The American Journal of Human Genetics, Volume *107*

Supplemental Data

Variants in *SCAF4* Cause a Neurodevelopmental Disorder and Are Associated with Impaired mRNA Processing

Anna Fliedner, Philipp Kirchner, Antje Wiesener, Irma van de Beek, Quinten Waisfisz, Mieke van Haelst, Daryl A. Scott, Seema R. Lalani, Jill A. Rosenfeld, Mahshid S. Azamian, Fan Xia, Marina Dutra-Clarke, Julian A. Martinez-Agosto, Hane Lee, UCLA Clinical Genomics Center, Grace J. Noh, Natalie Lippa, Anna Alkelai, Vimla Aggarwal, Katherine E. Agre, Ralitza Gavrilova, Ghayda M. Mirzaa, Rachel Straussberg, Rony Cohen, Brooke Horist, Vidya Krishnamurthy, Kirsty McWalter, Jane Juusola, Laura Davis-Keppen, Lisa Ohden, Marjon van Slegtenhorst, Stella A. de Man, Arif B. Ekici, Anne Gregor, Ingrid van de Laar, and Christiane Zweier

Supplemental Figures

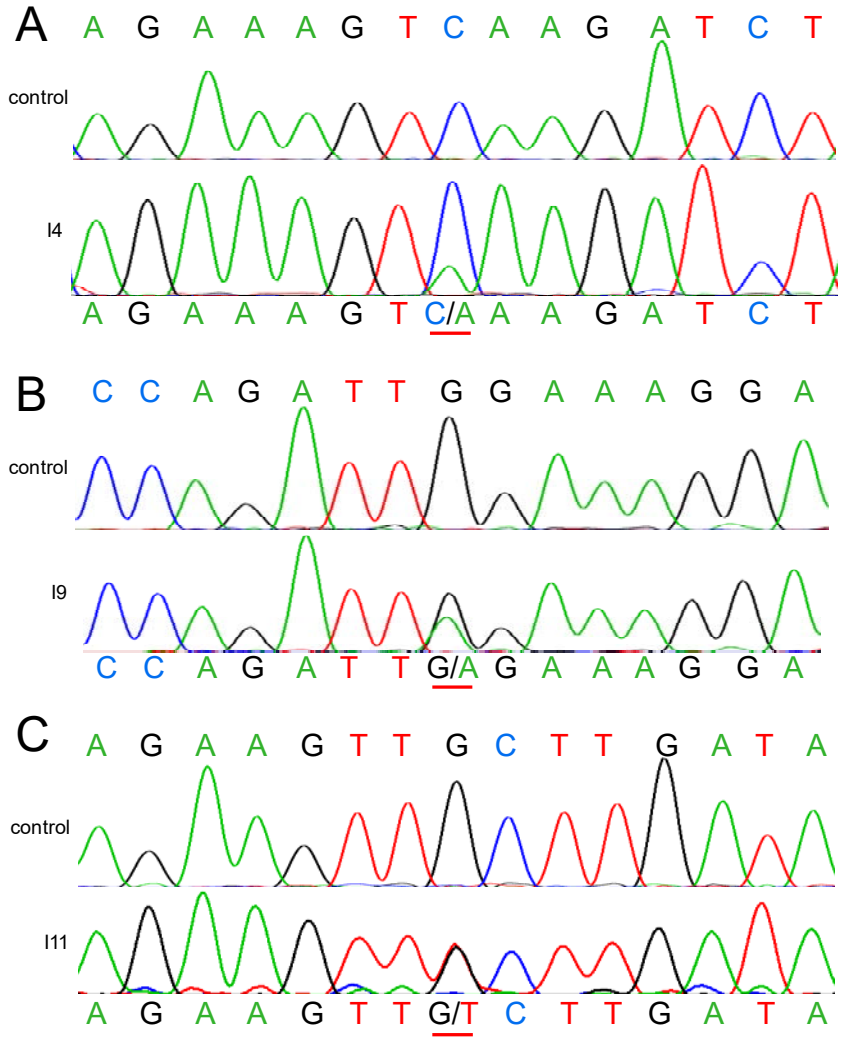


Figure S1. Sanger sequencing of cDNA from affected individuals

Sanger sequencing of cDNA from RNA of affected individuals with variants (A) c.1301C>A, (p.(Ser434*)), (B) c.1889G>A, (p.(Trp630*)), (C) c.783G>T, (p.(Leu261Phe)) after reverse-transcriptase (RT) PCR. In (A), the mutant allele (marked by red bar) is present in only a minor fraction, thus indicating nonsense mediated mRNA decay. This is in accordance with findings from RNA-sequencing in this individual, where a residual *SCAF4* level of 39.2% compared to controls was observed (Table S2). In (B) and (C) both wild type (WT) and mutant allele can be equally observed (marked by red bar). However, RNA-sequencing in both individuals revealed residual *SCAF4* levels of 29.6% for I9 and 60% for I11 (Table S2), respectively, indicating lower total *SCAF4* expression in these individuals.

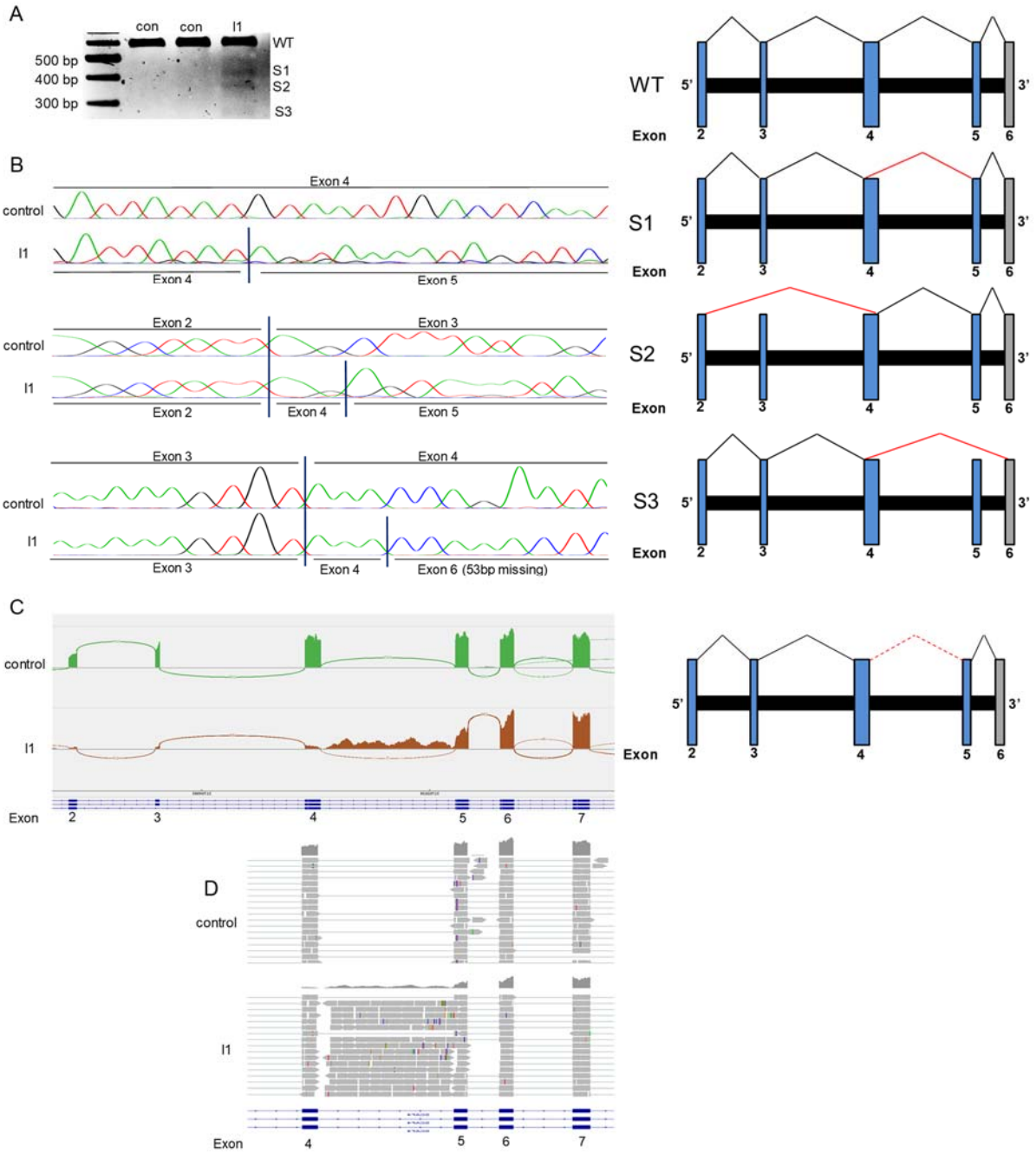
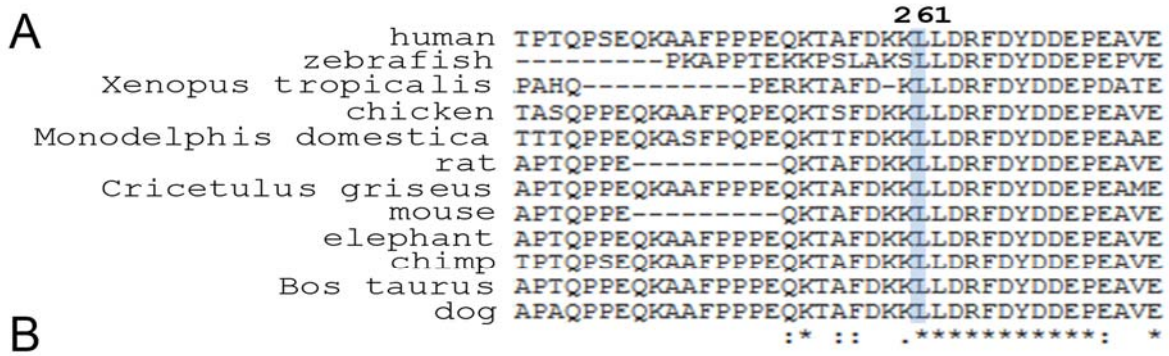


Figure S2. Analysis of the splice-site variant c.321+1G>T

(A) RT-PCR with primers spanning exons 2 to 7 of *SCAF4* revealed the occurrence of the WT isoform and three alternative splicing isoforms (S1, S2, S3) of *SCAF4* in I1 with the variant c.321+1G>T (r.spl?) in the splice donor site of intron 4. The WT isoform is most prominent, the alternatively spliced transcripts are weakly expressed. (B) Sanger sequencing of *SCAF4* transcripts after agarose gel purification. In S1, all but the first 30 bps of exon 4 are missing. In S2, exon 3 and exon 4 (except the last 3 bps) are missing. In S3, exon 4 (except the first

three bps), exon 5 and the first 53 bps of exon 6 are skipped. (C,D) Evaluation of RNA-sequencing data of *SCAF4* in I1 with the splice variant c.321+1G>T, where an increased expression of *SCAF4* (>300%) was found (Table S2). (C) Sashimi Plot, and (D) BAM files, representing aligned reads, confirm aberrant splicing of exon 5 plus intron 4 retention. In addition, exons 2, 3 and 4 are less expressed than exons 5, 6 and 7. Reduced expression of WT allele and upregulated levels of shorter, novel, non-functional transcripts might therefore explain the increased *SCAF4* expression in this individual. The Sashimi Plot was created using the IGV genome browser.[1]

A



B

Genomic position (hg19)	coding	protein	gnomAD	Mut Taster	PP2	SIFT	M-CAP
chr21:33069058C>A	c.783G>T	p.(Leu261Phe)	0	disease causing	Probably damaging score 0.999	damaging	possibly pathogenic

Figure S3. Analysis of the missense variant c.783G>T (p.(Leu261Phe))

(A) Conservation of amino acids around the missense variant c.783G>T (p.(Leu261Phe)). The position of the variant at amino acid 261 is indicated by a blue bar and highly conserved throughout all indicated species according to UCSC and depicted with clustal omega. (B) Prediction of the impact using Mutation Taster[2], PP2[3], SIFT[4] and M-CAP2.[5]

Mut Taster: Mutation Taster; PP2: Polyphen 2.

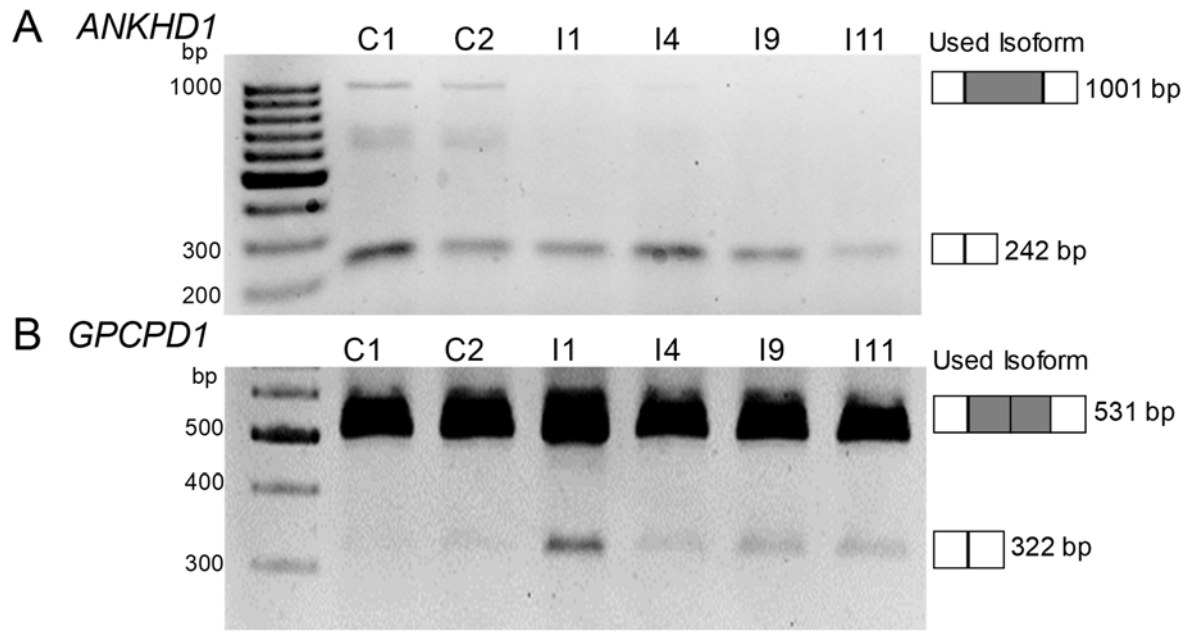


Figure S4. Validation of splicing analysis

Experimental verification of altered splicing from RNA-sequencing by RT-PCR for two examples, *ANKHD1* (MIM: 610500) and *GPCPD1* (MIM: 614124). (A) In individuals with variants in *SCAF4*, the usage of a long isoform of *ANKHD1* (ENST00000360839.7), containing exon 15, is decreased in comparison to controls. The usage of a shorter isoform (ENST00000246149.10), where exon 15 is spliced, is unchanged. (B) In individuals with variants in *SCAF4*, the usage of a short isoform of *GPCPD1* (ENST00000418646.5), with two spliced exons in comparison to longer isoforms, is increased. The usage of a longer isoform (ENST00000379019.6) is unchanged.

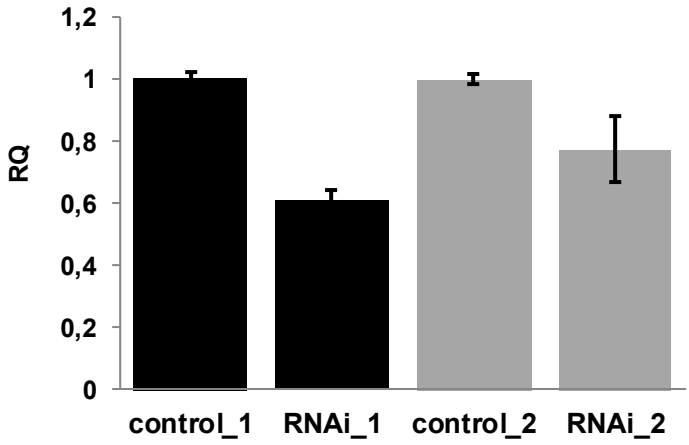


Figure S5. Confirmation of knockdown of *CG4266* in *Drosophila melanogaster*

We confirmed reduced *CG4266* expression to ca. 60% for ubiquitous knockdown (*actin*-Gal4/*Cyo*) using RNAi line 1 (BL#55354) and to 70% using RNAi line 2 (VDRC 26472) on surviving females. For male flies, lethality was observed. Quantitative real-time PCR was performed in technical quadruplicates. RQ values were calculated using the ΔΔCt method with controls as reference. Bars represent mean RQ values, Error bars depict ± (RQmax-RQmin)/2.

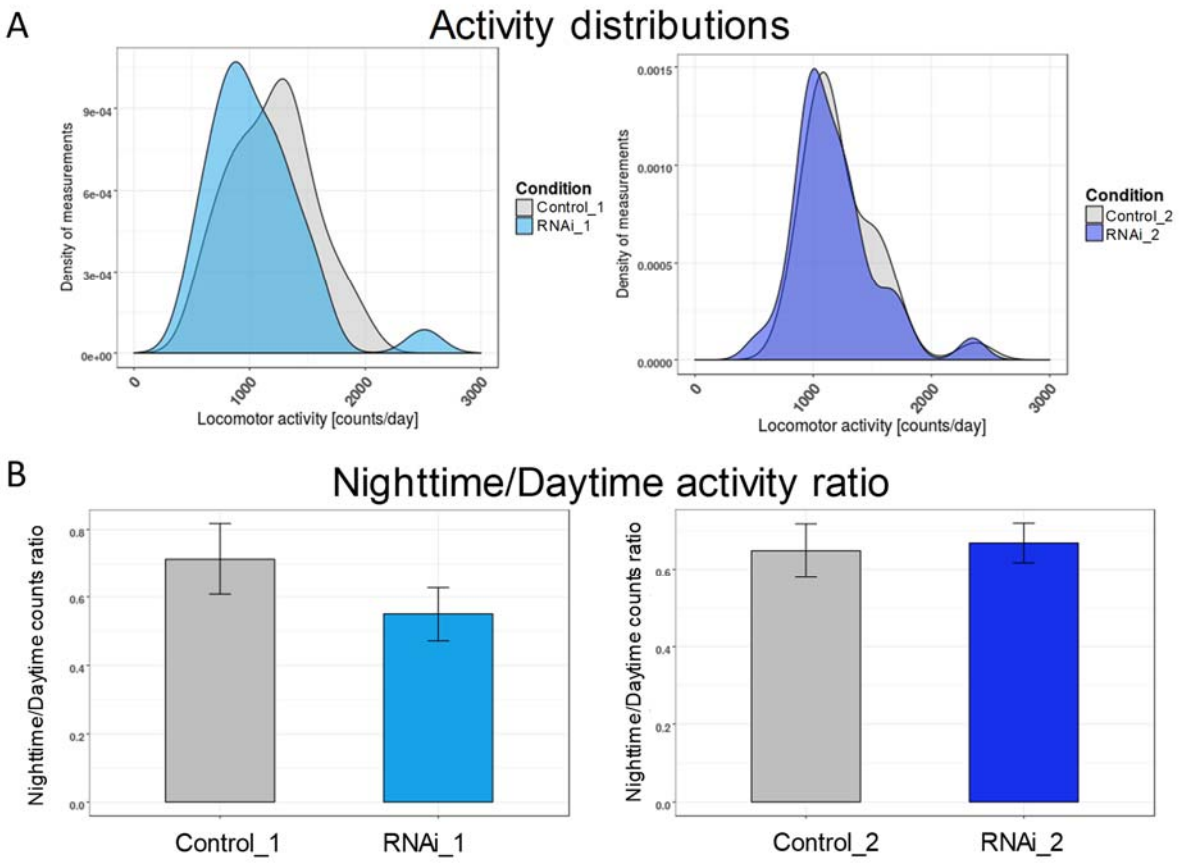


Figure S6. Normal locomotor activity of flies with pan-neural *CG4266* knockdown

The locomotor activity of flies with pan-neural (*elav*-Gal4/*Cyo*) *CG4266* knockdown at 12-hour day/night rhythm was monitored using the *Drosophila* Activity Monitor (DAM) system (Trikinetics, Waltham, MA). (A) The average daily activity was plotted as density plots. (B) Average daily nighttime/daytime activity was plotted as bar charts with SEM. Locomotor patterns in terms of total activity and activity/rest rhythm of the flies were indistinguishable from their genetic background controls.

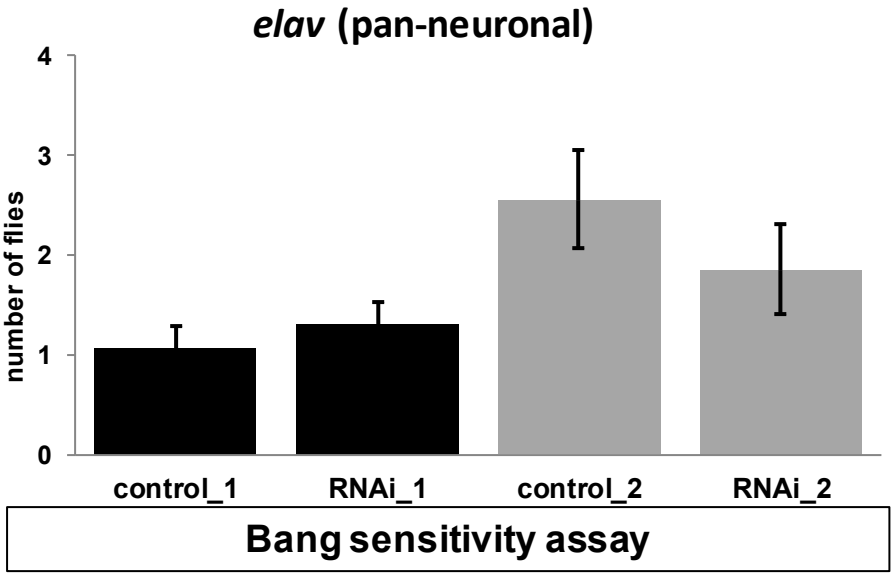


Figure S7. Bang sensitivity assay of flies with *CG4266* knockdown

Quantification of number of flies shaking on the bottom of a vial 2 seconds after vortexing. Flies with pan-neuronal knockdown of *CG4266* did not show bang sensitivity phenotypes. The diagram shows the mean value from a minimum number of 70 flies tested per genotype. Error bars represent the SEM.

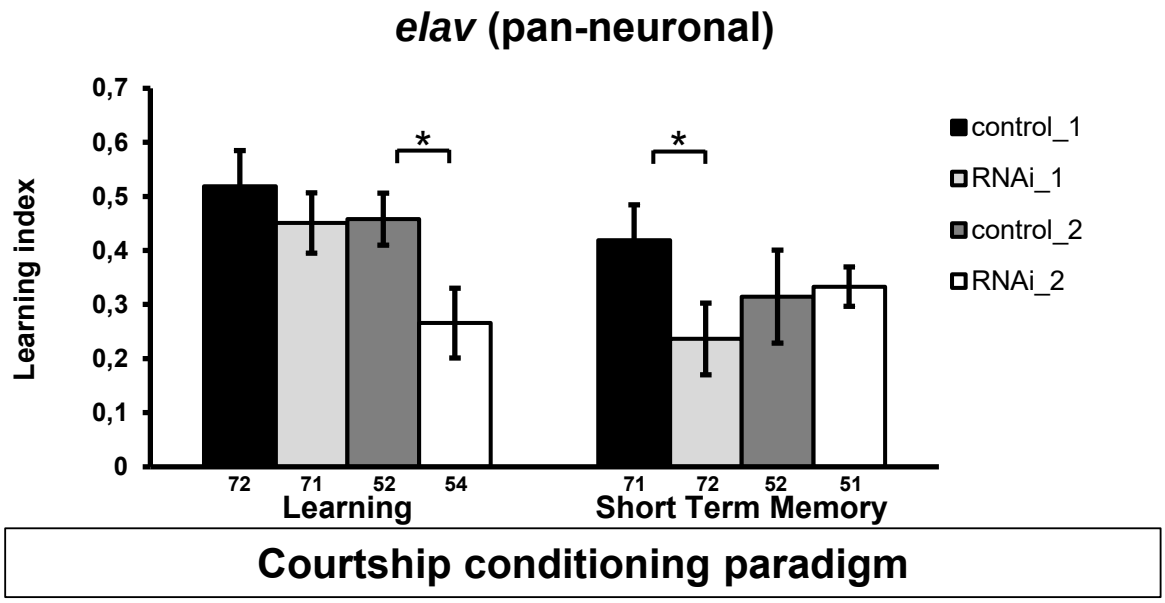


Figure S8. Courtship conditioning paradigm of flies with pan-neuronal *CG4266* knockdown

The courtship conditioning paradigm assays were performed as described previously.[6] Pan-neuronal knockdown of *CG4266* using the *elav*-x;*dicer* II driver led to significant impairment of learning in RNAi line 2 and showed a similar tendency in RNAi line 1. The short-term memory was reduced significantly in RNAi line 1, while there was no effect in RNAi line 2. Graphs display number of animals tested below the columns per genotype. Error bars represent the SEM. Asterisks indicate statistical significance (*$p < 0.05$).

Supplemental Tables

Individual	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Family History	learning difficulties in paternal half-sister	negative for ID, two olders siblings with normal development	unremarkable		negative	negative	ancestry: Dominican Republic; mother with a seizure at 9 months	maternal half sister of the mother with cleft palate and borderline ID	originally from Middle East	sister with developmental delay and behavioural anomalies	
variant in SCAF4 (NM_020706.2)	c.321+1G>T, r.spl?	c.453_456delTGAA, p.(Asn151Lysfs*8)	c.1028delC, p.(Pro343Hisfs*3)	c.1301C>A, p.(Ser434*)	c.1423C>T, p.(Arg475*)	c.1614+1G>C, r.spl?	c.1649dupT, p.(Met550Ilefs*4)	c.1812G>A, p.(Trp604*)	c. 1889G>A, p.(Trp630*)	c. 3200_3201delAG, p.(Glu1067Valfs*3)	c.783G>T, p.(Leu261Phe)
exon/intron number	intron 4	exon 5	exon 9	exon 11	exon 12	intron 13	exon 14	exon 15	exon 16	exon 20	exon 8
genomic position (GRCh37)	chr21:33076077	chr21:33074558_33074561	chr21:33068467	chr21:33066538	chr21:33065697	chr21:33064661	chr21:33064209	chr21:33063183	chr21:33060774	chr21:33043955_33043956	chr21:33069058
detected by	trio exome	exome	trio exome (Baylor)	trio exome (UCLA)	proband exome	exome	trio exome	proband exome	exome	exome	exome
test setting: clinical or research (which ethic committee?)	clinical+research (ethic committee FAU Erlangen-Nürnberg)	clinical	clinical	clinical	clinical	clinical	research, irb approval	clinical		clinical	clinical
inheritance	*de novo*	*de novo*	*de novo*	*de novo*	*de novo* (Sanger)	*de novo*	*de novo*	maternal (Sanger) (no indication for mosaicism)	*de novo*	unknown, probably inherited, as same variant in affected sister	*de novo*
additional (de novo) variants	*de novo* VUSs: *CARD11* (NM_032415.4): c.319A>T, p.(Thr107Ser); *MINK1* (NM_170663.4): c.1448_1462dup, p.(Gln483_Gln487dup)	*de novo* VUSs: *FAM83B* (NM_001010872.2): c.1424G>A, p.(Arg475His); *NDUFA5* (NM_001291304.1): c.122C>T, p.(Ser41Phe)	pathogenic comp het variants in *TPO*: c.1472G>A, p.(Arg491His), paternal, c.1184_1187dupGCCG, p.(Ala397Profs*76) maternal; *de novo* VUSs: *SLC4A4* (NM_003759.3): c.3091C>T, p.(Arg1031Cys); *DOT1L* (NM_032482.2): c.2765C>T, p.(Ala922Val);	Sotos syndrome due to a *NSD1* mutation; VUS in *ATP7A* (c.1301C>A)	*de novo* VUS: *FREM2*: c.8789A>G; pat VUS *FREM2*: c.6819A>C	no	inherited pathogenic variant in *EPM2A* (nonsense); inherited VUSs: *CHRNB2* (missense); *GRIN2A* (missense); *de novo* VUSs: *CCDC50* (missense); *CDK5RAP2* (synonymous); *CDHR5* (synonymous); *ZFC3H1* (missense)	VUS in *KANK1* (NM_015158) (unknown inheritance): c.3126C>A, p.(Asp1042Glu)	bi-allelic splicing variants in *PI4KA*	*SHANK2*; *CCDC120* (inheritance unknown)	*de novo* VUS: *MINK1*: c.3454A>G

			TMCC1 (NM_015008.5): c.578_579delAG, p.(Glu193Alafs*5); *UGT3A1* (NM_152404.3) c.223C>G, p.(Gln75Glu)								

Table S1. Variant details

Table S2. RNA-sequencing and splice analysis (excel file)

Fly Lines			
Name	**Notes**	**Identfier**	**Source**
actin-Gal4/*Cyo*	ubiquitous, Chr.2	BL#4414	BDSC
elav-x; *dicer* II	pan-neuronal, Chr. X	BL#25750	BDSC
w+, *dicer*; 247-Gal4/cyo	mushroom body, Chr.2		colleagues
elav-Gal4/*Cyo*	pan-neuronal, Chr. 2	BL#8765	BDSC
RNAi_1	*CG4266* knockdown, Chr.2	BL#55354	BDSC
RNAi _2	*CG4266* knockdown, Chr.2	v26472	VDRC
control_1	*CG4266* knockdown control, Chr.2	BL#36304	BDSC
control_2	*CG4266* knockdown control, Chr.2	v60000	VDRC

Table S3. Overview on *Drosophila* lines

BDSC, Bloomington Drosophila Stock Center; VDRC, Vienna Drosophila Resource Center

Supplemental Material and Methods

RNA-sequencing and transcriptome/splice analysis

RNA from peripheral blood from four individuals with *SCAF4* variants (I1, I4, I9, I11) was collected and extracted with the PaxGene system (PreAnalytiX, BD and Qiagen, Hombrechtikon, Switzerland). Using the TruSeq Stranded mRNA Kit (Illumina, San Diego, CA, USA), a cDNA library was constructed and sequenced (101 bp) paired-end on a HiSeq2500 Platform (Illumina, San Diego, CA, USA). Raw data were converted into reads and saved with a quality score (bcl2fastq v2.17). Reads were trimmed and filtered using cutadapt v1.18 and the abundance of individual transcripts was quantified using salmon v1.0[7; 8] together with the GENCODE GRCh38 version 32 gene reference including decoy sequences. Subsequent analyses were performed using R version 3.6.1.[9] Differential expression analysis was performed with the DESeq2 package v.1.24.0.[10] For the analysis of differentially used transcripts and the visualization of the output, the IsoformSwitchAnalyzeR package was used.[11-13]

Quantitative real-time PCR of *CG4266* RNAi fly lines

To confirm reduced *CG4266* expression for ubiquitous knockdown (*actin*-Gal4/*Cyo*) using both RNAi-lines (RNAi_1, BL#55354 and RNAi_2, vdrc 26472), quantitative real-time PCR was performed. Per sample, five surviving female flies were collected and frozen at -80°C for one hour. Total RNA was isolated according to the RNeasy Lipid Tissue Mini Kit (Qiagen), except that QIAzol was substituted by TRIzol™ (ThermoFisher Scientific). QIAshredder columns (Qiagen) were used for homogenization. cDNA was synthesized using a SuperScript™ II reverse transcriptase (ThermoFisher Scientific) and random primers. Expression analysis for *CG4266* was performed with exon-spanning primers (primer sequences available on request) using the PowerUp™ SYBR™ Green Master Mix (ThermoFisher Scientific) on a QuantStudio™ 12K Flex System (Life Technologies). *Actin* was used as an endogenous control, fly lines with the same genetic background but lacking the dsRNA element were used as reference. Quantitative real-time PCR was performed in technical quadruplicates. RQ values were calculated using the ΔΔCt method with controls as reference.

Neuromuscular junctions

Analysis of type 1b neuromuscular junctions (NMJs) of muscle 4 was performed as described previously.[14] *CG4266* RNAi lines and corresponding genetic background control lines were crossed to the UAS-*dicer*-II;*elav*-Gal4 (pan-neuronal) driver line. Briefly, male L3 larvae were dissected, fixated in 4% paraformaldehyde and stained with nc82 and anti-dlg antibodies (Developmental Studies Hybridoma Bank). Secondary antibodies used were Alexa 488 labeled anti mouse antibody and the Zenon® Alexa Fluor 546 Mouse IgG1 Labeling Kit (Life Technologies). Images were taken with a Zeiss Axio Imager Z2 microscope with 10x and 63x objectives, and NMJ pictures were subsequently stacked and analyzed in ImageJ. Synaptic area and length and numbers of synaptic branches, boutons, and active zones were determined, blinded to the genotype. For each genotype, at least 20 NMJs from a total of five to eight different animals were analyzed.

Locomotor activity profiling

The locomotor activity was analyzed with the *Drosophila* Activity Monitor system (DAM; Trikinetics, Waltham, MA). Three to five days old single male flies were transferred to monitor tubes with standard food. Activity data were collected every minute over the course of four days with a 12-hour light/dark cycle. Activity counts represent the amount of infrared beam passes by the fly. The analysis was performed using the ShinyR-DAM tool (https://karolcichewicz.shinyapps.io/shinyr-dam/).[15]

Bang sensitivity assay

Bang sensitivity is characterized by paralysis and hyperactivity after a mechanical shock and can be induced by vortexing at maximal strength for 10 seconds.[16] The bang sensitivity assay was performed as described previously.[6] *CG4266* lines and the corresponding genetic background control lines were crossed to the *elav*-Gal4/*Cyo* (pan-neuronal) driver line. A minimum of 70 flies per genotype and matching control were collected 0-48 hours after eclosion under CO_2 anesthesia in groups of 10 (balanced male/female ratio) and kept in normal

food vials for 24 hours. After transferal to testing vials and an adjustment time of one minute, flies were vortexed for 10 seconds. The number of flies displaying paralysis or spasms two seconds after vortexing was assessed.

Climbing assay

The climbing assay was carried out as described by Palladino *et al.* with some modifications.[6; 17] *CG4266* RNAi lines and the corresponding genetic background control lines were crossed to the *elav*-Gal4/*Cyo* (pan-neuronal) driver line. A minimum of 170 flies per genotype and matching control, respectively, were collected and prepared for trial as described for the bang sensitivity assay. After transferal to testing vials and an adjustment time of one minute, flies were tapped to the bottom of the vial and videotaped for 30 seconds. From the tapes, time was measured until 70% of flies had climbed 8.8 cm.

Courtship conditioning paradigm

The courtship conditioning paradigm assays were performed as described previously.[6] Flies were kept at 25°C and 70% humidity at a 12:12 light–dark cycle. Virgin males were trained individually by pairing them with premated females. Learning and short-term memory were tested immediately or one hour after a training period of one hour, respectively. The courtship index (CI), the percentage of time each male spent courting a non-receptive female was manually assessed from 10 minutes movies. By comparing the CI of naïve and trained males a learning index (LI) was calculated: LI = (CInaive – CItrained) / CInaive. Differences between learning indices of control and *CG4266* knockdown flies were statistically compared by a randomization test with 10,000 bootstrap replicates with a custom R script.[14]

Supplemental References

1. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nature biotechnology *29*, 24-26.
2. Schwarz, J.M., Rodelsperger, C., Schuelke, M., and Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods *7*, 575-576.
3. Adzhubei, I., Jordan, D.M., and Sunyaev, S.R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. Current protocols in human genetics *Chapter 7*, Unit7.20-Unit27.20.
4. Sim, N.-L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., and Ng, P.C. (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Research *40*, W452-W457.
5. Jagadeesh, K.A., Wenger, A.M., Berger, M.J., Guturu, H., Stenson, P.D., Cooper, D.N., Bernstein, J.A., and Bejerano, G. (2016). M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. Nature Genetics *48*, 1581-1586.
6. Straub, J., Konrad, E.D.H., Grüner, J., Toutain, A., Bok, L.A., Cho, M.T., Crawford, H.P., Dubbs, H., Douglas, G., Jobling, R., et al. (2018). Missense Variants in RHOBTB2 Cause a Developmental and Epileptic Encephalopathy in Humans, and Altered Levels Cause Neurological Defects in Drosophila. American journal of human genetics *102*, 44-57.
7. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal; Vol 17, No 1: Next Generation Sequencing Data AnalysisDO - 1014806/ej171200.
8. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nature methods *14*, 417.
9. (2018). RStudio: Integrated Development for R. http://wwwrstudiocom/*3.6.1*.
10. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology *15*, 550.

11. Soneson, C., Love, M., and Robinson, M. (2016). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences [version 2; peer review: 2 approved]. F1000Research *4*.

12. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research *43*, e47-e47.

13. Vitting-Seerup, K., and Sandelin, A. (2017). The Landscape of Isoform Switches in Human Cancers. Molecular Cancer Research *15*, 1206.

14. Gregor, A., Kramer, J.M., van der Voet, M., Schanze, I., Uebe, S., Donders, R., Reis, A., Schenck, A., and Zweier, C. (2014). Altered GPM6A/M6 dosage impairs cognition and causes phenotypes responsive to cholesterol in human and Drosophila. Hum Mutat *35*, 1495-1505.

15. Cichewicz, K., and Hirsh, J. (2018). ShinyR-DAM: a program analyzing Drosophila activity, sleep and circadian rhythms. Communications biology *1*, 25-25.

16. Kuebler, D., and Tanouye, M.A. (2000). Modifications of seizure susceptibility in Drosophila. J Neurophysiol *83*, 998-1009.

17. Palladino, M.J., Hadley, T.J., and Ganetzky, B. (2002). Temperature-sensitive paralytic mutants are enriched for those causing neurodegeneration in Drosophila. Genetics *161*, 1197-1208.