bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Title:

In-depth analysis reveals complex molecular etiology of idiopathic cerebral palsy

Authors:

Na Li [1, *], Pei Zhou [1, *], Hongmei Tang [2, *], Lu He [2, *], Xiang Fang [1, *], Jinxiang Zhao [3], Xin Wang [3], Yifei Qi [4], Chuanbo Sun [1], Yunting Lin [5], Fengying Qin [1], Miaomiao Yang [1], Zhan Zhang [1], Caihua Liao [1], Shuxin Zheng [1], Xiaofang Peng [1], Ting Xue [1], Qianying Zhu [1], Yan Li [1], Liru Liu [2], Jingyu Huang [2], Li Liu [5], Changgeng Peng [6], Dingding Han [1, #], Dong Liu [3, #], Kaishou Xu [2, #], (Cougar) Hao Hu [1, 7, 8, 9, #]

Affiliation:

1. Laboratory of Medical Systems Biology, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, 510623, Guangzhou, China
2. Department of Rehabilitation, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, 510120, Guangzhou, China
3. Key Laboratory of Neuroregeneration of Jiangsu and Ministry of Education, Nantong University, 226001, Nantong, China
4. Division of Uterine Vascular Biology, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, 510623, Guangzhou, China
5. Department of Genetics and Endocrinology, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, 510623, Guangzhou, China
6. The First Rehabilitation Hospital of Shanghai, Tongji University School of Medicine, 200029, Shanghai, China
7. Guangdong Provincial Key Laboratory of Research in Structural Birth Defect Disease, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, 510623, Guangzhou, China
8. Third Affiliated Hospital of Zhengzhou University, 450052, Zhengzhou, China
9. School of Medicine, South China University of Technology, 510006, Guangzhou, China

[*] Equal contribution

[#] Co-supervisor

Contact should be made to (Cougar) Hao Hu (ORCID: https://orcid.org/0000-0002-3567-609X) via huh@cougarlab.org, or Kaishou Xu via xksyi@126.com, or Dong Liu via liudongtom@gmail.com, or Dingding Han via handd@cougarlab.org.

Abstract:

Cerebral palsy (CP), the most prevalent physical disability in children, has long been ignored with regard to its inherent molecular mechanism. In this work, we performed in-depth clinical and molecular analysis on 120 idiopathic CP families, and identified in half of the patients the underlying risk factors. By a compilation of 117 CP-related genes, we recognized the characteristic features in terms of inheritance and function, and proposed a dichotomous classification system according to the expression patterns. In two patients with both CP and intellectual disability, we revealed that the defective TYW1, a tRNA hypermodification enzyme,

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

caused microcephaly and problems in motion and cognition by hindering neuronal proliferation and migration. We developed an algorithm and proved in mice brain that this malfunctioning hypermodification specifically perturbed the translation of a subset of proteins involved in cell cycling. In a CP patient with normal intelligence, we identified a mitochondrial enzyme GPAM, the hypomorphic form of which led to hypomyelination of corticospinal tract. We confirmed that the aberrant Gpam in mice perturbed the lipid metabolism in astrocyte, resulting in suppressed astrocyte proliferation and a shortage of lipid contents supplied for oligodendrocyte myelination. This work broadened the scope of understanding of CP etiology and paved a way for innovative therapeutic strategies.

Introduction:

Cerebral palsy (CP) is an umbrella term spanning a group of disorders with compromised ambulant performances, which is attributed to non-progressive disturbance in the developing fetal and infant brains [1]. CP can be categorized according to motor signs (spasticity, dyskinesia, ataxia, and hypotonia, among which spasticity manifests in ~80% of CP cases), extremities involvement (hemiplegia, monoplegia, diplegia, triplegia, and quadriplegia), and anatomical sites of the brain lesion (cerebral cortex, pyramidal tract, extrapyramidal system, and cerebellum) [2,3]. In CP children 2 years or older, severity is reliably classified by using the 5-level Gross Motor Function Classification System (GMFCS) [4]. The common comorbidities of CP include intellectual disability (ID), epilepsy, sensorial impairments, language deficit, autism spectrum disorders (ASD), sleep disorders, and secondary musculoskeletal defects [2,3]. CP is the most prevalent physical disability of childhood, with a prevalence of 2-3.5 per 1,000 live births in countries with varying socioeconomic development (more than 20 million patients around the world), and the incidence of CP at term remains consistent for 50 years [1-3,5-7]. CP poses a critical problem for rehabilitation and welfare systems, and the lifetime costs for a CP child in USA are estimated to be $1 million [5]. The conventional risk factors of CP are believed to be gestational and perinatal insults, including birth asphyxia, neonatal infections, and teratogens [1,5]. However, two thirds of CP patients are born at term, and birth asphyxia happens in less than 10% of CP cases [2,6,7]. Because the causality of CP is not confirmed in 80% of cases, it is reasonable that there exists uncharted mechanism accounting for the majority of CP events [3].

A growing body of evidence indicates prevailing genetic contributors to the etiology of CP, including the familial cases, the high concordance rate of monozygotic twins, the association with increased parental ages, and the prevalence of congenital malformation in CP patients [8-12]. Large cohorts of case-control studies assessing polymorphism in a variety of candidate genes did not identify any significantly associated variants [2,13,14]. Therefore, the genetic underpinnings of CP are likely similar to that of other neurodevelopmental disorders (NDD) such as ID and ASD, where causative variants are rarely detected in association studies but with large effect size [15-17]. In line with this hypothesis, clinically relevant copy number variation (CNV), both *de novo* and inherited, were recently identified in four collections of CP patients, resolving 4% - 31% of the cryptogenic cases [18-21]. Meanwhile, gene-panel sequencing or whole-exome sequencing revealed pathogenic variants, mostly *de novo*, in an aggregate of five CP studies, yielding molecular diagnosis for ~16% of cases [22-26].

So far, the genetic and molecular information on CP etiology has been insufficient, which

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

impedes an improvement in prophylaxis, prognosis, and treatment of CP patients in the era of precision medicine. Large cohorts need to be recruited to fully decipher the gene spectrum due to the high genetic heterogeneity of CP, and geographic and ethnic bias should be addressed, since most of the reported studies were conducted in the Caucasian populations. Whole-genome sequencing, instead of gene-panel sequencing or whole-exome sequencing, can render better coverage of predisposing alleles, and somatic mosaicism or post-zygotic mutations (PZM) in CP patients must be highlighted as in other NDDs [27,28]. An oriented investigation into the common characteristics of CP-related genes, plus detailed mechanistic studies, seems worthwhile, before the introduction of a gene-based classification system that will have profound impact on the clinical practice.

Results:

CP cohort recruitment, medical evaluation, and genetic screening

The 120 CP families, including 118 trios and 2 quartets, were recruited by the Department of Rehabilitation of the Guangzhou Women and Children's Medical Center (GWCMC) from August 2014 to June 2017 in a cohort study. This study was approved by the ethic committee of GWCMC, and the written informed consents were signed by the participating families. The following inclusion criteria were adopted: (1) a diagnosis of CP made by a board of pediatric neurologists, typically based on the early-onset disabling non-progressive pyramidal and/or extrapyramidal signs; (2) an age of 2 years and older; (3) uneventful pregnancy and delivery (most of the CP patients were born at term). The following exclusion criteria were implemented: (1) perinatal insults including asphyxia, head trauma, encephalitis, and brain tumor; (2) late-onset hereditary spastic paraplegia (HSP) and DOPA-responsive dystonia (DRD), due to the distinct progression trajectory and featured fluctuation of manifestation; (3) maternal infections such as rubella and cytomegalovirus; (4) hyperbilirubinemia. The classification and summary of the CP cohort were shown in Figure 1 (A) and (B). As expected, the majority of cohort patients belonged to spastic CP, followed by dyskinetic and ataxic CP. Patients with severe motor impairment (GMFCS level IV and V) were less than those mildly disabled in motion (GMFCS level I - III). There was a significant bias for male in this CP cohort, as observed previously in ID and ASD cohorts [29]. The prevalent comorbidities included ID/ASD and microcephaly. More details of the CP patients can be found in Supplementary table 1 and Supplementary figure 1.

As shown in Figure 1 (C) and Supplementary table 2, we initiated a comprehensive genetic assessment by multiple platforms, including targeted PCR for fragile-X, mitochondrial genome sequencing (6000X), high-density cytogenetic microarray, whole-exome sequencing (600X), and whole-genome sequencing (36X). In principle, we defined the disease-related variants by adopting the ACMG recommended standards, reinforced by an in-house variation database of 2,247 ethnic-matched unaffected individuals, and another database (https://db.cngb.org/cmdb/) of >50,000 individuals of East Asian descent [30,31].

In sum, we identified CP-related candidate genes and variants in 55 families (~45% of the 120 CP families), a comparable rate with that of other NDDs [16,32-35]. Interestingly, the unexpected abundance of recessive variants (about a quarter) in the candidate genes, was notably higher than that of the previous NDDs investigations in outbred populations [33,36,37]. Whether this was attributed to the specific genetic structure of ethnic group in this study, or due to the

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

characteristic etiology of CP *per se*, remained to be solved in future study [38].

Of the 82 CP-related variants identified in this cohort, germline point substitution took up ~72%, followed by germline indels (~17%) and CNVs (~5%) (Figure 1 (D)). Post-zygotic mutations (PZMs) were for the first time described in a CP cohort, in which ~6% of the CP-related variants were classified as PZMs. This finding met with the expectation derived from other NDDs studies, and filled a gap in the understanding of NDDs etiology [27].

A closer look at the phenotype-genotype correlation in the patients revealed that: patients with more severe physical impairment (GMFCS level IV or V) had a higher chance to harbor CP-related variants (P-value=0.00613, Fisher exact test), than those patients less disabled in movement (GMFCS level I - III); patients with comorbidity of ID were more likely to have deleterious variants (P-value=0.01065, Fisher exact test), compared to those without ID (Figure 1 (E)). We also noticed the raised diagnostic rates in the quadriplegia patients and the patients with poor rehabilitation outcome, although the statistical powers were not strong enough (P-value=0.09919 and 0.1884, respectively, Fisher exact test).

In this cohort, we found altogether 129 protein-changing *de novo* mutations (DNMs) in 117 CP patients, significantly higher than the in-house 2,247 ethnic-matched unaffected individuals (P-value=0.00375, unpaired t-test). Paternal ages were revealed to correlate positively with DNM counts of the patients (r=0.63, P=0.025, Pearson correlation test). We also detected 27 PZMs in 66 CP patients, but no significant correlation with paternal or maternal ages was established. When counting only private mutations of the CP families as the data input of a burden analysis, we concluded that the 146 private mutations found in 117 families were significantly higher than what we observed in the in-house 2,247 ethnic-matched controls (P-value=0.00126, unpaired t-test).

The details of disease-related variants identified in this study are shown in Table 1 and Supplementary table 3.

Figure 1: CP cohort study information. **(A)** Classification of the cohort patients based on motor signs, anatomical distribution, and motor severity. **(B)** Summary of the cohort patients in terms of family information, gender, delivery, neuroimaging, and comorbidities. **(C)** Project schematics, including cohort recruitment, genetic assessment, gene-oriented bioinformatics and functional analysis, and gene-based clinical practice. **(D)** Inheritance modes and disease-related variants identified in this study. **(E)** Correlations between diagnostic rate and clinical parameters. ID+: with intellectual disability. ID-: without intellectual disability. Microcephaly+: with microcephaly. Microcephaly-: without microcephaly.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

WGS / WES:

Patient ID	Gender	CP Subtype	GMFCS	Comorbidity	Gene	OMIM	Transcript	Mutation	Inheritance Mode	Known Disease	ACMG Classification	Note
CP_004_1	Male	Dyskinetic	IV	Intellectual disability	PURA	600473	NM_005859.4	c.683del, p.N228Tfs*19	Autosomal dominant	Mental retardation, autosomal dominant	Pathogenic	
CP_008_1	Male	Spastic quadriplegia	III	Epilepsy, intellectual disability, vision problem, microcephaly	COL4A1	120130	NM_001845.5	c.3629G>T, p.G1210V	Autosomal dominant	Brain small vessel disease with or without ocular anomalies, autosomal dominant	Pathogenic	
CP_022_1	Male	Dyskinetic	I	Vision problem	ZSWIM6	615951	NM_020928.1	c.492_493del,p.A165Rfs*88;c.495_519del,p.A166Pfs*46	Autosomal dominant	Neurodevelopmental disorder with movement abnormalities, abnormal gait, and autistic features, autosomal dominant	Pathogenic	
CP_027_1	Female	Hypotonia	III	Intellectual disability, macrocephaly	PPP2R5D	601646	NM_001270476.1	c.139G>A,p.E47K	Autosomal dominant	Mental retardation, autosomal dominant	Pathogenic	
CP_031_1	Male	Spastic diplegia	I	Intellectual disability	EFTUD2	603892	NM_001258353.1	c.2420A>G, p.Q807R	Autosomal dominant	Mandibulofacial dysostosis, Guion-Almeida type, autosomal dominant	Pathogenic	
CP_033_1	Male	Spastic hemiplegia	I	Epilepsy, intellectual disability	COL4A1	120130	NM_001845.5	c.145-1_150del, p.?	Autosomal dominant	Brain small vessel disease with or without ocular anomalies, autosomal dominant	Pathogenic	
CP_039_1	Female	Spastic quadriplegia	IV	Intellectual disability, microcephaly	KIF1A	601255	NM_001244008.1	c.506G>C, p.R169T	Autosomal dominant	Mental retardation, autosomal dominant	Pathogenic	
CP_050_1	Male	Spastic diplegia	II	None	SPAST	604277	NM_014946.3	c.1340T>G, p.L447W	Autosomal dominant	Spastic paraplegia, autosomal dominant	Pathogenic	
CP_053_1	Male	Spastic hemiplegia	I	None	CNNM2	607803	NM_017649.4	c.1686A>C, p.E562D	Autosomal dominant	Hypomagnesemia, seizures, and mental retardation, autosomal dominant	Pathogenic	
CP_055_1	Female	Spastic hemiplegia	II	Intellectual disability, microcephaly	TUBA1A	602529	NM_001270399.1	c.445T>G, p.F149V	Autosomal dominant	Lissencephaly, autosomal dominant	Pathogenic	
CP_061_1	Female	Spastic hemiplegia	I	None	TUBB2B	612850	NM_178012.4	c.799A>G, p.M267V	Autosomal dominant	Cortical dysplasia, complex, with other brain malformations, autosomal dominant	Likely pathogenic	Post-zygotic mutation
CP_066_1	Male	Spastic quadriplegia	IV	Epilepsy, intellectual disability, microcephaly	FOXG1	164874	NM_005249.4	c.506del, p.G169Afs*23	Autosomal dominant	Rett syndrome, congenital variant, autosomal dominant	Pathogenic	
CP_067_1	Male	Spastic quadriplegia	V	Epilepsy, intellectual disability, microcephaly	KIF1A	601255	NM_001244008.1	c.642C>A, p.S214R	Autosomal dominant	Mental retardation, autosomal dominant	Pathogenic	
CP_083_1	Male	Dyskinetic	III	Intellectual disability, microcephaly	CTNNB1	116806	NM_001904.3	c.824del,p.L275*	Autosomal dominant	Neurodevelopmental disorder with spastic diplegia and visual defects, autosomal dominant	Pathogenic	
CP_094_1	Male	Spastic diplegia	I	Intellectual disability, microcephaly	KIF1A	601255	NM_001244008.1	c.647G>A, p.R216H	Autosomal dominant	Mental retardation, autosomal dominant	Pathogenic	
CP_095_1	Male	Spastic diplegia	III	None	SPAST	604277	NM_014946.3	c.1496G>A, p.R499H	Autosomal dominant	Spastic paraplegia, autosomal dominant	Pathogenic	
CP_096_1	Female	Mixed	V	Intellectual disability	BCL11B	606558	NM_138576.3	c.1186C>T, p.Q396*	Autosomal dominant	Intellectual developmental disorder with dysmorphic facies, speech delay, and T-cell abnormalities, autosomal dominant	Pathogenic	
CP_107_1	Male	Dyskinetic	II	Intellectual disability	BCL11A	606557	NM_022893.3	c.1079T>C, p.L360P	Autosomal dominant	Dias-Logan syndrome, autosomal dominant	Likely pathogenic	Post-zygotic mutation

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

WGS / WES:

Patient ID	Gender	CP Subtype	GMFCS	Comorbidity	Gene	OMIM	Transcript	Mutation	Inheritance Mode	Known Disease	ACMG Classification	Note
CP_003_1	Male	Spastic hemiplegia	I	None	PCLO	604918	NM_033026.5	c.11945G>T,p.R3982L;c.9731G>A,p.R3244K;c.4314_4316del,p.H1439del	Autosomal recessive	Pontocerebellar hypoplasia, autosomal recessive	Likely pathogenic	
CP_005_1	Male	Spastic diplegia	II	None	CYP2U1	610670	NM_183075.2	c.200C>T,p.P67L;c.471del,p.I158Sfs*2	Autosomal recessive	Spastic paraplegia, autosomal recessive	Pathogenic	
CP_007_1	Male	Spastic quadriplegia	V	Intellectual disability, epilepsy, vision problem	RARS2	611524	NM_020320.4	c.1708C>T,p.L570F;c.282_285del,p.R94Sfs*3	Autosomal recessive	Pontocerebellar hypoplasia, autosomal recessive	Pathogenic	
CP_012_1	Male	Ataxia	III	Intellectual disability, microcephaly	TYW1	611243	NM_018264.3	c.616C>T,p.R206C;c.1166G>A,p.R389Q	Autosomal recessive	N.A.	Novel	
CP_012_2	Female	Mixed	IV	Intellectual disability, microcephaly	TYW1	611243	NM_018264.3	c.616C>T,p.R206C;c.1166G>A,p.R389Q	Autosomal recessive	N.A.	Novel	
CP_014_1	Male	Spastic hemiplegia	I	None	TARS	187790	NM_001258438.1	c.772A>T,p.T258S;c.2105A>G,p.Q702R	Autosomal recessive	N.A.	Novel	
CP_035_1	Female	Dyskinetic	II	Intellectual disability	MAG	159460	NM_080600.2	c.960_961del,p.S321Cfs*66;c.1441C>T,p.Q481*	Autosomal recessive	Spastic paraplegia, autosomal recessive	Pathogenic	
					PSAT1	610936	NM_058179.3	c.370G>A,p.V124I;c.916C>T,p.R306C	Autosomal recessive	Phosphoserine aminotransferase deficiency, autosomal recessive	Likely pathogenic	
CP_036_1	Male	Spastic diplegia	II	None	ALS2	606352	NM_020919.3	c.4853G>A,p.G1618D	Autosomal recessive	Amyotrophic lateral sclerosis, early-onset, autosomal recessive	Pathogenic	
CP_042_1	Male	Spastic quadriplegia	III	None	POMGNT1	606822	NM_001243766.1	c.1449_1454dup,p.Q484_R485dup;c.796C>T,p.R266W	Autosomal recessive	Muscular dystrophy-dystroglycanopathy, autosomal recessive	Pathogenic	
CP_044_1	Female	Spastic diplegia	III	Intellectual disability, microcephaly	CYP2U1	610670	NM_183075.2	c.1451G>A,p.G484E	Autosomal recessive	Spastic paraplegia, autosomal recessive	Pathogenic	
CP_049_1	Female	Spastic diplegia	II	None	DDHD2	615003	NM_015214.2	c.373dup,p.Y125Lfs*24;c.1546C>T,p.R516*	Autosomal recessive	Spastic paraplegia, autosomal recessive	Pathogenic	
CP_051_1	Female	Spastic quadriplegia	IV	Intellectual disability, microcephaly	WDR62	613583	NM_001083961.1	c.643A>C,p.T215P;c.1535G>A,p.R512Q	Autosomal recessive	Primary microcephaly, autosomal recessive	Pathogenic	
CP_052_1	Female	Spastic diplegia	I	None	PTK7	601890	NM_002821.4	c.1426G>A,p.G476R;c.2093C>G,p.P698R	Autosomal recessive	N.A.	Novel	
CP_057_1	Male	Spastic hemiplegia	III	Intellectual disability, microcephaly	PIGT	610272	NM_015937.5	c.594+2T>A,p.?;c.1363G>C,p.E455Q	Autosomal recessive	Multiple congenital anomalies-hypotonia-seizures syndrome, autosomal recessive	Pathogenic	
CP_063_1	Female	Spastic quadriplegia	IV	None	GPAM	602395	NM_001244949.1	c.2005C>T,p.P669S;c.1495G>C,p.G499R	Autosomal recessive	N.A.	Novel	
CP_064_1	Male	Spastic quadriplegia	IV	Intellectual disability	PTS	612719	NM_000317.2	c.155A>G,p.N52S;c.259C>T,p.P87S	Autosomal recessive	Hyperphenylalaninemia, BH4-deficient, autosomal recessive	Pathogenic	
CP_075_1	Female	Dyskinetic	III	Intellectual disability, microcephaly	CDK5RAP	608200	NM_016408.3	c.1481A>T,p.E494V;c.764T>C,p.M255T	Autosomal recessive	N.A.	Novel	
CP_091_1	Male	Spastic quadriplegia	IV	None	AMPD1	102770	NM_000036.2	c.1783C>T,p.R595*;c.1373G>A,p.R458H	Autosomal recessive	Myopathy due to myoadenylate deaminase deficiency, autosomal recessive	Likely pathogenic	
					CIT	605629	NM_001206999.1	c.1876A>G,p.K626E;c.382A>T,p.M128L	Autosomal recessive	Primary microcephaly, autosomal recessive	Likely pathogenic	
CP_098_1	Male	Dyskinetic	II	Dyslexia, microcephaly	ATP8A2	605870	NM_016529.5	c.102T>G,p.Y34*;c.2780C>G,p.T927S	Autosomal recessive	Cerebellar ataxia, mental retardation, and disequilibrium syndrome, autosomal recessive	Pathogenic	p.Y34* is a post-zygotic mutation; p.T927S is inherited from father

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

WGS / WES:

Patient ID	Gender	CP Subtype	GMFCS	Comorbidity	Gene	OMIM	Transcript	Mutation	Inheritance Mode	Known Disease	ACMG Classification	Note
CP_099_1	Male	Spastic quadriplegia	V	Epilepsy, intellectual disability	TREX1	606609	NM_016381.5	c.459dup, p.C154Mfs*3	Autosomal recessive	Aicardi-Goutieres syndrome, autosomal recessive	Pathogenic	
CP_105_1	Male	Ataxia	II	Intellectual disability, microcephaly	SQSTM1	601530	NM_001142298.1	c.571_572del, p.S191Ffs*17	Autosomal recessive	Neurodegeneration with ataxia, dystonia, and gaze palsy, childhood-onset, autosomal recessive	Pathogenic	
CP_110_1	Male	Dyskinetic	I	Intellectual disability, hearing problem	RARS2	611524	NM_020320.4	c.916G>A, p.D306N	Autosomal recessive	Pontocerebellar hypoplasia, autosomal recessive	Pathogenic	
CP_113_1	Female	Dyskinetic	I	None	HPCA	142622	NM_002143.2	c.209G>C, p.R70P	Autosomal recessive	Dystonia, torsion, autosomal recessive	Pathogenic	
CP_026_1	Female	Spastic quadriplegia	IV	Intellectual disability	WDR45	300526	NM_007075.3	c.400C>T, p.R134*	X-linked dominant	Neurodegeneration with brain iron accumulation, X-linked dominant	Pathogenic	
CP_108_1	Male	Mixed	IV	Epilepsy, intellectual disability	SMC1A	300040	NM_001281463.1	c.3402T>G, p.D1134E	X-linked dominant	Cornelia de Lange syndrome, X-linked dominant	Pathogenic	Post-zygotic mutation
CP_009_1	Male	Mixed	IV	Intellectual disability, epilepsy, microcephaly	SLC6A8	300036	NM_005629.3	c.1516G>A, p.D506N	X-linked recessive	Cerebral creatine deficiency syndrome, X-linked recessive	Likely pathogenic	
CP_010_1	Male	Spastic hemiplegia	I	None	FLNA	300017	NM_001456.3	c.3872C>T, p.P1291L	X-linked recessive	Intestinal pseudoobstruction, neuronal, X-linked recessive	Likely pathogenic	
CP_024_1	Male	Dyskinetic	IV	Intellectual disability, microcephaly	ATRX	300032	NM_000489.5	c.736C>T, p.R246C	X-linked recessive	Mental retardation-hypotonic facies syndrome, X-linked recessive	Pathogenic	
CP_029_1	Male	Spastic hemiplegia	I	None	FLNA	300017	NM_001456.3	c.1771G>A, p.V591I	X-linked recessive	Intestinal pseudoobstruction, neuronal, X-linked recessive	Likely pathogenic	
CP_069_1	Male	Dyskinetic	IV	Intellectual disability	HPRT1	308000	NM_000194.2	c.166G>T,p.E56*	X-linked recessive	Lesch-Nyhan syndrome, X-linked recessive	Pathogenic	
CP_077_1	Male	Spastic hemiplegia	I	Intellectual disability	PQBP1	300463	NM_005710.2	c.265G>T,p.A89S	X-linked recessive	Renpenning syndrome, X-linked recessive	Pathogenic	
CP_090_1	Male	Spastic diplegia	III	None	UBA1	314370	NM_003334.3	c.3085C>T, p.R1029C	X-linked recessive	Spinal muscular atrophy, X-linked 2, infantile, X-linked recessive	Pathogenic	
					PCDH19	300460	NM_001184880.1	c.1733C>G, p.P578R	X-linked recessive	Epileptic encephalopathy, early infantile, X-linked recessive	Likely pathogenic	
CP_102_1	Male	Ataxia	III	None	TAF1	313650	NM_004606.4	c.4880A>G, p.E1627G	X-linked recessive	Syndromic mental retardation, X-linked recessive	Likely pathogenic	Post-zygotic mutation

CMA-based CNV:

Patient ID	Gender	CP Subtype	GMFCS	Comorbidity	Microarray Nomenclature	Size (kb)	Gene Count / OMIM Gene Count	Probable Disease-related Gene	Inheritance Mode	Known Disease	ACMG Classification
CP_011_1	Male	Spastic quadriplegia	IV	Intellectual disability, hearing problem, microcephaly	arr[hg19]17p13.1(7,085,366-7,561,445)x1	476	47 / 35	*DLG4, PHF23, CTDNEP1, YBX2, NEURL4, NLGN2, ZBTB4 , POLR2A, TNFSF12, SENP3, EIF4A1, FXR2*	De novo	Intellectual disability, microcephaly, dysmorphism	Pathogenic
CP_046_1	Female	Spastic hemiplegia	I	Vision problem	arr[hg19]12p13.33(464,124-885,041)x3	421	7 / 4	*KDM5A, WNK1*	De novo	Intellectual disability	Likely pathogenic
CP_086_1	Female	Spastic diplegia	I	Intellectual disability, vision problem	arr[hg19]16p13.11(14,900,072-16,533,890)x3	1,634	45 / 11	*KIAA0430, MYH11*	De novo	Intellectual disability, autism spectrum disorder	Pathogenic
CP_119_1	Male	Ataxia	III	Intellectual disability, dyslexia	arr[hg19]Xq28(152,516,781-152,917,379)x2	401	10 / 6	*HAUS7, BGN, ATP2B3*	Inherited from mother	Intellectual disability, behavioral problem, movement problem	Likely pathogenic

Mitochondrial genome:

Patient ID	Gender	CP Subtype	GMFCS	Comorbidity	Gene	Transcript	Mutation	Known Disease	Abundance	ACMG Classification
CP_006_1	Male	Spastic hemiplegia	I	None	MT-TRNL1	NC_012920	m.3243A>G	Mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episodes (MELAS)	8%	Pathogenic

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1: CP-related variants identified in the cohort. The variants were sorted by the platforms and inheritance modes. Motion disability was scored by the Gross Motor Function Classification System (GMFCS). Known disease for a specific gene was retrieved from the Online Mendelian Inheritance in Man (www.omim.org).

Assessment of the role of *de novo* mutations

On aggregate, we characterized 26 CP-related *de novo* events in about one fifth of the cohort individuals, which included 18 autosomal dominant mutations (2 post-zygotic), 2 X-linked dominant mutations (1 post-zygotic), 3 *de novo* CNVs spanning multiple disease-related genes, and 1 *de novo* mutation in the mitochondrial genome (Supplementary figure 2 and 3).

It is noted that there was a PZM observed on a male patient's (CP_102_1) chromosome X while the affected gene *TAF1* (OMIM: 313650) is known to be related to X-linked recessive ID. Also interestingly, a patient (CP_098_1) had compound heterozygous variants in the gene of *ATP8A2* (OMIM: 605870), albeit with one allele of post-zygotic origin.

We found in two spastic diplegia patients (CP_050_1 and CP_095_1) DNMs of *SPAST* (OMIM: 604277), a gene frequently identified in CP patients that is related to early-onset autosomal dominant spastic paraplegia. We also identified in three CP patients (CP_039_1, CP_067_1, CP_094_1) DNMs of *KIF1A* (OMIM: 601255), which is known to cause autosomal dominant mental retardation and motor delay. These patients showed variable phenotype (spastic diplegia and quadriplegia, GMFCS level I - V). Nevertheless, all of the three patients manifested ID and severe microcephaly. Another recurrent gene in this cohort is *COL4A1* (OMIM: 120130), in which we identified DNMs in two CP patients (CP_008_1 and CP_033_1). Both patients presented spastic CP, epilepsy, and severe ID. In addition, one patient (CP_008_1) showed severe microcephaly and vision problem.

Some of the CP patients harbored disease-related variants in genes that are functionally relevant. For example, the patient CP_107_1 (dyskinetic CP with ID) had a post-zygotic *de novo* mutation of *BCL11A* (OMIM: 606557), while the patient CP_096_1 (mixed-type CP with ID) had a *de novo* truncating mutation of *BCL11B* (OMIM: 606558). *BCL11A* and *BCL11B* encode zinc-finger proteins that regulate transcription, highly expressed in brain and bone marrow, and were reported to be related to autosomal dominant developmental disorders. Similar scenario occurred in two CP patients (CP_055_1 and CP_061_1), where the former (spastic hemiplegia with severe ID and microcephaly) had a *de novo* missense mutation of *TUBA1A* (OMIM: 602529), and the latter (spastic hemiplegia with mild motor impairment) had a post-zygotic *de novo* mutation of *TUBB2B* (OMIM: 612850). *TUBA1A* and *TUBB2B* code for brain-enriched tubulin components as subunits of microtubules, and were known to be associated with autosomal dominant brain malformations.

Identification of the variants involved in recessive disease

Altogether we confirmed recessive-disease-related variants in 32 CP families (one quarter of the 120 cohort families), including 22 families with autosomal recessive disease, 9 families with X-linked recessive disease, and 1 CNV located at Xq28 of a male patient (CP_119_1) inherited from the unaffected mother, where a variety of CNVs were reported previously in patients of neuropsychiatric disorders, in some cases, with gender bias [39,40].

We noticed that, among the 22 CP families with autosomal recessive disease, the zygotic ratio

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of compound heterozygosity vs. homozygosity was 3:1, which was acceptable in a largely outbred country. Interestingly, among the 6 homozygous variants, only one was proved to be autozygous by the CMA-based genotyping and family history consulting (CP_036_1, second-cousin consanguinity). This observation indicated the existence of CP-related hotspot variants in a specific population.

The defects of *RARS2* (OMIM: 611524), a gene coding for mitochondrial arginyl-tRNA synthetase, led to manifestation in two patients (CP_007_1 and CP_110_1). The former patient had severe spastic quadriplegia with ID, epilepsy, and vision problem, probably caused by the highly pathogenic compound heterozygous variants of *RARS2*; whereas the latter patient, with mild dyskinetic CP, moderate ID and hearing problem, was likely associated with a homozygous missense variant that may be less damaging.

The patient CP_105_1 presented ataxic CP with mild ID and microcephaly. A homozygous truncating variant was found in the gene of *SQSTM1* (OMIM: 601530), encoding a ubiquitin-binding protein. According to a recent report, four families of autosomal recessive childhood-onset neurodegenerative disorder were found to have detrimental variants in *SQSTM1* [41]. However, the onset age in our case was much younger (in the first year since birth), therefore, this finding expanded the phenotype spectrum related to the gene of *SQSTM1*.

TARS (OMIM: 187790) codes for a threonyl-tRNA synthetase and plays dual roles both as an assembly scaffold of translation initiation components and as a target mRNA selector [42]. A very recent report introduced two individuals carrying *TARS* variants with trichothiodystrophy [43], while we found a spastic hemiplagia patient (CP_014_1) with biallelic *TARS* variants, who presented overlapping phenotype including developmental delay, but without manifestation in hair or skin (Supplementary figure 4).

PTK7 (OMIM: 601890) encodes a tyrosine kinase of Wnt-signaling pathway, and altered PTK7 activity in mice models caused perturbation of neural tube development [44]. Very recent studies found association of *PTK7* variants with neural tube defects and fetal anomaly in human [45,46]. We found that the patient CP_052_1 had compound heterozygous deleterious variants of this gene, and the patient showed spastic diplegia with mild motion impairment (Supplementary figure 5).

CDK5RAP1 (OMIM: 608200) encodes an enzyme involved in the mitochondrial tRNAs modification resulting in stabilization of codon/anticodon interaction, and is primarily expressed in neurons of the central nervous system. Defects of this gene were reported to cause myopathy in mice and humans [47]. Biallelic damaging variants of this gene were revealed in CP_075_1, a dyskinetic CP patient with ID and microcephaly (Supplementary figure 6).

Another two novel candidate genes, namely, *TYW1* (OMIM: 611243) and *GPAM* (OMIM: 602395), were identified in two CP families with distinct phenotypes (CP_012 and CP_063, respectively). The former family had multiple patients, manifesting both CP and ID; while the latter family had a patient with CP but without ID. Because this dichotomous manifestation epitomized the two major subgroups of CP (with or without ID), we carried out detailed functional studies, and presented the results in the subsequent parts of this manuscript.

Common features of CP-related genes

When lumping CP-related genes from this cohort and from literature, there have been in sum

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

117 CP-related genes identified up till now (Supplementary table 4). This number is far less than the 700+ genes involved in ID and other NDDs [29]. About half of the CP-related genes are defined to be ID/NDD genes, and this is not unexpected since CP is on a diagnostic continuum with ID/NDD (Figure 2 (A)). The inheritance modes of the 117 CP-related genes include autosomal dominant (42.7%), autosomal recessive (42.7%), X-linked recessive (12%), and X-linked dominant (2.6%). Compared to ID/NDD genes, CP-related genes lack significantly in autosomal recessive genes but abound in autosomal dominant genes (P-value=0.0006, Fisher exact test) (Figure 2 (A)). We thus speculated that there was a great chance for novel autosomal recessive CP-related genes to be identified in future studies.

We found that the 117 CP-related genes were enriched in a variety of pathways and protein-protein interaction modules, involving multiple aspects of brain development and functioning (Figure 2 (B). We noticed that there were at least 6 functional modules that reflected the essential and coordinated roles of CP-related genes in living organisms, i.e., lysosome, gap junction, dopaminergic synapse, focal adhesion, cellular senescence, and estrogen signaling modules (Figure 2 (C)). Estrogen signaling pathway regulates a plethora of physiological processes in mammals, including cellular homeostasis and behavior, given the significant gender bias in populations of CP, ID and ASD, to explore its involvement in the etiology of NDDs would be an interesting topic. On the other hand, we noticed the pathogen-related modules (pathogenic *escherichia coli* infection, human cytomegalovirus infection, and human papillomavirus infection) enriched in the CP/ID/NDD genes, comprising ~11% of the 117 CP-related genes, and occurring in the CP cohort patients (4%), such as *TUBA1A*, *TUBB2B*, *CTNNB1*, *COL4A1*. Although it is well known that CP and ID can be sequelae of intracranial pathogen infection, whether the variants of these genes increase the relevant disease risk remains elusive.

When data-mining in the public databases and literature of gene expression profiles in brain, we were aware that the CP-related genes were expressed in a variety of anatomic structures, including prefrontal cortex, motor cortex, cerebellar cortex, hippocampus, striatum, somatosensory cortex, etc, and in a variety of brain cells, including excitatory neuron, inhibitory neuron, interneuron, astrocyte, oligodendrocyte, cerebellar granule cell, Purkinje neuron, microglia, etc (Figure 2 (D)). The top3 brain regions were cerebellum, primary motor cortex, and striatum; the top3 cell types were Purkinje cell, astrocyte, and oligodendrocyte. For example, *TYW1* showed enriched expression in neuronal cells in each brain region, but not observable in pericytes or microglia cells. *GPAM* showed exclusively high expression in astrocytes, whereas untraceable in other brain cells. Expression of *CDK5RAP1* was observed in several subtypes of inhibitory neuron, vascular smooth muscle cells and endothelial cells, the latter two reflecting relationship with brain angiogenesis. *PTK7* showed higher expression in pericytes compared to other cell types, indicating a role in capillary formation. The specific expression patterns can serve as useful guidelines for further mechanistic studies on these genes.

As it was found, CP-related genes reached highest expression in different brain regions (forebrain 57%, hindbrain 43%) and at sequential developmental stages (embryonic 58%, postnatal 42%) (Figure 2 (E)). Interestingly, we noticed that CP-related genes could be roughly divided into two subgroups with distinctive features, namely, forebrain-expressed genes and hindbrain-expressed genes. The former genes, such as *TYW1*, highly expressed in forebrain

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

including motor cortex and striatum, had a significantly higher chance to have comorbidity of ID than the latter ones, such as *GPAM*, which were specifically expressed in hindbrain including brainstem and cerebellum (P-value=0.002, Fisher exact test). A spin-off observation was that, forebrain-expressed CP-related genes preferred to reach expression height during embryonic periods, while hindbrain-expressed CP-related genes approached expression climax throughout embryonic and postnatal periods, but with an inclination for postnatal days (P-value=0.05, Fisher exact test). It is an intuitive surmise that, the forebrain-expressed genes should participate in neurogenesis that is generally finished before birth, while the hindbrain-expressed genes may take part in myelination that is not complete until two years after birth in human [48,49].

Next, we will illustrate this dichotomous classification system by delving into two typical CP-related genes, newly identified in this study, namely, *TYW1* and *GPAM*.

Figure 2: CP-related gene summary. (A) Inheritance modes comparison between the CP-related genes and ID/NDD-related genes. The data of CP-related genes were from this study and a collection of literature. The data of ID/NDD-related genes were retrieved from literature [29]. **(B)** Enriched KEGG pathways and GO terms of the CP-related genes. Upper left: enriched KEGG pathways. Upper right: enriched GO terms of Biological Process. Bottom left: enriched GO terms of Molecular Function. Bottom right: enriched GO terms of Cellular Component. Red color indicated the pathways and GO terms also enriched in the ID/NDD-related genes. **(C)** Protein-protein interaction network modules linking CP-related genes (red), ID/NDD-related genes (blue), and shared genes (half red, half blue). **(D)** Expression of CP-related genes in different human brain regions and cell types. Upper left: expression heatmap in 16 human brain regions [50]. MFC: medial prefrontal cortex. OFC: orbital prefrontal cortex. DFC: dorsolateral prefrontal cortex. VFC: ventrolateral prefrontal cortex. M1C: primary motor cortex. S1C: somatosensory cortex. IPC: posterior inferior parietal cortex. A1C: primary auditory cortex. STC: superior temporal cortex. ITC: inferior temporal cortex. V1C: primary visual cortex. HIP: hippocampus. AMY: amygdala. STR: striatum. MD: mediodorsal nucleus of thalamus. CBC: cerebellar cortex. Upper right: expression heatmap in 8 cell types of human dorsolateral prefrontal cortex [50]. ExN: excitatory neuron. InN: interneuron. Astro: astrocyte. Oligo: oligodendrocyte. OPC: oligodendrocyte progenitor cell. Endo: endothelial cell. VSMC: vascular smooth muscle cell. Bottom left: expression heatmap in 9 cell types of human cerebellum [51]. Gran: cerebellar granule cell. Purk1: Purkinje neuron, subtype 1. Purk2: Purkinje neuron, subtype 2. Ast: astrocyte. Oli: oligodendrocyte. OPC: oligodendrocyte precursor cell. Mic: microglia. End: endothelial cell. Per: pericyte. Bottom right: expression heatmap in 24 cell types of human frontal cortex [51]. Ex1: excitatory neuron, subtype 1. Ex2: excitatory neuron, subtype 2. Ex3e: excitatory neuron, subtype 3e. Ex4: excitatory neuron, subtype 4. Ex5b: excitatory neuron, subtype 5b. Ex6a: excitatory neuron, subtype 6a. Ex6b: excitatory neuron, subtype 6b. Ex8: excitatory neuron, subtype 8. In1a: inhibitory neuron, subtype 1a. In1b: inhibitory neuron, subtype 1b. In1c: inhibitory neuron, subtype 1c. In3: inhibitory neuron, subtype 3. In4a: inhibitory neuron, subtype 4a. In4b: inhibitory neuron, subtype 4b. In6a: inhibitory neuron, subtype 6a. In6b: inhibitory neuron, subtype 6b. In7: inhibitory neuron, subtype 7. In8: inhibitory neuron, subtype 8. Ast: astrocyte. Oli: oligodendrocyte. OPC: oligodendrocyte precursor cell. Mic: microglia. End: endothelial cell. Per: pericyte. **(E)** Expression climax of CP-related genes in mice, with regard to brain regions (forebrain and hindbrain), developmental stages (11 time points before and after birth), and comorbidity (with or without ID). X-coordinate: CP-related genes; Y-coordinate: embryonic and postnatal days. Gene expression data were retrieved from

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

https://www.ebi.ac.uk/gxa/home.

Null and hypomorphic alleles of *TYW1* caused primary microcephaly and impairment in motion and cognition

The male proband CP_012_1, an ataxic CP patient, was the second of two children of a non-consanguineous, healthy Chinese couple with unremarkable family history. He was born by normal delivery at 40 weeks of gestational age after an uneventful pregnancy with a birth weight of 3.8 kg (+1.17 S.D.). At conception, the mother was aged 31 and the father was 34 years old. By clinical examination at age of 6, the patient's height, body weight, and head circumference were 103 cm (-3.20 S.D.), 16 kg (-2.06 S.D.), and 47.5 cm (-3.26 S.D.), respectively. At age of 8 years 7 months, the cranial MRI of the patient showed a slightly widened cerebral subarachnoid space, enlarged ventricle, and abnormal cerebellum morphology. Cognitive evaluation (Wechsler Preschool and Primary Scale of Intelligence) was performed at age of 7 years and 6 months, with full-scale IQ 40, indicating a moderate ID. The level of Gross Motor Function Classification System (GMFCS) was III, i.e., walk using a hand-held mobility device. The fine motor skill was poor since the Manual Ability Classification System (MACS) was evaluated at the level of IV (handles a limited selection of easily managed objects in adapted situations). Another patient CP_012_2, a three-year-older sister of the proband, was diagnosed mixed-type CP. The mother had a normal delivery after 40 weeks of uneventful gestation, and the birth weight of the patient was 3.2 kg (-0.02 SD). The parameters of clinical examination at age of 14 years of the patient were: height 150 cm (-1.51 S.D.), weight 41 kg (-1.03 S.D.), and head circumference 49 cm (-3.64 S.D.). Cognitive evaluation (Estimated Cognitive Level of Children with CP) was performed at age of 13 years and 1 month with IQ <50, indicating a moderate ID. The level of the GMFCS was IV, meaning self-mobility with limitations and needing powered mobility. The fine motor skill was poor since the MACS was at level III (handles objects with difficulty, need help to prepare and/or modify activities). The medical information of the two patients was presented in Figure 3 (A) and Supplementary table 1 and 5.

In both of the two patients, we identified compound heterozygous variants in the gene of *TYW1* (OMIM: 611243), which were co-segregating with the phenotype in the family and confirmed by targeted PCR and Sanger sequencing (Figure 3 (B)). The pRecessive value of *TYW1* (http://exac.broadinstitute.org/) is 0.9945, a very strong indicator of recessive-disease-related gene. The two variants (NM_018264: c.616C>T, p.R206C; c.1166G>A, p.R389Q) were absent from an in-house database with 2,247 healthy controls, and were very rare in the public databases (http://exac.broadinstitute.org/; https://db.cngb.org/cmdb/) with allele frequencies below 1 x 10^{-5}. Pathogenicity prediction classified the two variants as damaging or disease-causing (Supplementary table 3). Multiple protein sequences alignment among different species showed high conservation of both positions, i.e., R206 and R389, which were located in the functional domains (Figure 3 (C)). The protein structural modeling revealed the functional locations of both amino acids, and showed the reduced protein stability (R389Q) and disturbed substrate binding (R206C) (Figure 3 (C)). Compatible with the prediction, we found a dramatically lower protein level of TYW1 in the patient's peripheral blood sample, compared to those of the parents and healthy children with same age and gender (Supplementary figure 7).

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Using whole-mount *in situ* hybridization (WISH) of zebrafish, we found that *tyw1* expression was highly enriched in the developing central nervous system including brain and spinal cord (Figure 3 (D)). The morpholino-mediated *tyw1* knockdown (*tyw1*-MO) zebrafish models were established, by blocking mRNA splicing (MO1) and protein translation (MO2) of *tyw1* (Supplementary figure 8, 9 and 10). We observed significant head size reduction in the *tyw1*-MO zebrafish compared to the wild-type ones (Figure 3 (E)). It was revealed that *tyw1* deficiency in zebrafish resulted in ectopic neuronal cell migration in brain and undifferentiated motor neuronal cells in spinal cord (Figure 3 (F) and Supplementary figure 10). The phenotypic anomaly could be effectively rescued by zebrafish *tyw1* mRNA or human *TYW1* mRNA, but not by the *TYW1* mRNAs with variants identified in the CP patients (Figure 3 (G)). This result substantiated the pathogenicity of the two variants identified in the patients. By performing swimming behavior tests, we noticed that the swimming capacity of *tyw1*-MO zebrafish larvae was compromised significantly, specifically in swimming speed, time, and distance, but not in rotation (Figure 3 (H)).

By using CRISPR/Cas9 genome editing technology, we constructed a *Tyw1*-knockout mice model (Supplementary figure 11). In the *Tyw1*-/- mice, the brain size and weight were significantly below the level of those of the wild-type littermate, with abnormal intracranial morphology (Figure 3 (I) and Supplementary figure 12). Behavioral tests, i.e., the Morris water maze test, the rotarod test, and the grip strength test, all witnessed significantly reduced performances of the *Tyw1*-/- group compared to the wild-type controls (Figure 3 (J), (K), (L)). The deterioration of both motion and cognition was compatible with the manifestation of the CP patients with hypomorphic *TYW1* alleles, and reflected the underlying cerebral regions responsible for both functioning.

Figure 3: Correlation between phenotype and genotype of *TYW1*. (A) Clinical summary of the two index patients harboring *TYW1* variants, and the cranial magnetic resonance imaging of the patient II.2 (CP_012_1). **(B)** Compound heterozygous variants of *TYW1* were identified in the two index patients, which co-segregated with the phenotype in the family members. **(C)** The two identified amino-acid changes, i.e., R206C and R389Q, were located inside the protein domains of flavodoxin and radical SAM, respectively. The two changes took over the highly conserved amnio-acid positions throughout a series of species. Three-dimensional structural model of TYW1 protein displayed the stability alteration (ddG, Gibbs free energy) and local structural changes, before and after the introduction of two amino-acid changes. In the two zoom-in views, white bars stood for the wild-type amino acids and color bars for the mutated ones. **(D - H)** Morphological and behavioral assays on zebrafish models. **(D)** Whole mount *in situ* hybridization of zebrafish embryos at 36 hpf using zebrafish antisense probe of *tyw1* showed specific expression of *tyw1* in the central nervous system including brain and spinal cord. D': lateral view. D'': dorsal view. D''': cross section. hpf: hours post fertilization. **(E)** Head area, optic tectum area, medulla oblongata area, eyeball area, eyeball length, and eyeball width of the *tyw1* knockdown (*tyw1*-MO1, a splice-blocking morpholino) zebrafish larvae at 5 dpf were significantly decreased compared with the controls. N = 10 for each test group. ** $p < 0.01$, *** $p < 0.001$, unpaired t-test. dpf: days post fertilization. **(F)** Confocal imaging of egfp positive cells in the transgenic zebrafish *Tg(huC:egfp)* embryos at 48 hpf showed ectopic neuronal cells (indicated by red arrowheads) in the *tyw1*-MO1 zebrafish group, which could be rescued by the wild-type *tyw1* mRNA. Confocal imaging of gfp positive cells in the transgenic zebrafish *Tg(mnx1:gfp)*^{ml2} embryos at 72 hpf showed undifferentiated motor

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

neuronal cells (indicated by white stars) in the *tyw1*-MO1 zebrafish group, which could be rescued by the wild-type *tyw1* mRNA. **(G)** The *tyw1* knockdown (*tyw1*-MO1) brought about ectopic neuronal cells in the zebrafish *Tg(huC:mCherry)* embryos at 48 hpf, and there were only 9% of subjects that remained normal. This abnormal phenotype could be rescued by zebrafish *tyw1* mRNA (*ztyw1*), human *TYW1* mRNA (WT), mutated human *TYW1* leading to R389Q, and mutated human *TYW1* leading to R206C, to variable extents (81%, 86%, 11%, 22%, respectively). Ectopic neuronal cells were indicated by white arrowheads. Fish number ranged from 63 to 88 in different test groups. **(H)** Swimming behavior tests of the *tyw1*-MO1 zebrafish larvae at 5 dpf compared to the controls, on moving time, speed, distance, clockwise (CW) rotation, and counterclockwise (CCW) rotation. The former three parameters revealed significant reduction of the *tyw1*-MO1 group, while the latter two parameters showed no significant change. N = 9 to 12 for each test group. * $p < 0.05$, n.s.: no significant difference, unpaired t-test. **(I - L)** Morphological and behavioral assays on *Tyw1* knockout (*Tyw1*-/-) mice models. **(I)** The *Tyw1*-/- mice at postnatal 8 weeks showed significantly reduced brain size and weight compared to the wild-type (WT) ones. The orange arrow line, yellow arrow line, and red dash line indicated the width, length, and area of brain, respectively. In parallel, measurement of body weight revealed no significant difference between the *Tyw1*-/- mice and the WT ones. N = 6 per genotype with equal numbers of male and female mice. * $p < 0.05$, ** $p < 0.01$, n.s.: no significant difference, unpaired t-test. **(J)** Track plots depicted the trace of mice during the probe test in a Morris water maze. The *Tyw1*-/- mice showed worse performance compared to the WT ones in the test. N = 10 per genotype with equal numbers of male and female mice. * $p < 0.05$, ** $p < 0.01$, n.s.: no significant difference, unpaired t-test. **(K)** Rotarod tests showed significantly reduced motor coordination and balance of the *Tyw1*-/- mice compared to the WT ones. N = 10 per genotype with equal numbers of male and female mice. ** $p < 0.01$, unpaired t-test. **(L)** Grip strength tests revealed significantly weaker grip strength of the *Tyw1*-/- mice compared with the WT mice. N = 10 per genotype with equal numbers of male and female mice. *** $p < 0.001$, n.s.: no significant difference, unpaired t-test.

Defective Tyw1 hindered neuronal proliferation and migration due to increased ribosomal frameshift in a subset of cell-cycling-related proteins

The gene of *TYW1* (OMIM: 611243) encodes a tRNA-wybutosine (tRNA-yW) synthesizing protein, localized on the cytosolic surface of the endoplasmic reticulum and responsible for producing a hypermodified guanosine, i.e., wybutosine, at the position 37 of phenylalanine tRNA ($tRNA^{Phe}$) adjacent to the 3-prime of anticodon in archaea and eukaryotes [52,53]. This hypermodification at postion 37 of $tRNA^{Phe}$ is known to play a critical role in stabilizing the appropriate interaction between codon and anticodon during protein translation, absence of which incurs promoted ribosomal frameshift thus reducing translational accuracy [54-57]. Interestingly, although there exist two codons for phenylalanine, i.e., UUU and UUC, it seems that wybutosine exerts influence only upon the codon UUU [58].

The expression of TYW1 are enriched in the neuronal cells, especially during the embryonic development (Figure 2 (D) and (E); Figure 4 (A) and (B)). We observed significantly decreased neurogenesis in the E13.5 mice brain when comparing the *Tyw1*-/- mice with the wild-type ones (Figure 4 (C)). In the *Tyw1*-/- mice brain, the migration of cortical neuronal precursors between E15.5 and E18.5 was hindered, resulting in obvious reduction of superficial-layer neurons and noticeable accumulation of deep-layer neurons (Figure 4 (D)). Consequently, in both frontal cortex and motor cortex of the adult mice brain, the thickness of superficial layer was significantly less in the *Tyw1*-/- group compared with the wild-type one, while the trend

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

seemed reverse in the deep layer although the statistical power had not been gained (Figure 4 (E)). The simultaneous abnormalities in both frontal and motor cortices in the mice models provided a clue to explain the CP-plus-ID phenotype of patients with hypomorphic *TYW1* alleles in this study.

Primary neurons extracted from the E13.5 mice brain cortex showed significantly decreased proliferation in the *Tyw1*-/- group compared to the wild-type group, while the apoptosis intensity showed no obvious increase in the *Tyw1*-/- group (Figure 4 (F)). The migration tests on the primary neurons proved reduced migrating capacity of the *Tyw1*-/- group compared to the wild-type one (Figure 4 (G)). In parallel, we evaluated the performance of SH-SY5Y cells after introducing *TYW1*-knockout (Supplementary figure 13). Interestingly, as observed in the primary mice neurons, the *TYW1*-knockout SH-SY5Y cells presented significantly reduced abilities of proliferation, adhesion, and migration, compared to the wild-type ones, however, there was no observable change in neuron differentiation (Supplementary figure 14).

To explore the mechanism of epigenetic regulation by *TYW1*, we constructed a model to simulate how the levels of wybutosine affected the ribosomal frameshift (Online methods). We defined a parameter (attenuation coefficient) for each protein to depict the vulnerability of a specific protein-coding gene to the reduced level of wybutosine. Typically, if the mRNA sequence of a specific gene does not contain a codon UUU, this gene will not be affected by the ribosomal frameshift related to the wybutosine level. On the contrary, if the mRNA sequence of a gene holds many codon UUU, its protein translation will be sensitive to the wybutosine level, with the extent determined by the density and distribution of codon UUU. This algorithm thus generated a protein list with attenuation coefficients lower than expectation, due to malfunctioning TYW1. Subsequently, we filtered the list by two thumb rules (Supplementary table 6). First, since we observed microcephaly and neurological manifestation in the subjects of human, mouse, and zebrafish, we chose proteins with low attenuation coefficients in all the three species. Second, we paid attention to proteins that were known to be disease-related, and bore resemblance to TYW1 in terms of gene expression patterns, especially in developing brains (Supplementary figure 15). In this way, four candidate proteins emerged, namely, CENPE (OMIM: 117143), STIL (OMIM: 181590), NCAPD2 (OMIM: 615638), and SASS6 (OMIM: 609321) (Figure 4 (H)). The measurement of E13.5 mice brain samples showed significantly reduced levels of these four proteins in the *Tyw1*-/- group compared to the wild-type one, while the corresponding RNA levels did not change much (Figure 4 (I)). Therefore, it was reasonable that the defective Tyw1 hindered the production of these four proteins in brain, with ensuing reduction of neuronal proliferation and migration. Interestingly, these four proteins are all involved in cell cycling and mitosis, the defects of which were reported to be related to primary microcephaly. STIL and SASS6 are necessary for centriole duplication during the cell cycle [59-64]. CENPE is required for spindle microtubule capture and attachment at the kinetochore during cell division [65-67]. NCAPD2 is a component of the condensin multiprotein complex participating in mitotic chromosome condensation [68,69].

Figure 4: Functional mechanism of defective TYW1. (A) Brain section immunostaining of the wild-type mice at postnatal 8 weeks, where Tyw1 (green) and NeuN (red) were co-labeled, showed the Tyw1 expression in the neurons of cortex, hippocampus, and cerebellum. Blue: DAPI. Bar = 75 μm. **(B)** Immunostaining of the normal human brain tissue microarrays, where TYW1 (green) and NeuN (red)

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

were co-labeled, revealed the TYW1 expression in the neurons of a variety of brain regions. Blue: DAPI. Bar = 50 μm. **(C)** EdU was intraperitoneally injected 2h before sacrifice into pregnant mice with embryos on E13.5. Cells incorporated with EdU were determined in the embryonic brain sections under fluorescence microscope and analyzed with ImageJ software. Compared to the wild-type (WT) littermate, significantly reduced percentages of EdU^+ area were observed at neocortex (NCX) of the *Tyw1*-/- embryo brains, and at medial ganglionic eminence (MGE) and lateral ganglionic eminence (LGE) as well. Red: EdU. Blue: DAPI. N = 3 per genotype. For a specific sample, each measurement was performed twice in separate fields. ** $p < 0.01$, unpaired t-test. Bar = 100 μm. **(D)** EdU was injected into the pregnant mice carrying embryos on E15.5, which were observed after 72 h. The embryonic cortex was divided into ten equally spaced bins along the vertical axis. Percentage of EdU^+ area in each bin was measured. At the *Tyw1*-/- mice cortex, EdU^+ cells in the superficial layer were much less than that in the WT mice, while EdU^+ cells in the deep layer were a bit more than that in the WT mice. Light blue: EdU. Blue: DAPI. N = 3 per genotype. Bar = 100 μm. **(E)** Brain section immunostaining of mice at postnatal 8 weeks, where the $Cux1^+$ and $Foxp2^+$ cells were labeled for neurons in superficial layer (layer II-III) and deep layer (layer VI), respectively. Compared with the WT mice, the *Tyw1*-/- mice brain showed significantly reduced thickness of the superficial layers at both frontal and motor cortices, while in the deep layers, the reverse happened, albeit without statistical significance. WM: white matter. Green: Cux1. Red: Foxp2. Blue: DAPI. N = 3 per genotype. For a specific sample, each measurement was performed twice in separate fields. *** $p < 0.001$, n.s.: no significant difference, unpaired t-test. Bar = 200 μm. **(F - G)** Proliferation, apoptosis and migration analysis on primary neurons from the cortex of E13.5 mice. **(F)** Quantification of the proliferative and apoptotic primary neurons from the E13.5 mice brain cortex. The primary neurons from the *Tyw1*-/- mice showed significantly decreased proliferation by using the EdU staining, while the intensity of apoptosis did not change significantly by using the TUNEL staining. *** $p < 0.001$, n.s.: no significant difference, unpaired t-test. Bar = 100 μm. **(G)** Quantification of the migrating primary neurons from the E13.5 mice brain cortex. In the transwell assay, the primary neurons from the *Tyw1*-/- mice showed significantly decreased migration. Red arrows indicated cells that already migrated through pores, while arrowheads indicated cells that were migrating in pores. *** $p < 0.001$, n.s.: no significant difference, unpaired t-test. Bar = 250 μm. **(H)** Schematics of the mechanism of ribosomal frameshift under the conditions of wild-type and mutant alleles of *TYW1*. GAA is the anticodon of $tRNA^{Phe}$ for phenylalanine, and UUU is the codon of mRNA for phenylalanine. In the four mRNAs, with normalized lengths, of *SASS6*, *NCAPD2*, *CENPE*, and *STIL*, the positions of codon UUU were indicated by red bars. The introduction of mutated TYW1 led to promoted ribosomal frameshift at the interaction between $tRNA^{Phe}$ and mRNAs, resulting in the reduced production of a subset of proteins involved in cell cycling. **(I)** The measurement of mRNA and protein expression of the E13.5 mice brain cortex. In the *Tyw1*-/- group, the RNA levels of *Cenpe*, *Stil*, *Ncapd2*, and *Sass6* did not change much compared to the WT group. However, the protein levels of Cenpe, Stil, Ncapd2, and Sass6 showed significant reduction in reference to the WT group, with remaining levels of 0.38, 0.63, 0.36, and 0.14, respectively. β -actin was used as the control in both RNA and protein measurement. Neither RNA nor protein of Tyw1 could be detected in the *Tyw1*-/- group. MW (Tyw1) = 84 kDa. MW (Cenpe) =312 kDa. MW (Stil) =143 kDa. MW (Ncapd2) =157 kDa. MW (Sass6) =74 kDa and 65 kDa. MW (β -actin) = 43 kDa. The quantitative RT-PCR and western blot data were from three independent experiments. * $p < 0.05$, *** $p < 0.001$, unpaired t-test.

Null and hypomorphic alleles of *GPAM* incurred drastically reduced motor ability with unaffected cognition

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The female proband CP_063_1, a spastic quadriplegic CP patient, was the only child of unrelated healthy parents without relevant family history. She was born by the Caesarean section at 39 gestational week when her mother was 30 years old and father was 32 years old. Her birth weight was 3.05 kg (-0.44 S.D.). The clinical examination at age of 2 years and 4 months showed body height 83 cm (-2.45 S.D.), weight 10 kg (-1.7 S.D.), and head circumference 47 cm (-1.31 S.D.). The cranial neuroimaging taken at 4 years old revealed no apparent anatomical abnormality except for plagiocephaly, which was due to the elongated period of supine position after birth. However, there was a reduction of white matter fiber tracts, especially in the corticospinal tract. The level of Gross Motor Function Classification System (GMFCS) was IV, meaning 'self-mobility with limitations and may use powered mobility'. The fine motor skill was poor because the Manual Ability Classification System (MACS) was evaluated at level of III (handles objects with difficulty, need help to prepare and/or modify activities). Very interestingly, the cognitive evaluation (Wechsler Preschool and Primary Scale of Intelligence) was performed at age 3 years and 7 months, with full-scale IQ 116. The IQ scores ranked this CP patient above 70% of individuals in general population (IQ scores in a large unselected population follow a normal distribution with mean value 100 and S.D. 15). This patient epitomized a major subgroup, comprising half of the CP population, who were impaired in motion albeit normal in intelligence. The medical information of this patient was presented in Figure 5 (A) and Supplementary table 1 and 5.

Compound heterozygous variants of *GPAM* (OMIM: 602395) were identified in this patient, the association between which and the CP phenotype in this family was confirmed by using targeted-PCR and Sanger-sequencing (Table 1 and Figure 5 (B)). The pRecessive value 0.9998 of *GPAM* (http://exac.broadinstitute.org/) rendered a very strong indicator for a recessive-disease-related gene. Allele frequencies of the two variants (NM_001244949: c.1495G>C, p.G499R; c.2005C>T, p.P669S) were both below 1×10^{-5}, in the large-scale databases (https://db.cngb.org/cmdb/ and http://exac.broadinstitute.org/), and with no match in the in-house database of more than 2,247 healthy individuals. Pathogenicity prediction labeled the two variants damaging or disease-causing (Supplementary table 3). Multiple protein sequences alignment in a variety of species showed high conservation of both positions, i.e., G499 and P669, which were located inside the important protein domains. Three-dimensional structural models revealed that G499 and P669 were positioned at the pivotal regions, and both alterations of amino acids led to dramatic reduction of protein structural stability (Figure 5 (C)).

Using the whole-mount *in situ* hybridization (WISH) analysis of zebrafish, we found that *gpam* was specifically expressed in the developing central nervous system, mainly in the hindbrain (Figure 5 (D)). The morpholino-mediated *gpam* knockdown (*gpam*-MO) zebrafish models were established, by blocking mRNA splicing (MO1) and protein translation (MO2) of *gpam* (Supplementary figure 8 and 16). There was no apparent head size difference between *gpam*-MO zebrafish and the wild-type ones, but we observed obvious reduction of astrocyte in *gpam*-MO zebrafish, especially in the hindbrain (Figure 5 (E) and Supplementary figure 17). This reduction could be effectively rescued by the wild-type human *GPAM* or zebrafish *gpam* mRNA, but not by the *GPAM* mRNAs with variants causing G499R and P669S (Figure 5 (F)). This result corroborated the pathogenicity of the two variants identified in the patient. On the other hand, swimming behavior tests showed that the swimming capacity of *gpam*-MO

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

zebrafish was compromised significantly (Figure 5 (G)), mimicking the motor impairment of the respective CP patient. Very interestingly, we noticed that the swimming capacity patterns of *tyw1*-MO and *gpam*-MO zebrafish were somehow complementary to each other, i.e., the swimming distance, speed, and moving time were abnormal in the *tyw1*-MO zebrafish but normal in the *gpam*-MO zebrafish, in contrast, the rotation movement was normal in the *tyw1*-MO zebrafish but abnormal in the *gpam*-MO zebrafish. This finding may reflect the different underlying mechanism posed by the two genes.

Meanwhile, we constructed *Gpam* knockout mice models by using CRISPR/Cas9 genome editing technology (Supplementary figure 11). In *Gpam* knockout (*Gpam*-/-) mice, there was no significant alteration of brain size or weight compared to the wild-type littermates (Figure 5 (H)), compatible with the observation in human and zebrafish. Similarly, the Morris water maze tests showed no significant difference between the *Gpam*-/- mice and the wild-type ones (Figure 5 (I)). However, the rotarod test and the grip strength test revealed significant reduction of the motor capacity in the *Gpam*-/- mice compared to the wild-type ones (Figure 5 (J) and (K)).

Figure 5: Correlation between phenotype and genotype of *GPAM*. (A) Brief medical description of the index patient II.1 (CP_063_1) harboring *GPAM* variants, and the results of magnetic resonance imaging (MRI) and diffusion tensor imaging (DTI). No gross anatomical abnormality was observed except for plagiocephaly. Deterioration of corticospinal tract was indicated by white arrows. **(B)** Compound heterozygous variants of *GPAM* were identified in the index patient, which co-segregated with the phenotype in the family members. **(C)** The two identified amino-acid changes, i.e., G499R and P669S, were both located inside the conserved protein domains. The two changes took over the amino-acid positions that were highly conserved among different species. Three-dimensional structural model of GPAM protein displayed the reduced protein stability (ddG, Gibbs free energy), and the local structural alterations before and after the introduction of two amino-acid changes. In the two zoom-in views, white bars stood for the wild-type amino acids and color bars for the mutated ones. **(D - G)** Morphological and behavioral assays on zebrafish models. **(D)** Whole mount *in situ* hybridization of zebrafish embryos at 36 hpf by using zebrafish antisense *gpam* probe showed specific expression of *gpam* in the brain, especially in the hindbrain. D': lateral view. D'': lateral view with magnification of head region. D''': dorsal view. hpf: hours post fertilization. **(E)** Confocal imaging of transgenic zebrafish *Tg(gfap:egfp)* embryos at 48 hpf showed reduced number of $gfap^{+}$ cells in the *gpam* knockdown group (*gpam*-MO1, a splicing blocking morpholino), particularly in the hindbrain, labeled by white stars. Bright field lateral imaging showed no apparent general change in the brain. Left: lateral view. Middle: dorsal view. Right: lateral view. **(F)** The *gpam* knockdown (*gpam*-MO1) caused reduction of astrocyte in the zebrafish *Tg(gfap:egfp)* embryos at 48hpf, and there were only 17% of subjects that remained normal. This abnormal phenotype could be rescued by zebrafish *gpam* mRNA (*zgpam*), human *GPAM* mRNA (WT), mutated human *GPAM* leading to P669S, and mutated human *GPAM* leading to G449R, to variable extents (86%, 87%, 44%, 28%, respectively). Fish number ranged from 36 to 72 in different test groups. **(G)** Swimming behavior tests of the *gpam*-MO1 zebrafish larvae at 5 dpf compared to the controls, on moving time, speed, distance, clockwise (CW) rotation, and counterclockwise (CCW) rotation. The former three parameters showed no significant change, while the latter two parameters revealed significant changes. N = 10 to 12 for each test group. * $p < 0.05$, ** $p < 0.01$, n.s.: no significant difference, unpaired t-test. dpf: days post fertilization. **(H - K)** Morphological and behavioral assays on *Gpam* knockout (*Gpam*-/-) mice models. **(H)** The *Gpam*-/- mice at postnatal 8 weeks showed no significant alteration of brain size, brain weight, and body

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

weight compared to the wild-type (WT) ones. The red dash line indicated the area of brain. N = 6 per genotype with equal number of male and female. n.s.: no significant difference, unpaired t-test. **(I)** Track plots depicted the trace of mice during the probe test in a Morris water maze. The *Gpam*-/- mice showed no difference in performance compared to the WT ones. N = 10 per genotype with equal numbers of male and female mice. n.s.: no significant difference, unpaired t-test. **(J)** Rotarod tests showed significantly reduced motor coordination and balance of the *Gpam*-/- mice compared to the WT ones. N = 10 per genotype with equal numbers of male and female mice. ** $p < 0.01$, unpaired t-test. **(K)** Grip strength tests revealed significantly shorter grip duration and weaker grip strength of the *Gpam*-/- mice compared with the WT mice. N = 10 per genotype with equal numbers of male and female mice. *** $p < 0.001$, unpaired t-test.

Defective Gpam interfered with proliferation of astrocyte and oligodendrocyte myelination by perturbing lipid metabolism

The gene of *GPAM* encodes a mitochondrial enzyme that catalyzes the production of lysophosphatidic acid (LPA) by using glycerol-3-phosphate (G3P) as the substrate, which is the rate-limiting step in the synthesis of glycerolphospholipids and triacylglycerol (TAG) [70]. There are four isoenzymes in this step, among which GPAM occupies most of the activity in the liver, and the hepatic knockdown of Gpam in mice reduced the levels of TAG, diacylglycerol, fatty acid and cholesterol [70-72]. In the central nervous system, GPAM was specifically expressed in astrocyte, and reached highest expression level on P14 in mice hindbrain (Figure 2 (D) and (E); Figure 6 (A) and (B)). The enzymatic activity in astrocyte is maintained mainly by GPAM, supplemented by GPAT4, an isoenzyme ubiquitously expressed in a variety of brain cells including neurons and oligodendrocyte [73,74].

In the *Gpam*-/- mice, we observed the reduced number of astrocyte, especially in the medulla regions (Figure 6 (C)). The whole brain cell lysis of P1 mice showed significantly reduced proliferation of astrocyte in the *Gpam*-/- mice, but no apparent increase in the intensity of apoptosis was observed (Figure 6 (D)). Interestingly, brain section staining revealed an obviously decreased white matter thickness of the motor cortex in the *Gpam*-/- mice (Figure 6 (E)). In parallel with the decrease of astrocyte density, a myelin marker, namely, myelin binding protein (Mbp), was revealed to be significantly reduced in the medulla regions of the *Gpam*-/- mice brain (Figure 6 (F) and (G)). Given the manifestation and cranial MRI results of the respective CP patient, we assumed that, the phenotype of the *Gpam*-/- mice should be attributed to the hypomyelination of corticospinal tract that originates from the primary motor cortex, goes through and modulates in the medulla, and ends in the spinal cord [75].

It is known that myelin is a specialized form of oligodendrocyte membrane that tightly wraps and insulates the axons to accelerate the action potential conduction [48,76]. Compared with other membranes, myelin has a much higher content of lipid, up to 80% of the total dry weight, and the most abundant lipid groups in myelin are cholesterol, phospholipids, and glycolipids [76-78]. It was reported that in mice models, lipid synthesis blockage in the oligodendrocyte caused instant onset of severe demyelination and neurological symptoms, however, this damage could be spontaneously remedied to much extent after a period of months [79]. Since the blood brain barrier shields the central nervous system from circulating lipids, an exogenous lipid supply in the parenchyma of brain should have accounted for this phenomenon, and fortuitously, recent studies discovered that the major supplier of lipids in brain was astrocyte

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[80-82]. Actually, it was revealed that the horizontal lipid flux from astrocyte to oligodendrocyte was a major feature of myelination, and disruption of lipid synthesis in astrocyte led to lasting and more severe hypomyelination than inactivation of that in oligodendrocyte [83-85]. In astrocyte, GPAM is committed to a rate-limiting step from G3P to LPA, the subsequent lipid products of which include phosphatidic acid (PA), diacylglycerol (DAG), triacylglycerol (TAG), phosphatidylinositol (PI), phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylserine (PS), among others. PI, PE, PC, and PS are the major components of myelin [86,87] (Figure 6 (H)). Noticeably, a large body of evidence implicated the association between *GPAM* polymorphisms and plasma level of cholesterol [88-93]. Experiments on brain sections and extraction confirmed this hypothesis, with significantly reduced contents of lipids detected at medulla of the *Gpam*-/- mice (Figure 6 (I) and (J)). Interestingly, we also observed a significant reduction of 1,2-diacyl-G3P (PA) from the samples, which is a downstream product of GPAM-involved reaction (Figure 6 (H) and (J)). PA is not only a precursor of phosphatidylinositol (PI), a component of myelin, but also promotes cell proliferation in astrocyte by activating the MAPK and mTOR signaling pathways (Figure 6 (H)) [94-96]. Inhibition of PA was known to reduce the number of astrocyte and cause neurodevelopmental disorder [97-100].

Figure 6: Functional mechanism of defective GPAM. (A) Brain section immunostaining of the wild-type mice at postnatal 8 weeks, where Gpam (green) and Gfap (red) were co-labeled, showed the Gpam expression in the astrocyte of many brain regions. Blue: DAPI. Bar = 100 μm. **(B)** Immunostaining of the normal human brain tissue microarrays, where GPAM (green) and GFAP (red) were co-labeled, revealed the GPAM expression in the astrocyte of a variety of brain regions. Blue: DAPI. Bar = 50 μm. **(C)** Brain sections immunostaining of the *Gpam*-/- mice at postnatal 8 weeks revealed significantly reduced number of the Gfap+ cells in medulla, compared to the wild-type (WT) ones. The same tendency was observed in other brain regions albeit without statistical significance. Green: Gfap. Red: NeuN. Blue: DAPI. Bar = 100 μm. N = 3 per genotype. For a specific sample, each measurement was performed twice in separate fields. ** $p < 0.01$, n.s.: no significant difference, unpaired t-test. **(D)** Flow cytometry of whole brain cells lysis of P1 mice stained with Gfap, Ki67, and Caspase-3. Percentages of the $Gfap^+Ki67^+$ cells were significantly lower in the *Gpam*-/- mice brain compared to the WT ones, indicating inhibited proliferation of astrocyte. Percentages of the $Gfap^+Caspase\text{-}3^+$ cells showed no difference between the two groups, indicating no activated apoptosis. N = 8 per genotype with equal numbers of male and female mice. ** $p < 0.01$, n.s.: no significant difference, unpaired t-test. **(E)** Nissl staining of brain sagittal sections of mice at postnatal 8 weeks showed a significantly thinner white matter at the motor cortex of the *Gpam*-/- group, compared to the WT ones. The same tendency was observed at the frontal cortex albeit without statistical significance. Bar = 200 μm. N = 3 per genotype. For a specific sample, each measurement was performed four times in separate fields. ** $p < 0.01$, n.s.: no significant difference, unpaired t-test. **(F)** Brain sections immunostaining of mice at postnatal 8 weeks showed lower density of the myelin binding protein (Mbp) at medulla of the *Gpam*-/- mice, compared with the WT ones. Green: Mbp. Blue: DAPI. Bar = 100 μm. **(G)** Western blot of mice brain medulla showed significant reduction of Gfap and Mbp in the *Gpam*-/- group compared to the WT one. Gapdh and β -actin were used as controls. MW (Gfap) = 50 kDa. MW (Mbp) = 19 and 26 kDa. For a specific sample, measurement was performed three times independently. * $p < 0.05$, unpaired t-test. **(H)** Schematics of the mechanism of GPAM involvement in lipid metabolism and its influence on astrocyte proliferation and oligodendrocyte

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

myelination, by regulating the production of PA, PI, PC, PE, PS, etc. GPAT4: glycerol-3-phosphate acyltransferase 4. G3P: glycerol-3-phosphate. LPA: lysophosphatidic acid. PA: phosphatidic acid. DAG: diacylglycerol. TAG: triacylglycerol. CDP-DAG: cytidine diphosphate diacylglycerol. PI: phosphatidylinositol. PC: phosphatidylcholine. PE: phosphatidylethanolamine. PS: phosphatidylserine. MBP: myelin binding protein. PLP: proteolipid protein. **(I)** Brain medulla sections immunostaining showed significant reduction of Filipin, as well as Mbp and Gfap, in the *Gpam*-/- mice compared to the WT ones. Percentage of the Filipin^{+} area per imaging field was proportionate to the cholesterol abundance. Bar = 100 μm. N = 3 per genotype. For a specific sample, each measurement was performed four times in separate fields. * $p < 0.05$, unpaired t-test. **(J)** Quantification of PA, PC, and PE levels at the brain medulla of mice at postnatal 8 weeks revealed significant reduction in the *Gpam*-/- group compared to the WT one. N = 3 per genotype. For a specific sample, each measurement was performed twice or thrice. * $p < 0.05$, ** $p < 0.01$, unpaired t-test.

Discussion:

This work is, to the best of our knowledge, the first large-scale, full-dimensional, clinical genetics research on the etiology of cerebral palsy. We identified, in 45% of 120 CP families, detrimental variants of the established and novel genes related to CP and other NDDs. The meta-analysis on the 117 CP-related genes found in this study and garnered from literature, created an unprecedented compendium spanning a variety of attributes of CP-related genes. Intriguingly, we came up with a novel classification system for CP-related genes, according to their dichotomous spatiotemporal expression patterns. This can serve as a useful guideline for exploring the underlying mechanism and designing targeted treatment.

In parallel with this research, we managed to take full use of the etiological information in the clinical practice. The families were informed of the prognostic outcomes based on gene-wise knowledge bases, and received the adjusted medical and rehabilitation treatments, plus genetic counseling as they wished. For example, we offered serine and glycine replacement to CP_035_1 with *PSAT1* defects [101], Doconexent to CP_044_1 with *CYP2U1* defect [102,103], and magnesium carbonate to CP_053_1 with *CNNM2* defect [104]. Likewise, we implemented personalized rehabilitation training for countering the specific progression of motion disability (Figure 1 and Supplementary table 1).

TYW1 is a novel CP-related gene, defects of which were confirmed in this study to result in primary microcephaly and deterioration of both motion and cognition, by disturbing proliferation and migration of neurons in the developing brain, including motor cortex and frontal cortex. Blockage of neuronal proliferation and/or migration during brain development is a common theme in etiology of primary microcephaly [105,106]. Interestingly, most of the known MCPH, i.e., microcephaly primary hereditary, genes encode components of machinery of cell division and cell cycling, while reports on genes like *TYW1* that is in charge of a subset of MCPH genes, are rare, if not absent. TYW1 plays a critical role in generating wybutosine, a hypermodification at position 37 of tRNA^{Phe} responsible for stabilizing codon-anticodon binding, and this epigenetic procedure can be grouped into one of the three aspects of tRNA activation, namely, the biogenesis, charging, and modification of tRNAs [56,57]. We noticed that, a large proportion of genes involved in tRNA charging have been found to be associated with disease, especially NDDs, however, merely ~10% of genes participating in tRNA modification are related to human illness [54]. Because of the enduring existence of most tRNA modification

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

throughout species evolution, we believed that tRNA modification may be underestimated in its role in disease onset and progression. Future studies in this field can be modeled after this work on *TYW1*, and since the great majority of tRNA modification were identified at position 34 and 37 in the anticodon stem loop, the algorithm that we developed for wybutosin at postion 37 may find a multitude of scenarios to be applied to.

GPAM encodes a mitochondrial enzyme specifically expressed in astrocyte and plays a crucial role in lipid metabolism in brain. Aberrant GPAM found in this study reduced a variety of lipid contents in astrocyte and oligodendrocyte, causing hypomyelination in corticospinal tract and suppressing proliferation of astrocyte *per se*, and the latter deteriorated the interplay between oligodendrocyte and astrocyte thus aggravated the hypomyelination [85,107]. Lipid metabolism problems resulted in cases of neurodevelopmental disorders, including ID, ASD, attention deficit hyperactivity disorder, and schizophrenia [108-112]. And genetic defects in astrocyte were known to be related to lesions in brain white matter, such as *GFAP* and *NPC1* [113-116]. This work disentangled a complex mechanism between astrocyte and myelination, providing a chance to develop new therapeutic strategies. Very intriguingly, there were reports in mice models that, high-fat diet could promote circulating lipid contents and ameliorate hypomyelination in brain if the lipid synthesis in astrocyte was blocked; therefore, it is tantalizing to apply this to the *Gpam*-/- mice and human patients with similar lipid metabolic problems [84,117].

Data availability:

This study's samples and data are accessible, conforming to the principles of the International Cerebral Palsy Genomics Consortium (ICPGC) (https://icpgc.org/).

Acknowledgement:

We are very grateful to the families participating in this study. This work was supported by the Major Medical Collaboration and Innovation Program of Guangzhou Science Technology and Innovation Commission (201604020020), the National Natural Science Foundation of China (81671067, 81974163, 81672253, 31871063, and 81701451), the China Postdoctoral Science Foundation (2019M662852), the Natural Science Foundation of Guangdong Province of China (2018A030313538), and the Key-Area Research and Development Program of Guangdong Province (2019B020227001).

Author contribution:

H.H. conceived and coordinated the project. K.X., H.T., and L.H. recruited the cohort. D.L., J.Z., and X.W. performed the zebrafish modeling. D.H. and N.L. oversaw the cell and mice modeling with aids of P.Z., H.T., Y.Q., C.S., F.Q., M.Y., Z.Z., C.L., S.Z., Y.L., L.L., J.H., and C.P.. N.L. was in charge of the genetic analysis and bioinformatics with aids of D.H., X.F., Y.L., X.P., Z.Z., T.X., Q.Z., and L.L.. N.L. and H.H. wrote the manuscript with comments from all authors.

Competing interests:

The authors declare no competing interests.

References:

1. Colver, A., Fairhurst, C. & Pharoah, P.O. Cerebral palsy. *Lancet* **383**, 1240-9 (2014).
2. Moreno-De-Luca, A., Ledbetter, D.H. & Martin, C.L. Genetic [corrected] insights into the

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

causes and classification of [corrected] cerebral palsies. *Lancet Neurol* **11**, 283-92 (2012).

3. Novak, I. *et al.* Early, Accurate Diagnosis and Early Intervention in Cerebral Palsy: Advances in Diagnosis and Treatment. *JAMA Pediatr* **171**, 897-907 (2017).
4. Rosenbaum, P.L., Palisano, R.J., Bartlett, D.J., Galuppi, B.E. & Russell, D.J. Development of the Gross Motor Function Classification System for cerebral palsy. *Dev Med Child Neurol* **50**, 249-53 (2008).
5. Korzeniewski, S.J., Slaughter, J., Lenski, M., Haak, P. & Paneth, N. The complex aetiology of cerebral palsy. *Nat Rev Neurol* **14**, 528-543 (2018).
6. Graham, H.K. *et al.* Cerebral palsy. *Nat Rev Dis Primers* **2**, 15082 (2016).
7. MacLennan, A.H., Thompson, S.C. & Gecz, J. Cerebral palsy: causes, pathways, and the role of genetic variants. *Am J Obstet Gynecol* **213**, 779-88 (2015).
8. Petterson, B., Stanley, F. & Henderson, D. Cerebral palsy in multiple births in Western Australia: genetic aspects. *Am J Med Genet* **37**, 346-51 (1990).
9. Fletcher, N.A. & Foley, J. Parental age, genetic mutation, and cerebral palsy. *J Med Genet* **30**, 44-6 (1993).
10. Costeff, H. Estimated frequency of genetic and nongenetic causes of congenital idiopathic cerebral palsy in west Sweden. *Ann Hum Genet* **68**, 515-20 (2004).
11. Hemminki, K., Li, X., Sundquist, K. & Sundquist, J. High familial risks for cerebral palsy implicate partial heritable aetiology. *Paediatr Perinat Epidemiol* **21**, 235-41 (2007).
12. Garne, E. *et al.* Cerebral palsy and congenital malformations. *Eur J Paediatr Neurol* **12**, 82-8 (2008).
13. O'Callaghan, M.E., MacLennan, A.H., Haan, E.A., Dekker, G. & South Australian Cerebral Palsy Research, G. The genomic basis of cerebral palsy: a HuGE systematic literature review. *Hum Genet* **126**, 149-72 (2009).
14. Fahey, M.C., Maclennan, A.H., Kretzschmar, D., Gecz, J. & Kruer, M.C. The genetic basis of cerebral palsy. *Dev Med Child Neurol* **59**, 462-469 (2017).
15. Turner, T.N. *et al.* Genomic Patterns of De Novo Mutation in Simplex Autism. *Cell* **171**, 710-722 e12 (2017).
16. Hu, H. *et al.* Genetics of intellectual disability in consanguineous families. *Mol Psychiatry* (2018).
17. Satterstrom, F.K. *et al.* Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**, 568-584 e23 (2020).
18. Segel, R. *et al.* Copy number variations in cryptogenic cerebral palsy. *Neurology* **84**, 1660-8 (2015).
19. Oskoui, M. *et al.* Clinically relevant copy number variations detected in cerebral palsy. *Nat Commun* **6**, 7949 (2015).
20. 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).
21. Corbett, M.A. *et al.* Pathogenic copy number variants that affect gene expression contribute to genomic burden in cerebral palsy. *NPJ Genom Med* **3**, 33 (2018).
22. McMichael, G. *et al.* Whole-exome sequencing points to considerable genetic heterogeneity of cerebral palsy. *Mol Psychiatry* **20**, 176-82 (2015).
23. Parolin Schnekenberg, R. *et al.* De novo point mutations in patients diagnosed with ataxic cerebral palsy. *Brain* **138**, 1817-32 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

24. Matthews, A.M. *et al.* Atypical cerebral palsy: genomics analysis enables precision medicine. *Genet Med* (2018).
25. Takezawa, Y. *et al.* Genomic analysis identifies masqueraders of full-term cerebral palsy. *Ann Clin Transl Neurol* **5**, 538-551 (2018).
26. van Eyk, C.L. *et al.* Targeted resequencing identifies genes with recurrent variation in cerebral palsy. *NPJ Genom Med* **4**, 27 (2019).
27. D'Gama, A.M. & Walsh, C.A. Somatic mosaicism and neurodevelopmental disease. *Nat Neurosci* **21**, 1504-1514 (2018).
28. Campbell, I.M., Shaw, C.A., Stankiewicz, P. & Lupski, J.R. Somatic mosaicism: implications for disease and transmission genetics. *Trends Genet* **31**, 382-92 (2015).
29. Vissers, L.E., Gilissen, C. & Veltman, J.A. Genetic studies in intellectual disability and related disorders. *Nat Rev Genet* **17**, 9-18 (2016).
30. South, S.T. *et al.* ACMG Standards and Guidelines for constitutional cytogenomic microarray analysis, including postnatal and prenatal applications: revision 2013. *Genet Med* **15**, 901-9 (2013).
31. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**, 405-24 (2015).
32. Najmabadi, H. *et al.* Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature* **478**, 57-63 (2011).
33. Doan, R.N. *et al.* Recessive gene disruptions in autism spectrum disorder. *Nat Genet* **51**, 1092-1098 (2019).
34. Ruzzo, E.K. *et al.* Inherited and De Novo Genetic Risk for Autism Impacts Shared Networks. *Cell* **178**, 850-866 e26 (2019).
35. Gilissen, C. *et al.* Genome sequencing identifies major causes of severe intellectual disability. *Nature* **511**, 344-7 (2014).
36. Vissers, L.E. *et al.* A de novo paradigm for mental retardation. *Nat Genet* **42**, 1109-12 (2010).
37. Martin, H.C. *et al.* Quantifying the contribution of recessive coding variation to developmental disorders. *Science* **362**, 1161-1164 (2018).
38. Kahrizi, K. *et al.* Effect of inbreeding on intellectual disability revisited by trio sequencing. *Clin Genet* **95**, 151-159 (2019).
39. El-Hattab, A.W. *et al.* Clinical characterization of int22h1/int22h2-mediated Xq28 duplication/deletion: new cases and literature review. *BMC Med Genet* **16**, 12 (2015).
40. Yon, D.K., Park, J.E., Kim, S.J., Shim, S.H. & Chae, K.Y. A sibship with duplication of Xq28 inherited from the mother; genomic characterization and clinical outcomes. *BMC Med Genet* **18**, 30 (2017).
41. Haack, T.B. *et al.* Absence of the Autophagy Adaptor SQSTM1/p62 Causes Childhood-Onset Neurodegeneration with Ataxia, Dystonia, and Gaze Palsy. *Am J Hum Genet* **99**, 735-743 (2016).
42. Jeong, S.J. *et al.* A threonyl-tRNA synthetase-mediated translation initiation machinery. *Nat Commun* **10**, 1357 (2019).
43. Theil, A.F. *et al.* Bi-allelic TARS Mutations Are Associated with Brittle Hair Phenotype. *Am J Hum Genet* **105**, 434-440 (2019).
44. Williams, M., Yen, W., Lu, X. & Sutherland, A. Distinct apical and basolateral mechanisms

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

drive planar cell polarity-dependent convergent extension of the mouse neural plate. *Dev Cell* **29**, 34-46 (2014).

45. Lei, Y. *et al.* Variants identified in PTK7 associated with neural tube defects. *Mol Genet Genomic Med* **7**, e00584 (2019).
46. Meier, N. *et al.* Exome sequencing of fetal anomaly syndromes: novel phenotype-genotype discoveries. *Eur J Hum Genet* **27**, 730-737 (2019).
47. Wei, F.Y. *et al.* Cdk5rap1-mediated 2-methylthio modification of mitochondrial tRNAs governs protein translation and contributes to myopathy in mice and humans. *Cell Metab* **21**, 428-42 (2015).
48. Nave, K.A. & Werner, H.B. Myelination of the nervous system: mechanisms and functions. *Annu Rev Cell Dev Biol* **30**, 503-33 (2014).
49. Bercury, K.K. & Macklin, W.B. Dynamics and mechanisms of CNS myelination. *Dev Cell* **32**, 447-58 (2015).
50. Li, M. *et al.* Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* **362**(2018).
51. Lake, B.B. *et al.* Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat Biotechnol* **36**, 70-80 (2018).
52. Noma, A., Kirino, Y., Ikeuchi, Y. & Suzuki, T. Biosynthesis of wybutosine, a hyper-modified nucleoside in eukaryotic phenylalanine tRNA. *EMBO J* **25**, 2142-54 (2006).
53. Young, A.P. & Bandarian, V. TYW1: A Radical SAM Enzyme Involved in the Biosynthesis of Wybutosine Bases. *Methods Enzymol* **606**, 119-153 (2018).
54. El Yacoubi, B., Bailly, M. & de Crecy-Lagard, V. Biosynthesis and function of posttranscriptional modifications of transfer RNAs. *Annu Rev Genet* **46**, 69-95 (2012).
55. Landgraf, B.J., McCarthy, E.L. & Booker, S.J. Radical S-Adenosylmethionine Enzymes in Human Health and Disease. *Annu Rev Biochem* **85**, 485-514 (2016).
56. Rak, R., Dahan, O. & Pilpel, Y. Repertoires of tRNAs: The Couplers of Genomics and Proteomics. *Annu Rev Cell Dev Biol* **34**, 239-264 (2018).
57. Schaffer, A.E., Pinkard, O. & Coller, J.M. tRNA Metabolism and Neurodevelopmental Disorders. *Annu Rev Genomics Hum Genet* **20**, 359-387 (2019).
58. Carlson, B.A. *et al.* Transfer RNA modification status influences retroviral ribosomal frameshifting. *Virology* **255**, 2-8 (1999).
59. Izraeli, S. *et al.* The SIL gene is required for mouse embryonic axial development and left-right specification. *Nature* **399**, 691-4 (1999).
60. Kumar, A., Girimaji, S.C., Duvvari, M.R. & Blanton, S.H. Mutations in STIL, encoding a pericentriolar and centrosomal protein, cause primary microcephaly. *Am J Hum Genet* **84**, 286-90 (2009).
61. Vulprecht, J. *et al.* STIL is required for centriole duplication in human cells. *J Cell Sci* **125**, 1353-62 (2012).
62. Leidel, S., Delattre, M., Cerutti, L., Baumer, K. & Gonczy, P. SAS-6 defines a protein family required for centrosome duplication in C. elegans and in human cells. *Nat Cell Biol* **7**, 115-25 (2005).
63. Strnad, P. *et al.* Regulated HsSAS-6 levels ensure formation of a single procentriole per centriole during the centrosome duplication cycle. *Dev Cell* **13**, 203-13 (2007).
64. van Breugel, M. *et al.* Structures of SAS-6 suggest its organization in centrioles. *Science* **331**,

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1196-9 (2011).

65. Putkey, F.R. *et al.* Unstable kinetochore-microtubule capture and chromosomal instability following deletion of CENP-E. *Dev Cell* **3**, 351-65 (2002).
66. Yao, X., Abrieu, A., Zheng, Y., Sullivan, K.F. & Cleveland, D.W. CENP-E forms a link between attachment of spindle microtubules to kinetochores and the mitotic checkpoint. *Nat Cell Biol* **2**, 484-91 (2000).
67. Yen, T.J., Li, G., Schaar, B.T., Szilak, I. & Cleveland, D.W. CENP-E is a putative kinetochore motor that accumulates just before mitosis. *Nature* **359**, 536-9 (1992).
68. Schmiesing, J.A., Gregson, H.C., Zhou, S. & Yokomori, K. A human condensin complex containing hCAP-C-hCAP-E and CNAP1, a homolog of Xenopus XCAP-D2, colocalizes with phosphorylated histone H3 during the early stage of mitotic chromosome condensation. *Mol Cell Biol* **20**, 6996-7006 (2000).
69. Martin, C.A. *et al.* Mutations in genes encoding condensin complex proteins cause microcephaly through decatenation failure at mitosis. *Genes Dev* **30**, 2158-2172 (2016).
70. Karasawa, K., Tanigawa, K., Harada, A. & Yamashita, A. Transcriptional Regulation of Acyl-CoA:Glycerol-sn-3-Phosphate Acyltransferases. *Int J Mol Sci* **20**(2019).
71. Xu, H. *et al.* Hepatic knockdown of mitochondrial GPAT1 in ob/ob mice improves metabolic profile. *Biochem Biophys Res Commun* **349**, 439-48 (2006).
72. Hammond, L.E. *et al.* Mitochondrial glycerol-3-phosphate acyltransferase-deficient mice have reduced weight and liver triacylglycerol content and altered glycerolipid fatty acid composition. *Mol Cell Biol* **22**, 8204-14 (2002).
73. Papatheodorou, I. *et al.* Expression Atlas update: from tissues to single cells. *Nucleic Acids Res* **48**, D77-D83 (2020).
74. Tabula Muris, C. *et al.* Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367-372 (2018).
75. Iwasaki, Y. *et al.* An autopsied case of adult-onset bulbospinalform Alexander disease with a novel S393R mutation in the GFAP gene. *Clin Neuropathol* **34**, 207-14 (2015).
76. Aggarwal, S., Yurlova, L. & Simons, M. Central nervous system myelin: structure, synthesis and assembly. *Trends Cell Biol* **21**, 585-93 (2011).
77. O'Brien, J.S. Stability of the Myelin Membrane. *Science* **147**, 1099-107 (1965).
78. Ozgen, H., Baron, W., Hoekstra, D. & Kahya, N. Oligodendroglial membrane dynamics in relation to myelin biogenesis. *Cell Mol Life Sci* **73**, 3291-310 (2016).
79. Saher, G. *et al.* High cholesterol level is essential for myelin membrane growth. *Nat Neurosci* **8**, 468-75 (2005).
80. Saher, G. & Stumpf, S.K. Cholesterol in myelin biogenesis and hypomyelinating disorders. *Biochim Biophys Acta* **1851**, 1083-94 (2015).
81. Pfrieger, F.W. & Ungerer, N. Cholesterol metabolism in neurons and astrocytes. *Prog Lipid Res* **50**, 357-71 (2011).
82. Wang, H. & Eckel, R.H. What are lipoproteins doing in the brain? *Trends Endocrinol Metab* **25**, 8-14 (2014).
83. Molofsky, A.V. & Deneen, B. Astrocyte development: A Guide for the Perplexed. *Glia* **63**, 1320-9 (2015).
84. Camargo, N. *et al.* Oligodendroglial myelination requires astrocyte-derived lipids. *PLoS Biol* **15**, e1002605 (2017).

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

85. Molina-Gonzalez, I. & Miron, V.E. Astrocytes in myelination and remyelination. *Neurosci Lett* **713**, 134532 (2019).
86. Tracey, T.J., Steyn, F.J., Wolvetang, E.J. & Ngo, S.T. Neuronal Lipid Metabolism: Multiple Pathways Driving Functional Outcomes in Health and Disease. *Front Mol Neurosci* **11**, 10 (2018).
87. Chrast, R., Saher, G., Nave, K.A. & Verheijen, M.H. Lipid metabolism in myelinating glial cells: lessons from human inherited disorders and mouse models. *J Lipid Res* **52**, 419-34 (2011).
88. Teslovich, T.M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707-13 (2010).
89. Johansen, C.T. *et al.* An increased burden of common and rare lipid-associated risk alleles contributes to the phenotypic spectrum of hypertriglyceridemia. *Arterioscler Thromb Vasc Biol* **31**, 1916-26 (2011).
90. Tikkanen, E. *et al.* Association of known loci with lipid levels among children and prediction of dyslipidemia in adults. *Circ Cardiovasc Genet* **4**, 673-80 (2011).
91. Qi, Q., Liang, L., Doria, A., Hu, F.B. & Qi, L. Genetic predisposition to dyslipidemia and type 2 diabetes risk in two prospective cohorts. *Diabetes* **61**, 745-52 (2012).
92. Surakka, I. *et al.* The impact of low-frequency and rare variants on lipid levels. *Nat Genet* **47**, 589-97 (2015).
93. Yang, S. *et al.* Association between the GPAM rs1129555 SNP and serum lipid profiles in the Maonan and Han populations. *Int J Clin Exp Pathol* **11**, 1484-1498 (2018).
94. Brandenburg, L.O., Pufe, T. & Koch, T. Role of phospholipase d in g-protein coupled receptor function. *Membranes (Basel)* **4**, 302-18 (2014).
95. Foster, D.A., Salloum, D., Menon, D. & Frias, M.A. Phospholipase D and the maintenance of phosphatidic acid levels for regulation of mammalian target of rapamycin (mTOR). *J Biol Chem* **289**, 22583-8 (2014).
96. Guizzetti, M. & Costa, L.G. Effect of ethanol on protein kinase Czeta and p70S6 kinase activation by carbachol: a possible mechanism for ethanol-induced inhibition of glial cell proliferation. *J Neurochem* **82**, 38-46 (2002).
97. Samson, H.H. Microcephaly and fetal alcohol syndrome: Human and animal studies. *Alcohol and Brain Development*, 167-183 (1986).
98. Perez-Torrero, E. *et al.* Effects of acute prenatal ethanol exposure on Bergmann glia cells early postnatal development. *Brain Res* **746**, 305-8 (1997).
99. Guizzetti, M., Zhang, X., Goeke, C. & Gavin, D.P. Glia and neurodevelopment: focus on fetal alcohol spectrum disorders. *Front Pediatr* **2**, 123 (2014).
100. Burkhardt, U. & Klein, J. Chapter 26 - Ethanol Impairs Phospholipase D Signaling in Astrocytes. in *Molecular Aspects of Alcohol and Nutrition* (ed. Patel, V.B.) 325-335 (Academic Press, San Diego, 2016).
101. Hart, C.E. *et al.* Phosphoserine aminotransferase deficiency: a novel disorder of the serine biosynthesis pathway. *Am J Hum Genet* **80**, 931-7 (2007).
102. Yao, H.T. *et al.* The inhibitory effect of polyunsaturated fatty acids on human CYP enzymes. *Life Sci* **79**, 2432-40 (2006).
103. Westphal, C., Konkel, A. & Schunck, W.H. CYP-eicosanoids--a new link between omega-3 fatty acids and cardiac disease? *Prostaglandins Other Lipid Mediat* **96**, 99-108 (2011).
104. de Baaij, J.H. The art of magnesium transport. *Magnes Res* **28**, 85-91 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

105. Hu, W.F., Chahrour, M.H. & Walsh, C.A. The diverse genetic landscape of neurodevelopmental disorders. *Annu Rev Genomics Hum Genet* **15**, 195-213 (2014).
106. Jayaraman, D., Bae, B.I. & Walsh, C.A. The Genetics of Primary Microcephaly. *Annu Rev Genomics Hum Genet* **19**, 177-200 (2018).
107. Kiray, H., Lindsay, S.L., Hosseinzadeh, S. & Barnett, S.C. The multifaceted role of astrocytes in regulating myelination. *Exp Neurol* **283**, 541-9 (2016).
108. Irons, M., Elias, E.R., Salen, G., Tint, G.S. & Batta, A.K. Defective cholesterol biosynthesis in Smith-Lemli-Opitz syndrome. *Lancet* **341**, 1414 (1993).
109. Brookes, K.J., Chen, W., Xu, X., Taylor, E. & Asherson, P. Association of fatty acid desaturase genes with attention-deficit/hyperactivity disorder. *Biol Psychiatry* **60**, 1053-61 (2006).
110. Bell, J.G. *et al.* Essential fatty acids and phospholipase A2 in autistic spectrum disorders. *Prostaglandins Leukot Essent Fatty Acids* **71**, 201-4 (2004).
111. Pinto, D. *et al.* Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am J Hum Genet* **94**, 677-94 (2014).
112. Jungerius, B.J. *et al.* An association screen of myelin-related genes implicates the chromosome 22q11 PIK4CA gene in schizophrenia. *Mol Psychiatry* **13**, 1060-8 (2008).
113. Namekawa, M. *et al.* Identification of GFAP gene mutation in hereditary adult-onset Alexander's disease. *Ann Neurol* **52**, 779-85 (2002).
114. Patel, S.C. *et al.* Localization of Niemann-Pick C1 protein in astrocytes: implications for neuronal degeneration in Niemann- Pick type C disease. *Proc Natl Acad Sci U S A* **96**, 1657-62 (1999).
115. Schmitt, A. *et al.* The brain-specific protein MLC1 implicated in megalencephalic leukoencephalopathy with subcortical cysts is expressed in glial cells in the murine brain. *Glia* **44**, 283-95 (2003).
116. Dietrich, J. *et al.* EIF2B5 mutations compromise GFAP+ astrocyte generation in vanishing white matter leukodystrophy. *Nat Med* **11**, 277-83 (2005).
117. Yazdi, M. *et al.* The role of mitochondrial glycerol-3-phosphate acyltransferase-1 in regulating lipid and glucose homeostasis in high-fat diet fed mice. *Biochem Biophys Res Commun* **369**, 1065-70 (2008).

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Online methods

Cohort recruitment and sample preparation

Cohort recruitment

The family history and clinical characteristics of CP individuals were documented regarding the CP phenotypic profiles. The motor signs (spastic, ataxic, dyskinetic, etc.) and anatomical distribution (monoplegia, diplegia, triplegia, hemiplegia, and quadriplegia) of movement disorder were evaluated and categorized. Motor severity was measured by the Gross Motor Function Classification System (GMFCS) on 5 levels [1]. The common co-morbidities, i.e., ID/ASD, vision/hearing impairment, epilepsy, dyslexia, etc., were recorded for each patient. The fine motor skill was evaluated by the Manual Ability Classification System (MACS) [2]. Rehabilitation training outcome was determined by the Activity of Daily Living (ADL) [3], and the Gross Motor Function Measure (GMFM) [4], both of which were measured twice before and after treatment.

Magnetic Resonance Imaging (MRI) and diffusion tensor imaging (DTI)

All MRI imagings were performed on a 3.0T Siemens scanner (Magnetom Systems, Skyra, Munich, Germany). For DTI, an echoplanar sequence with diffusion gradients (b = 1000 s / mm^2) applied in 64 noncollinear directions was used. Basic parameters included: TR (10000 ms) / TE (92 ms), Voxel (2.0 × 2.0 × 2.0 mm), FOV (256 × 256), layer thickness (2.0 mm). DTI measurement was three-dimensional reconstruction of color-coded FA graphics through DTI Studio (DWI standard deviation).

Sample preparation

Genomic DNA from all the members in the CP cohort was extracted from the peripheral blood by using the QIAamp DNA blood mini kit (Qiagen, Frankfurt, Germany) according to the manufacturer's instructions.

PCR amplification was performed by using the Q5 High-Fidelity Polymerase (5X, NEB, MA, USA) with the specific primers targeted to the selected candidate variants listed as below:

TYW1-R206C-Forward: 5'-TTTCCTCTTCCTTAGCGGCA-3',

TYW1-R206C-Reverse: 5'-CAATCCTCCCAACTCCACCT-3';

TYW1-R389Q-Forward: 5'-CTCCAGGATGCATTCCCTTA-3',

TYW1-R389Q-Reverse: 5'-TCGATCACCATGACCTTTCA-3';

GPAM-P669S-Forward: 5'-GAGGGGTGCAGGTGTAGGTA-3',

GPAM-P669S-Reverse: 5'-AGGAGATGGCCTGGAAAAGT-3';

GPAM-G499R-Forward: 5'-CTGAGAACCCCAGGTCAAAA-3',

GPAM-G499R-Reverse: 5'-TGATGGAAGAGACCACACCA-3'.

Then, the PCR products were processed by the Sanger Sequencing to determine the carrier status of the specific variants.

Fragile-X screening and mitochondrial genome sequencing

Fragile-X screening

We performed the fragile X screening for the gene of *FMR1* on 120 families including patients, unaffected siblings, and parents. Fragile-X syndrome (FXS) is the most common monogenic

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

form of intellectual disability and autism spectrum disorder, predominately due to the expansion of a CGG repeat located at the 5' UTR of the *FMR1* gene (OMIM: 309550). Expansion to the full mutation (more than 200 repeats) leads to hypermethylation and silencing of *FMR1*, resulting in the absence of its protein product, FMRP, which causes FXS. A *FMR1* CGG repeat was amplified with the primers (Forward: 5'-GCGCTCAGCTCCGTTTCGGTTTCACTTCC-3' and Reverse: 5'-CCCAAGTCCAGTCCTTCCCTCCCAACAACA-3') by using the LA Taq polymerase kit (RR02BG, TaKaRa, Shiga, Japan) and 50 ng of total genome DNA, then the PCR products were analyzed by gel electrophoresis and Sanger sequencing. Thermal cycling was as follows: denaturation at 96 °C for 3 min and 10 cycles of 98 °C for 20 sec, 65 °C for 45 sec, and 72 °C for 3 min, followed by 22 cycles of 98 °C for 20 sec, 68 °C for 3.5 min, and a final extension at 68°C for 10 min [5]. There is no full mutation observed in this cohort.

Mitochondrial genome sequencing

We conducted very-high-coverage (6000X) mitochondrial DNA sequencing for all CP patients. The 16.6 kb circular mitochondrial genome encodes 13 protein subunits of the electron transport chain, 22 tRNA and 2 rRNA genes [6]. The mitochondrial genome was amplified by a single-amplicon LR-PCR using a LA Taq polymerase kit (RR02BG, TaKaRa, Shiga, Japan) and 100 ng of total genome DNA [7]. Indexed paired-end DNA libraries were prepared and sequenced by Nextseq500 sequencer (Illumina, CA, USA), according to the manufacturer instructions. The sequenced data was analyzed by the NextGENe (SoftGenetics, State College, PA) with the mitochondrial reference sequence (NC_012920). The average sequencing coverage of mitochondrial genome was 6000X, and the variants with abundance >2% were reported, because of the well-known heteroplasmy of mitochondrial variants. The pathogenicity of a specific variant was defined according to the guideline recommended by the American College of Medical Genetics and Genomics (ACMG) [8]. The pathogenic mitochondrial variants were validated by PCR and Sanger sequencing of the total DNAs from the patients and the biological mothers.

High-density cytogenetic microarray

High-density cytogenetic microarrays for all CP patients were performed by using the CytoScan HD Array platform (Affymetrix, CA, USA) with more than 2.6 million markers including 0.75 million single nucleotide polymorphisms. CNV detection was done by using the Chromosome Analysis Suite (version 3.0, Affymetrix, CA, USA). The genomic coordinates were based on the Human Genome Build GRCh37/hg19. Pathogenic CNVs and likely pathogenic CNVs were defined as the rules recommended by the American College of Medical Genetics and Genomics guidelines (ACMG) [9]. CNVs were compared in the Database of Genomic Variants (http://dgv.tcag.ca/dgv/app/), CAGdb (Cytogenomics Array Group Database) (http://www.cagdb.org/), ISCA (International Standards for Cytogenomic Arrays) (http://dbsearch.clinicalgenome.org/), DECIPHER (https://decipher.sanger.ac.uk/), and an in-house 2,247 control samples. Pathogenic and likely pathogenic CNVs were confirmed in the patients and parental samples using a SYBR Green-based real-time quantitative PCR assay (A25778, Thermo Fisher Scientific, MA, USA).

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Whole genome sequencing and ultra-high-depth whole exome sequencing

Whole genome sequencing (WGS) was conducted for all 385 members of the 120 CP families including 122 patients, their parents and unaffected siblings, on a HiSeq XTen Deep Sequencer (Illumina, CA, USA), with an average coverage of ~36X, according to the manufacturer's manuals. The primary data analysis was done by a combination of BWA sequence aligner for reads alignment to human reference genome GRCh37/hg19, GATK pipeline for calling single-nucleotide variant (SNV) and small insertion and deletion (indel), and ANNOVAR for variant characterization [10-12].

The 66 patients with no obvious causal variant underwent further assessment with the ultra-high-depth whole-exome sequencing (UWES) in order to capture post-zygotic mutation (PZM) and those variants located in poorly-represented coding regions in WGS. The UWES procedures were performed on a HiSeq4000 deep sequencer (Illumina, CA, USA) with a 150bp paired-end protocol after enrichment in the coding regions by the SureSelect Human All Exon V5 kits (Agilent, CA, USA). The average coverage of UWES for CP patients reached more than 600X. The generated raw sequences were processed by the MERAP package for alignment, quality control, calling single nucleotide variant (SNV), insertion and deletion (Indel), structural variation (SV), and copy number variation (CNV), plus the variants annotation and prioritization [13].

The sequencing performances of WGS and UWES were recorded in Supplementary table 2.

Analysis pipeline of deep sequencing data

The variant prioritization and evaluation of etiologic involvement were performed by two pipelines, namely, MERAP and ANNOVAR. In brief, MERAP pipeline first filters all identified variants through comparison with the disease-associated variants in the Human Gene Mutation Database (HGMD, 2020.2) and the Online Mendelian Inheritance in Man (OMIM) to collect those known disease-causing variants. In order to filter out neutral variants, MERAP uses entries from the dbSNP143 (http://www.ncbi.nlm.nih.gov/projects/SNP/), 1000 Genome (http://www.1000genomes.org/), NHLBI Exome Sequencing Project (ESP, http://evs.gs.washington.edu/EVS/), ExAc (http://exac.broadinstitute.org/), and the Chinese Millionome Database (CMDB) (https://db.cngb.org/cmdb/) as screening databases. In principle, candidate variants causing recessive traits should not occur in healthy controls as homozygotes, and the frequency of respective heterozygotes should not exceed 0.1%. MERAP uses the RefSeq genes (http://www.ncbi.nlm.nih.gov/refseq/) as reference, and nonsynonymous changes are described in terms of gene ID, base change, protein change, genomic coordinate, transcript coordinate, protein coordinate, protein length, affiliated with gene description from the Human Gene Nomenclature Committee (HGNC, http://www.genenames.org/). Changes destroying conventional splice sites or introducing novel splice sites are identified by MERAP's module called SSFinder. To assess the pathogenicity of missense mutations, MERAP generates a single score integrating the results of seven different algorithms, including the Grantham score, PhyloP, GERP, SIFT, PolyPhen2, Mutation-Taster, and the Conserved Domains Database (CDD, http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml). With empirical false discovery rate cutoffs, this score serves as dichotomized pathogenicity predictions even if any two of the

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

seven algorithms might not coincide, as is often the case. MERAP rules out candidate genes reported to harbor homozygous loss-of-function (LOF) variants in healthy individuals, which applies to >1% of the human genes. Typically, if more than three independent truncating variants are observed in >10 of the exomes listed in the 1000 Genome, Exome Variant Server (ESP6500), NHLBI GO Exome Sequencing Project (ESP) (http://evs.gs.washington.edu/EVS/), and ExAC (http://exac.broadinstitute.org/) databases, the relevant gene is flagged as LOF tolerant. To facilitate the choice between few remaining candidate genes, MERAP also provides a list of ~4,500 known disease genes extracted from OMIM (http://www.ncbi.nlm.nih.gov/omim), ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/), and HGMD (http://www.hgmd.org/), as well as more than 8,000 associated disorders and their symptoms. For novel candidate genes without prior link to disease, MERAP offers information on their interaction with known disease genes, based on data from Biogrid (http://thebiogrid.org/) and IntAct (http://www.ebi.ac.uk/intact/), assuming that genes implicated in clinically similar disorders tend to cluster in gene or protein interaction networks. Variants were considered *de novo* if neither parent had the variant, and candidate variants were selected by segregation analysis. The pathogenic and likely pathogenic genes/variants were defined according to the standards and guidelines of ACMG [8].

The ANNOVAR pipeline was also used, with additional considerations from the Residual Variation Intolerance Score (RVIS) and the Combined Annotation-Dependent Depletion (CADD) score [11]. The former ranks genes in terms of intolerance to functional genetic variation and the latter integrates several well-known tools. We empirically set a cutoff RVIS score of 50th percentile for known and novel genes and a cutoff CADD score of 20 for novel candidate genes. When defining likely causal variants/genes, we followed the guidelines designated by ACMG.

All variants of putative clinical relevance were confirmed by the conventional PCR and Sanger sequencing. Parent-child relationships were confirmed by using the PLINK with SNPs drawn from WGS which matched the CytoScan SNP repertoire [14]. The false positive rate of WGS was generated by the comparison between Sanger sequencing and WGS, and we found no false positive variant called by our pipeline. The false negative rate of WGS was evaluated by the SNP genotyping comparison between WGS and CytoScan high-density microarray, which resulted in an average false negative rate of less than 0.1%.

Gene-set enrichment, pathway analysis, gene expression analysis and protein-protein interaction analysis

Gene ontology and KEGG pathway analysis

Gene ontology analysis was performed on 117 CP-related genes and 706 NDD-related genes [15] separately using GO enrichment analysis tool of Omicshare (https://www.omicshare.com/tools/home/report/goenrich.html). The top 27 GO terms with p-value < 0.01 were selected and presented. To NDD-related genes, the same GO terms were filtered with p-value < 0.01, and only those with rich factor > 0.09 (for Biological process and cellular component) and 0.07 (for Molecular function) were indicated.

KEGG pathway analysis was performed on 117 CP-related genes and 706 NDD-related genes [15] separately using pathway enrichment analysis tool of Omicshare (https://www.omicshare.com/tools/home/report/koenrich.html). The top 27 pathways with

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

p-value < 0.05 were selected and presented. To NDD-related genes, same pathways were filtered with p-value < 0.05, and only those with rich factor (the ratio between number of genes from the present gene-set and all genes enriched in a certain pathway) > 0.1 were indicated.

Expression analysis of CP genes

To visualize the expression pattern of CP-related genes in different brain regions, we extracted the bulk RNA-seq data of CP-related genes from the published data of Li *et al.* [16]. On the other hand, to investigate the expression profile of CP-related genes in different cell types, we used the R package Seurat to import, filter, normalize and scale (with default settings) the single cell RNA-seq data for human brain obtained from the published data of Li *et al.* [16] and Lake *et al.* [17]. The expression level of CP-related genes in different brain regions and different cell types were displayed in the form of heatmaps by using TBtools software [18].

Protein-protein interaction network

Protein-protein interaction network between CP-related genes and NDD-related genes was obtained from the STRING database [19] (https://string-db.org/). The interaction network model was generated with Cytoscape [20].

Homology modeling

Domains of TYW1, GPAM, CDK5RAP1, PTK7 and TARS were predicted by the InterProScan protein domains identifier [21] (http://www.ebi.ac.uk/interpro/search/sequence-search). For GPAM, PTK7 and TARS, unknown domains were also predicted with the MEME (Multiple EM for Motif Elicitation) motif discovery tool [22] (http://meme-suite.org/tools/meme). Multiple sequence alignments of TYW1, GPAM, CDK5RAP1, PTK7 and TARS were generated by using the ClustalW program [23].

Three-dimensional structural models of TYW1 (TYW1_WT, TYW1_R206C and TYW1_R389Q), GPAM (GPAM_WT, GPAM_G499R and GPAM_P669S), CDK5RAP1 (CDK5RAP1_WT, CDK5RAP1_M255T and CDK5RAP1_E494V), PTK7 (PTK7_WT, PTK7_G476R and PTK7_P698R) and TARS (TARS_WT, TARS_T258S and TARS_Q702R) were predicted by the I-TASSER web tool [24] (http://zhang.bioinformatics.ku.edu/I-TASSER/). Visual representation of models and structural superposition were generated by a software package named visual molecular dynamics (VMD) [25].

The mutant stability change (ΔΔG) of variants of TYW1 (TYW1_R206C and TYW1_R389Q), GPAM (GPAM_G499R and GPAM_P669S), CDK5RAP1 (CDK5RAP1_M255T and CDK5RAP1_E494V), PTK7 (PTK7_G476R and PTK7_P698R) and TARS (TARS_T258S and TARS_Q702R) were predicted by using the STRUM server [26] (https://zhanglab.ccmb.med.umich.edu/STRUM/).

SH-SY5Y cell modeling

The human neuroblastoma cell line SH-SY5Y was retrieved from the China National Infrastructure of Cell Line Resource (resource code: 3131C0001000700097).

CRISPR/Cas9-mediated *TYW1* gene knockout in SH-SY5Y cells

Gene knockout was performed in SH-SY5Y. Briefly, all-in-one CRISPR/Cas9 system was used

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

to produce lentivirus expressing the single guide RNA (sgRNA) and Cas9 protein (U6-sgRNA-EF1α-Cas9-P2A-Puro). The sgRNA sequence is 5'-GCATCGTGTGATGAGTCGAG-3' for *TYW1* (Edigene, Beijing, China). Sequences were selected with online sgRNA design tool CRISPOR (http://crispor.tefor.net/). SH-SY5Y cells were lentiviral transduced. After 48 h, cells were selected with puromycin at a concentration of 3.5 mg/ml.

Genome editing status in the mixed clonal KO cells was determined by Sanger sequencing analysis with the TIDE web tool [27]. Knockout efficiency was confirmed by RT-qPCR using primers as followed:

TYW1-locus1-Forward: 5'-TGCAGGTGGACAAAGTCC-3',

TYW1-locus1-Reverse: 5'-TTATTAGCACACGCCAAGC;

TYW1-locus2-Forward: 5'- CAGACTGGAACAGCGAAGGGA-3',

TYW1-locus2-Reverse: 5'-TGTATGTCGCAACCAGGAAGA-3';

ACTB-Forward: 5'-TGCGTGACATTAAGGAGAAG-3',

ACTB-Reverse: 5'-CATGATGGAGTTGAAGGTAGT-3'.

Analysis of cell viability and adhesion

Cell viability was examined with CCK-8 detection kit (Doindo, kumamoto, Japan). Cells in each group were seeded into 96-well plate at a density of 5×10^4 cells/well. After 12 h, medium was removed. 100 μl CCK-8 working solution (1:10 in medium) was added in each well. Cells were incubated for 4 h. Cell viability which reflected cell proliferation was detected by OD_{450} with a microplate reader.

In adhesion experiment, 96-well plates were coated with matrigel (Corning, NY, USA). SH-SY5Y cells in the wild-type and *TYW1*-KO groups were both divided into control part and experiment part. Medium was added in background part. Plates were incubated at 37 ℃ for 1 h. In experimental part, cells were washed three times by using washing buffer. Then, staining solution (1:10 in medium) was added in control part, experimental part and also in background part. After 4 h, OD_{450} was measured. The adhesion rate was calculated as following: the adhesion rate of cells (%) = (OD_{450} of experimental part － background) / (OD_{450} of control part-background) × 100. Measurement was recorded on three independent experiments.

Detection of cell migration

Cells in each group were seeded on 6-well plates at a density of 2×10^6 cells/well. After reaching 90% confluence, a straight scratch wound was made using a pipette tip. Then, cells were washed with PBS to remove suspended cells gently and added with new medium. At 0 h, 24 h and 48 h, the wound area was observed under microscope. At 48 h, wound in the wild-type group almost closed. The cell migration was quantified according to the wound area at 48 h relative to the initial wound area by the ImageJ software. Measurement was recorded on three independent experiments.

Induction of cell differentiation

Cells in each group were seeded on 24-well plates coated with matrigel (Corning, NY, USA). After 24 h, 10 μM all-trans retinoid-acid (RA) (Stemcell technologies, Vancouver, Canada) was added to induce the cell differentiation. At day 5 in the presence of RA, cells were treated with

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

50 ng/ml BDNF (Stemcell technologies, Vancouver, Canada) in DMEM without serum. Morphology of cells was observed by using immuno-staining of MAP2 at day 10. Neurites were traced with the NeuroGrowth plugin of ImageJ software to examine the number and total length of branches. Measurement was recorded on three independent experiments.

Zebrafish modeling

All zebrafish experimentation was carried out in accordance with the NIH Guidelines for the care and use of laboratory animals (http://oacu.od.nih.gov/regs/index.htm) and ethically approved by the Administration Committee of Experimental Animals, Jiangsu Province, China [Approval ID: SYXK (SU) 2007–0021].

Zebrafish lines and whole mount *in situ* hybridization

The zebrafish embryos and adults were maintained in the Zebrafish Center of Nantong University under conditions in accordance with our previous protocols [28,29]. The fish lines *Tg(mnx1:GFP)*^{ml2} , *Tg(huC:egfp)* , *Tg(huC:mcherry)* and *Tg(gfap:egfp)* have been described in the previous work [28,29]. The whole mount *in situ* hybridization (WISH) was performed as previously described [30]. The templates for generating antisense RNA probes were amplified from the cDNA library, using specific primers targeting *tyw1* and *gpam* listed as below:

tyw1-Forward: 5'-AAGGCCAGCTGCAAGAATAA-3'

tyw1-Reverse: 5'-GTGTGGTGTCTCCAGCAGAA-3'

gpam-Forward: 5'-TCCTGTTCACAAGTCCCACA-3'

gpam-Reverse: 5'-AGTTCTTGCGGAGCATCCTA-3'

Morpholino and mRNAs Injections

tyw1 splice-blocking Morpholino (MO; Gene Tools, OR, USA) sequence was 5′-AACCTTATTCCCACTTAATGTTACC-3′. *gpam* splice-blocking Morpholino sequence was 5′-GGTGCTACTTTTCTCCAAGCTTACC-3′. The sequence of a standard control MO oligo was 5′-CCTCTTACCTCAGTTACAATTTATA-3′. *tyw1* Translation-blocking Morpholino sequence was 5'-CAGCATCTCATGTACTCTCTCCATC-3'. *gpam* Translation-blocking Morpholino sequence was 5'-ACGTCCATCCCCTCTCTTCAAACCA-3'. The MOs were diluted to 0.3 mM with RNase-free water and injected into the yolk of one to two-cell stage embryos and then raised in E3 medium at 28.5 ℃. The wild-type and mutated cDNAs (*TYW1* (NM_018264) mut1: p.R389Q; *TYW1* mut2: p.R206C; *GPAM* (NM_001244949) mut1: p.G499R; *GPAM* mut2: p.P669S) containing the open reading frame of the zebrafish and human *tyw1* or *gpam* genes were cloned into pCS2+ vector respectively and then were transcribed *in vitro* by using the mMESSAGE mMACHIN Kit (Thermo Fisher Scientific, MA, USA) after the recombinant plasmids linearized with NotI Restriction Enzyme (NEB, MA, USA), and then the capped mRNAs were purified by RNeasy Mini Kit (Qiagen, Frankfurt, Germany). Around 2 nl mRNA were injected at 50 ng/μl into 1/2-cell stage embryos.

RNA isolation, reverse transcription, and PCR

Total RNA was extracted from zebrafish embryos by TRIzol reagent according to the manufacturer's instructions (Invitrogen, Thermo Fisher Scientific, MA, USA). The reverse transcription was carried out according to the standard protocols from manufacturer

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(Fermentas, Thermo Fisher Scientific, MA, USA). The PCR was carried out as previously described to validate the efficiency of splicing-blocking effect of specific morpholino [28] by using the primer sequencing listed as below.

tyw1-Forward: 5'-TTATCGGTGTTGTCGGGTTT-3'

tyw1-Reverse: 5'-CCTCTGCCAATCTGTCTTCC-3'

gpam-Forward: 5'-ACAGTGCGCAAAAAGAGGTC-3'

gpam-Reverse: 5'-TAAGCACGTTCTCCACCACA-3'

Locomotion analysis in zebrafish larvae, microscopy and statistical analysis

The locomotion analysis of 5 dpf zebrafish larvae was carried out by using the DanioVision system (Noldus Information Technology, Wageningen, Netherlands). The confocal imaging was performed by using a TCS-SP8 LSM confocal imaging system (Leica, Wetzlar, Germany). The zebrafish embryos were embedded as we previously did [28]. Photographs of *in situ* hybridization results were taken by using a DP70 camera on an Olympus stereomicroscope MVX10. Statistical comparisons of the data were carried out by student's t-test, and $P < 0.05$ was considered statistically significant.

Mice modeling

All mice experiments in this study were approved by the institutional animal care and use committee in the Guangzhou Medical University (registration No. 2019-436, 2019-694). In this study, all mice, either wild-type or mutant, were generated from the C57BL/6J strain and provided by the Cyagen Biosciences (Guangzhou, China).

Generation of knockout mice by CRISPR/Cas9 system

Knockout mouse lines were created by CRISPR/Cas9-mediated genome engineering. To create a *Tyw1* knockout mouse model, exon 2~3 was selected as the target site. gRNA was designed as the following sequences:

gRNA1 (matching reverse strand of gene): GCCTAAGGACCACGTTTCGATGG.

gRNA2 (matching reverse strand of gene): CATGTAAGACCTCACGATCAAGG.

To create a *Gpam* knockout mouse model, exon 2~4 was selected as the target site. gRNA was designed as the following sequences:

gRNA1 (matching reverse strand of gene): CTCCCACGGGAAAACCACGCAGG.

gRNA2 (matching reverse strand of gene): CCCAAGGGGTCACGCACCACAGG.

Cas9 mRNA and gRNA were co-injected into the cytoplasm of fertilized eggs. The microinjected zygotes were cultured in medium until the two-cell stage, and then were transferred into the oviducts of pseudopregnant ICR females at 0.5 d after mating with vasectomized males. The mutant F0 mice were identified by genomic PCR and the DNA sequence was confirmed by Sanger sequencing. Subsequent breeding of the mutant F0 mice generated offspring with desired genotypes for experiments.

The primers used for genotyping and gene editing evaluation were listed as below. Gene knockout efficiency was identified using qPCR (supplementary figure 16).

Primers for knockout mice models:

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Gene	Allele	Forward primer	Reverse primer	Band size
Tyw1	WT	AGTTCGGCATTCAGTG GATAGC	AATGCCAGCAACC CTCTTTAGC	452 bp
	KO	AGTTCGGCATTCAGTG GATAGC	TGTGTGAACTTGCC AGAGATAGAC	617 or 452 bp
Gpam	WT	GAACATTGTATCCGTG GGATCTG	CTGAAAATGTGAA CCTCATAGAAGC	367 bp
	KO	GAACATTGTATCCGTG GGATCTG	CGACTTAGTACAGA CATGATCCAG	594 or 367 bp
Tyw1 (qPCR)		GTGGGACTTGTCGCCT TTG	GGGAACGAGTGAC CTGCTT	122 bp
Gpam (qPCR)		ACAGTTGGCACAATAG ACGTTT	CCTTCCATTTCAGT GTTGCAGA	139 bp
Actin (qPCR)		GGCTGTATTCCCCTCC ATCG	CCAGTTGGTAACAA TGCCATGT	154 bp

Behavioral tests

All behavioral tests were carried out on both male and female mice at 8 weeks of age, and the balanced gender was kept for both wild-type and *Tyw1*-/- or *Gpam*-/- mice. Each test was conducted at fixed day time (between 8:30 am to 18:30 pm) on each training day. Mice were moved to the testing room 1 h before behavioral testing for acclimation, and those participating in multiple tests were allowed to rest for at least 3 days between two tests. All experimental areas were cleaned with 70% ethanol before the tests and between subjects. All behavioral tests were carried out with the presence of two researchers blinded to the genotype.

Morris water maze test

Morris water maze test was performed to investigate the learning and memory ability of mice. Briefly, a platform was placed at the central zone of one quadrant of the pool below the surface of water. The mice were trained to learn the position of the hidden platform for 4 days. At day 5, the platform was removed. The mice were released into water and allowed to swim for 60 s to search a virtual quadrant centered on the location of the platform. Duration in zone of platform refers to amount of time mice remained in this virtual zone. Frequency of crossing platform refers to number of times mice crossed the virtual zone, while latency to first arrival refers to the time needed for mice to reach the virtual zone at first time, while distance travelled in target quadrant refers to the distance mice swam in the virtual quadrant. Videos were recorded and analyzed by using the SuperMaze V2.0 software (XinRuan, Shanghai, China).

Rotarod test

Rotarod test was performed to indicate motor coordination of mice in each group. Firstly, mice were trained three times at the rate of 30 rpm before test, and each training maintained for 5 min. In the test, the rotation rate increased gradually to reach 40 rpm within 300 s. The test was over when a mouse fell off. Each mouse was detected for 3 times and the mean latency

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

time at rotarod was recorded.

Grip strength test

Grip strength of mice forelimbs was measured with a grip strength test meter (BIOSEB, EB Instruments). During the test, the mice were placed over the grid allowing only forepaws to attach to the grid. After the mice paws grasped the grid, their tails were pulled horizontally until they completely released hold. Each mouse was tested for three times. The readings of grip strength and duration of grasping the grid were recorded and analyzed by using the SuperGSM software (XinRuan, Shanghai, China).

Tissue preparation

Mice were anaesthetized with 50 mg/kg sodium pentobarbital. After perfused with 0.9% NaCl and 4% paraformaldehyde, brains were extracted and post-fixed for 24 h, followed with dehydration by using sucrose. Brains were cut into two hemispheres, and then serial sagittal sections (30 μm thick) were sliced with freezing microtome (CM3050S, Leica, Germany).

Nissl staining

Sections were immersed into 75% ethanol (30 s), dH_2O (30 s) and cresyl violet (2 min). Then, sections were dehydrated with gradient ethanol (75%, 95%, 100%) for 30 s respectively, followed by incubation with xylene and mounted with neutral resins. The images were taken under an inverted microscope.

5-Ethynyl-2'-deoxyuridine labeling

For proliferation assays, pregnant dams on E13.5 were intraperitoneally injected with 5-Ethynyl-2'-deoxyuridine (EdU, 20 mg/kg body weight) 2 h before sacrifice. Embryonic brains were harvested, fixed in paraformaldehyde and immersed in 30% sucrose for dehydration. Brain slices were then stained with Click-iT EdU Imaging Kits (C10339, Invitrogen, Thermo Fisher Scientific, USA). Cells incorporated with EdU were determined under a fluorescence microscope. Percentage of EdU^+ fluorescence area (indicating the cell number of EdU^+) of E13.5 mouse brain cortex was calculated in 3 random fields per coverslip. For migration assays, pregnant dams on E15.5 were intraperitoneally injected with EdU. Brains were harvested after 72 h and stained as above.

Immunostaining

To detect expression of TYW1 and GPAM in human brain tissue, the normal brain tissue microarrays (BNC17011, Biomax, MD, USA) were prepared. According to the instruction from Biomax official website, BNC17011 were derived from normal brain tissue of male or female at the age of 2 to 50, including brain regions such as frontal lobe, apical lobe, occipital lobe, temporal lobe, midbrain, pons, medulla oblongata, thalamus opticus, cerebellum, hippocampus, callositas, optic nerve, and spinal cord. The slides were immersed in xylene and graded ethanol to deparaffinize and rehydrate. After antigen retrieval using microwave with sodium citrate buffer, 0.3% Triton-X-100 was added to permeabilize the tissue. The slides were blocked with goat serum and incubated with primary antibodies at 4 ℃ overnight. After washing with PBS, the sections were incubated with secondary antibodies.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

For mouse brain tissues, the perfused brain samples were fixed within 4% paraformaldehyde (PFA) in 0.01 M PBS at 4 ℃ for 24 h, then washed by PBS twice , cryoprotected with sucrose gradients, snap frozen, and sectioned with a cryostat (Leica, Wetzlar, Germany) to thickness of 30 μm. Then the sections were immuno-stained by using aforementioned methods, except deparaffinization and rehydration.

Number of $Cux1^+$ or $Foxp2^+$ cells in frontal cortex and motor cortex were counted on three sagittal cerebrum sections, respectively. $Gfap^+$ and $NeuN^+$ cells in different brain regions were counted in 3 random fields per section. Data were obtained from at least three independent experiments and analyzed with ImageJ software.

In order to visualize and measure the cholesterol level, Filipin staining was performed in tissue sections (10μm thick) by using a cell-based Cholesterol Assay Kit (ab133116, Abcam, UK). Briefly, the tissue sections were washed (3 × 5 min) with Cholesterol Detection wash buffer, and then the Filipin III was added to each section and incubated in the dark for 60 min at room temperature. After washing, the fluorescence images were obtained by a Leica DMi8 (Leica Microsystems, Wetzlar, Germany) fluorescence microscope by using × 20 HC PL FLUOTAR objective.

Western blot analysis

Total protein was extracted from peripheral blood using cell lysate containing RIPA and protease inhibitor cocktail. And the brain tissues were lysed with 2% SDS in PBS with PMSF and proteinase inhibitor cocktail. The BCA assay was used to determine protein concentration. Proteins were resolved on 7.5%, 10% or 15% tris-glycine gels based on different molecular weight and transferred to PVDF membranes. After blocking, the membranes were incubated with primary antibodies, followed by horseradish peroxidase (HRP)-labeled secondary antibodies. Then, blots were visualized by West Pico Plus Chemiluminescent Substrate (Thermo Fisher Scientific, MA, USA) and scanned using ChemiDoc™ MP system (Bio-Rad, CA, USA). Densitometries of individual blot signals were quantified using ImageJ software.

qRT-PCR analysis

Total RNA was extracted from the mice brain tissues by using TRIZOL reagent (Invitrogen, Thermo Fisher Scientific, MA, USA), and cDNA was synthesized from 1 μg total RNA by using PrimeScript™ RT Master Mix (RR036Q, Takara, Shiga, Japan). Quantitative PCR was carried out by using PowerUp™ SYBR™ Green Master Mix (A25778, Applied Biosystems, Thermo Fisher Scientific, MA, USA) on Biosystems QuantStudio 6 Flex Real-time PCR system (Applied Biosystems, Thermo Fisher Scientific, MA, USA). *β -Actin* was used as a reference gene. For comparison, mRNA expression level of knockout mice were normalized to those of wild-type mice for each subject. The primers were:

Tyw1-Forward: 5'-GTGGGACTTGTCGCCTTTG-3'

Tyw1-Reverse : 5'- GGGAACGAGTGACCTGCTT-3'

Stil-Forward: 5'- GACACAATTCAGGACTGGTAGAC-3'

Stil-Reverse : 5'- GGCATGATCCACTTTCTGTTCA-3'

Sass6-Forward: 5'- ATTCCTTTACGCGGACTTAGC-3'

Sass6-Reverse : 5'- AAGTAGGCTGAAGACGAGGAG-3'

Ncapd2-Forward: 5'- AGCCAGACAAGCCTCATTGAC-3'

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Ncapd2-Reverse: 5'- TCCATAGGTGACGGATGTCCA-3'

Cenpe-Forward: 5'- CTTCAGTGGCTGTCTGTGTTC-3'

Cenpe-Reverse: 5'- CCATCGCTCTGATAAATAGCGTT-3'

β -Actin -Forward: 5'- GGCTGTATTCCCCTCCATCG -3'

β -Actin -Reverse: 5'- CCAGTTGGTAACAATGCCATGT -3'

Culture of primary neurons

E13.5 mouse embryo brains were taken out from *Tyw1*+/- pregnant dam. Cortex in each brain was dissected separately and collected in Hibernate-E supplemented with 2% B27 on ice. Single cells were obtained by using 0.05% trypsin (containing 0.2 mM EDTA) digestion for 10 min at 37 ℃. After filtration with 70 μm strainer and centrifugation, cells resuspended in Neurobasal medium with 2% B27 Plus Supplement, 0.25% Glutamax and 25 μM glutamate were placed in poly-D-lysine coated plates. Adherent neurons were prepared for EdU and TUNEL staining. To prepare migration assay, isolated cells were initially cultured in ultra-low attachment plates to form neurospheres, which could be digested by using 0.05% trypsin without damage of neurites.

Proliferation and apoptosis assays of primary neurons

At 2 DIV, cells were labeled with 10 μM EdU solution and incubated for 4 h. Cells were treated by using 4% PFA, followed by 0.5% TritonX-100 permeabilization. After being washed twice with 3% BSA, cells were incubated with Click-iT reaction cocktail according to protocol of Click-iT EdU Imaging Kits (C10339, Invitrogen, Thermo Fisher Scientific, MA, USA) and stained with DAPI. Dead cells were labeled using *In Situ* Cell Death Detection Kit, POD (11684817910, Roche, Switzerland).

Migration test of primary neurons

Neurospheres were digested with 0.05% trypsin to obtain single neurons and resuspended in Neurobasal Medium with 1% B27 Plus Supplement. Cells were seeded on the upper layer of a cell culture insert with PET track-etched membrane (8 μm pore size, 353097, Corning, NY, USA) at density of 1×10^5 cells/well. Neurobasal Medium with 2% B27 Plus Supplement was added into the bottom of the lower chamber. After 16 h, the culture insert was taken out and the medium was removed carefully. Cells on the upper surface were wiped, while migrating cells in pore and on the lower surface of membrane were fixed, stained with 0.1% crystal violet and observed.

Flow cytometry

Whole brain tissues from *Gpam*-/- mice on P1 (postnatal day 1) were collected, including brain tissues of their wild-type littermate. After digestion with 4 mg/ml papain and 0.1 mg/ml DNase in 37 ℃ shaker for 30 min, suspension was diluted using DMEM/F12 medium containing 5% FBS and filtered with 70 μm strainers. Then, suspension was centrifuged and the cells were fixed in 4% PFA. 0.3% Triton-X100 was used for permeabilization. After being blocked with 5% goat serum, cells were incubated with mouse anti-Gfap antibody and rabbit anti-Ki67 antibody or rabbit anti-Caspase-3 antibody at 4 ℃ for 1 h. Cells were washed with DPBS and incubated with Alexa Fluor 488 goat anti-mouse IgG and Alexa Fluor 647 goat anti-rabbit IgG

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

at 4 ℃ for 30 min. After being washed for two times, cells were suspended by using DPBS and analyzed on the flow cytometer (BD FACSCanto, BD Biosciences, CA, USA).

Measurement of phosphatidic acid (PA)

Phosphatidic acid was measured by using the PicoProbe^{TM} Phosphatidic Acid Assay Kit (BioVision, CA, USA). In brief, medulla of wild-type and *Gpam*-/- mice was homogenized in PA assay buffer. Lipid extraction was obtained according to the protocol and solubilized in 5% Triton X-100 solution. “Sample background control” and “Sample” were prepared in parallel. Standard curve was generated by using PA standard solution. Converter mix was only added in sample and standard wells. All wells were incubated at 45 ℃ for 1 h. Reaction mix was added in each well and incubate at 37 ℃ for 30 min. Fluorescence was recorded at Ex/Em = 535/587 nm and PA concentration was expressed as nmol PA per mg tissue weight. Value of PA concentration was normalized.

Measurement of phosphatidylcholine (PC)

Phosphatidylcholine was measured by using the phosphatidylcholine assay kit (Abcam, UK). For each individual, about 10 mg brain tissues were washed with cold PBS, resuspended in the assay buffer, and homogenized on ice. After 10 min incubation, samples were centrifuged for 5 min at 4 °C at 16,000 g. The supernatant was incubated with the reaction mix including OxiRed Probe supplemented with hydrolysis enzyme for 30 min. The colorimetric reading was measured at OD_{570} nm on a microplate reader.

Measurement of Phosphatidylethanolamine (PE)

Phosphatidylethanolamine was measured by using the phosphatidylethanolamine assay kit (Abcam, UK) according to the manufacturer’s instruction. Briefly, solubilized lipids were extracted from brain tissues with 5% Triton X-100, incubated with converter mix at 45°C for 1 hour, and then with reaction mix at 40 °C for 3 hours. Fluorescence was recorded at Ex/Em 535/587 nm.

Modeling the effect of *TYW1* knockout on ribosomal frameshift and the calculation of attenuation coefficient for proteins

The modeling of *TYW1*-knockout effect on ribosomal frameshift was based upon the fact that TYW1 is a critical enzyme involved in the synthesis of wybutosine, so the hypomorphic and null alleles of *TYW1* could reduce and even remove the production of wybutosine accordingly [31,32]. It is known that wybutosine at the 37 position of $tRNA^{Phe}$ near the 3’ base of anticodon can stabilize the codon-anticodon interaction thus drastically reduce the ribosomal frameshift upon a specific codon UUU [33]. We, therefore, hypothesized that the density and distribution of codon UUU along a specific mRNA should determine the extent of influence exerted by wybutosine, and in parallel, the relative abundance of wybutosine was another major effector during this process. Our modeling procedure went as follows:

1. Download the UCSC RefSeq (refGene) gene models of human, mouse, and zebrafish from https://genome.ucsc.edu/cgi-bin/hgTables. Extract the coding domain sequences (CDS) of each protein-coding gene. If there exists multiple isoforms, choose the longest ones.

2. Label the location of codon UUU, normalize the location values by the length of CDS. By

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

doing this, we harvested a series of normalized location values, i.e., L(1), L(2), ..., for each gene, where 0 < L(n) < 1, n = 1, 2, ...

3. We made the following two assumptions: (1) for a codon UUU at any location, there is a fixed chance of ribosomal frameshift designated by Pf, given a fixed level of wybutosine; (2) for a given codon UUU located at L(n), after frameshifting, the remaining activity of a protein is proportionate to L(n). Thus, the consequence of translating a mRNA holding multiple codon UUU, i.e., the remaining activity of a protein, designated by R, can be estimated as:

$R = \sum (L(n) * Pf * Pr(n)) + Pr(n)$, where $Pr(1) = 1$ and $Pr(n) = Pr(n - 1) * (1 - Pf)$, $n = 1, 2, ...$

4. For the wild-type (WT) and *TYW1*-knockout mice (KO), we defined the R values for WT and KO to be R(WT) and R(KO), respectively. The R value is determined by the density and distribution of codon UUU in a given mRNA, also by Pf (the chance of ribosomal frameshift at a specific location of mRNA, given a fixed level of wybutosine). We set the values of Pf for R(WT) and R(KO) to be 0.12 and 0.35, respectively [33]. Subsequently, we defined the **attenuation coefficient** to be R(KO) / R(WT). An attenuation coefficient reflected the extent of protein activity reduction with *TYW1*-knockout, and the expected values of attenuation coefficient ranged from 0 to 1. The lower the value of attenuation coefficient, the more severe the ribosomal frameshift and protein degradation when *TYW1* is knocked out.

5. After generating a list of attenuation coefficients for each human / mouse / zebrafish protein, we focused on proteins meeting with these criteria: (1) a protein with homologues in all three species; (2) a protein with attenuation coefficient less than 0.34 (i.e., 0.12 / 0.35); (3) a protein with its corresponding gene matching *TYW1* in terms of expression profiles in normal brains; (4) a protein known to be disease-related, with records in HGMD or OMIM.

Reagents

Reagent type	Designation	Source	Identifiers	Dilution
Antibody	Self-generated anti-TYW1 (rabbit polyclonal)	GenScript	Cat no: 2469	1:20
Antibody	anti-GPAM (mouse monoclonal)	Santa Cruz	Cat no:sc-398135	1:500
Antibody	anti-NeuN (mouse monoclonal)	Millipore	Cat no: MAB377B RRID: AB_177621	1:150
Antibody	anti-GFAP (mouse monoclonal)	Millipore	Cat no: MAB360 RRID: AB_11212597	1:400
Antibody	anti-CUX1 (mouse monoclonal)	Abcam	Cat no: ab54583 RRID: AB_941209	1:40
Antibody	anti-FOXP2 (rabbit polyclonal)	Abcam	Cat no: ab16046 RRID: AB_2107107	1:1000
Antibody	anti-MAP2, clone AP20 (mouse monoclonal)	Millipore	Cat no: MAB3418 RRID: AB_94856	1:100
Antibody	anti-MBP (rat monoclonal)	Abcam	Cat no: ab7349 RRID: AB_305869	1:50
Antibody	anti-CNPase (mouse monoclonal)	Millipore	Cat no: MAB326 RRID: AB_2082608	1:200

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Antibody	anti-Ki-67 (rabbit monoclonal)	Cell Signaling Technology	Cat no: 9129 RRID: AB_2687446	1:400
Antibody	anti-Cleaved Caspase-3 (rabbit monoclonal)	Cell Signaling Technology	Cat no: 9664 RRID: AB_2070042	1:800
Antibody	anti-SAS6 (mouse monoclonal)	Santa Cruz Biotechnology	Cat no: sc-81431 RRID: AB_1128357	1:200
Antibody	anti-CENP-E (mouse monoclonal)	Santa Cruz Biotechnology	Cat no: sc-376685 RRID: AB_11150652	1:100
Antibody	anti-STIL (rabbit polyclonal)	Abcam	Cat no: ab89314 RRID: AB_2197878	1:2000
Antibody	anti-NCAPD2 (rabbit polyclonal)	Proteintech	Cat no: 13382-1-AP RRID: AB_2149840	1:600
Antibody	anti-β-actin (D6A8) (rabbit monoclonal)	Cell Signaling Technology	Cat no: 8457 RRID: AB_10950489	1:1000
Antibody	Goat anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 568	Thermo Fisher Scientific	Cat no: A-11036 RRID: AB_10563566	1:600
Antibody	Goat anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	Thermo Fisher Scientific	Cat no: A-11034 RRID: AB_2576217	1:600
Antibody	Goat anti-Rat IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	Thermo Fisher Scientific	A11006， RRID:AB_2534074	1:300
Antibody	Goat anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 568	Thermo Fisher Scientific	Cat no: A-11031 RRID: AB_144696	1:600
Antibody	Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	Thermo Fisher Scientific	Cat no: A-11001 RRID: AB_2534069	1:200
Antibody	Anti-rabbit IgG, HRP-linked Antibody	Cell signaling technology	Cat no: 7074 RRID: AB_2099233	1:4000
Antibody	Anti-rat IgG, HRP-linked Antibody	Cell signaling technology	Cat no: 7077 RRID: AB_10694715	1:1500
Antibody	Anti-mouse IgG, HRP-linked Antibody	Cell signaling technology	Cat no: 7076 RRID: AB_330924	1:4000
Nuclei stainning	4',6-Diamidino-2-phenylindole (DAPI)	Sigma-Aldrich	Cat no: 23670	1:5000
Other reagent	cell-based Cholesterol Assay Kit	Abcam	Cat no: ab133116	1:1000
Other reagent	Phosphatidylcholine Assay kit (Colorimetric/Fluorometric)	Abcam	Cat no: ab83377	
Other reagent	Phosphatidylethanolamine Assay kit	Abcam	Cat no: ab241005	
Other reagent	PicoProbeTM Phosphatidic Acid Assay Kit (Fluorometric)	Biovision	Cat no: K748-100	

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

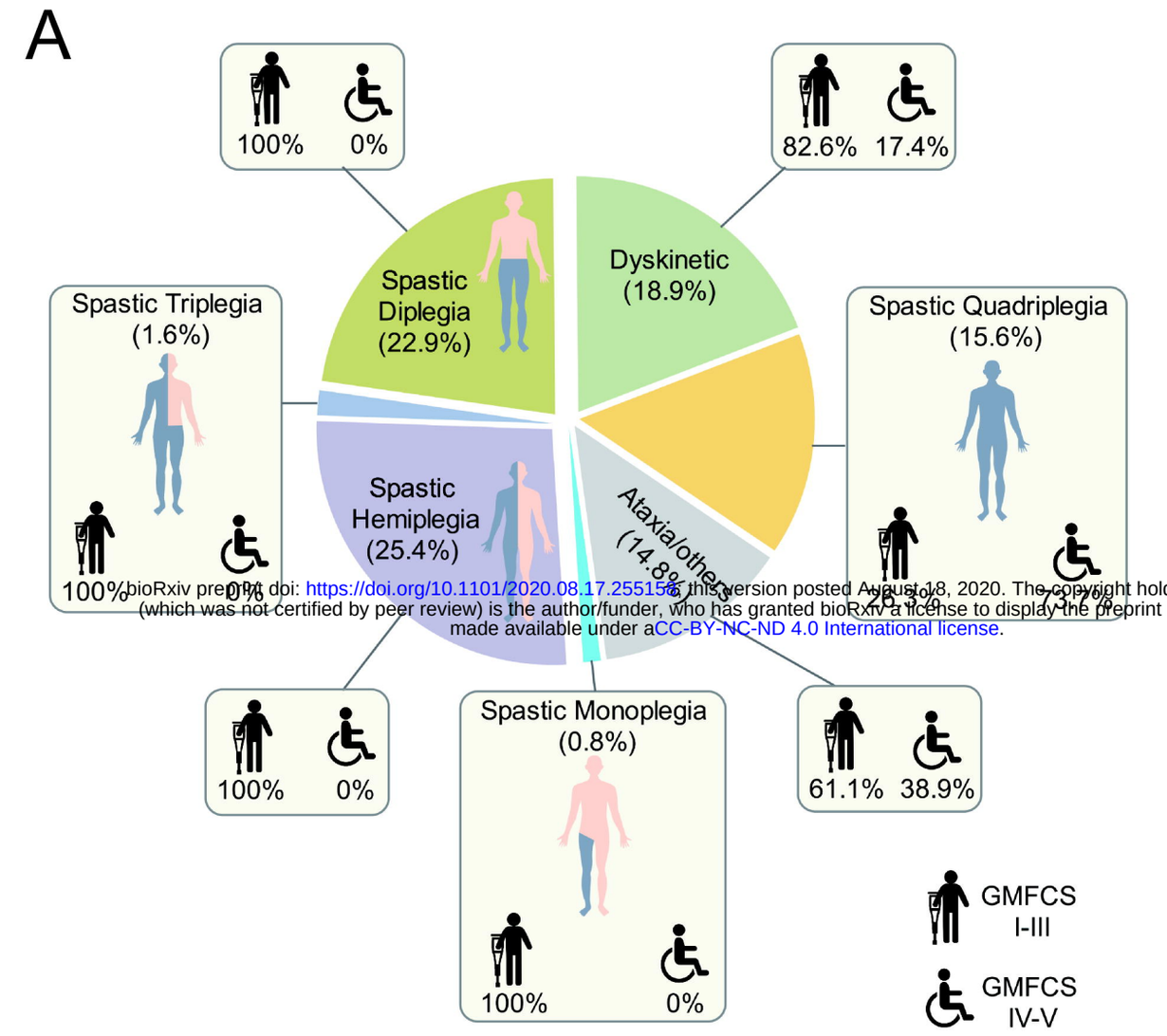
References:

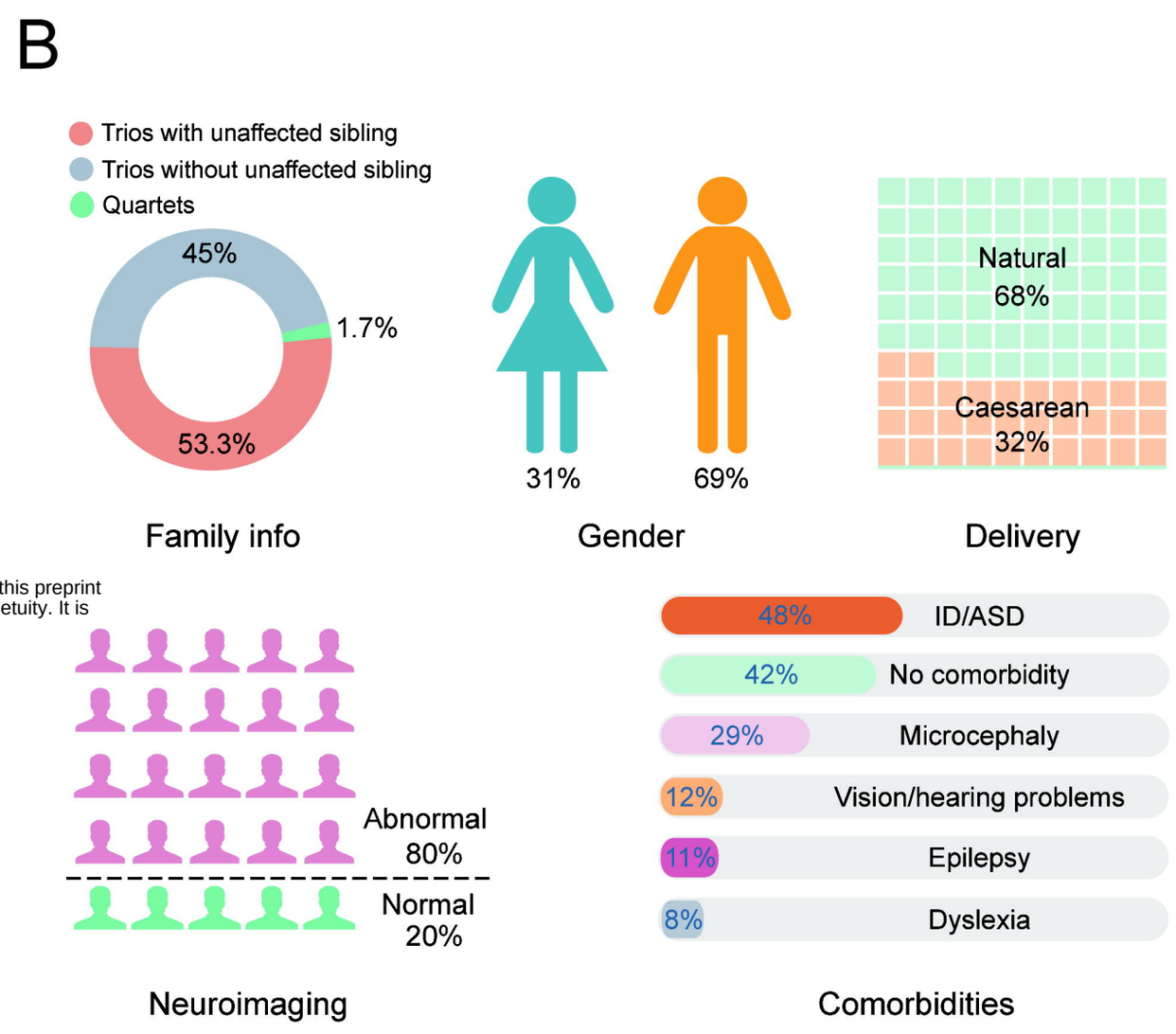
1. Palisano, R. *et al.* Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Dev Med Child Neurol* **39**, 214-23 (1997).
2. Eliasson, A.C. *et al.* The Manual Ability Classification System (MACS) for children with cerebral palsy: scale development and evidence of validity and reliability. *Dev Med Child Neurol* **48**, 549-54 (2006).
3. Xu, K., Wang, L., Mai, J. & He, L. Efficacy of constraint-induced movement therapy and electrical stimulation on hand function of children with hemiplegic cerebral palsy: a controlled clinical trial. *Disabil Rehabil* **34**, 337-46 (2012).
4. Russell, D.J. *et al.* The gross motor function measure: a means to evaluate the effects of physical therapy. *Dev Med Child Neurol* **31**, 341-52 (1989).
5. Wang, T., Bray, S.M. & Warren, S.T. New perspectives on the biology of fragile X syndrome. *Curr Opin Genet Dev* **22**, 256-63 (2012).
6. Wong, L.J. Diagnostic challenges of mitochondrial DNA disorders. *Mitochondrion* **7**, 45-52 (2007).
7. Cui, H. *et al.* Comprehensive next-generation sequence analyses of the entire mitochondrial genome reveal new insights into the molecular diagnosis of mitochondrial DNA disorders. *Genet Med* **15**, 388-94 (2013).
8. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**, 405-24 (2015).
9. South, S.T. *et al.* ACMG Standards and Guidelines for constitutional cytogenomic microarray analysis, including postnatal and prenatal applications: revision 2013. *Genet Med* **15**, 901-9 (2013).
10. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-95 (2010).
11. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).
12. DePristo, M.A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**, 491-8 (2011).
13. Hu, H. *et al.* Integrated sequence analysis pipeline provides one-stop solution for identifying disease-causing mutations. *Hum Mutat* **35**, 1427-35 (2014).
14. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
15. Vissers, L.E., Gilissen, C. & Veltman, J.A. Genetic studies in intellectual disability and related disorders. *Nat Rev Genet* **17**, 9-18 (2016).
16. Li, M. *et al.* Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* **362**(2018).
17. Lake, B.B. *et al.* Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat Biotechnol* **36**, 70-80 (2018).
18. Chen, C., Xia, R., Chen, H. & He, Y. TBtools, a Toolkit for Biologists integrating various biological data handling tools with a user-friendly interface. *BioRxiv*, 289660 (2018).
19. Szklarczyk, D. *et al.* The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* **45**, D362-D368 (2017).

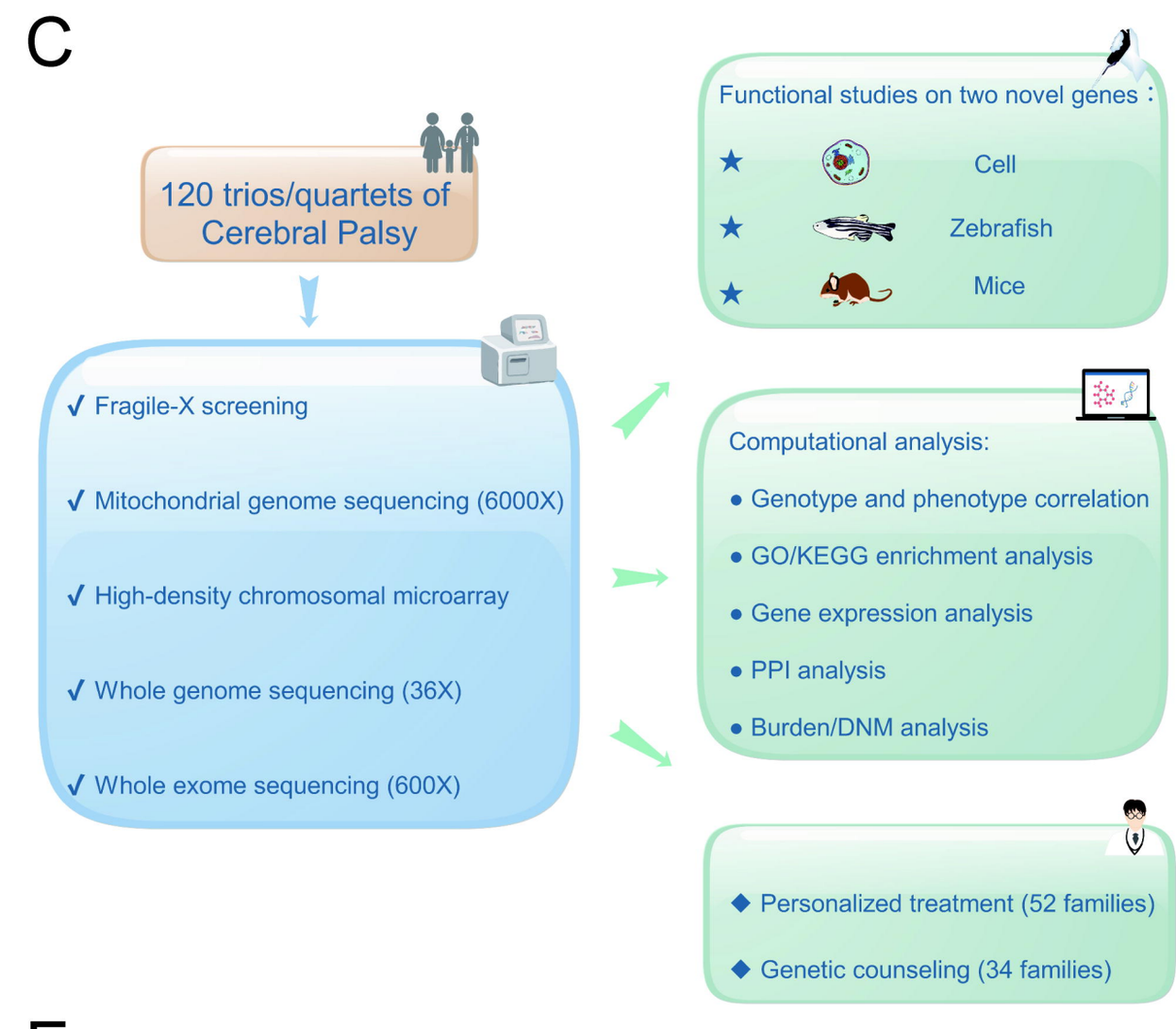
bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

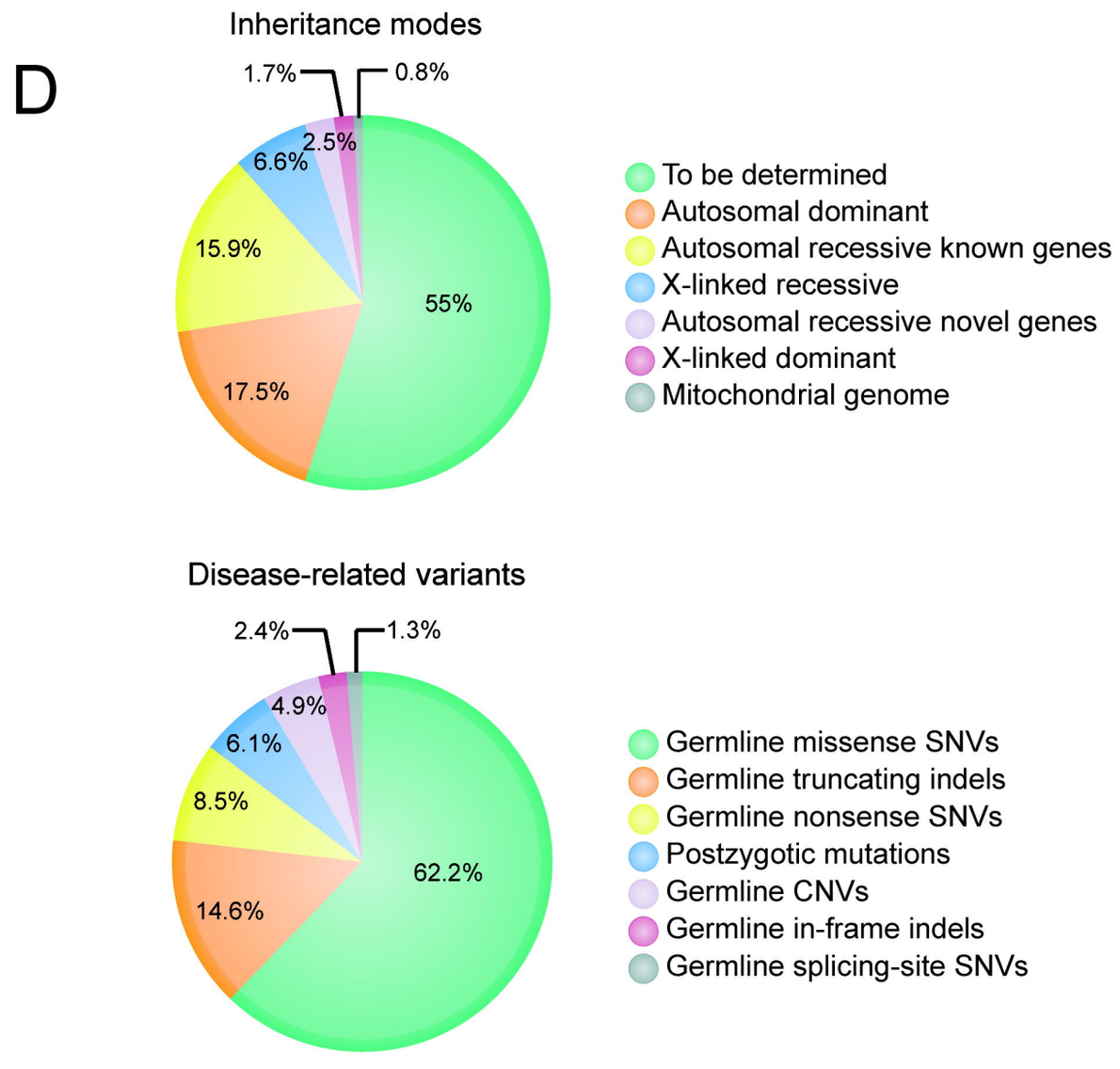
20. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-504 (2003).
21. Quevillon, E. *et al.* InterProScan: protein domains identifier. *Nucleic Acids Res* **33**, W116-20 (2005).
22. Bailey, T.L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**, W202-8 (2009).
23. Thompson, J.D., Gibson, T.J. & Higgins, D.G. Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics* **Chapter 2**, Unit 2 3 (2002).
24. Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nat Methods* **12**, 7-8 (2015).
25. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J Mol Graph* **14**, 33-8, 27-8 (1996).
26. Quan, L., Lv, Q. & Zhang, Y. STRUM: structure-based prediction of protein stability changes upon single-point mutation. *Bioinformatics* **32**, 2936-46 (2016).
27. Brinkman, E.K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res* **42**, e168 (2014).
28. Xu, M. *et al.* Kinesin-12 influences axonal growth during zebrafish neural development. *Cytoskeleton (Hoboken)* **71**, 555-63 (2014).
29. Gong, J. *et al.* Insm1a Regulates Motor Neuron Development in Zebrafish. *Front Mol Neurosci* **10**, 274 (2017).
30. Gong, J., Chai, L., Xu, G., Ni, Y. & Liu, D. The expression of natriuretic peptide receptors in developing zebrafish embryos. *Gene Expr Patterns* **29**, 65-71 (2018).
31. Young, A.P. & Bandarian, V. TYW1: A Radical SAM Enzyme Involved in the Biosynthesis of Wybutosine Bases. *Methods Enzymol* **606**, 119-153 (2018).
32. Noma, A., Kirino, Y., Ikeuchi, Y. & Suzuki, T. Biosynthesis of wybutosine, a hyper-modified nucleoside in eukaryotic phenylalanine tRNA. *EMBO J* **25**, 2142-54 (2006).
33. Carlson, B.A. *et al.* Transfer RNA modification status influences retroviral ribosomal frameshifting. *Virology* **255**, 2-8 (1999).

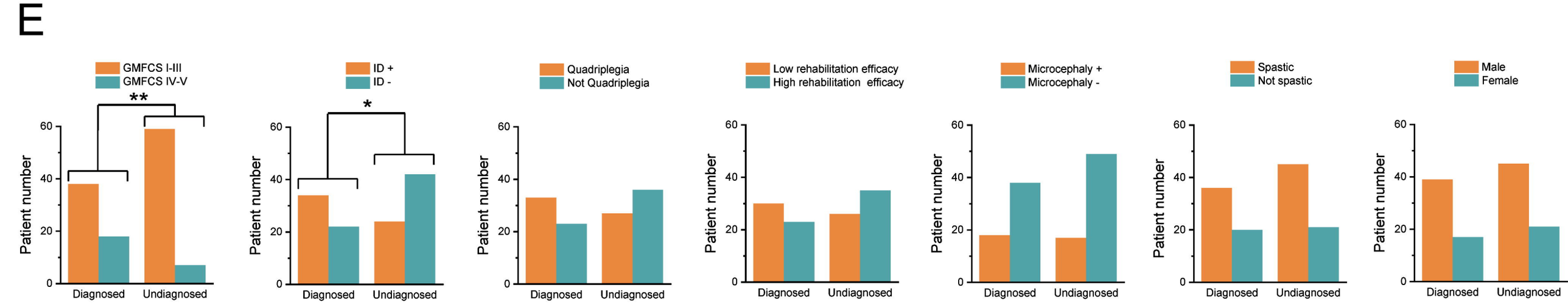
bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.







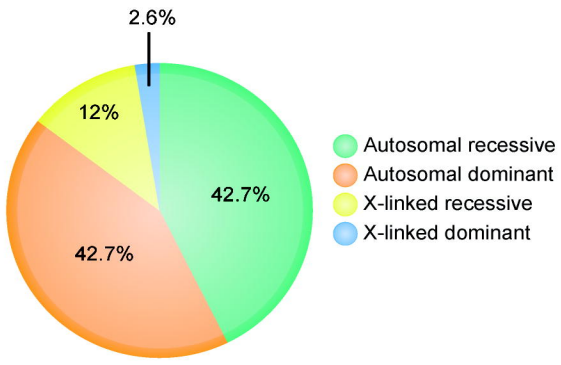




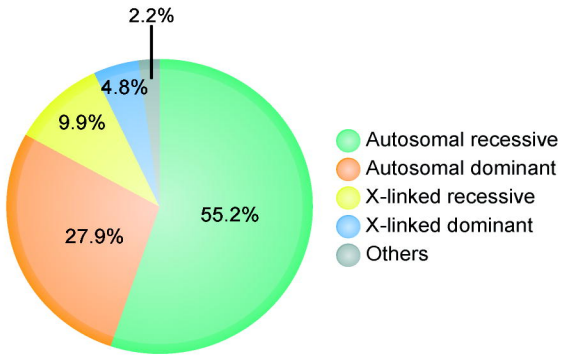
bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A

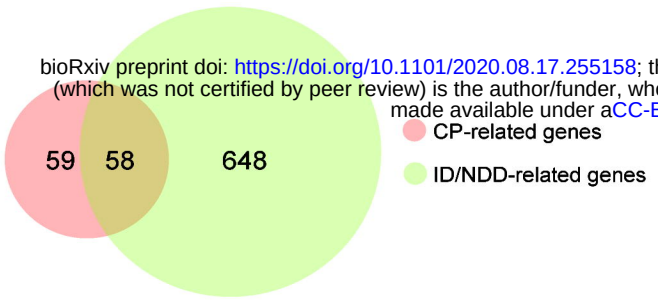
Inheritance modes of CP-related genes



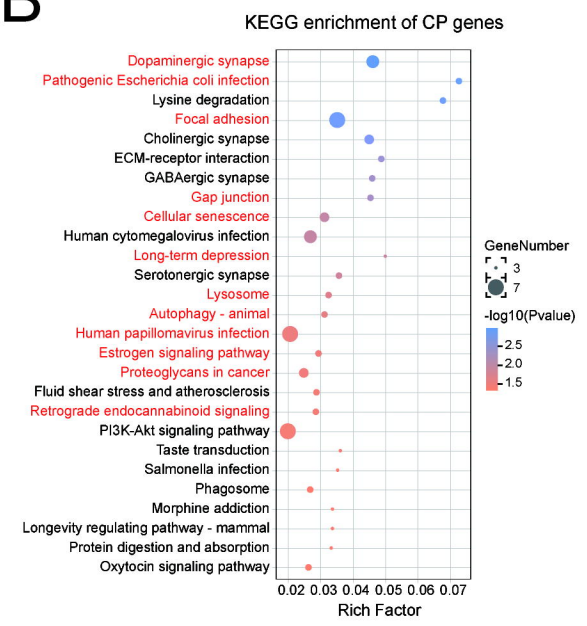
Inheritance modes of ID/NDD-related genes

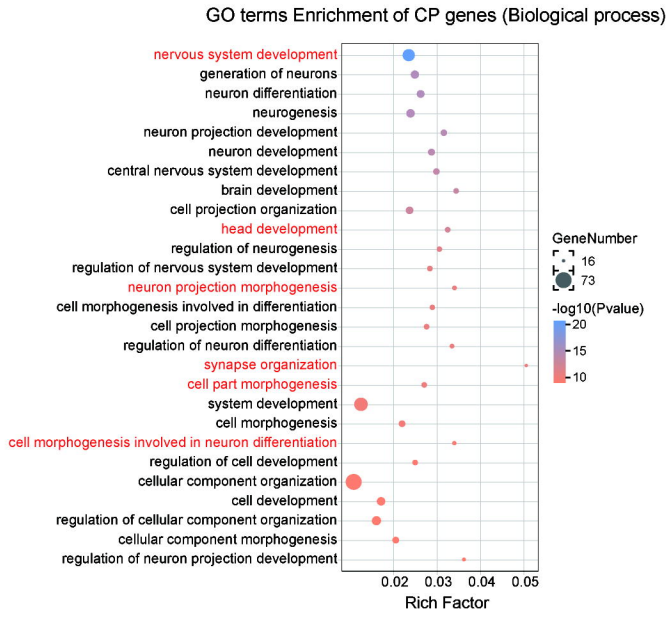


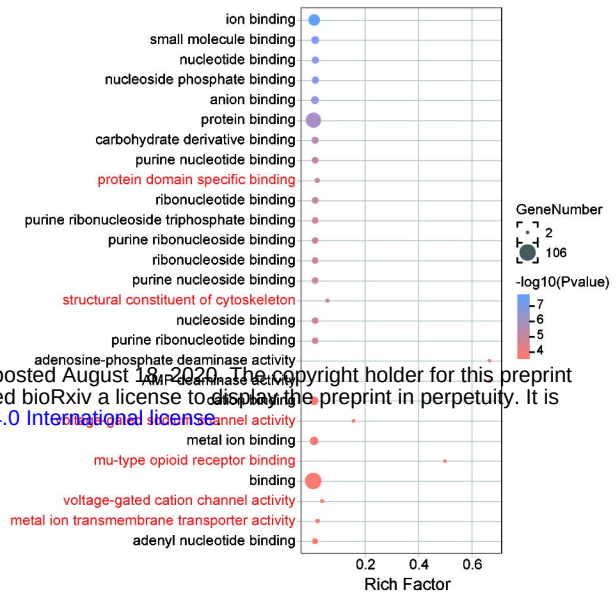
Overlap between CP- and ID/NDD-related genes

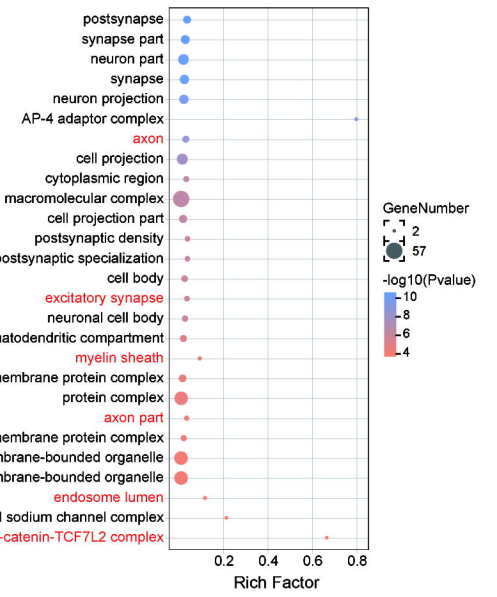


B

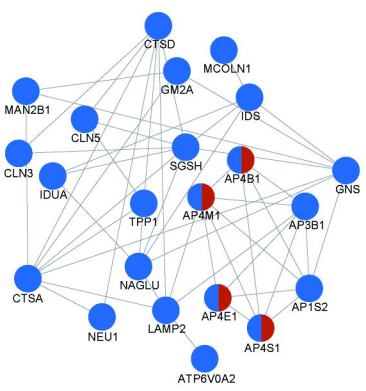




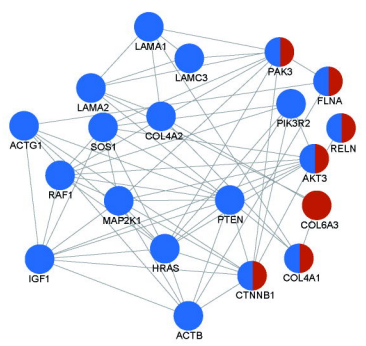




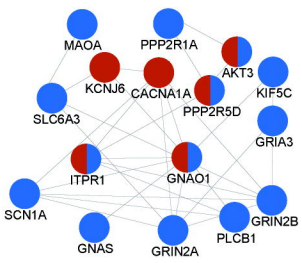
C



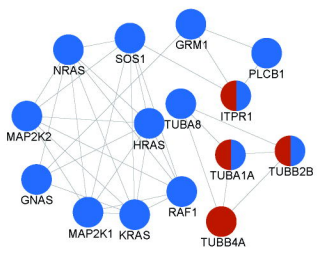
Lysosome



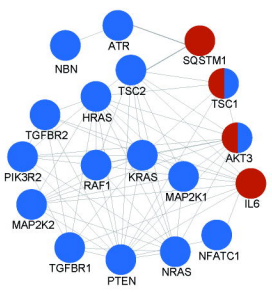
Focal adhesion



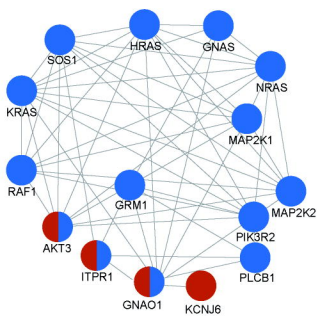
Dopaminergic synapse



Gap junction

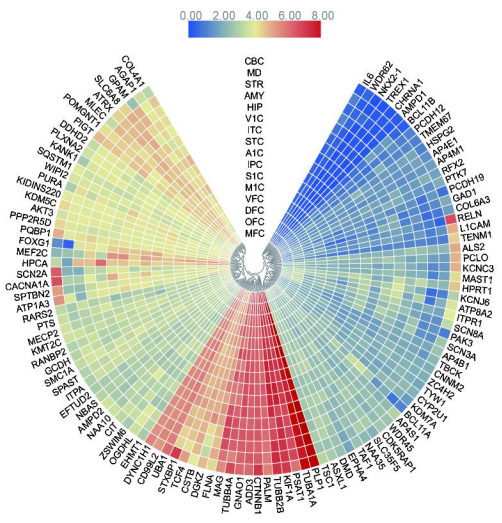


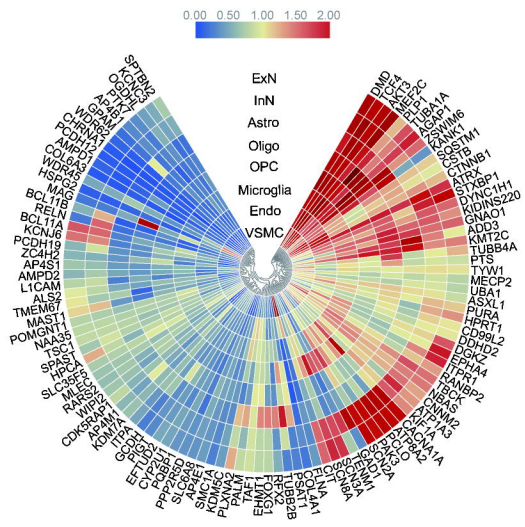
Cellular senescence

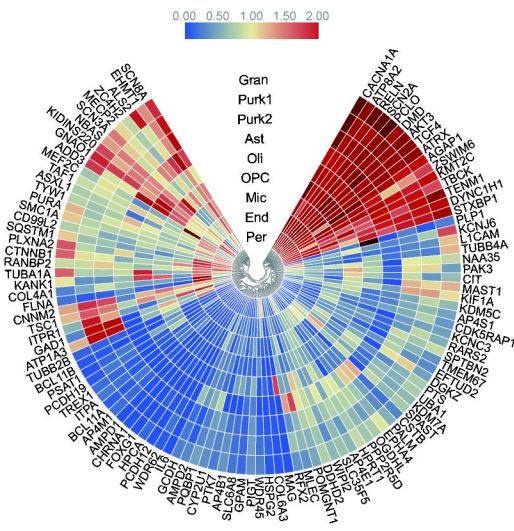


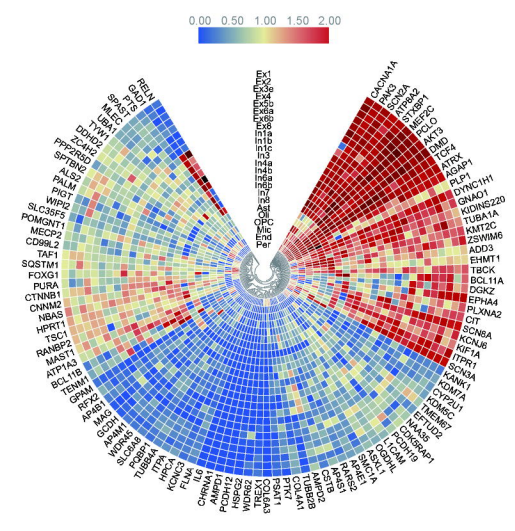
Estrogen signaling pathway

D

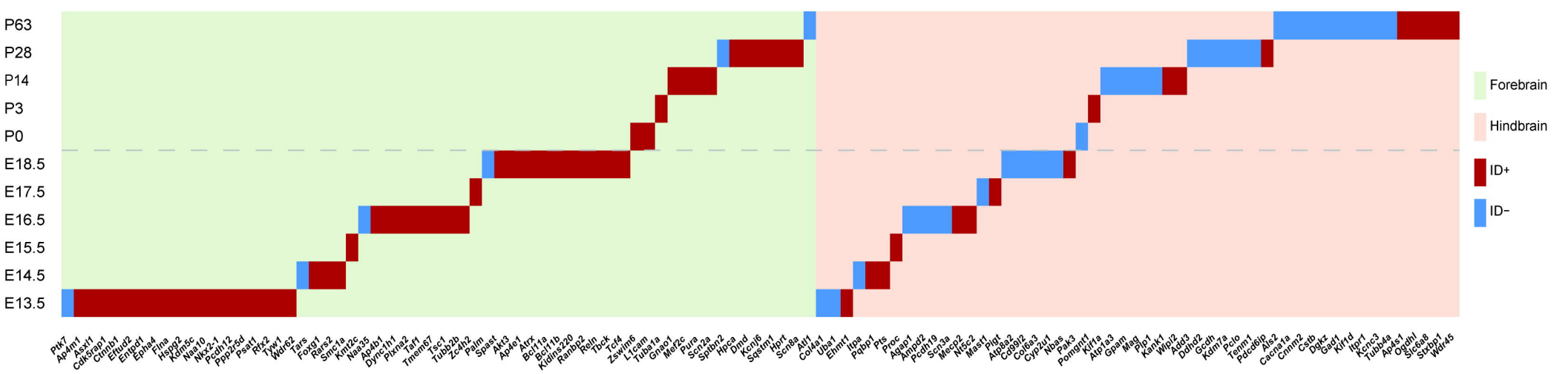




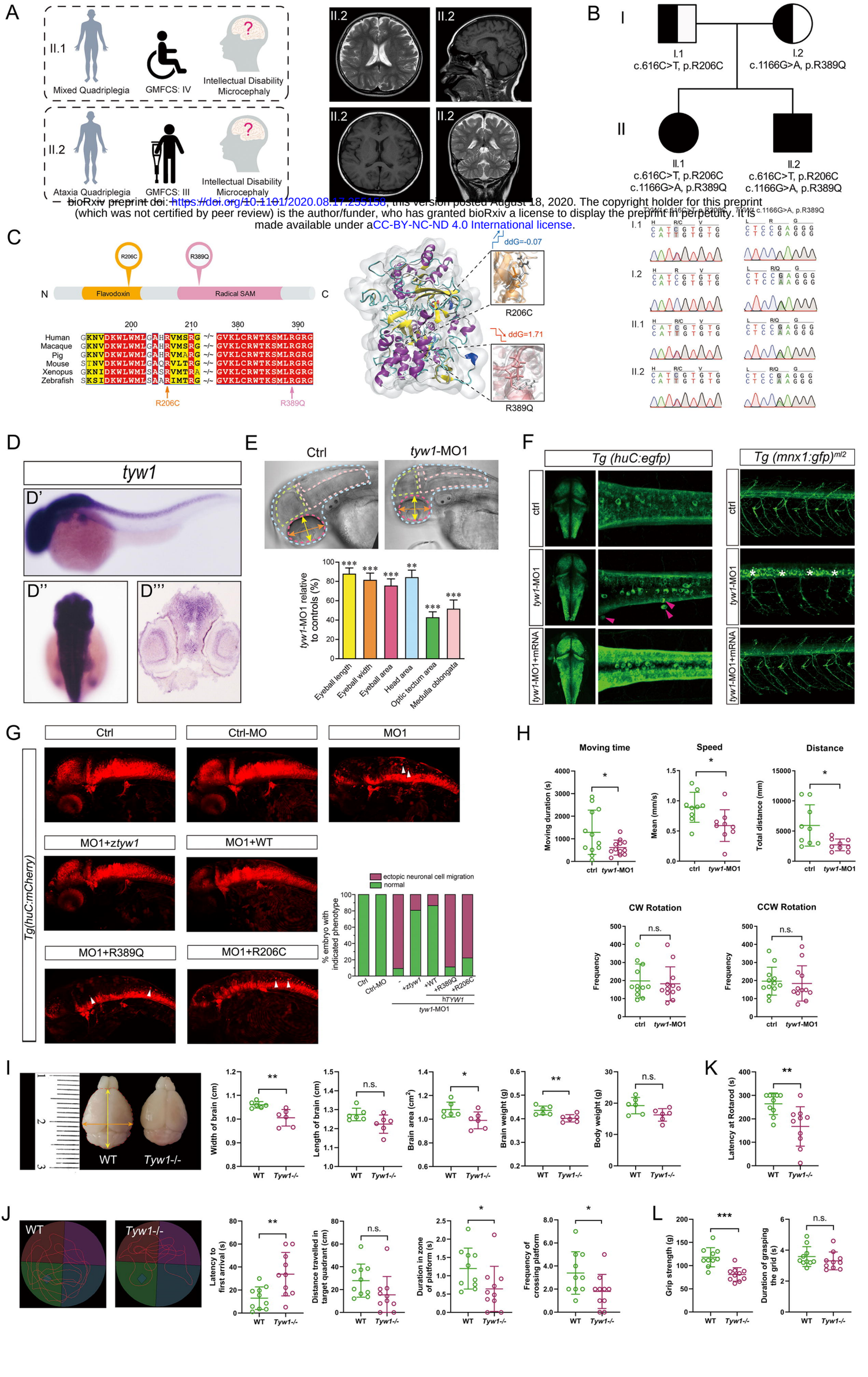




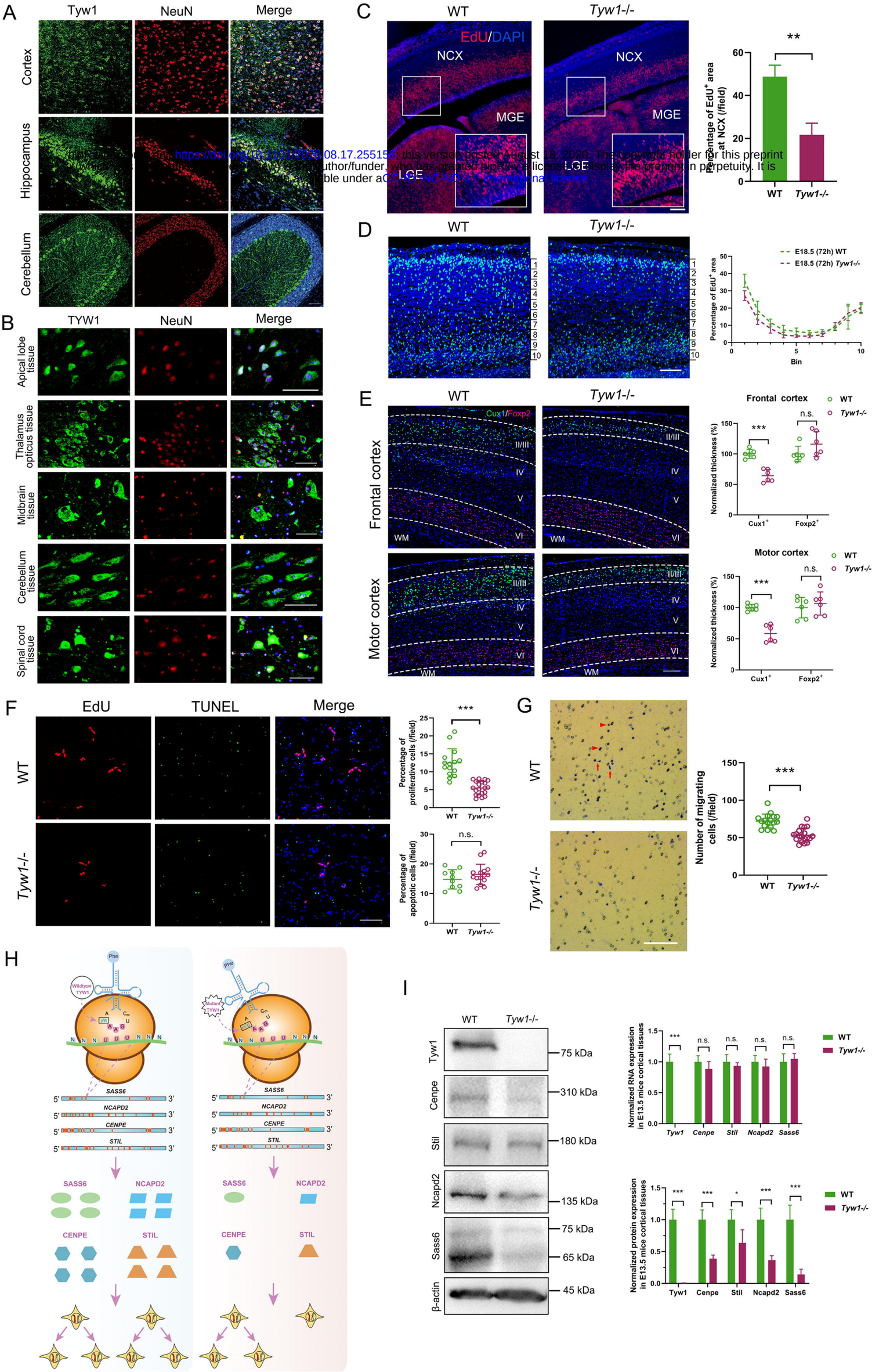
E



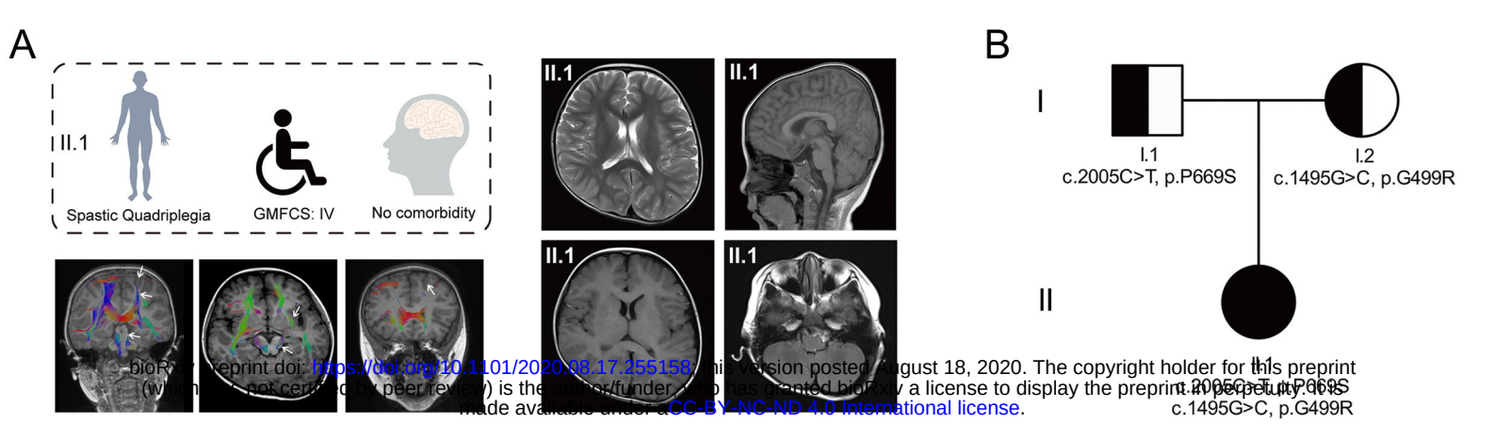
bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158, this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

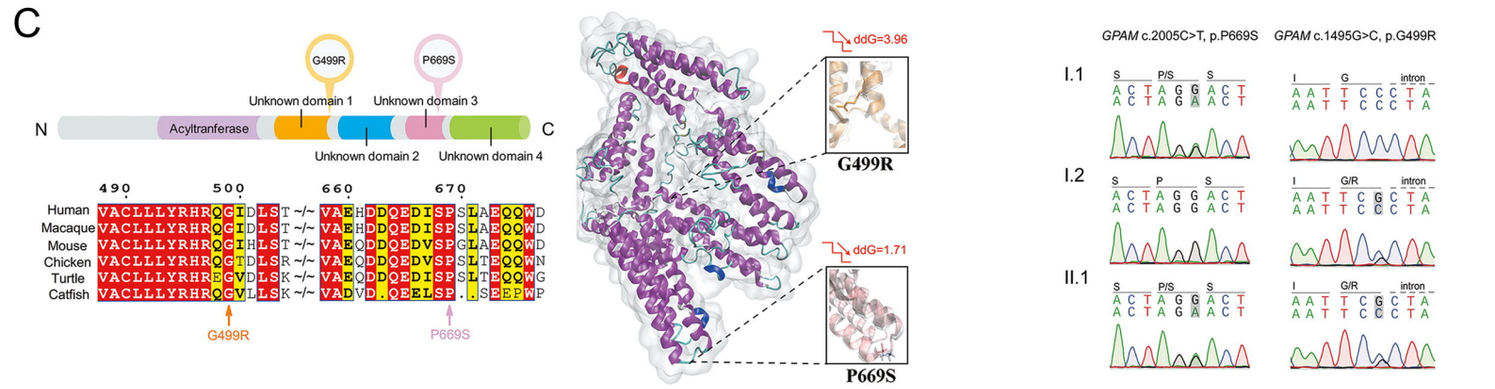


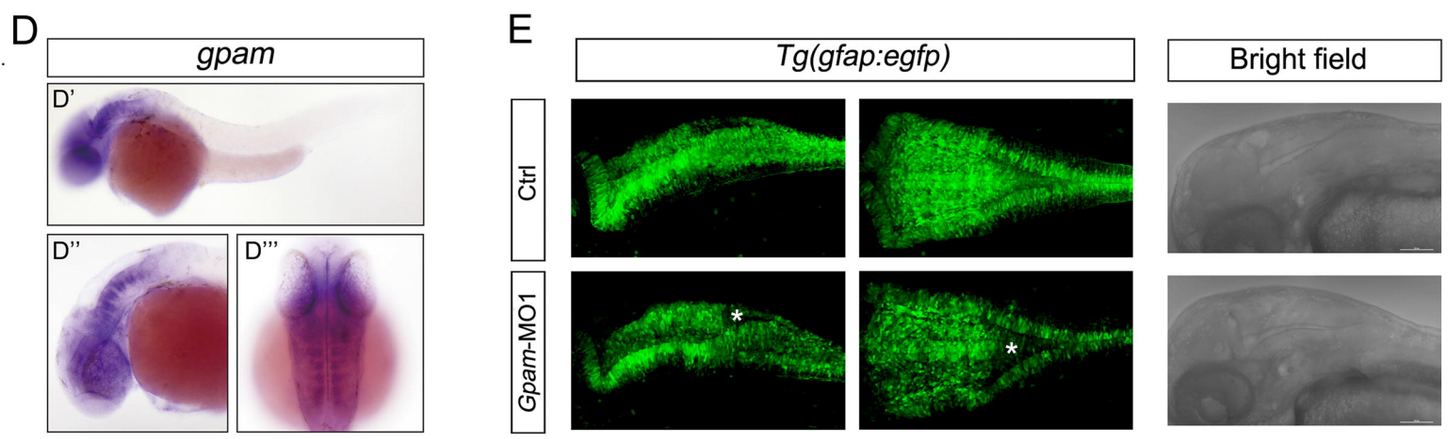
https://doi.org/10.1101/2020.08.17.255153 this version posted August 18, 2020. The copyright holder for this preprint
author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is
CC BY-NC-ND 4.0 International license.

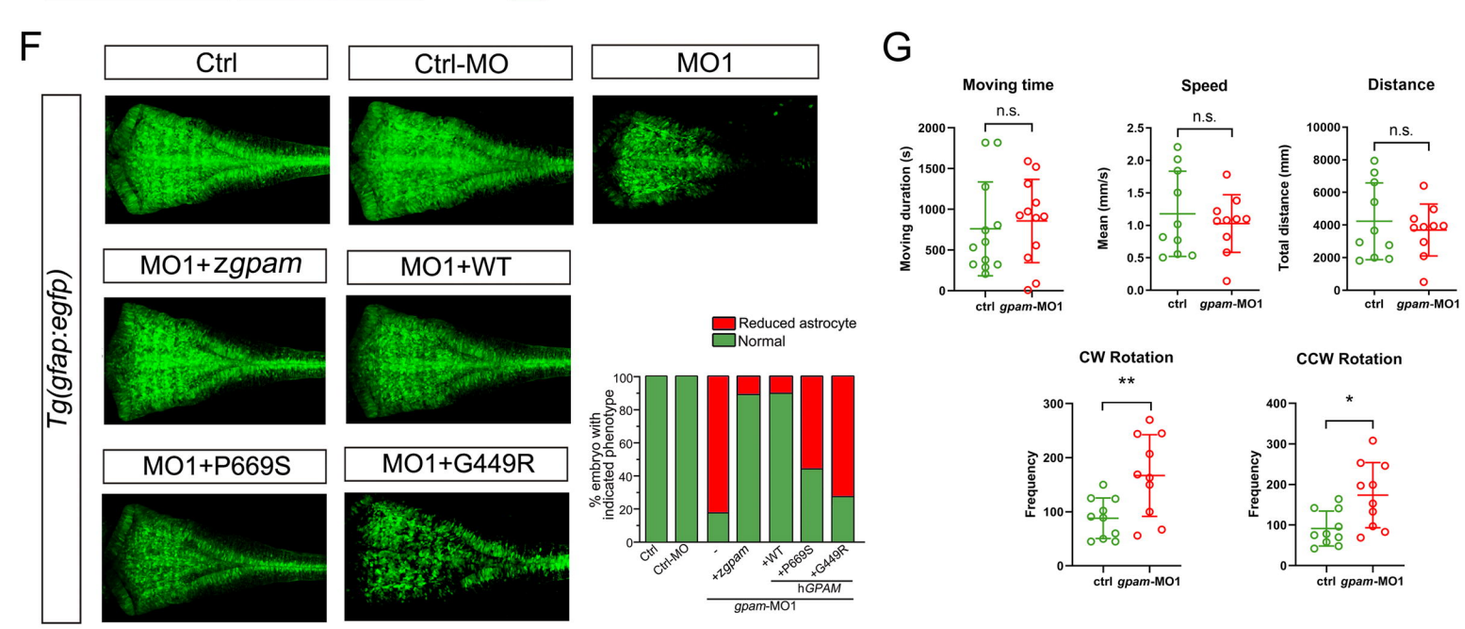


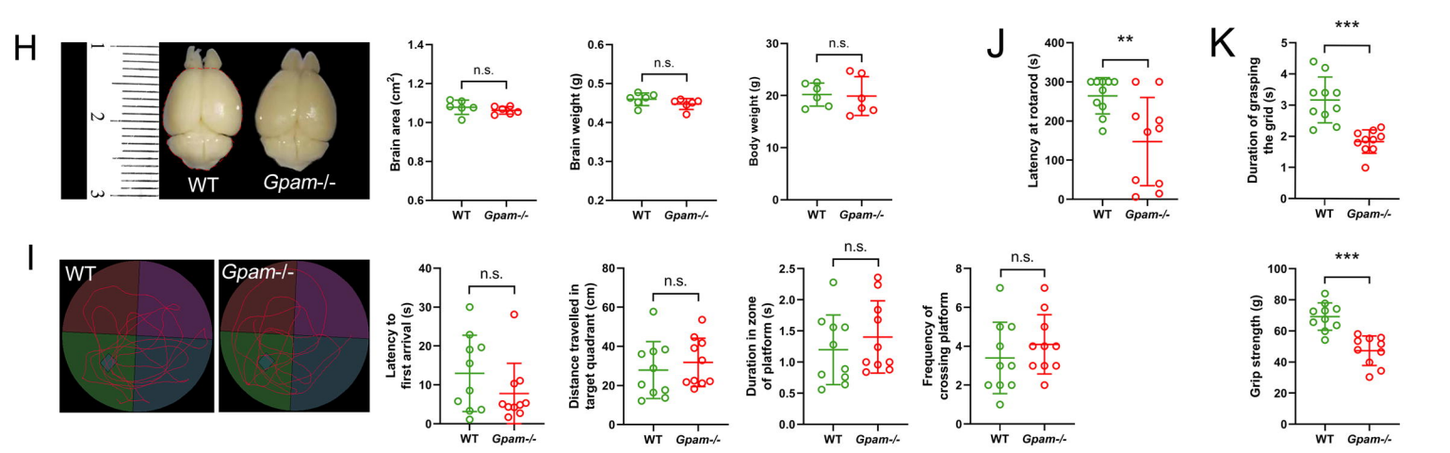
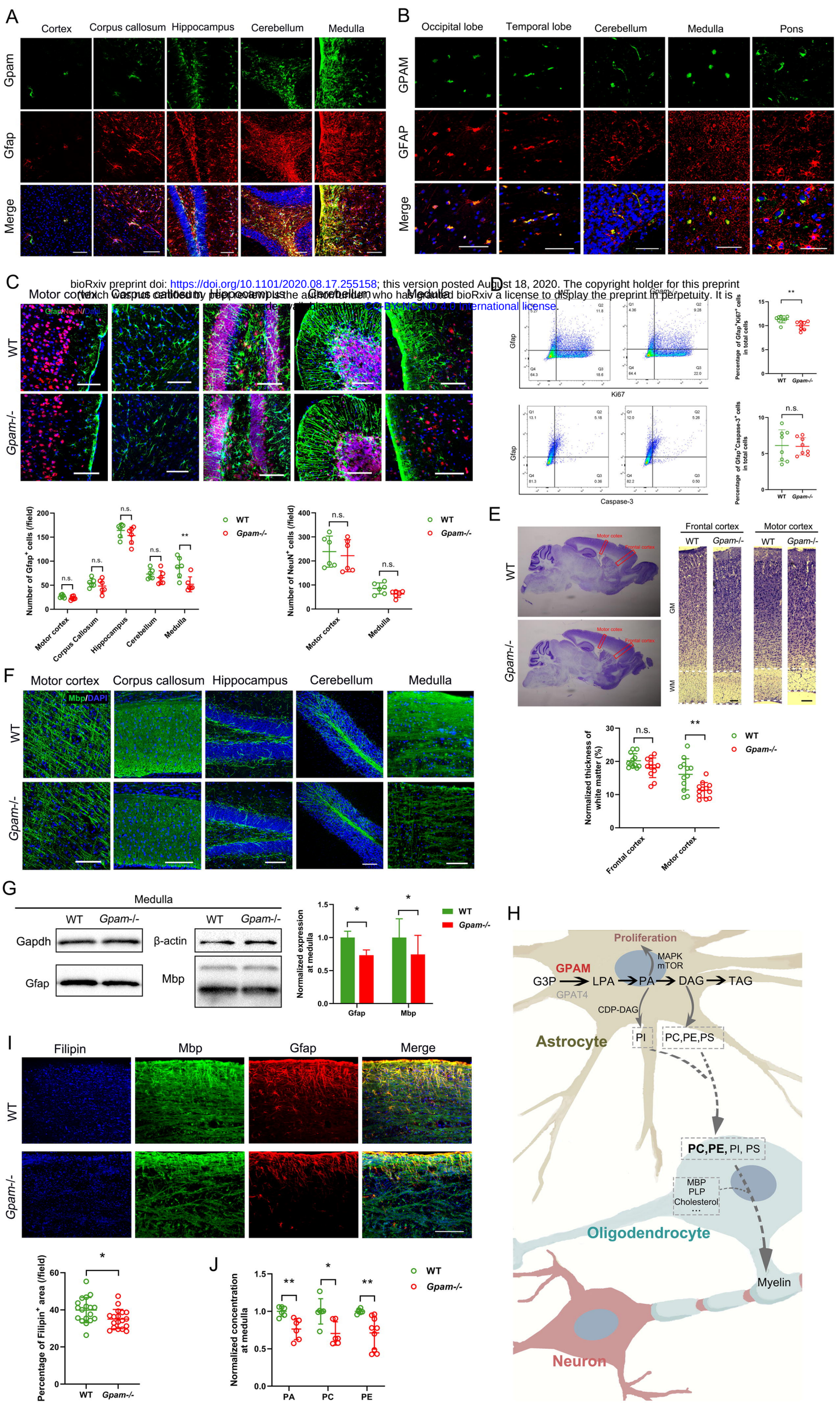
bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.











bioRxiv preprint doi: https://doi.org/10.1101/2020.08.17.255158; this version posted August 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.