Loss-of-function in *IRF2BPL* is associated with neurological phenotypes

Paul C. Marcogliese,[1,24] Vandana Shashi,[2,24] Rebecca C. Spillmann,[2] Nicholas Stong,[3] Jill A. Rosenfeld,[1] Mary Kay Koenig,[4] Julián A. Martínez-Agosto,[5,6,7] Matthew Herzog,[5] Agnes H. Chen,[6,8] Patricia I. Dickson,[8] Henry J. Lin,[8] Moin U. Vera,[8] Noriko Salamon,[9] Damara Ortiz,[10] Elena Infante,[10] Wouter Steyaert,[11] Bart Dermaut,[11] Bruce Poppe,[11] Hyung-Lok Chung,[1] Zhongyuan Zuo,[1] Pei-Tseng Lee,[1] Oguz Kanca,[1] Fan Xia,[1] Yaping Yang,[1] Edward C. Smith,[12] Joan Jasien,[12] Sujay Kansagra,[12] Gail Spiridigliozzi,[13] Mays El-Dairi,[14] Robert Lark,[15] Kacie Riley,[2] Dwight D. Koeberl,[2] Katie Golden-Grant,[16] Program for Undiagnosed Diseases (UD-PrOZA), Undiagnosed Diseases Network, Shinya Yamamoto,[1,17,18,19] Michael F. Wangler,[1,17,18] Ghayda Mirzaa,[20,21] Dimitri Hemelsoet,[22] Brendan Lee,[1] Stanley F. Nelson,[5] David B. Goldstein,[3] Hugo J. Bellen,[1,17,18,19,23]* Loren D.M. Pena[2]**

[1] Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, 77030, USA

[2] Division of Medical Genetics, Department of Pediatrics, Duke Health, Durham, NC, 27710, USA

[3] Institute for Genomic Medicine, Columbia University Medical Center, New York, NY, 10032, USA

[4] Division of Child & Adolescent Neurology, Department of Pediatrics, The University of Texas Health Science Center at Houston, Houston, TX, 77030, USA

[5] Department of Human Genetics, David Geffen School of Medicine, University of California-Los Angeles, Los Angeles, CA, 90095, USA

[6] Department of Pediatrics, David Geffen School of Medicine, University of California-Los Angeles, Los Angeles, CA, 90095, USA

[7] Department of Child and Adolescent Psychiatry, Resnick Neuropsychiatric Hospital, University of California-Los Angeles, Los Angeles, CA, 90095, USA.

[8] Division of Pediatric Neurology, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, 90502, USA

[9]Department of Radiology, David Geffen School of Medicine, University of California-Los Angeles, Los Angeles, CA, 90095 USA

[10] Children's Hospital of Pittsburgh of UPMC, University of Pittsburgh, Pittsburgh, PA, 15224, USA

[11] Department of Medical Genetics, Ghent University Hospital, Ghent, Belgium

[12] Division of Neurology, Department of Pediatrics, Duke Health, Durham, NC, 27710, USA

[13] Department of Psychiatry and Behavioral Sciences, Duke Health, Durham, NC, 27710, USA

[14] Department of Ophthalmology, Duke Health, Durham, NC, 27710, USA

[15] Department of Orthopedic Surgery, Duke Health, Durham, NC, 27710, USA

[16] Division of Genetic Medicine, Seattle Children's Hospital, Seattle, WA, 98105, USA

[17] Program in Developmental Biology, Baylor College of Medicine, Houston, TX, 77030, USA

[18] Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX, 77030, USA

[19] Department of Neuroscience, Baylor College of Medicine, Houston, TX, 77030, USA

[20] Center for Integrative Brain Research, Seattle Children's Research Institute, Seattle, WA, 98105, USA

[21] Department of Pediatrics, University of Washington, Seattle, WA, 98105, USA;

[22] Department of Neurology, Ghent University Hospital, Ghent, Belgium

[23] Howard Hughes Medical Institute, Baylor College of Medicine, Houston, TX, 77030, USA

[24] These authors contributed equally to this work

* Corresponding author: hbellen@bcm.edu

** Corresponding author: loren.pena@duke.edu

Key words: hypotonia, developmental regression, ataxia, seizures, *Drosophila*, *pits*, *CG11138*, neurodegeneration, C3HC4 Ring finger, EAP1

Abstract

The *Interferon Regulatory Factor 2 Binding Protein Like* (*IRF2BPL*) gene encodes a member of the IRF2BP family of transcriptional regulators. Currently the biological function of this gene is obscure, and the gene has not been associated with a Mendelian disease. Here we describe seven individuals affected with neurological symptoms who carry damaging heterozygous variants in *IRF2BPL.* Five cases carrying nonsense variants in *IRF2BPL* resulting in a premature stop codon display severe neurodevelopmental regression, hypotonia, progressive ataxia, seizures, and a lack of coordination. Two additional individuals, both with missense variants, display global developmental delay and seizures and a relatively milder phenotype than those with nonsense alleles. The bioinformatics signature for *IRF2BPL* based on population genomics is consistent with a gene that is intolerant to variation. We show that the *IRF2BPL* ortholog in the fruit fly, called *pits* (*protein interacting with Ttk69 and Sin3A*), is broadly expressed including the nervous system. Complete loss of *pits* is lethal early in development, whereas partial knock-down with RNA interference in neurons leads to neurodegeneration, revealing requirement for this gene in proper neuronal function and maintenance. The nonsense variants in *IRF2BPL* identified in patients behave as severe loss-of-function alleles in this model organism, while ectopic expression of the missense variants leads to a range of phenotypes. Taken together, *IRF2BPL* and *pits* are required in the nervous system in humans and flies, and their loss leads to a range of neurological phenotypes in both species.

Introduction

The etiology of neurodevelopmental disorders can vary and can include prenatal exposures, maternal disease, multifactorial causes and single genes. *De novo* genomic contributions were recently highlighted in a large cohort studied as part of the Deciphering Developmental Disorders Study, with an estimate that 42% of the cohort carried pathogenic *de novo* mutations in the coding region of genes.[1-3] The phenotypic and genomic heterogeneity of neurodevelopmental disorders can pose a diagnostic challenge. Several known Mendelian disorders such as Rett syndrome (MIM: 312750), the neuronal ceroid lipofuscinoses (NCLs), and X-linked adrenoleukodystrophy (X-ALD, MIM: 300100) display neurodevelopmental regression as a common element and illustrate the range of symptoms and pathologies associated with neurological symptoms. The known or proposed mechanisms of disease can include altered transcriptional control (Rett syndrome)[4,5], accumulation of a substrate with loss of neurons (juvenile infantile NCL, MIM: 204200)[6,7], or inflammation and demyelination (X-ALD).[8-10] More recently, new genes associated with severe developmental phenotypes and neurodegeneration have been discovered.[11,12] This has been largely possible due to the use of high-throughput sequencing methods such as next-generation sequencing (NGS), in conjunction with sequencing databases for control cohorts such as ExAC and gnomAD[13-21], variant prediction[22], model organism information (e.g. MARRVEL.org)[23] and crowdsourcing programs to identify additional cases, such as GeneMatcher.org.[24] These tools have greatly promoted gene discovery and assisted in ascertaining the role of the candidate variants for disease. Programs such as the Undiagnosed Diseases Network (UDN)[25,26], promote multi-site collaboration that combines NGS and functional data that facilitate the diagnosis of rare disorders.

Here we describe a cohort of seven cases, ascertained for neurological symptoms, who share predicted pathogenic variants in *IRF2BPL*, an intronless gene at 14q24.23.[27] The transcript is expressed in many organs, including in the central nervous system (CNS)

components such as the cerebellum (GTex, accessed 1/29/2018).[28] The IRF2BPL protein and its two mammalian paralogs, IRF2BP1 and IRF2BP2, share two highly conserved domains. These include a coiled-coil DNA binding domain (IRF2BP zinc finger domain) at the amino-terminus and a C3HC4-type RING finger domain at the carboxy-terminus. IRF2BPL also contains polyglutamine and polyalanine tracts. In between the two conserved domains is a variable region that contains a potential nuclear targeting signal.[27] IRF2BPL also contains several putative PEST (proline, glutamic acid, serine, and threonine-rich) sequences throughout the protein suggesting that the IRF2BPL protein is post-translationally regulated[29] (Figure 1A).

IRF2BPL has been proposed to have a role in the initiation of puberty in female non-human primates and rodents.[30] It acts as a transcriptional activator for Gonadotropin Releasing Hormone in the CNS.[30] Expression of *Irf2bpl* in the hypothalamus of female rats increases during puberty and site-specific reduction of *Irf2bpl* in the preoptic area disrupts the estrus cycle.[31] In addition, IRF2BPL has also recently been proposed to function as an E3 ubiquitin ligase that targets β-catenin for proteasome degradation in gastric cancer.[32] Despite these studies, the *in vivo* function of IRF2BPL in all species remains largely undefined. Here we describe a novel role for *IRF2BPL* in the functional and structural maintenance of the nervous system. We provide evidence in seven patients and use functional assays in fruit flies to support the findings that these variants cause dramatic changes to IRF2BPL function and that IRF2BPL plays a role in both development and neuronal maintenance.

Subjects and Methods

Demographics, ascertainment, and diagnoses (Table 1 and supplementary table 1). The individuals are all unrelated and between 2.5 and 43 years of age, including one patient who died at 15 years of age. Four are male, six are Caucasian, and one identified as Hispanic. They were evaluated by the genetics or neurology service as part of clinical care for neurological symptoms (subjects 2, 4, 5, 6, 7) or as participants (subjects 1 and 3) in the UDN.[26]

In non-UDN cases, use of the GeneMatcher website[24] facilitated ascertainment of two of the subjects (6 and 7); the UDN page for *IRF2BPL* facilitated contact with subject 5, and subjects 2 and 4 were ascertained through communication with the clinical lab that performed whole exome sequencing (WES). WES was performed as a trio in five subjects (1, 3, 5, 6, and 7) and proband-only in two (2 and 4). WES was performed with written informed consent for clinical sequencing and in accordance with institutional review board procedures for research sequencing for all subjects. Consent for publication and images were obtained from all guardians.

The seven cases reported here have a constellation of neurological findings of variable severity. While almost all suffered from epilepsy [Human Phenotype Ontology[33] term (HP: 0001250)], those with nonsense variants in *IRF2BPL* had severe, progressive neurodevelopmental regression (HP: 0002376), dysarthria (HP: 0001260), spasticity (HP: 0001257), and symptoms of movement disorders (HP: 0100022). There was also cerebral volume loss in the two oldest cases (3 and 5; HP: 0002506), and cerebellar volume loss in case 5 (HP:0007360). In contrast, the two cases with missense variants have a generally milder course, with symptoms of epilepsy, speech delay (HP: 0000750), and hypotonia (HP: 0009062). Each proband is described below, along with additional details in Table 1 and Supplementary Table 1.

<u>Subject #1</u>

This is a 7-year-old male who met early gross and fine developmental milestones in the first two years of life (Supplementary Figure 1). He had a minor speech delay, as he had his first word at 15 months of age. His development was normal until 2.5 years, when clumsiness and excessive falling were noted. At the age of 4 years, the patient was unable to walk without someone holding his hands or using a harness or a walker (Supplemental Movie 1). During an evaluation at 7 years, the patient had lost the ability to walk, was not able to grasp objects, was not able to feed himself and had no expressive language. He had progressive loss of gross and

fine motor function, speech and self-help skills. Additionally, he had a diagnosis of dystonia, choreoathetosis, and variable spasticity that primarily affected his lower extremities.

Previous evaluations included a normal brain magnetic resonance imaging (MRI), audiology, and electromyography (EMG) at 5 years of age. He developed esotropia at the age of 4 years that progressively became of a larger angle. Ophthalmic examination at the age of 5 years was notable for visual acuity better than 20/70 in either eye with full fields to distraction. There was evidence of large angle esotropia (higher than 65) and bilateral facial palsies with ophthalmoplegia sparing the eyelids and the pupils. There was a severe horizontal gaze palsy but vertical gaze was also limited. Saccades were decreased in both eyes. Funduscopic examination was normal and he had a normal electroretinogram with normal optical coherence tomography scan of the retina and the retinal nerve fiber layer (RNFL) (measured 85 µm on the RNFL, which is normal for age).

While his neurodevelopmental course demonstrated significant developmental regression, his cognition was tested with the Peabody Picture Vocabulary Test, Fourth Edition, at 6 years of age. He scored in the average range for cognition, consistent with a previous score on the Verbal Ability composite of the Differential Ability Scales- Second Edition, administered when he was 3 years, 9 months of age. Although he did not have clinical seizures at the time, an electroencephalogram (EEG) at 6 years of age showed diffuse slowing that was consistent with generalized brain dysfunction and interictal discharges from the left occipital and right temporal leads, indicating a possible area of epileptogenic potential. At almost 7 years of age, he had an episode of abnormal movements resembling myoclonus and was diagnosed with a seizure disorder, which was controlled with levetiracetam.

Prior biochemical, cytogenetic and molecular evaluations are summarized in Supplementary Table 2. Reanalysis of previously obtained research WES data (trio) at Columbia University identified heterozygosity for a *de novo* truncating variant in *IRF2BPL*, NM_024496.3, c.514G>T (p.E172X) that was Sanger confirmed. Additionally, he had a *de*

novo missense variant downstream and in *cis* to the nonsense variant in *IRF2BPL*, NM_024496.3, c.584G>T (p.G195V).

Subject #2

The patient is a female who met all of her early gross developmental and speech milestones. At 5-6 years of age her family began to notice progressive gait disturbance. Her initial diagnosis by the neurology service was dystonia, and she was given a trial of carbidopa-levodopa (Sinemet) with no improvement. Symptoms progressed, and by 8 years of age she had notable lower extremity spasticity with dysarthria, drooling, dysphagia, and incontinence. Examination at 9 years of age demonstrated an interactive girl with intact cognition, decreased speech fluency and severe dysarthria. Coordination exam demonstrated past pointing. Gait was wide-based with spastic diplegia and scissoring. She had an intrathecal baclofen pump placed at 9 years of age with no improvement. She had lost the ability to stand alone or walk by 10 years and was unable to use her hands or speak by 11 years. She had a dysconjugate gaze at 11 years. MRI spine studies were initially normal but at 10 years demonstrated marked cord thinning. MRI of the brain demonstrated mild cerebellar volume loss at 8 years. The cerebellum was small and the corpus callosum was ‘bulky’. Her respiratory function became compromised and ventilation by continuous positive airway pressure was recommended at 10 years. A G-tube was placed at 10 years for dysphagia and aspiration. During her last formal examination at 13 years of age, her cognition was difficult to assess, as she was no longer able to communicate. All symptoms continued to progress until her family chose to limit her medical interventions for feeding intolerance, and she died at 15 years of age. A nonsense variant in *IRF2BPL* was detected on whole exome sequencing (WES), NM_024496.3, c.562C>T (p.R188X), confirmed by Sanger sequencing.

Subject #3

This is a 20-year-old male who met gross and fine motor milestones early in life. Overall development through the first several years was felt to be normal, although some balance

problems with walking may have occurred around 2 years. However, he started stumbling and falling at school at approximately 5-6 years. He developed ataxia and began drooling. A neurological evaluation at age 7 years found nystagmus, hyperactive reflexes, bilateral Babinski signs, and dysmetria. Worse leg weakness and more falls were reported at 8 years. By age 10 his symptoms had progressively worsened, and he had begun experiencing 1-3 myoclonic jerks daily, feeding difficulties, facial weakness, and had lost his speech and ability to walk. By age 13 he was G-tube dependent (from worsening dysphagia) and had slowed saccadic eye movements and severe contractures of the hands and feet. He could not hold objects. He is wheelchair dependent with extremely limited range of motion of all joints. Spastic quadriplegia is managed with a baclofen pump. He has a roving gaze and no visual attention.

Previous evaluations have included MRI of the brain that was normal at age 7 years, but had diffuse cerebral atrophy with *ex vacuo* dilatation of the lateral ventricles but no evidence of white matter disease at 13 years. At 20 years of age, brain MRI demonstrated severe cerebral volume loss of the bilateral hemispheres, patchy periventricular subcortical white matter signal hyperintensity on fluid-attenuated inversion recovery, and marked thinning of the corpus callosum (Figure 2A). Neuropsychological evaluation with the Vineland Adaptive Behavior Scales (3rd edition) at age 20 years indicated continued global and significant delays across all aspects of the patient's functioning. A *de novo* nonsense variant in *IRF2BPL* was identified on WES, NM_024496.3, c.562C>T (p.R188X) and Sanger confirmed.

<u>Subject #4</u>

This is a 16-year-old female with global developmental, hypotonia, and speech delay. She had prenatal exposure to alcohol and controlled substances. She was diagnosed with subclinical epilepsy, due to staring episodes and abnormal EEGs, wide-based gait with high-arched feet at 6 years. Cognitive testing was completed at this age (Bracken Basics Concepts III Receptive and School Readiness Composite; Conner's Parent Rating Scale;

Child Behavior Checklist) and resulted in a diagnosis of ADHD and intellectual functioning at approximately the 1-3 year old level. Since 12 years old, there has been a decline in gross and fine motor abilities, speech and self-help skills with a recent diagnosis of catatonia and lower extremity dystonia in the past year. This was initially attributed to sexual abuse, but has been persistent since she was removed from the abusive environment. She had a normal brain MRI at 6 and 13 years of age. At 15, a thin corpus callosum, still within normal limits, was observed on brain MRI. A swallow study at 15 years of age revealed silent aspiration with all tested consistencies. She now is gastrostomy tube dependent, and requires assistance with all activities of daily living. She is largely wheelchair dependent, though able to walk with assistance for short periods. She has progressive macular degeneration that was stable upon her latest examination at 15 years.

Biochemical and cytogenetic testing were completed and are summarized in Supplementary Table 2. She had exome sequencing in 2016 that revealed a mutation in *BEST1* that explains her retinal symptoms (Macular dystrophy, vitelliform, 2, VMD2 [MIM: 153700]). A truncating variant in *IRF2BPL*, NM_024496.3, c.379C>T (p.Q127X) was also reported and was Sanger confirmed.

<u>Subject #5</u>

This is a 43-year-old male who had psychomotor retardation with hypotonia. He had febrile seizures between 2 and 4 years of age, was diagnosed with epilepsy at 10 years, and developed refractory myoclonic epilepsy by 12 years of age. His neurodevelopmental course showed regression, and at the age of 15 years he had ataxia, spastic rigidity, dystonia and dyskinesia, and a diagnosis of spastic-athetoid cerebral palsy. Progressive cognitive deterioration was noticed, and communication became difficult. At 28 years, he could barely walk independently and could not feed himself. Together with the refractory seizures, the progressive motor problems quickly led him to be wheelchair-bound. At 29 years, he had a vagal nerve stimulator placed to control his seizures, and an intrathecal baclofen pump was

placed at 30 years to treat his spasticity. Dystonic attacks involving the axial and appendicular musculature became more frequent. He was treated for recurrent aspiration pneumonia and spontaneous pneumothorax. At the age of 35, his epilepsy was under control, but his dystonia became generalized (limbs, axial, facial with tongue protrusion) and worsened gradually, leading to a completely bedridden state. Treatment with botulin toxin gave only limited benefit. At 42 he had a right-sided pallidotomy, also with limited and transient beneficial effect. Brain MRI at 34 showed global atrophy (cerebral, cerebellar and brainstem), thinning of corpus callosum, and no white matter or cortical lesions (Figure 2B). MR-spectroscopy showed normal findings.

Biochemical, cytogenetic and molecular analyses are summarized in Supplementary Table 2. Trio WES at Ghent University Hospital revealed a *de novo* heterozygous nonsense mutation (NM_024496.3, c.376C>T (p.Q126X)) in *IRF2BPL*, resulting in a premature stop codon, which was Sanger confirmed.

<u>Subject #6</u>

This is an 11-year-old male with gross motor and speech delays. He rolled over at 14 months and walked at 21 months. At approximately 14 months he received a diagnosis of generalized myoclonic epilepsy that was difficult to control. He had complete regression of speech at 2 years and remains nonverbal, though his receptive language skills appear to be appropriate. He was diagnosed with autism spectrum disorder (ASD) at 3.5 years. He has not had additional regression in developmental skills.

Evaluations are summarized in Supplementary Table 2 and include a normal brain MRI at 7 years of age. Trio WES detected a *de novo* missense variant in the candidate gene *IRF2BPL*, NM_024496.3, c.1115C>G (p.P372R) that was confirmed by Sanger sequencing.

<u>Subject #7</u>

The patient is a female who was diagnosed with infantile spasms at age 6 months. Seizures subsided with combination antiepileptics at approximately 9 months. Anti-seizure

medications were discontinued thereafter with no recurrence of episodes. Brain MRI and MR spectroscopy were both normal. She met gross motor milestones early in life, though on last assessment at 17 months, she continued to be non-verbal and hypotonic. Providers noted deceleration of developmental progress with onset of infantile spasms at 6 months of age, though the patient has not had frank regression. Facial examination revealed subtle dysmorphic facial features including epicanthal folds, mild telecanthus, and almond-shaped eyes with round facies. Developmental assessment using the Bayley Scales of Infant and Toddler Development, 3rd edition (Bayley-3) at 15 months revealed cognitive, expressive and receptive language, and fine and gross motor functioning below levels expected for age (< 1st percentile). There was no family history of seizures or developmental delay. A chromosomal microarray revealed a 400 kb interstitial duplication at 7q31.31 of uncertain clinical significance. Additional tests are listed in Supplementary Table 2. A *de novo* missense variant was detected on *IRF2BPL*, NM_024496.3, c.1254G>C, (p.K418N) on trio WES on 42/87 of the reads.

Methods

Exome sequencing: Subjects had WES performed on a clinical or research basis. Exome data are summarized in Supplementary Table 3. Across the performing labs, the minimum average depth of coverage was 100X across assays, and minimum proportion of the target at >10X coverage was 95%.

Exome reanalysis (subject 1): FASTQ files were obtained from the relevant source with parental consent. Alignment and variant calling have been previously described.[34] Novel genotypes were filtered for quality and control observations in public database controls. We highlighted variants in known OMIM genes or mouse essential genes, loss-of-function (LoF) variants that are in genes with known pathogenic LoF variants or reported as haploinsufficient by ClinGen[35],

and LoF intolerant by high probability of LoF intolerance (pLI) score (>0.9).[20] Conservation of the variant site is reported with the Genomic Evolutionary Rate Profiling (GERP) score.[36]

Generation of fly stocks: All fly strains used in this study were generated in house or obtained from the Bloomington Drosophila Stock Center (BDSC) and cultured at room temperature unless otherwise noted. The *pits*$^{MI02926\text{-}TG4.1}$ allele was generated by genetic conversion of the MiMIC[37,38] (Minos Mediated Integration Cassette) insertion line, *y*^{1} *w** *Mi{MIC}CG11138*^{MI02926} via recombination-mediated cassette exchange (RMCE) as described[39-41]. The recessive lethality associated with the *pits*$^{MI02926\text{-}TG4.1}$ allele was rescued using an 80 Kb P[acman] duplication (*w[1118]; Dp(1;3)DC256, PBac{y[+mDint2] w[+mC]=DC256}VK33*)[42] as well as by a 20 kb genomic rescue construct (see below). Expression pattern of *pits* was determined by crossing *pits*$^{MI02926\text{-}TG4.1}$ to *UAS-mCD8-eGFP* (BDSC_32184). In addition, the *y*^{1} *w** *Mi{MIC}CG11138*^{MI02926} line was converted to a protein trap line (*pits*::GFP) by injection of a construct that could either produce a splice acceptor (SA)-T2A-GAL4-polyA mutant allele or a SA-eGFP-splice donor (SD) protein trap allele depending on the inserted direction. The successful generation of the *pits*::GFP allele was confirmed by both PCR of genomic DNA and anti-GFP immunohistochemistry/Western blotting of fly tissue.

All transgenic constructs were generated by Gateway (Thermo Fisher) cloning into the pUASg-HA.attB plasmid.[43] The human *IRF2BPL* cDNA clone[30] was made to match the NM_024496.3 transcript. Flanking Gateway *attB* sites were added by PCR of the template cDNA clone, and then shuttled to the pDONR223 by BP clonase II (Thermo Fisher). Variants were generated by Q5 site directed mutagenesis (NEB), fully sequenced (Sanger) and finally cloned into pUASg-HA.attB via LR clonase II (Thermo Fisher). All expression constructs were inserted into the VK37 (*PBac{y[+]-attP}VK00037*) docking site by ϕC31 mediated transgenesis.[44] The 20Kb genomic rescue (GR) line was generated by inserting the P[acman] clone CH322-141N09 (BACPAC Resources)[45] into the VK37 docking site. The *UAS-pits-RC*

flies were generated by obtaining the *pits* cDNA (RE41430, RC isoform) clone from the Drosophila Genomics Resource Center (DGRC) and performing Gateway (Thermo Fisher) cloning into pUASg-HA.attB. The pits RNAi line (*P{KK108903}VIE-260B*) was obtained from Vienna Drosophila Resource Center (VDRC). The GAL4 lines used in this study from BDSC are: *nSyb-GAL4* (y^1 w^*; *P{*w^{+m*}*=nSyb-GAL4.S}3*) BDSC_51635, *Rh1-GAL4 (P{*$ry^{+t7.2}$*=rh1-GAL4}3,* ry^{506}) BDSC_8691*, Act-GAL4/CyO* (y^1 w^*; *P{*w^{+mC}*=Act5C-GAL4}25FO1/CyO,* y^+) BDSC_ 4414*.*

Drosophila behavioral assays: To perform the bang sensitivity assay[46], flies were anesthetized no sooner than 24 hours prior to testing with CO_2 and housed individually. At the time of testing, flies were transferred to a clean vial without food and vortexed at maximum speed for 15 seconds. The time required for flies to recover in an upright position was measured. Typically 25 flies were tested per data point. The climbing assay[47] was performed in a similar manner but at the time of testing flies were transferred to a fresh vial and given 1 minute to habituate before tapped to the bottom of the vial three times and examined for negative geotaxis (climbing) response to reach the 7 cm mark on the vial. Flies were given a maximum of 30 seconds to reach the top.

Histological and ultrastructural analysis of the fly retina: 45-day-old flies were dissected in ice cold fixative (2% PFA / 2.5% glutaraldehyde / 0.1 M sodium cacodylate). Heads were fixed, dehydrated and embedded in Embed-812 resin (Electron Microscopy Sciences) as described previously.[48] For histological sections (200 nm thick) were stained with toluidine blue and imaged with a Zeiss microscope (Axio Imager-Z2) equipped with an AxioCam MRm digital camera. For transmission electron microscopy (TEM), sections (50 nm thick) were stained with 1% uranyl acetate and 2.5% lead citrate. TEM images were obtained using a transmission

electron microscope (model 1010, JEOL). Images were processed with Photoshop (Adobe) and 45 day old toluidine blue images were color matched to the 5 day old images for clarity.

Overexpression of human *IRF2BPL* in *Drosophila*: We overexpressed reference and variant *IRF2BPL* cDNAs in flies by crossing the *UAS-IRF2BPL* males to virgin female flies from a ubiquitous (*Act-GAL4*) driver stock. The progeny of these crosses were cultured at various temperatures (18°C, 22°C, 25°C, and 29°C) to express the human proteins at different levels (lowest expression at 18°C and highest expression at 29°C).[38,49] Expected Mendelian ratios of flies was assessed. For temperatures ≥22°C over 100 flies were assessed for each cross. For 18°C crosses, over 50 flies were assessed from each group.

For additional methods please see supplemental material.

Results

Summary of Clinical Findings

The subjects described here are between 2.5 and 43 years of age, including one who died at 15 years of age. Comprehensive clinical information is included in Supplementary Table 1 and Supplementary Table 2.

All seven individuals have monoallelic variants in *IRF2BPL* (Table 1 and Supplementary Table 1). The nonsense variants identified are: p.Q126X, p.Q127X, p.E172X, and p.R188X (found in 2 individuals). There was remarkable similarity in the clinical course of the five cases that had nonsense variants. These include an unremarkable course in their initial development, followed by loss of developmental milestones and development of a seizure disorder at a variable age, with 2 years being the youngest. Additional findings included a movement disorder with dystonia and choreoathetosis, and cerebellar signs such as ataxia, dysarthria, dysmetria and dysdiadochokinesia. Three of the probands had had anomalies of eye movements but had normal retinae and optic nerves. Head circumference appeared to remain appropriate for age

in those with available measurements. Brain MRI was normal early in life, but cerebral and cerebellar atrophy was noted in the two oldest cases. Cognitive status was difficult to ascertain, particularly as verbal skills were lost. However, cognitive abilities appeared to remain intact over a short follow up period in one case and to deteriorate in an older individual. One case was deceased at 15 years of age when the family elected to limit care after continued decline. The variants were *de novo* in all for whom parental testing was available (three out of five cases).

Missense variants (p.P372R and p.K418N) were also reported in *IRF2BPL* for two individuals who had a variable phenotype of global developmental delay and seizures. The older case (subject 6) had a diagnosis of ASD. Despite the frequent finding of seizures, antiepileptic drugs were successfully weaned in subject 7, without recurrence over a short follow up period. In both missense cases, the variants were *de novo*. Neither case has developed a progressive loss of milestones, abnormal eye movements, or a movement disorder at 2 or 10 years of age.

***IRF2BPL de novo* variants are deleterious based on bioinformatics data**

IRF2BPL is a gene that is highly intolerant to variation with an Residual Variation Intolerance Score (RVIS)[50] of 9.3% and a pLI score of 0.97, with no observed LoF variants included in the calculation.[20] The only LoF variants present in ExAC and gnomAD are frameshifts, which either do not pass quality filtering or appear to be artifactual calls in repetitive regions using the browser -visualization tool.[20] *IRF2BPL* is also constrained to missense variants (z = 4.73).[20] All variants found in the subjects were absent from the gnomAD and ExAC databases. The Combined Annotation Dependent Depletion (CADD)[51] score for all of the variants was > 34, which places them among the variants predicted to be most deleterious. The four nonsense variants introduce a premature termination codon either downstream (p.E172, p.R188) or at the end (p.Q126, p.Q127) of the poly-glutamine tract and upstream of the first PEST sequence (Figure 1A), whereas the missense variants are within the variable region of

the protein. We used the DOMINO tool to assess the likelihood of monoallelic variants in *IRF2BPL* to cause a Mendelian disease. IRF2BPL had a DOMINO score of 0.962 out of 1 predicting that monoallelic variants would very likely cause disease (Figure 1B).[52]

IRF2BPL variants are severe loss-of-function mutations based on functional assays in flies

To validate the functional consequences of these variants, we utilized *Drosophila melanogaster* as a model organism. Experiments in fruit flies have previously provided experimental support in identifying causal variants for human disease[19,53-55] and these approaches have been an integral part of the UDN. [13,14,16,21,56] The fly ortholog of *IRF2BPL* is a poorly characterized gene, *CG11138*. This gene was studied in the context of epigenetic regulation through biochemical methodologies during embryogenesis, and was named *pits (protein interacting with Ttk69 and Sin3A*).[57] Although the overall identity (30%) and similarity (36%) between IRF2BPL and Pits may not seem high, the architecture of the protein is very similar and the sequences of the annotated domains show high conservation (79% identity for the zinc finger domain, 76% identity for the C3HC4 RING domain) (Figure 3A). The Pits protein also has a DIOPT[58] score of 12/15 suggesting that it is likely to be a true ortholog of IRF2BPL. Two other human paralogs of IRF2BPL share a similar high DIOPT score (12/15 for *IRF2BP1*; 11/15 for *IRF2BP2*). These data indicate that *pits* is the sole fly gene that is orthologous to the three IRF2BP family genes in humans.

To generate a *pits* mutant fly, we used a MiMIC insertion in an intron of *pits*, named *Mi{MIC}CG11138*^{MI02926} (Figure 3B).[37,38] MiMICs are engineered transposable elements that contain inverted *attP* sites derived from phage ΦC31 flanking a swappable cassette. This allows replacement of the content of the *MiMIC* insertion using Recombination-Mediated Cassette Exchange (RMCE) by expressing the ΦC31 integrase to swap the *MiMIC* cassette with a *SA-T2A-GAL4-polyA* (T2A-GAL4) cassette.[39-41] This insertion results in a truncated *pits*

transcript due to the polyA signal. During translation this short transcript produces a short protein that is truncated at the T2A site yet allows re-initiation of translation to produce the GAL4 protein. The GAL4 protein is expressed in a proper spatial and temporal fashion, i.e. those of the endogenous gene[38] (Figure 3B). This allows rescue of the *T2A-GAL4* induced allele, typically a null allele, with a UAS-(fly)cDNA for about 70% of the genes tested.[40]

The *pits* gene is on the X chromosome, and *pits*$^{MI02926\text{-}TG4.1}$ males are hemizygous lethal, as they fail to survive past the first instar larval stage, and most die as embryos. Lethality can be rescued to viable adults by introduction of an 80 Kb or a 20 Kb P[acman] genomic BAC rescue (GR) construct, the latter only carrying the *pits* gene[42,44] (Figure 3C and 3D). Hence, *pits* is an essential gene, and expression of the gene in the proper genomic context fully rescues the LoF of *pits.*

We attempted to rescue lethality observed in *pits*$^{MI02926\text{-}TG4.1}$ flies by overexpression of UAS-*pits* or UAS-*IRF2BPL* and failed to obtain viable flies. Intriguingly, we were also unable to obtain viable heterozygous female flies that contain both the *pits*$^{MI02926\text{-}TG4.1}$ allele and UAS-*pits* or *UAS-IRF2BP* (Figure 3D). These data show that overexpression of the fly or human IRF2BPL in the cells that endogenously express Pits is toxic to the fly. Indeed, expression of the *UAS-IRF2BPL* under the control of a ubiquitous driver (*Act-GAL4*) also causes lethality (Figure 3E).

The above data show that the lethality caused by overexpression of the reference *IRF2BPL* gene can be used as a functional assay to test if the variants are functional (toxic) as well. To modulate the levels of expression we ubiquitously expressed the variants with *Act-GAL4* as the UAS-cDNA expressed using this driver exhibits temperature dependence, with significantly greater expression at higher temperatures.[38,49] We conducted these experiments at 18°C, 22°C, 25°C and 29°C. Expression of the reference *IRF2BPL* cDNA consistently causes lethality at all temperatures tested, whereas the three nonsense variants (p.E172X, p.Q127X, and p.R188X) consistently produced viable animals at all temperatures tested (Figure 3E). This

provides evidence that the truncated proteins are not toxic and are very likely LoF alleles. Interestingly, the missense variant p.K418N is lethal when expressed at higher temperatures, whereas the flies are viable at lower temperatures, indicating that it retains some toxic function and hence behaves as a partial LoF allele. The p.P372R variant still remained lethal upon ubiquitous expression, even at lower temperatures indicating its toxic function is similar to the reference protein at least in this assay. We confirmed that the UAS-driven human cDNA in flies expresses IRF2BPL at relatively similar levels by making HA-tagged constructs for the reference and p.E172X truncation and performing Western blot using an anti-HA antibody, as commercially available antibodies for IRF2BPL recognize epitopes downstream of the premature termination codon, in the C-terminus of IRF2BPL. UAS expression of these constructs in neurons by *nSyb (*neuronal Synaptobrevin)-*GAL4* was relatively similar (Supplemental Figure 2A), Additionally, we confirmed that the untagged reference and two missense variants were expressed at similar levels using a commercial antibody against IRF2BPL and performing Western blot, suggesting that the reduced toxicity of p.K418N is not due to destabilization of the protein (Supplemental Figure 2B).

pits is expressed and required in the CNS

To examine the endogenous levels of *pits* as well as its subcellular localization, we generated a GFP protein trap line of *pits*. We integrated a protein trap (splice acceptor (SA)-linker-eGFP-linker-splice donor (SD)) cassette into the $pits^{MI02926}$ via RMCE. Although the SA-eGFP-SD functions as an artificial exon in the middle of the *pits* gene (Supplemental Figure 3A), *pits::GFP* flies are homozygous viable and do not display any obvious phenotype indicating that the internal GFP tag is not deleterious. This is in concert with our previous findings that most (75%) proteins tolerate the presence of an internal GFP incorporated by the SA-eGFP-SD MiMIC insertions.[38] The GFP fusion proteins also reflect the localization of the endogenous proteins.[37,38] We first confirmed the presence of a single protein of the appropriate size on SDS-

PAGE in males (Figure 4A). As shown in Figure 4B and 4C, the tagged protein is widely expressed in the brain. In the third instar larval brain the protein is widely expressed and is enriched in the mushroom body (MB, Figure 4B, yellow arrow). In the adult brain Pits::GFP is localized in most neurons, including the cell bodies and nuclei (co-staining with a pan-neuronal nuclear marker, Elav) of many neurons as well as their axons (Figure 4C). This is in agreement with a previous study showing *pits* is expressed in the nucleus.[59] Pits::GFP does not show expression in the glia of the adult CNS upon pan-glial staining with Repo (Supplemental Figure 3B). Additionally, we also determined the expression of pits using the GAL4 reporter from *pits*$^{MI02926\text{-}TG4.1}$ allele crossed to a UAS-membrane bound GFP (Supplemental Figure 3C) which allow detection of cells that express the gene of interest at low levels [40] We also generated z-stacks confocal image movies of pits::GFP co-stained with Elav (Supplemental Movie 2) and Repo (Supplemental Movie 3). These data reveal that *pits* is highly expressed in the MB, the learning and memory center of the flies[60] as well as in the antennal mechanosensory and motor center, which are required for balance and hearing and motor coordination, a cerebellar insect equivalent.[61] Again, no expression was observed in glia from these analyses.

Due to the lethality observed in *pits*$^{MI02926\text{-}TG4.1}$ flies, we determined the function of pits in the brain through gene knock-down studies using RNA interference (RNAi).[62] Ubiquitous knock-down of pits using *Act-GAL4* resulted in semi-lethality (Figure 5A). This RNAi has specificity to pits as it consistently reduces the endogenous Pits protein level to ~50% of control RNAi when assessed via Western blot (Figure 5B). To determine if partial knock-down of pits results in neurological defects, we expressed the pits RNAi with the *nSyb-GAL4* driver. Young flies (~5 days post-eclosion) displayed normal climbing (Figure 5C) and mechanical stress tolerance (bang sensitivity) (Figure 5D). However, aged flies that were 30 and 45 days post-eclosion displayed progressive abnormalities in climbing, and became bang sensitive at 30 days post-eclosion when compared to controls (Figure 5C and 5D).

To determine if an age-dependent deterioration in neural morphology can be observed, we reduced *pits* expression in photoreceptors of the fly eye using rhodopsin (*Rh1*)-*GAL4*. The fly retina is a well-characterized system to explore neurodegeneration in *Drosophila* due to the highly stereotypical organization of neurons (photoreceptors) and glia cells (pigment cells, cone cells) in this tissue[19,63]. Indeed, Pits::GFP shows a robust nuclear signal in the photoreceptors of the fly retina when co-stained with Elav (Supplemental Figure 4). Flies were raised in a 12 hr light/dark cycle for 45 days, and histological examination of retina sections stained with toluidine blue was performed. We observed a severe disorganization of the ommatidia (units of photoreceptors) and notable rhabdomere (light-sensing organelle) loss (Figure 6A). This loss was not observed in young flies (5 days post-eclosion), nor in aged control RNAi flies. To examine ultrastructure, we performed transmission electron microscopy (TEM) of the same tissue and observed numerous abnormalities in the ultrastructure of the *Rh1-GAL4>pits-RNAi* retina compared to age matched *Rh1-GAL4>control-RNAi* eyes (Figure 6B and 6C, Supplemental Figure 5 and Supplemental Figure 6): 1) a significant decrease in intact rhabdomeres (Figure 6D); 2) a significant increase in the presence of tubulovesicular like structures (TVS), associated with some neurodegenerative models[64] (Figure 6C, red arrow); and 3) the abnormal presence of neuronal lipid droplets in photoreceptors (Figure 6C, yellow arrow, Supplemental Figure 7A), often associated with mutants that cause high reactive oxygen species (ROS) accumulation.[65] However, we did not detect obvious changes in mitochondrial morphology or the number of mitochondria per photoreceptor (Supplemental Figure 7B). In summary, *pits* is required for proper maintenance of neuronal function and structure in flies.

Discussion

We present seven subjects with rare heterozygous variants in *IRF2BPL*, a gene that has not previously been associated with disease in humans. The cases with nonsense variants exhibited a remarkably similar, progressive course of neurological regression that eventually led

to severe disability. In contrast, the missense variants observed in two subjects are associated with milder neurological symptoms such as seizures, developmental delay, and ASD. Interestingly, the nonsense variants in this cohort cluster in or just downstream of the polyglutamine tract within the variable region, while the missense variants map further downstream (Figure 1A). The bioinformatics signature of the gene suggests that *IRF2BPL* is highly intolerant to variation, and that the variants reported in the cohort are among the most deleterious. The DOMINO score was suggestive of dominant inheritance. RNA sequencing from a blood sample for subject 1 (c.514G>T (p.E172X) revealed expression of the transcript (data not shown). The lack of nonsense-mediated decay of *IRF2BPL* transcript with a nonsense mutation is consistent with the fact that this is a single exon gene.[66] Western blotting on subject samples with nonsense variants could not be pursued due to lack of a commercially-available antibody that recognizes an epitope upstream of the premature truncation.

To date, a role for *IRF2BPL* in humans is limited to an association with developmental phenotypes. For example, *IRF2BPL* has been identified in the top 1,000 genes that are significantly lacking in functional coding variation in non-ASD samples and are enriched for *de novo* LoF mutations identified in ASD cases.[67] Other large scale sequencing studies have identified *de novo* variants in *IRF2BPL* in ASD (2 individuals) and major developmental disorders (2 individuals).[1,3] Intriguingly the ASD variants include a missense in the conserved DNA binding domain (p.F30L) and a frameshift near the end of the protein (p.A701fs*65). The developmental disorder cohort also has a missense (p.R391C) and frameshift that similarly terminates near the end of the protein (p.L713Pfs*54).[1,3] The missense variants in these cases compared to our two cases indicate that the full spectrum of IRF2BPL related phenotypes is still developing. The p.A701fs*65 frameshift, in particular, suggests that truncations near the end of the protein may cause variable phenotypes. Another possibility is the presence of mosacism in these large cohort studies. Our findings offer additional evidence that variants in *IRF2BPL* are implicated in neurological symptoms, and additionally extend the phenotype into neurological

regression. Furthermore, our model organism experiments using fruit flies supports an important role for *IRF2BPL* in embryologic development as well as neuronal maintenance. *IRF2BPL* is well-conserved and the fly gene, *pits,* is expressed widely in the nervous system during development and in adulthood. The *pits*$^{MI02926\text{-}TG4.1}$ LoF mutant fails to survive past early larval stages and mostly die as embryos. Although we were not able to rescue the *pits*$^{MI02926\text{-}TG4.1}$ allele with either the human or fly cDNA, the *pits*$^{MI02926\text{-}TG4}$ is rescued with a genomic rescue construct specific to *pits,* showing that this chromosome does not carry other lethal or second site mutations. Interestingly, ubiquitous overexpression of *IRF2BPL* or *pits* in flies is toxic, suggesting that the gene is highly dosage sensitive. Many genes implicated in neurodegenerative disorders cause overexpression phenotypes in *Drosophila*.[68,69] In our experiments, however, overexpression of the nonsense variants was not toxic. These results suggest that the mechanism of disease may be through loss of protein function and haploinsufficiency of *IRF2BPL*. Overexpression of the missense variants showed a range of effects. While overexpression of p.K418N was only lethal at higher temperatures (i.e., higher levels of expression), p.P372R was lethal at any temperature. These results suggest that some of the proteins expressed with the missense variants retained a function that was toxic when overexpressed, or perhaps suggest a gain of function mechanism. Understanding the precise molecular function of Pits/IRF2BPL will allow one to design experiments to test this possibility. In summary, either excess or loss of *pits* or *IRF2BPL* is highly detrimental to survival. Given that expression at low levels (*Act-GAL4* at 18°C) is still toxic suggests that the level of expression of the gene product is highly regulated *in vivo* through mechanisms such as ubiquitination through the PEST domain.

There are compelling parallels in phenotypes between a partial Pits reduction in flies and symptoms observed in patients, which may suggest evolutionary conservation in neuronal mechanisms. 1) Almost all patients had seizures or EEG abnormalities, and a reduction of Pits in neurons leads to a bang-sensitive phenotype in flies. Bang-sensitivity is associated with

seizure-like paralysis that has phenotypic as well as genetic parallels with human epilepsy.[70,71] 2) The five patients carrying nonsense variants in *IRF2BPL* displayed progressive motor dysfunction that manifested after early childhood. Similarly, neuronal reduction of Pits in flies caused a progressive decline in climbing ability that was not observed in young flies. 3) The two oldest patients, who had nonsense variants, had evidence of cerebral atrophy in adulthood. Correspondingly, we found that a reduction in pits expression in photoreceptors leads to a slow and age-dependent loss of neuronal integrity. 4) The cerebellar symptoms and cerebellar atrophy may correspond to a requirement for Pits expression in the antennal mechanosensory and motor center. These neurons are required for balance, auditory and motor coordination, a cerebellar insect equivalent [61] and Pits is abundantly expressed in these cells. These results support the notion that IRF2BPL / Pits have fundamental roles in central nervous system development and maintenance.

IRF2BPL has similarity to its mammalian paralogs IRF2BP1 and IRF2BP2 and has been shown to interact with IRF2 (Interferon Regulatory Factor 2).[72,73] However, flies lack an obvious homolog to IRF2, indicating that *pits* and *IRF2BPL* may have additional conserved functions that are currently unknown. Recently, a study in *Drosophila* indicated that the fly homologue, Pits, regulates transcription during early embryogenesis by interacting with a histone deacetylase Sin3A (SIN3A and SIN3B in human) and a co-repressor Tramtrack 69 (no identified human ortholog).[57] Consistent with previous studies in flies and rats, our data suggest both Pits and IRF2BPL are found predominantly in the nucleus.[59,74] We showed robust expression in the nucleus of photoreceptors and numerous neurons, but also observed Pits::GFP in axons and cell bodies, and little to no staining in the dendrites of the mushroom body. This could indicate that Pits may have a non-nuclear role in specific subcellular compartments in neurons. Finally, through neuron-specific knock-down experiments in flies, we demonstrate that *pits* is important for function and/or neuronal maintenance over time.

We observed different phenotypes within the cohort based on the type of variant. All four nonsense variants truncate the IRF2BPL protein upstream of the putative nuclear localization signal, the conserved C' terminal RING domain, and multiple putative PEST sequences. Although little is known about IRF2BPL/Pits, most studies implicate its expression and function in the nucleus.[30,57,59,74,75]. However, the RING domain of IRF2BPL has recently been shown to act as an E3-ligase ubiquitinating β-catenin and supressing Wnt signalling in gastric cancer.[32] It is currently unclear whether this pathway could be altered in IRF2BPL-associated disease. Finally, the predicted PEST sequences suggest that IRF2BPL is highly regulated and support our data that overexpression of the protein can be detrimental. Toxicity was observed by overexpression of fly or human cDNA constructs in all cells (*Act-GAL4*) or within the cell-types in which pits is endogenously expressed (*pits*$^{MI02926\text{-}TG4.1}$). While *IRF2BPL* ubiquitous expression or overexpression within the cell types that pits is endogenously expressed caused lethality, we did not observe lethality or any remarkable phenotypes by overexpression of IRF2BPL or the variants when expressed specifically in neurons (*nSyb-GAL4*) (data not shown). Therefore, increased Pits/IRF2BPL protein expression may be detrimental to only certain cells. Our functional assays of nonsense and missense variants support haploinsufficiency and tight control of protein expression and/or turn-over. Future studies will determine IRF2BPL targets for ubiquitination and binding partners that mediate neurological phenotypes.

In summary, we have implicated dominant *de novo* variation in *IRF2BPL* to a new neurological disorder in humans. We observed that patients with nonsense variants in this single exon gene suffer from a progressive and devastating neurological regression, and individuals with rare missense variants also show a spectrum of neurological phenotypes. The bioinformatics signature supports a deleterious effect in IRF2BPL at the gene and variant level. We provide functional analysis in flies to support a loss-of-function model of *IRF2BPL*-associated disease. Future studies examining the mechanism of early death in *pits* mutant flies, as well as molecular mechanisms of neurodegeneration in *pits* knockdown animals, will likely

shed light on conserved pathways essential for neurological development and provide guidance to design an effective therapeutic or preventive strategy.

Conflicts of Interest

The Department of Molecular and Human Genetics at Baylor College of Medicine receives revenue from clinical genetic testing conducted by Baylor Genetics. David Goldstein is a founder of and holds equity in Pairnomix and Praxis, serves as a consultant to AstraZeneca, and has research supported by Janssen, Gilead, Biogen, AstraZeneca, and UCB.

Consortia

The Undiagnosed Diseases Network co-investigators are David R. Adams, Mercedes E. Alejandro, Patrick Allard, Mahshid S. Azamian, Carlos A. Bacino, Ashok Balasubramanyam, Hayk Barseghyan, Gabriel F. Batzli, Alan H. Beggs, Babak Behnam, Anna Bican, David P. Bick, Camille L. Birch, Devon Bonner, Braden E. Boone, Bret L. Bostwick, Lauren C. Briere, Donna M. Brown, Matthew Brush, Elizabeth A. Burke, Lindsay C. Burrage, Shan Chen, Gary D. Clark, Terra R. Coakley, Joy D. Cogan, Cynthia M. Cooper, Heidi Cope, William J. Craigen, Precilla D'Souza, Mariska Davids, Jyoti G. Dayal, Esteban C. Dell'Angelica, Shweta U. Dhar, Ani Dillon, Katrina M. Dipple, Laurel A. Donnell-Fink, Naghmeh Dorrani, Daniel C. Dorset, Emilie D. Douine, David D. Draper, David J. Eckstein, Lisa T. Emrick, Christine M. Eng, Ascia Eskin, Cecilia Esteves, Tyra Estwick, Carlos Ferreira, Brent L. Fogel, Noah D. Friedman, William A. Gahl, Emily Glanton, Rena A. Godfrey, David B. Goldstein, Sarah E. Gould, Jean-Philippe F. Gourdine, Catherine A. Groden, Andrea L. Gropman, Melissa Haendel, Rizwan Hamid, Neil A. Hanchard, Lori H. Handley, Matthew R. Herzog, Ingrid A. Holm, Jason Hom, Ellen M. Howerton, Yong Huang, Howard J. Jacob, Mahim Jain, Yong-hui Jiang, Jean M. Johnston, Angela L. Jones, Isaac S. Kohane, Donna M. Krasnewich, Elizabeth L. Krieg, Joel B. Krier, Seema R. Lalani, C. Christopher Lau, Jozef Lazar, Brendan H. Lee, Hane Lee, Shawn E. Levy, Richard A. Lewis, Sharyn A. Lincoln, Allen Lipson, Sandra K. Loo, Joseph Loscalzo, Richard L. Maas, Ellen F. Macnamara, Calum A. MacRae, Valerie V. Maduro, Marta M. Majcherska, May Christine V. Malicdan, Laura A. Mamounas, Teri A. Manolio, Thomas C. Markello, Ronit Marom,

Julian A. Martínez-Agosto, Shruti Marwaha, Thomas May, Allyn McConkie-Rosell, Colleen E. McCormack, Alexa T. McCray, Matthew Might, Paolo M. Moretti, Marie Morimoto, John J. Mulvihill, Jennifer L. Murphy, Donna M. Muzny, Michele E. Nehrebecky, Stan F. Nelson, J. Scott Newberry, John H. Newman, Sarah K. Nicholas, Donna Novacic, Jordan S. Orange, J. Carl Pallais, Christina G.S. Palmer, Jeanette C. Papp, Neil H. Parker, Loren D.M. Pena, John A. Phillips III, Jennifer E. Posey, John H. Postlethwait, Lorraine Potocki, Barbara N. Pusey, Chloe M. Reuter, Amy K. Robertson, Lance H. Rodan, Jill A. Rosenfeld, Jacinda B. Sampson, Susan L. Samson, Kelly Schoch, Molly C. Schroeder, Daryl A. Scott, Prashant Sharma, Vandana Shashi, Rebecca Signer, Edwin K. Silverman, Janet S. Sinsheimer, Kevin S. Smith, Rebecca C. Spillmann, Kimberly Splinter, Joan M. Stoler, Nicholas Stong, Jennifer A. Sullivan, David A. Sweetser, Cynthia J. Tifft, Camilo Toro, Alyssa A. Tran, Tiina K. Urv, Zaheer M. Valivullah, Eric Vilain, Tiphanie P. Vogel, Colleen E. Wahl, Nicole M. Walley, Chris A. Walsh, Patricia A. Ward, Katrina M. Waters, Monte Westerfield, Anastasia L. Wise, Lynne A. Wolfe, Elizabeth A. Worthey, Shinya Yamamoto, Yaping Yang, Guoyun Yu, Diane B. Zastrow, and Allison Zheng.

The Program for Undiagnosed Diseases (UD-PrOZA) co-investigators are Steven Callens, Paul Coucke, Bart Dermaut, Dimitri Hemelsoet, Bruce Poppe, Wouter Steyaert, Wim Terryn, and Rudy Van Coster.

Acknowledgments

This work is funded by the Undiagnosed Diseases Network U01HG007672 to VS and DBG, U01HG007703 to SFN, and U54NS093793 to HJB, SY & MJW. HJB is also supported by R01GM067858 and R24OD022005, and SY & MFW by a Simons Foundation Functional Screen Award (368479). We would like to thank Alejandro Lomniczi (OHSU) for sharing the *IRF2BPL* cDNA and Hongling Pan for technical assistance. Research reported in this

publication was supported by the National Institute of Neurological Disorders and Stroke (NINDS) under award number K08NS092898 and Jordan's Guardian Angels (to G.M.). Confocal microscopy at BCM is supported in part by U54HD083092 to the Intellectual and Developmental Disabilities Research Center (IDDRC) Neurovisualization Core. HJB is an Investigator of the Howard Hughes Medical Institute. The content is solely the responsibility of the authors, and does not necessarily represent the official views of the National Institutes of Health. The funding sources had no role in the design and conduct of the study, collection, management, analysis and interpretation of the data, preparation, review or approval of the manuscript, or decision to submit the manuscript for publication. The authors would like to thank the subjects and their families for their participation.

Web Resources:

GeneMatcher, http://www.genematcher.org/

OMIM, http://www.omim.org/

Undiagnosed Diseases Network, https://undiagnosed.hms.harvard.edu

UDN *IRF2BPL* page, https://undiagnosed.hms.harvard.edu/genes/irf2bpl/

MARRVEL, http://www.marrvel.org/

ExAC, http://exac.broadinstitute.org/

gnomAD, http://gnomad.broadinstitute.org/

References

1. Iossifov, I., O'Roak, B.J., Sanders, S.J., Ronemus, M., Krumm, N., Levy, D., Stessman, H.A., Witherspoon, K.T., Vives, L., Patterson, K.E., et al. (2014). The contribution of de novo coding mutations to autism spectrum disorder. Nature *515*, 216–221.

2. de Ligt, J., Willemsen, M.H., van Bon, B.W.M., Kleefstra, T., Yntema, H.G., Kroes, T., Vulto-van Silfhout, A.T., Koolen, D.A., de Vries, P., Gilissen, C., et al. (2012). Diagnostic exome sequencing in persons with severe intellectual disability. N. Engl. J. Med. *367*, 1921–1929.

3. Deciphering Developmental Disorders Study (2017). Prevalence and architecture of de novo mutations in developmental disorders. Nature *542*, 433–438.

4. Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U., and Zoghbi, H.Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat. Genet. *23*, 185–188.

5. Chen, L., Chen, K., Lavery, L.A., Baker, S.A., Shaw, C.A., Li, W., and Zoghbi, H.Y. (2015). MeCP2 binds to non-CG methylated DNA as neurons mature, influencing transcription and the timing of onset for Rett syndrome. Proc. Natl. Acad. Sci. U.S.a. *112*, 5509–5514.

6. Kohlschütter, A., and Schulz, A. (2016). CLN2 Disease (Classic Late Infantile Neuronal Ceroid Lipofuscinosis). Pediatr Endocrinol Rev *13 Suppl 1*, 682–688.

7. Rakheja, D., Narayan, S.B., and Bennett, M.J. (2007). Juvenile neuronal ceroid-lipofuscinosis (Batten disease): a brief review and update. Curr. Mol. Med. *7*, 603–608.

8. Engelen, M., Kemp, S., and Poll-The, B.-T. (2014). X-linked adrenoleukodystrophy: pathogenesis and treatment. Curr Neurol Neurosci Rep *14*, 486.

9. Berger, J., and Gärtner, J. (2006). X-linked adrenoleukodystrophy: clinical, biochemical and pathogenetic aspects. Biochim. Biophys. Acta *1763*, 1721–1732.

10. Contreras, M., Mosser, J., Mandel, J.L., Aubourg, P., and Singh, I. (1994). The protein coded by the X-adrenoleukodystrophy gene is a peroxisomal integral membrane protein. FEBS Lett. *344*, 211–215.

11. Edvardson, S., Nicolae, C.M., Agrawal, P.B., Mignot, C., Payne, K., Prasad, A.N., Prasad, C., Sadler, L., Nava, C., Mullen, T.E., et al. (2017). Heterozygous De Novo UBTF Gain-of-Function Variant Is Associated with Neurodegeneration in Childhood. Am. J. Hum. Genet. *101*, 267–273.

12. Toro, C., Hori, R.T., Malicdan, M.C.V., Tifft, C.J., Goldstein, A., Gahl, W.A., Adams, D.R., Harper, F., Wolfe, L.A., Xiao, J., et al. (2018). A recurrent de novo missense mutation in UBTF causes developmental neuroregression. Hum. Mol. Genet. *27*, 691–705.

13. Chao, H.-T., Davids, M., Burke, E., Pappas, J.G., Rosenfeld, J.A., McCarty, A.J., Davis, T., Wolfe, L., Toro, C., Tifft, C., et al. (2017). A Syndromic Neurodevelopmental Disorder Caused by De Novo Variants in EBF3. The American Journal of Human Genetics *100*, 128–137.

14. Luo, X., Rosenfeld, J.A., Yamamoto, S., Harel, T., Zuo, Z., Hall, M., Wierenga, K.J.,

Pastore, M.T., Bartholomew, D., Delgado, M.R., et al. (2017). Clinically severe CACNA1A alleles affect synaptic function and neurodegeneration differentially. PLoS Genet. *13*, e1006905.

15. Lupski, J.R., Reid, J.G., Gonzaga-Jauregui, C., Rio Deiros, D., Chen, D.C.Y., Nazareth, L., Bainbridge, M., Dinh, H., Jing, C., Wheeler, D.A., et al. (2010). Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. N. Engl. J. Med. *362*, 1181–1191.

16. Oláhová, M., Yoon, W.H., Thompson, K., Jangam, S., Fernandez, L., Davidson, J.M., Kyle, J.E., Grove, M.E., Fisk, D.G., Kohler, J.N., et al. (2018). Biallelic Mutations in ATP5F1D, which Encodes a Subunit of ATP Synthase, Cause a Metabolic Disorder. Am. J. Hum. Genet. *102*, 494–504.

17. Schoch, K., Meng, L., Szelinger, S., Bearden, D.R., Stray-Pedersen, A., Busk, O.L., Stong, N., Liston, E., Cohn, R.D., Scaglia, F., et al. (2017). A Recurrent De Novo Variant in NACC1 Causes a Syndrome Characterized by Infantile Epilepsy, Cataracts, and Profound Developmental Delay. Am. J. Hum. Genet. *100*, 343–351.

18. Shashi, V., Pena, L.D.M., Kim, K., Burton, B., Hempel, M., Schoch, K., Walkiewicz, M., McLaughlin, H.M., Cho, M., Stong, N., et al. (2016). De Novo Truncating Variants in ASXL2 Are Associated with a Unique and Recognizable Clinical Phenotype. Am. J. Hum. Genet. *99*, 991–999.

19. Yoon, W.H., Sandoval, H., Nagarkar-Jaiswal, S., Jaiswal, M., Yamamoto, S., Haelterman, N.A., Putluri, N., Putluri, V., Sreekumar, A., Tos, T., et al. (2017). Loss of Nardilysin, a Mitochondrial Co-chaperone for α-Ketoglutarate Dehydrogenase, Promotes mTORC1 Activation and Neurodegeneration. Neuron *93*, 115–131.

20. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.

21. Wangler, M.F., Yamamoto, S., Chao, H.-T., Posey, J.E., Westerfield, M., Postlethwait, J., Members of the Undiagnosed Diseases Network (UDN), Hieter, P., Boycott, K.M., Campeau, P.M., et al. (2017). Model Organisms Facilitate Rare Disease Diagnosis and Therapeutic Research. Genetics *207*, 9–27.

22. Liu, X., Wu, C., Li, C., and Boerwinkle, E. (2016). dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. Hum. Mutat. *37*, 235–241.

23. Wang, J., Al-Ouran, R., Hu, Y., Kim, S.-Y., Wan, Y.-W., Wangler, M.F., Yamamoto, S., Chao, H.-T., Comjean, A., Mohr, S.E., et al. (2017). MARRVEL: Integration of Human and Model Organism Genetic Resources to Facilitate Functional Annotation of the Human Genome. Am. J. Hum. Genet. *100*, 843–853.

24. 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.

25. Gahl, W.A., Wise, A.L., and Ashley, E.A. (2015). The Undiagnosed Diseases Network of the National Institutes of Health: A National Extension. Jama *314*, 1797–1798.

26. Ramoni, R.B., Mulvihill, J.J., Adams, D.R., Allard, P., Ashley, E.A., Bernstein, J.A., Gahl, W.A., Hamid, R., Loscalzo, J., McCray, A.T., et al. (2017). The Undiagnosed Diseases Network: Accelerating Discovery about Health and Disease. Am. J. Hum. Genet. *100*, 185–192.

27. Rampazzo, A., Pivotto, F., Occhi, G., Tiso, N., Bortoluzzi, S., Rowen, L., Hood, L., Nava, A., and Danieli, G.A. (2000). Characterization of C14orf4, a novel intronless human gene containing a polyglutamine repeat, mapped to the ARVD1 critical region. Biochem. Biophys. Res. Commun. *278*, 766–774.

28. GTEx Consortium (2013). The Genotype-Tissue Expression (GTEx) project. Nat. Genet. *45*, 580–585.

29. Rogers, S., Wells, R., and Rechsteiner, M. (1986). Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. Science *234*, 364–368.

30. Heger, S., Mastronardi, C., Dissen, G.A., Lomniczi, A., Cabrera, R., Roth, C.L., Jung, H., Galimi, F., Sippell, W., and Ojeda, S.R. (2007). Enhanced at puberty 1 (EAP1) is a new transcriptional regulator of the female neuroendocrine reproductive axis. J. Clin. Invest. *117*, 2145–2154.

31. Matagne, V., Mastronardi, C., Shapiro, R.A., Dorsa, D.M., and Ojeda, S.R. (2009). Hypothalamic expression of Eap1 is not directly controlled by ovarian steroids. Endocrinology *150*, 1870–1878.

32. Higashimori, A., Dong, Y., Zhang, Y., Kang, W., Nakatsu, G., Ng, S.S.M., Arakawa, T., Sung, J.J.Y., Chan, F.K.L., and Yu, J. (2018). Forkhead Box F2 Suppresses Gastric Cancer through a Novel FOXF2-IRF2BPL-β-Catenin Signaling Axis. Cancer Res. *78*, 1643–1656.

33. Köhler, S., Vasilevsky, N.A., Engelstad, M., Foster, E., McMurry, J., Aymé, S., Baynam, G., Bello, S.M., Boerkoel, C.F., Boycott, K.M., et al. (2017). The Human Phenotype Ontology in 2017. Nucleic Acids Res. *45*, D865–D876.

34. Zhu, X., Petrovski, S., Xie, P., Ruzzo, E.K., Lu, Y.-F., McSweeney, K.M., Ben-Zeev, B., Nissenkorn, A., Anikster, Y., Oz-Levi, D., et al. (2015). Whole-exome sequencing in undiagnosed genetic diseases: interpreting 119 trios. Genet. Med. *17*, 774–781.

35. Rehm, H.L., Berg, J.S., Brooks, L.D., Bustamante, C.D., Evans, J.P., Landrum, M.J., Ledbetter, D.H., Maglott, D.R., Martin, C.L., Nussbaum, R.L., et al. (2015). ClinGen--the Clinical Genome Resource. N. Engl. J. Med. *372*, 2235–2242.

36. Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R., and Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. *20*, 110–121.

37. Venken, K.J.T., Schulze, K.L., Haelterman, N.A., Pan, H., He, Y., Evans-Holm, M., Carlson, J.W., Levis, R.W., Spradling, A.C., Hoskins, R.A., et al. (2011). MiMIC: a highly versatile transposon insertion resource for engineering Drosophila melanogaster genes. Nature Methods *8*, 737–743.

38. Nagarkar-Jaiswal, S., Lee, P.-T., Campbell, M.E., Chen, K., Anguiano-Zarate, S., Gutierrez, M.C., Busby, T., Lin, W.-W., He, Y., Schulze, K.L., et al. (2015). A library of MiMICs allows tagging of genes and reversible, spatial and temporal knockdown of proteins in Drosophila.

eLife Sciences *4*, 2743.

39. Diao, F., Ironfield, H., Luan, H., Diao, F., Diao, F., Shropshire, W.C., Ewer, J., Marr, E., Potter, C.J., Landgraf, M., et al. (2015). Plug-and-Play Genetic Access to Drosophila Cell Types using Exchangeable Exon Cassettes. Cell Reports *10*, 1410–1421.

40. Lee, P.-T., Zirin, J., Kanca, O., Lin, W.-W., Schulze, K.L., Li-Kroeger, D., Tao, R., Devereaux, C., Hu, Y., Chung, V., et al. (2018). A gene-specificT2A-GAL4library forDrosophila. eLife Sciences *7*, e35574.

41. Gnerer, J.P., Venken, K.J.T., and Dierick, H.A. (2015). Gene-specific cell labeling using MiMIC transposons. Nucleic Acids Res. *43*, e56–e56.

42. Venken, K.J.T., Popodi, E., Holtzman, S.L., Schulze, K.L., Park, S., Carlson, J.W., Hoskins, R.A., Bellen, H.J., and Kaufman, T.C. (2010). A molecularly defined duplication set for the X chromosome of Drosophila melanogaster. Genetics *186*, 1111–1125.

43. Bischof, J., Maeda, R.K., Hediger, M., Karch, F., and Basler, K. (2007). An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. Proc. Natl. Acad. Sci. U.S.a. *104*, 3312–3317.

44. Venken, K.J.T., 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.

45. Venken, K.J.T., Carlson, J.W., Schulze, K.L., Pan, H., He, Y., Spokony, R., Wan, K.H., Koriabine, M., de Jong, P.J., White, K.P., et al. (2009). Versatile P[acman] BAC libraries for transgenesis studies in Drosophila melanogaster. Nature Methods *6*, 431–434.

46. Burg, M.G., and Wu, C.-F. (2012). Mechanical and temperature stressor-induced seizure-and-paralysis behaviors in Drosophila bang-sensitive mutants. J. Neurogenet. *26*, 189–197.

47. Madabattula, S.T., Strautman, J.C., Bysice, A.M., O'Sullivan, J.A., Androschuk, A., Rosenfelt, C., Doucet, K., Rouleau, G., and Bolduc, F. (2015). Quantitative Analysis of Climbing Defects in a Drosophila Model of Neurodegenerative Disorders. J Vis Exp e52741.

48. Jaiswal, M., Haelterman, N.A., Sandoval, H., Xiong, B., Donti, T., Kalsotra, A., Yamamoto, S., Cooper, T.A., Graham, B.H., and Bellen, H.J. (2015). Impaired Mitochondrial Energy Production Causes Light-Induced Photoreceptor Degeneration Independent of Oxidative Stress. PLOS Biology *13*, e1002197.

49. Duffy, J.B. (2002). GAL4 system in Drosophila: a fly geneticist's Swiss army knife. Genesis *34*, 1–15.

50. Petrovski, S., Wang, Q., Heinzen, E.L., Allen, A.S., and Goldstein, D.B. (2013). Genic intolerance to functional variation and the interpretation of personal genomes. PLoS Genet. *9*, e1003709.

51. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. *46*, 310–315.

52. Quinodoz, M., Royer-Bertrand, B., Cisarova, K., Di Gioia, S.A., Superti-Furga, A., and Rivolta, C. (2017). DOMINO: Using Machine Learning to Predict Genes Associated with Dominant Disorders. Am. J. Hum. Genet. *101*, 623–629.

53. Harel, T., Yoon, W.H., Garone, C., Gu, S., Coban-Akdemir, Z., Eldomery, M.K., Posey, J.E., Jhangiani, S.N., Rosenfeld, J.A., Cho, M.T., et al. (2016). Recurrent De Novo and Biallelic Variation of ATAD3A, Encoding a Mitochondrial Membrane Protein, Results in Distinct Neurological Syndromes. Am. J. Hum. Genet. *99*, 831–845.

54. Yamamoto, S., Jaiswal, M., Charng, W.-L., Gambin, T., Karaca, E., Mirzaa, G., Wiszniewski, W., Sandoval, H., Haelterman, N.A., Xiong, B., et al. (2014). A drosophila genetic resource of mutants to study mechanisms underlying human genetic diseases. Cell *159*, 200–214.

55. Tan, K.L., Haelterman, N.A., Kwartler, C.S., Regalado, E.S., Lee, P.-T., Nagarkar-Jaiswal, S., Guo, D.-C., Duraine, L., Wangler, M.F., University of Washington Center for Mendelian Genomics, et al. (2018). Ari-1 Regulates Myonuclear Organization Together with Parkin and Is Associated with Aortic Aneurysms. Dev. Cell *45*, 226–244.e228.

56. Liu, N., Schoch, K., Luo, X., Pena, L., Bhavana, V.H., Kukolich, M.K., Stringer, S., Powis, Z., Radtke, K., Mroske, C., et al. (2018). Functional variants in TBX2 are associated with a syndromic cardiovascular and skeletal developmental disorder. Hum. Mol. Genet.

57. Liaw, G.-J. (2016). Pits, a protein interacting with Ttk69 and Sin3A, has links to histone deacetylation. Sci Rep *6*, 33388.

58. Hu, Y., Flockhart, I., Vinayagam, A., Bergwitz, C., Berger, B., Perrimon, N., and Mohr, S.E. (2011). An integrative approach to ortholog prediction for disease-focused and other functional studies. BMC Bioinformatics *12*, 357.

59. Rohrbaugh, M., Clore, A., Davis, J., Johnson, S., Jones, B., Jones, K., Kim, J., Kithuka, B., Lunsford, K., Mitchell, J., et al. (2013). Identification and characterization of proteins involved in nuclear organization using Drosophila GFP protein trap lines. PLoS ONE *8*, e53091.

60. McGuire, S.E., Le, P.T., and Davis, R.L. (2001). The role of Drosophila mushroom body signaling in olfactory memory. Science *293*, 1330–1333.

61. Hsu, C.T., and Bhandawat, V. (2016). Organization of descending neurons in Drosophila melanogaster. Sci Rep *6*, 20259.

62. Dietzl, G., Chen, D., Schnorrer, F., Su, K.-C., Barinova, Y., Fellner, M., Gasser, B., Kinsey, K., Oppel, S., Scheiblauer, S., et al. (2007). A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature *448*, 151–156.

63. Yamamoto, S., Jaiswal, M., Charng, W.-L., Gambin, T., Karaca, E., Mirzaa, G., Wiszniewski, W., Sandoval, H., Haelterman, N.A., Xiong, B., et al. (2014). A Drosophila Genetic Resource of Mutants to Study Mechanisms Underlying Human Genetic Diseases. Cell *159*, 200–214.

64. Sumi-Akamaru, H., Beck, G., Kato, S., and Mochizuki, H. (2015). Neuroaxonal dystrophy in PLA2G6 knockout mice. Neuropathology *35*, 289–302.

65. Liu, L., Zhang, K., Sandoval, H., Yamamoto, S., Jaiswal, M., Sanz, E., Li, Z., Hui, J., Graham, B.H., Quintana, A., et al. (2015). Glial lipid droplets and ROS induced by mitochondrial defects promote neurodegeneration. Cell *160*, 177–190.

66. Cusack, B.P., Arndt, P.F., Duret, L., and Roest Crollius, H. (2011). Preventing dangerous nonsense: selection for robustness to transcriptional error in human genes. PLoS Genet. *7*, e1002276.

67. Samocha, K.E., Robinson, E.B., Sanders, S.J., Stevens, C., Sabo, A., McGrath, L.M., Kosmicki, J.A., Rehnström, K., Mallick, S., Kirby, A., et al. (2014). A framework for the interpretation of de novo mutation in human disease. Nat. Genet. *46*, 944–950.

68. Marcogliese, P.C., Abuaish, S., Kabbach, G., Abdel-Messih, E., Seang, S., Li, G., Slack, R.S., Emdadul Haque, M., Venderova, K., and Park, D.S. (2017). LRRK2(I2020T) Functional Genetic Interactors that Modify Eye Degeneration and Dopaminergic Cell Loss in Drosophila. Hum. Mol. Genet.

69. Chouhan, A.K., Guo, C., Hsieh, Y.-C., Ye, H., Senturk, M., Zuo, Z., Li, Y., Chatterjee, S., Botas, J., Jackson, G.R., et al. (2016). Uncoupling neuronal death and dysfunction in Drosophila models of neurodegenerative disease. Acta Neuropathologica Communications 2013 1:1 *4*, 62.

70. Fergestad, T., Bostwick, B., and Ganetzky, B. (2006). Metabolic disruption in Drosophila bang-sensitive seizure mutants. Genetics *173*, 1357–1364.

71. Pavlidis, P., and Tanouye, M.A. (1995). Seizures and failures in the giant fiber pathway of Drosophila bang-sensitive paralytic mutants. J. Neurosci. *15*, 5810–5819.

72. Li, S., Wang, L., Berman, M., Kong, Y.-Y., and Dorf, M.E. (2011). Mapping a dynamic innate immunity protein interaction network regulating type I interferon production. Immunity *35*, 426–440.

73. Huttlin, E.L., Bruckner, R.J., Paulo, J.A., Cannon, J.R., Ting, L., Baltier, K., Colby, G., Gebreab, F., Gygi, M.P., Parzen, H., et al. (2017). Architecture of the human interactome defines protein communities and disease networks. Nature *545*, 505–509.

74. Xu, J., and Li, P. (2016). Expression of EAP1 and CUX1 in the hypothalamus of female rats and relationship with KISS1 and GnRH. Endocr. J. *63*, 681–690.

75. Li, C., and Li, P. (2017). Enhanced at Puberty-1 (Eap1) Expression Critically Regulates the Onset of Puberty Independent of Hypothalamic Kiss1 Expression. Cell. Physiol. Biochem. *43*, 1402–1412.

Figure Legends

Figure 1: IRF2BPL protein structure and gene constraint.

(A) IRF2BPL is 796 amino acids (AA) long and contains two highly conserved domains (IRF2BP zinc finger and C3HC4 RING) and the N- and C- termini. Within the variable region are multiple polyalanine and PEST sequences and a 25 AA polyglutamine tract (AA 103 to 127). All four nonsense variants occur early in the transcript before predicted PEST sequences, and the two missense variants (highlighted in orange) occur in the middle of the protein.

(B) *IRF2BPL* is highly constrained based on the lack of LoF variants in ExAC [20] resulting in a high probability of LoF intolerance (pLI) score and missense constraint z-score. Predictions based on the DOMINO algorithm indicate that variation in IRF2BPL is likely to lead to a dominantly inherited disease.

Figure 2: Progressive cerebral atrophy in patients with *IRF2BPL* nonsense truncations.

(A) Brain MRI for subject 3 at 7, 13, and 20 years (top row, axial FLAIR; bottom row, sagittal T1). Brain MRI was normal at 7 years. However, at age 13, there was severe diffuse cerebral atrophy with *ex-vacuo* dilatation of the lateral ventricles. There may be slightly increased white matter signal in the peritrigonal region, but otherwise the white matter appears intact and does not suggest a leukodystrophy. The cerebellum has only minimal atrophy. There is mild atrophy of the basal ganglia and brainstem (not shown). At age 20 years, there is further atrophy, including severe volume loss in the bilateral cerebral hemispheres, further thinning of the corpus callosum, mild worsening of the increased white matter signal in the peritrigonal region, further atrophy of the cerebellum and brainstem.

(B) MRI images of subject 5 at 34 years depicting global cerebral and cerebellar atrophy, thinning of corpus callosum and brainstem without focal brain lesions (axial T2-weighted and sagittal T1-weighted). Images were taken on a 1.5T Siemens.

Figure 3: The IRF2BPL homolog, Pits, is highly conserved, and human disease variants display dramatic loss-of-function

(A) The two annotated domains (IRF2BP zinc finger and C3HC4 ring) for IRF2BPL (as well as IRF2BP1 and IRF2BP2) display very high conservation with the fly homolog Pits.

(B). The *pits*$^{MI02926\text{-}TG4.1}$ allele was generated by genetic conversion of *y*^{1} *w** *Mi{MIC}CG11138*^{MI02926} by recombination-mediated cassette exchange (RMCE) *in vivo*. The resulting mutant incorporates a SA-T2A-GAL4 that acts as an artificial exon resulting in early truncation of the *pits* transcript and cellular replacement with expression of GAL4 under the endogenous pits regulatory elements.

(C) Genomic location of genomic rescue (GR) constructs inserted on chromosome 2 (VK37) of the fly. Note that the 20 Kb rescue line is specific to only the *pits* gene.

(D) Reintroduction of either GR construct (Figure 3C) rescues lethality for *pits*$^{MI02926\text{-}TG4.1}$ flies but rescue is not observed by overexpression of the fly or human cDNA. Female *pits*$^{MI02926\text{-}TG4.1}$*/FM7* virgins are crossed to males of either GR or UAS lines, and progeny are examined for males containing *pits*$^{MI02926\text{-}TG4.1}$ and the rescue construct (minimum progeny examined n=91). Examination of female flies heterozygous for the presence of *pits*$^{MI02926\text{-}TG4.1}$ and the rescue construct reveals a lack of toxicity in female *pits*$^{MI02926\text{-}TG4.1}$ */+; UAS-IRF2BPL-E172X/*+ flies indicating loss-of-function.

(E) The ubiquitous expression of *UAS-IRF2BPL* or variants with *Act-GAL4* reveals that all nonsense variants are strong loss-of-function mutations and the p.K418N causes partial loss of function.

Figure 4: The *IRF2BPL* homolog, *pits*, is expressed in both the developing and adult CNS and is present in the nucleus of a wide subset of neurons.

(A) Male fly heads were lysed and run on SDS-PAGE to determine the presence of Pits::GFP.

(B) Pits is widely expressed in 3rd instar larvae, assessed by immunostaining of homozygous pits::GFP animals and viewed by confocal microscopy (z-stack - max projection). Note the enrichment in the mushroom body (yellow arrow).

(C) Single slice confocal images of the adult CNS show *pits* expression in neurons (co-localized with Elav). Notably *pits* is expressed in the cell bodies of the adult mushroom body (top left panel), is enriched in the central complex (yellow arrow) and is not present in the dendrites of the mushroom body (bottom left panel). Scale bar is 50 μm.

Figure 5: Neuronal knock-down of *pits* leads to progressive behavioral deficits.

(A) Ubiquitous knock-down of *pits* results in semi-lethality, as shown by lower than expected genotypic ratios of survival into adulthood. Flies were compared to *Act-GAL4>control-RNAi* and *y w Act-GAL4/+*.

(B) The *pits* RNAi can partially knock-down ~50-60% of the *pits* isoforms consistently observed in female flies.

(C) Pan-neuronal knock-down of *pits* (*nSyb-GAL4>pits-RNAi*) leads to a bang-sensitive paralytic phenotype in aged animals that is not observed in control flies or young animals. Multiple cohorts of flies were anesthetized and single housed 24 hours prior to testing with 15 seconds of vortexing in an empty vial. Statistical analyses were with one-way ANOVA followed by Tukeys post-hoc test. Results are means ±SEM (*$p<0.05$; NS, not significant)

(D) Pan-neuronal knock-down of *pits* using RNAi leads to progressive climbing deficits that are only observed in aged flies. Singly housed flies (similar to above) were given 1 minute to habituate to an empty vial before being tapped 3 times. Flies were given 30 seconds to cross the 7 cm mark on the vial. Flies that failed to cross the line were given a score of 30 (only seen at day 45). One-way ANOVA followed by Tukeys post-hoc test. Data are mean ±SEM (*$p<0.05$, **$p<0.01$; NS, not significant).

Figure 6: Knock-down of pits in the photoreceptor leads to degenerative phenotypes in aged animals.

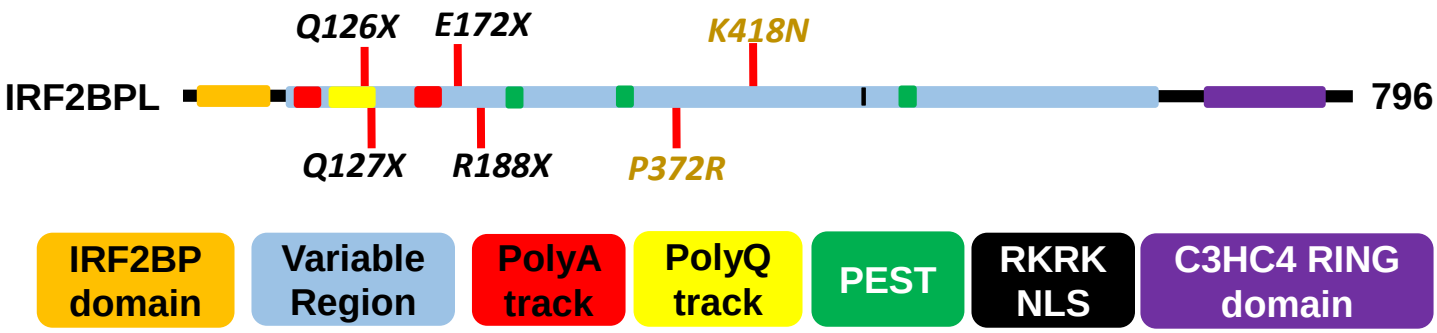
(A) Toluidine blue staining of 5- and 45-day-old retina from *Rh1-GAL4>luciferase-RNAi* and *Rh1-GAL4>pits-RNAi flies* revealing disorganization of the ommatidtia and photoreceptor loss. Scale bar is 5 μm.

(B to C) TEM images showing the photoreceptors and the ommatidia (B, scale bar is 2 μm) and photoreceptors (C, scale bar is 1 μm) section. The red arrow indicates the presence of tubulovesicular like structures (TVS), and the yellow arrow indicates neuronal lipid droplets (C). Further TEM images are available in Supplemental Figures 5 and 6. Quantification of rhabdomere loss (D) and TVS structures (E) on a minimum of n=27 randomized sections from the retina from three animals per genotype. Statistical analyses were with unpaired two-tailed t-tests. Results are mean ±SEM (**$p<0.01$, ****$p<0.0001$).

Table 1:

Salient clinical features for each subject.

A. IRF2BPL Protein Domains and Nonsense and Missense Variants

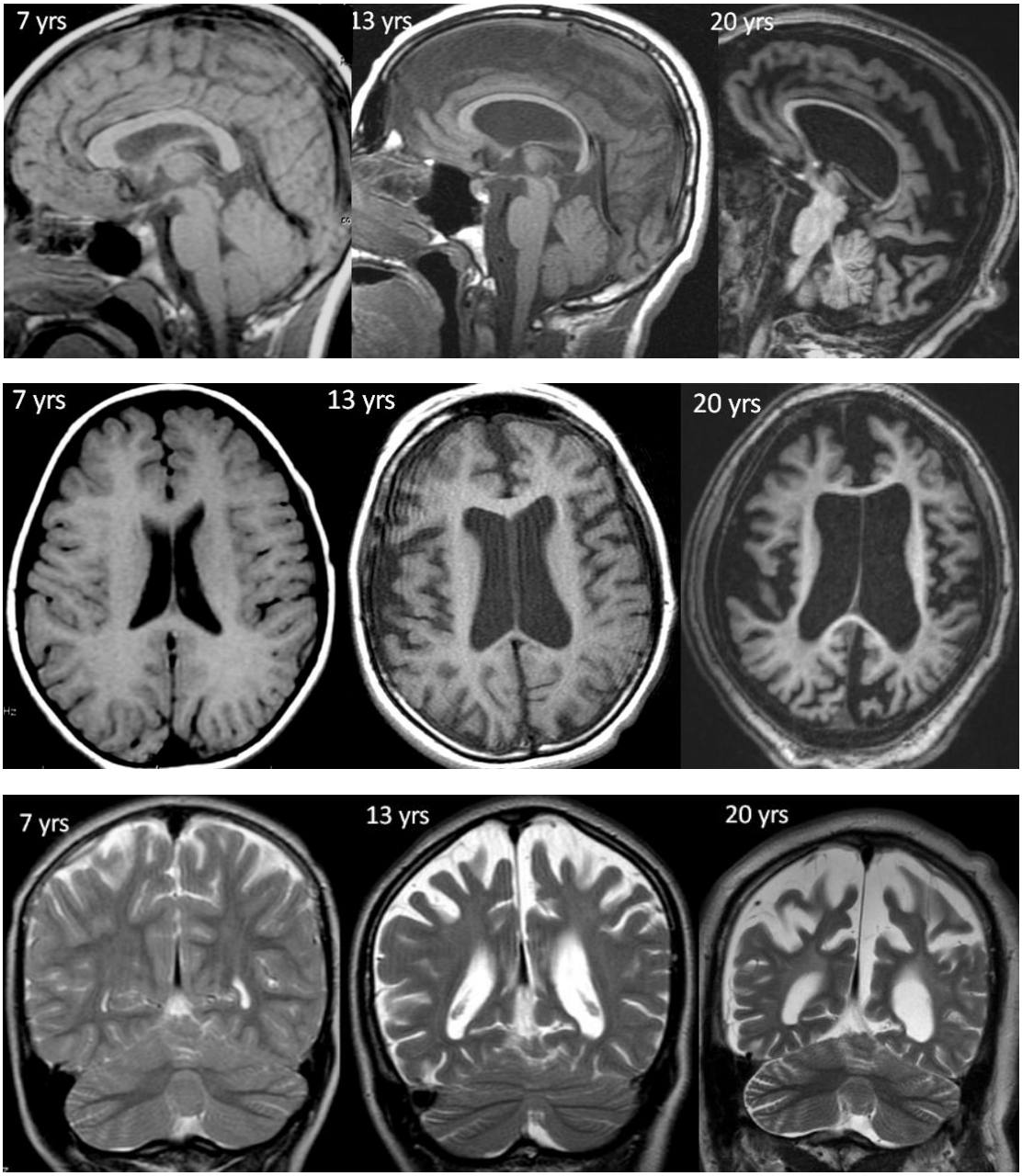


B. IRF2BPL tolerance to change and predicted inheritance

Probability of LoF intolerance (pLI)			Missense Constraint			DOMINO	
Expected	Observed	pLI score	Expected	Observed	z score	Score	Inheritance prediction
10.5	0	0.97	384.8	195	4.73	0.962	Very likely dominant

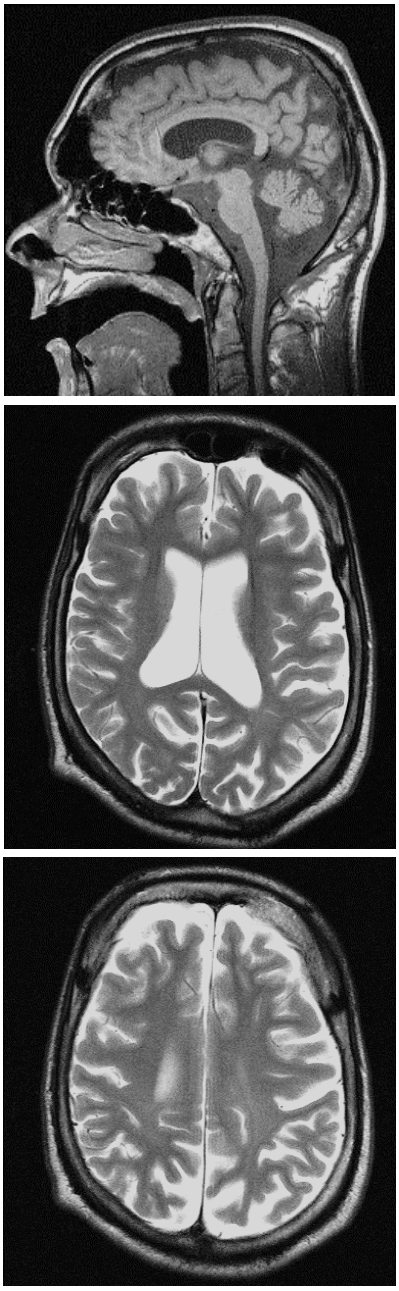
A. **Subject 3 MRI measurements**

7 yo (2005) **13 yo (2011)** **20 yo (2017)**

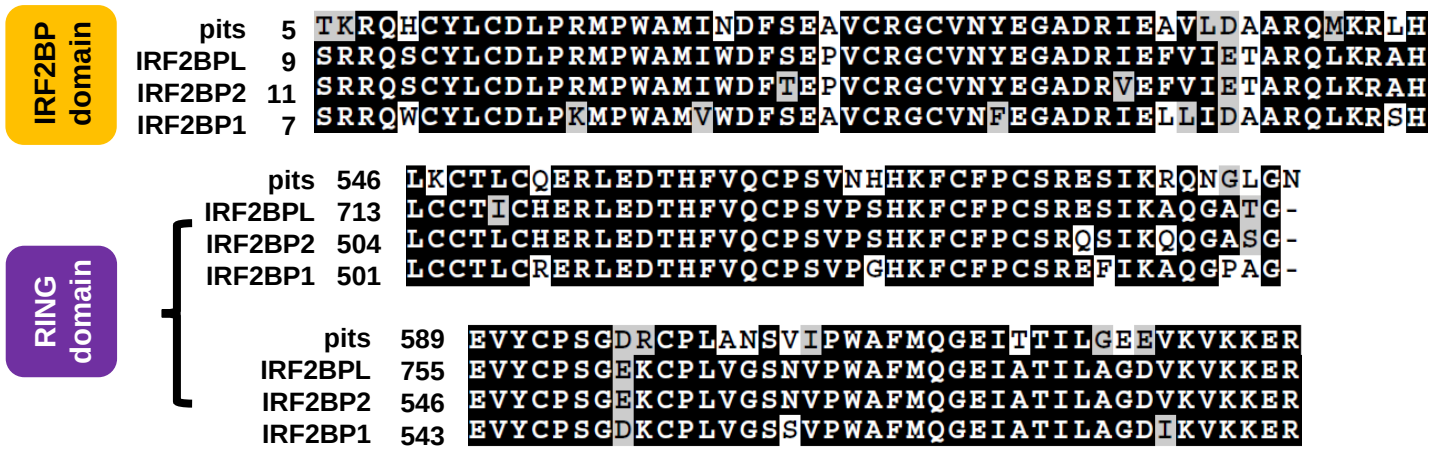


B. **Subject 5 MRI measurements**

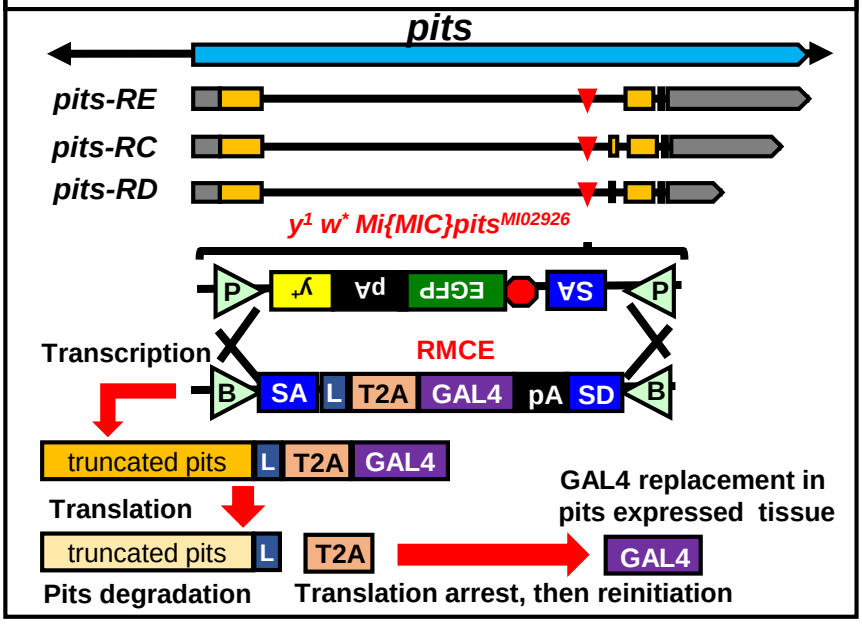
34 yo (2009)



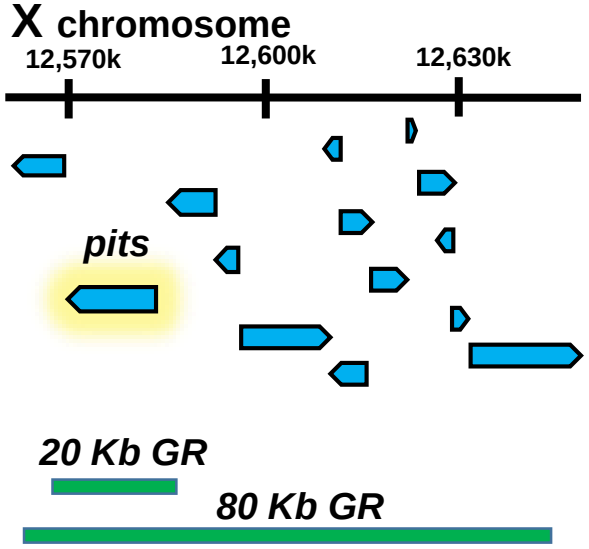
A. pits domain conservation of IRF2BP proteins



B. *pits*$^{MI02926\text{-}TG4.1}$ generation



C. Genomic Rescue lines



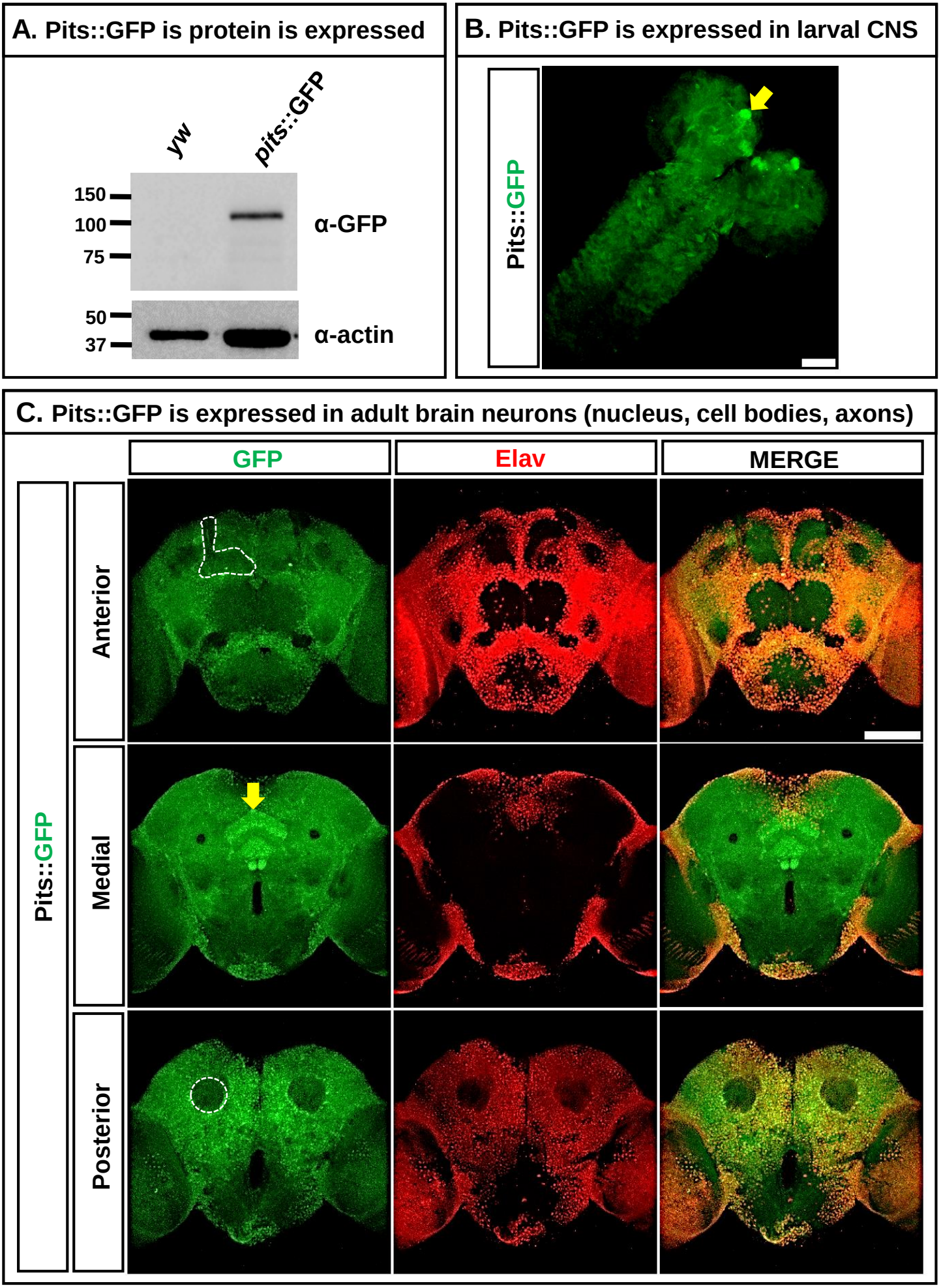
D. IRF2BPL fails to rescue *pits*$^{MI02926\text{-}TG4.1}$

UAS	*pits*$^{MI02926\text{-}TG4.1}$ *Hemi* ♂	
	Lethal	
80 Kb GR	Viable	
20 Kb GR	Viable	*Het* ♀
pits-RC	Lethal	Lethal
IRF2BPL	Lethal	Lethal
IRF2BPL-E172X	Lethal	Viable

E. *IRF2BPL* nonsense variants lose toxic function

Act-GAL4 mediated transgene expression

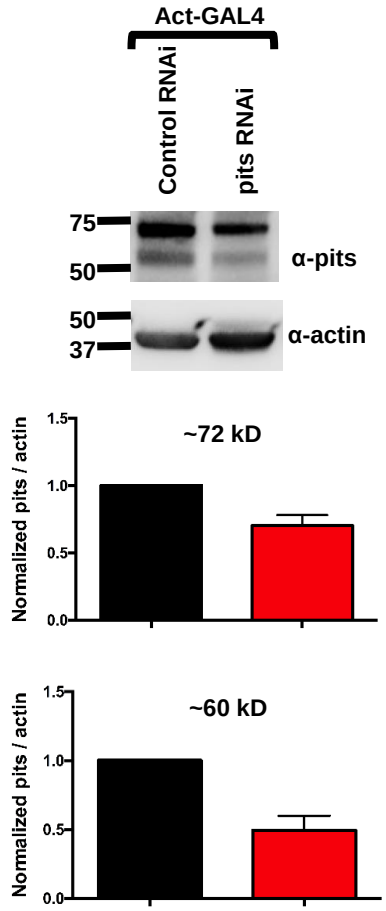
IRF2BPL	18°C	22°C	25°C	29°C
Reference	Lethal	Lethal	Lethal	Lethal
E172X	Viable	Viable	Viable	Viable
Q127X	Viable	Viable	Viable	Viable
R188X	Viable	Viable	Viable	Viable
P372R	Lethal	Lethal	Lethal	Lethal
K418N	Viable	Viable	Lethal	Lethal


A. Pits::GFP is protein is expressed
yw
pits::GFP
150
100
75
α-GFP
50
37
α-actin
B. Pits::GFP is expressed in larval CNS
Pits::GFP
C. Pits::GFP is expressed in adult brain neurons (nucleus, cell bodies, axons)
GFP
Elav
MERGE
Pits::GFP
Anterior
Medial
Posterior

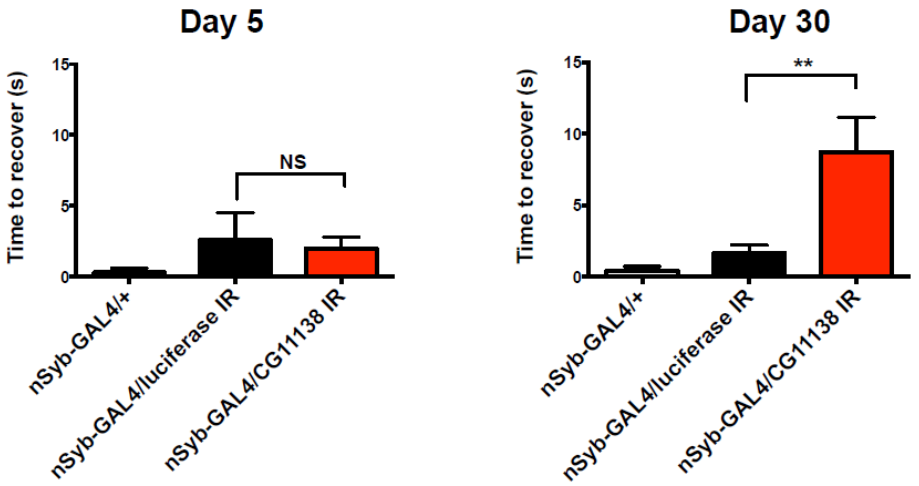
A. Ubiquitous pits knock-down is semi-lethal

Act-GAL4	25°C	29°C
yw	Viable	Viable
Control RNAi	Viable	Viable
Pits RNAi	Semi-lethal (<50% viability)	Semi-lethal (<20% viability)

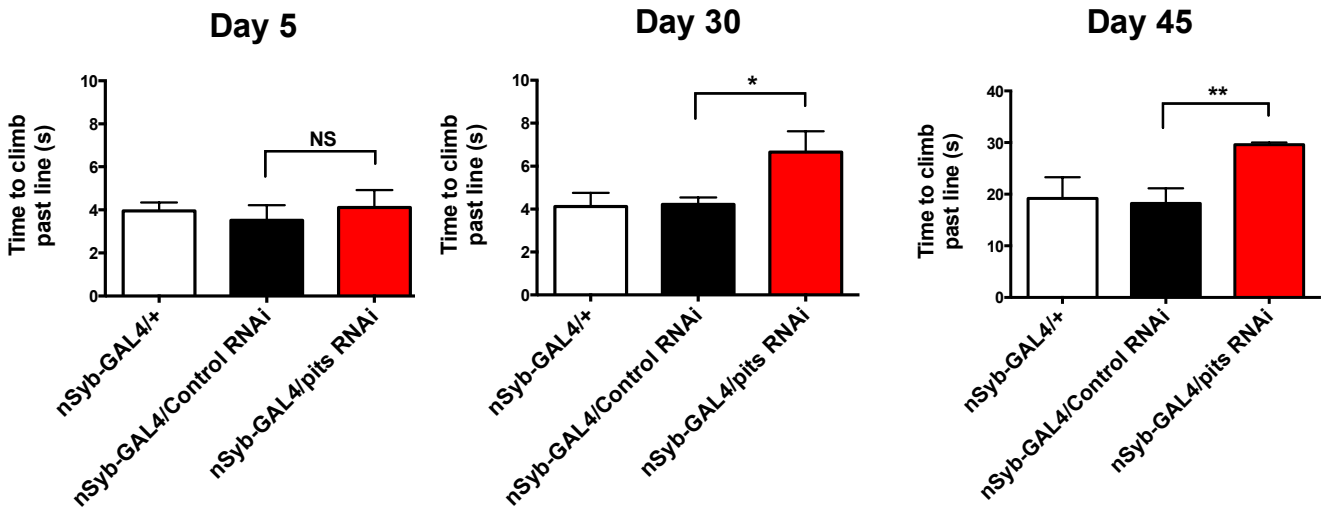
B. pits-RNAi knock-down efficiency



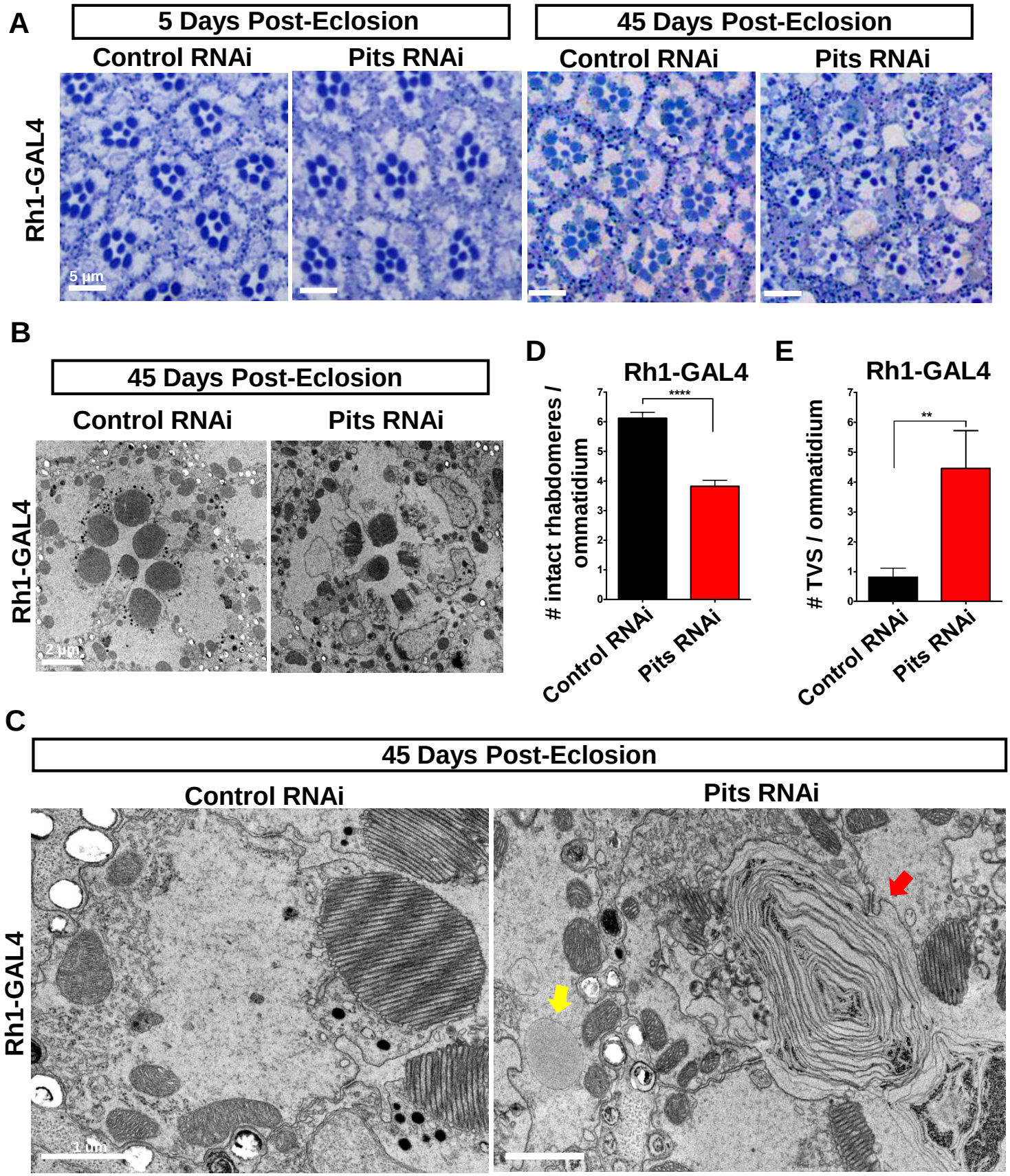
C. Neuronal knock-down of pits causes progressive bang-sensitivity defects



D. Neuronal knock-down of pits causes progressive climbing defects



Knock-down of pits in the photoreceptors causes neurodegeneration in aged flies



	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7
***IRF2BPL* variant**	c.584G>T (p.G195V) AND c.514G>T (p.E172X) (complex rearrangement Sanger confirmed *de novo*	c.562C>T p.R188X Sanger confirmed No parents available for testing	c.562C>T p.R188X Sanger confirmed *de novo*	c. 379C>T p.Q127X Sanger confirmed, no parents available for testing	c.376C>T p.Q126X Sanger confirmed *de novo*	c.1115C>G p.P372R Sanger confirmed *de novo*	c.1254G>C p.K418N No Sanger confirmation, present in 42/87 reads *de novo*
Gender	M	F	M	F	M	M	F
Current age	7 years	Deceased at 15 years	20 years	15 years	43 years	10 years	2.5 years
Growth for age at most recent visit (Z-score)*	6 yr 6 mo Wt 20.1kg (-0.42) Lt 116.8cm (-0.67) HC-51.4cm -0.17)	12 years Wt 45 kg (0.5) Ht-168 cm (2)	20 years Wt 66 kg (-0.43) H 141 cm (-5.0) HC 59 cm (2.4)	15 years Wt 46.9kg (-0.5) Ht 162.4cm (0) HC-52cm (-2)	NA	11 yr 2 mo Wt 32.6kg (-0.62) Ht 125cm (-2.81) HC 51cm (-1.13)	2 yr 1 mo Wt 13.30kg (0.5) Ht 86.30cm (-1) HC 49.40cm (1)
Developmental delays preceding regression	Present	None	None	Present	NA	NA	NA
Age of onset of motor regression	2.5-3 years	7 years	5 years	Age at onset unknown	5-10 years	3-4 years	No regression
Current speech and language skills	No speech at 6.5 years of age	Lack of speech since 12 years	Lack of speech since 10 years	NA	NA	Has few words, utilizes augmentive communication device	Echolalia, sign language
Current gross motor skills	Wheelchair-bound since since 5.5 years	In a wheelchair since 9 years of age	Non-ambulatory since 11 years	Unsteady gait, clumsy	NA	Normal stance and gait, tandem and reciprocal.	Walking easily up and down stairs, starting to run
Current oromotor skills	Mild-moderate oropharyngeal dysphagia, silent aspiration & tube feeds	Dysphagia at 10 years	Progressive dysphagia since 10 years of age	Silent aspiration at 15 years	NA	Normal	Normal
Seizures	Diagnosed with seizures at 6 years	'Staring spells'	Myoclonus at 9-10 years	Seizures at 13 years	Febrile, photosensitive and myoclonic epilepsy	Myoclonic, atonic and absence seizures at 14 years	Infantile spasms at 6 months
Movement abnormalities	Ataxia, dystonia, choreoathetosis	Dystonia, no ataxia	Ataxic gait at 6 years	Lower extremity dystonia.at 15	Ataxia, choreoathetosis,	None	None

				years	generalized dystonia		
Other neurological findings	Spasticity, cerebellar signs	NA	Spasticity, cerebellar signs, positive Romberg sign by 7 years	NA	Hyperreflexia	None	None
EEG	Abnormal at 5 years, with diffuse slowing, interictal discharges from left occipital and right temporal lobes	EEG with mild background slowing at 9 and 13 years; normal EEGs at 9 and 14 years	EEG normal at 13 years	Abnormal since 6.5 years	Abnormal at 31 years	Abnormal with generalized spike and wave	Abnormal at 6 months
Brain MRI	Normal at 5 years	'Bulky' corpus callosum, mild cerebellar volume loss, large left middle cranial fossa arachnoid cyst at 8 years. Marked cord thinning on spine MRI at 10 years	Normal at 7 years; at 13 years, diffuse cerebral atrophy and at 20 years cerebral atrophy	Normal at 6 & 13 years.	Cerebral, cerebellar, brainstem and corpus callosum atrophy at 26 years	Probable Rathke's cyst at 16 months, otherwise normal.	Normal at 6 months and at 2 years

*Z-score per Centers for Disease Control growth charts in case of weight (Wt) and height/length (Ht/Lt), or Nellhaus growth charts for head circumference (HC).

NA, not available