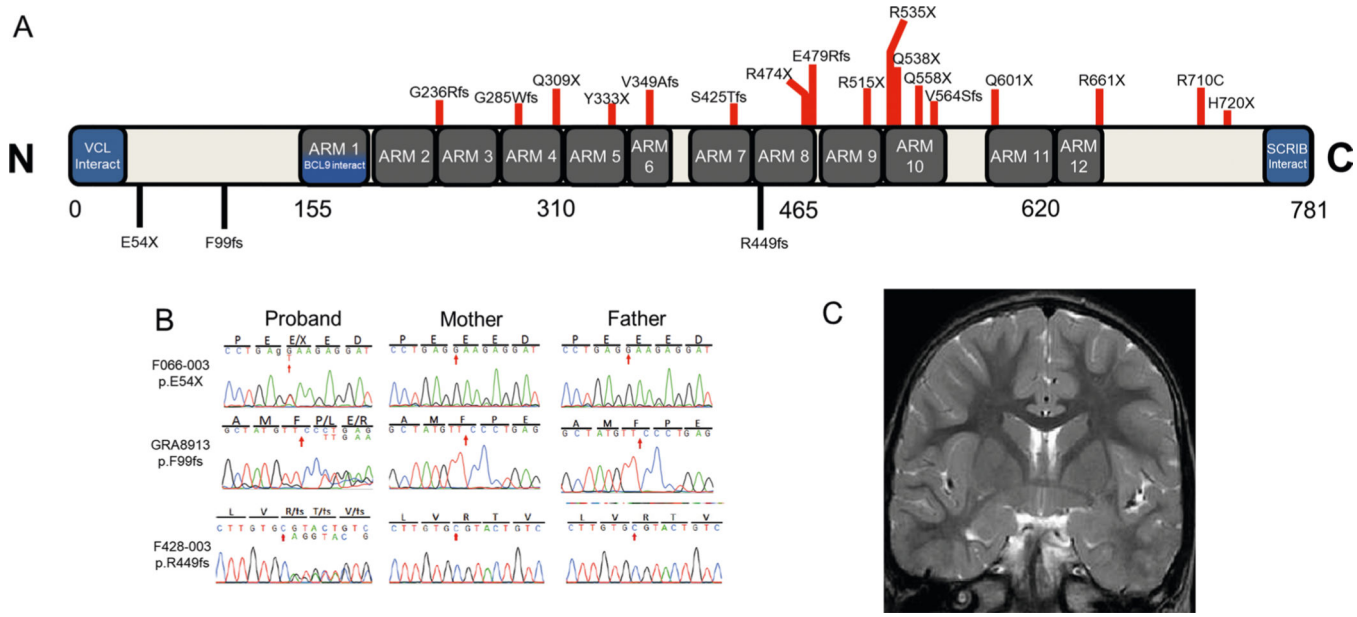
Extended Data Fig. 2 |. De novo mutation rate closely approximates Poisson distribution in cases and controls.

Observed number of *de novo* mutations per subject (bars) compared to the numbers expected (line) from the Poisson distribution in the case (red) and control (blue) cohorts. Here, 'P' denotes chi-squared *P*-value.

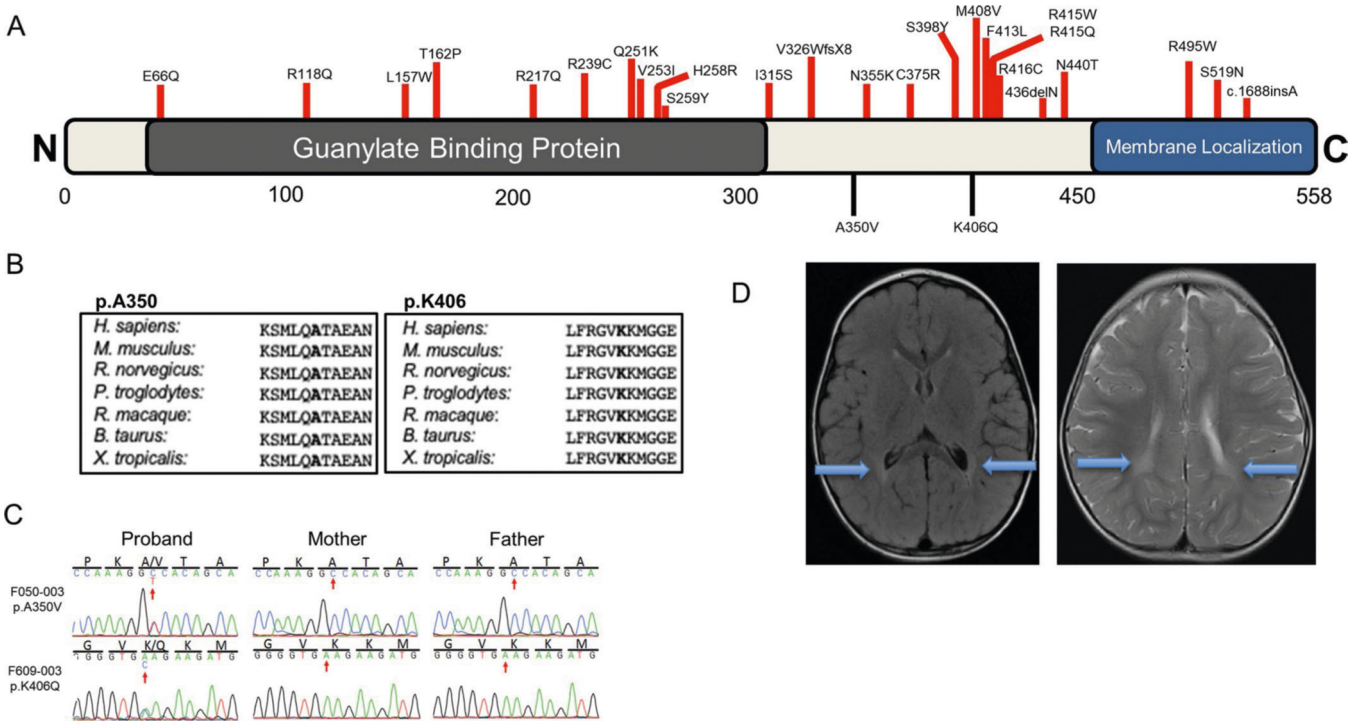


Extended Data Fig. 3 |. *De novo* mutation in *TUBA1A* encoding α-tubulin.

a, TUBA1A functional domains schematic with locations of previously-described pathogenic variants (red) compared to those from this work (black). **b**, Phylogenetic conservation of reference amino acid at each mutated position described in this work. **c**, Sanger-verified mutated base (red arrow) with the corresponding reference bases. **d**, MRI of the brain (F356) demonstrates evidence of bilateral perisylvian pachygyria (blue arrows). conserved Domain Annotations: TNBDL (AA 1–244) as IPro36525; SD (AA 418–451) annotated as per[39].

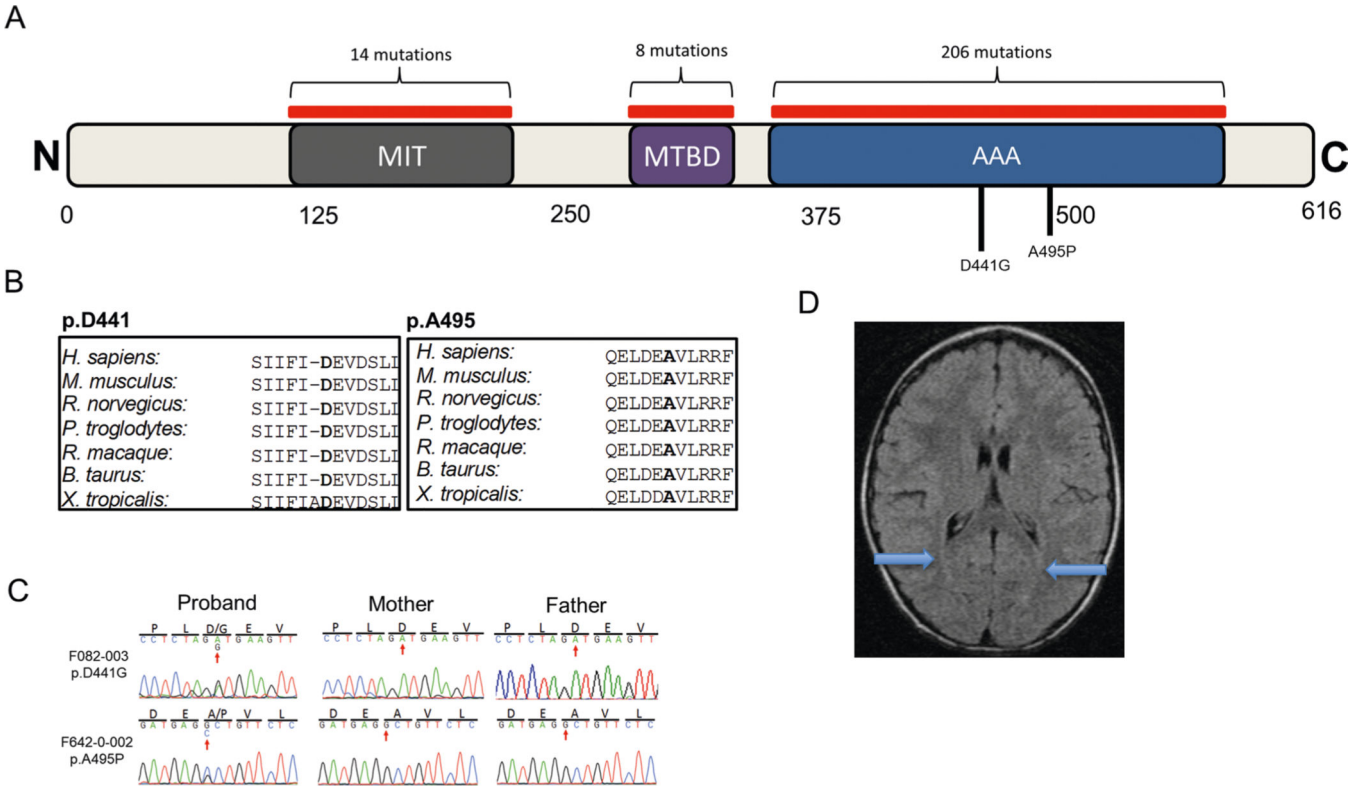


Extended Data Fig. 4 |. *De novo* mutations in *CTNNB1* encoding β-catenin.
a, CTNNB1 functional domain with location of previously reported pathogenic variants (red) and those identified in this work (black). (Given the loss-of-function nature of the identified variants, phylogenetic alignments were not performed; however, 100% identify is seen at these loci (p.E54, p.F99, and p.R449) in primates). **b**, Sanger-verified mutated base (red arrow) with corresponding reference bases. **c**, Brain MRI (F066) was unremarkable. conserved Domain Annotations: ARM, Armadillo/beta-catenin-like repeats from UniProtKB/Swiss-Prot (P35222.1); SCRIB, interaction with SCRIB (AA 772–781, by similarity, experimental evidence); BCL9, interaction with BCL9 (AA 156–178, by similarity, experimental evidence); VCL, interaction with VCL (AA 2–23, by similarity, experimental evidence).



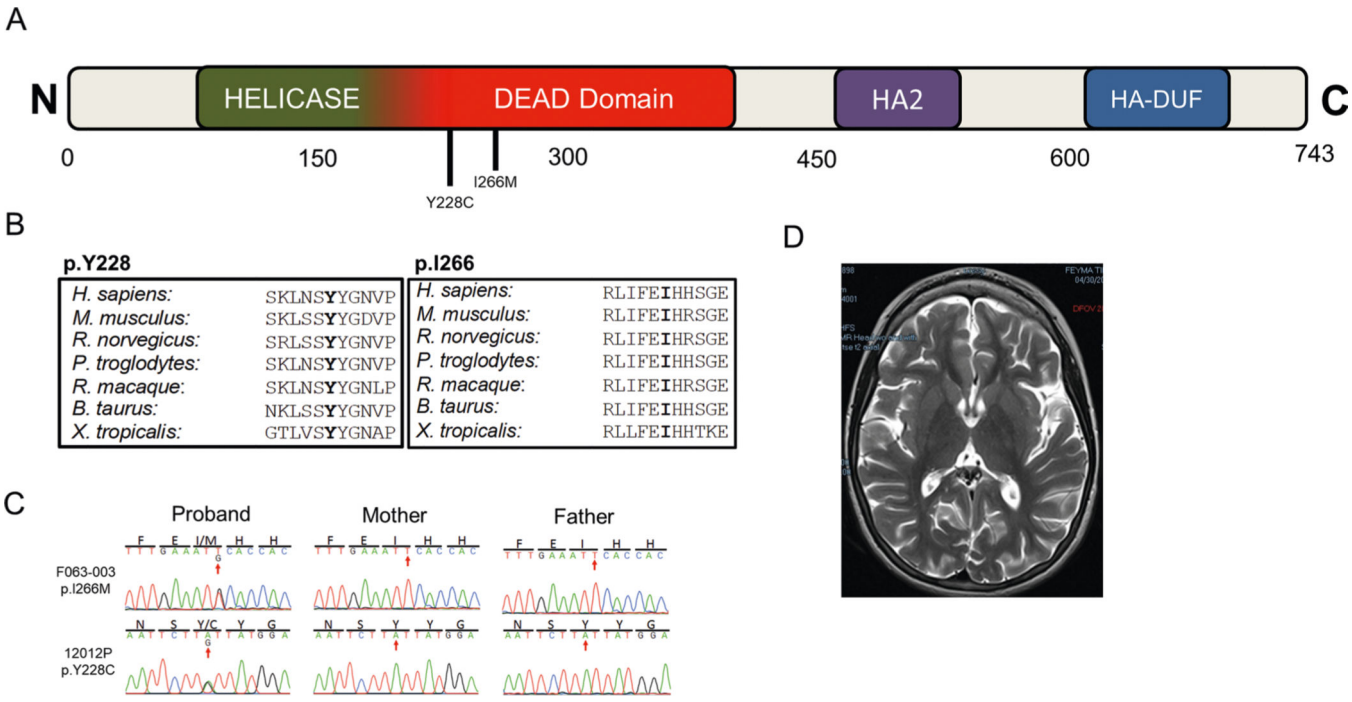
Extended Data Fig. 5 |. *De novo* mutations in *ATL1* encoding atlastin-1.

a, ATL1 functional domain with location of previously reported variants (red) as well as those identified in this work (black). **b**, Phylogenetic conservation of reference amino acid at each affected position. **c**, Sanger-verified mutated base (red arrow) with the corresponding reference bases. **d**, Brain MRI images from F050 and F609 demonstrate mild periventricular T2 hyperintensity (blue arrows). conserved Domain Annotations: GBP (AA 43–314) as pfam02263; Membrane localization domain (AA 448–558) from UniProtKB (Q8WXF7.1).



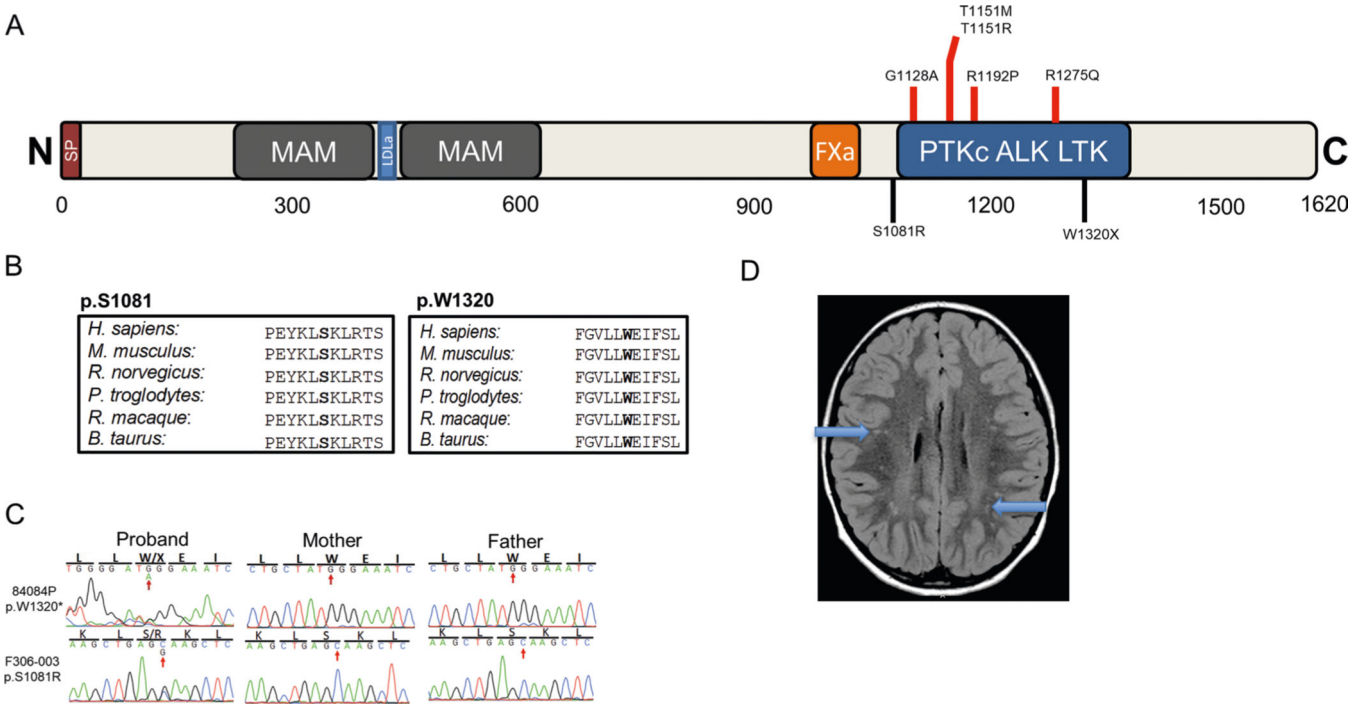
Extended Data Fig. 6 |. *De novo* mutations in *SPAST* encoding spastin.

a, SPAST functional domains with location of CP-associated damaging variants identified in this study (black); 277 pathological mutations[58] have previously been identified in *SPAST* with the majority (82%) located within the conserved domains (red). **b**, Phylogenetic conservation of wild-type amino acid at each mutated position. **c**, Sanger-verified mutated base indicated by red arrow with corresponding reference bases. **d**, Brain MRI (F082) showed mild subcortical T2 hyperintensities (blue arrows). conserved Domain Annotations: MIT (AA 116–196) as CDD:239142; Microtubule binding domain (AA 270–328) from UniProtKB/Swiss-Prot (Q9UbP0.1); ATPase AAA core and Lid domains (378–567) from IPR003959 and IPR041569, respectively.

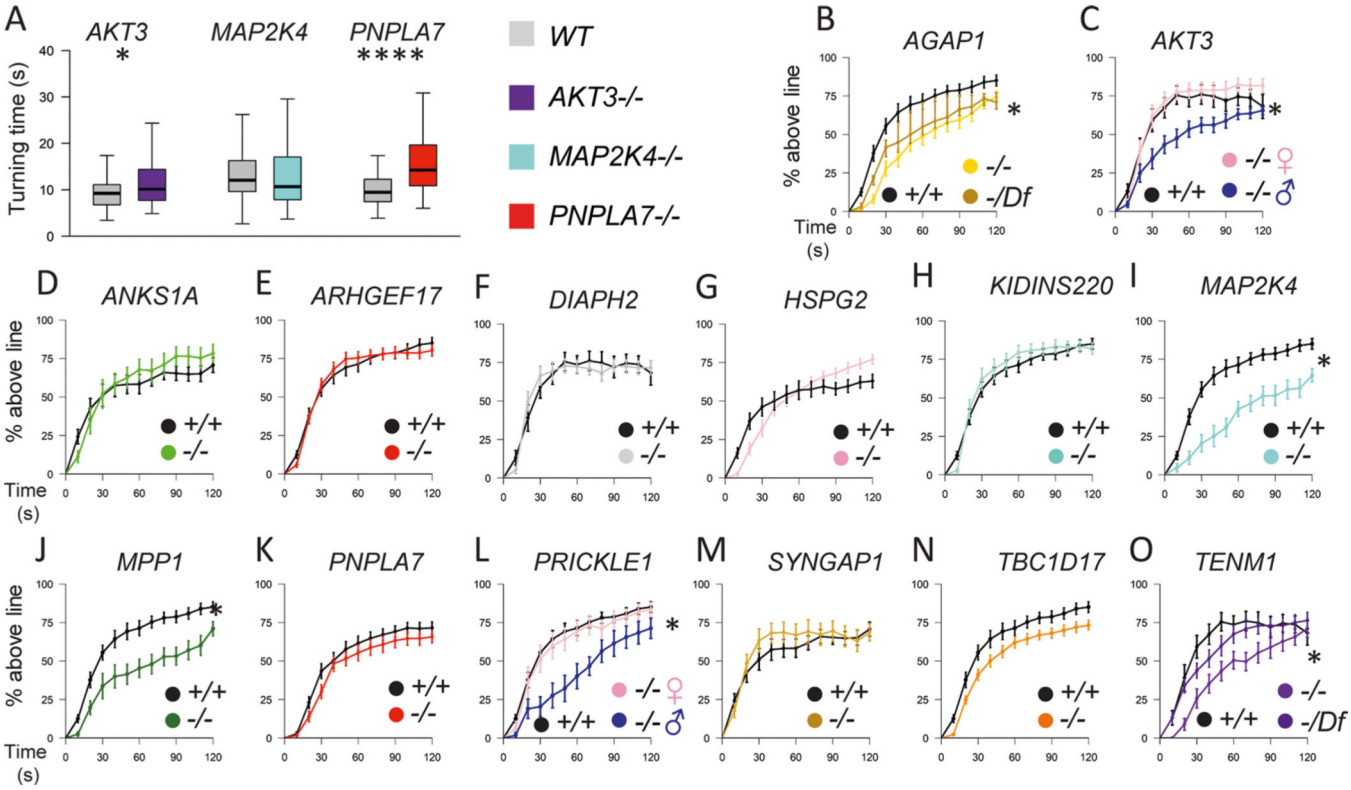


Extended Data Fig. 7 |. *De novo* mutations in *DHX32* encoding the DEAH box polypeptide 32.

a, DHX32 functional domains with location of CP-associated damaging variants from this work (black). Germline *DHX32* variants have not been previously associated with human disease although somatic variants (>40) have been associated with variants cancers (COSMIC). **b**, Phylogenetic conservation of wild-type amino acid at each mutated position. **c**, Sanger-verified mutated base indicated by red arrow with corresponding reference bases. **d**, Brain MRI (F063) showed diffusely diminished cortical volume. conserved Domain Annotations: Helicase and DEAD domains overlap (72–378 and 146–403) from IPR014001 and cd17912, respectively; HA2 domain (AA 458–547) as IPR007502; Helicase associated domain of unknown function (AA 616–696) from IPR011709.

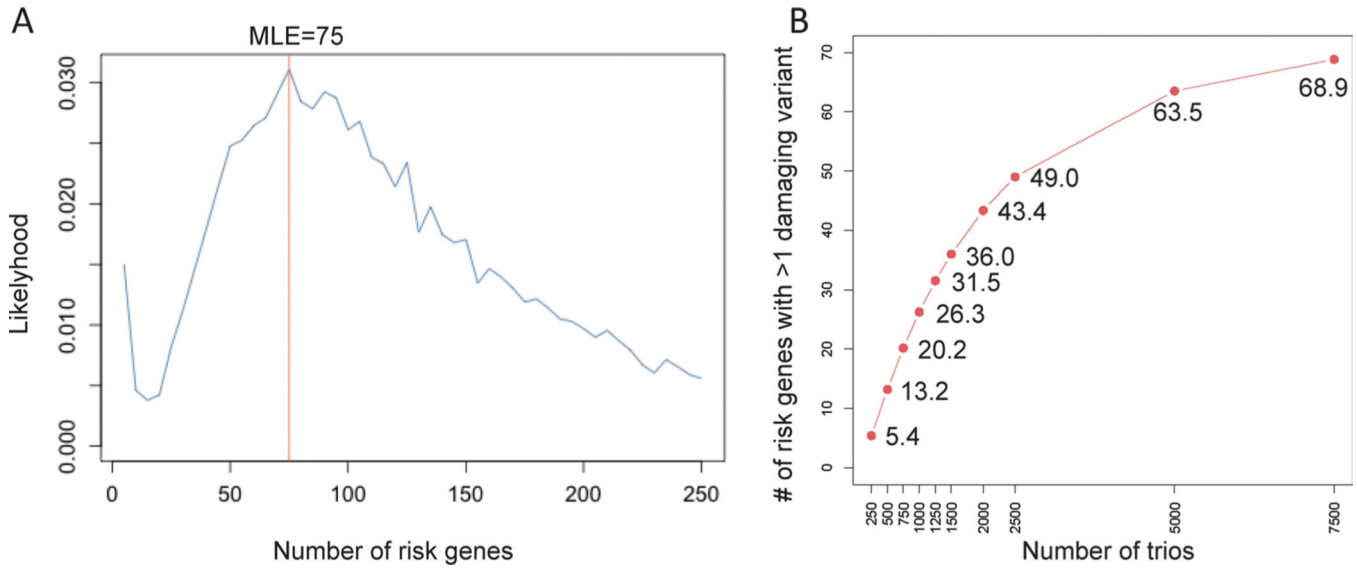


Extended Data Fig. 8 |. *De novo* mutations in *ALK* encoding the anaplastic lymphoma kinase. **a**, ALK functional domain with location of previously reported pathogenic variants associated with susceptibility to neuroblastoma (OMIM# 613014) (red) as well as CP-associated damaging variants identified in this work (black). **b**, Phylogenetic conservation of wild-type amino acid at each mutated position. **c**, Sanger-verified mutated base indicated by red arrow with corresponding reference bases. **d**, Brain MRI (F306) demonstrates punctate subcortical T2 hyperintensities of both hemispheres. conserved Domain Annotations: Signal Peptide (AA 1–18) by SignalP 4.0; MAM (AA 266–427, 480–636) as pfam #00629; LDLa (AA 441–467) as smart#00192; Fxa (AA 987–1021) as pfam#14670; PtKc ALK LTK (AA 1109–1385) as CDD#05036.



Extended Data Fig. 9 |. Additional locomotor phenotypes of loss of function mutations in Drosophila orthologs of candidate cerebral palsy risk genes.

Drosophila mutant and control genotypes are shown in Supplementary Table 9. **a**, Turning time, a measure of coordinated movements, is increased in larva with mutations in *AKT3* and *PNPLA7* orthologs, but not in *MAP2K4*. **b-o**, Distance threshold assay examining negative geotaxis climbing defects in for 14 day-old adult flies with mutations in orthologs of *AGAP1* (**b**), *AKT3* (**c**), *ANKS1A* (**d**), *ARHGEF17* (**e**), *DIAPH2* (**f**), *HSPG2* (**g**), *KIDINS220* (**h**), *MAP2K4* (**i**), *MPP1* (**j**), *PNPLA7* (**k**), *PRICKLE1* (**l**), *SYNGAP1* (**m**), *TBC1D17* (**n**), and *TENM1* (**o**). Impairments in the climbing assay was detected for males with mutations in *AKT3* and *PRICKLE1* (**c**,**l**) and for both sexes with mutations in *MAP2K4* and *MPP1* (**i**,**j**) orthologs. climbing phenotype mapped to gene using deficiency chromosome for *AGAP1* (**b**), but did not map for *TENM1* (**o**). there was no locomotor impairment in the two negative control genotypes, *ARHGEF15* and *ANKS1A*, where the patient variant did not pass our deleteriousness filters (**d**). For larval turning, box indicates 75th and 25th percentile with median line; whiskers indicate 10th and 90th percentile (n = 50 larvae). Locomotor curve represents average of all trials and bars indicate standard error (n = 10–21 trials). Statistics between larval turning times determined using unpaired 2-tailed t-test. Locomotor curves considered to be significantly different from each other if $P < 0.05$ for Kolomogrov-Smirnov test in addition to a significant difference at one or more time bins by Mann-Whitney rank sum 2-tailed test. *$P < 0.05$, ****$P < 1 \times 10^{-6}$. exact genotypes, n, and P values are provided in Supplementary Table 9.



Extended Data Fig. 10 |. Cerebral palsy gene discovery projections.

a, estimation of the number of cerebral palsy risk genes via *de novo* mechanism. Monte carlo simulation performed was performed based on observed damaging *de novo* mutations in 3,049 loss-of-function intolerant genes (pLI ≥ 0.9 in gnomAD (v2.1.1)) using 20,000 iterations. We estimate that the number of risk genes via *de novo* events to be ~75 (95% confidence interval = (26.5, 123.5)). **b**, estimation of the number of recurrent genes. the number of trios and the number of genes with more than one damaging *de novo* mutation are specified on the *x* and *y*-axis, respectively. We modeled the expected rate of damaging *de novo* mutations given an increasing sample size. A total of 10,000 iterations were performed to estimate the number of genes with more than one damaging *de novo* mutations taking into account of the damaging *de novo* mutation probability. WeS of 2,500 and 7,500 trios are expected to yield a 65.3% and 91.8% saturation rate, respectively, for all cerebral palsy risk genes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Sheng Chih Jin[1,2,3,34], Sara A. Lewis[4,5,34], Somayeh Bakhtiari[4,5,34], Xue Zeng[1,2,34], Michael C. Sierant[1,2], Sheetal Shetty[4,5], Sandra M. Nordlie[4,5], Aureliane Elie[4,5], Mark A. Corbett[6], Bethany Y. Norton[4,5], Clare L. van Eyk[6], Shozeb Haider[7], Brandon S. Guida[4,5], Helen Magee[4,5], James Liu[4,5], Stephen Pastore[8], John B. Vincent[8], Janice Brunstrom-Hernandez[9], Antigone Papavasileiou[10], Michael C. Fahey[11], Jesia G. Berry[6], Kelly Harper[6], Chongchen Zhou[12], Junhui Zhang[1], Boyang Li[13], Jennifer Heim[4], Dani L. Webber[6], Mahalia S. B. Frank[6], Lei Xia[14], Yiran Xu[14], Dengna Zhu[14], Bohao Zhang[14], Amar H. Sheth[1], James R. Knight[15], Christopher Castaldi[15], Irina R. Tikhonova[15], Francesc López-Giráldez[15], Boris Keren[16], Sandra Whalen[17], Julien Buratti[16], Diane Doummar[18], Megan Cho[19],

Kyle Retterer[19], Francisca Millan[19], Yangong Wang[20], Jeff L. Waugh[21], Lance Rodan[22], Julie S. Cohen[23], Ali Fatemi[23], Angela E. Lin[24], John P. Phillips[25], Timothy Feyma[26], Suzanna C. MacLennan[27], Spencer Vaughan[28], Kylie E. Crompton[29], Susan M. Reid[29], Dinah S. Reddihough[29], Qing Shang[12], Chao Gao[30], Iona Novak[31], Nadia Badawi[31], Yana A. Wilson[31], Sarah J. McIntyre[31], Shrikant M. Mane[15], Xiaoyang Wang[14,32], David J. Amor[29], Daniela C. Zarnescu[28], Qiongshi Lu[33], Qinghe Xing[20,35], Changlian Zhu[14,32,35], Kaya Bilguvar[1,15,35], Sergio Padilla-Lopez[4,5,35], Richard P. Lifton[1,2,35], Jozef Gecz[6,35], Alastair H. MacLennan[6,35], Michael C. Kruer[4,5,35]

Affiliations

[1]Department of Genetics, Yale University School of Medicine, New Haven, CT, USA.

[2]Laboratory of Human Genetics and Genomics, Rockefeller University, New York, NY, USA.

[3]Department of Genetics, Washington University School of Medicine, St Louis, MO, USA.

[4]Pediatric Movement Disorders Program, Division of Pediatric Neurology, Barrow Neurological Institute, Phoenix children's Hospital, Phoenix, AZ, USA.

[5]Departments of child Health, Neurology, and cellular & Molecular Medicine, and Program in Genetics, University of Arizona college of Medicine–Phoenix, Phoenix, AZ, USA.

[6]Robinson Research Institute, The University of Adelaide, Adelaide, South Australia, Australia.

[7]Department of Pharmaceutical and Biological Chemistry, UCL School of Pharmacy, London, UK.

[8]Molecular brain Sciences, Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON, Canada.

[9]One CP Place, Plano, TX, USA.

[10]Division of Paediatric Neurology, Iaso Children's Hospital, Athens, Greece.

[11]Department of Pediatrics, Monash University, Melbourne, Victoria, Australia.

[12]Henan Key Laboratory of Child Genetics and Metabolism, Rehabilitation Department, Children's Hospital of Zhengzhou University, Zhengzhou, China.

[13]Department of biostatistics, Yale School of Public Health, New Haven, CT, USA.

[14]Henan Key Laboratory of Child Brain Injury, Third Affiliated Hospital of Zhengzhou University, Zhengzhou, china.

[15]Yale Center for Genome Analysis, Yale University, New Haven, CT, USA.

[16]Department of Genetics, Pitié-Salpêtrière Hospital, APHP.Sorbonne Université, Paris, France.

[17]UF de Génétique Clinique et Centre de Référence Anomalies du Développement et Syndromes Malformatifs, APHP.Sorbonne Université, Hôpital Armand trousseau, Paris, France.

[18]Sorbonne Université, APHP, Service de Neurologie Pédiatrique et centre de Référence Neurogénétique, Hôpital Armand Trousseau, Paris, France.

[19]GeneDx, Gaithersburg, MD, USA.

[20]Institute of biomedical Science and children's Hospital, and Key Laboratory of Reproduction Regulation of the National Population and Family Planning Commission (NPFPC), Shanghai Institute of Planned Parenthood research (SIPPR), IRD, Fudan University, Shanghai, China.

[21]Departments of Pediatrics & Neurology, University of Texas Southwestern and Children's Medical Center of Dallas, Dallas, TX, USA.

[22]Departments of Genetics & Genomics and Neurology, Boston Children's Hospital, Boston, MA, USA.

[23]Division of Neurogenetics and Hugo W. Moser Research Institute, Kennedy Krieger Institute, Baltimore, MD, USA.

[24]Medical Genetics, Department of Pediatrics, MassGeneral Hospital for Children, Boston, MA, USA.

[25]Departments of Pediatrics and Neurology, University of New Mexico, Albuquerque, NM, USA.

[26]Division of Pediatric Neurology, Gillette children's Hospital, St Paul, MN, USA.

[27]Department of Paediatric Neurology, Women's & Children's Hospital, Adelaide, South Australia, Australia.

[28]Departments of Molecular & Cellular Biology and Neuroscience, University of Arizona, Tucson, AZ, USA.

[29]Murdoch Children's Research Institute and University of Melbourne Department of Paediatrics, Royal children's Hospital, Melbourne, Victoria, Australia.

[30]Rehabilitation Department, children's Hospital of Zhengzhou University/Henan children's Hospital, Zhengzhou, China.

[31]cerebral Palsy Alliance research Institute, University of Sydney, Sydney, New South Wales, Australia.

[32]Institute of Neuroscience and Physiology, Sahlgrenska Academy, Gothenburg University, Gothenburg, Sweden.

[33]Department of biostatistics & Medical Informatics, University of Wisconsin-Madison, Madison, WI, USA.

[34]These authors contributed equally: Sheng Chih Jin, Sara A. Lewis, Somayeh Bakhtiari, Xue Zeng.

[35]These authors jointly supervised this work: Qinghe Xing, Changlian Zhu, Kaya Bilguvar, Sergio Padilla-Lopez, Richard P. Lifton, Jozef Gecz, Alastair H. MacLennan, Michael C. Kruer.

Acknowledgements

We gratefully acknowledge the support of the patients and families who have graciously and patiently supported this work from its inception. Without their partnership, these studies would not have been possible. We acknowledge the support of the clinicians who generously provided their expertise in support of this study, including M.-C. Waugh, M. Axt and V. Roberts of the Children's Hospital Westmead; K. Lowe of Sydney Children's Hospital; R. Russo, J. Rice and A. Tidemann of the Women's and Children's Hospital, Adelaide; T. Carroll and L. Copeland of the Lady Cilento Children's Hospital, Brisbane; and J. Valentine of Perth Children's Hospital. We appreciate the collaboration of S. Knoblach and E. Hoffman (Children's National Medical Center). This work was supported in part by the Cerebral Palsy Alliance Research Foundation (M.C.K.), the Yale-NIH Center for Mendelian Genomics (U54 HG006504–01), Doris Duke Charitable Foundation CSDA 2014112 (M.C.K.), the Scott Family Foundation (M.C.K.), Cure CP (M.C.K.), NHMRC grant 1099163 (A.H.M., C.L.v.E., J.G. and M.A.C.), NHMRC Senior Principal Research Fellowship 1155224 (J.G.), Channel 7 Children's Research Foundation (J.G.), a Cerebral Palsy Alliance Research Foundation Career Development Award (M.A.C.), the Tenix Foundation (A.H.M., J.G., C.L.v.E. and M.A.C.), the National Natural Science Foundation of China (U1604165, X.W.), Henan Key Research Program of China (171100310200, C. Zhu), VINNOVA (2015–04780, C. Zhu), the James Hudson Brown–Alexander Brown Coxe Postdoctoral Fellowship at the Yale University School of Medicine (S.C.J.), an American Heart Association Postdoctoral Fellowship (18POST34060008 to S.C.J.), the NIH K99/R00 Pathway to Independence Award (R00HL143036–02 to S.C.J.) and NIH grants R01NS091299 (D.C.Z.) and NIH R01NS106298 (M.C.K.).

Data availability

Sequencing data from University of Adelaide Robinson Research Institute (n = 154 trios) are available from the corresponding author on request, subject to human research ethics approval and patient consent. Data from PCH (n = 52 trios) are available from the corresponding author on request, subject to patient consent. Data from Zhengzhou City Children's Hospital (n = 44 trios) are available in the CNSA of China National GeneBank DataBase repository (https://db.cngb.org/cnsa/). Source data are provided with this paper.

References

1. 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 (2014). [PubMed: 24117446]
2. Oskoui M, Coutinho F, Dykeman J, Jette N & Pringsheim T An update on the prevalence of cerebral palsy: a systematic review and meta-analysis. Dev. Med. Child Neurol. 55, 509–519 (2013). [PubMed: 23346889]
3. Cans C Surveillance of cerebral palsy in Europe: a collaboration of cerebral palsy surveys and registers. Dev. Med. Child Neurol. 42, 816–824 (2000). [PubMed: 11132255]
4. Longo LD & Ashwal S William Osler, Sigmund Freud and the evolution of ideas concerning cerebral palsy. J. Hist. Neurosci. 2, 255–282 (1993). [PubMed: 11618461]
5. Panteliadis C, Panteliadis P & Vassilyadi F Hallmarks in the history of cerebral palsy: from antiquity to mid-20th century. Brain Dev. 35, 285–292 (2013). [PubMed: 22658818]
6. Tan S Fault and blame, insults to the perinatal brain may be remote from time of birth. Clin. Perinatol. 41, 105–117 (2014). [PubMed: 24524449]
7. Donn SM, Chiswick ML & Fanaroff JM Medico-legal implications of hypoxic–ischemic birth injury. Semin. Fetal Neonatal Med. 19, 317–321 (2014). [PubMed: 25150792]
8. Korzeniewski SJ, Slaughter J, Lenski M, Haak P & Paneth N The complex aetiology of cerebral palsy. Nat. Rev. Neurol. 14, 528–543 (2018). [PubMed: 30104744]

9. Numata Y et al. Brain magnetic resonance imaging and motor and intellectual functioning in 86 patients born at term with spastic diplegia. Dev. Med. Child Neurol. 55, 167–172 (2013). [PubMed: 23121133]
10. Segel R et al. Copy number variations in cryptogenic cerebral palsy. Neurology 84, 1660–1668 (2015). [PubMed: 25817843]
11. McIntyre S et al. Congenital anomalies in cerebral palsy: where to from here? Dev. Med. Child Neurol. 58, 71–75 (2016). [PubMed: 26762782]
12. Petterson B, Stanley F & Henderson D Cerebral palsy in multiple births in Western Australia: genetic aspects. Am. J. Med. Genet. 37, 346–351 (1990). [PubMed: 2260563]
13. Costeff H Estimated frequency of genetic and nongenetic causes of congenital idiopathic cerebral palsy in west Sweden. Ann. Hum. Genet. 68, 515–520 (2004). [PubMed: 15469428]
14. Hallmayer J et al. Genetic heritability and shared environmental factors among twin pairs with autism. Arch. Gen. Psychiatry 68, 1095–1102 (2011). [PubMed: 21727249]
15. Sandin S et al. The heritability of autism spectrum disorder. J. Am. Med. Assoc. 318, 1182–1184 (2017).
16. McMichael G et al. Rare copy number variation in cerebral palsy. Eur. J. Hum. Genet. 22, 40–45 (2014). [PubMed: 23695280]
17. Oskoui M et al. Clinically relevant copy number variations detected in cerebral palsy. Nat. Commun. 6, 7949 (2015). [PubMed: 26236009]
18. Zarrei M et al. De novo and rare inherited copy-number variations in the hemiplegic form of cerebral palsy. Genet. Med. 20, 172–180 (2018). [PubMed: 28771244]
19. Corbett MA et al. Pathogenic copy number variants that affect gene expression contribute to genomic burden in cerebral palsy. NPJ Genom. Med. 3, 33 (2018). [PubMed: 30564460]
20. Takezawa Y et al. Genomic analysis identifies masqueraders of full-term cerebral palsy. Ann. Clin. Transl. Neurol. 5, 538–551 (2018). [PubMed: 29761117]
21. Parolin Schnekenberg R et al. De novo point mutations in patients diagnosed with ataxic cerebral palsy. Brain 138, 1817–1832 (2015). [PubMed: 25981959]
22. McMichael G et al. Whole-exome sequencing points to considerable genetic heterogeneity of cerebral palsy. Mol. Psychiatry 20, 176–182 (2015). [PubMed: 25666757]
23. Rosenbaum P et al. A report: the definition and classification of cerebral palsy April 2006. Dev. Med. Child Neurol. Suppl. 109, 8–14 (2007). [PubMed: 17370477]
24. Jin SC et al. Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. Nat. Genet. 49, 1593–1601 (2017). [PubMed: 28991257]
25. Krumm N et al. Excess of rare, inherited truncating mutations in autism. Nat. Genet. 47, 582–588 (2015). [PubMed: 25961944]
26. McKenna A et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303 (2010). [PubMed: 20644199]
27. Van der Auwera GA et al. From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. Curr. Protoc. Bioinformatics 43, 11.10.1–11.10.33 (2013). [PubMed: 25431634]
28. Dong C et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum. Mol. Genet. 24, 2125–2137 (2015). [PubMed: 25552646]
29. Kircher M et al. A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. 46, 310–315 (2014). [PubMed: 24487276]
30. Wei Q et al. A Bayesian framework for de novo mutation calling in parents–offspring trios. Bioinformatics 31, 1375–1381 (2015). [PubMed: 25535243]
31. Karczewski KJ et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 581, 434–443 (2020). [PubMed: 32461654]
32. Rainier S, Sher C, Reish O, Thomas D & Fink JK De novo occurrence of novel SPG3A/atlastin mutation presenting as cerebral palsy. Arch. Neurol. 63, 445–447 (2006). [PubMed: 16533974]
33. Blom N, Gammeltoft S & Brunak S Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J. Mol. Biol. 294, 1351–1362 (1999). [PubMed: 10600390]

34. McNair K et al. A role for RHOB in synaptic plasticity and the regulation of neuronal morphology. J. Neurosci. 30, 3508–3517 (2010). [PubMed: 20203211]
35. Deshaies RJ & Joazeiro CA RING domain E3 ubiquitin ligases. Annu. Rev. Biochem. 78, 399–434 (2009). [PubMed: 19489725]
36. Li Y et al. Structural basis of the phosphorylation-independent recognition of cyclin D1 by the SCF^{FBXO31} ubiquitin ligase. Proc. Natl Acad. Sci. USA 115, 319–324 (2018). [PubMed: 29279382]
37. Vadhvani M, Schwedhelm-Domeyer N, Mukherjee C & Stegmuller J The centrosomal E3 ubiquitin ligase FBXO31-SCF regulates neuronal morphogenesis and migration. PLoS ONE 8, e57530 (2013).
38. Mir A et al. Truncation of the E3 ubiquitin ligase component FBXO31 causes non-syndromic autosomal recessive intellectual disability in a Pakistani family. Hum. Genet. 133, 975–984 (2014). [PubMed: 24623383]
39. Lefevre J et al. The C terminus of tubulin, a versatile partner for cationic molecules: binding of Tau, polyamines, and calcium. J. Biol. Chem. 286, 3065–3078 (2011). [PubMed: 21062741]
40. Hebebrand M et al. The mutational and phenotypic spectrum of TUBA1A-associated tubulinopathy. Orphanet J. Rare Dis. 14, 38 (2019). [PubMed: 30744660]
41. Song DH et al. CK2 phosphorylation of the armadillo repeat region of beta-catenin potentiates Wnt signaling. J. Biol. Chem. 278, 24018–24025 (2003). [PubMed: 12700239]
42. Panagiotou ES et al. Defects in the cell signaling mediator beta-catenin cause the retinal vascular condition FEVR. Am. J. Hum. Genet. 100, 960–968 (2017). [PubMed: 28575650]
43. de Ligt J et al. Diagnostic exome sequencing in persons with severe intellectual disability. N. Engl. J. Med. 367, 1921–1929 (2012). [PubMed: 23033978]
44. Tucci V et al. Dominant beta-catenin mutations cause intellectual disability with recognizable syndromic features. J. Clin. Invest. 124, 1468–1482 (2014). [PubMed: 24614104]
45. Kharbanda M et al. Clinical features associated with *CTNNB1* de novo loss of function mutations in ten individuals. Eur. J. Med. Genet. 60, 130–135 (2017). [PubMed: 27915094]
46. Chen J, Knowles HJ, Hebert JL & Hackett BP Mutation of the mouse hepatocyte nuclear factor/forkhead homologue 4 gene results in an absence of cilia and random left-right asymmetry. J. Clin. Invest. 102, 1077–1082 (1998). [PubMed: 9739041]
47. Orso G et al. Homotypic fusion of ER membranes requires the dynamin-like GTPase atlastin. Nature 460, 978–983 (2009). [PubMed: 19633650]
48. Zhu PP, Denton KR, Pierson TM, Li XJ & Blackstone C Pharmacologic rescue of axon growth defects in a human iPSC model of hereditary spastic paraplegia SPG3A. Hum. Mol. Genet. 23, 5638–5648 (2014). [PubMed: 24908668]
49. Guelly C et al. Targeted high-throughput sequencing identifies mutations in atlastin-1 as a cause of hereditary sensory neuropathy type I. Am. J. Hum. Genet. 88, 99–105 (2011). [PubMed: 21194679]
50. Zhao X et al. Mutations in a newly identified GTPase gene cause autosomal dominant hereditary spastic paraplegia. Nat. Genet. 29, 326–331 (2001). [PubMed: 11685207]
51. Hazan J et al. Spastin, a new AAA protein, is altered in the most frequent form of autosomal dominant spastic paraplegia. Nat. Genet. 23, 296–303 (1999). [PubMed: 10610178]
52. Burger J et al. Hereditary spastic paraplegia caused by mutations in the *SPG4* gene. Eur. J. Hum. Genet. 8, 771–776 (2000). [PubMed: 11039577]
53. Hazan J et al. A fine integrated map of the SPG4 locus excludes an expanded CAG repeat in chromosome 2p-linked autosomal dominant spastic paraplegia. Genomics 60, 309–319 (1999). [PubMed: 10493830]
54. de la Cruz J, Kressler D & Linder P Unwinding RNA in *Saccharomyces cerevisiae*: DEAD-box proteins and related families. Trends Biochem. Sci. 24, 192–198 (1999). [PubMed: 10322435]
55. Della Corte CM et al. Role and targeting of anaplastic lymphoma kinase in cancer. Mol. Cancer 17, 30 (2018). [PubMed: 29455642]
56. Chen Y et al. Oncogenic mutations of ALK kinase in neuroblastoma. Nature 455, 971–974 (2008). [PubMed: 18923524]

57. Janoueix-Lerosey I et al. Somatic and germline activating mutations of the ALK kinase receptor in neuroblastoma. Nature 455, 967–970 (2008). [PubMed: 18923523]
58. Schule R et al. Hereditary spastic paraplegia: clinicogenetic lessons from 608 patients. Ann. Neurol. 79, 646–658 (2016). [PubMed: 26856398]
59. Parodi L et al. Spastic paraplegia due to SPAST mutations is modified by the underlying mutation and sex. Brain 141, 3331–3342 (2018). [PubMed: 30476002]
60. Solowska JM, Rao AN & Baas PW Truncating mutations of SPAST associated with hereditary spastic paraplegia indicate greater accumulation and toxicity of the M1 isoform of spastin. Mol. Biol. Cell 28, 1728–1737 (2017). [PubMed: 28495799]
61. Ji Z et al. Spastin interacts with CRMP5 to promote neurite outgrowth by controlling the microtubule dynamics. Dev. Neurobiol. 78, 1191–1205 (2018). [PubMed: 30257070]
62. Gao Y et al. Atlastin-1 regulates dendritic morphogenesis in mouse cerebral cortex. Neurosci. Res. 77, 137–142 (2013). [PubMed: 23999326]
63. Romeo DM et al. Sex differences in cerebral palsy on neuromotor outcome: a critical review. Dev. Med. Child Neurol. 58, 809–813 (2016). [PubMed: 27098195]
64. Reid SM, Meehan EM, Arnup SJ & Reddihough DS Intellectual disability in cerebral palsy: a population-based retrospective study. Dev. Med. Child Neurol. 60, 687–694 (2018). [PubMed: 29667705]
65. Pinero J et al. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. Nucleic Acids Res. 45, D833–D839 (2017). [PubMed: 27924018]
66. Szklarczyk D et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 47, D607–D613 (2019). [PubMed: 30476243]
67. Al-Mubarak B et al. Whole exome sequencing reveals inherited and de novo variants in autism spectrum disorder: a trio study from Saudi families. Sci. Rep. 7, 5679 (2017). [PubMed: 28720891]
68. Giacopuzzi E et al. Exome sequencing in schizophrenic patients with high levels of homozygosity identifies novel and extremely rare mutations in the GABA/glutamatergic pathways. PLoS ONE 12, e0182778 (2017).
69. Huang da W, Sherman BT & Lempicki RA Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57 (2009). [PubMed: 19131956]
70. Liberzon A et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 1, 417–425 (2015). [PubMed: 26771021]
71. Mi H et al. Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). Nat. Protoc. 14, 703–721 (2019). [PubMed: 30804569]
72. Fang H & Gough J DcGO: database of domain-centric ontologies on functions, phenotypes, diseases and more. Nucleic Acids Res. 41, D536–D544 (2013). [PubMed: 23161684]
73. Novarino G et al. Exome sequencing links corticospinal motor neuron disease to common neurodegenerative disorders. Science 343, 506–511 (2014). [PubMed: 24482476]
74. Stessman HA et al. Targeted sequencing identifies 91 neurodevelopmental- disorder risk genes with autism and developmental-disability biases. Nat. Genet. 49, 515–526 (2017). [PubMed: 28191889]
75. Estes PS et al. Wild-type and A315T mutant TDP-43 exert differential neurotoxicity in a *Drosophila* model of ALS. Hum. Mol. Genet. 20, 2308–2321 (2011). [PubMed: 21441568]
76. Madabattula ST et al. Quantitative analysis of climbing defects in a Drosophila model of neurodegenerative disorders. J. Vis. Exp. 10.3791/52741 (2015).
77. Kim M et al. Mutation in ATG5 reduces autophagy and leads to ataxia with developmental delay. eLife 5, e12245 (2016).
78. Aleman-Meza B, Loeza-Cabrera M, Pena-Ramos O, Stern M & Zhong W High-content behavioral profiling reveals neuronal genetic network modulating *Drosophila* larval locomotor program. BMC Genet. 18, 40 (2017). [PubMed: 28499390]
79. Hemminki K, Li X, Sundquist K & Sundquist J High familial risks for cerebral palsy implicate partial heritable aetiology. Paediatr. Perinat. Epidemiol. 21, 235–241 (2007). [PubMed: 17439532]

80. MacLennan AH et al. Cerebral palsy and genomics: an international consortium. Dev. Med. Child Neurol. 60, 209–210 (2018). [PubMed: 29336076]
81. Himmelmann K & Uvebrant P The panorama of cerebral palsy in Sweden part XII shows that patterns changed in the birth years 2007–2010. Acta Paediatr. 107, 462–468 (2018). [PubMed: 29121418]
82. van Eyk CL et al. Analysis of 182 cerebral palsy transcriptomes points to dysregulation of trophic signalling pathways and overlap with autism. Transl. Psychiatry 8, 88 (2018). [PubMed: 29681622]
83. Martinelli S et al. Functional dysregulation of CDC42 causes diverse developmental phenotypes. Am. J. Hum. Genet. 102, 309–320 (2018). [PubMed: 29394990]
84. Englander ZA et al. Brain structural connectivity increases concurrent with functional improvement: evidence from diffusion tensor MRI in children with cerebral palsy during therapy. Neuroimage Clin. 7, 315–324 (2015). [PubMed: 25610796]
85. Loubet D et al. Neuritogenesis: the prion protein controls beta1 integrin signaling activity. FASEB J. 26, 678–690 (2012). [PubMed: 22038049]
86. Colombo S et al. G protein-coupled potassium channels implicated in mouse and cellular models of GNB1 Encephalopathy. Preprint at bioRxiv 10.1101/697235 (2019).
87. Pipo-Deveza J et al. Rationale for dopa-responsive CTNNB1/β-catenin deficient dystonia. Mov. Disord. 33, 656–657 (2018). [PubMed: 29436745]
88. Akizu N et al. AMPD2 regulates GTP synthesis and is mutated in a potentially treatable neurodegenerative brainstem disorder. Cell 154, 505–517 (2013). [PubMed: 23911318]
89. van Eyk CL et al. Targeted resequencing identifies genes with recurrent variation in cerebral palsy. NPJ Genom. Med. 4, 27 (2019). [PubMed: 31700678]
90. Miller SP, Shevell MI, Patenaude Y & O'Gorman AM Neuromotor spectrum of periventricular leukomalacia in children born at term. Pediatr. Neurol. 23, 155–159 (2000). [PubMed: 11020641]
91. Li H & Durbin R Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics 26, 589–595 (2010). [PubMed: 20080505]
92. Wang K, Li M & Hakonarson H ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38, e164 (2010). [PubMed: 20601685]
93. Lek M et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285–291 (2016). [PubMed: 27535533]
94. 1000 Genomes Project Consortium A global reference for human genetic variation. Nature 526, 68–74 (2015). [PubMed: 26432245]
95. Ware JS, Samocha KE, Homsy J & Daly MJ Interpreting de novo variation in human disease using denovolyzeR. Curr. Protoc. Hum. Genet. 87, 7.25.1–7.25.15 (2015). [PubMed: 26439716]
96. Homsy J et al. De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. Science 350, 1262–1266 (2015). [PubMed: 26785492]
97. Huang da W, Sherman BT & Lempicki RA Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37, 1–13 (2009). [PubMed: 19033363]
98. Mi H, Muruganujan A, Ebert D, Huang X & Thomas PD PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res. 47, D419–D426 (2019). [PubMed: 30407594]
99. Subramanian A et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl Acad. Sci. USA 102, 15545–15550 (2005). [PubMed: 16199517]

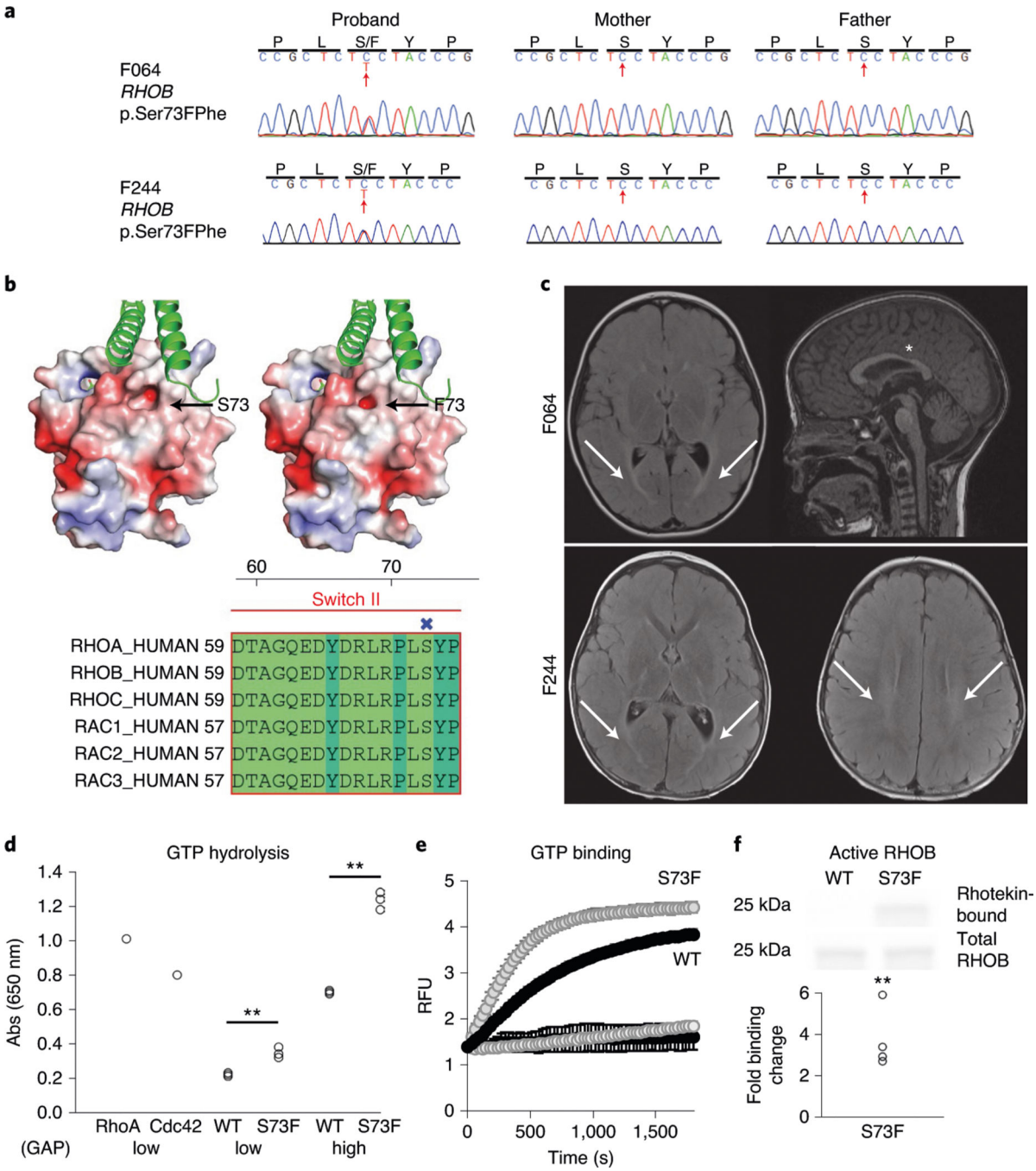


Fig. 1 |. Functional validation of the CP-associated RHOB variant S73F.

a, Sanger traces of the mother, father and proband from families F064 and F244 verify de novo inheritance and the position of the variant (red arrow). **b**, Top: Poisson–Boltzmann electrostatic maps of wild-type RHOB (left) and the F73 variant (right) showing changes to the kinase-binding site (arrow) and the surface charge of the protein. bottom: alignment of human Rho family proteins shows high conservation of the RHOB 73 residue in the Switch II domain. the site of S73/F73 has been labeled (X). **c**, Top: brain MRI from F064 demonstrates bilateral periventricular T2/FLAIR hyperintensity (arrows) on axial imaging

(left), while the sagittal view (right) reveals equivocal thinning of the isthmus of the corpus callosum (asterisk). bottom: MRI from F244 demonstrates T2 hyperintensity of the posterior limb of the internal capsule and optic radiations (arrows; left image) and hyperintensity of the periventricular white matter (arrows; right image). **d**, GTP hydrolysis is enhanced ~1.5-fold in the S73F RHOB variant in a GAP assay. the plot shows absorbance measurements of hydrolyzed GTP in the presence of either a low (5 μg; $P = 0.003$) or high (13 μg; $P = 5.6 \times 10^{-5}$) level of RHOA GAP. there was no change in the endogenous GTPase activity with the S73F variant without GAP added (not shown; $n = 3$). **e**, GTP binding is enhanced in the S73F RHOB variant in a GEF assay. the *N*-methylantraniloyl–GTP fluorophore increases its fluorescence emission when bound to Rho family GTPases, indicating nucleotide uptake by the GTPase. both the wild type and S73F have low endogenous GTP binding (bottom curves). In the presence of the GEF protein Dbs, GTP binding is enhanced, and the Michaelis constant (K_m) of S73F is significantly reduced compared to that of wild-type RHOB ($n = 5$; mean 243 versus 547 s, $P = 0.0017$; top curves). **f**, S73F GTP binding is increased fourfold in a pulldown assay with rhotekin, an interactor with active GTP-bound Rho proteins. Top: a sample western blot cropped to show RHOB from the bead-bound fraction and the total input detected using an antibody against the V5 tag. bottom: quantification of the ratio of rhotekin-bound/total RHOB ($n = 5$), $P = 0.001$. RFU, relative fluorescence units (10^6) at 360 nm excitation. the statistics were determined by a two-tailed unpaired *t*-test. $^{**}P < 0.003$. Full-length blots are provided as source data.

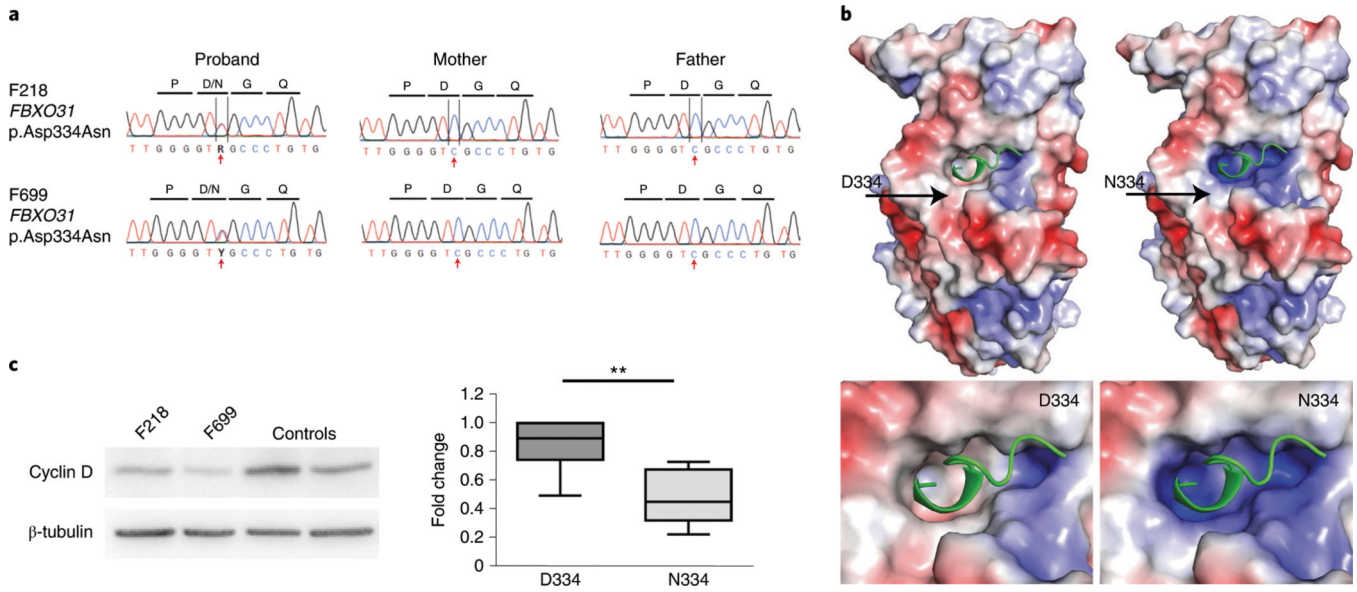
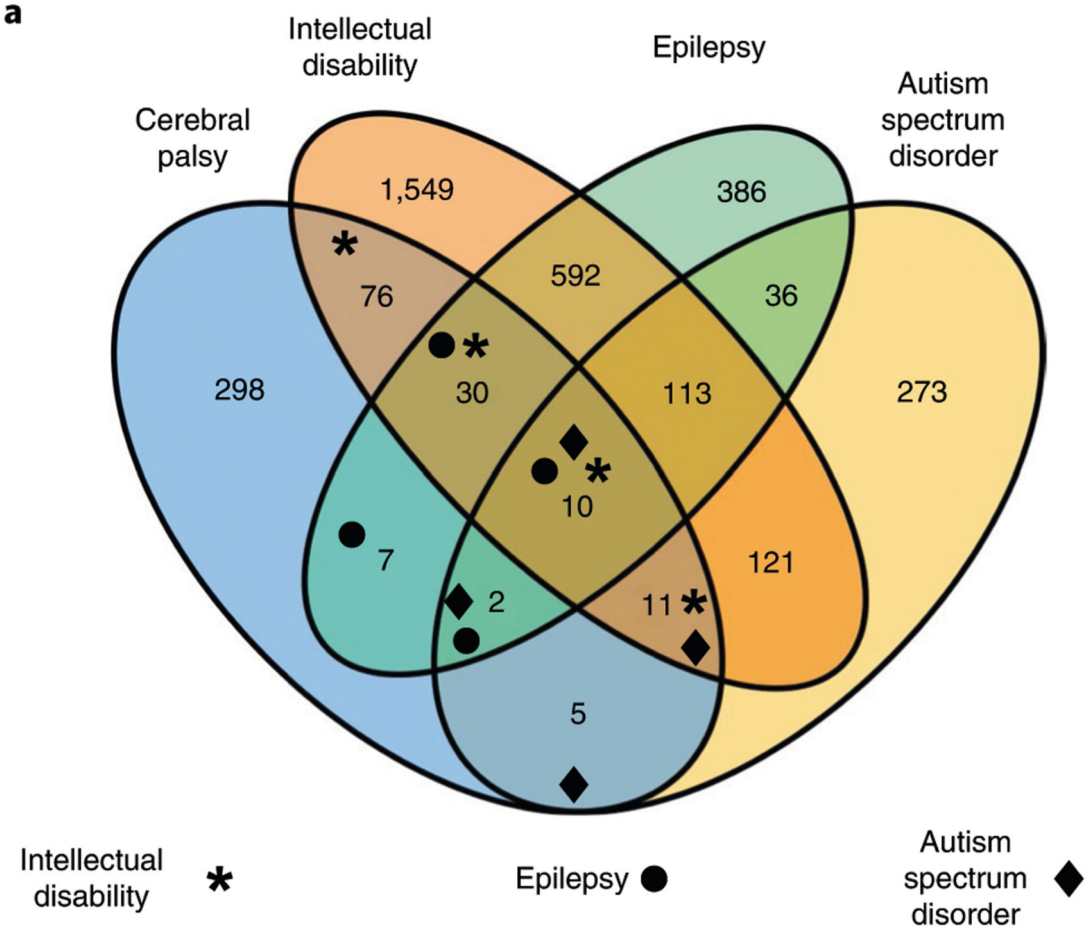


Fig. 2 |. Functional validation of the CP-associated FBXO31 variant p.Asp334Asn shows alterations in cyclin D regulation.

a, Sanger traces of the mother, father and proband from families F218 and F699 verify de novo inheritance and the position of the variant (red arrow). **b**, Poisson–Boltzmann electrostatic maps of wild-type FbXO31 (left) and the p.Asp334Asn variant (right). D334 is positioned around the cyclin D1 (green)-binding pocket on FbXO31. the mutation alters the surface electrostatic charge around the cyclin D1-binding site with a predicted effect on cyclin D1 binding to FbXO31. the site of D334/N334 has been labeled (arrow). the bottom panels are magnified views showing the alterations to the surface charge in the cyclin D1-binding site. **c**, A representative western blot cropped to show the decreased cyclin D expression in patient-derived fibroblasts with the *FBXO31* p.Asp334Asn variant. Quantification of cyclin D is normalized to in-lane β-tubulin and the within-experiment control GMO8398. both patients had reduced cyclin D compared to pooled controls. the data are averaged for three independent cell culture experiments ($n = 7$ controls, $n = 6$ patient measurements). the box indicates the 75th and 25th percentiles with a center line indicating the median; the whiskers indicate the 10th and 90th percentiles. $^{**}P = 0.004$ calculated using a two-tailed unpaired *t*-test. Full-length blots are provided as source data.

a

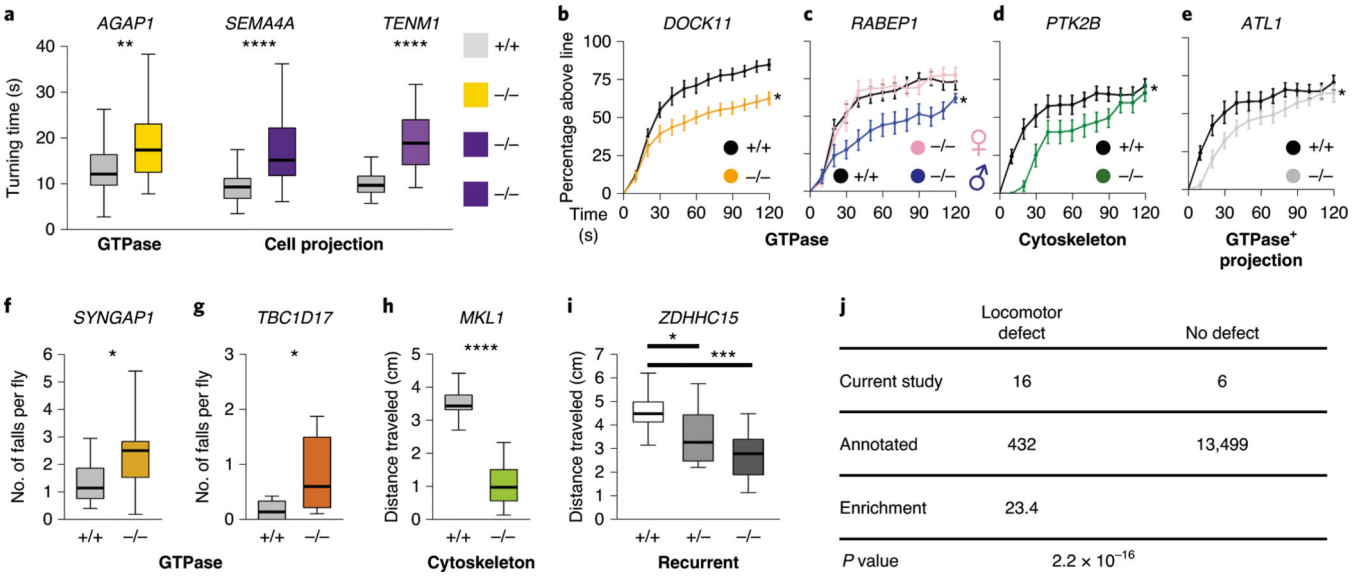


b

Disease	Genes in DisGeNET	CP candidate genes overlapping with DisGeNET	*P* value
ASD	571	28	◆ 1.20×10^{-5}
ID	2,502	127	✱ 2.56×10^{-16}
Epilepsy	1,176	49	● 0.00016
Alzheimer's	1,981	47	0.617

Fig. 3 |. Genetic overlap among common NDDs.

a, A Venn diagram showing the number of overlapping genes between candidate CP genes and genes linked to other NDDs, ID, epilepsy and ASD. CP risk genes were identified as having one or more damaging variants across modes of inheritance with overlap determined using DisGeNET. **b**, Overlap between CP and other NDDs was significant by hypergeometric two-tailed test, while overlap between CP and Alzheimer's disease was not. total number of genes in DisGeNET = 17,549; total number of genes in our gene set = 439.



j	Locomotor defect	No defect
Current study	16	6
Annotated	432	13,499
Enrichment	23.4	
P value	2.2×10^{-16}	

Fig. 4 |. Locomotor phenotypes of LoF mutations in *Drosophila* orthologs of candidate CP risk genes.

a, Turning time, a measure of coordinated movements, is increased in larvae with mutations in *AGAP1*, *SEMA4A* and *TENM1* orthologs. *Drosophila* mutant and control genotypes are provided in Supplementary Table 9. **b**–**i**, 14-day-old adult flies have locomotor impairments. **b**–**e**, Negative geotaxis climbing defects in distance threshold assays for flies with mutations in orthologs of *DOCK11* (**b**), *RABEP1* (**c**), *PTK2B* (**d**) and *ATL1* (**e**). Some genotypes have a male-specific locomotor defect (**c**). **f**,**g**, Increased number of falls for flies with mutations in *SYNGAP1* (**f**) and *TBC1D17* (**g**) orthologs, although the percentage reaching the threshold distance was normal (extended Data Fig. 10). **h**,**i**, Impairments in the average distance traveled of flies with mutations in *MKL1* (**h**) and *ZDHHC15* (**i**) orthologs. related GO terms for genes are shown in bold. For the box and whisker plots, the box indicates the 75th and 25th percentiles with a median line, and the whiskers indicate the 10th and 90th percentiles. the locomotor curve represents the average of all trials and the error bars indicate standard error. $n = 50$ larvae, $n = 10$–21 trials for falls and distance traveled assays, and $n = 10$–21 trials for locomotor curves. the differences in larval turning times, distances traveled and numbers of falls were determined by unpaired two-tailed *t*-tests. the locomotor curves were considered to be significantly different from each other if $P < 0.05$ for a Kolmogorov–Smirnov test in addition to a significant difference at one or more time bins by a Mann–Whitney rank sum two-tailed test. $^{*}P < 0.05$, $^{**}P < 0.005$, $^{***}P < 0.001$, $^{****}P < 1 \times 10^{-6}$. exact genotypes, *n* and *P* values are provided in Supplementary Table 9. **j**, Enrichment of locomotor phenotypes detected in studies of putative CP genes (observed) compared to genome-wide rates annotated in https://flybase.org (expected, 3.1%). The *P* value was calculated by Fisher's exact two-tailed test.

Table 1 |

Significant enrichment of DNMs in CP cases

	Cases, *n* = 250						Controls, *n* = 1,789					
	Observed		Expected		Enrichment	*P*		Observed		Expected		Enrichment	*P*
	n	Rate	*n*	Rate				*n*	Rate	*n*	Rate		
All genes (*n* = 19,347)							**All genes (*n* = 19,347)**						
Total	298	1.19	276.8	1.11	1.08	0.11	Total	1,834	1.03	1,967.2	1.10	0.93	1.00
Synonymous	68	0.27	78.4	0.31	0.87	0.89	Synonymous	484	0.27	557	0.31	0.87	1.00
T-Mis	63	0.25	61.3	0.25	1.03	0.43	T-Mis	410	0.23	431.6	0.24	0.95	0.86
D-Mis	132	0.53	113.1	0.45	1.17	0.04	D-Mis	790	0.44	808.1	0.45	0.98	0.74
LoF	35	0.14	24.1	0.10	1.46	0.02	LoF	150	0.08	170.4	0.10	0.88	0.95
Protein-altering	230	0.92	198.5	0.79	1.16	0.02	Protein-altering	1,350	0.75	1,410.2	0.79	0.96	0.95
Damaging	167	0.67	137.2	0.55	1.22	7.4×10^{-3}	Damaging	940	0.53	978.5	0.55	0.96	0.89
LoF-intolerant genes (gnomAD v2.1.1 pLI ≥ 0.9; *n* = 3,049)							**LoF-intolerant genes (gnomAD v2.1.1 pLI ≥ 0.9; *n* = 3,049)**						
Total	99	0.40	66.4	0.27	1.49	1.1×10^{-4}	Total	456	0.25	473.8	0.26	0.96	0.80
Synonymous	20	0.08	18.7	0.07	1.07	0.41	Synonymous	113	0.06	133.4	0.07	0.85	0.97
T-Mis	13	0.05	10.7	0.04	1.21	0.28	T-Mis	86	0.05	75.5	0.04	1.14	0.13
D-Mis	53	0.21	31.1	0.12	1.70	2.3×10^{-4}	D-Mis	222	0.12	222.7	0.12	1.00	0.53
LoF	13	0.05	5.9	0.02	2.19	8.1×10^{-3}	LoF	35	0.02	42.2	0.02	0.83	0.89
Protein-altering	79	0.32	47.7	0.19	1.66	2.1×10^{-5}	Protein-altering	343	0.19	340.4	0.19	1.01	0.45
Damaging	66	0.26	37.1	0.15	1.78	1.2×10^{-5}	Damaging	257	0.14	264.9	0.15	0.97	0.69

A one-tailed Poisson test was used to test the enrichment of DNMs for each functional class. A marginal enrichment of DNMs was observed for LoF, protein-altering and damaging DNMs. Strikingly, when we restricted our analysis to LoF-intolerant genes, stronger enrichment was observed for protein-altering and damaging DNMs, suggesting a significant contribution of DNMs in this gene set to CP pathogenesis. No enrichment was found in controls. *n*, number of DNMs; rate, number of DNMs divided by the number of individuals in the cohort; enrichment, ratio of observed to expected numbers of mutations; D-Mis, damaging missense mutations as predicted by MetaSVM and cADD algorithms; Protein-altering, missense + LoF; Damaging, D-Mis + LoF.

Table 2 |

Eight genes with two or more damaging (LoF + D-Mis) DNMs

Gene	No. of LoF	No. of D-Mis	Poisson *P* value	pLI	mis_Z
CTNNB1	**3**	**0**	**9.8×10^{-10}**	**1.00**	**3.85**
TUBA1A	**0**	**3**	**4.8×10^{-8}**	**0.97**	**5.58**
RHOB	0	2	7.6×10^{-6}	0.12	2.51
ATL1	0	2	2.0×10^{-5}	0.98	2.63
DHX32	0	2	3.5×10^{-5}	0.00	1.26
SPAST	0	2	3.5×10^{-5}	1.00	1.24
FBXO31	0	2	5.1×10^{-5}	0.44	2.46
ALK	1	1	2.5×10^{-4}	0.00	0.01

A one-tailed Poisson test was performed for damaging and LoF DNMs for each gene independently. The Bonferroni correction for genome-wide significance is 1.3×10^{-6} (= 0.05/(19,347 genes × 2 tests)). pLI, intolerance score for loss-of-function variation; mis_Z, *Z* score for missense constraint.

Table 3 |

Idiopathic CP cases show enrichment of damaging RGs in HSP-associated genes

Gene set (no. of genes)	Observed				Expected	Enrichment	*P*
	Homozygotes	Compound heterozygous	Unique genes	RGs	RGs		
250 CP cases							
All genes (19,347)	63	133	187	196	–	–	–
Recessive known HSP genes (52)	3	3	6	6	0.78	**7.74**	**1.5×10^{-4}**
Known HSP genes (73)	3	3	6	6	0.97	**6.20**	**4.8×10^{-4}**
157 idiopathic cases							
All genes (19,347)	49	89	136	138	–	–	–
Recessive known HSP genes (52)	3	2	5	5	0.54	**9.22**	**2.4×10^{-4}**
Known HSP genes (73)	3	2	5	5	0.68	**7.37**	**6.5×10^{-4}**
84 environmental cases							
All genes (19,347)	14	41	40	55	–	–	–
Recessive known HSP genes (52)	0	1	1	1	0.22	4.48	0.20
Known HSP genes (73)	0	1	1	1	0.28	3.60	0.24
1,789 controls							
All genes (19,347)	81	687	610	768	–	–	–
Recessive known HSP genes (52)	0	3	3	3	2.46	1.22	0.45
Known HSP genes (73)	0	3	3	3	2.94	1.02	0.56

The expected number of recessive genotypes was determined on the basis of fitted values from the polynomial regression model by using the damaging de novo probabilities. *P* values were calculated by using the one-tailed binomial probability. Values in bold are *P* values exceeding the Bonferroni multiple-testing cutoff ($0.05/(3 \times 4) = 4.2 \times 10^{-3}$).

Table 4 |

CP risk gene pathway enrichment

	Term names and IDs	Overlap per set	Observed	Expected	FDR
Database					
DAVID	Non-integrin membrane–ECM interactions (R-HSA-3000171)	10/40	10/218	40/9,075	0.00045
	Laminin interactions (R-HSA-3000157)	8/30	8/218	30/9,075	0.0075
	ECM–receptor interaction (R-HSA-04512)	12/87	12/168	87/6,879	0.00888
PANTHER	Non-integrin membrane–ECM interactions (R-HSA-3000171)	12/59	12/447	59/20,851	6.02×10^{-5}
	Laminin interactions (R-HSA-3000157)	8/30	8/447	30/20,851	0.00114
	Signaling by Rho GTPases (R-HSA-194315)	24/408	24/447	408/20,851	0.00721
	Extracellular matrix organization (R-HSA-1474244)	20/299	20/447	299/20,851	0.00826
MSigDB	Non-integrin membrane–ECM interactions (R-HSA-30000171)	12/59	12/439	59/38,055	5.53×10^{-9}
	Laminin interactions (R-HSA-30000157)	8/30	8/439	30/38,055	4.65×10^{-7}
	Signaling by Rho GTPases (R-HSA-194315)	26/450	26/439	450/38,055	2.15×10^{-8}
	Extracellular matrix organization (R-HSA-1474244)	20/301	20/439	301/38,055	1.97×10^{-7}
Biological processes	**Cell projection and organization**				
	Regulation of cell projection organization (GO:0031344)	47/695	47/447	695/20,851	1.35×10^{-7}
	Positive regulation of cell projection organization (GO:0031346)	30/395	30/447	395/20,851	1.84×10^{-5}
	Positive regulation of neuron projection development (GO:0010976)	19/294	19/447	294/20,851	0.0087
	Microtubule-based movement				
	Movement of cell or subcellular component (GO:0006928)	66/1,544	66/447	1,544/20,851	1.22×10^{-4}
	Microtubule-based process (GO:0007017)	32/667	32/447	667/20,851	0.00875
	Microtubule-based movement (GO:0007018)	18/271	18/447	271/20,851	0.00966
Cell components	**Axonal cell projection**				
	Plasma membrane bounded cell projection part (GO:0120025)	89/2,197	89/447	2,197/20,851	7.15×10^{-6}
	Axon (GO:0030424)	34/641	34/447	641/20,851	0.000882
	Actin-based cell projection (GO:0098858)	15/214	15/447	214/20,851	0.00898
	Microtubule-associated components				
	Cytoskeleton (GO:0005856)	82/2,274	82/447	2,274/20,851	0.000894

	Term names and IDs	Overlap per set	Observed	Expected	FDR
	Microtubule cytoskeleton (GO:0015630)	46/1,246	46/447	1,246/20,851	0.0213
	Microtubule associated complex (GO:0005875)	11/154	11/447	154/20,851	0.0302
Molecular functions	**GTPase activity**				
	Small GTPase binding (GO:0031267)	28/421	28/421	421/20,851	7.98×10^{-5}
	Rho GTPase binding (GO:0017048)	19/145	19/447	145/20,851	9.63×10^{-7}
	GTPase regulator activity (GO:0030695)	18/307	18/447	307/20,851	0.0211
	Actin cytoskeleton regulation				
	Extracellular matrix structural constituent (GO:0005201)	15/165	15/447	165/20,851	0.00108
	Actin binding (GO:0003779)	23/443	23/447	443/20,851	0.0207

Key pathways and terms overlapping among DAVID, PANTHER and MSigDB bioinformatics tools. PANTHER GO terms include cell projections, cytoskeleton and Rho GTPase signaling. GO terms were extracted from the total set (Supplementary Datasets 6–15) using hierarchical nesting, or functions that were represented by multiple GO terms. Overlap per set refers to the number of genes overlapping between CP risk genes and a given database term/the total number of genes in the database for that term. FDR = *q* value (FDR cutoff = 0.05) from two-tailed Fisher and hypergeometric tests. FDR differences are due to differences in tool methodologies.

Nat Genet. Author manuscript; available in PMC 2022 May 29.