Supplementary Materials

Materials and Methods

Screening of Genetic/Genomic Variants in Patients with ASD or ASD/ID/DD

Previously, we collaboratively screened 892 patients at Boston Children's Hospital (BCH; 437 samples of the western population) and the Children's Hospital of Fudan University (CHFU; 455 samples of Han Chinese) with autism spectrum disorder (ASD), or ASD with ID (intellectual disability) or DD (developmental delay) (a.k.a., ASD/ID/DD) for possible associated genetic/genomic variants. The patient's referring clinicians, including the developmental behavioral pediatrician, neurologist, pediatric psychologist, or psychiatrist, made the initial clinical diagnosis of ASD, or ASD with ID or DD, according to the criteria of the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR). All samples were de-identified prior to collect for research under Institutional Review Board protocols for non-identifiable biological specimens, which were deemed "leftover" from patient samples after clinical tests. Of them, we selected samples with a diagnosis of ASD or ASD/ID/DD based on clinical test requisition and medical record review to confirm that DSM-IV-TR criteria were used. Each institute approved this study. For more information of the screening criterion and procedure, please see Shen *et al*. 2010 [1] and Dang *et al*. 2018 [2].

All de-identified DNA samples were obtained with the informed consent of patients and their families according to the requirements of the Internal Review Board (BCH) and Ethics Committee (CHFU). If a child participant was identified with a genetic abnormality, such as copy number variant (CNV) or single nucleotide variants (SNV), then parental testing was recommended, the

parents' samples with the written consent were also provided for participation. A total of 576 de-identified control DNA samples were obtained from the healthy Chinese individuals undergoing physical examination (CHFU), and 8,045 anonymous non-neurological/psychiatric patients listed in the BCH-GDL database were used as additional controls. We conducted a burden test using the Fisher test in ASDs and healthy controls, the *P*-value is 6.24×10^{-8} for the combined cohort (patients: 892, and controls: 8,621).

Genomic DNA was isolated from EDTA peripheral blood samples of the patients; by using the QIAamp DNA blood Kit (Qiagen, Dusseldorf, Germany), according to the manufacturer's protocol. The promoter, UTR, and coding region of *PLCL1* were amplified by PCR, and sequenced with the ABI Prism Big Dye System and run on an ABI 3730 automated sequencer (ABI, Carlsbad, USA). The primer sequences are available upon request. For array comparative genomic hybridization (array CGH) analysis, DNA was fragmented, labeled, and hybridized to the CHB Array CGH (version 1.0) according to the manufacturer's protocol.

The *PLCL1* gene spans 343.7 kb (GenBank NC_000002.11, ID: 5334) on chromosome 2q33.1, and the *PLCL1* transcript comprises 6 exons (5.1 kb mRNA sequence, GenBank NM_006226). Localization of detected variants in protein domains was assessed by Uniprot (http://www.uniprot.org/). The potential detrimental effects of missense variants on protein function were predicted using three software programs: PolyPhen-2 (Polymorphism Phenotyping V2, http://genetics.bwh.harvard.edu/pph2/), SIFT (Sorting Intolerant From Tolerant Program, http://sift.bii.a-star.edu.sg/), and Mutation Taster (http://www.mutationtaster.org/). For *PLCL1* promoter prediction, we used two online promoter prediction software (Fig. S4A): Promoter 2.0

(http://www.cbs.dtu.dk/services/promoter/), and Fruitfly (http://www.fruitfly.org/seq_tools/promoter.html).

Clinical Information on Patients Harboring *PLCL1* Variants

A 4-year-old boy with a 542 kb microdeletion was diagnosed with ASD and ID (also called mental retardation) with motor deficits, and had significant impairments in verbal language (primarily single words), paucity of nonverbal communication compensation, and aggressive behaviors. The p.(Met974Ile) (c.2922G>T) was detected in a 10-year-old girl born to healthy parents. The same p.(Met974Ile) substitution was found in the father and was absent in the mother. The affected girl had an indifferent personality and was expressionless at the age of three, with no other complications. Her family had no psychiatric history. The patient who carried the *PLCL1* 5'-UTR variant c.−136_−134 delGCC, was a 12-year-old boy who has been sensitive, irritable, and quirky since the age of 10, accompanied by mental retardation. The variant was inherited from his unaffected mother. The variant c.−636 C>A was detected in a boy, and was absent in his healthy mother. The information of his father was not available because of a divorced family.

Plasmids

For neuronal function study, the full-length open reading frame of human *PLCL1* was synthesized by GenScript (Nanjing, China). Then we cloned HA-PLCL1 into the vector pCAGGS-IRES-GFP. The predicted *PLCL1* gene promoter region (from c.−901 bp to c.−1 bp) was cloned into the luciferase reporter gene pGL3-basic vector to construct a wild-type plasmid pGL3-WT. The mutants were generated using KOD-Plus according to the manufacturer's instructions (no. 132000; Toyobo, Osaka,

Japan). Four transcription factors E2F4 (NM_001950.3), E2F6 (NM_198256.3), SP1 (NM_138473.2), and GATA1 (NM_002049) were cloned into the mammalian expression vector pCMV plasmid and carried the HA or Myc tag.

The mouse *Plcl1* short hairpin RNAs and were introduced into the pFUGW-H1 vector and were directed against the following sequences: 5'-GCGGATAGATTCCAGTAAT-3' (shRNA1) and 5'-GCGCAAATACAAAGGGCAT-3' (shRNA2); the NC sequence was used as a control: 5'-TTCTCCGAACGTGTCACGT-3'. *Plcl1* shRNA lentivirus was provided by OBIO (Shanghai, China). The human GATA1 siRNA1 (small interfering RNA 1) and siRNA2 sequences were: 5'-GCCUCUAUCACAAGAUGAATT-3' and 5'-CCAUGCGGAAGGAUGGUAUTT-3'; NC: 5'-UUCUCCGAACGUGUCACGUTT-3'.

Antibodies

Antibodies used in this study were as follows: anti-PLCL1 (no. ab190225; Abcam, Cambridge, UK), α-tubulin (no. ab7291; Abcam), anti-GFP (no. G10362; Invitrogen, Carlsbad, USA), SMI 312 (no. SMI-312R; Covance, Princeton, USA), anti-VGAT (no. AB2257; Millipore, Boston, USA), anti-VGLUT1 (no. AB5905; Millipore), anti-Myc (no. M20002; Abmart, Berkeley Heights, USA), anti-HA (no. M20003; Abmart).

Cell Culture

For neurodevelopment study, embryonic day 15.5 (E15.5) mouse (C57BL/6 J) cortical neurons were transfected *via* electroporation with an Amaxa Nucleofector (Amaxa, Cologne, Germany) at 0 days

in vitro (DIV) before plating on poly-L-ornithine- and fibronectin-coated plates. Cells were harvested at DIV3 for further axon outgrowth analysis.

For VGLUT1 and VGAT rescue experiments, we co-transfected each group with two plasmids, using a 1:3 ratio of shRNAs and wild-type (WT) or mutant variants of PLCL1, *via* calcium phosphate transfection at DIV5 according to Tong *et al*., 2019 [3]. Cells were collected for immunofluorescence analysis at DIV14.

Immunofluorescence

Cells were washed with 1 × phosphate-buffered saline (PBS) for 5 min, then fixed in 4% paraformaldehyde (PFA) for 20 min and blocked in 1 × PBS buffer with 3% BSA and 0.1% Triton-X-100 for 60 min at room temperature (RT). Cells were incubated in primary antibodies overnight at 4℃, washed three times in 1 × PBS, then incubated in the secondary antibodies at RT for 2 h and the signals were observed *via* fluorescence microscopy. DAPI (1:200) was used to visualize cell nuclei.

Western Blot

To examine PLCL1-specific expression in the mouse cortex, cortical tissue homogenates were prepared from the mouse brains at different developmental stages, immersed in Radio Immunoprecipitation Assay (RIPA) lysis buffer, boiled in 1 × sodium dodecyl sulfate loading buffer, and then resolved by 8% sodium dodecyl sulfate polyacrylamide gel electrophoresis. Proteins on the gel were transferred onto polyvinylidene difluoride membranes (Millipore), and the membranes were blocked with 5% non-fat milk in Tris-buffered saline and Tween 20 (TBST) buffer for 1 h and then

incubated in primary antibodies overnight at 4°C. The membranes were washed three times in TBST and incubated in secondary antibodies for 1 h at RT. The signals were revealed by the horseradish peroxidase reaction using SuperSignal Chemiluminescent Substrate (Thermo Fisher Scientific, Waltham, USA).

Luciferase Reporter Assays

For luciferase assay, the reporter constructs were co-transfected into HEK-293T cells with pCMV-E2F4 (E2F6, SP1, or GATA1), or the control pCMV-Myc vector alone using Lipofectin 3000 (Thermo Fisher Scientific). Thirty-six hours after transfection, cell lysates were prepared and luciferase activity was measured by using a Dual Reporter Assay System (Promega, Wisconsin, USA) according to the manufacturer's instructions. The pRL-TK plasmid (Promega) encoding *Renilla* luciferase was used for internal normalization in each transfection.

Data Analysis

Images were obtained using a Nikon E80i fluorescence microscope (Nikon, Tokyo, Japan) equipped with a ×20 objective (for neurite development, axonal polarity formation), a Leica P5 confocal microscope equipped with a ×20 objective (for neurite outgrowth) and a ×60 oil immersion lens (for spines, VGLUT1, and VGAT density). GFP-positive cells were selected randomly from each condition and all the images were analyzed with ImageJ software (http://imagej.net/Fiji/Downloads). Unpaired *t*-test or one-way ANOVA followed by Tukey's multiple comparison tests were used for 2 or more groups. Two-way ANOVA was used for luciferase analysis. The data are presented as the mean ± SEM. The statistical significance was set at a $P < 0.05$.

Supplementary Figures

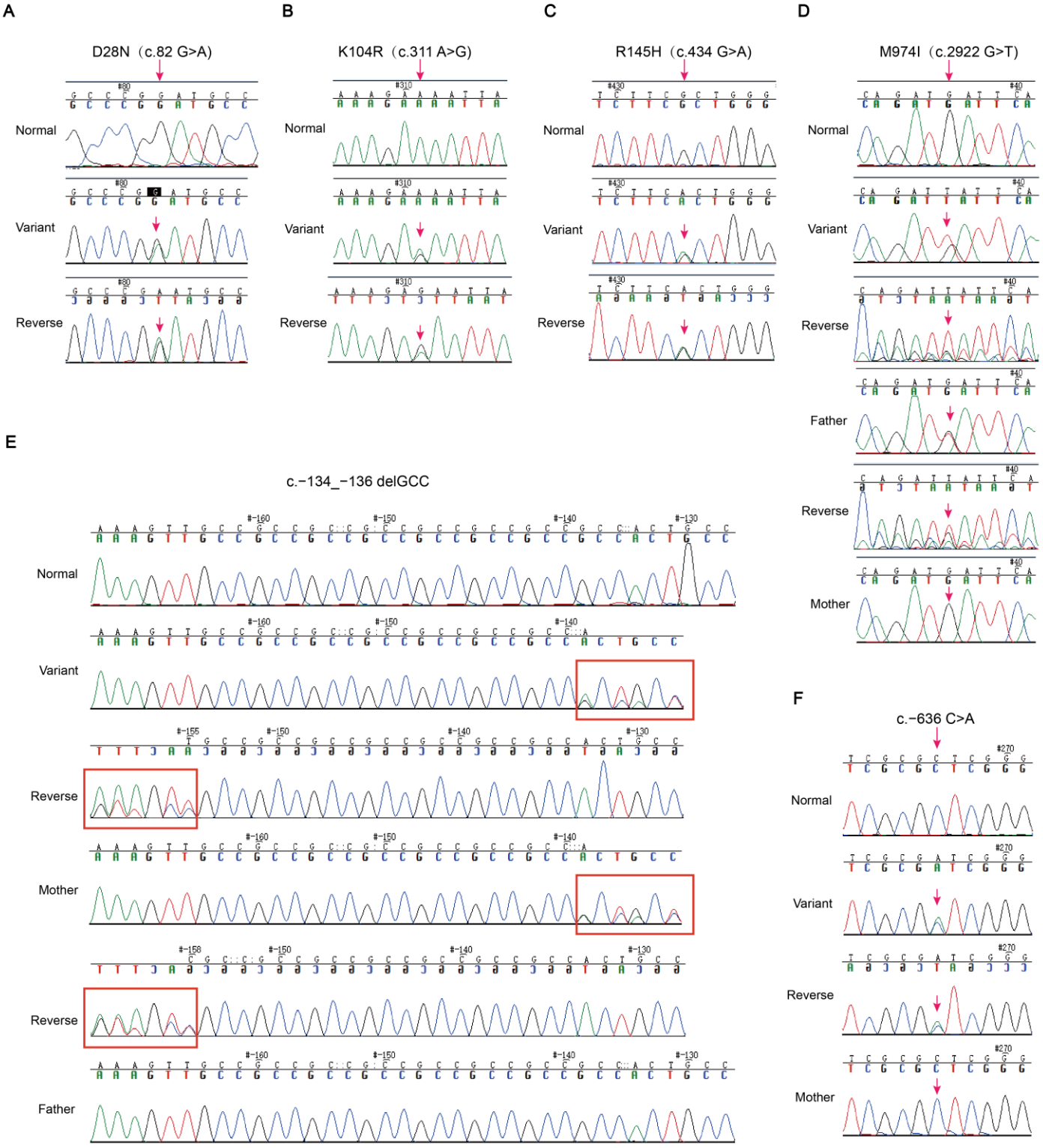

Fig. S1 Forward and reverse sequencing of the *PLCL1* variants. **A–D** Sequencing of the variants in the *PLCL1* coding region and parental sample validation with Sequencher 4.9 software. The red arrow indicates the variant site. Parental samples of p.(Asp28Asn), p.(Lys104Arg), and p.(Arg145His) were not obtained. **E, F** Sequencing of the variants in the *PLCL1* non-coding region and the verification of parental samples with Sequencher 4.9 software. The red arrow indicates the variant site and the red box indicates the variant region.

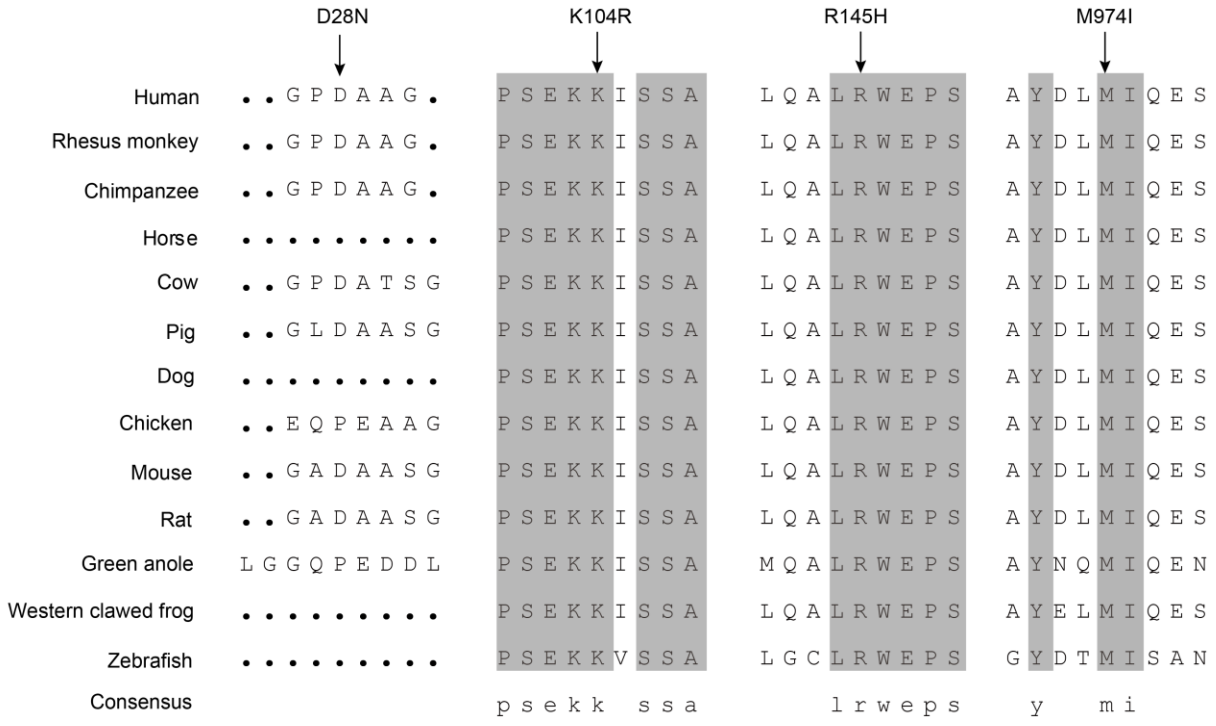

Fig. S2 Alignment of PLCL1 protein sequences using DNAMAN-multiple sequence alignment. The following sequences were used: human, NP_006217.3; chimpanzee, XP_001169560.1; rhesus monkey, XP_001089483.1; pig, XP_003133613.2; horse, XP_001500322.3; chicken, XP_421916.3; green anole, XP_003227045.1; dog, XP_536020.3; cow, NP_001178208.1; mouse, NP_001108135.1; rat, NP_445908.1; western clawed frog, XP002937769.1; zebrafish,

XP_001921798.3. Conserved amino-acid residues are shaded in grey; amino-acid residues altered in the human sequence are arrowed.

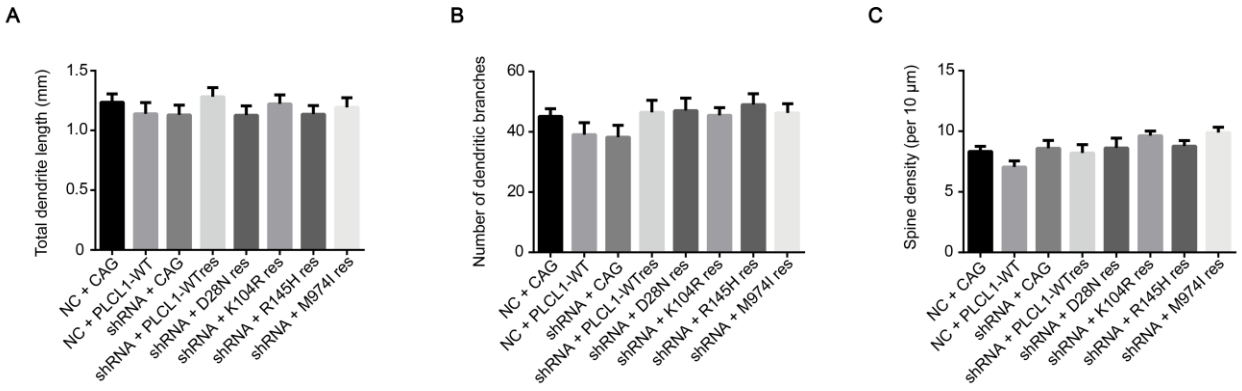

Fig. S3 PLCL1 dosage imbalance does not affect the total dendrite length (**A**), branching number (**B**) and spine density (**C**). Neurons expressing the vectors (pFUGW and CAG), shRNA, PLCL1-WT, or mutant forms of *PLCL1* along with the shRNA. All neurons were co-labeled with VGLUT1, VGAT, and GFP. Approximately 30 cells from three independent experiments were randomly selected and counted. Statistical significance was determined by one-way ANOVA analysis. Error bars, ± SEM.

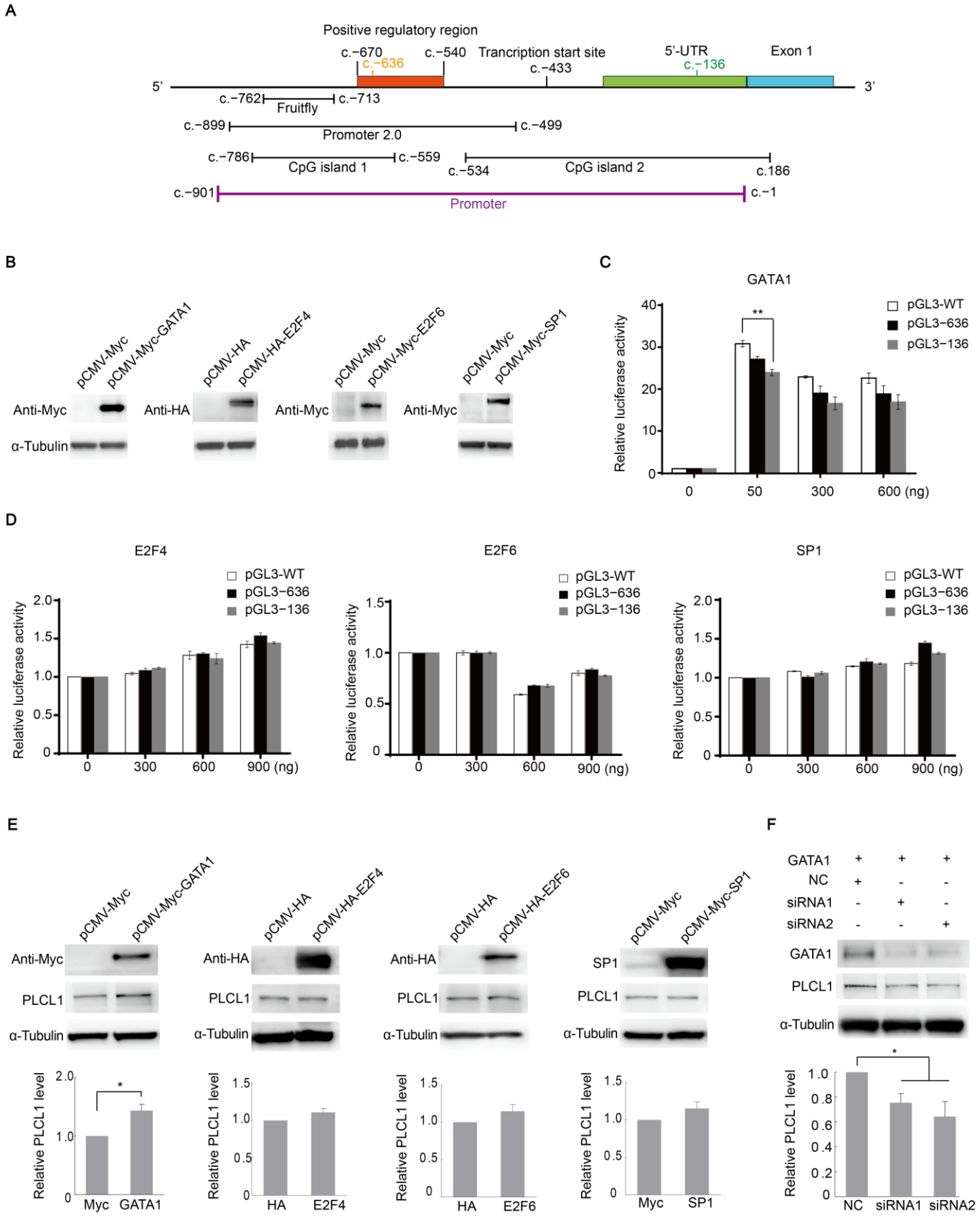

Fig. S4 Rare variant c.−136_−134 delGCC at the non-coding sequence affects the transcriptional activation of *PLCL1* by GATA1, which regulates PLCL1 protein expression. **A** *PLCL1* promoter

prediction. The predicted regions with software are shown with solid lines, as well as predicted the cytosine-phosphate-guanine (CpG) islands. **B** Western blot analysis of ectopic expression of transcription factors GATA1, E2F4, E2F6, and SP1 in HEK-293T cells. **C** The effect of GATA1 on wild-type and mutant *PLCL1* promoter activity was measured by luciferase reporter assays. HEK-293T cells were transfected with pGL3-WT *PLCL1* promoter (pGL3-WT) or mutant types (pGL3−136 and pGL3−636) plus various amounts pCMV-Myc-GATA1 vector. The relative luciferase activities were obtained against the promoter activities of transfecting pGL3-WT plus mock pCMV-Myc vectors. *P*-value was obtained with the two-way ANOVA statistical method. **D** The effect of E2F4, E2F6, and SP1 on wild-type and mutant *PLCL1* promoter activity by luciferase reporter assays. **E** Western blot examined the effect of ectopic expression of GATA1 E2F4, E2F6, and SP1 on PLCL1 protein expression in HEK-293T cells. Three independent experiments were performed. The intensity of each band was quantified using ImageJ software and was compared by Student's *t*-test, $*P < 0.05$. **F** Knockdown of ectopic GATA1 expression reduced GATA1-induced PLCL1 protein expression in HEK-293T cells. GATA1 siRNA1 and siRNA2 were co-transfected into HEK-293T cells with pCMV-Myc-GATA1. Error bars, ± SEM.

Supplementary Tables

Table S1 The microdeletions in the DECIPHER database that overlapped with the 542 kb deletion in this study[#]

Patients	Gender	Genome position	Size	Coverage of *PLCL1*	Phenotype	MAF in ASD	Inheritance
ASD-09-034 (this study)	Male	Chr2: 198,030,536–198,573,140	542 kb	*PLCL*1 exons 2–6	ASD with ID, motor deficits, impaired verbal language, paucity of nonverbal communication compensation, aggressive behaviors.	0.0023 (1/437)	*de novo*
Decipher250327	Male	Chr2: 191,678,618–198,619,858	6.94 Mb	Including whole *PLCL1* gene	Intellectual disability.	NA	*de novo*
Decipher258554	Male	Chr2: 197,714,432–201,492,307	3.78 Mb	Including whole *PLCL1* gene	Delayed speech and language development, intellectual disability.	NA	*de novo*
Decipher259644	Female	Chr2: 196,572,755–199,111,073	2.54 Mb	Including whole *PLCL1* gene	Abnormality of the face, intellectual disability.	NA	*de novo*
Decipher263623	Male	Chr2: 196,169,516–199,585,862	3.42 Mb	Including whole *PLCL1* gene	Delayed speech and language development, Intellectual disability.	NA	Unknown
Decipher350717	Female	Chr2: 194,934,603–201,002,109	6.07 Mb	Including whole *PLCL1* gene	Delayed speech and language development, Misalignment of teeth, self-mutilation.	NA	Unknown
Decipher248613	Female	Chr2: 196,640,039–201,015,790	4.38 Mb	Including whole *PLCL1* gene	Abnormality of the dentition, Cleft palate, Feeding difficulties in infancy, Intellectual disability, Joint laxity, Seizures, Short stature.	NA	*de novo*

Abbreviations: MAF, minor allele frequency; NA, not available.

[#]The 542 kb deletion compared with 6 similar deletions listed in a publicly available DECIPHER database (website: decipher.Sanger.ac.uk).

Table S2 The genomic and bioinformatic analysis of *PLCL1* missense and non-coding region variants in the studied patients

Variant type	Nucleotide change	Genome position	Amino acid change	Location	Data base search	Domain	Function prediction by bioinformatic tools	MAF in ASD	MAF in controls	Inheritance*
Missense	c.82 G>A	Chr2: 197,805,181	p.(D28N)	Exon 1	0	-	SIFT: tolerated Polyphen: benign MutationTaster: polymorphism	0.0022 (1/455)	0	NA
Missense	c.311 A>G	Chr2: 198,083,828	p.(K104R)	Exon 2	0	Region: Interaction with PPP1C	SIFT: deleterious Polyphen: probably damaging MutationTaster: disease causing	0.0022 (1/455)	0	NA
Missense	c.434 G>A	Chr2: 198,083,951	p.(R145H)	Exon 2	dbSNP ID: rs764761564 0.00011 (13/121122)	PH Region: Interaction with PPP1C	SIFT: Deleterious Polyphen: probably damaging MutationTaster: disease causing Molecular modeling: benign	0.0022 (1/455)	0	NA
Missense	c.2922 G>T	Chr2: 198,101,287	p.(M974I)	Exon 4	0	-	SIFT: tolerated Polyphen: benign MutationTaster: disease causing	0.0022 (1/455)	0	Father
SNV	c.−636 C>A	Chr2: 197,804,464	-	Predicted promoter	0	Positive regulatory region CpG island	Transcription Factor Binding Prediction: TF Search: C: E2F; A: GATA1, GATA2, GATA3 TESS:C: E2F; A: none Alibaba 2.1: C: SP1;A: none	0.0022 (1/455)	0	Not mother
STR	c.−136_−134 delGCC	Chr2: 197,804,964–197,804,966	-	5’-UTR predicted promoter	0	CpG island	Transcription Factor Binding Prediction: NA	0.0022 (1/455)	0	Mother

*The inheritance pattern of some variants in this table could not be determined because no parental samples were available.

Abbreviations: MAF, minor allele frequency; NA, not available; STR, short tandem repeats; UTR, untranslated region.

Note: (1) The variant c.434 G>A located in the exon 2 of the *PLCL1* coding region has been defined as SNP (rs764761564) in dbSNP (build 150), but the frequency of allele A in the population is much lower (0.011%, 13/121122) than the frequency detected in 455 Chinese ASD patients (0.22%, 1/455).

Considering the potential disease-causing risk of this site, we include the variant in subsequent functional studies. (2) The variant c.−136_ c.−134 delGCC, Rs531185771 in the dbSNP database (Build 150) shows that there are four forms of polymorphism in the c.−158_−157 locus of the *PLCL1* gene: -/GCC/GCCGCC/GCCGCCGCCGCC, that is, c.−157 to c.−134 in the 5'-UTR region of normal people, there are 8/9/10/12 tandem base GCC repeats. One ASD patient harbored a heterozygous deletion of one tandem GCC in this c.−157 to c.−134 region, i.e., 7 repeats of GCC, which has not been reported in any literature or database, so we included it in this study.

References

1. Shen Y, Dies KA, Holm IA, Bridgemohan C, Sobeih MM, Caronna EB, *et al*. Clinical genetic testing for patients with autism spectrum disorders. Pediatrics 2010, 125: e727–e735.
2. Dang T, Duan WY, Yu B, Tong DL, Cheng C, Zhang YF, *et al*. Autism-associated Dyrk1a truncation mutants impair neuronal dendritic and spine growth and interfere with postnatal cortical development. Mol Psychiatry 2018, 23: 747–758.
3. Tong DL, Chen RG, Lu YL, Li WK, Zhang YF, Lin JK, *et al*. The critical role of ASD-related gene *CNTNAP3* in regulating synaptic development and social behavior in mice. Neurobiol Dis 2019, 130: 104486.