Journal of Visualized Experiments

In vivo functional study of disease-associated rare human variants using Drosophila

--Manuscript Draft--

Article Type:	Invited Methods Article - JoVE Produced Video
Manuscript Number:	JoVE59658R3
Full Title:	In vivo functional study of disease-associated rare human variants using Drosophila
Keywords:	Human genetics and genomics; Mendelian diseases; rare and undiagnosed diseases; Undiagnosed Diseases Network (UDN); Drosophila melanogaster; variant of unknown significance (VUS); gene of uncertain significance (GUS); Functional genomics; transgenic flies; UAS/GAL4 system; T2A-GAL4; electroretinogram (ERG)
Corresponding Author:	Shinya Yamamoto, D.V.M., Ph.D., Baylor College of Medicine Houston, TX UNITED STATES
Corresponding Author's Institution:	Baylor College of Medicine
Corresponding Author E-Mail:	yamamoto@bcm.edu
Order of Authors:	J. Michael Harnish
	Samantha L. Deal
	Hsiao-Tuan Chao
	Michael F. Wangler
	Shinya Yamamoto, D.V.M., Ph.D.,
Additional Information:	
Question	**Response**
Please indicate whether this article will be Standard Access or Open Access.	Standard Access (US$2,400)
Please indicate the **city, state/province, and country** where this article will be **filmed**. Please do not use abbreviations.	Houston, TX, USA

TITLE:

In Vivo Functional Study of Disease-Associated Rare Human Variants using Drosophila

AUTHORS AND AFFILIATIONS:

J. Michael Harnish[1*], Samantha L. Deal[2*], Undiagnosed Diseases Network[%], Hsiao-Tuan Chao[1,3,4,5], Michael F. Wangler[1,2,4], Shinya Yamamoto[1,2,4,5#]

[1]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA
[2]Program in Developmental Biology, Baylor College of Medicine, Houston, TX, USA
[3]Department of Pediatrics, Section of Neurology and Developmental Neuroscience, Baylor College of Medicine, Houston, TX, USA
[4]Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX, USA
[5]Department of Neuroscience, Baylor College of Medicine, Houston, TX, USA

*These authors contributed equally.
[%]A list of members of the Undiagnosed Diseases Network is provided in **Supplemental Table 1**.

Corresponding Author:

Shinya Yamamoto (yamamoto@bcm.edu)
Tel: +1-832-824-8119

E-mail Addresses of Co-Authors:

J. Michael Harnish (Jacob.Harnish@bcm.edu)
Samantha L. Deal (Samantha.Deal@bcm.edu)
Hsiao-Tuan Chao (hc140077@bcm.edu)
Michael F. Wangler (mw147467@bcm.edu)

KEYWORDS:

human genetics and genomics, Mendelian diseases, rare and undiagnosed diseases, Undiagnosed Diseases Network, *Drosophila melanogaster*, variant of unknown significance, VUS, gene of uncertain significance, GUS, functional genomics, transgenic flies, UAS/GAL4 system, T2A-GAL4, electroretinogram, ERG

SUMMARY:

The goal of this protocol is to outline the design and performance of in vivo experiments in *Drosophila melanogaster* to assess the functional consequences of rare gene variants associated with human diseases.

ABSTRACT:

Advances in sequencing technology have made whole-genome and whole-exome datasets more accessible for both clinical diagnosis and cutting-edge human genetics research. Although a number of in silico algorithms have been developed to predict the pathogenicity of variants identified in these datasets, functional studies are critical to determining how specific genomic

variants affect protein function, especially for missense variants. In the Undiagnosed Diseases Network (UDN) and other rare disease research consortia, model organisms (MO) including *Drosophila*, *C. elegans*, zebrafish, and mice are actively used to assess the function of putative human disease-causing variants. This protocol describes a method for the functional assessment of rare human variants used in the Model Organisms Screening Center *Drosophila* Core of the UDN. The workflow begins with gathering human and MO information from multiple public databases, using the MARRVEL web resource to assess whether the variant is likely to contribute to a patient's condition as well as design effective experiments based on available knowledge and resources. Next, genetic tools (e.g., T2A-GAL4 and UAS-human cDNA lines) are generated to assess the functions of variants of interest in *Drosophila*. Upon development of these reagents, two-pronged functional assays based on rescue and overexpression experiments can be performed to assess variant function. In the rescue branch, the endogenous fly genes are "humanized" by replacing the orthologous *Drosophila* gene with reference or variant human transgenes. In the overexpression branch, the reference and variant human proteins are exogenously driven in a variety of tissues. In both cases, any scorable phenotype (e.g., lethality, eye morphology, electrophysiology) can be used as a read-out, irrespective of the disease of interest. Differences observed between reference and variant alleles suggest a variant-specific effect, and thus likely pathogenicity. This protocol allows rapid, in vivo assessments of putative human disease-causing variants of genes with known and unknown functions.

INTRODUCTION:

Patients with rare diseases often undergo an arduous journey referred to as the "diagnostic odyssey" to obtain an accurate diagnosis[1]. Most rare diseases are thought to have a strong genetic origin, making genetic/genomic analyses critical elements of the clinical workup. In addition to candidate gene panel sequencing and copy number variation analysis based on chromosomal microarrays, whole-exome (WES) and whole-genome sequencing (WGS) technologies have become increasingly valuable tools over the past decade[2,3]. Currently, the diagnostic rate for identifying a known pathogenic variant in WES and WGS is ~25% (higher in pediatric cases)[4,5]. For most cases that remain undiagnosed after clinical WES/WGS, a common issue is that there are many candidate genes and variants. Next-generation sequencing often identifies novel or ultra-rare variants in many genes, and interpreting whether these variants contribute to disease phenotypes is challenging. For example, although most nonsense or frameshift mutations in genes are thought to be loss-of-function (LOF) alleles due to nonsense-mediated decay of the encoded transcript, truncating mutations found in the last exons escape this process and may function as benign or gain-of-function (GOF) alleles[6].

Moreover, predicting the effects of a missense allele is a daunting task, since it can result in a number of different genetic scenarios as first described by Herman Muller in the 1930s (i.e., amorph, hypomorph, hypermorph, antimorph, neomorph, or isomorph)[7]. Numerous in silico programs and methodologies have been developed to predict the pathogenicity of missense variants based on evolutionary conservation, type of amino acid change, position within a functional domain, allele frequency in the general population, and other parameters[8]. However, these programs are not a comprehensive solution to solving the complicated problem of variant

interpretation. Interestingly, a recent study demonstrated that five broadly used variant pathogenicity prediction algorithms (Polyphen[9], SIFT[10], CADD [11], PROVEAN [12], Mutation Taster) agree on pathogenicity ~80% of the time[8]. Notably, even when all algorithms agree, they return an incorrect prediction of pathogenicity up to 11% of the time. This not only leads to flawed clinical interpretation but also may dissuade researchers from following up on new variants by falsely listing them as benign. One way to complement the current limitation of in silico modeling is to provide experimental data that demonstrates the effect of variant function in vitro, ex vivo (e.g., cultured cells, organoids), or in vivo.

In vivo functional studies of rare disease associated variants in MO have unique strengths[13] and have been adopted by many rare disease research initiatives around the world, including the Undiagnosed Diseases Network (UDN) in the United States and Rare Diseases Models & Mechanisms (RDMM) Networks in Canada, Japan, Europe, and Australia[14]. In addition to these coordinated efforts to integrate MO researchers into the workflow of rare disease diagnosis and mechanistic studies at a national scale, a number of individual collaborative studies between clinical and MO researchers have led to the discovery and characterization of many new human disease-causing genes and variants[82-84].

In the UDN, a centralized Model Organisms Screening Center (MOSC) receives submissions of candidate genes and variants with a description of the patient's condition and assesses whether the variant is likely to be pathogenic using informatics tools and in vivo experiments. In Phase I (2015–2018) of the UDN, the MOSC comprised of a *Drosophila* Core [Baylor College of Medicine (BCM)] and Zebrafish Core (University of Oregon) that worked collaboratively to assess cases. Using informatics analysis and a number of different experimental strategies in *Drosophila* and zebrafish, the MOSC has so far contributed to the diagnosis of 132 patients, identification of 31 new syndromes[55], discovery of several new human disease genes (e.g., *EBF3*[15], *ATP5F1D*[16], *TBX2*[17], *IRF2BPL*[18], *COG4*[19], *WDR37*[20]) and phenotypic expansion of known disease genes (e.g., *CACNA1A*[21], *ACOX1*[22]).

In addition to projects within the UDN, MOSC *Drosophila* Core researchers have contributed to new disease gene discoveries in collaboration with the Centers for Mendelian Genomics and other initiatives (e.g., *ANKLE2*[23], *TM2D3*[24], *NRD1*[25], *OGDHL*[25], *ATAD3A*[26], *ARIH1*[27], *MARK3*[28], *DNMBP*[29]) using the same set of informatics and genetic strategies developed for the UDN. Given the significance of MO studies on rare disease diagnosis, the MOSC was expanded to include a *C. elegans* Core and second Zebrafish core (both at Washington University at St. Louis) for Phase II (2018–2022) of the UDN.

This manuscript describes an in vivo functional study protocol that is actively used in the UDN MOSC *Drosophila* Core to determine if missense variants have functional consequences on the protein of interest using transgenic flies that express human proteins. The goal of this protocol is to help MO researchers work collaboratively with clinical research groups to provide experimental evidence that a candidate variant in a gene of interest has functional consequences, thus facilitating clinical diagnosis. This protocol is most useful in a scenario in

which a *Drosophila* researcher is approached by a clinical investigator who has a rare disease patient with a specific candidate variant in a gene of interest.

This protocol can be broken down into three elements: (1) gathering information to assess the likelihood of the variant of interest being responsible for the patient phenotype and the feasibility of a functional study in *Drosophila*, (2) gathering existing genetic tools and establishing new ones, and (3) performing functional studies in vivo. The third element can further be subdivided into two sub-elements based on how the function of a variant of interest can be assessed (rescue experiment or overexpression-based strategies). It is important to note that this protocol can be adapted and optimized to many scenarios outside of rare monogenic disease research (e.g., common diseases, gene-environment interactions, and pharmacological/genetic screens to identify therapeutic targets). The ability to determine the functionality and pathogenicity of variants will not only benefit the patient of interest by providing accurate molecular diagnosis but will also have broader impacts on both translational and basic scientific research.

PROTOCOL:

1. Gathering human and MO information to assess: likelihood of a variant of interest being responsible for disease phenotypes and feasibility of functional studies in *Drosophila*

1.1. Perform extensive database and literature searches to determine whether the specific genes and variants of interest are good candidates to explain the phenotype of the patient of interest. Specifically, gather the following information.

1.1.1 Assess if the gene of interest has been previously implicated in other genetic disorders (phenotypic expansion of known disease gene) or this is an entirely new disease candidate gene [gene variant of uncertain significance (GVUS)].

1.1.2. Assess the allele frequency of the variant of interest in disease or control population databases.

1.1.3. Assess whether there are copy number variations (CNVs) that include this gene in disease or control population databases.

1.1.4. Assess what the orthologous genes are in different MO species including mouse, zebrafish, *Drosophila*, *C. elegans,* and yeast, then further investigate the known functions and expression patterns of these orthologous genes.

1.1.5. Assess whether the variant of interest is present in a functional domain of the protein and if the amino acid of interest is evolutionarily conserved.

NOTE: Answers to these five questions (steps 1.1.1–1.1.5) can be obtained by accessing a number of human and MO databases individually or by using the MARRVEL (Model organism

Aggregated Resources for Rare Variant Exploration) web resource (see **Table 1** for online resources)[30], which is described in-depth in an accompanying article[31]. See the representative results section for specific examples. The Monarch Initiative website[32] and Gene2Function[33] also provide useful information.

1.2. Gather additional information to further assess whether the variant is a good disease candidate from a protein function and structure point-of-view.

1.2.1. Assess if the variant of interest is predicted to be damaging based on in silico prediction algorithms.

NOTE**:** Variant pathogenicity algorithms have been developed by many research groups over the past ~15 years, and some are also displayed in the MARRVEL search results. More recent programs, including the two listed below, combine multiple variant pathogenicity prediction algorithms and machine learning approaches to generate a pathogenicity score. For more information on variant prediction algorithms and their performance, refer to Ghosh, et al.[8]. (i) CADD (combined annotation-dependent depletion): integrative annotation tool built from more than 60 genomic features, which provides scores for human SNVs as well as short insertions and deletions[11]. (ii) REVEL (rare exome variant ensemble learner): combines multiple variant pathogenicity algorithms (MutPred, FATHMM, VEST, PolyPhen, SIFT, PROVEAN, MutationAssessor, MutationTaster, LRT, GERP, SiPhy, phyloP, and phastCons) to provide an integrated score for all possible human missense variants[34].

1.2.2. Determine if the human gene/protein of interest or its MO orthologs have been shown to genetically or physically interact with genes/proteins previously linked to genetic diseases. If so, assess if the patient of interest exhibits overlapping phenotypes with these disorders.

NOTE: Several tools have been developed to analyze genetic and protein-protein interactions based on MO publications as well as large-scale proteomics from multiple species screens. STRING (search tool for recurring instances of neighboring genes)[35]: a database for known and predicted protein-protein interactions. It integrates genetic interaction and co-expression datasets as well as text-mining tools to identify genes and proteins that may function together in a variety of organisms. MIST (molecular interaction search tool)[36]: a database that integrates genetic and protein-protein interaction data from core genetic MOs (yeast, *C. elegans*, *Drosophila*, zebrafish, frog, rat, mouse) and humans. Prediction of interactions inferred from orthologous genes/proteins (interlogs) are also displayed.

1.2.3. Determine if the 3-D structure of the protein of interest has been solved or modeled. If so, determine where the variant of interest map relative to key functional domains.

NOTE: Protein structures solved by X-ray crystallography, nuclear magnetic resonance (NMR), and cryo-electron microscopy can be found in public databases including the PDB (protein data bank) and EMDatabank[37]. Although there is no single database for

predicted/modeled protein structures, a number of algorithms (i.e., SWISS-MODEL[38], Modeller[39], and Phyre$_2$[40]) are available for users to perform protein modeling.

1.3. Communicate with clinical collaborators to discuss information gathered from the informatics analyses in sections 1.1–1.2. If clinical collaborators also feel that the variant and gene of interest are good candidates to explain the phenotypes seen in the patient, proceed to section 2. If there are specific questions about the patient's genotype and phenotype, discuss them with the clinical collaborators before moving forward.

NOTE: If it is felt that the variant of interest is unlikely to explain the phenotype of interest (e.g., identical variant found in high frequency in control population), discuss this with clinical collaborators to determine whether the variant is a good candidate, as there may not be the appropriate expertise to interpret clinical phenotypes.

2. Gathering existing genetic tools and establishing new reagents to study a specific variant of interest

NOTE: Once the variant of interest has been determined a good candidate to pursue experimentally, gather or generate reagents to perform in vivo functional studies. For functional studies described in this protocol, some key *Drosophila melanogaster* reagents are needed: 1) upstream activation sequence-regulated human cDNA transgenic strains that carry the reference or variant sequence, 2) a loss-of-function allele of a fly gene of interest, and 3) a GAL4 line that can be used for rescue experiments.

2.1. Generation of UAS-human cDNA constructs and transgenic flies

2.1.1. Identify and obtain the appropriate human cDNA constructs. Many clones are available from the MGC (mammalian gene collection)[43] and can be purchased from selected venders. For genes that are alternatively spliced, check which isoform cDNA corresponds to using Ensembl or RefSeq.

NOTE: Many cDNAs are available in recombinase-mediated cloning system compatible reagents[44], which simplifies the subcloning step. cDNAs may come in an "open (no stop codon)" or "closed (with endogenous or artificial stop codon)" format. While open clones allow C'-tagging of proteins that are useful for biochemical (e.g., western blot) and cell biological (e.g., immunostaining) assays to monitor expression of the protein of interest, it may interfere with protein function in some cases.

2.1.2. Sub-clone the reference and variant cDNA into the *Drosophila* transgenic vector. Use the ϕC31-mediated transgenesis system, since this allows the reference and variant cDNAs to be integrated into the same location in the genome[45]. For this project, the MOSC *Drosophila* Core routinely uses the pGW-HA.attB vector[46].

NOTE: If the human cDNA is in recombinase-mediated cloning system compatible vectors (e.g., pDONR221, pENTR221), skip to section 2.1.4, which explains LR reactions to subclone the cDNAs into pGW-HA.attB.

2.1.2.1. If the human cDNA is not in a recombinase-mediated cloning system compatible plasmid, subclone the human cDNAs into a Gateway entry vector using standard molecular biological techniques (an example of such protocol is documented below).

2.1.2.1.1. Perform an overhang PCR to introduce *attB1* and *attB2* arms. The forward primer should have the *attB1* sequence 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCACC-3' followed by the first 22 nucleotides of the target cDNA. The reverse primer should have the *attB2* sequence 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTCCTA-3' followed by the reverse complement of the last 25 nucleotides of the cDNA of interest. Exclude the stop codon of the cDNA, then add a C' tag if it is desired to "open" a clone, or add a stop codon to "close" a clone.

2.1.2.1.2. Prepare a 100 µL of high-fidelity PCR mix consisting of 50 µL of high-fidelity PCR master mix, 36 µL of distilled water, 5 µL of each forward and reverse primers listed in step 2.1.2.1.1 diluted to 10 µM, and 4 µL of target cDNA (150 ng/µL).

2.1.2.1.3. Perform PCR using standard mutagenesis protocol to add *attB1* and *attB2* arms onto the cDNA of interest. Conditions will vary depending on the construct and variants of interest.

2.1.2.1.4. Isolate the target cDNA with added homology arms via gel electrophoresis and gel extraction. Create 1% agarose gel and perform electrophoresis using standard methods. Excise the band that corresponds to the size of the cDNA plus the additional length of the homology arms. Extract DNA from the gel through standard methods[95]. Commercial gel extraction kits are available from several companies.

2.1.2.1.5. Perform an in vitro recombinase reaction based on the recombinase-mediated cloning protocol according to the system that is used.

2.1.2.1.6. Transform the BP reaction mix into chemically competent *E. coli* cells. Competent cells can be made in-house or purchased from commercial vendors. Culture the transformed cells overnight on an LB plate containing appropriate antibiotics for colony selection. The next day, select several colonies and grow them in independent liquid cultures overnight.

2.1.2.1.7. Isolate DNA from the overnight cultures through miniprep. Sanger sequence the positive clones to ensure that the cDNA has the correct sequence[96]. Maintain cells from the cultures that are positive for the desired sequence in 25% glycerol stored at -80 °C.

2.1.3. Perform site-directed mutagenesis to introduce the variant of interest into the Gateway plasmid with the reference human cDNA[97]. A detailed protocol for this method can be found in the vendor's website[48,49]. Validate the presence of the variant in the mutated plasmid via

Sanger sequencing of the entire open reading frame (ORF) to ensure that there are no additional variants introduced through this mutagenesis step.

2.1.4. Subclone the reference and variant human cDNAs in the donor plasmid (Gateway plasmids with *attL1* and *attL2* sites) into the transgenic plasmid (e.g., pGW-HA.attB with *attR1* and *attR2* sites) via a LR clonase reaction.

NOTE: There are UAS ϕC31 vectors designed for conventional restriction enzyme-based subcloning (e.g., pUAST.attB[50]) if it is preferred to subclone human cDNAs via restriction enzyme methods.

2.1.5. Select ϕC31 docking sites in which to integrate the UAS-human cDNA transgenes. A number of docking sites have been generated by several laboratories and are publically available from stock centers[50,52,53].

NOTE: Since it is convenient to have the human transgene on a chromosome that does not contain the fly ortholog of the gene of interest, it is recommended to use a second chromosome docking site [VK37 (BDSC stock #24872] when the fly ortholog is on the X, third, or fourth chromosomes, and use a third chromosome docking site [VK33 (BDSC stock #24871] when the fly ortholog is on the second chromosome.

2.1.6. Inject UAS-human cDNA constructs into flies expressing ϕC31 integrase in their germline (e.g., vas-ϕC31, nos-ϕC31).

NOTE: Microinjection can be performed in-house or sent to core facilities or commercial entities for transgenesis. Detailed protocol for generating transgenic flies can be found in the cited book chapter[51].

2.1.7. Establish stable transgenic strains from the injected embryos. Inject ~100–200 embryos per construct[94]. A representative crossing scheme for a transgene insertion into a second chromosome docking site (VK37) is depicted in **Figure 1A**. Refer to the cited books[54,55] for basic *Drosophila* genetics information.

2.2. Obtain or generate a T2A-GAL4 line that facilitates rescue-based functional assays (see **Figure 2** and section 3.1).

NOTE: This line will serve two purposes. First, most T2A-GAL4 lines tested behave as strong LOF alleles by functioning as a gene trap allele. Second, T2A-GAL4 lines function as a GAL4 driver that allows expression of UAS constructs (e.g. UAS-GFP, UAS-human cDNAs) under the endogenous regulation elements of the gene of interest[56,57] (**Figure 2A–C**).

2.2.1 Search public stock collections for available T2A-GAL4 lines including the *Drosophila* Gene Disruption Project (GDP)[58] in which ~1,000 T2A-GAL4 lines have been generated[59]. These strains

are currently available from the Bloomington *Drosophila* Stock Center (BDSC) and are searchable through both the GDP and BDSC websites.

2.2.2. If a T2A-GAL4 line for the fly gene of interest is not available, check if a suitable coding intronic MiMIC (*Minos*-mediated integration cassette) line is available for conversion into a T2A-GAL4 line using recombinase-mediated cassette exchange (RMCE)[60] (**Figure 2A**).

NOTE**:** RMCE allows intronic MiMIC elements that are in between two coding exons to be converted into a T2A-GAL4 line through microinjection of a donor construct (an example of a crossing scheme is shown in **Figure 1B**) or series of crosses, as described in detail[57,59].

2.2.3. If a T2A-GAL4 line is not available and an appropriate coding intronic MiMIC does not exist, explore the possibility of generating a T2A-GAL4 line via the CRIMIC (CRISPR-mediated integration cassette) system[59].

NOTE: This methodology uses CRISPR-mediated DNA cleavage and homology-directed repair (HDR) to integrate a MiMIC-like cassette into a coding intron in a gene of interest.

2.2.4. If the gene of interest lacks a large intron (>150 bp) or has no introns, attempt to knock-in a GAL4 transgene into the fly gene with the CRISPR/Cas9 system using HDR as described[20,61,62].

NOTE: If generation of a T2A-GAL4 or GAL4 knock-in allele is difficult, attempt to perform rescue experiments using these pre-existing alleles or RNAi lines and ubiquitous or tissue-specific GAL4 drivers as described[92] (REF).

3. Performing functional analysis of human variant of interest in vivo in *Drosophila*

NOTE: Perform a rescue-based analysis (section 3.1) as well as overexpression studies (section 3.2) using the tools gathered or generated in section 2 to assess consequences of the variant of interest in vivo in *Drosophila*. Consider utilizing both approaches, since the two are complementary.

3.1. Performing functional analysis through rescue-based experiments

NOTE: Heterologous rescue-based experiments in *Drosophila* using human proteins determine whether the molecular function of the two orthologous genes have been conserved over ~500 million years of evolution. They also assess the function of the variant in the context of the human protein[63]. Although a systematic analysis studying hundreds of gene pairs has not been reported, several dozen human and mammalian (e.g., mouse) genes are able to replace the function of *Drosophila* genes[13].

3.1.1. In the rescue-based approach, first determine whether there are obvious, scorable, and reproducible phenotypes in LOF mutants in the fly ortholog before assessing functions of the variants.

NOTE: Previous literature on the fly gene is a good place to data mine first, and it can be found using databases including FlyBase and PubMed. Additional databases such as MARRVEL, Monarch Initiative, and Gene2Funcion are also useful in gathering this information.

3.1.2. Perform a global survey of scorable phenotypes in homozygous and hemizygous (T2A-GAL4 allele over a molecularly defined chromosomal deficiency animals, especially if the T2A-GAL4 allele is the first mutation to be characterized for a specific gene. Assess phenotypes such as lethality, sterility, longevity, morphological (e.g., size and morphology of the eye) and behavior (e.g., courtship, flight, climbing, and bang sensitivity defects).

NOTE: If there are no major phenotypes identified from this primary screen, more subtle phenotypes such as neurological defects measured by electrophysiological recordings can be used if they are highly reproducible and specific. As an example, functional studies using electroretinogram (ERG) are described in step 3.2.3. If there is failure to detect any scorable phenotype, move to section 3.2 and perform the overexpression-based functional study.

3.1.3. Once a scorable phenotype is identified in the fly LOF mutant, test whether the reference human cDNA can replace the function of the fly ortholog by attempting to use human cDNA to rescue the mutant fly line. The phenotypic assay to be performed here depends on the results from step 3.1.2 and will be specific to the gene being studied.

NOTE: If “humanization” of the fly gene is successful, there is now a platform to compare the efficiency of rescue for the variant of interest compared to the reference counterpart. The rescue seen with reference human cDNA does not have to be perfect. Partial rescue of the fly mutant phenotype using a human cDNA still provides a reference point to perform comparative studies using the variant human cDNA strain.

3.1.4. Using the assay system selected in step 3.1.2, compare the rescue observed with the reference human cDNA to the rescue observed with the variant human cDNA to determine if the variant of interest has consequences on the gene of interest.

NOTE: If the variant human cDNA performs worse than the reference allele, this suggests that the variant of interest is deleterious to the protein function. If the variant and reference cannot be functionally distinguished, then 1) the allele may be an isomorph (a variant that does not affect protein function) or 2) the assay is not sensitive enough to detect subtle differences.

3.1.5. If the variant is found to be a deleterious allele, then further assess the expression and intracellular localization of the reference and variant proteins of interest in vivo via western blot, immunofluorescence staining, or other methods[93].

NOTE: If the UAS-human cDNA is generated from an open clone in a pGW-HA.attB vector, use an anti-HA antibody to perform these biochemical and cellular assays. If the original clone is a closed clone, test whether commercial antibodies against the human proteins can be used for

these assays. A difference in expression levels and intracellular localization may reveal how the variant of interest affects protein function.

3.2. Performing functional analysis through overexpression studies

NOTE: Ubiquitous or tissue-specific overexpression of human cDNAs in otherwise wild-type flies can provide information that is complementary to the rescue-based experiments discussed in section 3.1. While rescue-based assays are primarily designed to detect LOF variants (amorphic, hypomophic), overexpression-based assays may also reveal gain-of-function (GOF) variants that are more difficult to assess (hypermorphic, antimorphic, neomorphic).

3.2.1. Select a set of GAL4 drivers to overexpress the human cDNAs of interest. A number of ubiquitous and tissue-/stage-specific GAL4 drivers are available from public stock centers (e.g., BDSC), some of which are more frequently used than others. First focus on ubiquitous drivers and easily scorable phenotypes (lethality, sterility, morphological phenotypes), then move on to tissue-specific drivers and more specific phenotypes.

NOTE: Validate the expression of GAL4 drivers with a reporter line (e.g., UAS-GFP) to confirm expression patterns before use in the experiments.

3.2.2. Express the reference and variant human cDNAs using the same driver under the same condition (e.g., temperature) by crossing 3–5 virgin females from the GAL4 line (e.g. *ey-GAL4* for eye imaginal disc expression) with 3–5 male flies from the UAS lines [e.g., *UAS-TBX2(+)* or *UAS-TBX2(p.R305H)* for the example shown in section 4.2] in a single vial.

3.2.2.1. Transfer the crosses every 2–3 days to have many animals eclosing from a single cross.

3.2.3. Examine the progeny and detect any differences between the reference and variant strains (e.g., eye morphology) under a dissection microscope. Image the flies using a camera attached to a dissection microscope to document phenotype.

NOTE: If a phenotype is only seen in the reference but not in the variant line, then the variant may be an amorphic or a strong hypomorphic allele. If the phenotype is seen in both genotypes but the reference causes a stronger defect, then the variant may be a mild-to-weak hypomorphic allele. If the reference does not show a phenotype or only exhibits a weak phenotype but the variant shows a strong defect, then the variant may be a GOF allele.

3.2.4. If a phenotype is not seen in standard culture conditions (RT or at 25°C, then set the crosses at different temperatures ranging 18–29 °C, since the UAS/GAL4 system is known to be temperature-sensitive[64,65]). Typically, the expression of UAS transgenes is higher at higher temperatures.

3.3 Perform additional functional studies related to the genes and protein of interest.

NOTE: In addition to examining general defects, an assay system can be selected to probe into molecular functions of the gene and variant. In one example discussed in the representative results (*TBX2* case), ERG recordings were used to determine effects of the variant on photoreceptor function, since the fly gene of interest (*bifid*) had been studied extensively in the context of visual system development. Detailed protocols for ERG in *Drosophila* can be found as previously published[66–68].

3.3.1. Generate flies to test for functional defects in the visual system. Cross virgin females from the *Rhodopsin 1 (Rh1)-GAL4* line to males with reference or variant UAS-human cDNA transgenes to express the human proteins of interest in the R1-R6 photoreceptors.

NOTE: Cross 3–5 virgin females to 3–5 male flies in a single vial and transfer the crosses every 2–3 days to have many animals eclosing from a single cross. All crosses must be kept in an incubator set at the experimental temperature to obtain consistent results.

3.3.2. Once flies begin to eclose (at 25°C, ~10 days after setting the initial cross), gather the progeny (*Rh1-GAL4/+; UAS-human cDNA/+*) into fresh vials and place them back into the incubator set at the experimental temperature for an additional 3 days for the visual system to mature.

NOTE: Although ERG can be performed on 1–2 day old flies, newly eclosed flies may have large fluctuations in their ERG signal. If it is desired to examine an age-dependent phenotype, these flies can be aged for several weeks as long as they are regularly (e.g., every ~5 days) transferred to a new vial to avoid drowning in wet food.

3.3.3. Prepare the flies for ERG recording by first anesthetizing the flies using CO_2 or placing them into a vial on ice. Gently glue one side of the fly onto a glass microscope slide to immobilize them.

NOTE: Multiple reference and variant flies can be glued on to a single slide. Place all flies in approximately the same orientation with one eye being accessible for the recording electrode. Be careful not to get glue on the eye and to leave the proboscis free.

3.3.4. Prepare the recording and reference electrodes. Place a 1.2 mm glass capillary into a needle puller and activate. Break the capillary tube to obtain two sharp tapered electrodes. The resulting electrodes should be hollow and have final diameters of <0.5mm.

3.3.5. Fill the capillaries with saline solution (100 mM NaCl), making sure there are no air bubbles. Slide the glass capillaries over the silver wire electrodes (both the recording electrode and reference electrode, see **Figure 4**) and secure the capillaries in place.

3.3.6. Configure the stimulator and amplifier. A detailed set-up can be found in Lauwers, et al.[67]. The UDN *Drosophila* MOSC set-up consists of equipment listed in the materials section.

3.3.6.1. Set the amplifier to 0.1 Hz high-pass filter, 300 Hz low-pass filter, and 100 gain.

3.3.6.2. Set the stimulator to 1 s period, 500 ms pulse width, 500 ms pulse delay, run mode, and 7 Amp.

3.3.6.3. Prepare the light source for photostimulation. Use a halogen light source to activate the fly photoreceptors.

3.3.6.4. Prepare the recording software on a computer connected to the ERG set-up. Create a stimulation protocol with acquisition model “fixed length events” and 20 s duration.

3.3.7. Acclimate the flies to complete darkness before initiating ERG recordings. Place the flies into complete darkness for at least 10 min before beginning the experiment.

NOTE: Since flies cannot detect red light, use a red light source during the period of dark habituation.

3.3.8. Place the slide containing the flies onto the recording apparatus and move the micromanipulators carrying the reference and recording electrodes to a point close to the fly of interest on the slide. Watch the tip of the electrode and carefully place the reference electrode into the thorax of the fly (penetrate the cuticle). Once the reference electrode is stably inserted, place the recording electrode on the surface of the eye.

NOTE: The exact position of this reference electrode does not have a major impact on the ERG signal. The recording electrode should be placed at the surface of the eye since ERG is a field potential recording rather than an intracellular recording. The proper amount of pressure applied to the recording electrode causes a small dimple without penetrating the eye.

3.3.9. Turn off all lights for another 3 min to acclimate the flies again to the dark environment. If using a halogen light source with a manual shutter, turn on the light source at this point with the shutter closed (flies are still in dark).

3.3.10. Set up the recording software and begin the recording the signal.

3.3.11. Expose the fly eyes to light by opening and closing the shutter every 1 s for the 20 s duration of a single run.

NOTE: The on/off of the halogen light source can be controlled manually or programmed to become automated using a white LED light source. More robust and reliable ERG can be obtained by using a halogen light source compared to a white light LED, likely due to the broader light spectrum emitted from the halogen light source.

3.3.12. Record ERGs from all flies that are mounted on the glass slide. Perform ERGs from 15 flies per genotype per condition.

NOTE: Some parameters that can be altered to find a condition that shows robust differences between reference and variant cDNAs may include temperature, age, or environmental conditions (e.g., reared in light-dark cycle or constant light/darkness).

3.3.13. Perform data analysis: compare the ERGs from the reference, variant, and controls to determine if there are differences. Assess the ERG data for changes in on-transients, depolarization, off-transients, and repolarization[69] (**Figure 4B**).

NOTE: Depolarization and repolarization reflect the activation and inactivation of the phototransduction cascade within photoreceptors, whereas on- and off-transients are measures of the activities of post-synaptic cells that receive signals from the photoreceptors. Decreased amplitude and altered kinetics of repolarization are often associated in defects with photoreceptor function and health, whereas defects in on- and off-transients are found in mutants with defective synapse development, function, or maintenance[70].

3.3.14. Upon identification of differences in ERG phenotypes with overexpression of reference vs. variant human cDNAs, determine whether this electrophysiological phenotype is associated with structural and ultrastructural defects in photoreceptors and their synapses by performing histological analysis and transmission electron microscopy.

NOTE: Further discussion on interpretation of ERG defects and structural/ultrastructural analysis has been previously described[69].

REPRESENTATIVE RESULTS:

Functional study of *de novo* missense variant in *EBF3* linked to neurodevelopmental phenotypes

In a 7-year-old male with neurodevelopmental phenotypes including hypotonia, ataxia, global developmental delay, and expressive speech disorder, physicians and human geneticists at the National Institutes of Health Undiagnosed Diseases Project (UDP) identified a *de novo* missense variant (p.R163Q) in *EBF3 (Early B-Cell Factor 3*)[15], a gene that encodes a COE (Collier/Olfactory-1/Early B-Cell Factor) family transcription factor. This case was submitted to the UDN MOSC in March 2016 for functional studies. To assess whether this gene was a good candidate for this case, the MOSC gathered human genetic and genomic information from OMIM, ClinVar, ExAC (now expanded to gnomAD), Geno$_2$MP, DGV, and DECIPHER. In addition, the orthologous genes in key MO species were identified using the DIOPT tool. Gene expression and phenotypic information from individual MO databases (e.g., Wormbase, FlyBase, ZFIN, MGI) were then obtained. The informatics analyses performed for *EBF3* and other pioneering studies in the UDN MOSC formed the basis for later development of the MARRVEL resource[30].

The information gathered using this methodology indicated *EBF3* was not associated with any known human genetic disorder at the time of analysis, and it was concluded that the p.R163Q variant was a good candidate based on the following information. (1) This variant had not been

previously reported in control population databases (ExAC) and disease population database ($Geno_2MP$), indicating that this is a very rare variant. (2) Based on ExAC, the pLI (probability of LOF intolerance) score of this gene is 1.00 (pLI scores range from 0.00–1.00). This indicates that there is selective pressure against LOF variants in this gene in the general population and suggests that haploinsufficiency of this gene may cause disease. For more information on pLI score and its interpretation, an accompanying MARRVEL tutorial article in JoVE[31] and related papers provide details[30,71].

The p.R163Q variant was also considered a good candidate because (3) it is located in the evolutionarily conserved DNA binding domain of this protein, suggesting that it may affect DNA binding or other protein functions. (4) The p.R163 residue is evolutionarily conserved from *C. elegans* and *Drosophila* to humans, suggesting that it may be critical for protein functional across species. (5) *EBF3* orthologs have been implicated in neuronal development in multiple MO[72] including *C. elegans*[73], *Drosophila*[74], Xenopus[75], and mice[76]. (6) During brain development in mice, *Ebf3* has been shown to function downstream of *Arx* (*Aristaless-related homebox*)[77], a gene associated with several epilepsy and intellectual disability syndromes in humans[78]. Hence, these data together suggest that *EBF3* is highly likely to be crucial to human neurodevelopment and that the p.R163Q variant may have functional consequences.

To assess whether p.R163Q affects EBF3 function, a T2A-GAL4 line for *knot (kn*; the fly ortholog of human *EBF3*[79]) was generated via RMCE of a coding intronic MiMIC insertion[15]. The $kn^{T2A\text{-}GAL4}$ line is recessive lethal and failed to complement the lethality of a classic *kn* allele ($kn^{col\text{-}1}$) as well as molecularly defined deficiency that covers *kn [Df(2R)BSC429]*[80]. Expression patterns of the GAL4 also reflected previously reported patterns of *kn* expression in the brain as well as in the wing imaginal disc[15]. UAS transgenic flies were then generated to allow the expression of reference and variant human *EBF3* cDNA as well as wild-type fly *kn* cDNA. All three proteins were tagged with a C-terminal 3xHA tag. Importantly, UAS wild-type fly *kn* (kn^{+}) or reference human *EBF3* ($EBF3^{+}$) transgenes rescued the lethality of $kn^{T2A\text{-}GAL4}/Df(2R)BSC429$ to a similar extent (**Figure 3C,** left panel)[81].

In contrast, UAS-human *EBF3* transgene with the p.R163Q variant ($EBF3^{p.R163Q}$) was not able to rescue this mutant, suggesting that the p.R163Q variant affects EBF3 function in vivo[15]. Interestingly, when assessed using an anti-HA antibody, the $EBF3^{p.R163Q}$ protein was successfully expressed in the fly tissues, and its levels and subcellular localization (primarily nuclear) were indistinguishable from those of $EBF3^{+}$ and Kn^{+}. This suggests that the variant is not causing a LOF phenotype due to protein instability or mis-localization. To further assess whether the p.R163Q variant affected the transcriptional activation function of *EBF3*, a luciferase-based reporter assay was performed in HEK293 cells[15]. This experiment in cultured human cells revealed that the $EBF3^{p.R163Q}$ variant failed to activate transcription of the reporter constructs, supporting the LOF model obtained from *Drosophila* experiments.

In parallel to the experimental studies, collaborations with physicians, human geneticists, and genetic counselors at BCM led to the identification of two additional individuals with similar symptoms. One patient carried the identical p.R163Q variant, and another carried a missense

variant that affected the same residue (p.R163L). The p.R163L variant also failed to rescue the fly *kn* mutant[93] suggesting that this allele also affected EBF3 function. Interestingly, this work was published back-to-back with two independent human genetics studies that reported additional individuals with *de novo* missense, nonsense, frameshift, and splicing variants in *EBF3* linked to similar neurodevelopmental phenotypes[82,83]. Subsequently, three additional papers were published reporting additional cases of *de novo EBF3* variants and copy number deletion[84–86]. This novel neurodevelopmental syndrome is now known as the Hypotonia, Ataxia, and Delayed Development Syndrome (HADDS, MIM #617330) in the Online Mendelian Inheritance in Man (OMIM, an authoritative database for genotype-phenotype relationships in humans).

Functional study of dominantly inherited missense variant in *TBX2* linked to a syndromic cardiovascular and skeletal developmental disorder

In a small family affected with overlapping spectrums of craniofacial dysmorphism, cardiac anomalies, skeletal malformation, immune deficiency, endocrine abnormalities, and developmental impairment, the UDN Duke Clinical Site identified a missense variant (p.R20Q) in *TBX2* that segregates with disease phenotypes[87]. Three (son, daughter, mother) out of the four family members are affected by this condition, and the son exhibited the most severe phenotype. Clinically, he met a diagnosis of "complete DiGeorge syndrome", a condition often caused by haploinsufficiency of *TBX1*. While there were no mutations identified in *TBX1* in this family, the clinicians and human geneticists focused on a variant in *TBX2,* since previous studies in mice showed that these genes have overlapping functions during development[88]. TBX1 and TBX2 both belong to T-box (TBX) family of transcription factors that can act as transcriptional repressors as well as activators depending on the context.

Previously, variants in 12 out of 17 members of the *TBX* family genes were linked to human diseases. The MOSC decided to experimentally pursue this variant based on the following information gathered through MARRVEL and other resources. (1) This variant was reported only once in a cohort of ~90,000 "control" individuals in gnomAD (this variant was filtered out in a default view, likely due to low coverage reads). Considering the milder phenotypic presentation of the mother, this can still be considered as a rare variant that may be responsible for the disease phenotypes. (2) The pLI scores of *TBX2* in ExAC/gnomAD are 0.96/0.99, which is high (maximum = 1.00). In addition, the o/e (observed/expected) LOF score in gnomAD is 0.05 (only 1/18.6 expected LOF variant is observed in gnomAD). These numbers suggest that LOF variants in this gene are selected against in the general population.

Additionally, (3) the p.R20 is evolutionarily conserved from *C. elegans* and *Drosophila* to humans, suggesting that this may be an important residue for TBX2 function. (4) Multiple programs predict that the variant is likely damaging (polyphen: possibly/probably damaging, SIFT: deleterious, CADD score: 24.4, REVEL score: 0.5). (5) MO mutants exhibit defects in tissues affected in patients (e.g., knockout mice exhibiting defects in cardiovascular system, digestive/alimentary systems, craniofacial, limbs/digit). Hence, together with the biological links between *TBX1* and *TBX2* and phenotypic links between these patients and DiGeorge Syndrome, it was optimal to perform functional studies of variants in this gene using *Drosophila*.

To assess whether the p.R20Q variant affects *TBX2* function, a T2A-GAL4 line in *bifid* (*bi*; the *Drosophila* ortholog of human *TBX2*), was generated via RMCE of a coding intronic MiMIC (**Figure 2**)[87]. This allele, $bi^{T2A\text{-}GAL4}$, was recessive pupal lethal and behaved as a strong LOF mutant, similar to previously reported *bi* LOF alleles (e.g., bi^{D2}, bi^{D4}; **Figure 2E**). The lethality of these classic and newly generated *bi* alleles was rescued by an ~80 kb genomic rescue construct carrying the entire *bi* locus, indicating that these reagents are indeed clean LOF alleles. The expression pattern of GAL4 in the $bi^{T2A\text{-}GAL4}$ line also matched well with previously reported patterns of *bi* expression in multiple tissues including in the wing imaginal disc (**Figure 2D**).

In parallel, UAS-transgenic lines for *TBX2* carrying the reference or variant (p.R20Q) sequences were generated. Unfortunately, neither transgene was able to rescue lethality of the $bi^{T2A\text{-}GAL4}$ line. Importantly, a wild-type fly *UAS-bi* transgene also failed to rescue the $bi^{T2A\text{-}GAL4}$ allele, likely due to the dosage-sensitivity of this gene. Indeed, overexpression of *UAS-*bi^{+} and *UAS-*$TBX2^{+}$ caused some degree of lethality when overexpressed in a wild-type animal. This toxic effect of *bi/TBX2* overexpression was utilized as a functional assay to assess whether the p.R20Q variant may affect TBX2 function. Since the *Drosophila bi* gene has been extensively studied in the context of the visual system [gene is also known as *optomotor blind* (*omb*)], phenotypes related to the visual system were investigated extensively. When the reference *TBX2* was expressed using an *ey-GAL4* driver that expresses UAS-transgenes in the eye and parts of the brain relevant to the visual system, ~85% lethality (**Figure 3C, right panel**) and significant reduction of eye size (**Figure 4B**) were observed. This phenotype was stronger than the phenotype observed when a wild-type fly *UAS-bi* transgene was expressed, suggesting that the human TBX2 is more detrimental to the fly when overexpressed.

Interestingly, the p.R20Q TBX2 was less potent in causing lethality (**Figure 3C,** right panel) and in inducing a small eye phenotype (**Figure 4B**) using the same driver under the identical condition[87], suggesting that the variant affects protein function. Moreover, the function of photoreceptors overexpressing reference and variant *TBX2* using a different GAL4 driver, *(Rh1-GAL4)* that specifically expresses UAS transgenes in R1-R6 photoreceptors, revealed that the variant TBX2 exhibited a much milder ERG phenotype compared to reference TBX2 (**Figure 4B**)[87]. Interestingly, most of the p.R20Q TBX2 protein was still found in the nucleus, similar to the reference protein, suggesting that the variant did not affect nuclear localization. A luciferase-based transcriptional repression assay in HEK293T cells showed that the p.R20Q was not able to effectively repress transcription of a reporter construct with palindromic T-box sites[87]. In addition, decreases in protein levels of $TBX2^{p.R20Q}$ were observed compared to $TBX2^{+}$, suggesting that the variant may affect translation or protein stability of TBX2, which in turn affects its abundance within a cell.

Additional patients with rare variants in *TBX2* were identified by clinicians at the UDN Duke Clinical Site in parallel with these experimental studies. An 8-year-old boy with a *de novo* missense (p.R305H) variant from an unrelated family exhibited many of the features found in the first family[87]. Additional functional studies in *Drosophila* and human cell lines revealed that the p.R305H variant also affects TBX2 function and protein levels, strongly suggesting that

defects in this gene likely underlie many phenotypes found in the two families. This disorder was recently curated as "vertebral anomalies and variable endocrine and T cell dysfunction" (VETD, MIM #618223) in OMIM. Identification of additional individuals with damaging variants in *TBX2* with overlapping phenotypes is critical to understanding the full spectrum of genotype-phenotype relationships for this gene in human disease.

FIGURE & TABLE LEGENDS:

Figure 1: **Injection and crossing scheme to generate UAS-human cDNA and T2A-GAL4 lines. (A)** Generation of UAS-human cDNA transgenes through microinjections and crosses. Crossing scheme to integrate the transgenes into a second chromosome docking site (VK37) using male flies in the first and second generation are shown as an example. Upon injection of the human cDNA ϕC31 transgenic construct (pGW-HA.attB) into early embryos that contain a germline source of ϕC31 integrase (labeled with 3xP3-GFP and 3xP3-RFP) and VK37 docking site [labeled with a *yellow*$^{+}$ (*y*$^{+}$) marker], transgenic events can be followed with the *white*$^{+}$ (*w*$^{+}$) minigene that is present in the transgenic vector. It is recommended to cross out the ϕC31 integrase by selecting against flies with GFP and RFP. The final stable stock can be kept as homozygotes or as a balanced stock if the chromosome carries a second site lethal/sterile hit mutation. Presence of second site lethal/sterile mutations on a transgenic constructs usually does not affect the outcome of functional studies as long as these transgenes are used in a heterozygous state (**Figure 3**). **(B)** Generation of T2A-GAL4 lines through microinjection and crosses. Crossing scheme to convert a second chromosome MiMIC insertion into a T2A-GAL4 element is shown as an example. By microinjecting an expression vector for ϕC31 integrase and RMCE vector for T2A-GAL4 (pBS-KS-attB2-SA-T2A-Gal4-Hsp70, an appropriate reading frame for the MiMIC of interest is selected. See the following papers for details[57,59] into embryos carrying a MiMIC in a coding intron in gene of interest, one can convert the original MiMIC into a T2A-GAL4 line. **Figure 2A** shows a schematic diagram of the RMCE conversion. The conversion event can be selected by screening against the *y*$^{+}$ marker in the original MiMIC cassette[60]. Since RMCE can occur in two directions, only 50% of the successful conversion event leads to successful production of GAL4, which can be detected by a UAS-GFP reporter transgene in the next generation. The final stable stock can be kept as homozygotes or as a balanced stock if the LOF of the gene is lethal/sterile.

Figure 2: **Conversion of MiMIC elements into T2A-GAL4 lines via RMCE**. **(A)** ϕC31 integrase facilitates the recombination between the two *attP* sites in the fly (top) and two *attB* sites flanking a T2A-GAL4 cassette shown as a circular vector (bottom). **(B)** Successful RMCE events lead to a loss of a selectable marker (*yellow+*) and insertion of the T2A-GAL4 cassette in the same orientation of the gene of interest. Since the RMCE event can happen in two orientations, only 50% of the RMCE reaction yields a desired product. An RMCE product inserted in the opposite orientation will not function as a gene-trap allele or express GAL4. Directionality of the construct must be confirmed via Sanger sequencing. **(C)** Transcription (top) and translation (bottom) of the gene of interest leads to generation of a truncated mRNA and protein due to the polyA signal present at the 3' end of the T2A-GAL4 cassette. The T2A is a ribosome skipping signal, which allows the ribosome to halt and reinitiate translation after this signal. This is used

to generate a GAL4 element that is not covalently attached to the truncated gene product of interest. The GAL4 will enter the nucleus and will facilitate the transcription of transgenes that are under control of UAS elements. UAS-GFP can be used as a gene expression reporter, and UAS-human cDNA can be used for rescue experiments via gene "humanization". **(D)** Shown is an example of a T2A-GAL4 element in *bi* driving expression of UAS-GFP (top). This expression pattern resembles a previously generated enhancer trap line for the same gene ($bi^{omb\text{-}GAL4}$; bottom). **(E)** Comparison of T2A-GAL4 allele of *bi* with previously reported LOF *bi* alleles. This figure has been adopted and modified from previous publications[57,87].

Figure 3: **Functional analysis of human variants using rescue-based (left) and overexpression-based (right) studies**. **(A)** (left panel): The function of *EBF3* variants was assessed with a rescue-based analysis of the fly *knot (kn)* LOF allele focusing on lethality/viability; (right panel): the function of variants in *TBX2* was assessed by performing overexpression of human *TBX2* transgenes in wild-type flies, focusing on lethality/viability, eye morphology, and electrophysiology phenotypes (**Figure 4**). **(B)** Crossing schemes to obtain the flies to be tested in the functional studies. It is advised to always use a neutral UAS element (e.g., *UAS-lacZ*, *UAS-GFP*) as a control experiment. **(C)** Representative results from functional studies of $EBF3^{p.R163Q}$ and $TBX2^{p.R20Q}$ variants, respectively, along with appropriate control experiments that are necessary to interpret the results. Both rescue-based analysis and overexpression studies reveal that the variants behave as amorphic or hypomorphic alleles. The lethality/viability data shown here are based on experimental data presented in previous publications[15,87].

Figure 4: **Functional analysis of a rare missense variant in human *TBX2* based on eye morphology and electroretinogram in *Drosophila***. **(A)** A schematic image showing the typical placement of recording and reference electrodes on the fly eye, along with a representative electroretinogram recording with four major components (on-transient, depolarization, off-transient, repolarization). **(B)** *TBX2* variant (p.R20Q) functions as a partial LOF allele based on overexpression studies in the fly eye using GAL4 drivers specific to the visual system (*ey-GAL4* and *Rh1-GAL4*). This showed that the reference TBX2 caused a strong morphological and electrophysiological phenotype compared to the variant protein. (Top panels): a severe reduction in eye size is seen upon overexpression of $UAS\text{-}TBX2^{+}$ with *ey-GAL4*. $UAS\text{-}TBX2^{p.R20Q}$. Driven with *ey-GAL4* also causes a smaller eye, but the phenotype is much milder. (Bottom panels): when $UAS\text{-}TBX2^{+}$ is expressed in core R1-R6 photoreceptors using Rh1-GAL4, there is a loss of on- and off-transients, reduced depolarization, and large abnormal prolonged depolarization after potential (PDA) phenotype, which is not seen in control flies. These phenotypes are not as severe as when UAS-$TBX2^{p.R20Q}$ is expressed using the same *Rh1-GAL4*. This figure has been adopted and modified from previous publications[69,87].

Table 1: Online resources related to this protocol.

DISCUSSION:

Experimental studies using *Drosophila melanogaster* provide a robust assay system to assess the consequences of disease-associated human variants. This is due to the large body of

knowledge and diverse genetic tools that have been generated by many researchers in the fly field over the past century[89]. Just like any other experimental system, however, it is important to acknowledge the caveats and limitations that exist.

Caveats associated with data mining

Although the first step in this protocol is to mine databases for information pertaining to a gene of interest, it is important to use it only as a starting point. For example, although *in silico* prediction of variant function provides valuable insights, these data should always be interpreted with caution. There are some instances in which all major algorithms predict that a human variant is benign, yet functional studies in *Drosophila* clearly demonstrated the damaging nature of such variant[24]. Similarly, although protein-protein interactions, co-expression, and structural modeling data are all insightful pieces of information, there may be pseudo-positive and pseudo-negative information present in these large -omics data sets. For example, some of the previously identified or predicted protein-protein interactions may be artificial or only seen in certain cell and tissue types.

In addition, there may be many false negative interactions not captured in these data sets, since certain protein-protein interactions are transient (e.g., enzyme-substrate interactions). Experimental validation is critical to demonstrating that certain genes or proteins genetically or physically interact in vivo in the biological context of interest. Similarly, structures predicted based on homology modeling should be treated as a model rather than solved structure. Although this information may be useful if it is found that an amino acid of interest is present in a structurally important part of the protein, negative data does not rule out the possibility that the variant may be damaging. Finally, some of the previously reported genotype-phenotype information should also be treated with caution, since some information archived in public databases may not be accurate. For example, some information in MO databases are based on experiments that have been well-controlled and performed rigorously, whereas others may come from a large screen paper with no further follow-up studies and stringent controls.

“Humanization” experiments using T2A-GAL4 strategy not always successful

While rescue- and overexpression-based functional studies using human cDNAs allow assessment of variants in the context of the human protein, this approach is not always successful. If a reference human cDNA cannot rescue the fly mutant phenotype, there are two probable explanations. The first possibility is that the human protein is nonfunctional or has significantly reduced activity in the context of a fly cell. This may be due to 1) reduced protein expression, stability, activity and/or localization or 2) a lack of compatibility with fly proteins that work in a multi-protein complex. Since the UAS/GAL4 system is temperature sensitive, the flies can be raised at a relatively high temperature (e.g., 29 °C) to see the possibility of a rescue in this condition. In addition, a UAS-fly cDNA construct and transgene as a positive control can be generated. If the variant of interest affects a conserved amino acid, the analogous variant can be introduced into the fly cDNA for functional study of the variant in the context of the fly ortholog. Although this is not necessary, it greatly helps the study in cases that using human cDNA transgenic lines give negative or inconclusive results (**Figure 3**).

The second possibility is that expression of the human protein causes some cellular- or organism-level toxicity. This may be due to antimorphic (e.g., acting as a dominant negative protein), hypermorphic (e.g., too much activity), or neomorphic (e.g., gain of a novel toxic function such as protein aggregation that is not always related to the endogenous function of the gene of interest) effects. In this case, keeping the flies in a low temperature (e.g., 18 °C) may alleviate some of these problems. Finally, there are some scenarios in which overexpression of a fly cDNA may not rescue the fly T2A-GAL4 line as seen in the *TBX2* example, likely due to trict dosage dependence of the gene product. To avoid overexpression of a protein of interest, the fly gene of interest can be modified directly via CRISPR, a genomic rescue construct can be engineered that contains the variant of interest, or rescue experiments can be performed using a LOF allele[21]. For small genes, "humanizing" the fly genomic rescue construct can be considered to test human variants that affect non-conserved amino acids[24]. In summary, alternate strategies should be considered when the humanization experiment does not allow for functional assessment of the variant of interest.

Interpreting negative and positive results

If 1) both the reference and variant human cDNAs rescue the fly mutant phenotypes to a similar degree and 2) there is no difference observed in all conditions tested, then it can be assumed that the variant is functionally indistinguishable in *Drosophila* in vivo. It is important to note, however, that this information is not sufficient to rule out that the variant of interest is non-pathogenic, since the *Drosophila* assay may not be sensitive enough or capture all potential functions of the gene/protein of interest relevant to humans. Positive data, on the other hand, is a strong indication that the variant has damaging consequences on protein function, but this data alone is still not sufficient to claim pathogenicity. The American College of Medical Genetics and Genomics (ACMG) has published a set of standards and guidelines to classify variants in human disease associated genes into "benign", "likely benign", "variant of unknown significance (VUS)", "likely pathogenic", and "pathogenic"[90]. Although this classification only applies to established disease-associated genes and is not directly applicable to variants in "genes of uncertain significance" (GUS), all individuals involved in human variant functional studies are strongly encouraged to adhere to this guideline when reporting variant function.

Extracting useful biological information when MO phenotypes do not model a human disease condition

It is important to keep in mind that overexpression-based functional assays have limitations, especially since some of the phenotypes being scored may have little relevance to the disease condition of interest. Similarly, phenotypes that are being assessed through rescue experiments may not have any direct relevance to the disease of interest. Since these experiments are conducted outside the endogenous contexts in an invertebrate system, they should not be considered disease models but rather as a gene function test using *Drosophila* as a "living test tube".

Even if the model organism does not mimic a human disease condition, scorable phenotypes used in rescue experiments can often provide useful biological insights into disease conditions. The concept of "phenologs (non-obvious homologous phenotypes)"[91] can be used to further

determine underlying molecular connections between *Drosophila* and human phenotypes. For example, morphological phenotypes in the fly wing, thorax, legs, and eyes are excellent phenotypic readouts for defects in Notch signaling pathway, an evolutionarily conserved pathway linked to many congenital disorders, including cardiovascular defects in humans[62]. By understanding the molecular logic behind certain phenotypes in *Drosophila*, it is possible to identify hidden biological links between genes and phenotypes in humans that are not yet understood.

Continuous communication with clinical collaborators

When working with clinicians to study the function of a rare variant found in patient, it is important to establish a strong collaborative relationship. Although clinical and basic biomedical researchers may share interests in the same genes/genetic pathways, there is a large cultural and linguistic (e.g., medical jargon, model organism-specific nomenclature) gap between the clinical and scientific fields. A strong, trust-based relationship between the two parties can be built through extensive communication. Furthermore, bidirectional communication is critical to establishing and maintaining this relationship. For example, in the two cases described in the representative results section, identification of additional patients with similar genotypes and phenotypes, as well as subsequent functional study, were critical to prove pathogenicity of the variants of interest. Even with strong functional data, researchers and clinicians often have difficulties convincing human geneticists that a variant identified in "n = 1" cases is the true cause of disease.

Once the MO researcher identifies that a variant of interest is damaging, it is critical to communicate back to clinical collaborators as soon as possible so they can actively try to identify matching cases by networking with other clinicians and human geneticists. Tools such as $Geno_2MP$ [Genotypes to Mendelian Phenotypes: a de-identified database of 9,650 individuals enrolled in the University of Washington's Center for Mendelian Genomics Study[41]; includes patients and family members suspected of having genetic disorders] can be searched to assess individuals that may have the same disorder. Then, the lead clinician can be contacted using a messaging feature.

Alternatively, GeneMatcher can be used, which is a matchmaking website for clinicians, basic researchers, and patients who share interests in the same genes to identify additional patients that carry rare variants. Since GeneMatcher is part of a larger integrative network of matchmaking websites called Matchmaker Exchange[42], additional databases around the world can be searched, including the Australian Genomics Health Alliance Patient Archive, Broad Matchbox, DECIPHER, MyGene2, and PhenomeCentral in a single GeneMatcher gene submission. Although participation in GeneMatcher is possible as a "researcher", it is recommended that basic scientists utilize this website with their clinical collaborators, since communication with other clinicians after a match requires certain levels of medical expertise.

ACKNOWLEDGMENTS:

We thank Jose Salazar, Julia Wang, and Dr. Karen Schulze for critical reading of the manuscript. We acknowledge Drs. Ning Liu and Xi Luo for the functional characterization of the *TBX2*

variants discussed here. Undiagnosed Diseases Network Model Organisms Screening Center was supported through the National Institutes of Health (NIH) Common Fund (U54 NS093793). H. T. C. was further supported by the NIH[CNCDP-K12 and NINDS (1K12 NS098482)], American Academy of Neurology (Neuroscience Research grant), Burroughs Wellcome Fund (Career Award for Medical Scientists), Child Neurology Society and Child Neurology Foundation (PERF Elterman grant), and the NIH Director's Early Independence Award (DP5 OD026426). M. F. W. was further supported by Simons Foundation (SFARI Award: 368479). S. Y. was further supported by the NIH (R01 DC014932), the Simons Foundation (SFARI Award: 368479), the Alzheimer's Association (New Investigator Research Grant: 15-364099), Naman Family Fund for Basic Research, and Caroline Wiess Law Fund for Research in Molecular Medicine. Confocal microscopy at BCM is supported in part by NIH Grant U54HD083092 to the Intellectual and Developmental Disabilities Research Center (IDDRC) Neurovisualization Core.

DISCLOSURES:

The authors have nothing to disclose.

REFERENES:

1. Boycott, K. M., et al. International Cooperation to Enable the Diagnosis of All Rare Genetic Diseases. *The American Journal of Human Genetics.* **100**, (5) 695–705 (2017).
2. Lupski, J. R., et al. Whole-Genome Sequencing in a Patient with Charcot–Marie–Tooth Neuropathy. *New England Journal of Medicine.* **362**, (13) 1181–1191 (2010).
3. Boycott, K. M., Vanstone, M. R., Bulman, D. E., MacKenzie, A. E. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nature Reviews Genetics.* **14**, (10) 681–691 (2013).
4. Yang, Y., et al. Molecular Findings Among Patients Referred for Clinical Whole-Exome Sequencing. *JAMA.* **312**, (18) 1870 (2014).
5. Lee, H., et al. Clinical Exome Sequencing for Genetic Identification of Rare Mendelian Disorders. *JAMA.* **312**, (18) 1880 (2014).
6. Coban-Akdemir, Z., et al. Identifying Genes Whose Mutant Transcripts Cause Dominant Disease Traits by Potential Gain-of-Function Alleles. *The American Journal of Human Genetics.* **103**, (2) 171–187 (2018).
7. Muller, H. J. Further studies on the nature and causes of gene mutations. *Proceedings of the Sixth International Congress of Genetics.* 213–255 (1932).
8. Ghosh, R., Oak, N., Plon, S. E. Evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. *Genome Biology.* **18**, (1) 225 (2017).
9. Adzhubei, I. A., et al. A method and server for predicting damaging missense mutations. *Nature Methods.* **7**, (4) 248–249 (2010).
10. Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M., Ng, P. C. SIFT missense predictions for genomes. *Nature Protocols.* **11**, (1) 1–9 (2016).
11. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J.l Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research.* doi:10.1093/nar/gky1016 (2018).
12. Choi, Y., Sims, G. E., Murphy, S., Miller, J. R.. Chan, A. P. Predicting the functional effect of amino acid substitutions and indels. *PloS ONE.* **7**, (10) e46688 (2012).

13. Wangler, M. F., et al. Model Organisms Facilitate Rare Disease Diagnosis and Therapeutic Research. *Genetics.* **207**, (1) 9–27 (2017).
14. Oriel, C., Lasko, P. Recent Developments in Using Drosophila as a Model for Human Genetic Disease. *International Journal of Molecular Sciences.* **19**, (7) 2041 (2018).
15. Chao, H.-T., et al. A Syndromic Neurodevelopmental Disorder Caused by De Novo Variants in EBF3. *American Journal of Human Genetics.* **100**, (1) 128–137 (2017).
16. Oláhová, M., et al. Biallelic Mutations in ATP5F1D, which Encodes a Subunit of ATP Synthase, Cause a Metabolic Disorder. *American Journal of Human Genetics.* **102**, (3) 494–504 (2018).
17. Liu, N., et al. Functional variants in TBX2 are associated with a syndromic cardiovascular and skeletal developmental disorder. *Human Molecular Genetics.* **27**, (14) 2454–2465 (2018).
18. Marcogliese, P. C., et al. IRF2BPL Is Associated with Neurological Phenotypes. *American Journal of Human Genetics.* **103**, (2) 245–260 (2018).
19. Ferreira, C. R., et al. A Recurrent De Novo Heterozygous COG4 Substitution Leads to Saul-Wilson Syndrome, Disrupted Vesicular Trafficking, and Altered Proteoglycan Glycosylation. *The American Journal of Human Genetics.* **103**, (4) 553–567 (2018).
20. Kanca, O., et al. *De novo* variants in *WDR37* are associated with epilepsy, colobomas and cerebellar hypoplasia. *Americal Journal of Human Genetics.* **Submitted**, (2019).
21. Luo, X., et al. Clinically severe CACNA1A alleles affect synaptic function and neurodegeneration differentially. *PLOS Genetics.* **13**, (7) e1006905 (2017).
22. Chung, H., et al. ACOX1 induces autoimmunity whereas a *de novo* gain of function variant induces elevated ROS and glial loss in humans and flies. *Cell Metabolism.* **Submitted**, (2019).
23. Yamamoto, S., et al. A Drosophila Genetic Resource of Mutants to Study Mechanisms Underlying Human Genetic Diseases. *Cell.* **159**, (1) 200–214 (2014).
24. Jakobsdottir, J., et al. Rare Functional Variant in TM2D3 is Associated with Late-Onset Alzheimer's Disease. *PLoS Genetics.* **12**, (10) e1006327 (2016).
25. Yoon, W. H., et al. Loss of Nardilysin, a Mitochondrial Co-chaperone for α-Ketoglutarate Dehydrogenase, Promotes mTORC1 Activation and Neurodegeneration. *Neuron.* **93**, (1) 115–131 (2017).
26. Harel, T., et al. Recurrent De Novo and Biallelic Variation of ATAD3A, Encoding a Mitochondrial Membrane Protein, Results in Distinct Neurological Syndromes. *American Journal of Human Genetics.* **99**, (4) 831–845 (2016).
27. Tan, K. L., et al. Ari-1 Regulates Myonuclear Organization Together with Parkin and Is Associated with Aortic Aneurysms. *Developmental Cell.* **45**, (2) 226–244.e8 (2018).
28. Ansar, M., et al. Visual impairment and progressive phthisis bulbi caused by recessive pathogenic variant in MARK3. *Human Molecular Genetics.* **27**, (15) 2703–2711 (2018).
29. Ansar, M., et al. Bi-allelic Loss-of-Function Variants in DNMBP Cause Infantile Cataracts. *The American Journal of Human Genetics.* **103**, (4) 568–578 (2018).
30. Wang, J. , et al. MARRVEL: Integration of Human and Model Organism Genetic Resources to Facilitate Functional Annotation of the Human Genome. *The American Journal of Human Genetics.* **100**, (6) 843–853 (2017).
31. Wang, J., Liu, Z., Bellen, H., Yamamoto, S. MARRVEL, a web-based tool that integrates human and model organism genomics information. *Journal of Visualized Experiments.* **Accepted**, (2019).

32. Mungall, C. J., et al. The Monarch Initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species. *Nucleic Acids Research.* **45**, (D1) D712–D722 (2017).
33. Hu, Y., Comjean, A., Mohr, S. E., Perrimon, N., Perrimon, N. Gene2Function: An Integrated Online Resource for Gene Function Discovery. *Genes|Genomes|Genetics.* **7**, (8) 2855–2858 (2017).
34. Ioannidis, N. M., et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *The American Journal of Human Genetics.* **99**, (4) 877–885 (2016).
35. Szklarczyk, D., et al. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Research.* **45**, (D1) D362–D368 (2017).
36. Hu, Y., et al. Molecular Interaction Search Tool (MIST): an integrated resource for mining gene and protein interaction data. *Nucleic Acids Research.* **46**, (D1) D567–D574 (2018).
37. Lawson, C. L., et al. EMDataBank unified data resource for 3DEM. *Nucleic Acids Research.* **44**, (D1) D396–D403 (2016).
38. Bienert, S., et al. The SWISS-MODEL Repository—new features and functionality. *Nucleic Acids Research.* **45**, (D1) D313–D319 (2017).
39. Webb, B., Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Current Protocols in Bioinformatics.* **54**, 5.6.1-5.6.37 (2016).
40. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols.* **10**, (6) 845–858 (2015).
41. Bamshad, M. J., et al. The Centers for Mendelian Genomics: A new large-scale initiative to identify the genes underlying rare Mendelian conditions. *American Journal of Medical Genetics Part A.* **158A**, (7) 1523–1525 (2012).
42. Sobreira, N. L. M., et al. Matchmaker Exchange. *Current Protocols in Human Genetics.* **95**, 9.31.1-9.31.15 (2017).
43. Temple, G., et al. The completion of the Mammalian Gene Collection (MGC). *Genome Research.* **19**, (12) 2324–2333 (2009).
44. Katzen, F. Gateway ®Recombinational cloning: a biological operating system. *Expert Opinion on Drug Discovery.* **2**, (4) 571–589 (2007).
45. Venken, K. J. T., He, Y., Hoskins, R. A., Bellen, H. J. P[acman]: A BAC Transgenic Platform for Targeted Insertion of Large DNA Fragments in D. melanogaster. *Science.* **314**, (5806) 1747–1751 (2006).
46. Bischof, J., et al. A versatile platform for creating a comprehensive UAS-ORFeome library in Drosophila. *Development (Cambridge, England).* **140**, (11) 2434–42 (2013).
47. Bischof, J., Sheils, E. M., Björklund, M., Basler, K. Generation of a transgenic ORFeome library in Drosophila. *Nature Protocols.* **9**, (7) 1607–1620 (2014).
48. Laible, M., Boonrod, K. Homemade site directed mutagenesis of whole plasmids. *Journal of Visualized Experiments.*(27) doi:10.3791/1135 (2009).
49. Balana, B., Taylor, N., Slesinger, P. A. Mutagenesis and Functional Analysis of Ion Channels Heterologously Expressed in Mammalian Cells. *Journal of Visualized Experiments* (44) doi:10.3791/2189 (2010).
50. Bischof, J., Maeda, R. K., Hediger, M., Karch, F., Basler, K. An optimized transgenesis system for Drosophila using germ-line-specific C31 integrases. *Proceedings of the National*

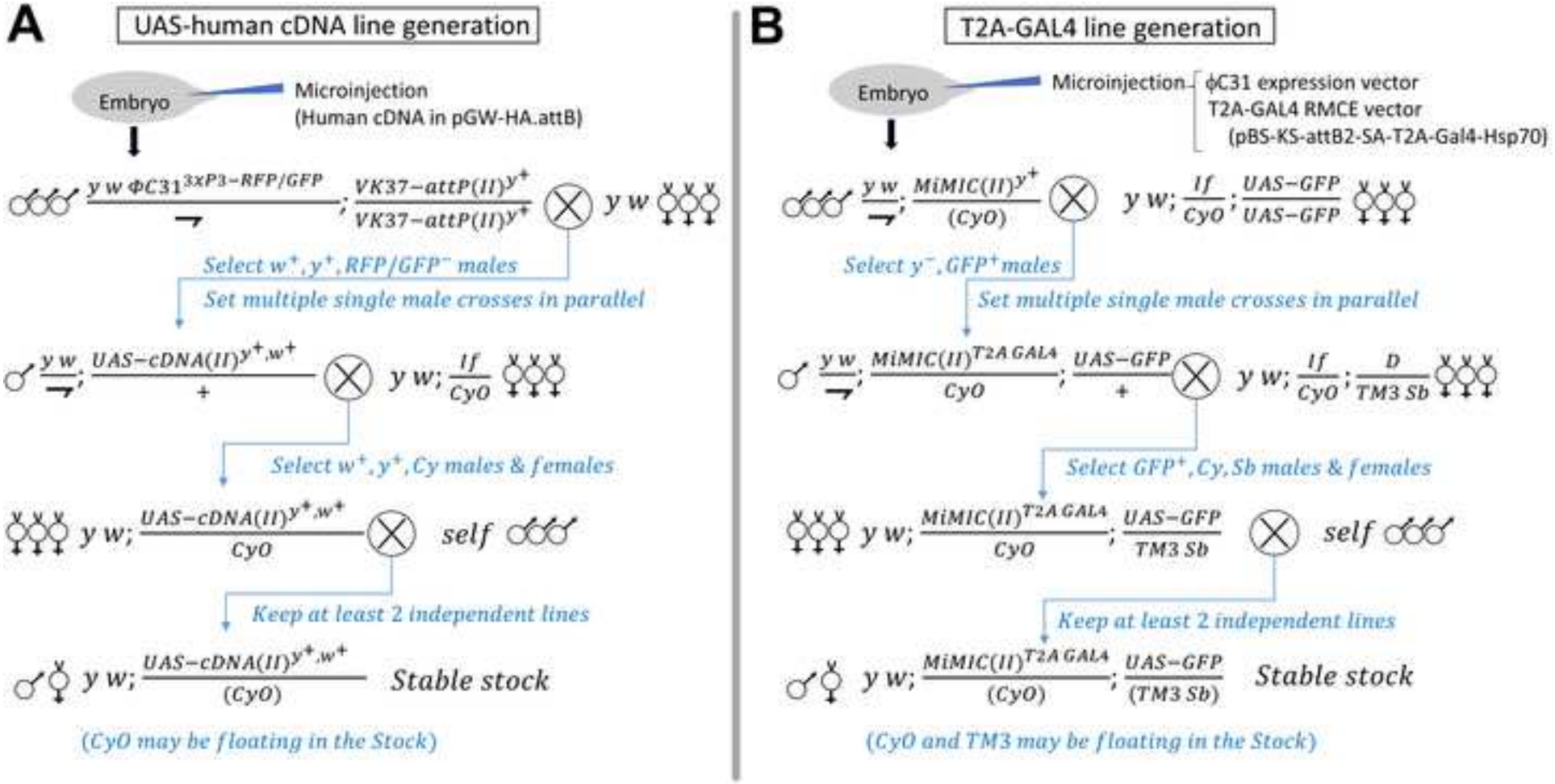
Academy of Sciences. **104**, (9) 3312–3317 (2007).
51. Ringrose, L. Transgenesis in Drosophila melanogaster. *Methods in Molecular Biology (Clifton, N.J.).* **561**, 3–19 (2009).
52. Venken, K. J. T., He, Y., Hoskins, R. A., Bellen, H. J. P[acman]: A BAC Transgenic Platform for Targeted Insertion of Large DNA Fragments in D. melanogaster. *Science.* **314**, (5806) 1747–1751 (2006).
53. Groth, A. C., Fish, M., Nusse, R., Calos, M. P. Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. *Genetics.* **166**, (4) 1775–82 (2004).
54. Greenspan, R. *Fly Pushing: The Thory and Practice of Drosophila Genetics*. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York (2004).
55. Ashburner, M., Golic, K., Hawley, R. S. *Drosophila: A Laboratory Handbook*. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York (2005).
56. Diao, F., White, B. H. A Novel Approach for Directing Transgene Expression in Drosophila: T2A-Gal4 In-Frame Fusion. *Genetics.* **190**, (3) 1139–1144 (2012).
57. Diao, F., et al. Plug-and-Play Genetic Access to Drosophila Cell Types using Exchangeable Exon Cassettes. *Cell Reports.* **10**, (8) 1410–1421 (2015).
58. Bellen, H. J., et al. The *Drosophila* Gene Disruption Project: Progress Using Transposons With Distinctive Site Specificities. *Genetics.* **188**, (3) 731–743 (2011).
59. Lee, P.-T., et al. A gene-specific T2A-GAL4 library for Drosophila. *eLife.* **7**, (2018).
60. Venken, K. J. T., et al. MiMIC: a highly versatile transposon insertion resource for engineering Drosophila melanogaster genes. *Nature Methods.* **8**, (9) 737–43 (2011).
61. Li-Kroeger, D., et al. An expanded toolkit for gene tagging based on MiMIC and scarless CRISPR tagging in Drosophila. *eLife.* **7**, (2018).
62. Salazar, J. L., Yamamoto, S. Integration of Drosophila and Human Genetics to Understand Notch Signaling Related Diseases. *Advances in Experimental Medicine and Biology.* **1066**, 141–185 (2018).
63. Wangler, M. F., Yamamoto, S., Bellen, H. J. Fruit Flies in Biomedical Research. *Genetics.* **199**, (3) 639–653 (2015).
64. Duffy, J. B. GAL4 system indrosophila: A fly geneticist's swiss army knife. *Genesis.* **34**, (1–2) 1–15 (2002).
65. Nagarkar-Jaiswal, S., et al. A library of MiMICs allows tagging of genes and reversible, spatial and temporal knockdown of proteins in Drosophila. *eLife.* **4**, (2015).
66. Dolph, P., Nair, A., Raghu, P. Electroretinogram Recordings of Drosophila. *Cold Spring Harbor Protocols*. (1) pdb.prot5549-pdb.prot5549 (2011).
67. Lauwers, E., Verstreken, P. Assaying Mutants of Clathrin-Mediated Endocytosis in the Fly Eye. *Methods in Molecular Biology (Clifton, N.J.).* **1847**, 109–119 (2018).
68. Rhodes-Mordov, E., Samra, H., Minke, B. Electroretinogram (ERG) Recordings from Drosophila. *Bio-Protocol.* **5**, (21) (2015).
69. Deal, S., Yamamoto, S. Unraveling novel mechanisms of neurodegeneration through a large-scale forward genetic screen in Drosophila. *Frontiers in Genetics.* **In press**, (2019).
70. Chouhan, A. K., et al. Uncoupling neuronal death and dysfunction in Drosophila models of neurodegenerative disease. *Acta Neuropathologica Communications.* **4**, (1) 62 (2016).
71. Lek, M., et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, (7616) 285–291 (2016).

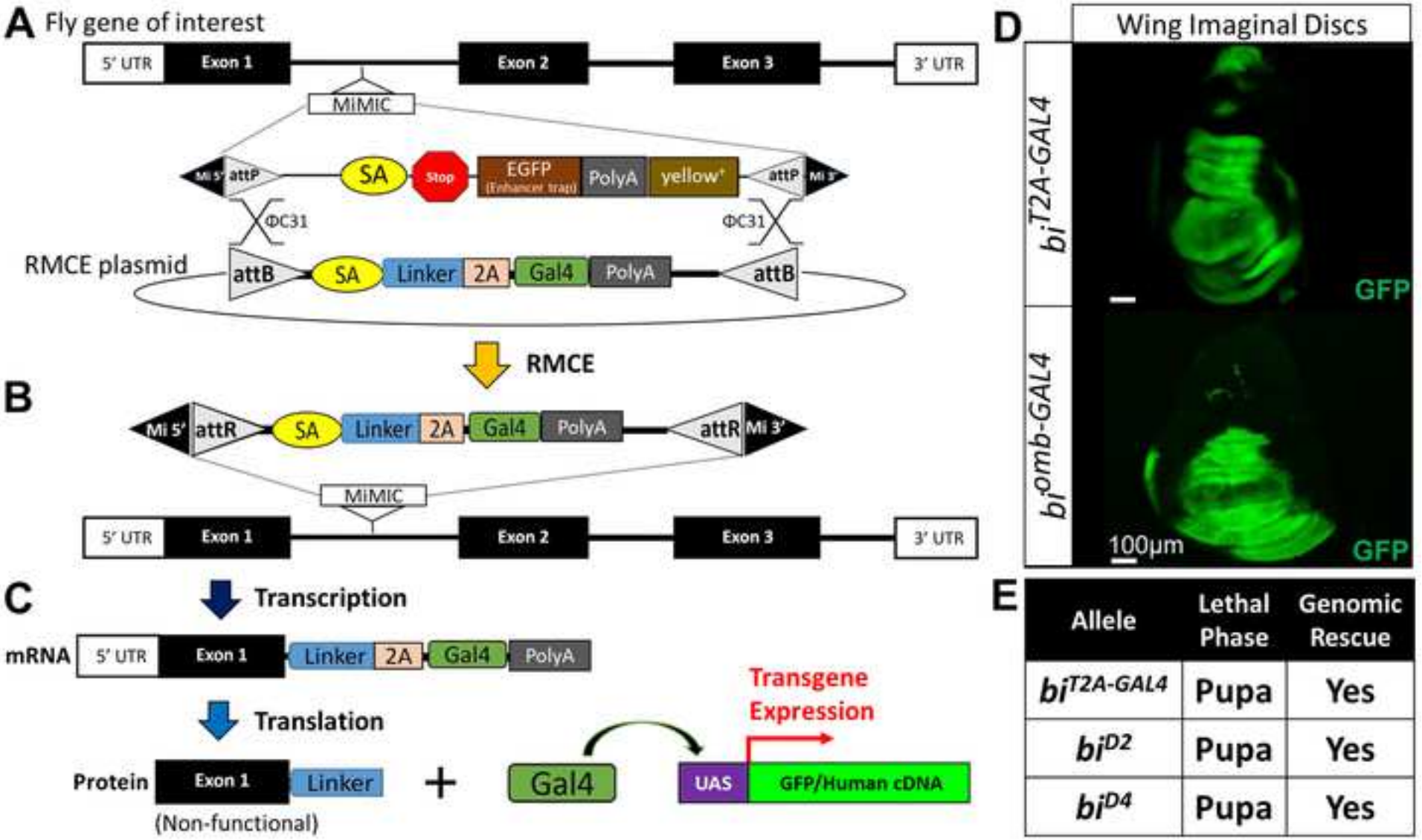
72. Liberg, D., Sigvardsson, M., Akerblad, P. The EBF/Olf/Collier family of transcription factors: regulators of differentiation in cells originating from all three embryonal germ layers. *Molecular and Cellular Biology.* **22**, (24) 8389–97 (2002).
73. Prasad, B. C., et al. Unc-3, a gene required for axonal guidance in Caenorhabditis elegans, encodes a member of the O/E family of transcription factors. *Development (Cambridge, England).* **125**, (8) 1561–8 (1998).
74. Jinushi-Nakao, S., et al. Knot/Collier and Cut Control Different Aspects of Dendrite Cytoskeleton and Synergize to Define Final Arbor Shape. *Neuron.* **56**, (6) 963–978 (2007).
75. Pozzoli, O., Bosetti, A., Croci, L., Consalez, G. G., Vetter, M. L. Xebf3 is a regulator of neuronal differentiation during primary neurogenesis in Xenopus. *Developmental Biology.* **233**, (2) 495–512 (2001).
76. Wang, S. S., Lewcock, J. W., Feinstein, P., Mombaerts, P., Reed, R. R. Genetic disruptions of O/E2 and O/E3 genes reveal involvement in olfactory receptor neuron projection. *Development.* **131**, (6) 1377–1388 (2004).
77. Fulp, C. T., et al. Identification of Arx transcriptional targets in the developing basal forebrain. *Human Molecular Genetics.* **17**, (23) 3740–3760 (2008).
78. Gécz, J., Cloosterman, D., Partington, M. ARX: a gene for all seasons. *Current Opinion in Genetics & Development.* **16**, (3) 308–316 (2006).
79. Dubois, L., Vincent, A. The COE--Collier/Olf1/EBF--transcription factors: structural conservation and diversity of developmental functions. *Mechanisms of Development.* **108**, (1–2) 3–12 (2001).
80. Cook, R. K., et al. The generation of chromosomal deletions to provide extensive coverage and subdivision of the Drosophila melanogaster genome. *Genome Biology.* **13**, (3) R21 (2012).
81. Chao, H.-T., et al. A Syndromic Neurodevelopmental Disorder Caused by De Novo Variants in EBF3. *The American Journal of Human Genetics.* **100**, (1) 128–137 (2017).
82. Sleven, H., et al. De Novo Mutations in EBF3 Cause a Neurodevelopmental Syndrome. *The American Journal of Human Genetics.* **100**, (1) 138–150 (2017).
83. Harms, F. L., et al. Mutations in EBF3 Disturb Transcriptional Profiles and Cause Intellectual Disability, Ataxia, and Facial Dysmorphism. *The American Journal of Human Genetics.* **100**, (1) 117–127 (2017).
84. Tanaka, A. J., et al. De novo variants in *EBF3* are associated with hypotonia, developmental delay, intellectual disability, and autism. *Molecular Case Studies.* **3**, (6) a002097 (2017).
85. Blackburn, P. R., et al. Novel de novo variant in *EBF3* is likely to impact DNA binding in a patient with a neurodevelopmental disorder and expanded phenotypes: patient report, in silico functional assessment, and review of published cases. *Molecular Case Studies.* **3**, (3) a001743 (2017).
86. Lopes, F., Soares, G., Gonçalves-Rocha, M., Pinto-Basto, J., Maciel, P. Whole Gene Deletion of EBF3 Supporting Haploinsufficiency of This Gene as a Mechanism of Neurodevelopmental Disease. *Frontiers in Genetics.* **8**, 143 (2017).
87. Liu, N., et al. Functional variants in TBX2 are associated with a syndromic cardiovascular and skeletal developmental disorder. *Human Molecular Genetics.* **27**, (14) 2454–2465 (2018).
88. Mesbah, K., et al. Identification of a Tbx1/Tbx2/Tbx3 genetic pathway governing pharyngeal and arterial pole morphogenesis. *Human Molecular Genetics.* **21**, (6) 1217–1229 (2012).

89. Bellen, H. J., Yamamoto, S. Morgan's legacy: fruit flies and the functional annotation of conserved genes. *Cell.* **163**, (1) 12–4 (2015).
90. Richards, S., et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine.* **17**, (5) 405–423 (2015).
91. McGary, K. L., Park, T. J., Woods, J. O., Cha, H. J., Wallingford, J. B., Marcotte, E. M. Systematic discovery of nonobvious human disease models through orthologous phenotypes. *Proceedings of the National Academy of Sciences of the United States of America.* **107**, (14) 6544–9 (2010).
92. Yamamoto S., et al., A Drosophila Genetic Resource of Mutants to Study Mechanisms Underlying Human Genetic Diseases. *Cell.* **159**, 200-214 (2014).
93. Ausubel, F. M. *Current Protocols in Molecular Biology*. Greene Publishing and Wiley-Interscience: New York (1989).
94. Rubin, G., Spradling, A. Genetic transformation of Drosophila with transposable element vectors. *Science.* **218** (4570), 348–353, doi:10.1126/science.6289436 (1982).
95. Sun, Y., Sriramajayam, K., Luo, D., Liao, D. J. A Quick, Cost-Free Method of Purification of DNA Fragments from Agarose Gel. *Journal of Cancer.* **3**, 93–95, doi:10.7150/jca.4163 (2012).
96. Ronaghi, M. DNA Sequencing :A Sequencing Method Based on Real-Time Pyrophosphate. *Science.* **281** (5375), 363–365, doi:10.1126/science.281.5375.363 (1998).
97. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., Pease, L. R. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene.* **77** (1), 51–59, doi:10.1016/0378-1119(89)90358-2 (1989).

Fig1

Click here to access/download;Figure;Figure 1.tif

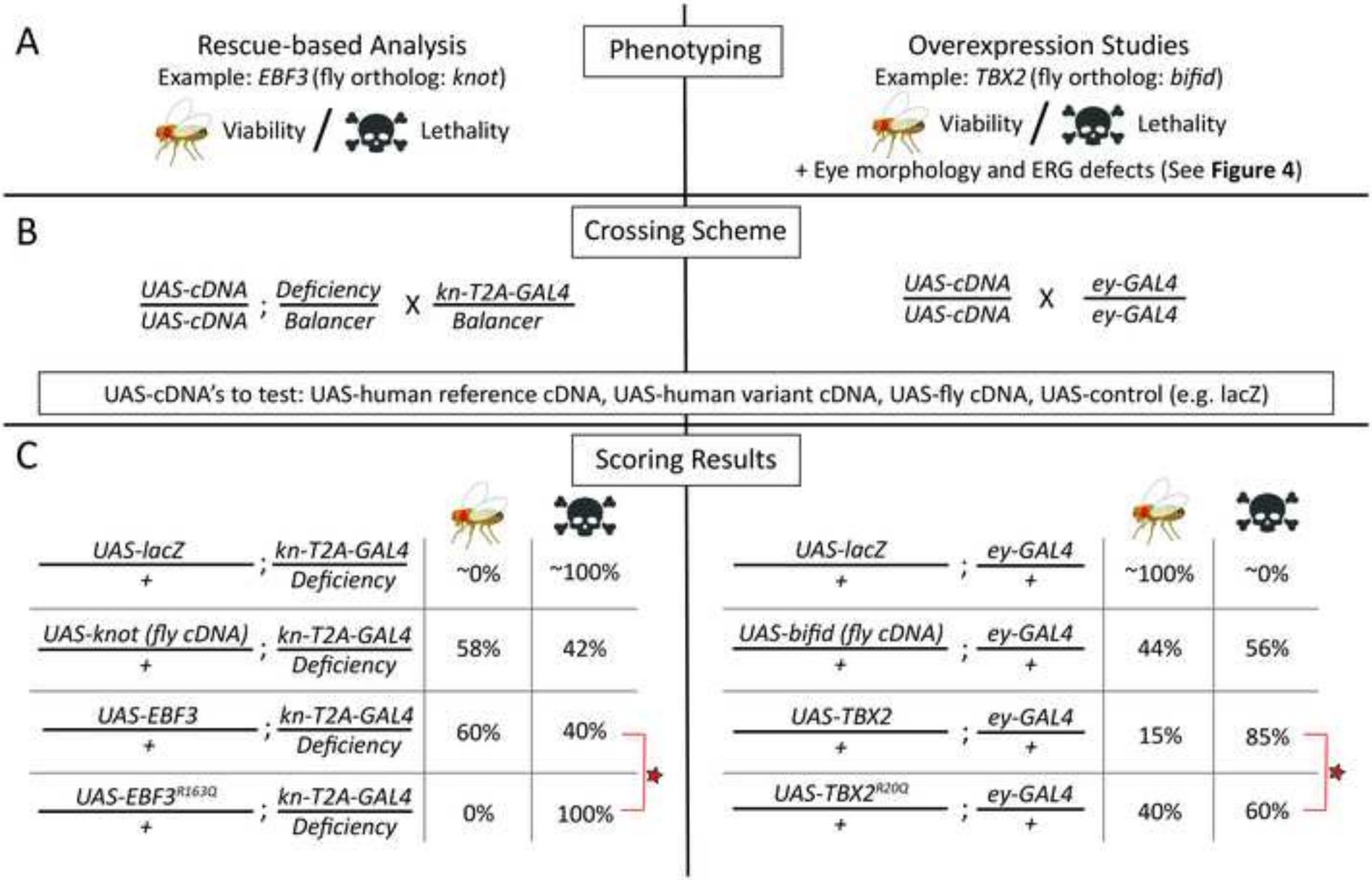


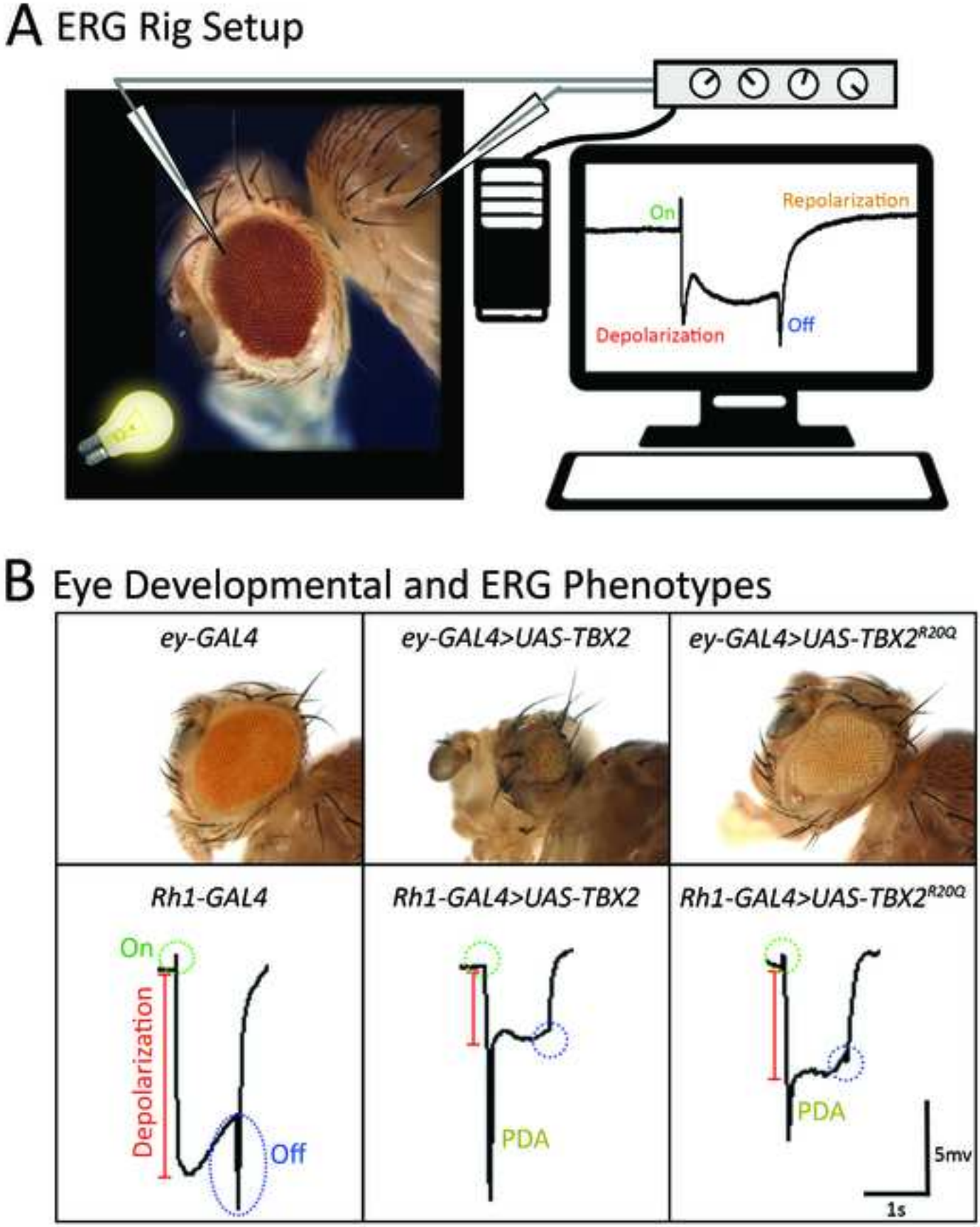


E

Allele	Lethal Phase	Genomic Rescue
$bi^{T2A\text{-}GAL4}$	Pupa	Yes
bi^{D2}	Pupa	Yes
bi^{D4}	Pupa	Yes

Fig3





Purpose	Tool
Variant function prediction algorithms	PolyPhen-2
	SIFT
	CADD
	PROVEAN
	MutationTaster
	REVEL
Rare and undiagnosed disease research consortia	UDN
	RDMM
	IRUD
	SOLVE-RD
	AFGN
Integrative database for human and model organism Information	MARRVEL
	Monarch Initiative
	Gene2Function
	Phenologs
Human Genetic and Genomics Databases	OMIM
	ClinVar
	ExAC
	gnomAD
	GenoMP
	DGV
	DECIPHER
Ortholog Identification Tool	DIOPT
Model Organism Databases and Biomedical Literature Search	WormBase (*C elegans*)
	FlyBase (*Drosophila*)
	ZFIN (Zebrafish)
	MGI (Mouse)
	Pubmed
Genetic and protein interaction databases	STRING
	MIST
Protein structure databases and modeling tools	WWPBD
	SWISS-MODEL
	Modeller
	Phyre^2
Patient matchmaking platforms	Matchmaker Exchange
	GeneMatcher
	AGHA Archive
	matchbox
	DECIPHER
	MyGene^2
	Phenome Central
Human transcript annotation and cDNA clone information	Mammalian Gene Collection
	Ensembl
	Refseq

URL
http://genetics.bwh.harvard.edu/pph2
https://sift.bii.a-star.edu.sg
https://cadd.gs.washington.edu
http://provean.jcvi.org/index.php
http://www.mutationtaster.org
https://sites.google.com/site/revelgenomics
https://undiagnosed.hms.harvard.edu
http://www.rare-diseases-catalyst-network.ca
https://irudbeyond.nig.ac.jp/en/index.html
http://solve-rd.eu
https://www.functionalgenomics.org.au
http://marrvel.org
https://monarchinitiative.org
http://www.gene2function.org
http://www.phenologs.org
https://www.omim.org/
https://www.ncbi.nlm.nih.gov/clinvar/
http://exac.broadinstitute.org/
http://gnomad.broadinstitute.org/
http://geno2mp.gs.washington.edu/Geno2MP/#/
http://dgv.tcag.ca/dgv/app/home
https://decipher.sanger.ac.uk/
https://www.flyrnai.org/cgi-bin/DRSC_orthologs.pl
https://www.wormbase.org
http://flybase.org
https://zfin.org
http://www.informatics.jax.org
https://www.ncbi.nlm.nih.gov/pubmed/
https://string-db.org
http://fgrtools.hms.harvard.edu/MIST/
http://www.wwpdb.org
https://swissmodel.expasy.org/
https://salilab.org/modeller/
http://www.sbg.bio.ic.ac.uk/phyre2
http://www.matchmakerexchange.org
https://www.genematcher.org
https://mme.australiangenomics.org.au/#/home
https://seqr.broadinstitute.org/matchmaker/matchbox
https://decipher.sanger.ac.uk
https://www.mygene2.org/MyGene2
https://phenomecentral.org
https://genecollections.nci.nih.gov/MGC
http://useast.ensembl.org
http://www.ncbi.nlm.nih.gov/refseq

Reagent Purpose	Specific Reagent
***Drosophila* Stocks for UAS-human cDNA transgenesis**	
Injection strains for transgenesis (*D. melanogaster*)	VK33 (3^{rd} chromosome) Injection line
	VK37 (2^{nd} chromosome) Injection line
Plasmid DNA	
Cloning vector	pDONR221
Drosophila transgenesis vector	pGW-HA.attB
Molecular biology kits and reagents	
Agarose	Agarose (molecular biology grade)
Chemically Competent Cells (*E. coli*)	DH5α
DNA Gel Extraction kit	PureLink Gel Extraction Kit
DNA Isolation and purification kit	QIAprep Spin Miniprep Kit
High Fidelity Polymerase	Q5 Polymerase kit
Recombinase mediated cloning system	Gateway BP Clonase kit
	Gateway LR Clonase II Enzyme kit
Site Directed Mutagenesis kit	Quick Change II Mutagenesis kit
Electroretinogram Rig related equipment	
ERG Analysis	Axon pCLAMP 10 Data Software Package
ERG Data Collection	ISO-DAM Isolated Biologic Amplifier
ERG Stimulator	Square Pulse Stimulator

Vendor	Catolog #
BDSC	#24871
BDSC	#24872
Thermo Fisher	#12536-017
Gift from Drs. Johannes Bischof and Konrad Basler (Bischof et al., 2013 PNAS)	
Sigma-Aldrich	#A2790
Thermo Fisher	#18265017
Thermo Fisher	#K210012
Qiagen	#27104
NEB	#M0491
Thermo Fisher	#11789020
Thermo Fisher	#11791100
Agilent	#200523
Molecular Devices	N/A
LabX	#R150358
Astro-Med	#S48

jove JOURNAL OF VISUALIZED EXPERIMENTS

1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

Title of Article: In vivo functional Study of disease-associated human variants using Drosophila

Author(s): J. Michael Hamish, Samantha L. Deal, UDN, Hsiao Tuan Chao, Michael F. Wangler, Shinya Yamamoto

Item 1 (check one box): The Author elects to have the Materials be made available (as described at http://www.jove.com/author) via: ☑ Standard Access ☐ Open Access

Item 2 (check one box):

☑ The Author is NOT a United States government employee.

☐ The Author is a United States government employee and the Materials were prepared in the course of his or her duties as a United States government employee.

☐ The Author is a United States government employee but the Materials were NOT prepared in the course of his or her duties as a United States government employee.

ARTICLE AND VIDEO LICENSE AGREEMENT

1. Defined Terms. As used in this Article and Video License Agreement, the following terms shall have the following meanings: "**Agreement**" means this Article and Video License Agreement; "**Article**" means the article specified on the last page of this Agreement, including any associated materials such as texts, figures, tables, artwork, abstracts, or summaries contained therein; "**Author**" means the author who is a signatory to this Agreement; "**Collective Work**" means a work, such as a periodical issue, anthology or encyclopedia, in which the Materials in their entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole; "**CRC License**" means the Creative Commons Attribution-Non Commercial-No Derivs 3.0 Unported Agreement, the terms and conditions of which can be found at: http://creativecommons.org/licenses/by-nc-nd/3.0/legalcode; "**Derivative Work**" means a work based upon the Materials or upon the Materials and other pre-existing works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Materials may be recast, transformed, or adapted; "**Institution**" means the institution, listed on the last page of this Agreement, by which the Author was employed at the time of the creation of the Materials; "**JoVE**" means MyJove Corporation, a Massachusetts corporation and the publisher of *The Journal of Visualized Experiments;* "**Materials**" means the Article and / or the Video; "**Parties**" means the Author and JoVE; "**Video**" means any video(s) made by the Author, alone or in conjunction with any other parties, or by JoVE or its affiliates or agents, individually or in collaboration with the Author or any other parties, incorporating all or any portion of the Article, and in which the Author may or may not appear.

2. Background. The Author, who is the author of the Article, in order to ensure the dissemination and protection of the Article, desires to have the JoVE publish the Article and create and transmit videos based on the Article. In furtherance of such goals, the Parties desire to memorialize in this Agreement the respective rights of each Party in and to the Article and the Video.

3. Grant of Rights in Article. In consideration of JoVE agreeing to publish the Article, the Author hereby grants to JoVE, subject to **Sections 4** and **7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Article in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Article into other languages, create adaptations, summaries or extracts of the Article or other Derivative Works (including, without limitation, the Video) or Collective Works based on all or any portion of the Article and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. If the "Open Access" box has been checked in **Item 1** above, JoVE and the Author hereby grant to the public all such rights in the Article as provided in, but subject to all limitations and requirements set forth in, the CRC License.

4. Retention of Rights in Article. Notwithstanding the exclusive license granted to JoVE in **Section 3** above, the Author shall, with respect to the Article, retain the non-exclusive right to use all or part of the Article for the non-commercial purpose of giving lectures, presentations or teaching classes, and to post a copy of the Article on the Institution's website or the Author's personal website, in each case provided that a link to the Article on the JoVE website is provided and notice of JoVE's copyright in the Article is included. All non-copyright intellectual property rights in and to the Article, such as patent rights, shall remain with the Author.

5. Grant of Rights in Video – Standard Access. This **Section 5** applies if the "Standard Access" box has been checked in **Item 1** above or if no box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby acknowledges and agrees that, Subject to **Section 7** below, JoVE is and shall be the sole and exclusive owner of all rights of any nature, including, without limitation, all copyrights, in and to the Video. To the extent that, by law, the Author is deemed, now or at any time in the future, to have any rights of any nature in or to the Video, the Author hereby disclaims all such rights and transfers all such rights to JoVE.

6. Grant of Rights in Video – Open Access. This **Section 6** applies only if the "Open Access" box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby grants to JoVE, subject to **Section 7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Video in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Video into other languages, create adaptations, summaries or extracts of the Video or other Derivative Works or Collective Works based on all or any portion of the Video and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. For any Video to which this Section 6 is applicable, JoVE and the Author hereby grant to the public all such rights in the Video as provided in, but subject to all limitations and requirements set forth in, the CRC License.

7. Government Employees. If the Author is a United States government employee and the Article was prepared in the course of his or her duties as a United States government employee, as indicated in **Item 2** above, and any of the licenses or grants granted by the Author hereunder exceed the scope of the 17 U.S.C. 403, then the rights granted hereunder shall be limited to the maximum rights permitted under such statute. In such case, all provisions contained herein that are not in conflict with such statute shall remain in full force and effect, and all provisions contained herein that do so conflict shall be deemed to be amended so as to provide to JoVE the maximum rights permissible within such statute.

8. Likeness, Privacy, Personality. The Author hereby grants JoVE the right to use the Author's name, voice, likeness, picture, photograph, image, biography and performance in any way, commercial or otherwise, in connection with the Materials and the sale, promotion and distribution thereof. The Author hereby waives any and all rights he or she may have, relating to his or her appearance in the Video or otherwise relating to the Materials, under all applicable privacy, likeness, personality or similar laws.

9. Author Warranties. The Author represents and warrants that the Article is original, that it has not been published, that the copyright interest is owned by the Author (or, if more than one author is listed at the beginning of this Agreement, by such authors collectively) and has not been assigned, licensed, or otherwise transferred to any other party. The Author represents and warrants that the author(s) listed at the top of this Agreement are the only authors of the Materials. If more than one author is listed at the top of this Agreement and if any such author has not entered into a separate Article and Video License Agreement with JoVE relating to the Materials, the Author represents and warrants that the Author has been authorized by each of the other such authors to execute this Agreement on his or her behalf and to bind him or her with respect to the terms of this Agreement as if each of them had been a party hereto as an Author. The Author warrants that the use, reproduction, distribution, public or private performance or display, and/or modification of all or any portion of the Materials does not and will not violate, infringe and/or misappropriate the patent, trademark, intellectual property or other rights of any third party. The Author represents and warrants that it has and will continue to comply with all government, institutional and other regulations, including, without limitation all institutional, laboratory, hospital, ethical, human and animal treatment, privacy, and all other rules, regulations, laws, procedures or guidelines, applicable to the Materials, and that all research involving human and animal subjects has been approved by the Author's relevant institutional review board.

10. JoVE Discretion. If the Author requests the assistance of JoVE in producing the Video in the Author's facility, the Author shall ensure that the presence of JoVE employees, agents or independent contractors is in accordance with the relevant regulations of the Author's institution. If more than one author is listed at the beginning of this Agreement, JoVE may, in its sole discretion, elect not take any action with respect to the Article until such time as it has received complete, executed Article and Video License Agreements from each such author. JoVE reserves the right, in its absolute and sole discretion and without giving any reason therefore, to accept or decline any work submitted to JoVE. JoVE and its employees, agents and independent contractors shall have

jove JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

full, unfettered access to the facilities of the Author or of the Author's institution as necessary to make the Video, whether actually published or not. JoVE has sole discretion as to the method of making and publishing the Materials, including, without limitation, to all decisions regarding editing, lighting, filming, timing of publication, if any, length, quality, content and the like.

11. Indemnification. The Author agrees to indemnify JoVE and/or its successors and assigns from and against any and all claims, costs, and expenses, including attorney's fees, arising out of any breach of any warranty or other representations contained herein. The Author further agrees to indemnify and hold harmless JoVE from and against any and all claims, costs, and expenses, including attorney's fees, resulting from the breach by the Author of any representation or warranty contained herein or from allegations or instances of violation of intellectual property rights, damage to the Author's or the Author's institution's facilities, fraud, libel, defamation, research, equipment, experiments, property damage, personal injury, violations of institutional, laboratory, hospital, ethical, human and animal treatment, privacy or other rules, regulations, laws, procedures or guidelines, liabilities and other losses or damages related in any way to the submission of work to JoVE, making of videos by JoVE, or publication in *JoVE* or elsewhere by JoVE. The Author shall be responsible for, and shall hold JoVE harmless from, damages caused by lack of sterilization, lack of cleanliness or by contamination due to the making of a video by JoVE its employees, agents or independent contractors. All sterilization, cleanliness or decontamination procedures shall be solely the responsibility of the Author and shall be undertaken at the Author's expense. All indemnifications provided herein shall include JoVE's attorney's fees and costs related to said losses or damages. Such indemnification and holding harmless shall include such losses or damages incurred by, or in connection with, acts or omissions of JoVE, its employees, agents or independent contractors.

12. Fees. To cover the cost incurred for publication, JoVE must receive payment before production and publication the Materials. Payment is due in 21 days of invoice. Should the Materials not be published due to an editorial or production decision, these funds will be returned to the Author. Withdrawal by the Author of any submitted Materials after final peer review approval will result in a US$1,200 fee to cover pre-production expenses incurred by JoVE. If payment is not received by the completion of filming, production and publication of the Materials will be suspended until payment is received.

13. Transfer, Governing Law. This Agreement may be assigned by JoVE and shall inure to the benefits of any of JoVE's successors and assignees. This Agreement shall be governed and construed by the internal laws of the Commonwealth of Massachusetts without giving effect to any conflict of law provision thereunder. This Agreement may be executed in counterparts, each of which shall be deemed an original, but all of which together shall be deemed to me one and the same agreement. A signed copy of this Agreement delivered by facsimile, e-mail or other means of electronic transmission shall be deemed to have the same legal effect as delivery of an original signed copy of this Agreement.

A signed copy of this document must be sent with all new submissions. Only one Agreement required per submission.

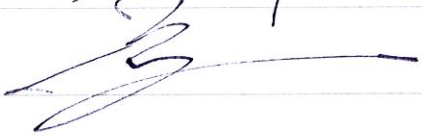
CORRESPONDING AUTHOR:

Name: SHINYA YAMAMOTO

Department: MOLECULAR & HUMAN GENETICS

Institution: BAYLOR COLLEGE OF MEDICINE

Article Title: In vivo functional study of disease-associated rare human variants using Drosophila

Signature: 

Date: 1/4/2019

Please submit a signed and dated copy of this license by one of the following three methods:

1) Upload a scanned copy of the document as a pfd on the JoVE submission site;
2) Fax the document to +1.866.381.2236;
3) Mail the document to JoVE / Attn: JoVE Editorial / 1 Alewife Center #200 / Cambridge, MA 02139

For questions, please email submissions@jove.com or call +1.617.945.9051

Dear Dr. Dsouza,

Please see below for the additional edits we made. A copy without comments and edit-track are uploaded as the new "Manuscript file".

The comments regarding the points you raised are shown as comments in this file.

Sincerely,

Shinya

Shinya Yamamoto, DVM, PhD
Assistant Professor, Department of Molecular and Human Genetics
Investigator, Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital
Baylor College of Medicine

CC: "Jacob Michael Harnish" jacob.harnish@bcm.edu, "Samantha L. Deal" samantha.deal@bcm.edu, "Hsiao-Tuan Chao" hc140077@bcm.edu, "Michael F. Wangler" mw147467@bcm.edu

Dear Dr. Yamamoto,

Your manuscript, JoVE59658R1 " In vivo functional study of disease-associated rare human variants using Drosophila ," has been editorially reviewed and the following comments need to be addressed. Please track the changes to identify all of the manuscript edits. After revising the submission, please also upload a separate document that addresses each of the editorial comments individually with the revised manuscript.

Your revision is due by **May 01, 2019**.

To submit a revision, go to the JoVE submission site and log in as an author. You will find your submission under the heading "Submission Needing Revision".

Best,

Alisha DSouza, Ph.D.
Senior Review Editor
JoVE
617.674.1888
Follow us: Facebook | Twitter | LinkedIn

About JoVE

Editorial comments:

Please see the attached word document. In-text comments have been made; these require your attention. Please address the comments by editing the attached manuscript/figures. Please maintain the current format and track all your edits.

Title:

In vivo* functional study of disease-associated rare human variants using *Drosophila

Authors and Affiliations:

J. Michael Harnish[1*], Samantha L. Deal[2*], Undiagnosed Diseases Network[%], Hsiao-Tuan Chao[1,3,4,5], Michael F. Wangler[1,2,4], Shinya Yamamoto[1,2,4,5#]

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

[2] Program in Developmental Biology, BCM, Houston, TX 77030, USA

[3] Department of Pediatrics, Section of Neurology and Developmental Neuroscience, BCM, Houston, TX 77030, USA

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

[5] Department of Neuroscience, BCM, Houston, TX 77030, USA

* These authors contributed equally

[#] Corresponding Author

E-mail: yamamoto@bcm.edu

phone: +1-832-824-8119

e-mail addresses of co-authors: JMH Jacob.Harnish@bcm.edu; SLD Samantha.Deal@bcm.edu; HTC hc140077@bcm.edu; MFW mw147467@bcm.edu

[%] List of members of the 'Undiagnosed Diseases Network' is provided in Supplemental Table 1.

KEYWORDS:

Human genetics and genomics, Mendelian diseases, rare and undiagnosed diseases, Undiagnosed Diseases Network (UDN), *Drosophila melanogaster*, variant of unknown significance (VUS), gene of uncertain significance (GUS), functional genomics, transgenic flies, UAS/GAL4 system, T2A-GAL4, electroretinogram (ERG)

SUMMARY:

The goal of this protocol is to outline a process to design and perform in vivo experiments in *Drosophila melanogaster* to assess the functional consequences of rare gene variants associated with human diseases.

ABSTRACT:

Advances in sequencing technology have made whole-genome and whole-exome datasets more accessible for both clinical diagnosis and cutting-edge human genetics research. Although a number of in silico algorithms have been developed to predict the pathogenicity of variants identified in these datasets, functional studies are critical to determine if specific genomic variants affect protein function, especially for missense variants. In the Undiagnosed Diseases Network (UDN) and other rare disease research consortia, model organisms (MO) including *Drosophila*, *C. elegans*, zebrafish, and mice are actively used to assess the function of putative

human disease-causing variants. This protocol describes a method for the functional assessment of rare human variants used in the Model Organisms Screening Center *Drosophila* Core of the UDN. The workflow begins with gathering human and MO information from multiple public databases, using the MARRVEL web resource to assess whether the variant is likely to contribute to patients' conditions, and design effective experiments based on available knowledge and resources. Next, one generates genetic tools (e.g., T2A-GAL4 and UAS-human cDNA lines) to assess the function of variants of interest in *Drosophila*. Upon development of these reagents, two-pronged functional assays based on rescue and over-expression experiments can be performed to assess variant function. In the rescue branch, the endogenous fly genes are "humanized" by replacing the orthologous *Drosophila* gene with reference or variant human transgenes. In the over-expression branch, the reference and variant human proteins are exogenously driven in a variety of tissues. In both cases, any scorable phenotype (e.g., lethality, eye morphology, electrophysiology) can be used as a read-out, irrespective of the disease of interest. Differences observed between reference and variant alleles suggest a variant-specific effect, and thus likely pathogenicity. This protocol allows rapid in vivo assessments of putative human disease-causing variants for genes of both known and unknown functions.

INTRODUCTION:

Patients with rare diseases often undergo an arduous journey referred to as the 'diagnostic odyssey' to obtain an accurate diagnosis[1]. Most rare diseases are thought to have a strong genetic origin, making genetic and genomic analyses critical elements of the clinical workup. In addition to candidate gene panel sequencing and copy number variation analysis based on chromosomal microarrays, whole-exome (WES) and whole-genome sequencing (WGS) technologies have become increasingly valuable tools over the past decade[2,3]. Currently, the diagnostic rate for identifying a known pathogenic variant in WES and WGS is ~25% (higher for pediatric cases)[4,5]. For most cases that remain undiagnosed after clinical WES/WGS, a common issue is that there are many candidate genes and variants. Next generation sequencing often identifies novel or ultra-rare variants in many genes and interpreting whether these variants contribute to disease phenotypes is challenging. For example, although most nonsense or frameshift mutations in genes are thought to be loss-of-function (LOF) alleles due to nonsense-mediated decay of the encoded transcript, truncating mutations found in the last exons escape this process and may function as benign or gain-of-function (GOF) alleles[6]. Moreover, predicting the effect of a missense allele is a daunting task since it can result in a number of different genetic scenarios as first described by Herman Muller in the 1930s: amorph, hypomorph, hypermorph, antimorph, neomorph, or isomorph[7]. Numerous in silico programs and methodologies have been developed to predict the pathogenicity of missense variants based on evolutionary conservation, type of amino acid change, position within a functional domain, allele frequency in the general population, and other parameters[8]. However, these programs are not a comprehensive solution to solving the complicated problem of variant interpretation. Interestingly, a recent study demonstrated that five broadly used variant pathogenicity prediction algorithms [Polyphen (genetics.bwh.harvard.edu/pph2)[9], SIFT (sift.bii.a-star.edu.sg)[10], CADD (cadd.gs.washington.edu)[11], PROVEAN (provean.jcvi.org/index.php)[12], and Mutation Taster (www.mutationtaster.org)] agree on pathogenicity ~80% of the time[8]. Notably, even when all algorithms agree, they return an incorrect prediction of pathogenicity up to 11% of the time. This

Commented [A1]: Please move all weblinks into the reference list and remove them from the main text.

Commented [A2R1]: Done.

not only leads to flawed clinical interpretation but also may dissuade researchers from following up on new variants by falsely listing them as benign. One way to complement the current limitation of in silico modeling is to provide experimental data that demonstrates the effect of variant function in vitro, ex vivo (e.g. cultured cells, organoids), or in vivo.

In vivo functional studies of rare disease associated variants in MO have unique strengths[13], and have been adopted by many rare disease research initiatives around the world including the Undiagnosed Diseases Network (UDN) in the United States ~~(undiagnosed.hms.harvard.edu)~~ and the Rare Diseases Models & Mechanisms (RDMM) Networks in Canada (~~www.rare-diseases-catalyst-network.ca), ~~, Japan ~~(irudbeyond.nig.ac.jp)~~, Europe ~~(solve-rd.eu)~~ and Australia ~~(www.functionalgenomics.org.au)~~[14]. In addition to these coordinated efforts to integrate MO researchers into the workflow of rare disease diagnosis and mechanistic studies at a national scale, a number of individual collaborative studies between clinical and MO researchers have led to the discovery and characterization of many new human disease-causing genes and variants[82-84]. In the UDN, a centralized Model Organisms Screening Center (MOSC) receives submissions of candidate genes and variants with a description of the patients' condition and assesses whether the variant is likely to be pathogenic using informatics tools and *in vivo* experiments. In Phase I (2015-2018) of the UDN, the MOSC comprised of a *Drosophila* Core [Baylor College of Medicine (BCM)] and a Zebrafish Core (University of Oregon) that worked collaboratively to assess cases. Using informatics analysis and a number of different experimental strategies in *Drosophila* and zebrafish, the MOSC has so far contributed to the diagnosis of 132 patients, the identification of 31 new syndromes[55], the discovery of several new human disease genes (e.g., *EBF3*[15], *ATP5F1D*[16], *TBX2*[17], *IRF2BPL*[18], *COG4*[19], *WDR37*[20]) and phenotypic expansion of known disease genes (e.g., *CACNA1A*[21], *ACOX1*[22]). In addition to projects within the UDN, MOSC *Drosophila* Core researchers have contributed to new disease gene discoveries in collaboration with the Centers for Mendelian Genomics and other initiatives (e.g., *ANKLE2*[23], *TM2D3*[24], *NRD1*[25], *OGDHL*[25], *ATAD3A*[26], *ARIH1*[27], *MARK3*[28], *DNMBP*[29]) using the same set of informatics and genetic strategies that were developed for the UDN. Given the significance of MO studies on rare disease diagnosis, the MOSC was expanded to include a *C. elegans* Core and a second Zebrafish core (both at Washington University at St. Louis) for the Phase II (2018-2022) of the UDN.

This manuscript describes an in vivo functional study protocol that is actively used in the UDN MOSC *Drosophila* Core to determine if missense variants have a functional consequence on the protein of interest using transgenic flies that express human proteins. The goal of this protocol is to help MO researchers work collaboratively with clinical research groups to provide experimental evidence that a candidate variant in a gene of interest has functional consequences and thus facilitates clinical diagnosis. This protocol will be most useful in a scenario in which a *Drosophila* researcher is approached by a clinical investigator who has a rare disease patient with a specific candidate variant in a gene of interest. This protocol can be broken down into three elements: (1) gathering information to assess the likelihood of the variant of interest being responsible for the patient phenotype and the feasibility of a functional study in *Drosophila*, (2) gathering existing genetic tools and establishing new ones, and (3) performing functional studies in vivo. The third element can further be subdivided into two sub-elements based on how one can assess the function of a variant of interest (rescue experiment or over-expression-based

Commented [A3]: Please move all weblinks into the reference list and remove them from the main text.

Commented [A4R3]: Done.

strategies). It is important to note that this protocol can be adapted and optimized to many scenarios outside of rare monogenic disease research (e.g., common diseases, gene-environment interaction, pharmacological and genetic screens to identify therapeutic targets). The ability to determine the functionality and pathogenicity of variants will not only benefit the patient of interest via providing accurate molecular diagnosis but will also have broader impacts on both translational and basic scientific research.

PROTOCOL:

1. Gather human and MO information to assess the likelihood of a variant of interest being responsible for disease phenotypes and the feasibility of functional studies in *Drosophila*

1.1. Perform extensive database and literature searches to determine whether the specific genes and variants of interest are good candidates to explain the phenotype of the patient of interest. Specifically, one should gather the following information.

1.1.1 Assess if the gene of interest has been previously implicated in other genetic disorders (phenotypic expansion of known disease gene) or this is an entirely new disease candidate gene [gene variant of uncertain significance (GVUS)].

1.1.2. Assess the allele frequency of the variant of interest in disease or control population databases.

1.1.3. Assess whether there are copy number variations (CNVs) that include this gene in disease or control population databases.

1.1.4. Assess what the orthologous genes are in different MO species including mouse, zebrafish, *Drosophila*, *C. elegans* and yeast, and further investigate the known functions and expression patterns of these orthologous genes.

1.1.5. Assess whether the variant of interest is present in a functional domain of the protein and if the amino acid of interest is evolutionarily conserved.

Note: Answers to these five questions (**1.1.1-1.1.5**) can be obtained by accessing a number of human and MO databases individually or by using the MARRVEL (Model organism Aggregated Resources for Rare Variant ExpLoration, ~~http://marrvel.org/)~~ web resource[30], which is described in-depth in an accompanying article[31]. See the representative results section for specific examples. The Monarch Initiative website ~~(https://monarchinitiative.org/)~~[32] and Gene2Function ~~(http://www.gene2function.org/search/)~~[33] also provide useful information.

1.2 Gather additional information to further assess whether the variant is a good disease candidate from a protein function and structure point of view.

1.2.1 Assess if the variant of interest is predicted to be damaging based on *in silico* prediction algorithms.

Notes: A number of variant pathogenicity algorithms have been developed by many research groups over the past ~15 years, and some are also displayed in the MARRVEL search result. More recent programs, including the two listed below, combine multiple variant pathogenicity

Commented [A5]: Minor edits were made to the structure of the protocol to meet JoVE's style requirements.

Commented [A6]: The protocol section currently has too many descriptive paragraphs between steps. Please edit it to limit long notes as much as possible. Please also ensure that the protocol length is under 10 pages after all edits.

Commented [A7R6]: We tried our best best to trim down the text between the steps. We feel that the descriptive paragraphs that are still remaining in the current version are essential to the readers' understanding of the protocol and their ability to reproduce our work. Any reduction in these paragraphs compromises the clarity of the protocol. If there are some text the editor thinks can be removed, please let us know so that we can consider further modifications.

Commented [A8]: I have unhighlighted steps in section 1 as they lack filmable content.

Commented [A9R8]: This is OK, but since this is a critical step in the whole protocol, you may still want to have some text with bullet points that highlight the important steps are projected along with a narrative. Also, we can have one shot where we have a screenshot of MARRVEL.org and encourage the viewers to go to this website and watch the accompanying JoVE video by Wang et al.,

Commented [A10]: Only accepted articles can be referenced.

Commented [A11R10]: The paper has since been accepted and in press; citation updated

Commented [A12]: Please move all weblinks into the reference list and remove them from the main text.

Commented [A13R12]: Done.

prediction algorithms and machine learning approaches to generate a pathogenicity score. For more information on variant prediction algorithms and their performance, please refer to Ghosh *et al*[8]. (i) CADD (Combined Annotation-Dependent Depletion): Integrative annotation tool built from more than 60 genomic features, which provides scores for human SNVs as well as short insertions and deletions. ~~(cadd.gs.washington.edu)~~[11]. (ii) REVEL (Rare Exome Variant Ensemble Learner): Combines multiple variant pathogenicity algorithms (MutPred, FATHMM, VEST, PolyPhen, SIFT, PROVEAN, MutationAssessor, MutationTaster, LRT, GERP, SiPhy, phyloP, and phastCons) to provide an integrated score for all possible human missense variants. ~~(sites.google.com/site/revelgenomics)~~[34]

1.2.2. Determine if the human gene/protein of interest or its MO orthologs have been shown to genetically or physically interact with genes/proteins previously linked to genetic diseases. If so, assess if the patient of interest exhibits overlapping phenotypes with these disorders.

Notes: Several tools have been developed to analyze genetic and protein-protein interactions based on MO publications as well as large-scale proteomics from multiple species screens. STRING (Search Tool for Recurring Instances of Neighboring Genes) ~~(string-db.org)~~[35]: A database for known and predicted protein-protein interactions. It integrates genetic interaction and co-expression datasets as well as text-mining tools to identify genes and proteins that may function together in a variety of organisms. MIST (Molecular Interaction Search Tool) ~~(fgrtools.hms.harvard.edu/MIST)~~[36]: A database that integrates genetic and protein-protein interaction data from core genetic MOs (yeast, *C. elegans*, *Drosophila*, zebrafish, frog, rat and mouse) and humans. Prediction of interactions inferred from orthologous genes/proteins (interlogs) are also displayed.

1.2.3. Determine if the three-dimensional structure of the protein of interest has been solved or modeled. If so, determine where the variant of interest map relative to key functional domains.

Notes: Protein structures that have been solved by X-ray crystallography, nuclear magnetic resonance (NMR) and cryo-electron microscopy can be found in public databases including the PDB (Protein Data bank) ~~(www.wwpdb.org)~~ and EMDatabank ~~(www.emdatabank.org)~~[37]. Although there is no single database for predicted/modeled protein structures, a number of algorithms including SWISS-MODEL ~~(swissmodel.expasy.org)~~[38], Modeller ~~(salilab.org/modeller)~~[39] and Phyre$_2$ ~~(www.sbg.bio.ic.ac.uk/phyre2)~~[40] are available for users to perform protein modeling.

1.3 Communicate with clinical collaborators to discuss the information gathered from the informatics analyses in steps **1.1** and **1.2**. If clinical collaborators also feel the variant and gene of interest are good candidates to explain the phenotypes seen in the patient, proceed to **Section 2**. If one has specific questions about the patient's genotype and phenotype, make sure to discuss with the clinical collaborators before moving forward.

Note: If one feels the variant of interest is unlikely to explain the patient's phenotype of interest (e.g. identical variant found in high frequency in control population), the reader must discuss this

Commented [A14]: Please move all weblinks into the reference list and remove them from the main text.

Commented [A15R14]: Done.

Commented [A16]: Please move all weblinks into the reference list and remove them from the main text.

Commented [A17R16]: Done.

Commented [A18]: Please move all weblinks into the reference list and remove them from the main text.

Commented [A19R18]: Done.

with clinical collaborators to determine whether the variant is a good candidate as the reader may not have the expertise to interpret the clinical phenotype.

2. Gather existing genetic tools and establish new reagents to study a specific variant of interest

Note: Once the variant(s) of interest has been determined to be a good candidate to pursue experimentally, the next step is to gather or generate reagents to perform in vivo functional studies. For functional studies described in this protocol, one will need a few key *Drosophila melanogaster* reagents: 1) U~~AS~~pstream Activation Sequence-regulated Human cDNA transgenic strains that carry the reference or variant sequence, 2) a loss of function~~LOF~~ allele of a fly gene of interest, and 3) a GAL4 line that can be used for rescue experiments.

2.1. Generate UAS-human cDNA constructs and transgenic flies

2.1.1. Identify and obtain the appropriate human cDNA constructs. Many clones are available from the MGC (Mammalian Gene Collection)[43] and can be purchased from selected venders ~~(genecollections.nci.nih.gov/MGC)~~. For genes that are alternatively spliced, check which isoform cDNA corresponds to using Ensembl ~~(useast.ensembl.org)~~ or RefSeq ~~(www.ncbi.nlm.nih.gov/refseq)~~.

Note: Many cDNAs are available in recombinase mediated cloning system ~~Gateway~~ compatible reagents[44], which simplifies the subcloning step. cDNAs may come in an "open (no stop codon)" or "closed (with endogenous or artificial stop codon)" format. While open clones allow C' tagging of proteins that are useful for biochemical (e.g. western blot) and cell biological (e.g. immunostaining) assays to monitor expression of the protein of interest, it may interfere with protein function in some cases.

2.1.2 Sub-clone the reference and variant cDNA into the *Drosophila* transgenic vector. Use the ϕC31-mediated transgenesis system since this allows the reference and variant cDNAs to be integrated into the same location in the genome[45]. For this project, the MOSC *Drosophila* Core routinely uses the pGW-HA.attB vector[46].

Note: If the human cDNA is in recombinase mediated cloning system ~~Gateway~~ compatible vectors (e.g. pDONR221, pENTR221), one can skip to **2.1.4** which explains LR reactions to subclone the cDNAs into pGW-HA.attB.

2.1.2~~a.~~.1 If the human cDNA is not in a recombinase mediated cloning system ~~Gateway~~ compatible plasmid, subclone the human cDNAs into a Gateway entry vector using standard molecular biological techniques. An example of such protocol is documented below.

2.1.2.1~~a.1~~ Perform an overhang PCR to introduce *attB1* and *attB2* arms. The forward primer should have the *attB1* sequence **5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCACC-3'** followed by the first 22 nucleotides of the target cDNA and the reverse primer should have the *attB2* sequence **5'-GGGGACCACTTTGTACAAGAAAGCTGGGTCCTA-3'** followed by the reverse

Commented [A20]: substeps would need to be highlighted to film this step. However, most steps in the section are too vague for filming.

Commented [A21R20]: We agree with this but seems the viewers/readers somehow needs to make these reagents before going to Step 3. Again, a simple screenshot with text and narrative saying please see main protocol or other methods may be necessary.

Commented [A22]: Please move all weblinks into the reference list and remove them from the main text.

Commented [A23R22]: Done.

Commented [A24]: Please avoid commercial names. These can be added to the table of materials instead. I have flagged them all using red fonts.

Commented [A25R24]: While Gateway is technically a commercial name, it is also the accepted way in the field to refer to this style of cloning as no generic reagents or methods exist.
https://en.wikipedia.org/wiki/Gateway_Technology
If you have recommendations or previous ways you have referred to this cloning technique without using the trade name, please let us know and we will be happy to make changes. Have changed it to "(e.g. Recombinase mediated cloning system" for the time being.

complement of the last 25 nucleotides of the cDNA of interest. Exclude the stop codon of the cDNA if it is desirable to "open" a clone to add a C' tag, or add a stop codon if one wishes to "close" a clone.

2.1.2.1a.2 Prepare a 100 µL high-fidelity~~Q5~~ PCR mix consisting of 50 µL high-fidelity PCR master mix~~Q5 mastermix (NEB #M0491)~~, 36 µL distilled~~milliQ~~ water, 5 µL of each forward and reverse primers listed in **2.1.2.1~~2.1~~** diluted to 10 µM, and 4 µL of target cDNA (150 ng/µl).

2.1.2.1a.3 Perform ~~the~~ PCR using standard ~~Q5~~ mutagenesis protocol to add *attB1* and *attB2* arms onto the cDNA of interest. Conditions will vary depending on the construct and the variants of interest. ~~(NEB #M0491)~~

2.1.2.1a.4 Isolate the target cDNA with added homology arms via gel electrophoresis and gel extraction. Create 1% agarose gel and perform electrophoresis using standard methods. Excise the band that corresponds to the size of the cDNA plus the additional length of the homology arms. Extract DNA from the gel through standard methods[95]. Commercial gel extraction kits are available from several companies. ~~(Qiagen #28704).~~

2.1.2.1a.5 Perform a ~~BP clonase~~ n in vitro recombinase reaction ~~using the Invitrogen Gateway~~based on the recombinase mediated cloning ~~cloning~~ protocol according to the system that is used. ~~(Thermo Fisher #11789020)~~

2.1.2.1a.6 Transform the BP reaction mix into chemically competent *E. coli* cells. Competent cells can be made in house or purchased from commercial vendors ~~(e.g. NEB #C2987H)~~. Culture the transformed cells overnight on an LB plate containing appropriate antibiotics for colony selection. The next day, select several colonies and grow them up in independent liquid cultures overnight.

2.1.2.1a.7 Isolate DNA from the overnight cultures through miniprep ~~(e.g. Qiagen #27104)~~. Sanger sequence the positive clones to ensure that the cDNA has the correct sequence[96]. ~~Generate glycerol stocks~~Maintain cells from the cultures that were positive for the desired sequence in 25% glycerol stored at -80°C.

2.1.3 Perform ~~Q5~~site-directed mutagenesis to introduce the variant of interest into the Gateway plasmid with the reference human cDNA[97]. A detailed protocol for this method can be found in the vendor's website ~~(NEB #E0554S)~~[48,49]. Validate the presence of the variant in the mutated plasmid via Sanger sequencing of the entire open reading frame (ORF) in order to make sure there are no additional variants introduced through this mutagenesis step.

2.1.4 Subclone the reference and variant human cDNAs in the donor plasmid (Gateway plasmids with *attL1* and *attL2* sites) into the transgenic plasmid (e.g. pGW-HA.attB with *attR1* and *attR2* sites) via the LR clonase reaction ~~(Thermo Fisher #11791100)~~.

Commented [A26]: mention reaction conditions.

Formatted: Font: Italic

Formatted: Font: Italic

Commented [A27]: cite a reference

Commented [A28R27]: Done.

Commented [A29]: cite a reference

Commented [A30R29]: Done.

Commented [A31]: cite a reference

Commented [A32R31]: Done.

Note: There are UAS ϕC31 vectors that are designed for conventional restriction enzyme based subcloning (e.g. pUAST.attB[50]) if one prefers to subclone human cDNAs via restriction enzyme methods.

2.1.5 Select the ϕC31 docking sites in which to integrate the UAS-human cDNA transgenes. A number of docking sites have been generated by several laboratories and are publically available from stock centers ~~(bdsc.indiana.edu/stocks/phic31/phic31_attp.html, kyotofly.kit.jp/stocks/documents/phiC31.html)~~[50,52,53].

Note: Since it is convenient to have the human transgene on a chromosome that does not contain the fly ortholog of the gene of interest, we recommend the use of a 2nd chromosome docking site [VK37 (BDSC stock #24872~~, flybase.org/reports/FBst0024872.html)~~] when the fly ortholog is on the X, 3rd or 4th chromosomes, and a 3rd chromosome docking site [VK33 (BDSC stock #24871~~, flybase.org/reports/FBst0024871.html)~~] when the fly ortholog is on the 2nd chromosome.

2.1.6 Inject the UAS-human cDNA constructs into flies expressing the ϕC31 integrase in their germline (e.g. vas-ϕC31, nos-ϕC31) ~~(bdsc.indiana.edu/stocks/phic31/phic31_int.html)~~.

Note: Microinjection can be performed in house, or can be sent to core facilities or commercial entities for transgenesis. Detailed protocol for generating transgenic flies can be found in the cited book chapter[51].

2.1.7 Establish stable transgenic strains from the injected embryos. Inject ~100-200 embryos per construct[94]. A representative crossing scheme for a transgene insertion into a 2nd chromosome docking site (VK37) is depicted in **Figure 1A**. Refer to the cited books[54,55] for basic *Drosophila* genetics information.

2.2 Obtain or generate a T2A-GAL4 line that facilitates rescue-based functional assays (see **Figure 2** and **Section 3.1**).

Note: This line will serve two purposes. First, most T2A-GAL4 lines tested behave as strong LOF alleles by functioning as a gene trap allele. Second, T2A-GAL4 lines function as a GAL4 driver that allows expression of UAS constructs (e.g. UAS-GFP, UAS-human cDNAs) under the endogenous regulation elements of the gene of interest[56,57] (**Figure 2A-C**).

2.2.1 Search public stock collections for available T2A-GAL4 lines including the *Drosophila* Gene Disruption Project (GDP)[58] in which ~1,000 T2A-GAL4 lines have been generated[59]. These strains are currently available from the Bloomington *Drosophila* Stock Center (BDSC) and are searchable through both the GDP ~~(flypush.imgen.bcm.tmc.edu/pscreen)~~ and BDSC ~~(bdsc.indiana.edu)~~ websites.

2.2.2. If a T2A-GAL4 line for the fly gene of interest is not available, check if a suitable coding intronic MiMIC (*Minos* mediated Integration Cassette) line is available for conversion into a T2A-GAL4 line using recombinase mediated cassette exchange (RMCE)[60] (**Figure 2A**).

Commented [A33]: Please move all weblinks into the reference list and remove them from the main text.

Commented [A34R33]: Done.

Commented [A35]: Please move all weblinks into the reference list and remove them from the main text.

Commented [A36R35]: Done.

Commented [A37]: Please move all weblinks into the reference list and remove them from the main text.

Commented [A38R37]: Done.

Commented [A39]: needs references

Commented [A40R39]: Done.

Commented [A41]: Substeps would need to be highlighted to film this step. However, most steps in the section are too vague for filming or lack filmable content.

Commented [A42R41]: We agree but this is an essential step for step 3.1. Again, maybe a slide showing bullet points with some narration will be needed.

Commented [A43]: Please move all weblinks into the reference list and remove them from the main text.

Commented [A44R43]: Done.

Note: RMCE allows intronic MiMIC elements that are in between two coding exons to be converted into a T2A-GAL4 line through microinjection of a donor construct (an example of a crossing scheme is shown in **Figure 1B**) or series of crosses as described in detail in the following papers[57,59].

2.2.3. If a T2A-GAL4 line is not available and an appropriate coding intronic MiMIC does not exist, explore the possibility of generating a T2A-GAL4 line via the CRIMIC (CRISPR-mediated Integration Cassette) system[59].

Note: This methodology uses CRISPR-mediated DNA cleavage and homology directed repair (HDR) to integrate a MiMIC-like cassette into a coding intron in a gene of interest.

2.2.4. If the gene of interest lacks a large intron (>150bp) or has no introns, attempt to knock-in a GAL4 transgene into the fly gene with the CRISPR/Cas9 system using HDR as described in the following papers[20,61,62].

Note: If generation of a T2A-GAL4 or GAL4 knock-in allele is difficult, one can attempt to perform rescue experiments using these pre-existing alleles or RNAi lines and ubiquitous or tissue-specific GAL4 drivers as described in the following papers[92] (REF).

3. Perform Functional Analysis of the human variant of interest *in vivo* in *Drosophila*

NOTE: Perform a rescue-based analysis (**Section 3.1**) as well as over-expression studies (**Section 3.2**) using the tools gathered or generated in **Section 2** to assess the consequence of the variant of interest *in vivo* in *Drosophila*. Consider utilizing both approaches since the two are complementary.

3.1. Perform functional analysis through rescue based experiments.

NOTE: Heterologous rescue-based experiments in *Drosophila* using human proteins determine whether the molecular function of the two orthologous genes have been conserved over ~500 million years of evolution, and further assess the function of the variant in the context of the human protein[63]. Although a systematic analysis studying hundreds of gene pairs has not been reported, several dozen human and mammalian (e.g. mouse) genes are able to replace the function of *Drosophila* genes[13].

3.1.1 In the rescue-based approach, first determine whether there are obvious, scorable, and reproducible phenotypes in LOF mutants in the fly ortholog before assessing the function of variants.

Note: Previous literature on the fly gene is the first place to datamine and can be found using databases including FlyBase ~~(flybase.org/)~~ and PubMed ~~(www.ncbi.nlm.nih.gov/pubmed/)~~. Additional databases such as MARRVEL ~~(marrvel.org)~~, Monarch Initiative

Commented [A45]: Substeps would need to be highlighted to film this step. However, most steps in the section are too vague for filming or lack filmable content.

Commented [A46R45]: We agree but this is an important component of the protocol. Maybe we can have one slide that says one can perform functional studies based on rescue or over-expression strategies. Here, the video will focus on the latter but the viewers are encouraged to read the text for the former.

Commented [A47]: Please move all weblinks into the reference list and remove them from the main text.

Commented [A48R47]: Done.

(monarchinitiative.org/), and Gene2Funcion (http://www.gene2function.org) are also useful in gathering this information.

3.1.2 Perform a global survey of scorable phenotypes in homozygous and hemizygous (T2A-GAL4 allele over a molecularly defined chromosomal deficiency; e.g. bdsc.indiana.edu/stocks/df/index.html) animals, especially if the T2A-GAL4 allele is the first mutation to be characterized for a specific gene. Assess phenotypes such as lethality, sterility, longevity, morphological (e.g. size and morphology of the eye) and behavior (e.g. courtship, flight, climbing and bang sensitivity defects).

Note: If there are no major phenotypes identified from this primary screen, more subtle phenotypes such as neurological defects measured by electrophysiological recordings can be used if they are highly reproducible and specific. As an example, functional studies using electroretinogram (ERG) are described in **3.2.3**. If one fails to detect any scorable phenotype, move on to 3.2 to perform the over-expression based functional study.

3.1.32 Once a scorable phenotype is identified in the fly LOF mutant, test whether the reference human cDNA can replace the function of the fly ortholog by attempting to use the human cDNA to rescue the mutanta fly line lacking the predicted ortholog. The phenotypic assay to be performed here will depend on the result of **3.1.2.** and will be specific to the gene that is under study

Note: If this "humanization" of the fly gene is successful, then one now has a platform to compare the efficiency of rescue for the variant of interest compared to the reference counterpart. The rescue seen with reference human cDNA does not have to be perfect. Partial rescue of the fly mutant phenotype using a human cDNA still provides a reference point to perform comparative studies using the variant human cDNA strain.

3.1.34 Using the assay system selected in **3.1.2**, compare the rescue observed with the reference human cDNA to the rescue observed with the variant human cDNA to determine if the variant of interest has consequences on the gene of interest.

Note: If the variant human cDNA performs worse than the reference allele, then this suggests that the variant of interest is deleterious to the protein function. If the variant and reference cannot be functionally distinguished, then the allele may be an isomorph (variant that doesn't affect protein function) or the assay is not sensitive enough to detect subtle differences.

3.1.45. If the variant is found to be a deleterious allele, then further assess the expression and intracellular localization of the reference and variant protein of interest *in vivo* via western blot, immunofluorescence staining or other methods[93].

Note: If the UAS-human cDNA was generated from an open clone in a pGW-HA.attB vector, then one can use an anti-HA antibody to perform these biochemical and cellular assays. If the original clone was a closed clone, then one can test whether commercial antibodies against the human

Commented [A49]: How? Needs a reference.

Commented [A50R49]: We modified the text to explain this a bit more. The method and assay will depend on the gene of interest.

Formatted: Font: Bold

Commented [A51]: References?

Commented [A52R51]: Added

proteins can be used for these assays. A difference in expression levels and intracellular localization may reveal how the variant of interest affects protein function.

3.2 Perform functional analysis through over-expression studies

NOTE: Ubiquitous or tissue-specific over-expression of human cDNAs in otherwise wild-type flies can provide information that is complementary to the rescue-based experiments discussed in **3.1**. While rescue-based assays are primarily designed to detect LOF variants (amorphic, hypomophic), over-expression based assays may also reveal gain-of-function (GOF) variants that may be more difficult to assess (hypermorphic, antimorphic, neomorphic).

3.2.1 Select a set of GAL4 drivers to over-express the human cDNAs of interest. A number of ubiquitous, tissue, and stage specific GAL4 drivers are available from public stock centers ~~(e.g. bdsc.indiana.edu/stocks/gal4/index.html, kyotofly.kit.jp/stocks/documents/GAL4.html)~~, some of which are more frequently used than others ~~(http://flybase.org/GAL4/freq_used_drivers/)~~. Validate drivers with a reporter line (e.g. UAS-GFP) to confirm their expression pattern upon receiving the stocks.

3.2.2 Express the reference and variant human cDNAs using the same driver under the same condition (e.g. temperature) and ascertain if there is a difference between them. First focus on ubiquitous drivers and easily scorable phenotypes (lethality, sterility, morphological phenotypes), and move on to tissue specific drivers and more specific phenotypes.

Note: If a phenotype is only seen in the reference but not in the variant line, then the variant may be an amorphic or a strong hypomorphic allele. If the phenotype is seen in both genotypes, but the reference causes a stronger defect, then the variant may be a mild to weak hypomorphic allele. If the reference does not show a phenotype, or only exhibits a weak phenotype, but the variant shows a strong defect, then the variant may be a GOF allele.

3.2.3. If one does not see a phenotype in standard culture conditions (room temperature or at 25°C, then set the crosses at different temperatures ranging between 18°C to 29°C since the UAS/GAL4 system is known to be temperature-sensitive[64,65]. Typically, the expression of UAS transgenes is higher at higher temperatures.

3.3 Perform additional functional studies related to the genes and protein of interest.

Note: In addition to examining general defects, one can select an assay system to probe into the molecular function of the gene and variant. In one of the examples discussed under "Representative Results" section (*TBX2* case), ERG recordings were used to determine the effect of the variant on photoreceptor function since the fly gene of interest (*bifid*) had been studied extensively in the context of visual system development. Detailed protocols for ERG in *Drosophila* can be found in the following papers[66–68].

Commented [A53]: I'm not sure what we would film here. Please describe the specific steps in order to film this.

Commented [A54R53]: Our thoughts for filming this would be to primarily look at the differences noticed under the scope between reference and variant alleles using the TBX2 over-expression lines as an example. We may also use some sort of graphic or animation to show the crossing scheme. We look forward to working with the film side to determine the best method to present this protocol since it is a bit different than most JoVE protocols

Commented [A55]: How? What would we film? All specific details must be described to ensure proper scripting and filming. All work performed in a browser or software must be described with all button clicks and menu selections

Commented [A56R55]: See above

Commented [A57]: How? What would we film? All specific details must be described to ensure proper scripting and filming. All work performed in a browser or software must be described with all button clicks and menu selections

Commented [A58R57]: See above

3.3.1 Generate flies to test for functional defects in the visual system. Cross virgin females from the Rhodopsin 1 (Rh1)-GAL4 line to males with reference or variant UAS-human cDNA transgenes to express the human proteins of interest in the R1-R6 photoreceptors.

Note: Cross 3-5 virgin females to 3-5 male flies in a single vial and transfer the crosses every 2-3 days to have many animals eclosing from a single cross. All crosses must be kept in an incubator set at the experimental temperature to obtain consistent results.

3.3.2 Once flies begin to eclose (at 25°C, ~10 days after setting the initial cross), gather the progeny (*Rh1-GAL4/+; UAS-human cDNA/+*) into fresh vials and place them back into the incubator set at the experimental temperature for an additional 3 days.

Note: Although ERG can be performed on 1-2 day old flies, newly eclosed flies may have large fluctuations in their ERG signal. If one wants to examine an age-dependent phenotype, these flies can be aged for several weeks as long as they are regularly (e.g. every ~5 days) transferred to a new vial to avoid the flies from drowning in wet food.

3.3.3 Prepare the flies for ERG recording by first anesthetizing the flies using CO_2 or placing them into a vial on ice. Gently glue one side of the fly onto a glass microscope slide to immobilize them.

Note: Multiple reference and variant flies can be glued on to a single slide. Place all flies in approximately the same orientation with one eye being accessible for the recording electrode. Be careful not to get glue on the eye and to leave the proboscis free.

3.3.4 Prepare the recording and reference electrodes. Place a 1.2mm glass capillary into a needle puller ~~(e.g. NARISHIGE Model PP-830)~~ and ~~switch on the filament~~activate. B~~B~~reak the capillary tube to obtain two sharp tapered electrodes. The resulting electrodes should be hollow and have a final diameter <0.5mm. ~~As soon as the weight drops, turn off the puller and detach the pulled capillary tube from the machine. Break the capillary tube to obtain two sharp tapered electrodes.~~

3.3.5 Fill the capillaries with saline solution (100 mM NaCl), making sure there are no air bubbles. Slide the glass capillaries over the silver wire electrodes (both the recording electrode and reference electrode, see **Figure 4**) and secure the capillaries in place.

3.3.6 Configure the stimulator and amplifier. Detailed set up can be found in Lauwers *et al.*[67] The UDN *Drosophila* MOSC set up consists of the ~~following equipment~~equipment listed in the materials section of the protocol:

- ~~Iso-Dam Isolated Biological Amplifier (World Precision Instruments, Sarasota, FL, USA):~~

3.3.6.1. Set the amplifier to 0.1 Hz high pass filter, 300 Hz low pass filter, and 100 gain.

- ~~S48 Stimulator (Astro-Med Inc. GRASS Instrument Division; West Warwick, RI, USA):~~

3.3.6.2. Set the stimulator to 1 s period, 500 ms pulse width, 500 ms pulse delay, run mode, and 7 amplitude.

Commented [A59]: Mention puller settings and desired capillary specifications.

Commented [A60R59]: Information added.

Formatted: No bullets or numbering

Formatted: Font: Bold

Formatted: Font: Bold

- **3.3.6.3.** Prepare the Light source for photostimulation.~~:~~ Use a halogen light source ~~(ACE Light Source, SCHOTT North America Inc., Southbridge, MA, USA)~~ to ~~stimulate~~ activate the fly photoreceptors

- ~~Axoscope 10.5 data acquisition software (Molecular Devices, San Jose, CA, USA):~~ **3.3.6.4.** Prepare the recording software on a computer connected to the ERG setup. Create a stimulation protocol with acquisition model "fixed length events" and 20 seconds duration.

3.3.7 Acclimate the flies to complete darkness before initiating the ERG recordings. Place the flies into complete darkness for at least 10 minutes before beginning the experiment.

Note: Since flies cannot detect red light, one can use a red light source during the period of dark habituation.

3.3.8 Place the slide containing the flies onto the recording apparatus and move the micromanipulators carrying the reference and recording electrodes to a point that is close to the fly of interest on the slide. Watch the tip of the electrode and carefully place the reference electrode into the thorax of the fly (penetrate the cuticle). Once the reference electrode is stably inserted, place the recording electrode on the surface of the eye.

Note: The exact position of this reference electrode does not have a major impact on the ERG signal. The recording electrode should be placed at the surface of the eye since ERG is a field potential recording rather than an intracellular recording. The proper amount of pressure applied to the recording electrode causes a small dimple without penetrating the eye.

3.3.9 Turn off all lights for another 3 minutes to acclimate the flies again to the dark environment. If using a halogen light source with a manual shutter, turn on the light source at this point with the shutter closed (flies are still in dark).

3.3.10 Set up the recording software ~~(Axoscope)~~ and begin the recording the signal.

3.3.11 Expose the fly eyes to light by opening and closing the shutter every 1 second for the 20 second duration of a single run.

Note: One can control the on/off of the halogen light source manually or this can be programmed to have it automated using a white LED light source. Note that one can obtain much more robust and reliable ERG by using a halogen light source compared to a white light LED, likely due to the broader light spectrum emitted from the halogen light source.

3.3.12 Record ERGs from all of the flies that are mounted on the glass slide. Perform ERGs from 15 flies per genotype per condition.

Formatted: Font: Bold

Formatted: Font: Bold

Note: Some parameters that can be altered to find a condition that shows robust differences between reference and variant cDNAs may include temperature, age, or environmental conditions (e.g. reared in light-dark cycle or constant light/darkness).

3.3.13 Perform data analysis: Compare the ERGs from the reference, variant, and controls to determine if there are differences. Assess the ERG data for changes in on-transients, depolarization, off-transients, and repolarization[69] (**Figure 4B**).

Note: Depolarization and repolarization reflects the activation and inactivation of the phototransduction cascade within the photoreceptors, whereas the on- and off- transients are measures of the activities of post-synaptic cells that receive signals from the photoreceptors. Decreased amplitude and altered kinetics of repolarization are often associated in defects with photoreceptor function and health, whereas defects in on- and off-transients are found in mutants with defective synapse development, function or maintenance[70].

3.3.14 Upon identification of differences in ERG phenotypes with over-expression of reference versus variant human cDNAs, further determine whether this electrophysiological phenotype is associated with structural and ultrastructural defects in photoreceptors and their synapses by performing histological analysis as well as transmission electron microscopy.

Note: Further discussion on interpretation of ERG defects and structural/ultrastructural analysis can be found in the following article[69].

Representative Results:

Functional Study of a *de novo* missense variant in *EBF3* linked to neurodevelopmental phenotypes

In a 7-year-old male with neurodevelopmental phenotypes including hypotonia, ataxia, global developmental delay and expressive speech disorder, physicians and human geneticists at the National Institutes of Health Undiagnosed Diseases Project (UDP) identified a *de novo* missense variant (p.R163Q) in *EBF3 (Early B-Cell Factor 3)*[15], a gene that encodes a COE (Collier/Olfactory-1/Early B-Cell Factor) family transcription factor. This case was submitted to the UDN MOSC in March 2016 for functional studies. To assess whether this gene was a good candidate for this case, the MOSC gathered human genetic and genomic information from OMIM (~~www.omim.org/~~), ClinVar (~~www.ncbi.nlm.nih.gov/clinvar/~~), ExAC [~~exac.broadinstitute.org/~~ (now expanded to gnomAD, ~~gnomad.broadinstitute.org/~~)], Geno$_2$MP (~~geno2mp.gs.washington.edu/Geno2MP/#/~~), DGV (~~dgv.tcag.ca/dgv/app/home~~), and DECIPHER (~~decipher.sanger.ac.uk/~~). In addition, the orthologous genes in key MO species were identified using the DIOPT tool (~~www.flyrnai.org/cgi-bin/DRSC_orthologs.pl~~), and further obtained gene expression and phenotypic information from individual MO databases [e.g. Wormbase (~~www.flyrnai.org/cgi-bin/DRSC_orthologs.pl~~), FlyBase (~~flybase.org/~~), ZFIN (~~zfin.org/~~) and MGI (~~www.informatics.jax.org/~~)]. The informatics analyses performed for *EBF3* and other pioneering studies in the UDN MOSC formed the basis for the later development of the MARRVEL resource (~~marrvel.org/~~) in 2017[30].

Commented [A61]: Please move all weblinks into the reference list and remove them from the main text.

Commented [A62R61]: Done.

The information gathered via this methodology indicated *EBF3* was not associated with any known human genetic disorder at the time of analysis, and it was concluded that the p.R163Q variant was a good candidate based on the following information. (1) This variant had not been previously reported in control population databases (ExAC) and disease population database ($Geno_2MP$), indicating that this is a very rare variant. (2) Based on ExAC, pLI (probability of LOF intolerance) score of this gene is 1.00 (pLI score ranges from 0.00 to 1.00). This indicates that there is a selective pressure against LOF variants in this gene in the general population and suggests that haploinsufficiency of this gene may cause disease. For more information on pLI score and its interpretation, please refer to the accompanying MARRVEL tutorial article in JoVE[31] as well as related papers[30,71]. (3) The p.R163Q variant is located in the evolutionarily conserved DNA binding domain of this protein, suggesting that it may affect DNA binding or other protein function. (4) The p.R163 residue is evolutionarily conserved from *C. elegans* and *Drosophila* to human, suggesting that it may be critical for protein functional across species. (5) *EBF3* orthologs have been implicated in neuronal development in multiple MO[72] including *C. elegans*[73], *Drosophila*[74], Xenopus[75]and mice[76]. (6) During brain development in mice, *Ebf3* was shown to function downstream of *Arx* (*Aristaless-related homebox*)[77], a gene known to be associated with several epilepsy and intellectual disability syndromes in human[78]. Hence, these data together suggested that *EBF3* is highly likely to be crucial to human neurodevelopment and that the p.R163Q variant may have functional consequences.

To assess whether p.R163Q affects EBF3 function, a T2A-GAL4 line for *knot (kn)*, the fly ortholog of human *EBF3*[79] was generated via RMCE of a coding intronic MiMIC insertion[15]. The $kn^{T2A\text{-}GAL4}$ line was recessive lethal and failed to complement the lethality of a classic *kn* allele ($kn^{col\text{-}1}$) as well as a molecularly defined deficiency that covers *kn [Df(2R)BSC429]*[80]. Expression pattern of the GAL4 also reflected previously reported patterns of *kn* expression in the brain as well as in the wing imaginal disc[15]. UAS transgenic flies were generated to allow the expression of reference and variant human *EBF3* cDNA as well as a wild-type fly *kn* cDNA. All three proteins were tagged with a C-terminal 3xHA tag. Importantly, UAS-wild type fly *kn* (kn^+) or reference human *EBF3* ($EBF3^+$) transgenes rescued the lethality of $kn^{T2A\text{-}GAL4}/Df(2R)BSC429$ to a similar extent (**Figure 3C, left panel**)[81]. In contrast, UAS-human *EBF3* transgene with the p.R163Q variant ($EBF3^{p.R163Q}$) was not able to rescue this mutant, suggesting that the p.R163Q variant affects EBF3 function *in vivo*[15]. Interestingly, when assessed using an anti-HA antibody, the $EBF3^{p.R163Q}$ protein was successfully expressed in the fly tissues and its levels and subcellular localization (primarily nuclear) was indistinguishable from that of $EBF3^+$ and Kn^+. This suggests that the variant is not causing a LOF phenotype due to protein instability or mis-localization. To further assess whether the p.R163Q variant affected the transcriptional activation function of *EBF3*, a luciferase based reporter assay was performed in HEK293 cells[15]. This experiment in cultured human cells revealed that the $EBF3^{p.R163Q}$ variant failed to activate transcription of the reporter constructs, supporting the LOF model obtained from *Drosophila* experiments.

In parallel to the experimental studies, collaborations with physicians, human geneticists, and genetic counselors at Baylor College of Medicine led to the identification of two additional individuals with similar symptoms. One patient carried the identical p.R163Q variant, and

another patient carried a missense variant that affected the same residue (p.R163L). The p.R163L variant also failed to rescue the fly *kn* mutant[93] suggesting that this allele also affected EBF3 function. Interestingly, this work was published back-to-back with two independent human genetics studies that reported additional individuals with *de novo* missense, nonsense, frameshift and splicing variants in *EBF3* linked to similar neurodevelopmental phenotypes[82,83]. Subsequently, three additional papers were published reporting additional cases of *de novo EBF3* variants and copy number deletion[84–86]. This novel neurodevelopmental syndrome is now known as the Hypotonia, Ataxia, and Delayed Development Syndrome (HADDS, MIM #617330) in the Online Mendelian Inheritance in Man (OMIM~~, www.omim.org~~), an authoritative database for genotype-phenotype relationships in human.

Functional Study of a dominantly inherited missense variant in *TBX2* linked to a syndromic cardiovascular and skeletal developmental disorder

In a small family affected with overlapping spectrums of craniofacial dysmorphism, cardiac anomalies, skeletal malformation, immune deficiency, endocrine abnormalities and developmental impairment, the UDN Duke Clinical Site identified a missense variant (p.R20Q) in *TBX2* that segregates with disease phenotypes[87]. Three (son, daughter and mother) out of four family members are affected by this condition, and the son exhibited the most severe phenotype. Clinically, he met a diagnosis of 'complete DiGeorge syndrome', a condition that is often caused by haploinsufficiency of *TBX1*. While there were no mutations identified in *TBX1* in this family, the clinicians and human geneticists focused on a variant in *TBX2* since previous studies in mice showed that these genes have overlapping functions during development[88]. TBX1 and TBX2 both belong to T-box (TBX) family of transcription factors that can act as transcriptional repressors as well as activators depending on the context. Previously, variants in 12 out of 17 members of the *TBX* family genes were linked to human diseases. The MOSC decided to experimentally pursue this variant based on the following information gathered through MARRVEL and other resources. (1) This variant was reported only once in a cohort of ~90,000 'control' individuals in gnomAD (note that this variant was filtered out in a default view, likely due to low coverage reads). Considering the milder phenotypic presentation of the mother, this still can be considered as a very rare variant that may be responsible for the disease phenotypes. (2) The pLI score of *TBX2* in ExAC/gnomAD are 0.96/0.99 which is high (Max for pLI is 1.00). In addition, the o/e (observed/expected) LOF score in gnomAD is 0.05 (only 1/18.6 expected LOF variant is observed in gnomAD). These numbers suggest that LOF variants in this gene are selected against in the general population. (3) The p.R20 is evolutionarily conserved from *C. elegans* and *Drosophila* to human, suggesting that this may be an important residue for TBX2 function. (4) Multiple programs predict that the variant is likely damaging. Polyphen: Possibly/Probably Damaging, SIFT: Deleterious, CADD Score: 24.4, REVEL Score: 0.5. (5) MO mutants exhibit defects in tissues affected in patients (e.g. knockout mice exhibit defects in cardiovascular system, digestive/alimentary systems, craniofacial, limbs/digit). Hence, together with the biological links between *TBX1* and *TBX2* and the phenotypic links between these patients and DiGeorge Syndrome, it was determined best to perform functional studies of variants in this gene using *Drosophila*.

To begin to assess whether the p.R20Q variant affects *TBX2* function, a T2A-GAL4 line in *bifid* (*bi*), the *Drosophila* ortholog of human *TBX2*, was generated via RMCE of a coding intronic MiMIC (**Figure 2**)[87]. This allele, $bi^{T2A\text{-}GAL4}$, was recessive pupal lethal and behaved as a strong LOF mutant similar to previously reported *bi* LOF alleles (e.g. bi^{D2}, bi^{D4}) (**Figure 2E**). The lethality of these classic and newly generated *bi* alleles were rescued by an ~80kb genomic rescue construct carrying the entire *bi* locus, indicating that these reagents are indeed clean LOF alleles. The expression pattern of GAL4 in the $bi^{T2A\text{-}GAL4}$ line also matched well with previously reported patterns of *bi* expression in multiple tissues including in the wing imaginal disc (**Figure 2D**). In parallel, UAS-transgenic lines for *TBX2* carrying the reference or variant (p.R20Q) sequences were generated. Unfortunately, neither transgene was able to rescue lethality of the $bi^{T2A\text{-}GAL4}$ line. Importantly, a wild-type fly *UAS-bi* transgene also failed to rescue the $bi^{T2A\text{-}GAL4}$ allele, likely due to the dosage-sensitivity of this gene. Indeed, over-expression of *UAS-bi*$^{+}$ as well as *UAS-TBX2*$^{+}$ caused some degree of lethality when overexpressed in a wild-type animal. This toxic effect of *bi/TBX2* over-expression was utilized as a functional assay to assess whether the p.R20Q variant may affect TBX2 function. Since the *Drosophila bi* gene has been extensively studied in the context of the visual system (gene is also known as *optomotor blind* (*omb*)), phenotypes related to the visual system were investigated extensively. When the reference *TBX2* was expressed using an *ey-GAL4* driver that expresses UAS-transgenes in the eye as well as in parts of the brain relevant to the visual system, an ~85% lethality was observed (**Figure 3C, right panel**) and significant reduction of eye size (**Figure 4B**). This phenotype was stronger than the phenotype observed when a wild-type fly *UAS-bi* transgene was expressed, suggesting that the human TBX2 is more detrimental to the fly when overexpressed. Interestingly, the p.R20Q TBX2 was less potent in causing lethality (**Figure 3C, right panel**) as well as inducing a small eye phenotype (**Figure 4B**) using the same driver under the identical condition[87], suggesting the variant affects protein function. Moreover, the function of photoreceptors over-expressing reference and variant *TBX2* using a different GAL4 driver, *(Rh1-GAL4)* that specifically expresses UAS transgenes in R1-R6 photoreceptors, revealed that the variant TBX2 exhibited a much milder ERG phenotype compared to reference TBX2 (**Figure 4B**)[87]. Interestingly, most of the p.R20Q TBX2 protein was still found in the nucleus similar to the reference protein, suggesting that the variant did not affect nuclear localization. A luciferase based transcriptional repression assay in HEK293T cells further showed the p.R20Q was not able to effectively repress transcription of a reporter construct with palindromic T-box sites[87]. In addition, decreases in protein levels of $TBX2^{p.R20Q}$ were observed compared to $TBX2^{+}$, suggesting the variant may affect translation or protein stability of TBX2, which in turn affects its abundance within a cell.

Additional patients with rare variants in *TBX2* were identified by clinicians at the UDN Duke Clinical Site in parallel with these experimental studies. An 8-year-old boy with a *de novo* missense (p.R305H) variant from an unrelated family exhibited many of the features found in the first family[87]. Additional functional studies in *Drosophila* and human cell lines revealed that this p.R305H variant also affects TBX2 function and protein levels, strongly suggesting that defect in this gene is likely to underlie many of the phenotypes found in the two families. This disorder was recently curated as 'Vertebral anomalies and variable Endocrine and T-cell Dysfunction (VETD, MIM #618223)' in OMIM. Identification of additional individuals with damaging variants

in *TBX2* with overlapping phenotypes is critical to understand the full spectrum of genotype-phenotype relationship for this gene in human disease.

Figure Legends:

Figure 1: **Injection and crossing scheme to generate UAS-human cDNA and T2A-GAL4 lines. (A)** Generation of UAS-human cDNA transgenes through microinjections and crosses. Crossing scheme to integrate the transgenes into a 2nd chromosome docking site (VK37) using male flies in the 1st and 2nd generation are shown as an example. Upon injection of the human cDNA ϕC31 transgenic construct (pGW-HA.attB) into early embryos that contain a germline source of ϕC31 integrase (labeled with both 3xP3-GFP and 3xP3-RFP) and VK37 docking site [labeled with a *yellow*$^{+}$ (y^{+}) marker], one can follow the transgenic events with the *white*$^{+}$ (w^{+}) minigene that is present in the transgenic vector. We recommend the readers to cross out the ϕC31 integrase by selecting against flies with GFP and RFP. The final stable stock can be kept as homozygotes or as a balanced stock if the chromosome carries a 2nd site lethal/sterile hit mutation. Presence of 2nd site lethal/sterile mutations on a transgenic constructs usually does not affect the outcome of functional studies as long as these transgenes are used in a heterozygous state (see **Figure 3**). **(B)** Generation of T2A-GAL4 lines through microinjection and crosses. Crossing scheme to convert a 2nd chromosome MiMIC insertion into a T2A-GAL4 element is shown here as an example. By microinjecting an expression vector for ϕC31 integrase and a RMCE vector for T2A-GAL4 (pBS-KS-attB2-SA-T2A-Gal4-Hsp70, must select appropriate reading frame for the MiMIC of interest. See the following papers for details[57,59]) into embryos carrying a MiMIC in a coding intron in gene of interest, one can convert the original MiMIC into a T2A-GAL4 line. See **Figure 2A** for a schematic diagram of the RMCE conversion. The conversion event can be selected by screening against the y^{+} marker in the original MiMIC cassette[60]. Since RMCE can happen in two directions, only 50% of the successful conversion event will lead to successful production of GAL4, which can be detected by a UAS-GFP reporter transgene in the next generation. The final stable stock can be kept as homozygotes or as a balanced stock if the LOF of the gene is lethal/sterile.

Figure 2: **Conversion of MiMIC elements into T2A-GAL4 lines via RMCE**. **(A)** ϕC31 integrase facilitates the recombination between the two *attP* sites in the fly **(A-top)** and the two *attB* sites flanking a T2A-GAL4 cassette shown as a circular vector **(A-bottom)**. **(B)** Successful RMCE events lead to a loss of a selectable marker (*yellow+*), and insertion of the T2A-GAL4 cassette in the same orientation of the gene of interest. Since the RMCE event can happen in two orientations, only 50% of the RMCE reaction will give a desired product. RMCE product inserted in the opposite orientation will not function as a gene-trap allele and will not express GAL4. The directionality of the construct must be confirmed via Sanger sequencing. **(C)** Transcription **(C-top)** and translation **(C-bottom)** of the gene of interest leads to generation of a truncated mRNA and protein due to the polyA signal present at the 3' end of the T2A-GAL4 cassette. The T2A is a ribosome skipping signal, which allows the ribosome to halt and reinitiate translation after this signal. This is used to generate a GAL4 element that is not covalently attached to the truncated gene product of interest. The GAL4 will enter the nucleus and will facilitate the transcription of transgenes that are under the control of UAS elements. UAS-GFP can be used as a gene expression reporter, and UAS-human cDNA can be used for rescue experiments via gene 'humanization'. **(D)** Example of a T2A-GAL4 element in *bi* driving expression of UAS-GFP shown on the top. This expression pattern

Commented [A63]: Please move up the figure legends so that they appear after the representative results and before the discussion. Please re-order the references accordingly.

Commented [A64R63]: Done

resembles a previously generated enhancer trap line for the same gene (*bi*$^{omb\text{-}GAL4}$) shown on the bottom. **(E)** Comparison of T2A-GAL4 allele of *bi* with previously reported LOF *bi* alleles. This figure has been adopted and modified from [57,87].

Figure 3: **Functional analysis of human variants using rescue-based (left) and over-expression-based (right) studies**. **(A-left panel)**. The function of *EBF3* variants was assessed with a rescue-based analysis of the fly *knot (kn)* LOF allele focusing on lethality/viability. **(A-right panel)** The function of variants in *TBX2* was assessed by performing over-expression of human *TBX2* transgenes in wild-type flies, focusing on lethality/viability as well as eye morphology and electrophysiology phenotypes (see **Figure 4**). **(B)** Crossing schemes to obtain the flies that would be tested in the functional studies. One should always use a neutral UAS element (e.g. *UAS-lacZ*, *UAS-GFP*) as a control experiment. **(C)** Representative results from functional studies of *EBF3*$^{p.R163Q}$ and *TBX2*$^{p.R20Q}$ variants, respectively, along with appropriate control experiments that are necessary to interpret the results. Both the rescue-based analysis and over-expression studies reveal that the variants behave as amorphic or hypomorphic alleles. The lethality/viability data shown here are based on the experimental data presented in [15,87].

Figure 4: **Functional analysis of a rare missense variant in human *TBX2* based on eye morphology and electroretinogram in *Drosophila***. **(A)** A schematic image showing the typical placement of recording and reference electrodes on the fly eye along with a representative electroretinogram recording with four major components (on-transient, depolarization, off-transient, repolarization). **(B)** *TBX2* variant (p.R20Q) functions as a partial LOF allele based on over-expression studies in the fly eye using GAL4 drivers specific to the visual system (*ey-GAL4* and *Rh1-GAL4*) showed that the reference TBX2 caused a strong morphological and electrophysiological phenotype compared to the variant protein. **(B-top panels)** A severe reduction in eye size is seen upon over-expression of *UAS-TBX2*$^{+}$ with *ey-GAL4*. *UAS-TBX2*$^{p.R20Q}$ driven with *ey-GAL4* also causes a smaller eye but the phenotype is much milder. **(B-bottom panels)** When *UAS-TBX2*$^{+}$ is expressed in core R1-R6 photoreceptors using Rh1-GAL4, there is a loss of the on-transient and off-transient, reduced depolarization, and a large abnormal prolonged depolarization after potential (PDA) phenotype that is not seen in control flies. These phenotypes are not as severe when UAS-TBX2$^{p.R20Q}$ is expressed using the same *Rh1-GAL4*. This figure has been adopted and modified from [69,87].

~~**Table 1:**~~

Commented [A65]: Please add a legend and also cite it in text.

Commented [A66R65]: The two tables we have are Tables for materials and online resources.

Supplemental Table 1 lists contributing members of the UDN since the author list would be several dozen members long if presented there. It is mentioned under "%" under the Authors section.

Discussion:

Experimental studies using *Drosophila melanogaster* provide a robust assay system to assess the consequence of disease associated human variants thanks to the large body of knowledge and diverse genetic tools that have been generated by many researchers in the fly field over the past century[89]. Just like any other experimental system, however, it is important to acknowledge the caveats and limitations that exist.

Caveats associated with data mining

Although the first step in this protocol is to mine databases for information pertaining to a gene of interest, it is important to use this information only as a starting point. For example, although *in silico* prediction of variant function provides valuable insights, these data should always be interpreted with caution. There are some instances in which all major algorithms predict that a human variant is benign, yet functional studies in *Drosophila* clearly demonstrated the damaging nature of such variant[24]. Similarly, although protein-protein interaction, co-expression and structural modeling data are all insightful pieces of information, there may be pseudo-positive and pseudo-negative information present in these large 'omics' data sets. For example, some of the previously identified or predicted protein-protein interactions may be artificial or can only be seen in certain cell or tissue types. In addition, there may be many false negative interactions that are not captured in these data sets since certain protein-protein interactions are transient (e.g. enzyme-substrate interactions). Experimental validation is critical to demonstrate that certain genes or proteins genetically or physically interact *in vivo* and in the biological context of interest. Similarly, structures predicted based on homology modeling should truly be treated as a 'model' rather than a solved structure. Although this information could be useful if one finds that an amino acid of interest is present in a structurally important part of the protein, negative data does not rule out the possibility that the variant may be damaging. Finally, some of the previously reported genotype-phenotype information may also need to be treated with caution since some information archived in public database may not be accurate. For example, some information in MO databases are based on experiments that have been well controlled and performed rigorously, whereas others may have come from a large screen paper that did not perform further follow-up studies with stringent controls.

'Humanization' experiments using T2A-GAL4 strategy may not always be successful

While rescue and over-expression based functional studies using human cDNAs allow assessment of variants in the context of the human protein, this approach is not always successful. If a reference human cDNA cannot rescue the fly mutant phenotype, there are two probable explanations. The first possibility is that the human protein is nonfunctional or has significantly reduced activity in the context of a fly cell. This could be due to reduced protein expression, stability, activity and/or localization, or could be due to the lack of compatibility with fly proteins that work in a multi-protein complex. Since the UAS/GAL4 system is temperature sensitive, one can raise the flies at a relatively high temperature (e.g. 29°C) to see if one may be able to see a rescue in this condition. In addition, one can also generate a UAS-fly cDNA construct and transgene as a positive control. If the variant of interest affects a conserved amino acid, the

analogous variant can be introduced into the fly cDNA for functional study of the variant in the context of the fly ortholog. Although this is not absolutely necessary, it greatly helps the study in case the experiments using human cDNA transgenic lines give negative or inconclusive results (**Figure 3**). The second possibility is that the expression of the human protein causes some sort of cellular or organism level toxicity. This could be due to an antimorphic effect (e.g. acting as a dominant negative protein), hypermorphic effect (e.g. too much activity), or neomorphic effect (e.g. gain of a novel toxic function such as protein aggregation that is not always related to the endogenous function of the gene of interest). In this case, keeping the flies in a low temperature (e.g. 18°C) may alleviate some of these problems. Finally, there are some scenarios in which the over-expression of a fly cDNA may not even rescue the fly T2A-GAL4 line as seen in the *TBX2* example, likely due to the strict dosage dependence of the gene product. To avoid the over-expression of a protein of interest, one can modify the fly gene of interest directly via CRISPR or engineer a genomic rescue construct that contains the variant of interest and perform rescue experiments using a LOF allele[21]. For small genes, one can also consider 'humanizing' the fly genomic rescue construct to test human variants that affect non-conserved amino acids[24]. In summary, one should always look for alternate strategies when the humanization experiment does not allow functional assessment of the variant of interest.

Things to note when interpreting negative and positive results

If both the reference and variant human cDNAs rescue the fly mutant phenotypes to a similar degree, and there is no difference observed in all conditions tested, then one may assume that the variant is functionally indistinguishable in *Drosophila in vivo*. It is important to note, however, that this information is not sufficient to rule out that the variant of interest is non-pathogenic since the *Drosophila* assay may not be sensitive enough or may not capture all potential functions of the gene/protein of interest that matter in humans. Positive data, on the other hand, is a strong indication that the variant has damaging consequences on protein function, but this data alone is still not sufficient to claim pathogenicity. American College of Medical Genetics and Genomics (ACMG) has published a set of standards and guidelines to classify variants in human disease associated genes into "benign", "likely benign", "variant of unknown significance (VUS)", "likely pathogenic" and "pathogenic"[90]. Although this classification only applies to established disease-associated genes and not directly applicable to variants in 'genes of uncertain significance (GUS)', all individuals who are involved in human variant functional studies are strongly encouraged to read and adhere to this guideline when reporting variant function.

Extracting useful biological information when MO phenotypes do not 'model' the human disease condition

It is important to keep in mind that over-expression based functional assays have their own limitations, especially since some of the phenotypes being scored may have little relevance to the disease condition of interest. Similarly, the phenotypes that are being assessed through rescue experiments may not have any direct relevance to the disease of interest. Since these experiments are conducted outside the endogenous contexts in an invertebrate system, they should not be considered as 'disease models' but rather as a gene function test using *Drosophila* as a 'living test tube'.

Even if the model organism does not mimic a human disease condition, scorable phenotypes used in rescue experiments can often provide useful biological insights into the disease conditions. The concept of 'phenologs (non-obvious homologous phenotypes)' (www.phenologs.org)[91] can be used to further determine the underlying molecular connection between the *Drosophila* and human phenotypes. For example, morphological phenotypes in the fly wing, thorax, legs and eyes are excellent phenotypic readouts for defects in Notch signaling pathway, an evolutionarily conserved pathway linked to many congenital disorders including cardiovascular defects in humans[62]. By understanding the molecular logic behind certain phenotypes in *Drosophila*, one may identify hidden biological links between genes and phenotypes in humans that have yet to be understood.

Continuous communication with your clinical collaborator

When working with clinicians to study the function of a rare variant found in patient, it is very important to establish a strong collaborative relationship. Although clinical and basic biomedical researchers may share interests in the same gene or genetic pathways, there is a large cultural and linguistic (e.g. medical jargon, model organism specific nomenclature) gap between the clinical and scientific fields. A strong, trust-based relationship between the two parties can be built through extensive communication. Furthermore, bidirectional communication is critical to establish and maintain this relationship. For example, in the two cases described in the representative results section, identification of additional patients with similar genotype and phenotype, and subsequent functional study was critical to prove the pathogenicity of the variants of interest. Even with strong functional data, researchers and clinicians often have a hard time to convince human geneticists that a variant identified in "n=1" cases is the true cause of disease.

Once the MO researcher identifies that a variant of interest is damaging, it is critical to communicate the information back to the clinical collaborators as soon as possible so they can more actively try to identify matching cases by networking with other clinicians and human geneticists around the world. Tools such as Geno$_2$MP (Genotypes to Mendelian Phenotypes, geno2mp.gs.washington.edu), a de-identified database of 9,650 individuals (as of Sep 2018) enrolled in the University of Washington's Center for Mendelian Genomics Study[41] that includes patients and family members suspected of having genetic disorders, can be searched to assess whether there are individuals that may have the same disorder and contact the lead clinician using their messaging feature. Alternatively, one can use GeneMatcher (www.genematcher.org), a matchmaking website for clinicians, basic researchers and patients who share interest in the same gene, to identify additional patients that carry rare variants in the same gene. Since GeneMatcher is part of a larger integrative network of matchmaking websites called Matchmaker Exchange (www.matchmakerexchange.org)[42], one can also search additional databases around the world including Australian Genomics Health Alliance Patient Archive (mme.australiangenomics.org.au/#/home), Broad Matchbox (seqr.broadinstitute.org/matchmaker/matchbox), DECIPHER (decipher.sanger.ac.uk), MyGene2 (www.mygene2.org/MyGene2) and PhenomeCentral (phenomecentral.org) in a single GeneMatcher gene submission. Although one can participate in GeneMatcher as a "researcher", we recommend the basic scientists to utilize this website together with their clinical collaborators

Commented [A67]: Please move all weblinks into the reference list.

Commented [A68R67]: Done.

since communication with other clinicians that must happen after the match is made requires certain level of medical expertise.

Acknowledgements:

We thank Jose Salazar, Julia Wang and Dr. Karen Schulze for critical reading of the manuscript. We acknowledge Drs. Ning Liu and Xi Luo for the functional characterization of the *TBX2* variants discussed here. Undiagnosed Diseases Network Model Organisms Screening Center was supported through the National Institutes of Health (NIH) Common Fund (U54 NS093793). HTC was further supported by the NIH[CNCDP-K12 and NINDS (1K12 NS098482)], American Academy of Neurology (Neuroscience Research grant), Burroughs Wellcome Fund (Career Award for Medical Scientists), Child Neurology Society and Child Neurology Foundation (PERF Elterman grant), and the NIH Director's Early Independence Award (DP5 OD026426). MFW was further supported by Simons Foundation (SFARI Award: 368479). SY was further supported by the NIH (R01 DC014932), the Simons Foundation (SFARI Award: 368479), the Alzheimer's Association (New Investigator Research Grant: 15-364099), Naman Family Fund for Basic Research and Caroline Wiess Law Fund for Research in Molecular Medicine. Confocal microscopy at BCM is supported in part by NIH Grant U54HD083092 to the Intellectual and Developmental Disabilities Research Center (IDDRC) Neurovisualization Core.

Disclosures:

The authors have nothing to disclose.

Figure Legends:

Figure 1: **Injection and crossing scheme to generate UAS-human cDNA and T2A-GAL4 lines. (A)** Generation of UAS-human cDNA transgenes through microinjections and crosses. Crossing scheme to integrate the transgenes into a 2nd chromosome docking site (VK37) using male flies in the 1st and 2nd generation are shown as an example. Upon injection of the human cDNA ϕC31 transgenic construct (pGW-HA.attB) into early embryos that contain a germline source of ϕC31 integrase (labeled with both 3xP3-GFP and 3xP3-RFP) and VK37 docking site [labeled with a *yellow*$^{+}$ (*y*$^{+}$) marker], one can follow the transgenic events with the *white*$^{+/-}$ (*w*$^{+}$) minigene that is present in the transgenic vector. We recommend the readers to cross out the ϕC31 integrase by selecting against flies with GFP and RFP. The final stable stock can be kept as homozygotes or as a balanced stock if the chromosome carries a 2nd site lethal/sterile hit mutation. Presence of 2nd site lethal/sterile mutations on a transgenic constructs usually does not affect the outcome of functional studies as long as these transgenes are used in a heterozygous state (see **Figure 3**). **(B)** Generation of T2A-GAL4 lines through microinjection and crosses. Crossing scheme to convert a 2nd chromosome MiMIC insertion into a T2A-GAL4 element is shown here as an example. By microinjecting an expression vector for ϕC31 integrase and a RMCE vector for T2A-GAL4 (pBS-KS-attB2-SA-T2A-Gal4-Hsp70, must select appropriate reading frame for the MiMIC of interest. See the following papers for details[57,59]) into embryos carrying a MiMIC in a coding intron in gene of interest, one can convert the original MiMIC into a T2A-GAL4 line. See **Figure 2A** for a schematic diagram of the RMCE conversion. The conversion event can be selected by screening against the *y*$^{+}$ marker in the original MiMIC cassette[60]. Since RMCE can happen in two directions, only 50% of the successful conversion event will lead to successful production of GAL4, which can

be detected by a UAS-GFP reporter transgene in the next generation. The final stable stock can be kept as homozygotes or as a balanced stock if the LOF of the gene is lethal/sterile.

Figure 2: **Conversion of MiMIC elements into T2A-GAL4 lines via RMCE**. **(A)** ϕC31 integrase facilitates the recombination between the two *attP* sites in the fly **(A-top)** and the two *attB* sites flanking a T2A-GAL4 cassette shown as a circular vector **(A-bottom)**. **(B)** Successful RMCE events lead to a loss of a selectable marker (*yellow+*), and insertion of the T2A-GAL4 cassette in the same orientation of the gene of interest. Since the RMCE event can happen in two orientations, only 50% of the RMCE reaction will give a desired product. RMCE product inserted in the opposite orientation will not function as a gene-trap allele and will not express GAL4. The directionality of the construct must be confirmed via Sanger sequencing. **(C)** Transcription **(C-top)** and translation **(C-bottom)** of the gene of interest leads to generation of a truncated mRNA and protein due to the polyA signal present at the 3' end of the T2A-GAL4 cassette. The T2A is a ribosome skipping signal, which allows the ribosome to halt and reinitiate translation after this signal. This is used to generate a GAL4 element that is not covalently attached to the truncated gene product of interest. The GAL4 will enter the nucleus and will facilitate the transcription of transgenes that are under the control of UAS elements. UAS-GFP can be used as a gene expression reporter, and UAS-human cDNA can be used for rescue experiments via gene 'humanization'. **(D)** Example of a T2A-GAL4 element in *bi* driving expression of UAS-GFP shown on the top. This expression pattern resembles a previously generated enhancer trap line for the same gene (*bi*$^{omb\text{-}GAL4}$) shown on the bottom. **(E)** Comparison of T2A-GAL4 allele of *bi* with previously reported LOF *bi* alleles. This figure has been adopted and modified from [57,87].

Figure 3: **Functional analysis of human variants using rescue-based (left) and over-expression-based (right) studies**. **(A-left panel)**. The function of *EBF3* variants was assessed with a rescue-based analysis of the fly *knot (kn)* LOF allele focusing on lethality/viability. **(A-right panel)** The function of variants in *TBX2* was assessed by performing over-expression of human *TBX2* transgenes in wild-type flies, focusing on lethality/viability as well as eye morphology and electrophysiology phenotypes (see **Figure 4**). **(B)** Crossing schemes to obtain the flies that would be tested in the functional studies. One should always use a neutral UAS element (e.g. *UAS-lacZ*, *UAS-GFP*) as a control experiment. **(C)** Representative results from functional studies of *EBF3*$^{p.R163Q}$ and *TBX2*$^{p.R20Q}$ variants, respectively, along with appropriate control experiments that are necessary to interpret the results. Both the rescue-based analysis and over-expression studies reveal that the variants behave as amorphic or hypomorphic alleles. The lethality/viability data shown here are based on the experimental data presented in [15,87].

Figure 4: **Functional analysis of a rare missense variant in human *TBX2* based on eye morphology and electroretinogram in *Drosophila***. **(A)** A schematic image showing the typical placement of recording and reference electrodes on the fly eye along with a representative electroretinogram recording with four major components (on-transient, depolarization, off-transient, repolarization). **(B)** *TBX2* variant (p.R20Q) functions as a partial LOF allele based on over-expression studies in the fly eye using GAL4 drivers specific to the visual system (*ey-GAL4* and *Rh1-GAL4*) showed that the reference TBX2 caused a strong morphological and electrophysiological phenotype compared to the variant protein. **(B-top panels)** A severe

reduction in eye size is seen upon over-expression of *UAS-TBX2*$^{+}$ with *ey-GAL4*. *UAS-TBX2*$^{p.R20Q}$ driven with *ey-GAL4* also causes a smaller eye but the phenotype is much milder. **(B-bottom panels)** When *UAS-TBX2*$^{+}$ is expressed in core R1-R6 photoreceptors using Rh1-GAL4, there is a loss of the on-transient and off-transient, reduced depolarization, and a large abnormal prolonged depolarization after potential (PDA) phenotype that is not seen in control flies. These phenotypes are not as severe when UAS-TBX2$^{p.R20Q}$ is expressed using the same *Rh1-GAL4*. This figure has been adopted and modified from [69,87].

Table 1:

References Cited

1. Boycott, K. M., Rath, A., *et al.* International Cooperation to Enable the Diagnosis of All Rare Genetic Diseases. *The American Journal of Human Genetics* **100**, (5)695–705 (2017).
2. Lupski, J. R., Reid, J. G., *et al.* Whole-Genome Sequencing in a Patient with Charcot–Marie–Tooth Neuropathy. *New England Journal of Medicine* **362**, (13)1181–1191 (2010).
3. Boycott, K. M., Vanstone, M. R., Bulman, D. E. & MacKenzie, A. E. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nature Reviews Genetics* **14**, (10)681–691 (2013).
4. Yang, Y., Muzny, D. M., *et al.* Molecular Findings Among Patients Referred for Clinical Whole-Exome Sequencing. *JAMA* **312**, (18)1870 (2014).
5. Lee, H., Deignan, J. L., *et al.* Clinical Exome Sequencing for Genetic Identification of Rare Mendelian Disorders. *JAMA* **312**, (18)1880 (2014).
6. Coban-Akdemir, Z., White, J. J., *et al.* Identifying Genes Whose Mutant Transcripts Cause Dominant Disease Traits by Potential Gain-of-Function Alleles. *The American Journal of Human Genetics* **103**, (2)171–187 (2018).
7. Muller, H. J. Further studies on the nature and causes of gene mutations. *Proceedings of the Sixth International Congress of Genetics* 213–255 (1932).
8. Ghosh, R., Oak, N. & Plon, S. E. Evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. *Genome Biology* **18**, (1)225 (2017).
9. Adzhubei, I. A., Schmidt, S., *et al.* A method and server for predicting damaging missense mutations. *Nature Methods* **7**, (4)248–249 (2010).
10. Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M. & Ng, P. C. SIFT missense predictions for genomes. *Nature Protocols* **11**, (1)1–9 (2016).
11. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research* (2018).doi:10.1093/nar/gky1016
12. Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. & Chan, A. P. Predicting the functional effect of amino acid substitutions and indels. *PloS one* **7**, (10)e46688 (2012).
13. Wangler, M. F., Yamamoto, S., *et al.* Model Organisms Facilitate Rare Disease Diagnosis and Therapeutic Research. *Genetics* **207**, (1)9–27 (2017).
14. Oriel, C. & Lasko, P. Recent Developments in Using Drosophila as a Model for Human Genetic Disease. *International Journal of Molecular Sciences* **19**, (7)2041 (2018).
15. Chao, H.-T., Davids, M., *et al.* A Syndromic Neurodevelopmental Disorder Caused by De Novo Variants in EBF3. *American journal of human genetics* **100**, (1)128–137 (2017).
16. Oláhová, M., Yoon, W. H., *et al.* Biallelic Mutations in ATP5F1D, which Encodes a Subunit of ATP Synthase, Cause a Metabolic Disorder. *American journal of human genetics* **102**, (3)494–504 (2018).
17. Liu, N., Schoch, K., *et al.* Functional variants in TBX2 are associated with a syndromic cardiovascular and skeletal developmental disorder. *Human molecular genetics* **27**, (14)2454–2465 (2018).
18. Marcogliese, P. C., Shashi, V., *et al.* IRF2BPL Is Associated with Neurological Phenotypes. *American journal of human genetics* **103**, (2)245–260 (2018).
19. Ferreira, C. R., Xia, Z.-J., *et al.* A Recurrent De Novo Heterozygous COG4 Substitution Leads to Saul-Wilson Syndrome, Disrupted Vesicular Trafficking, and Altered Proteoglycan

Formatted: Indent: Left: 0", Hanging: 0.25"

Glycosylation. *The American Journal of Human Genetics* **103**, (4)553–567 (2018).
20. Kanca, O., Andrews, J., *et al. De novo* variants in *WDR37* are associated with epilepsy, colobomas and cerebellar hypoplasia. *Americal Journal of Human Genetics* **Submitted**, (2019).
21. Luo, X., Rosenfeld, J. A., *et al.* Clinically severe CACNA1A alleles affect synaptic function and neurodegeneration differentially. *PLOS Genetics* **13**, (7)e1006905 (2017).
22. Chung, H., Wangler, M., *et al.* ACOX1 induces autoimmunity whereas a *de novo* gain of function variant induces elevated ROS and glial loss in humans and flies. *Cell Metabolism* **Submitted**, (2019).
23. Yamamoto, S., Jaiswal, M., *et al.* A Drosophila Genetic Resource of Mutants to Study Mechanisms Underlying Human Genetic Diseases. *Cell* **159**, (1)200–214 (2014).
24. Jakobsdottir, J., van der Lee, S. J., *et al.* Rare Functional Variant in TM2D3 is Associated with Late-Onset Alzheimer's Disease. *PLoS genetics* **12**, (10)e1006327 (2016).
25. Yoon, W. H., Sandoval, H., *et al.* Loss of Nardilysin, a Mitochondrial Co-chaperone for α-Ketoglutarate Dehydrogenase, Promotes mTORC1 Activation and Neurodegeneration. *Neuron* **93**, (1)115–131 (2017).
26. Harel, T., Yoon, W. H., *et al.* Recurrent De Novo and Biallelic Variation of ATAD3A, Encoding a Mitochondrial Membrane Protein, Results in Distinct Neurological Syndromes. *American journal of human genetics* **99**, (4)831–845 (2016).
27. Tan, K. L., Haelterman, N. A., *et al.* Ari-1 Regulates Myonuclear Organization Together with Parkin and Is Associated with Aortic Aneurysms. *Developmental Cell* **45**, (2)226–244.e8 (2018).
28. Ansar, M., Chung, H., *et al.* Visual impairment and progressive phthisis bulbi caused by recessive pathogenic variant in MARK3. *Human Molecular Genetics* **27**, (15)2703–2711 (2018).
29. Ansar, M., Chung, H., *et al.* Bi-allelic Loss-of-Function Variants in DNMBP Cause Infantile Cataracts. *The American Journal of Human Genetics* **103**, (4)568–578 (2018).
30. Wang, J., Al-Ouran, R., *et al.* MARRVEL: Integration of Human and Model Organism Genetic Resources to Facilitate Functional Annotation of the Human Genome. *The American Journal of Human Genetics* **100**, (6)843–853 (2017).
31. Wang, J., Liu, Z., Bellen, H. & Yamamoto, S. MARRVEL, a web-based tool that integrates human and model organism genomics information. *Journal of Visualized Experiments* **Accepted**~~**Submitted**~~, (2019).
32. Mungall, C. J., McMurry, J. A., *et al.* The Monarch Initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species. *Nucleic Acids Research* **45**, (D1)D712–D722 (2017).
33. Hu, Y., Comjean, A., Mohr, S. E., Perrimon, N. & Perrimon, N. Gene2Function: An Integrated Online Resource for Gene Function Discovery. *G3: Genes|Genomes|Genetics* **7**, (8)2855–2858 (2017).
34. Ioannidis, N. M., Rothstein, J. H., *et al.* REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *The American Journal of Human Genetics* **99**, (4)877–885 (2016).
35. Szklarczyk, D., Morris, J. H., *et al.* The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Research* **45**, (D1)D362–

D368 (2017).
36. Hu, Y., Vinayagam, A., *et al.* Molecular Interaction Search Tool (MIST): an integrated resource for mining gene and protein interaction data. *Nucleic Acids Research* **46**, (D1)D567–D574 (2018).
37. Lawson, C. L., Patwardhan, A., *et al.* EMDataBank unified data resource for 3DEM. *Nucleic Acids Research* **44**, (D1)D396–D403 (2016).
38. Bienert, S., Waterhouse, A., *et al.* The SWISS-MODEL Repository—new features and functionality. *Nucleic Acids Research* **45**, (D1)D313–D319 (2017).
39. Webb, B. & Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Current Protocols in Bioinformatics* **54**, 5.6.1-5.6.37 (2016).
40. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10**, (6)845–858 (2015).
41. Bamshad, M. J., Shendure, J. A., *et al.* The Centers for Mendelian Genomics: A new large-scale initiative to identify the genes underlying rare Mendelian conditions. *American Journal of Medical Genetics Part A* **158A**, (7)1523–1525 (2012).
42. Sobreira, N. L. M., Arachchi, H., *et al.* Matchmaker Exchange. *Current Protocols in Human Genetics* **95**, 9.31.1-9.31.15 (2017).
43. Temple, G., Gerhard, D. S., *et al.* The completion of the Mammalian Gene Collection (MGC). *Genome Research* **19**, (12)2324–2333 (2009).
44. Katzen, F. Gateway ® recombinational cloning: a biological operating system. *Expert Opinion on Drug Discovery* **2**, (4)571–589 (2007).
45. Venken, K. J. T., He, Y., Hoskins, R. A. & Bellen, H. J. P[acman]: A BAC Transgenic Platform for Targeted Insertion of Large DNA Fragments in D. melanogaster. *Science* **314**, (5806)1747–1751 (2006).
46. Bischof, J., Björklund, M., Furger, E., Schertel, C., Taipale, J. & Basler, K. A versatile platform for creating a comprehensive UAS-ORFeome library in Drosophila. *Development (Cambridge, England)* **140**, (11)2434–42 (2013).
47. Bischof, J., Sheils, E. M., Björklund, M. & Basler, K. Generation of a transgenic ORFeome library in Drosophila. *Nature Protocols* **9**, (7)1607–1620 (2014).
48. Laible, M. & Boonrod, K. Homemade site directed mutagenesis of whole plasmids. *Journal of visualized experiments : JoVE* (27) (2009).doi:10.3791/1135
49. Balana, B., Taylor, N. & Slesinger, P. A. Mutagenesis and Functional Analysis of Ion Channels Heterologously Expressed in Mammalian Cells. *Journal of Visualized Experiments* (44) (2010).doi:10.3791/2189
50. Bischof, J., Maeda, R. K., Hediger, M., Karch, F. & Basler, K. An optimized transgenesis system for Drosophila using germ-line-specific C31 integrases. *Proceedings of the National Academy of Sciences* **104**, (9)3312–3317 (2007).
51. Ringrose, L. Transgenesis in Drosophila melanogaster. *Methods in molecular biology (Clifton, N.J.)* **561**, 3–19 (2009).
52. Venken, K. J. T., He, Y., Hoskins, R. A. & Bellen, H. J. P[acman]: A BAC Transgenic Platform for Targeted Insertion of Large DNA Fragments in D. melanogaster. *Science* **314**, (5806)1747–1751 (2006).
53. Groth, A. C., Fish, M., Nusse, R. & Calos, M. P. Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. *Genetics* **166**, (4)1775–82 (2004).

54. Greenspan, R. *Fly Pushing: The Thory and Practice of Drosophila Genetics*. (Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York, 2004).
55. Ashburner, M., Golic, K. & Hawley, R. S. *Drosophila: A Laboratory Handbook*. (Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York, 2005).
56. Diao, F. & White, B. H. A Novel Approach for Directing Transgene Expression in Drosophila: T2A-Gal4 In-Frame Fusion. *Genetics* **190**, (3)1139–1144 (2012).
57. Diao, F., Ironfield, H., *et al.* Plug-and-Play Genetic Access to Drosophila Cell Types using Exchangeable Exon Cassettes. *Cell Reports* **10**, (8)1410–1421 (2015).
58. Bellen, H. J., Levis, R. W., *et al.* The *Drosophila* Gene Disruption Project: Progress Using Transposons With Distinctive Site Specificities. *Genetics* **188**, (3)731–743 (2011).
59. Lee, P.-T., Zirin, J., *et al.* A gene-specific T2A-GAL4 library for Drosophila. *eLife* **7**, (2018).
60. Venken, K. J. T., Schulze, K. L., *et al.* MiMIC: a highly versatile transposon insertion resource for engineering Drosophila melanogaster genes. *Nature methods* **8**, (9)737–43 (2011).
61. Li-Kroeger, D., Kanca, O., *et al.* An expanded toolkit for gene tagging based on MiMIC and scarless CRISPR tagging in Drosophila. *eLife* **7**, (2018).
62. Salazar, J. L. & Yamamoto, S. Integration of Drosophila and Human Genetics to Understand Notch Signaling Related Diseases. *Advances in experimental medicine and biology* **1066**, 141–185 (2018).
63. Wangler, M. F., Yamamoto, S. & Bellen, H. J. Fruit Flies in Biomedical Research. *Genetics* **199**, (3)639–653 (2015).
64. Duffy, J. B. GAL4 system indrosophila: A fly geneticist's swiss army knife. *genesis* **34**, (1–2)1–15 (2002).
65. Nagarkar-Jaiswal, S., Lee, P.-T., *et al.* A library of MiMICs allows tagging of genes and reversible, spatial and temporal knockdown of proteins in Drosophila. *eLife* **4**, (2015).
66. Dolph, P., Nair, A. & Raghu, P. Electroretinogram Recordings of Drosophila. *Cold Spring Harbor Protocols* **2011**, (1)pdb.prot5549-pdb.prot5549 (2011).
67. Lauwers, E. & Verstreken, P. Assaying Mutants of Clathrin-Mediated Endocytosis in the Fly Eye. *Methods in molecular biology (Clifton, N.J.)* **1847**, 109–119 (2018).
68. Rhodes-Mordov, E., Samra, H. & Minke, B. Electroretinogram (ERG) Recordings from Drosophila. *BIO-PROTOCOL* **5**, (21) (2015).
69. Deal, S. & Yamamoto, S. Unraveling novel mechanisms of neurodegeneration through a large-scale forward genetic screen in Drosophila. *Frontiers in Genetics* **In press**, (2019).
70. Chouhan, A. K., Guo, C., *et al.* Uncoupling neuronal death and dysfunction in Drosophila models of neurodegenerative disease. *Acta Neuropathologica Communications* **4**, (1)62 (2016).
71. Lek, M., Karczewski, K. J., *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, (7616)285–291 (2016).
72. Liberg, D., Sigvardsson, M. & Akerblad, P. The EBF/Olf/Collier family of transcription factors: regulators of differentiation in cells originating from all three embryonal germ layers. *Molecular and cellular biology* **22**, (24)8389–97 (2002).
73. Prasad, B. C., Ye, B., Zackhary, R., Schrader, K., Seydoux, G. & Reed, R. R. unc-3, a gene required for axonal guidance in Caenorhabditis elegans, encodes a member of the O/E family of transcription factors. *Development (Cambridge, England)* **125**, (8)1561–8 (1998).
74. Jinushi-Nakao, S., Arvind, R., Amikura, R., Kinameri, E., Liu, A. W. & Moore, A. W. Knot/Collier

and Cut Control Different Aspects of Dendrite Cytoskeleton and Synergize to Define Final Arbor Shape. *Neuron* **56**, (6)963–978 (2007).
75. Pozzoli, O., Bosetti, A., Croci, L., Consalez, G. G. & Vetter, M. L. Xebf3 is a regulator of neuronal differentiation during primary neurogenesis in Xenopus. *Developmental biology* **233**, (2)495–512 (2001).
76. Wang, S. S., Lewcock, J. W., Feinstein, P., Mombaerts, P. & Reed, R. R. Genetic disruptions of O/E2 and O/E3 genes reveal involvement in olfactory receptor neuron projection. *Development* **131**, (6)1377–1388 (2004).
77. Fulp, C. T., Cho, G., Marsh, E. D., Nasrallah, I. M., Labosky, P. A. & Golden, J. A. Identification of Arx transcriptional targets in the developing basal forebrain. *Human Molecular Genetics* **17**, (23)3740–3760 (2008).
78. Gécz, J., Cloosterman, D. & Partington, M. ARX: a gene for all seasons. *Current Opinion in Genetics & Development* **16**, (3)308–316 (2006).
79. Dubois, L. & Vincent, A. The COE--Collier/Olf1/EBF--transcription factors: structural conservation and diversity of developmental functions. *Mechanisms of development* **108**, (1–2)3–12 (2001).
80. Cook, R. K., Christensen, S. J., *et al.* The generation of chromosomal deletions to provide extensive coverage and subdivision of the Drosophila melanogaster genome. *Genome Biology* **13**, (3)R21 (2012).
81. Chao, H.-T., Davids, M., *et al.* A Syndromic Neurodevelopmental Disorder Caused by De Novo Variants in EBF3. *The American Journal of Human Genetics* **100**, (1)128–137 (2017).
82. Sleven, H., Welsh, S. J., *et al.* De Novo Mutations in EBF3 Cause a Neurodevelopmental Syndrome. *The American Journal of Human Genetics* **100**, (1)138–150 (2017).
83. Harms, F. L., Girisha, K. M., *et al.* Mutations in EBF3 Disturb Transcriptional Profiles and Cause Intellectual Disability, Ataxia, and Facial Dysmorphism. *The American Journal of Human Genetics* **100**, (1)117–127 (2017).
84. Tanaka, A. J., Cho, M. T., *et al.* De novo variants in *EBF3* are associated with hypotonia, developmental delay, intellectual disability, and autism. *Molecular Case Studies* **3**, (6)a002097 (2017).
85. Blackburn, P. R., Barnett, S. S., *et al.* Novel de novo variant in *EBF3* is likely to impact DNA binding in a patient with a neurodevelopmental disorder and expanded phenotypes: patient report, in silico functional assessment, and review of published cases. *Molecular Case Studies* **3**, (3)a001743 (2017).
86. Lopes, F., Soares, G., Gonçalves-Rocha, M., Pinto-Basto, J. & Maciel, P. Whole Gene Deletion of EBF3 Supporting Haploinsufficiency of This Gene as a Mechanism of Neurodevelopmental Disease. *Frontiers in Genetics* **8**, 143 (2017).
87. Liu, N., Schoch, K., *et al.* Functional variants in TBX2 are associated with a syndromic cardiovascular and skeletal developmental disorder. *Human molecular genetics* **27**, (14)2454–2465 (2018).
88. Mesbah, K., Rana, M. S., *et al.* Identification of a Tbx1/Tbx2/Tbx3 genetic pathway governing pharyngeal and arterial pole morphogenesis. *Human Molecular Genetics* **21**, (6)1217–1229 (2012).
89. Bellen, H. J. & Yamamoto, S. Morgan's legacy: fruit flies and the functional annotation of conserved genes. *Cell* **163**, (1)12–4 (2015).

90. Richards, S., Aziz, N., *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine* **17**, (5)405–423 (2015).
91. McGary, K. L., Park, T. J., Woods, J. O., Cha, H. J., Wallingford, J. B. & Marcotte, E. M. Systematic discovery of nonobvious human disease models through orthologous phenotypes. *Proceedings of the National Academy of Sciences of the United States of America* **107**, (14)6544–9 (2010).
92. Yamamoto S., Jaiswal M., et al., A Drosophila Genetic Resource of Mutants to Study Mechanisms Underlying Human Genetic Diseases
Cell **159**, 200-214 (2014).
93. Ausubel, F. M. *Current protocols in molecular biology*. (Greene Publishing and Wiley-Interscience: New York, 1989).
94. Rubin, G. & Spradling, A. Genetic transformation of Drosophila with transposable element vectors. *Science* **218**(4570), 348–353, doi:10.1126/science.6289436 (1982).
95. Sun, Y., Sriramajayam, K., Luo, D. & Liao, D. J. A Quick, Cost-Free Method of Purification of DNA Fragments from Agarose Gel. *Journal of Cancer* **3**, 93–95, doi:10.7150/jca.4163 (2012).
96. Ronaghi, M. DNA SEQUENCING:A Sequencing Method Based on Real-Time Pyrophosphate. *Science* **281** (5375), 363–365, doi:10.1126/science.281.5375.363 (1998).
97. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77** (1), 51–59, doi:10.1016/0378-1119(89)90358-2 (1989).

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: (Default) +Body (Calibri)

Members of the Undiagnosed Diseases Network

Maria T. Acosta
David R. Adams
Pankaj Agrawal
Mercedes E. Alejandro
Patrick Allard
Euan A. Ashley
Mahshid S. Azamian
Carlos A. Bacino
Guney Bademci
Eva Baker
Ashok Balasubramanyam
Dustin Baldridge
Deborah Barbouth
Gabriel F. Batzli
Alan H. Beggs
Hugo J. Bellen
Jonathan A. Bernstein
Gerard T. Berry
Anna Bican
David P. Bick
Camille L. Birch
Stephanie Bivona
Carsten Bonnenmann
Devon Bonner
Braden E. Boone
Bret L. Bostwick
Lauren C. Briere
Elly Brokamp
Donna M. Brown
Matthew Brush
Elizabeth A. Burke
Lindsay C. Burrage
Manish J. Butte
Olveen Carrasquillo
Ta Chen Peter Chang
Hsiao-Tuan Chao
Gary D. Clark
Terra R. Coakley
Laurel A. Cobban
Joy D. Cogan
F. Sessions Cole
Heather A. Colley
Cynthia M. Cooper
Heidi Cope
William J. Craigen
Precilla D'Souza
Surendra Dasari
Mariska Davids
Jean M. Davidson

Jyoti G. Dayal
Esteban C. Dell'Angelica
Shweta U. Dhar
Naghmeh Dorrani
Daniel C. Dorset
Emilie D. Douine
David D. Draper
Annika M. Dries
Laura Duncan
David J. Eckstein
Lisa T. Emrick
Christine M. Eng
Gregory M. Enns
Cecilia Esteves
Tyra Estwick
Liliana Fernandez
Carlos Ferreira
Elizabeth L. Fieg
Paul G. Fisher
Brent L. Fogel
Irman Forghani
Noah D. Friedman
William A. Gahl
Rena A. Godfrey
Alica M. Goldman
David B. Goldstein
Jean-Philippe F. Gourdine
Alana Grajewski
Catherine A. Groden
Andrea L. Gropman
Melissa Haendel
Rizwan Hamid
Neil A. Hanchard
Nichole Hayes
Frances High
Ingrid A. Holm
Jason Hom
Alden Huang
Yong Huang
Rosario Isasi
Fariha Jamal
Yong-hui Jiang
Jean M. Johnston
Angela L. Jones
Lefkothea Karaviti
Emily G. Kelley
Dana Kiley
David M. Koeller
Isaac S. Kohane
Jennefer N. Kohler
Deborah Krakow

Donna M. Krasnewich
Susan Korrick
Mary Koziura
Joel B. Krier
Jennifer E. Kyle
Seema R. Lalani
Byron Lam
Brendan C. Lanpher
Ian R. Lanza
C. Christopher Lau
Jozef Lazar
Kimberly LeBlanc
Brendan H. Lee
Hane Lee
Roy Levitt
Shawn E. Levy
Richard A. Lewis
Sharyn A. Lincoln
Pengfei Liu
Xue Zhong Liu
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
Martin G. Martin
Julian A. Martínez-Agosto
Shruti Marwaha
Thomas May
Jacob McCauley
Allyn McConkie-Rosell
Colleen E. McCormack
Alexa T. McCray
Jason D. Merker
Thomas O. Metz
Matthew Might
Eva Morava-Kozicz
Paolo M. Moretti
Marie Morimoto
John J. Mulvihill
David R. Murdock
Avi Nath
Stan F. Nelson
J. Scott Newberry

John H. Newman
Sarah K. Nicholas
Donna Novacic
Devin Oglesbee
James P. Orengo
Stephen Pak
J. Carl Pallais
Christina GS. Palmer
Jeanette C. Papp
Neil H. Parker
John A. Phillips III
Jennifer E. Posey
John H. Postlethwait
Lorraine Potocki
Barbara N. Pusey
Genecee Renteria
Chloe M. Reuter
Lynette Rives
Amy K. Robertson
Lance H. Rodan
Jill A. Rosenfeld
Robb K. Rowley
Ralph Sacco
Jacinda B. Sampson
Susan L. Samson
Mario Saporta
Judy Schaechter
Timothy Schedl
Kelly Schoch
Daryl A. Scott
Lisa Shakachite
Prashant Sharma
Vandana Shashi
Kathleen Shields
Jimann Shin
Rebecca Signer
Catherine H. Sillari
Edwin K. Silverman
Janet S. Sinsheimer
Kathy Sisco
Kevin S. Smith
Lilianna Solnica-Krezel
Rebecca C. Spillmann
Joan M. Stoler
Nicholas Stong
Jennifer A. Sullivan
David A. Sweetser
Cecelia P. Tamburro
Queenie K.-G. Tan
Mustafa Tekin
Fred Telischi

Willa Thorson
Cynthia J. Tifft
Camilo Toro
Alyssa A. Tran
Tiina K. Urv
Tiphanie P. Vogel
Daryl M. Waggott
Colleen E. Wahl
Nicole M. Walley
Chris A. Walsh
Melissa Walker
Jennifer Wambach
Jijun Wan
Lee-kai Wang
Michael F. Wangler
Patricia A. Ward
Katrina M. Waters
Bobbie-Jo M. Webb-Robertson
Daniel Wegner
Monte Westerfield
Matthew T. Wheeler
Anastasia L. Wise
Lynne A. Wolfe
Jeremy D. Woods
Elizabeth A. Worthey
Shinya Yamamoto
John Yang
Amanda J. Yoon
Guoyun Yu
Diane B. Zastrow
Chunli Zhao
Stephan Zuchner