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

Anna Fliedner,[1] Philipp Kirchner,[1] Antje Wiesener,[1] Irma van de Beek,[2] Quinten Waisfisz,[2] Mieke van Haelst,[2] Daryl A. Scott,[3,4] Seema R. Lalani,[3] Jill A. Rosenfeld,[3,5] Mahshid S. Azamian,[3] Fan Xia,[3,5] Marina Dutra-Clarke,[6] Julian A. Martinez-Agosto,[6,7] Hane Lee,[7,8] UCLA Clinical Genomics Center, Grace J. Noh,[9] Natalie Lippa,[10] Anna Alkelai,[10] Vimla Aggarwal,[11] Katherine E. Agre,[12] Ralitza Gavrilova,[12,13] Ghayda M. Mirzaa,[14,15,16] Rachel Straussberg,[17,18] Rony Cohen,[18,19] Brooke Horist,[20] Vidya Krishnamurthy,[20] Kirsty McWalter,[21] Jane Juusola,[21] Laura Davis-Keppen,[22] Lisa Ohden,[22] Marjon van Slegtenhorst,[23] Stella A. de Man,[24] Arif B. Ekici,[1] Anne Gregor,[1] Ingrid van de Laar,[23] and Christiane Zweier[1,*]

RNA polymerase II interacts with various other complexes and factors to ensure correct initiation, elongation, and termination of mRNA transcription. One of these proteins is SR-related CTD-associated factor 4 (SCAF4), which is important for correct usage of polyA sites for mRNA termination. Using exome sequencing and international matchmaking, we identified nine likely pathogenic germline variants in *SCAF4* including two splice-site and seven truncating variants, all residing in the N-terminal two thirds of the protein. Eight of these variants occurred *de novo*, and one was inherited. Affected individuals demonstrated a variable neurodevelopmental disorder characterized by mild intellectual disability, seizures, behavioral abnormalities, and various skeletal and structural anomalies. Paired-end RNA sequencing on blood lymphocytes of *SCAF4*-deficient individuals revealed a broad deregulation of more than 9,000 genes and significant differential splicing of more than 2,900 genes, indicating an important role of SCAF4 in mRNA processing. Knockdown of the *SCAF4* ortholog *CG4266* in the model organism *Drosophila melanogaster* resulted in impaired locomotor function, learning, and short-term memory. Furthermore, we observed an increased number of active zones in larval neuromuscular junctions, representing large glutamatergic synapses. These observations indicate a role of CG4266 in nervous system development and function and support the implication of *SCAF4* in neurodevelopmental phenotypes. In summary, our data show that heterozygous, likely gene-disrupting variants in *SCAF4* are causative for a variable neurodevelopmental disorder associated with impaired mRNA processing.

The regulation of protein-coding gene transcription is a highly complex process that is crucial for proper gene expression. RNA polymerase II (RNAPII) interacts with many other complexes and factors to ensure accurate pre-initiation, initiation, elongation, and termination of mRNA transcription (reviewed in Cramer,[1] Orphanides and Reinberg,[2] Kornberg,[3] and Lee and Young[4]). Mutations in several genes and proteins involved in mRNA processing have been implicated in human diseases. For example, heterozygous pathogenic variants in *POLR2A* (MIM: 180660), which encodes the largest subunit of RNAPII, have recently been identified as the cause of a neurodevelopmental disorder (NDD) with hypotonia and variable intellectual and behavioral anomalies (NEDHIB [MIM: 618603]).[5] Similarly, pathogenic variants in several subunits of the transcription factor IID complex, which plays a key role in transcriptional initiation,[6,7] have been implicated in NDDs. These include pathogenic variants in TATA binding protein associated factors, e.g., an X-linked NDD (MRXS33 [MIM: 300966]) is caused by variants in *TAF1* (MIM: 313650) and an autosomal-recessive NDD (MRT60 [MIM: 617432]) by variants in *TAF13* (MIM: 600774).[8,9] Furthermore, bi-allelic variants in two subunits of the RNAPII interacting Integrator complex, *INTS1* (MIM: 611345) and *INTS8* (MIM: 611351), have been shown to be associated with NDDs (NDCAGF [MIM: 618571] and NEDCHS [MIM: 618572]) and lead to both altered splicing patterns and differential

[1]Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91054 Erlangen, Germany; [2]Department of Clinical Genetics, Amsterdam UMC, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, the Netherlands; [3]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; [4]Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, TX 77030, USA; [5]Baylor Genetics Laboratories, Houston, TX 77021, USA; [6]Division of Medical Genetics, Department of Pediatrics, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA 90095, USA; [7]Department of Human Genetics, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA 90095, USA; [8]Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA 90095, USA; [9]Department of Genetics, Southern California Permanente Medical Group, Fontana, CA 92335, USA; [10]Institute for Genomic Medicine, Columbia University Irving Medical Center, New York, NY 10019, USA; [11]Institute for Genomic Medicine and Department of Pathology and Cell Biology, Columbia University Irving Medical Center, New York, NY 10019, USA; [12]Department of Clinical Genomics, Mayo Clinic, Rochester, MN 55905, USA; [13]Department of Neurology, Mayo Clinic, Rochester, MN 55905, USA; [14]Center for Integrative Brain Research, Seattle Children's Research Institute, Seattle, WA 98195, USA; [15]Department of Pediatrics, University of Washington, Seattle, WA 98195, USA; [16]Brotman Baty Institute for Precision Medicine, Seattle, WA 98195, USA; [17]Neurogenetics Clinic, Neurology Unit, Schneider Children's Medical Center, Petah Tikva 49202, Israel; [18]Sackler School of Medicine, Tel Aviv University, Ramat Aviv, Tel Aviv 69978, Israel; [19]Neurology Institute and Epilepsy Center, Schneider Children's Medical Center, Petah Tikva 49202, Israel; [20]Pediatrics and Genetics, Alpharetta, GA 30005, USA; [21]Clinical Genomics, GeneDx, Gaithersburg, MD 20877, USA; [22]University of South Dakota, Sanford School of Medicine, Sioux Falls, SD 57105, USA; [23]Department of Clinical Genetics, Erasmus MC, University Medical Center Rotterdam, 3015 GD Rotterdam, the Netherlands; [24]Department of Pediatrics, Amphia Hospital, 4818 CK Breda, the Netherlands
*Correspondence: christiane.zweier@uk-erlangen.de
https://doi.org/10.1016/j.ajhg.2020.06.019.

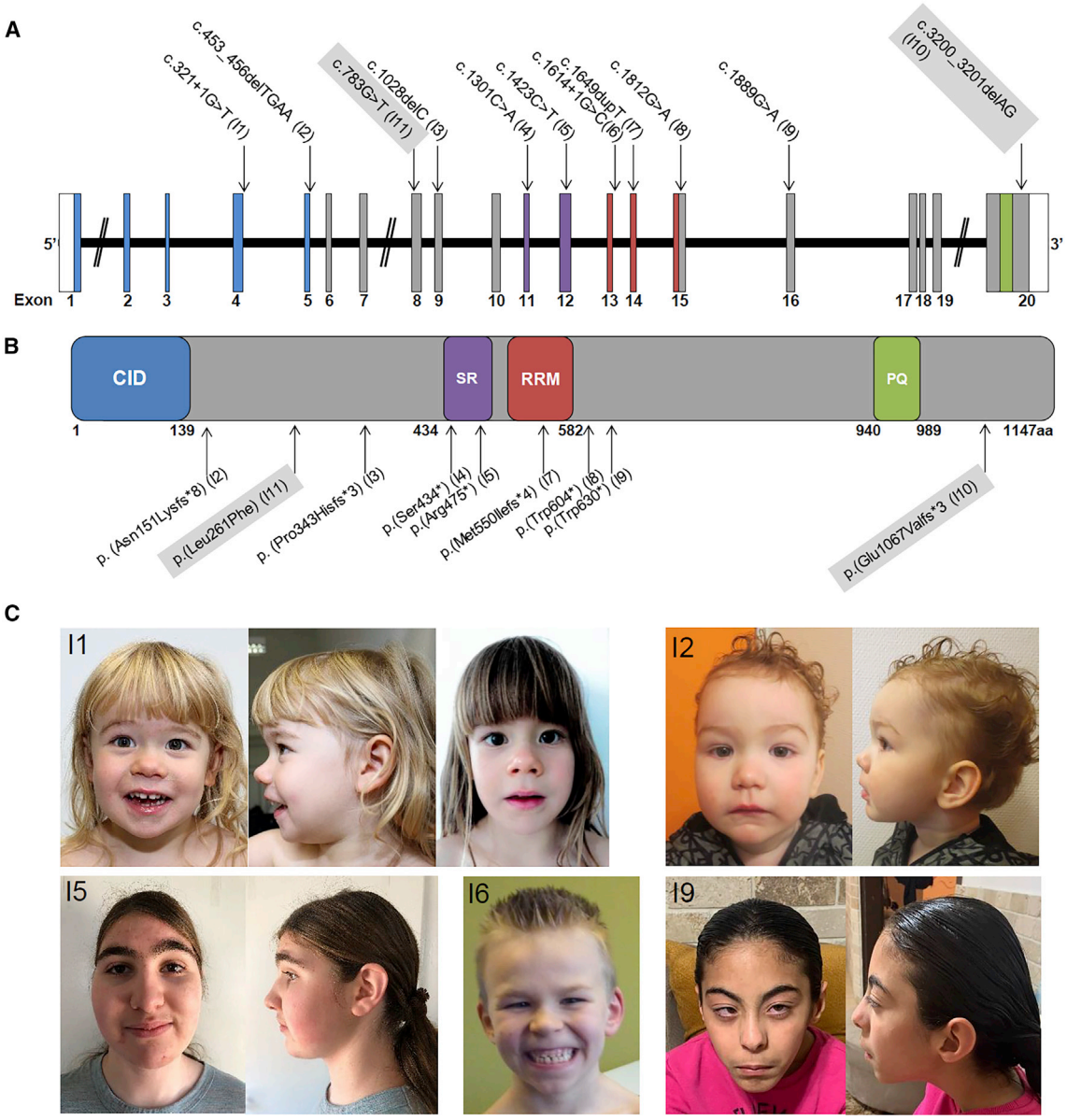


Figure 1. *SCAF4* Variants Observed in Individuals with NDDs

(A) Schematic drawing of *SCAF4* (longest isoform GenBank: NM_020706.2) with identified variants. Non-coding exonic regions are displayed in white, coding exons in gray, and encoded domains (according to Ensembl[12]) in color.

(B) Schematic drawing of the SCAF4 protein. One missense and eight likely gene-disrupting variants (excluding two splice-site variants) are displayed below the scheme. Variants of unknown significance are shaded in gray.

(C) Clinical pictures of individuals with pathogenic variants in *SCAF4* (I1 at age 3 and 5 years; I2 at age 20 months; I5 at age 13 years; I6 at age 6.5 years; I9 at age 9 years). Note common facial features such as epicanthus, a deep nasal bridge, bulbous nasal tip, and deep philtrum, particularly in I1 and I2.

Abbreviations: CID, conserved CTD-interacting domain; SR, Ser/Arg-rich domain; RRM, RNA recognition motif; PQ, Pro/Gln-rich domain.

gene expression in cells derived from affected individuals.[10]

Using trio exome sequencing on an Illumina HiSeq 2500 platform and data analysis with an in-house pipeline as described previously,[11] we identified the *de novo* splice-site variant c.321+1G>T in *SCAF4* (MIM: 616023) (GenBank: NM_020706.2) in an individual with mild intellectual disability (ID) and seizures. *SCAF4* consists of 20 exons and encodes SR (serine and arginine)-related CTD (C-terminal domain)-associated factor 4, consisting of 1,147 amino acids and containing an N-terminal conserved CTD-interacting domain and an RNA recognition motif (Figures 1A and 1B).[13] SCAF4 interacts with the C-terminal domain of the largest subunit of RNAPII, and together with SCAF8, is required for correct polyA site selection and mRNA termination.[13,14] So far, *SCAF4* has not been implicated in NDDs or other diseases. According to gnomAD[15] constraint scores, *SCAF4* is intolerant

toward loss-of-function variants (pLI = 1, o/e = 0.03 (0.01–0.11)), thus supporting a pathogenic relevance of the identified splice-site variant.

We used GeneMatcher[16] to identify eight other unrelated individuals with NDDs and likely pathogenic variants in *SCAF4* and two individuals with variants of unknown significance (Table 1, Table S1). Variants were identified through routine diagnostic testing or in a research setting approved by the ethical review boards of the respective institutions (Table S1). Informed consent was obtained from the parents or legal guardians of the affected individuals.

The nine variants that were considered likely pathogenic included two splice-site and seven truncating variants (Figures 1A and 1B). None of them were present in gnomAD.[15] Eight occurred *de novo*, and one nonsense variant c.1812G>A (p.Trp604*) was inherited from a healthy mother. gnomAD[15] contains 14 presumably truncating variants in *SCAF4*, of which 10 are located in the last exon and probably are benign due to escaping nonsense-mediated mRNA decay. In contrast, the two splice-site and seven truncating variants in our cohort are predicted to affect the N-terminal two thirds of the protein, and therefore likely trigger nonsense-mediated mRNA-decay or produce a severely truncated protein. Nonsense-mediated mRNA decay was indicated by reverse transcription PCR (RT-PCR) on cDNA/RNA from a PaxGene (PreAnalytiX, BD and QIAGEN) blood sample of I4 with the c.1301C>A (p.Ser434*) variant (Figure S1A). In contrast, the mutant c.1889G>A (p.Trp630*) allele in I9 was still equally visible. However, as RNA sequencing (see below) indicated reduced *SCAF4* expression also in this individual (Figure S1B, Table S2), a later decay process might occur. Furthermore, we confirmed aberrant splicing for the splice-site variant in I1 by showing several aberrant transcripts with complete or partial loss of exons 3, 4, or 5 (Figures S2A and S2B). RNA sequencing in I1 additionally revealed another isoform with intron 4 retained (Figures S2C and S2D). For all truncating variants in the N-terminal part of *SCAF4*, a general loss-of-function mechanism or haploinsufficiency is therefore likely.

The Decipher database[17,18] contains several deletions including *SCAF4*, but all encompass a large number of additional genes, thereby preventing specific deductions regarding *SCAF4* haploinsufficiency in these individuals.

Additionally, we identified two variants of unknown significance. The frameshifting variant c.3200_3201del (p.Glu1067Valfs*3) segregated in two siblings with neurodevelopmental phenotypes. Parents were not available for testing. However, it is located in the C-terminal part of the gene/protein, where most of the truncating variants in gnomAD reside. The *de novo* missense variant c.783G>T (p.Leu261Phe) was identified in an individual with a consistent NDD phenotype. Though *SCAF4* is not particularly intolerant toward missense variants (gnomAD constraint scores: z = 1.94, o/e = 0.79 (0.73–0.85)) and though this variant is not located in any of the known functional domains, there is some evidence pointing to a possible pathogenic relevance. This variant is not present in gnomAD, affects a highly conserved amino acid (Figure S3A), is predicted to be deleterious by Mutation Taster,[19] PP2,[20] SIFT,[21] and M-CAP[22] (Figure S3B), and resulted in reduced *SCAF4* expression in RNA sequencing in I11, while mutant and wild-type alleles were equally visible in RT-PCR (Figure S1C).

The clinical features of the affected individuals are summarized in Table 1. For phenotypic delineation, we excluded I10 and I11 due to their unclear variant status and I4 who additionally has Sotos syndrome (SOTOS1 [MIM: 117550]) due to a pathogenic variant in *NSD1* (MIM: 606681), confounding the clinical picture. The remaining eight individuals all had developmental delay and intellectual disability, mostly in the mild range. Speech was more severely affected than motor development. While age of walking was within a normal range of 12–18 months in seven individuals, only two individuals had normal speech development (first words before 15 months). Speech was severely delayed (first words after 2 years) in four individuals. Developmental stagnation co-occurring with seizures was reported in one and possible developmental regression in two individuals. Behavioral anomalies were reported in five individuals (63%) and included autistic features, hyperactivity, and aggressive behavior. Seizures occurred in four individuals (50%) and included myoclonic seizures in two and intractable seizures in one. Brain MRIs were performed in five individuals showing nonspecific white matter anomalies in three of them. I9 was diagnosed with pontocerebellar hypoplasia and a thin corpus callosum. Variable other features included renal (38%), cardiac (38%), or skeletal (63%) anomalies. I3 presented with multiple malformations and anomalies. Minor but rather unspecific facial dysmorphism were noted in most of the individuals. Shared facial features in two individuals (I1, I2) were epicanthus, a flat nasal bridge, a bulbous nasal tip, and a deep philtrum (Figure 1C).

Taken together, likely gene-disruptive variants in *SCAF4* are associated with a variable neurodevelopmental phenotype with predominantly mild developmental delay and intellectual disability and frequently with seizures and behavioral anomalies. As increasingly observed for other genes associated with mild NDDs, putatively pathogenic, autosomal-dominant variants in *SCAF4* may be inherited from a mildly affected or presumably healthy parent, as observed for I8.

SCAF4 interacts with RNAPII. Interestingly, also variants in *POLR2A*, encoding the largest subunit of RNAPII, have been identified to cause a neurodevelopmental disorder, possibly due to a dominant-negative effect on RNA transcription.[5] While variable developmental delay, behavioral abnormalities, and white matter anomalies in brain imaging overlap between individuals with either *SCAF4* or *POLR2A* variants, the *POLR2A*-associated phenotype

Table 1. Clinical Details

Individual	1	2	3	4	5	6	7	8	9	10	11
Gender	f	m	m	m	f	m	m	m	f	m	m
Age	3y 3mo	20mo	16y	18y	13y	6y 6mo	6y 5mo	11y	4.5y	4y	10y
Variant cDNA	c.321+1G>T	c.453_456delTGAA	c.1028delC	c.1301C>A	c.1423C>T	c.1614+1G>C	c.1649dupT	c.1812G>A	c. 1889G>A	c.3200_3201delAG	c.783G>T
Variant RNA/ protein	r.spl?	p.Asn151Lysfs*8	p.Pro343Hisfs*3	p.Ser434*	p.Arg475*	r.spl?	p.Met550Ilefs*4	p.Trp604*	p.Trp630*	p.Glu1067Valfs*3	p.Leu261Phe
De novo	yes	yes	yes	yes	yes	yes	yes	maternal	yes	unknown (also in affected sister)	yes
Height cm/ SD	101 / 0.66	84.5 / −0.28	162.6 / −1.6	165.5 / −1.51	157.5 / −0.41	126.5 / 1	114 / −0.22 (5y)	136 / −1.69	125 / −1.24 (8.5y)	N/A	142 / −0.25
Weight kg/SD	17.2 / 1.04	12.5 / +0.10	67.5 / 0.40	56.9 / −1.25	44.4 / −0.61	24.9 / 0	26 / 1.56 (5y)	38.2 / −0.14	21 / 1.19	N/A	62.7 / 2.39
OFC cm/SD	49.2 / −0.38	47.5 / −0.8	N/A	59 / 1.42	51.8 / −1.9	52.8 / 0	N/A	55 / 0.54	49.2 / −0.92	N/A	57.6 / 2.77
DD/ID	mild-moderate (SON-IQ 1x67, 1x50)	mild	mild	severe	mild	mild (tIQ 91 WISC-V)	language delay	learning diff.	mild-moderate (DQ 56)	moderate	yes
Walking	14mo	17mo	14mo	4–5y	14mo	12mo	18mo	14–15mo	29mo	N/A	15mo
First words	12mo	18mo	2.5y	5y	14mo	16mo	3y	2–3y	3.7y	not yet	3–4y
Speech	3y 3mo: 30 words	N/A	3y: speech therapy	10 words	>2y: 2-word comb.	delay	N/A	11y: 20 words	simple sentences	4y: no speech	N/A
Regression	stagnation with seiz.	no	no	N/A	possibly in speech	no	no	no	yes (10–12mo)	N/A	possibly
Seizures, onset age	4x, 27–30mo	no	no	intractable (myoclonic/ tonic), 10mo	myoclonus, 12y	no	myoclonic astatic epilepsy, 4y	intractable, 18mo	no	N/A	yes, 2y
MRI anom.	subcort., periventr. hypomyelination	N/D	N/D (normal US)	possible focal cortical dysplasia, volume loss	normal	N/D	nonspecific FLAIR hyperintensities frontal subcortical area left > right	white matter changes, cerebellar atrophy	PCH, HCC	N/A	HCC
Muscular hypotonia	yes	yes	no	yes	no	N/R	N/R	hypertonia	yes	N/A	yes
Behavioral anom.	autistic features	N/A	autistic features in infancy	autism, aggression, hyperactivity	autism	autism, hyperactivity, aggression	no	self-injurous, aggression	aggression	autism, hyperactivity	disruptive behavior, autism

(*Continued on next page*)

Table 1. ***Continued***

Individual	1	2	3	4	5	6	7	8	9	10	11
Gender	f	m	m	m	f	m	m	m	f	m	m
Cardiac defect	N/D	murmur, US normal	VSD, bicuspid aortic valve, hypoplastic aortic arch, dilated cardiomyopathy	N/A	VSD (self-resolved)	ND	N/R	N/R	VSD (self-resolved), PFO	N/A	N/R
Renal ano.	N/D	N/D	unilateral agenesis	multicystic kidneys	unilateral hydronephrosis	no	N/R	N/R	multicystic kidneys	N/A	N/R
Urogenital anom.	no	no	cryptorchidism	cryptorchidism	no	no	N/R	N/R	inguinal hernia	N/A	N/R
GI anomalies	no	no	TE fistula, imperforate anus	s/p pyloroplasty, nissen fundoplication, G-tube	none	no	N/R	N/R	no	N/A	N/R
Skeletal anom.	no	no	sacrum segmentation anom., brachydactyly	kyphosis, scoliosis	lordosis, hallux valgus, toe syndactyly II/III	kyphosis	N/R	bilateral ankle rotation, pectus excavatum	antevertion of femur	N/A	scoliosis, pronation of feet
Other	none	none	tethered cord, hypothyroidism	sleep apnea, chronic lung disease, bronchopulmonary dysplasia, Sotos syndrome	premature adrenarche	none	none	delayed teeth eruption	none	none	none

f, female; m, male; y, years; mo, months; SD, standard deviation; OFC, occipito-frontal head circumference; DD, developmental delay; ID, intellectual disability; N/A, not available or not applicable; N/D, not done; N/R, not reported; anom., anomaly; HCC, hypoplasia of corpus callosum; VSD, ventricle septum defect; PFO, persistent foramen ovale; PCH, pontocerebellar hypoplasia; TE, tracheesophageal; US, ultrasound.

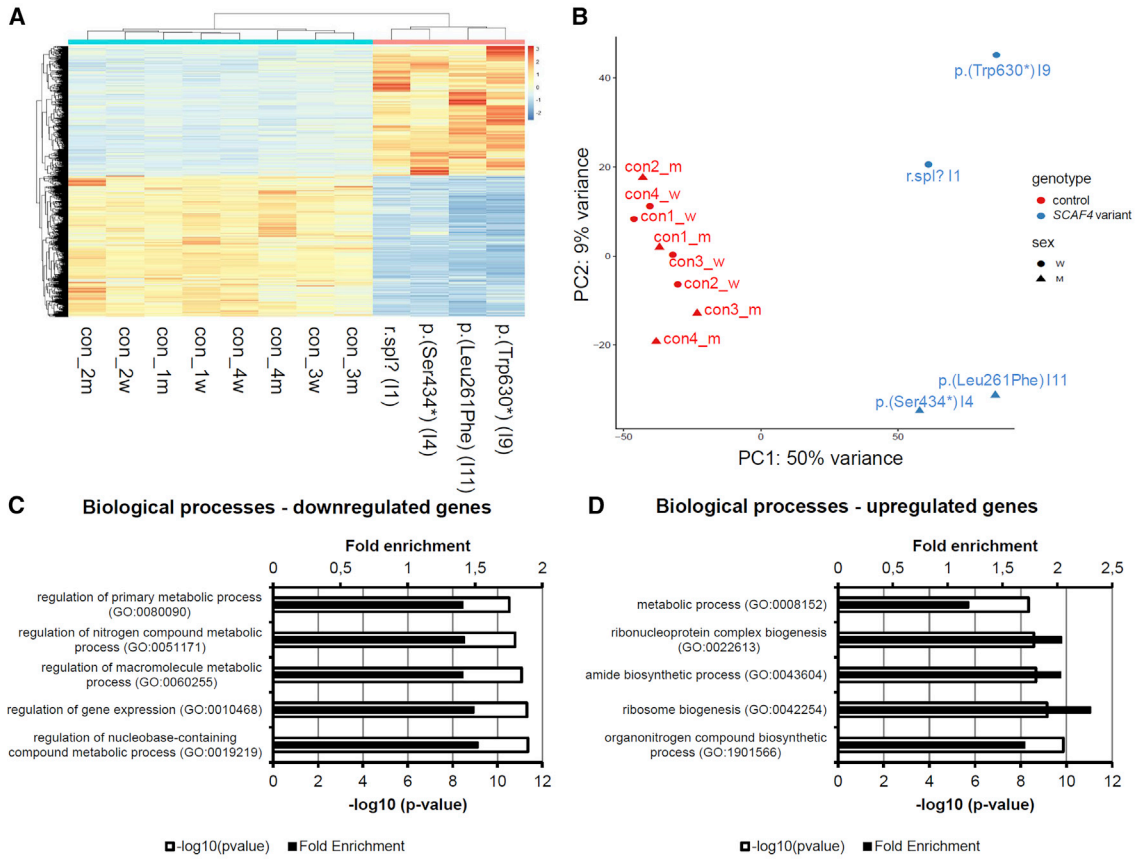


Figure 2. Transcriptome Analysis in Four Individuals with Variants in *SCAF4*

(A) Heatmap displaying 9,038 differentially expressed genes between four affected individuals and eight healthy control subjects and depicting the expression similarity between individuals with adjusted p value < 0.01. A total of 4,710 genes were downregulated and 4,328 genes were upregulated in the affected individuals. Heatmap was created using the pheatmap package v.1.0.12 (see Web Resources) applying the standard settings (complete linkage method for hierarchical clustering for both columns and rows). Expression values were normalized, and genes were sorted by adjusted p value. Blue, downregulated; red, upregulated.

(B) Principal component analysis (PCA) of affected individuals and control subjects separated by genotype and sex. Affected individuals and control subjects readily clustered within their group. PCA plot was created using the DESeq2 package version 1.24.0.[23]

(C and D) Gene Ontology (GO) term analysis of (C) downregulated and (D) upregulated genes depicting the top five GO terms for biological processes.

appears to be more severe with profound muscular hypotonia as a prominent feature.[5]

In the context of its interaction with RNAPII, an important role for SCAF4 in mRNA polyA recognition and mRNA termination was characterized recently.[14] CRISPR-Cas9-mediated knockout of *SCAF4* in HEK293 cells resulted in increased transcriptional read-through or in the usage of alternative last exons. This, in turn, was visible as an increase in the number of genes with an elongated 3′ sequence over genes with a truncated one.[14]

To investigate whether SCAF4 defects also result in incorrect mRNA transcription termination and mRNA processing in cells from affected individuals, we performed paired-end RNA sequencing on RNA from peripheral blood, collected and extracted with the PaxGene system (PreAnalytiX, BD and QIAGEN), of four individuals with *SCAF4* variants (I1, I4, I9, I11) and eight healthy control subjects. In the affected individuals, we then determined deregulated genes, performed principal component analysis (PCA) using DESeq2 package v.1.24.0,[23] and determined differentially used isoforms using salmon v.1.0.[24,25] More details are available in Supplemental Material and Methods. Of note, this approach detects only known and annotated isoforms and indicates shifts in their respective usage.

We found that a very large number of genes (n = 9,038) were differentially expressed between the groups of four affected individuals and the eight control subjects, with an adjusted p value < 0.01 (FDR corrected). Of those, 4,328 were upregulated and 4,710 were downregulated (Table S2). Affected individuals and control subjects readily clustered within their group (Figures 2A and 2B), including I11 with the missense variant, thus pointing to a similar loss-of-function mechanism as for the truncating variants and supporting its pathogenic relevance. However, additional data as well as direct comparison to expression profiles of individuals with other NDDs would be required to draw any further conclusions on the possible pathogenicity of this or other missense variants in *SCAF4*.

Enrichment of Gene Ontology (GO) terms as well as Reactome Pathways[26] among down- and upregulated genes were analyzed using the PANTHER v.14.0[27] enrichment tool from Gene Ontology[28,29] with the following settings: test type: "Fisher's Exact;" correction: "Bonferroni correction for multiple testing." A list of all expressed genes within control subjects and the affected individuals with a base mean > 2 was used as background. In this analysis, we found that downregulated genes were enriched for GO terms that included regulation of gene expression and regulation of nucleobase-containing compound metabolic processes (Figure 2C). Furthermore, they were enriched for Reactome pathways that included translation, gene expression, and RNA Polymerase II Transcription (Table S2). Upregulated genes were enriched for GO terms like ribosome biogenesis and organonitrogen compound biosynthetic processes (Figure 2D), as well as for Reactome pathways like translation, metabolism of RNA, and rRNA processing (Table S2).

In addition to broad transcriptional deregulation, we also detected differential splicing of 2,942 genes (adjusted p value < 0.05) between affected individuals and control subjects, supporting the role of SCAF4 in mRNA processing. We selected two alternatively spliced genes and confirmed a shift in the usage of alternative transcripts using RT-PCR with primers flanking differentially spliced regions (Figure S4).

The most prominent splicing effect observed was the increased fraction of short transcripts in more than 70% of all events (2,095 out of 2,942 total) (Figure 3A). This was mainly due to usage of alternative down-stream transcription start sites (TSSs) (70%) and/or alternative upstream transcription termination sites (TTSs) (37%) and/or exon loss (75%) (Figure 3B). Interestingly, for most of the transcripts, truncation was caused by a combination of two or more of these alterations with all three modifications occurring simultaneously in 17.8% of all altered and in 25.1% of shortened transcripts. As an example, the isoform switch of *BTLA* (MIM: 607925) is shown. Here, two shorter isoforms due to usage of downstream TSS, upstream TTS, and exon loss have an increased expression level (Figure 3C).

Similarly to the observations from differentially expressed genes, differentially spliced genes were enriched for GO terms like nucleic acid metabolic processes and gene expression as well as for Reactome pathways like chromatin modifying enzymes, chromatin expression, and gene expression (transcription) (Figure 3D, Table S2).

We compared the differentially spliced genes in individuals with variants in *SCAF4* with the 565 differentially spliced genes after *SCAF4* knockout in HEK293 cells[14] and found an overlap of 25.7% (145 out of 565). A total of 66 (45.5%) of these genes used an alternative last exon (ALE). As various cell types usually have unique transcriptome profiles, usage of HEK293 versus human blood cells might contribute to discrepancies between differentially spliced genes. Also, complete *SCAF4* knockout in HEK293 cells versus a heterozygous variant in affected individuals could play a role, since the remaining allele might still cover some of the SCAF4 function. Nevertheless, the overlap between differentially expressed genes in HEK293 and blood cells is larger than expected by chance (95 out of 565) as determined in a hypergeometric test (p value $<$ 0.001), and the general consequence of altered mRNA termination is similar in both systems. Furthermore, we frequently found usage of alternative transcription start sites highlighting the potential that SCAF4 is involved not only in correct mRNA termination but also in TSS recognition. Such a function could directly contribute to the large number of changes in gene expression.

To find more evidence that *SCAF4* truncation or haploinsufficiency causes NDDs, we utilized *Drosophila melanogaster* as a model organism and investigated the effect of knockdown of *CG4266*, the ortholog of both human *SCAF4* and *SCAF8* (MIM: 616024), with regard to nervous system development and function. We obtained two *CG4266* UAS-RNAi lines (RNAi_1, BL#55354 and RNAi_2, VDRC 26472) from the Transgenic RNAi Project[33] from the Bloomington Stock Center (BL) and from the Vienna *Drosophila* Resource Center (VDRC), respectively. Ubiquitous or tissue-specific knockdown was induced using the UAS/Gal4 system[34] with different promoter lines (Table S3). By real-time quantitative PCR we confirmed reduced *CG4266* expression to approximately 60%–70% upon ubiquitous knockdown (Figure S5). We first investigated development and morphology of larval neuromuscular junctions (NMJ), an established model for vertebrate glutamatergic synapses,[35,36] upon pan-neuronal knock-down. Specifically, we dissected stage 3 larvae and stained NMJs in larval muscle 4 as described previously,[37] to assess parameters such as NMJ length and area, number of synaptic boutons, and the number of active zones where neurotransmitters are released (Supplemental Material and Methods, Figure 4A). Pan-neuronal knockdown of *CG4266* with either of the two RNAi lines resulted in an increased number of active zones, while the number of branches or the number of boutons was elevated only in one of the two lines, respectively (Figures 4A and 4B). These findings suggest a role of CG4266 in synapse development and morphology.

We then tested activity, seizure susceptibility, and gross neurological and locomotor function in adult flies upon pan-neuronal knockdown. Evaluation of locomotor activity with a monitor system (Trikinetics) and a 12 h day-night rhythm for 4 days did not reveal any alterations in spontaneous activity or sleeping behavior (Figure S6). After inducing a mechanical shock by vortexing the flies for 10 s in the bang sensitivity assay,[38,39] we did not observe a seizure phenotype (Figure S7). However, when tapping flies down in a vial and assessing their innate behavior to walk up immediately in the climbing assay,[39,40] locomotor ability was severely impaired for both RNAi lines (Figure 4C).

Furthermore, we tested whether knockdown of *CG4266* also impairs complex learning and memory behavior by

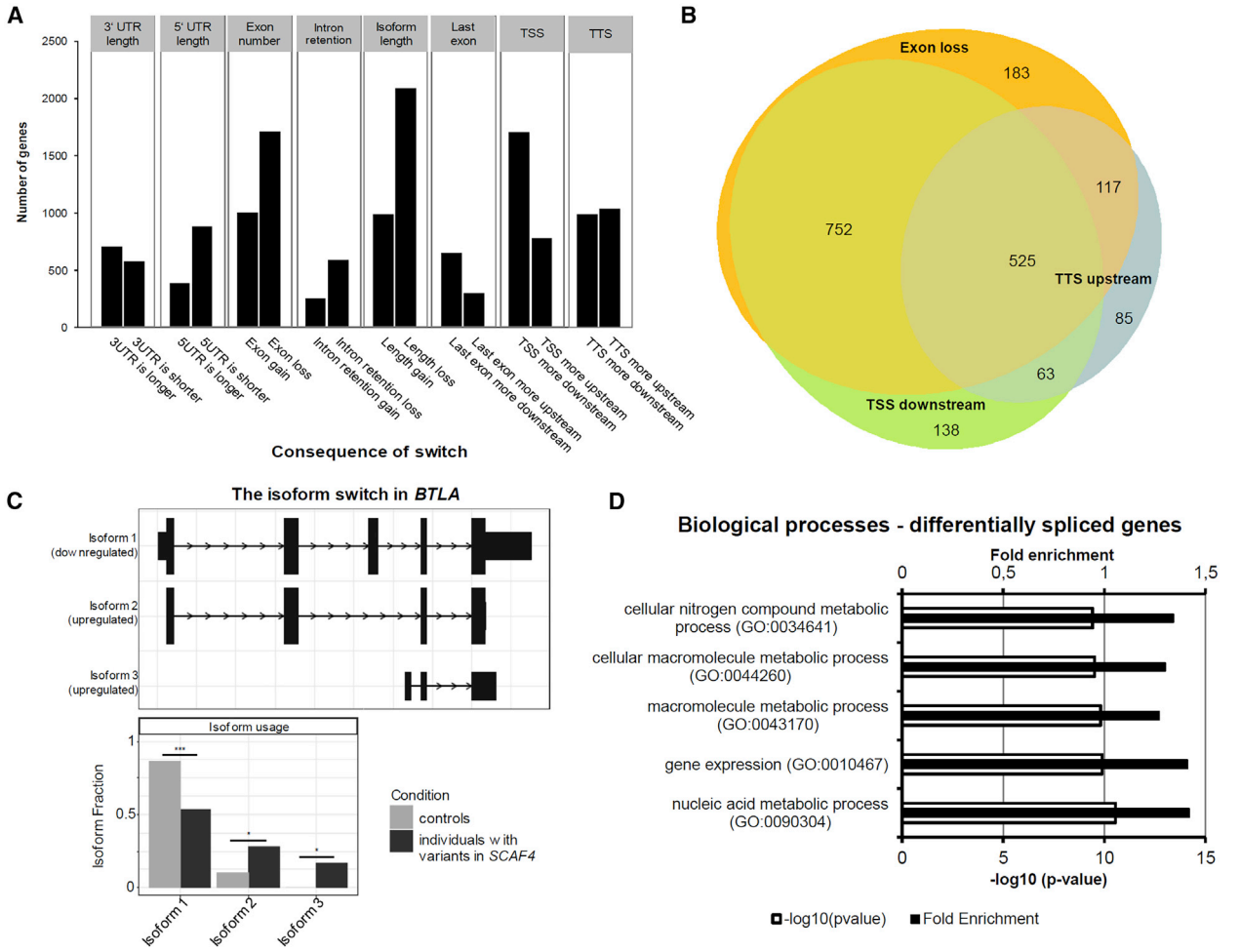


Figure 3. Impact of *SCAF4* Variants on mRNA Processing and Splicing

(A) mRNA isoform expression changes of four individuals with variants in *SCAF4* detected by IsoformSwitchAnalyzeR[30–32] analysis of RNA-sequencing data. Most frequent changes include length loss, exon loss, and usage of an alternative transcription start site (TSS) more downstream.

(B) Venn diagram depicting the most prominent causes for a length loss of isoforms. Truncation results from a combination of exon loss, alternative down-stream transcription start sites (TSS downstream), and alternative up-stream transcription termination sites (TTS up-stream). Venn diagram was created using the eulerr package version 6.0.0 (see Web Resources).

(C) Example of an alternative spliced gene (*BTLA*). All expressed transcripts are shown in the upper panel. The isoform usage is shown in the lower panel. In individuals with variants in *SCAF4*, the two shorter isoforms with downstream TSSs, upstream TTSs, and exon loss showed increased expression while the longest isoform 1 is used less.

(D) Gene Ontology (GO) term analysis of differentially spliced genes depicting the top five GO terms for biological processes.

Abbreviations: TSS, transcription start site; TTS, transcription termination site. Asterisks indicate statistical significance (*$p < 0.05$, ***$p < 0.001$).

utilizing the courtship conditioning paradigm. This assay is based on the reduction of male courtship behavior in response to sexual rejection of a non-receptive pre-mated female,[41] as described in more detail previously[42] and in Supplemental Material and Methods. By adjusting the time between training and testing periods, both learning (test immediately after training) and short-term memory (test 1 h after training) can be assessed. Knockdown of *CG4266* in the mushroom body, the fly center for learning and memory,[43] resulted in a significantly impaired learning performance for both RNAi lines and a significantly impaired short-term memory upon knockdown with RNAi line 1. RNAi line 2 showed the same tendency but without reaching significance (Figure 4D). Also pan-neuronal knockdown of *CG4266* resulted in significant but less consistent learning and memory impairment for the two different RNAi lines (Figure S8).

Taken together, these results suggest that CG4266 plays a crucial role in neurological functioning and complex learning and memory processes. This is consistent with the neurodevelopmental phenotypes and cognitive impairments seen in humans with *SCAF4* loss-of-function variants.

In conclusion, our results suggest that heterozygous truncating/likely gene-disrupting variants in *SCAF4* cause a variable neurodevelopmental disorder and that impaired SCAF4 function results in altered mRNA processing and splicing.

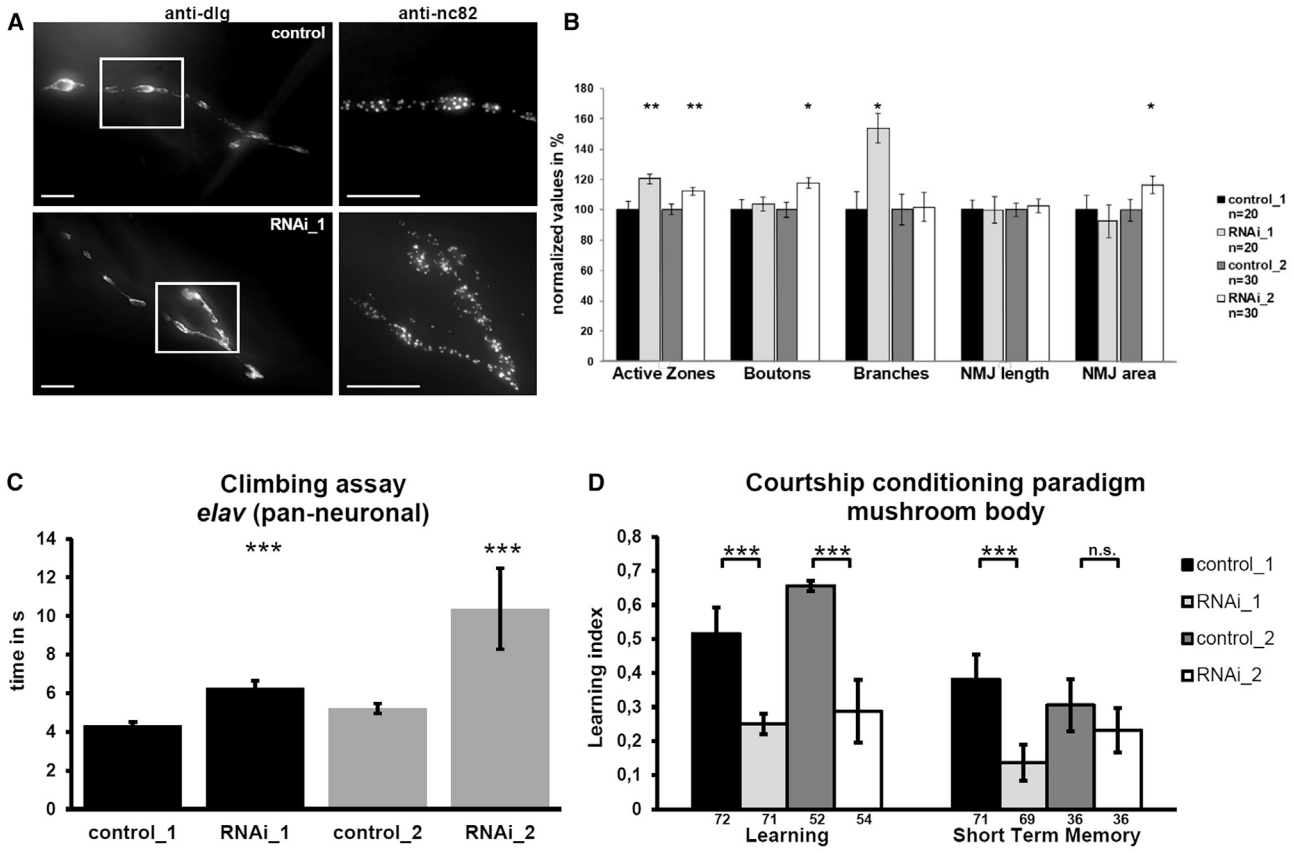


Figure 4. Impact of *CG4266* Knockdown on *Drosophila* Nervous System

(A) Representative pictures of neuromuscular junctions (NMJs) from L3 control (upper panel) and pan-neuronal *CG4266* knockdown (lower panel) larvae. The white-framed box indicates the cutout on the right side. NMJ area and length as well as the number of synaptic boutons, islands, and branches were determined. In the RNAi_1 larva, an increased number of AZs can be observed in comparison to control larva. Scale bars represent 10 μm.

(B) Quantification revealed a significant increase in the number of AZs upon *CG4266* knockdown with two RNAi fly lines using the *elav*-x;*dicer* II driver (BL#25750).

(C) Flies with knockdown of *CG4266* in all neurons (*elav*-Gal4/*Cyo*, BL#8765) showed significant locomotor impairment in the climbing assay, as measured by the amount of time that 70% of flies in a vial needed to crawl up 8.8 cm after being tapped down. Data represent the mean from a minimum of 170 flies tested per genotype.

(D) In the courtship conditioning paradigm, both *CG4266* RNAi lines showed significant impairment of learning upon knockdown with a mushroom-body-specific driver line (UAS-*Dcr*-2;247-Gal4). The short-term memory was reduced significantly in RNAi line 1 with RNAi line 2 showing the same tendency. Graphs display number of animals below the columns per genotype. Differences between learning indices of control and mutant flies were statistically compared by a randomization test with 10,000 bootstrap replicates with a custom R script.[37]

Error bars represent the SEM. Asterisks indicate statistical significance (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

Data and Code Availability

The raw datasets supporting the current study have not been deposited in a public repository because of data safety and privacy regulations.

Consortia

UCLA Clinical Genomics Center: Stanley F. Nelson, Wayne W. Grody, Hane Lee, Joshua L. Deignan, Sung-Hae Kang, Valerie A. Arboleda, T. Niroshi Senaratne, Naghmeh Dorrani, Marina S. Dutra-Clarke, Jessica Kianmahd, Franceska L. Hinkamp, Ahna M. Neustadt, Julian A. Martinez-Agosto, Brent L. Fogel, and Fabiola Quintero-Rivera.

Supplemental Data

Supplemental Data can be found online at https://doi.org/10.1016/j.ajhg.2020.06.019.

Acknowledgment

We thank all individuals and families for participating in this study. We especially thank Laila Distel, Christine Suchy, and Petra Rothe for excellent technical assistance. We thank Christian T. Thiel for the in-house NGS tool and André Reis from the NGS facility at the institute of Human Genetics, Erlangen. Fly stocks were obtained from the Bloomington Drosophila Stock Center (NIH P40OD018537) and the Vienna Drosophila Resource Center. Antibodies from the Developmental Studies Hybridoma Bank (created by the NICHD of the NIH) were used in this study. C.Z. is supported

by grants from the German Research Foundation (DFG) (ZW184/3-1, ZW184/6-1, and 270949263/GRK2162) and by the IZKF Erlangen (E31). G.M.M. is supported by the National Institute of Neurological Disorders and Stroke (NINDS) under award number K08NS092898, Jordan's Guardian Angels, and the Brotman Baty Institute.

Declaration of Interests

The Department of Molecular and Human Genetics at Baylor College of Medicine receives revenue from clinical genetic testing completed at Baylor Genetics Laboratories. J.J. and K.Mc. are employees of GeneDx, Inc. The remaining authors declare no competing interests.

Received: April 17, 2020
Accepted: June 25, 2020
Published: July 29, 2020

Web Resources

eulerr, https://cran.r-project.org/web/packages/eulerr/index.html
GenBank, https://www.ncbi.nlm.nih.gov/genbank/
Gene Ontology, http://geneontology.org/
gnomAD Browser, https://gnomad.broadinstitute.org/
M-CAP, http://bejerano.stanford.edu/mcap/
MutationTaster, http://www.mutationtaster.org/
OMIM, https://www.omim.org/
pheatmaps, https://cran.r-project.org/package=pheatmap
PolyPhen-2, http://genetics.bwh.harvard.edu/pph2/
SIFT, http://sift.bii.a-star.edu.sg/
UCSC Genome Browser, https://genome.ucsc.edu
Vienna Drosophila Research Center, https://stockcenter.vdrc.at/control/main

References

1. Cramer, P. (2004). RNA polymerase II structure: from core to functional complexes. Curr. Opin. Genet. Dev. *14*, 218–226.
2. Orphanides, G., and Reinberg, D. (2002). A unified theory of gene expression. Cell *108*, 439–451.
3. Kornberg, R.D. (1999). Eukaryotic transcriptional control. Trends Cell Biol. *9*, M46–M49.
4. Lee, T.I., and Young, R.A. (2000). Transcription of eukaryotic protein-coding genes. Annu. Rev. Genet. *34*, 77–137.
5. Haijes, H.A., Koster, M.J.E., Rehmann, H., Li, D., Hakonarson, H., Cappuccio, G., Hancarova, M., Lehalle, D., Reardon, W., Schaefer, G.B., et al. (2019). De Novo Heterozygous POLR2A Variants Cause a Neurodevelopmental Syndrome with Profound Infantile-Onset Hypotonia. Am. J. Hum. Genet. *105*, 283–301.
6. Cormack, B.P., and Struhl, K. (1992). The TATA-binding protein is required for transcription by all three nuclear RNA polymerases in yeast cells. Cell *69*, 685–696.
7. Mizzen, C.A., Yang, X.-J., Kokubo, T., Brownell, J.E., Bannister, A.J., Owen-Hughes, T., Workman, J., Wang, L., Berger, S.L., Kouzarides, T., et al. (1996). The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. Cell *87*, 1261–1270.
8. O'Rawe, J.A., Wu, Y., Dörfel, M.J., Rope, A.F., Au, P.Y., Parboosingh, J.S., Moon, S., Kousi, M., Kosma, K., Smith, C.S., et al. (2015). TAF1 Variants Are Associated with Dysmorphic Features, Intellectual Disability, and Neurological Manifestations. Am. J. Hum. Genet. *97*, 922–932.
9. Tawamie, H., Martianov, I., Wohlfahrt, N., Buchert, R., Mengus, G., Uebe, S., Janiri, L., Hirsch, F.W., Schumacher, J., Ferrazzi, F., et al. (2017). Hypomorphic Pathogenic Variants in TAF13 Are Associated with Autosomal-Recessive Intellectual Disability and Microcephaly. Am. J. Hum. Genet. *100*, 555–561.
10. Oegema, R., Baillat, D., Schot, R., van Unen, L.M., Brooks, A., Kia, S.K., Hoogeboom, A.J.M., Xia, Z., Li, W., Cesaroni, M., et al. (2017). Human mutations in integrator complex subunits link transcriptome integrity to brain development. PLoS Genet. *13*, e1006809.
11. Hauer, N.N., Popp, B., Schoeller, E., Schuhmann, S., Heath, K.E., Hisado-Oliva, A., Klinger, P., Kraus, C., Trautmann, U., Zenker, M., et al. (2018). Clinical relevance of systematic phenotyping and exome sequencing in patients with short stature. Genet. Med. *20*, 630–638.
12. Cunningham, F., Achuthan, P., Akanni, W., Allen, J., Amode, M.R., Armean, I.M., Bennett, R., Bhai, J., Billis, K., Boddu, S., et al. (2019). Ensembl 2019. Nucleic Acids Res. *47* (D1), D745–D751.
13. Yuryev, A., Patturajan, M., Litingtung, Y., Joshi, R.V., Gentile, C., Gebara, M., and Corden, J.L. (1996). The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. Proc. Natl. Acad. Sci. USA *93*, 6975–6980.
14. Gregersen, L.H., Mitter, R., Ugalde, A.P., Nojima, T., Proudfoot, N.J., Agami, R., Stewart, A., and Svejstrup, J.Q. (2019). SCAF4 and SCAF8, mRNA Anti-Terminator Proteins. Cell *177*, 1797–1813.e18.
15. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2019). Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv. https://doi.org/10.1101/531210.
16. Sobreira, N., Schiettecatte, F., Valle, D., and Hamosh, A. (2015). GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum. Mutat. *36*, 928–930.
17. Firth, H.V., Richards, S.M., Bevan, A.P., Clayton, S., Corpas, M., Rajan, D., Van Vooren, S., Moreau, Y., Pettett, R.M., and Carter, N.P. (2009). DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am. J. Hum. Genet. *84*, 524–533.
18. Deciphering Developmental Disorders Study (2017). Prevalence and architecture of de novo mutations in developmental disorders. Nature *542*, 433–438.
19. Schwarz, J.M., Rödelsperger, C., Schuelke, M., and Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. Nat. Methods *7*, 575–576.
20. Adzhubei, I., Jordan, D.M., and Sunyaev, S.R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. Curr. Protocols Human Genet. *Chapter 7*, Unit7.20-Unit27.20.
21. 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 Res. *40*, W452-7.
22. 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. Nat. Genet. *48*, 1581–1586.
23. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.
24. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal *17*, 10–12.
25. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods *14*, 417–419.
26. Fabregat, A., Sidiropoulos, K., Viteri, G., Forner, O., Marin-Garcia, P., Arnau, V., D'Eustachio, P., Stein, L., and Hermjakob, H. (2017). Reactome pathway analysis: a high-performance in-memory approach. BMC Bioinformatics *18*, 142–142.
27. Mi, H., Muruganujan, A., Ebert, D., Huang, X., and Thomas, P.D. (2019). PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res. *47* (D1), D419–D426.
28. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al.; The Gene Ontology Consortium (2000). Gene ontology: tool for the unification of biology. Nat. Genet. *25*, 25–29.
29. The Gene Ontology Consortium (2017). Expansion of the Gene Ontology knowledgebase and resources. Nucleic Acids Res. *45* (D1), D331–D338.
30. Soneson, C., Love, M., and Robinson, M. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res. *4*, 1521.
31. 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 Res. *43*, e47–e47.
32. Vitting-Seerup, K., and Sandelin, A. (2017). The Landscape of Isoform Switches in Human Cancers. Mol. Cancer Res. *15*, 1206–1220.
33. Perkins, L.A., Holderbaum, L., Tao, R., Hu, Y., Sopko, R., McCall, K., Yang-Zhou, D., Flockhart, I., Binari, R., Shim, H.S., et al. (2015). The Transgenic RNAi Project at Harvard Medical School: Resources and Validation. Genetics *201*, 843–852.
34. Brand, A.H., and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development *118*, 401–415.
35. Jan, L.Y., and Jan, Y.N. (1976). Properties of the larval neuromuscular junction in Drosophila melanogaster. J. Physiol. *262*, 189–214.
36. Ehmann, N., Owald, D., and Kittel, R.J. (2018). Drosophila active zones: From molecules to behaviour. Neurosci. Res. *127*, 14–24.
37. 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.
38. Kuebler, D., and Tanouye, M.A. (2000). Modifications of seizure susceptibility in Drosophila. J. Neurophysiol. *83*, 998–1009.
39. 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.; Deciphering Developmental Disorders Study (2018). Missense Variants in RHOBTB2 Cause a Developmental and Epileptic Encephalopathy in Humans, and Altered Levels Cause Neurological Defects in Drosophila. Am. J. Hum. Genet. *102*, 44–57.
40. 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.
41. Siegel, R.W., and Hall, J.C. (1979). Conditioned responses in courtship behavior of normal and mutant Drosophila. Proc. Natl. Acad. Sci. USA *76*, 3430–3434.
42. Konrad, E.D.H., Nardini, N., Caliebe, A., Nagel, I., Young, D., Horvath, G., Santoro, S.L., Shuss, C., Ziegler, A., Bonneau, D., et al.; DDD Study (2019). CTCF variants in 39 individuals with a variable neurodevelopmental disorder broaden the mutational and clinical spectrum. Genet. Med. *21*, 2723–2733.
43. Zars, T. (2000). Behavioral functions of the insect mushroom bodies. Curr. Opin. Neurobiol. *10*, 790–795.

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

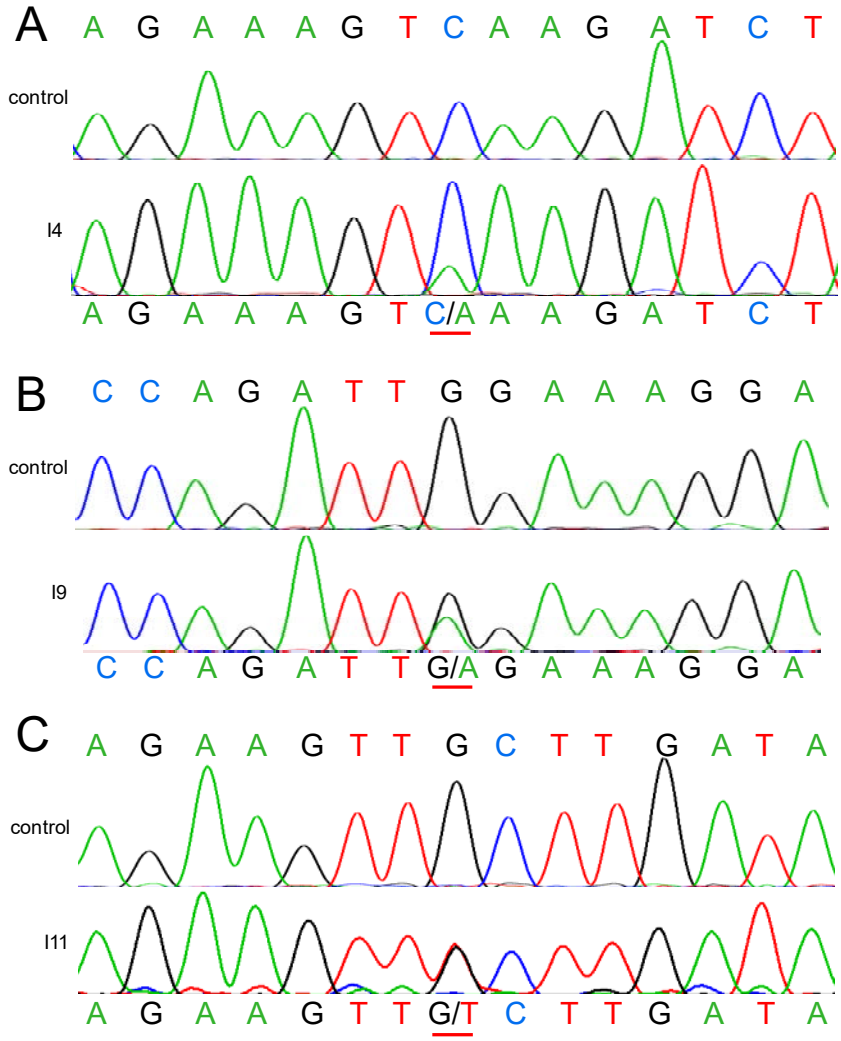


Figure S1. Sanger sequencing of cDNA from affected individuals

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.

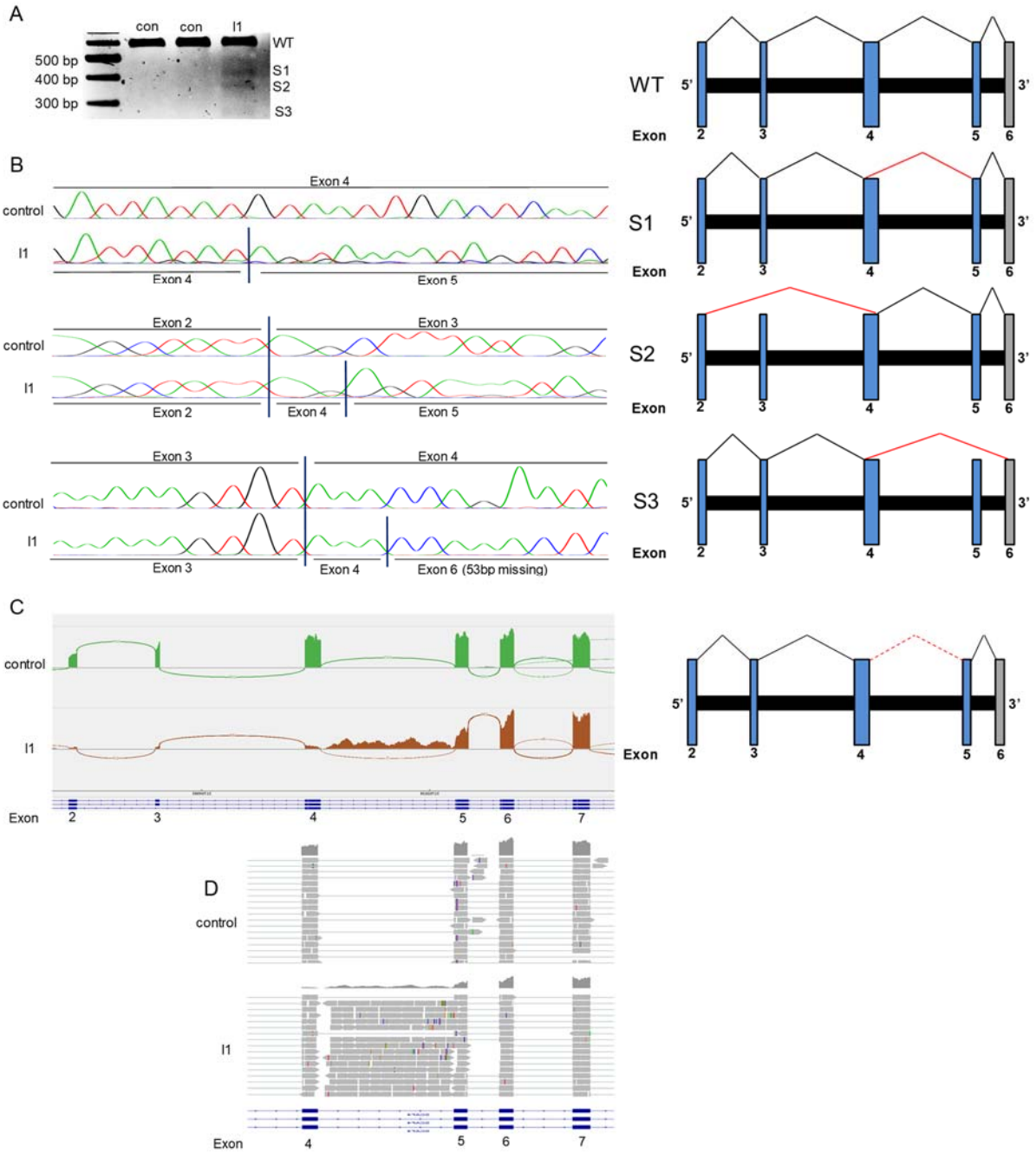
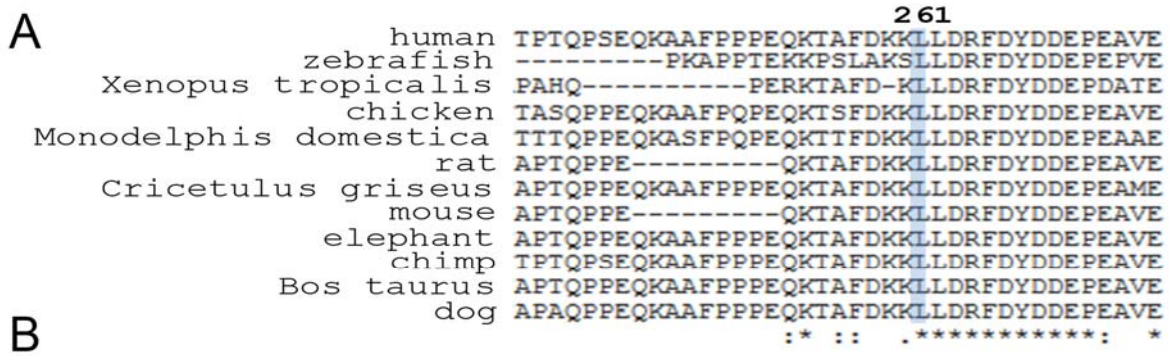


Figure S2. Analysis of the splice-site variant c.321+1G>T

(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.

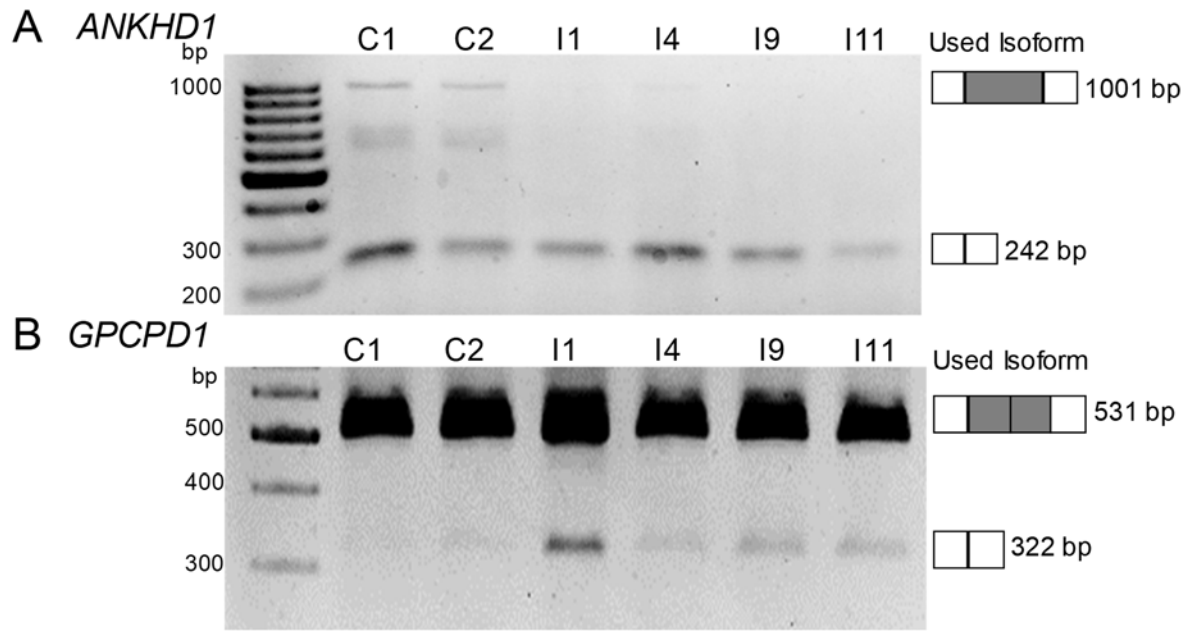


Figure S4. Validation of splicing analysis

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.

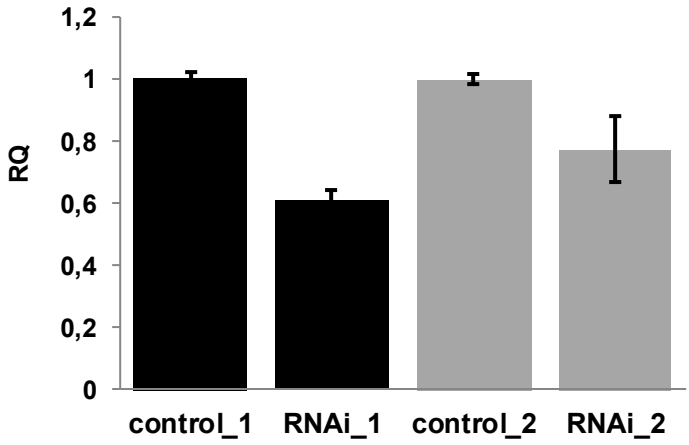


Figure S5. Confirmation of knockdown of *CG4266* in *Drosophila melanogaster*

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.

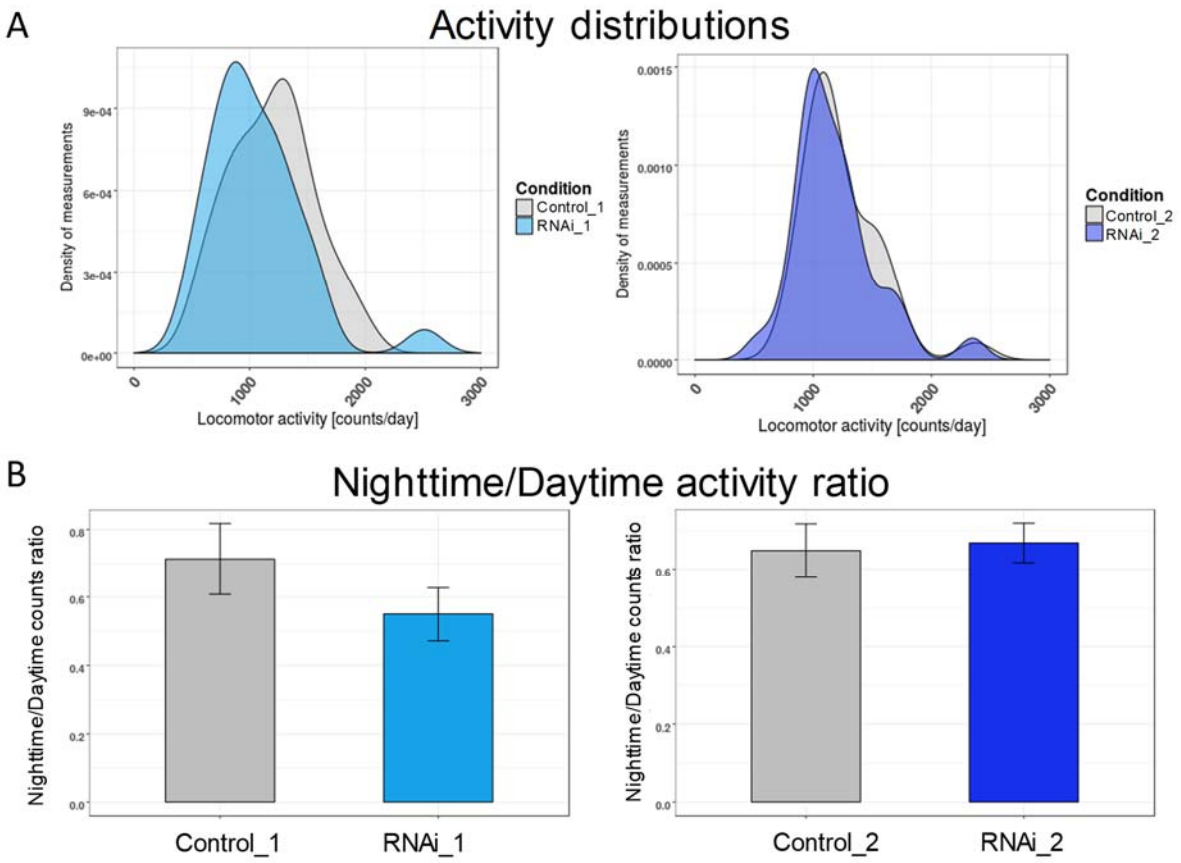


Figure S6. Normal locomotor activity of flies with pan-neural *CG4266* knockdown

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.

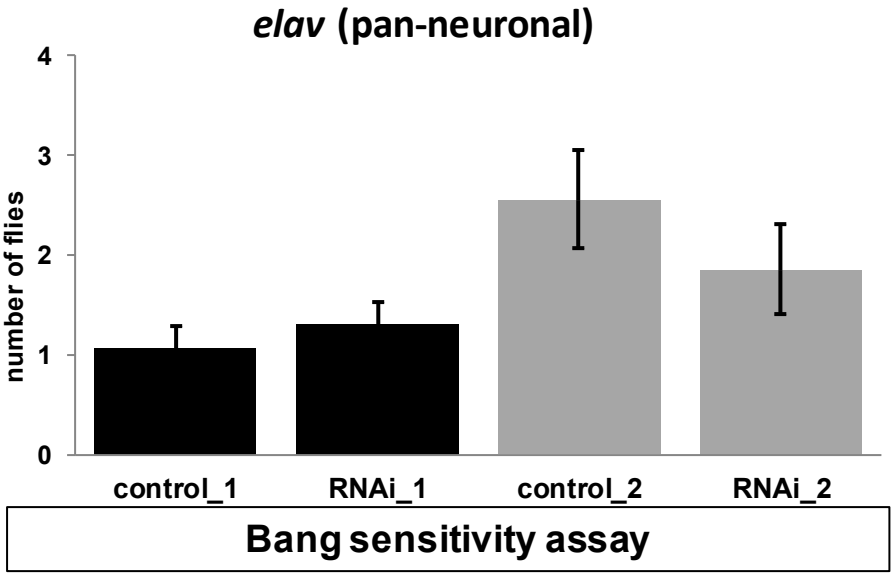


Figure S7. Bang sensitivity assay of flies with *CG4266* knockdown

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.

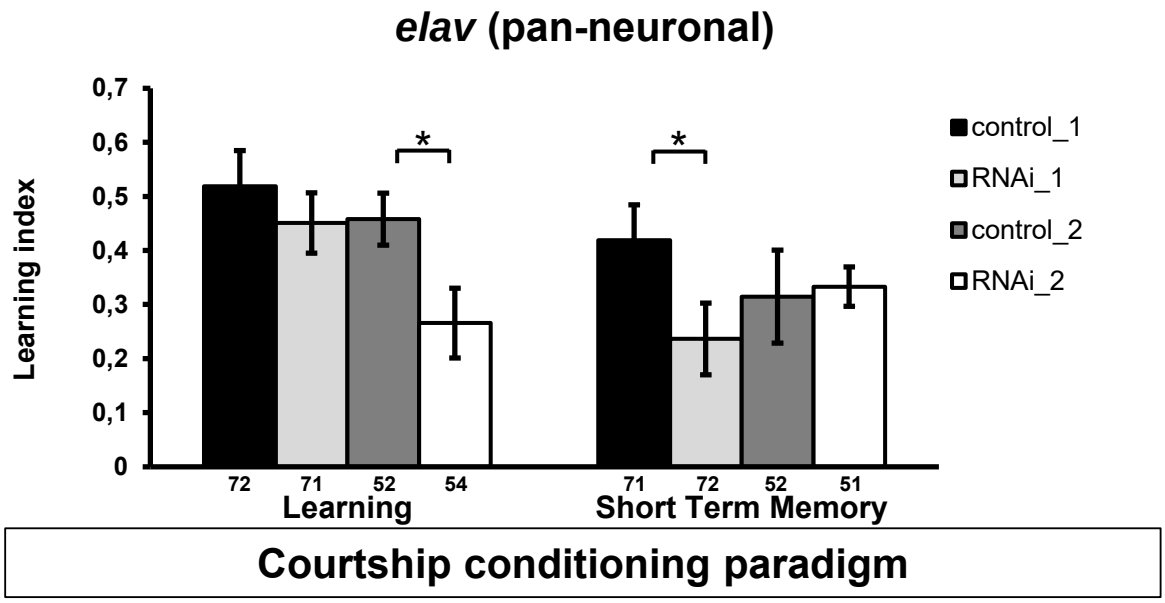


Figure S8. Courtship conditioning paradigm of flies with pan-neuronal *CG4266* knockdown

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.