The American Journal of Human Genetics, Volume *106*

Supplemental Data

De Novo Variants in *CDK19* Are Associated with a Syndrome Involving Intellectual Disability and Epileptic Encephalopathy

Hyung-lok Chung, Xiao Mao, Hua Wang, Ye-Jin Park, Paul C. Marcogliese, Jill A. Rosenfeld, Lindsay C. Burrage, Pengfei Liu, David R. Murdock, Shinya Yamamoto, Michael F. Wangler, Undiagnosed Diseases Network, Hsiao-Tuan Chao, Hongyu Long, Li Feng, Carlos A. Bacino, Hugo J. Bellen, and Bo Xiao

SUPPLEMENTAL DATA

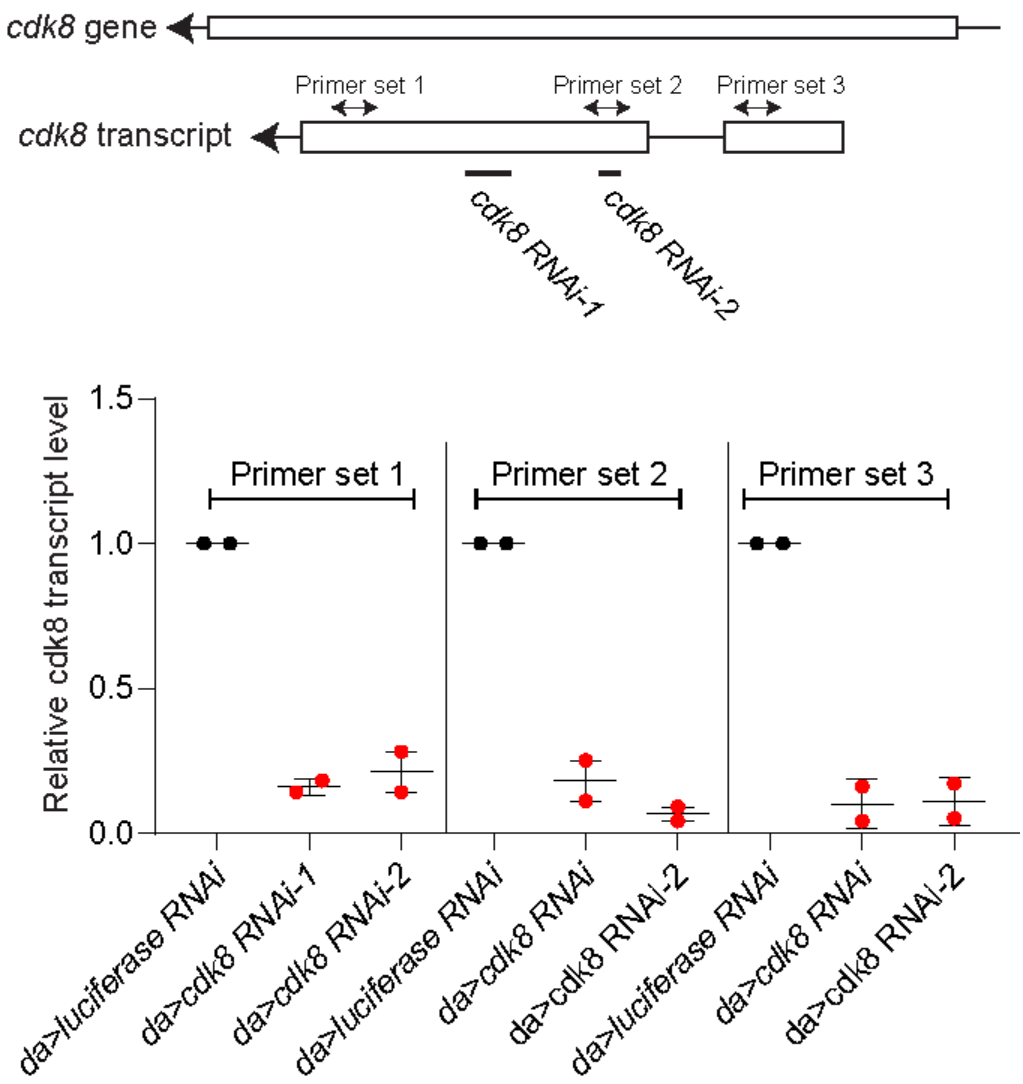

Figure S1: Two independent RNAis can efficiently reduce the transcripts level.

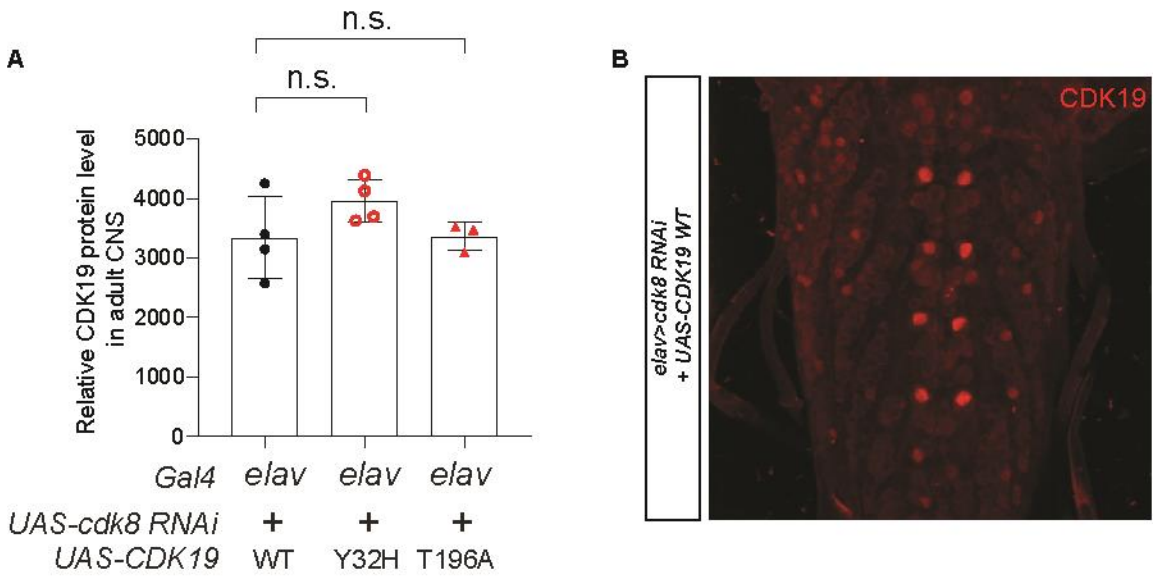

Figure S2: *CDK19* reference and variants are similarly expressed in CNS of flies.

(A) The relative intensity of CDK19 signal in adult fly CNS is comparable when we overexpressed the CDK19 reference or CDK19 variants (n = 4 for CDK19 reference, n = 4 for CDK19 p. Tyr32His and n = 3 for CDK19 p. Thr196Ala). Statistical analyses are one-way ANOVA followed by a Tukey post-hoc test. Results are mean ± s.e.m; n.s., not significant.

Supplemental Materials and Methods

Informed consent

Informed consent was obtained from the parents of proband 1 for participation in the Undiagnosed Diseases Network study at Baylor College of Medicine. This protocol has been approved under the oversight of the National Institutes of Health Institutional Review Board. Additionally, informed consent was obtained for publication of photos shown here. Patient 2 and 3 were recruited to the Xiangya hospital Epilepsy Cohort which aims to elucidate the genetic basis

of epilepsy and epilepsy related developmental disease. Patients were recruited to the study during 2018 to 2019. Both of them underwent detailed medical history collection and physical examination by neurologists. DNA was extracted from peripheral venous blood using the Genomic DNA kit from Aidlab. This study was approved by the Ethics Committee of Xiangya hospital, Central South University (No.201605577). Informed consents were obtained from both enrolled participants for publication of photos shown here.

Genetic analysis

Trio whole genome sequencing (WGS) was performed for proband 1 at Baylor Genetics (BG) as part of the Undiagnosed Diseases Network (UDN). The library is prepared using a PCR free 550-bp insert size protocol by the KAPA Hyper Prep kit. The library is subjected to sequence analysis on Illumina NovaSeq 6000 platform for 150 bp paired-end reads. The following quality control metrics of the sequencing data are generally achieved: average sequenced coverage over the genome > 40X, >97.5% target base (digital exome) covered at >20X, SNP concordance to genotype array: >95%. BG WGS data were aligned to the hg38 reference sequence on the Illumina Dragen pipeline. For compatibility with the BCM UDN annotation pipeline based on the Codified Genomics (Houston, TX) interpretation tool, the BAM files were first converted back to FASTQ files and then realigned to the hg19 reference. Variant calling was then performed on the realigned BAM files and subsequent VCFs were annotated. We filtered for rare variants that had a frequency of <0.01 in gnomAD (https://gnomad.broadinstitute.org/) and in our local database which consists of over 750 samples. For our initial analysis, we focused on *de novo* variants, and genes with biallelic variants where at least one variant was in the coding region or at a canonical splice site.

Trio ES of probands 2 and 3 was performed using the HiSeq2500 system (Illumina, San Diego, CA, USA) with a mean depth of 100X. The ES data was analyzed by using a customized pipeline,[32]

and ANNOVAR software was used to annotate the data[33]. Public databases (ESP6500, 1000genomes, ExAC, gnomAD) were used to filter variants with frequencies higher than 0.001. Variants in exonic regions and splicing regions were remained after removing variants in UTR regions, non-coding regions and intronic regions. Synonymous variants were also filtered. Sanger sequencing was used to validate trio ES results. A *de novo* variant in *EFCAB13 (c.583C>G)* and compound heterozygous variants in *ADGRE1* (*c.587C>T, c.662G>A*) were additionally identified, however, all these 3 variants were predicted to be tolerable or not conserved.

A *de novo* variant in WDR63 (c.449A>G) was identified in proband 3, however this variant was predicted to be tolerable or not conserved. We also identified compound heterozygous variants in *CEP164* (*c.548T>A, c.3965C>T*) and *PKD1* (*c.191G>C, c.4343C>T*). All the four variants were predicted to be tolerable, and the patient 3 showed no clinical features to related the diseases related to *CEP164* or *PKD1* (Nephronophthisis (*CEP164*, OMIM #614845) and polycystic kidney disease (*PKD1*, OMIM #173900)).

Drosophila **genetics**

The following stocks were obtained from the Bloomington *Drosophila* Stock Center (BDSC) at Indiana University: y^1 *w*; P {Act5C-GAL4} 25FO1/CyO, y+* (RRID: BDSC_4414[34]), *w[*];da-GAL4* (RRID : BDSC_5460), *w[1118]; P{GAL4}repo/TM3, Sb1* (RRID: BDSC_7415), *P{GAL4-elav.L}2/CyO* (RRID: BDSC_8765[35]), y^1 sc^* v^1 sev^{21}*; P{TRiP.GL00231}attP2* (RRID: BDSC_35324[22]), y^1 *sc* v*^{1} sev^{21}*;* P{TRiP.HMS05476}attP40 (RRID: BDSC_67010;[22]).

Generation of ***CDK19 Drosophila*** **transgenes**

UAS-CDK19 reference and variant transgenic flies were generated as previously described[36; 37].

Using Gateway cloning (Thermo Fisher Scientific), the *CDK19* cDNA entry clone (GenBank: NM_015076.3) in the pDONR221 vector was shuttled to the pGW-attB-HA[38]. Site-directed mutagenesis was performed with the Q5 site-directed mutagenesis kit (NEB) followed by Sanger verification. The following forward and reverse primers were used to make *CDK19* variants:

CDK19 p. Y32H: FW- ACGCGGCACCCACGGTCACGT,

RV- CCCACTTTGCACCCTTCGTACTC

CDK19 p. T196A: FW- AGTAGTTGTGGCATTTTGGTATC,

RV- GGATCCAAATCTGCTAGTG

All UAS-cDNA Constructs were inserted into the VK37 (PBac{y[+]-attP}VK00037) docking site by ϕC31 mediated transgenesis[39].

Quantitative real-time PCR

Total RNA was extracted from adult flies using RNeasy Plus Mini Kit (QIAGEN) and 1.5 µg RNA was reverse transcribed to synthesize cDNA using a 5X All-In-One RT MasterMix (abm). qPCR was performed with iTaq Universal SYBR Green Supermix (BIO-RAD) and CFX96 Touch Real-Time PCR Detection System (BIO-RAD). The level of Rp49 transcripts was used for normalization.

The following primers were used for PCR:

Rp49 FW (5′- ACAGGCCCAAGATCGTGAAGA -3′)

Rp49-RV (5′- CGCACTCTGTTGTCGATACCCT-3′)

Cdk8-1-FW (5′- CCAGCAAGATTTTCACCACCA -3′)

Cdk8-1-RV (5′- CAGTTGAAGCGCTGGAAGTTCT -3′)

Cdk8-2-FW (5′- CATCCGGGTGTTTCTGTCG -3′)

Cdk8-2-RV (5′- CAGCCCGATGGAACTTAATGAT-3′)

Cdk8-3-FW (5′- GTCTACAAGGCGAAATGGAAGG -3′)

Cdk8-3-RV (5′- CGGACATGGACAATCCGGTG -3′)

Immunohistochemistry of adult fly brains

Fly brains were dissected in cold 1X PBS, and were fixed with 4% paraformaldehyde for 30 minutes at room temperature (RT). Tissues were washed in PBST solution (1X PBS, 0.1%Triton X-100) and incubated with 5% NGS (Normal Goat Serum) at room temperature (RT) for 2 hours and incubated with primary antibodies with 5% NGS solution on a rotating platform overnight at 4°C. The brains were washed again with PBST prior to incubation with secondary antibodies on a rotating platform overnight at 4°C in the dark and thoroughly rinsed in PBST and mounted with Vectashield for imaging. Leica SP8X Confocal Microscope was used for imaging.

Primary antibodies used in this study were as follows: Mouse anti-Elav (9F8A9, 1:50, DSHB) and Rabbit anti-CDK19 (SAB4301196, 1:200, Sigma-aldrich). For secondary antibodies, goat anti-mouse-IgG with Alexa Fluor 488 and anti-rabbit Cy5 (Jackson ImmunoResearch) were used.

Dissecting and staining for larval NMJ.

NMJs were dissected and fixed from wandering third instar larvae, as described previously[40]. The tissues were incubated with 5% Normal Goat Serum solution for blocking at RT for 2 hours and then incubated with primary antibodies with 5% NGS solution on a rotating platform overnight at 4°C. The tissues were washed again with PBST before incubation with secondary antibodies on a rotating platform overnight at 4°C in the dark and then thoroughly rinsed in PBST and mounted with Vectashield for imaging. The NMJs between muscles 6/7 of the A3 segment were imaged by

Leica SP8 X Confocal Microscope. ImageJ was used for quantification of boutons, branches, and NMJ length. Synaptic branches with two or more synaptic boutons were considered branches, according to Miller et al., 2012[41].

Supplemental references

37. Ansar, M., Chung, H.L., Al-Otaibi, A., Elagabani, M.N., Ravenscroft, T.A., Paracha, S.A., Scholz, R., Abdel Magid, T., Sarwar, M.T., Shah, S.F., et al. (2019). Bi-allelic Variants in IQSEC1 Cause Intellectual Disability, Developmental Delay, and Short Stature. Am J Hum Genet. 105, 907.
38. Bischof, J., Bjorklund, M., Furger, E., Schertel, C., Taipale, J., and Basler, K. (2013). A versatile platform for creating a comprehensive UAS-ORFeome library in Drosophila. Development 140, 2434-2442.
39. Venken, K.J., He, Y., Hoskins, R.A., and Bellen, H.J. (2006). P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in D. melanogaster. Science 314, 1747-1751.
40. Smith, R., and Taylor, J.P. (2011). Dissection and imaging of active zones in the Drosophila neuromuscular junction. J Vis Exp.
41. Miller, D.L., Ballard, S.L., and Ganetzky, B. (2012). Analysis of synaptic growth and function in Drosophila with an extended larval stage. J Neurosci 32, 13776-13786.