A Genetic Screen for Genes That Impact Peroxisomes in *Drosophila* Identifies Candidate Genes for Human Disease

Hillary K. Graves,* Sharayu Jangam,* Kai Li Tan,* Antonella Pignata,* Elaine S. Seto,*
Shinya Yamamoto,*,†,‡,§,1 and Michael F. Wangler*,‡,§,1
*Department of Molecular and Human Genetics, †Department of Neuroscience, ‡Program in Developmental Biology, Baylor College of Medicine, Houston, TX 77030, and §Jan and Dan Duncan Neurological Research Institute, Texas Children Hospital, Houston, TX 77030

ORCID IDs: 0000-0001-7389-0890 (S.J.); 0000-0002-1450-7916 (K.L.T.); 0000-0003-2172-8036 (S.Y.)

ABSTRACT Peroxisomes are subcellular organelles that are essential for proper function of eukaryotic cells. In addition to being the sites of a variety of oxidative reactions, they are crucial regulators of lipid metabolism. Peroxisome loss or dysfunction leads to multi-system diseases in humans that strongly affect the nervous system. In order to identify previously unidentified genes and mechanisms that impact peroxisomes, we conducted a genetic screen on a collection of lethal mutations on the *X* chromosome in *Drosophila*. Using the number, size and morphology of GFP tagged peroxisomes as a readout, we screened for mutations that altered peroxisomes based on clonal analysis and confocal microscopy. From this screen, we identified eighteen genes that cause increases in peroxisome number or altered morphology when mutated. We examined the human homologs of these genes and found that they are involved in a diverse array of cellular processes. Interestingly, the human homologs from the *X*-chromosome collection are under selective constraint in human populations and are good candidate genes particularly for dominant genetic disease. This *in vivo* screening approach for peroxisome defects allows identification of novel genes that impact peroxisomes *in vivo* in a multicellular organism and is a valuable platform to discover genes potentially involved in dominant disease that could affect peroxisomes.

Peroxisomes are subcellular organelles that mediate crucial biological processes in eukaryotic cells, including oxidative reactions, catabolism of very-long-chain fatty acids, catabolism of branched chain fatty acids, synthesis of bile acids, and biosynthesis of plasmalogen lipids (Fujiki *et al.* 2014; Wanders 2014). Human diseases caused by lack of peroxisomes are devastating multisystem diseases that result in severe brain, liver, bone and kidney disease (Wanders and Waterham 2005). These conditions, called peroxisome biogenesis disorders Zellweger-spectrum disorders (PBD-ZSD), are a group of multi-system autosomal recessive disorders with severe central nervous system (CNS) manifestations and as yet no effective treatments exist for the hearing, visual and CNS phenotypes (Klouwer *et al.* 2015; Braverman *et al.* 2016).

Historically, genetic screens for biochemical phenotypes have identified genes implicated in peroxisome-biogenesis, pexophagy, and peroxisomal biochemistry (Subramani 1998; Kao *et al.* 2018). More recently, microscopy-based screens have uncovered genes implicated in peroxisome morphology (Baron *et al.* 2016; Yofe *et al.* 2017). These studies have shown that the pathways that regulate peroxisome dynamics (*i.e.*, peroxisome size and number) remain incompletely understood (Mast *et al.* 2015). This is especially true with regards to the information gained from multicellular organisms as opposed to yeast and cell models because most essential genes are not amenable to peroxisomal studies due to early lethality in development. Although null alleles in some of the *Pex* genes in the fruit fly *Drosophila melanogaster* produce viable

■ Table 1 Hits from the Peroxisome X-pex screen

Fly Gene	Allele (s)	Peroxisomal Phenotype	Human Gene(s)	Biological Function in Fly (UniProt)
fs(1)h (FBgn0004656)	*fs[1]h[A], [B], [C]*	**Category A & B**	**BRD2, BRDT, BRD3, BRD4**	**Transcriptional regulation**
Rbcn-3B (FBgn0023510)	*Rbcn-3B[A], [B]*	**Category A**	**WDR7**	**Vacuolar acidification, Notch signaling**
Coq8 (FBgn0052649)	*Coq8[A],[B]*	**Category A**	**COQ8B**	**Protein kinase, electron transport**
Usp16-45 (FBgn0029763)	*Usp16-45[A],[B]*	**Category A**	**USP45**	**Protein deubiquitination**
mxc (FBgn0260789)	*mxc[A],[B],[C],[D],[E]*	Category A	No Human Ortholog	Transcriptional, hemocyte differentiation and proliferation
Cp7Fb (FBgn0014465)	*Cp7Fb[B]*	Category A	No Human Ortholog	Chorion
Upf1 (FBgn0030354)	*Upf1[A],[B]*	Category A	UPF1	Nonsense mediated decay
Upf2 (FBgn0029992)	*Upf2[A],[B]*	Category A	UPF2	Nonsense mediated decay
Nrg (FBgn0264975)	*Nrg[XB]*	Category A	NRCAM, NFASC, L1CAM, CHL1	Cell adhesion
Fum1 (FBgn0286222)	*Fum1[A]*	Category A	FH	TCA cycle enzyme (mitochondria)
Coq7 (FBgn0029502)	*Coq7[B]*	Category A	COQ7	Ubiquinone biosynthesis (mitochondria)
sgg (FBgn0003371)	*sgg[A],[B],[E]*	Category A	GSK3B, GSK3A	Protein kinase
CG17829 (FBgn0025635)	*CG17829[A],[B]*	Category A	HINFP	Transcriptional regulation
CG3149 (FBgn0027564)	*CG3149[B]*	Category A	RFT1	Glycolipid translocation
Smox (FBgn0025800)	*Smox[B]*	Category A	SMAD3	Transcriptional regulation
PI4KIIIα (FBgn0267350)	*PI4KIIIα [E], [W]*	Category A	PI4KA	Synaptic growth, cell polarity, membrane organization
MTPAP (FBgn0024360)	*MTPAP[A],[B]*	Category A	MTPAP	Mitochondrial transcription
temp (FBgn0027296)	*temp [A], [B] & [D]*	Category A	PTAR1	Rab geranylgeranyltransferase activity
TOTAL	**37 hits**	**18 fly genes**	**23 human genes**	

The Fly gene and specific allele are listed along with the phenotype observed in the screen (Category A, B and C). The human homologs of each gene were identified using DIOPT or HCOP (Hu *et al.* 2011; Braschi *et al.* 2019). Known biological function of the fly protein is listed according to the annotation in UniProt (Uniprot Consortium 2019).

adult flies that that exhibit locomotor defects and reduced longevity (Nakayama *et al.* 2011; Faust *et al.* 2014; Wangler *et al.* 2017a), a number of essential genes that regulate peroxisome biogenesis and homeostasis may remain unstudied. To identify mechanisms that impact peroxisome number and size, we conducted a forward genetic screen on the *X* chromosome in *Drosophila* using lines that we had previously generated and analyzed representing a large collection of recessive lethal mutant lines on Drosophila X-chromosome (Yamamoto *et al.* 2014). We report an analysis of 215 lines from this collection that correspond to 100 genes and screened for peroxisomal phenotypes using GFP tagged peroxisomes (Chao *et al.* 2016; Wangler *et al.* 2017a) in conjunction with clonal analysis, allowing generation of homozygous mutant cells within *Drosophila* larval fat body in an otherwise heterozygous animal to bypass early lethality.

Our screen identified a number of genes not previously implicated in peroxisome dynamics or regulation. In previous studies, we've shown that the genes from this collection are enriched for human disease genes (Yamamoto *et al.* 2014). Based on this we propose our screen results as identifying candidate human disease genes particularly for dominant disease that may impact peroxisomes.

METHODS & MATERIALS

Drosophila X-Chromosome peroxisome (X-Pex) screen

All *X*-linked recessive lethal mutant alleles utilized in this paper listed in Table 1 and Supplemental Material, Table 1 were generated on an isogenized *y¹ w* FRT19A* chromosome using ethyl methanesulfonate (EMS) mutagenesis as described (Haelterman *et al.* 2014; Yamamoto *et al.* 2014; Deal and Yamamoto 2019). These fly strains are publically available from the Kyoto Stock Center (https://kyotofly.kit.jp/stocks/documents/EMS_X_lethals.html) or the

Bloomington Drosophila Stock Center (https://bdsc.indiana.edu/stocks/chemically_induced_mutations/xlethals.html). Heterozygous females (*y¹ w* mut* FRT19A/FM7c Kr-GAL4, UAS-GFP*, mut* indicates the mutation of interest) from these lethal lines were crossed to males of the genotype *hsFLP, Ubi-RFP FRT19A; Actin-GAL4,UAS-GFP-SKL/CyO*, and their embryonic progeny were heat shocked at 0-4 hr after egg laying at 37° for 1 hr. Third larval instar wandering larvae were dissected in PBS and fixed in 4% paraformaldehyde for 20-30 min. Fat bodies were mounted in DAPI (4',6-diamidino-2-phenylindole) containing mounting media (Vectashield), confocal microscopy images were captured on a LSM 710 laser scanning confocal microscope (Zeiss), and processed in Photoshop (Adobe). Dissection of the lines shown in Supplemental Table 1 were performed by a team of three co-authors who each dissected and examined and imaged multiple clones. All three also examined each other's images and a group consensus on the classification was reached. This was then independently verified by two independent co-authors through imaging review. Image quantification on select groups was performed using Image J to ensure reproducibility. All the "hits" presented were also independently dissected and imaged. We did observe in the course of our screen that there were isolated cells with very low GFP in our dissections and this effect appeared random and we attributed it to somatic events, for example a recombination or mutation that mutates the UAS or GAL4. Other possibilities were changes in protein produrence, new somatic mutations that impact the UAS-GAL4 system or somatic mutations in peroxisomal genes occurring in single cells. We dealt with these "background" effects in the fat body through screening multiple clones and having independent observers select the hits.

The fat body experiments were a peroxisome focused secondary screen on a subset of mutants of the larger X-lethal collection reported
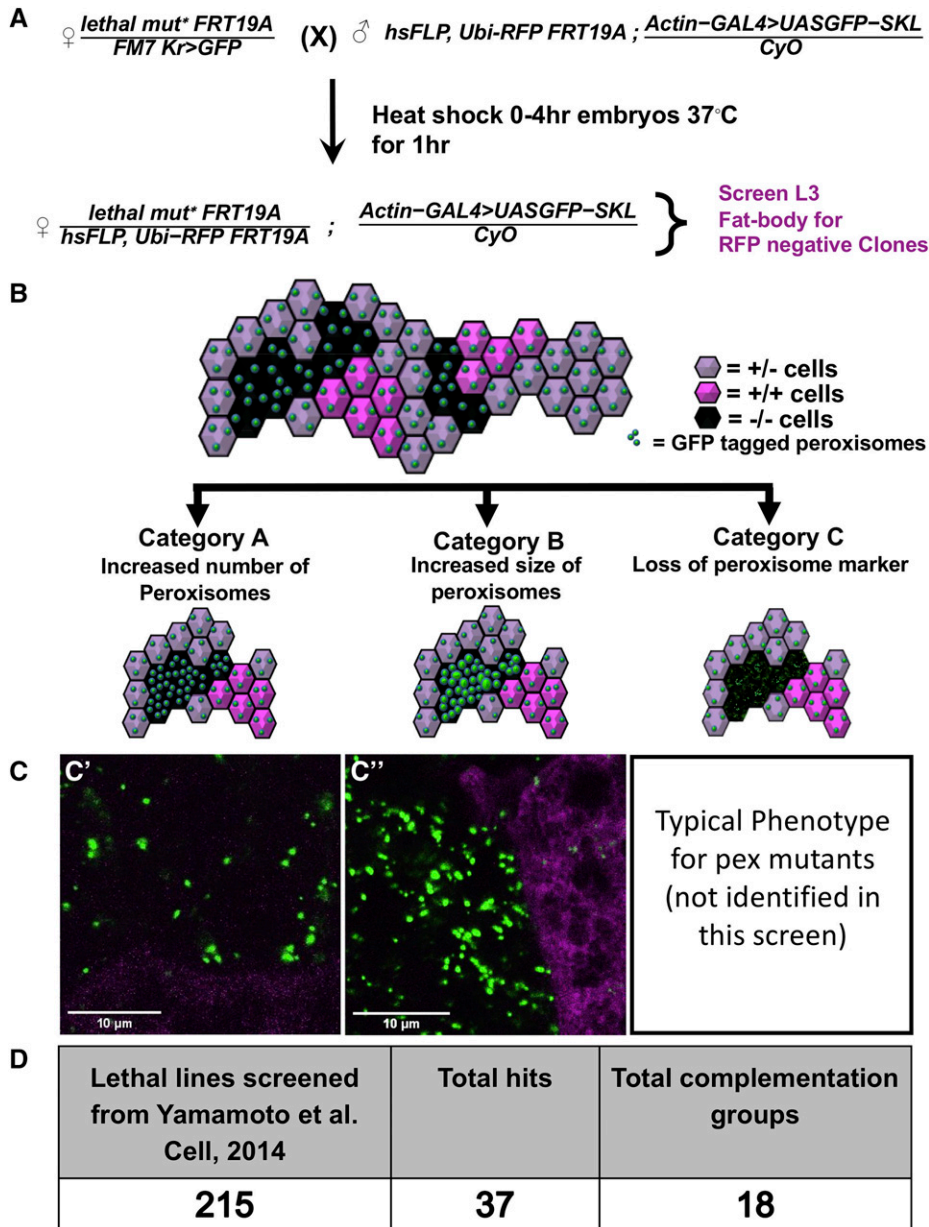
Figure 1 Overall study design and outcomes from the Drosophila X-Pex screen. A. Procedure for the X-Pex screen. Drosophila Crossing Scheme shown in-detail. "*lethal mut**" represents the different X-chromosome recessive lethal alleles used for the screen as listed in Supplemental Table 1. Males and females used for the experiment were crossed at room temperature. Females were allowed to lay eggs for 4 hr and the embryos were then heat-shocked at 37°C in a water-bath for 1hr and then kept at 25°C. The Fat bodies of the wandering third instar larvae were dissected, fixed and imaged by confocal microscopy. The homozygous mutant cells were identified through the absence of RFP (RFP-). B. Schematic representation of fat bodies expressing the GFP tagged peroxisome marker (GFP-SKL) with clones of mutant and wild-type cells. While homozygous mutant cells are marked by the absence of RFP, the sibling homozygous wild-type cell are marked by two doses of RFP (dark magenta). Heterozygous cells are marked with one dose of RFP (pale magenta) Category A represents an increase in peroxisomal numbers, Category B represents enlargement of peroxisomes, and Category C represents a loss of mislocalization of peroxisomal markers. C. Representative images of the categories explained in the Fig. B are shown in C' (*Usp16-45*) and C'' (*fs(h)1*). Detailed analysis and each individual channel for these two images are shown in Figure 3. D. Table representing the overall results from the screen. 215 total lines were screened, 37 total allele hits from 18 genes were identified.

(Yamamoto *et al.* 2014), similar to an Atg8 and *LAMP1* based screen to identify novel autophagy regulators conducted on the same collection (Fang *et al.* 2016). Similar secondary screens were performed on the same X-lethal collection to identify regulators of other biological processes such as ring canal formation and somatic stem cell maintenance during oogenesis (Yamamoto *et al.* 2013; Cook *et al.* 2017) demonstrating the value of this collection in screening for genes involved in diverse cellular processes. We named our screen the "X-Pex" putting the fly X-chromosome together with an abbreviation of peroxisome. We will refer to the gene set and the screen as "X-Pex" throughout the manuscript.

ImageJ Protocol for peroxisomal area counting

Different scans of the original Z-stack images were used to calculate the area of the peroxisomes. The individual scan is divided in GFP-SKL(green layer) and RFP(red layer), converted to black and white and then by finding the edges the area of each peroxisome is calculated with freehand selection tool or wand tool. The detailed protocol is presented in the supplementary file.

Human gene candidate analysis

Human homologs from the fly genes were determined using the Human Gene Nomenclature Orthology Prediction (HCOP, https://www.gene-names.org/tools/hcop/) and the Drosophila RNAi Screening center Integrative Ortholog Prediction Tool (DIOPT, https://www.flyrnai.org/cgi-bin/DRSC_orthologs.pl) tools (Hu *et al.* 2011; Gray *et al.* 2015). These genes were further examined in a series of public human and model organism databases using the MARRVEL tool (http://marrvel.org/) to gather information about the homologous proteins in human and other model organisms (Wang *et al.* 2017). Human gene nomenclature was confirmed using the HGNC (HUGO Gene Nomenclature Committee, https://www.genenames.org) database (Braschi *et al.* 2019). Mendelian disease links were explored in the OMIM (https://www.omim.org/) database (Amberger *et al.* 2015), and each gene was

| DAPI/RFP/GFP-SKL | RFPminus -/- clone | GFP-SKL |

iso FRT19A — A, A', A''
fs(1)h — B, B', B''
Rbcn-3B — C, C', C''
Coq8 — D, D', D''
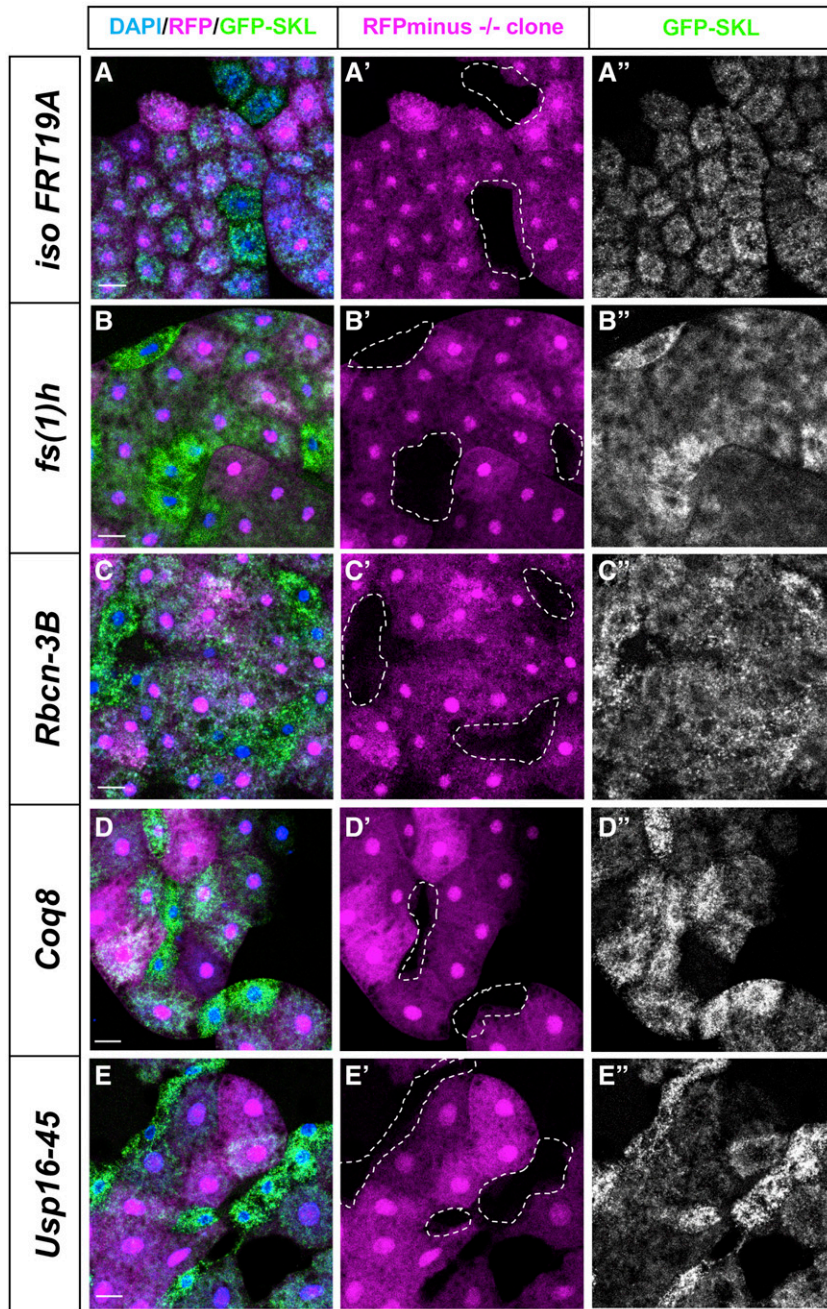**Usp16-45** — E, E', E''

Figure 2 Peroxisomal morphological phenotypes observed *in-vivo*. Third instar fat body clones are shown in merge **DAPI (in blue) / RFP (magenta) / GFP-SKL (green)** in first from left column. Homozygous mutant cells that lack RFP are shown with dotted lines in the middle column and the right most column shows the same cells showing the **GFP-SKL** signal. The images in A-A"- *iso FRT19A* clones are clones of the non-mutagenized chromosome and serve as negative controls. Images of some of the strong hits are shown in panels B-E" as *fs(1)h* clones (from B-B"), *Rbcn-3B* clones (from C-C"), *Coq8 clones* (from D-D") and *Usp16-45* clones (from E-E"). Scale bars represent 50μm. Images are taken on Zeiss710-LSM microscope with 10X objective of NA 0.3 with 8 bit depth and 2200x2200 resolution.

examined using the gnomAD (https://gnomad.broadinstitute.org/) browser (Lek *et al.* 2016). Each gene was also examined using the DOMINO tool for predicted likelihood of a gene having dominant impact on disease (Quinodoz *et al.* 2017). In addition, *de novo* events were examined in denovo-db (http://denovo-db.gs.washington.edu/denovo-db/) website (Turner *et al.* 2017).

Data availability

All the supplemental data files are available on the GSA figshare portal including the Supplemental Tables listed in the manuscript. Supplemental Table 1 lists all the Drosophila reagents that are available from the X-screen through public stock centers including the Bloomington stock center and Kyoto Stock center. Supplemental material available at figshare: https://doi.org/10.25387/g3.8872547.

RESULTS AND DISCUSSION

Identification of genes involved in peroxisomal dynamics

To identify new regulators of peroxisomal morphology and dynamics, we performed a screen with the collection of recessive lethal X-chromosome mutants that have been extensively studied by our group (Yamamoto *et al.* 2014). We had previously generated a large collection of recessive lethal mutant lines on an isogenized chromosome and these lines have been extensively screened for developmental and neurological phenotypes (Yamamoto *et al.* 2014; Deal and Yamamoto 2019). Moreover, we previously utilized this collection of mutations to uncover new human disease genes and showed that human orthologs of these fly essential genes are enriched for
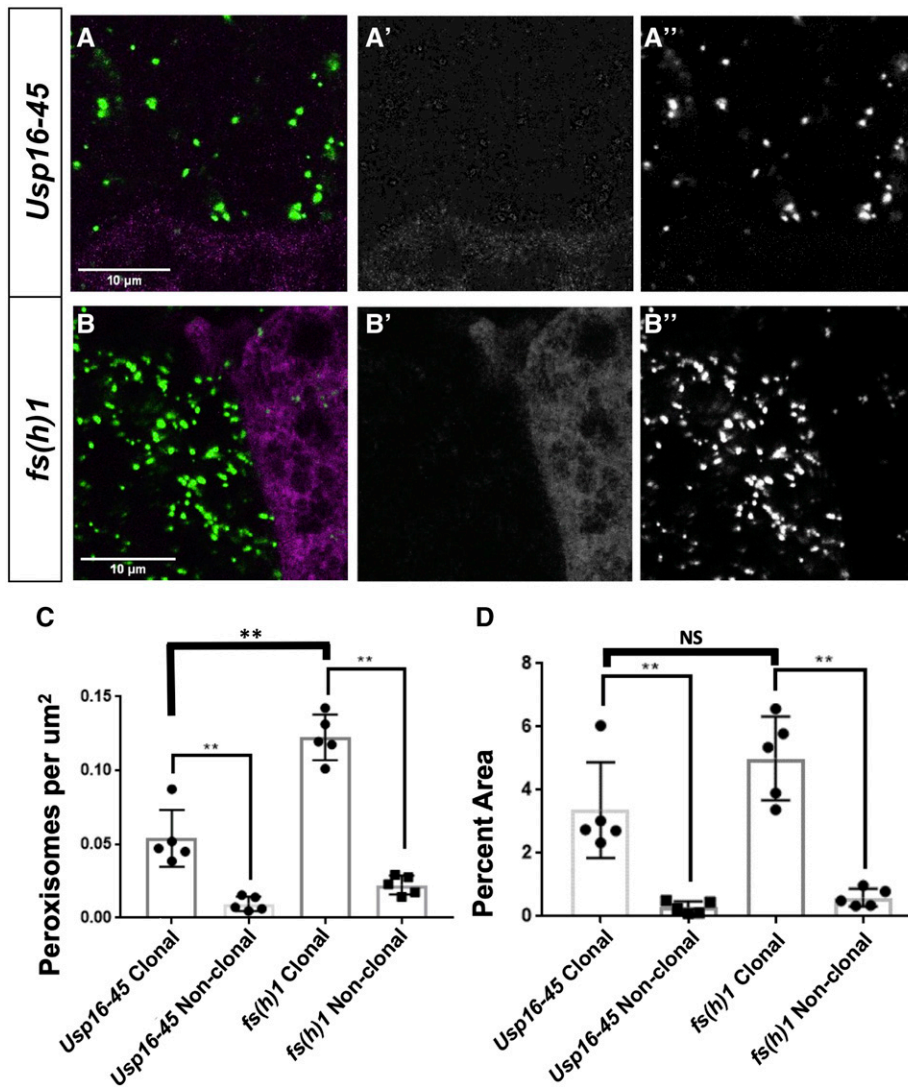
G3 Genes | Genomes | Genetics

Figure 3 Peroxisomal morphological quantification. A. Higher magnification images of fat bodies of *Usp16-45* Fig. 3A-A'' were taken on Zeiss710-LSM microscope with 63X objective of NA 1.4 with 16 bit depth and 512x512 resolution. The scale bars noted in the image are 10μm. Note A is the image shown in Figure 1C' to illustrate the Category A phenotype, and A' is the RFP channel while A'' is the GFP channel. B. Higher magnification images of fat bodies of *fs(h)1* Fig 3B-B'' were taken on Zeiss710-LSM microscope with 63X objective of NA 1.4 with 16 bit depth and 512x512 resolution. The scale bars noted in the image are 10μm. B is the image shown in Figure 1C''to illustrate the Category B phenotype, and B' is the RFP channel while B''is the GFP channel. C. Quantification of number of peroxisomes present per μm^2 of Clonal and non-clonal area. This data represents the number of peroxisomes present per μm^2 of clonal/non-clonal area. Peroxisomes from five different stacks were counted with ImageJ and then divided with the total clonal/non-clonal area in that individual stack to get these data points. D. Quantification of percent area of total peroxisomes. The areas covered by peroxisomes in clonal as well as non-clonal regions from five different stacks were counted with ImageJ. The total of those peroxisomal areas in the same stack is divided by clonal/non-clonal area as shown in A' and B' of the same stack. The percent value of this is counted as one data point.

genes listed in the Online Mendelian Inheritance in Man (OMIM) disease database (Yamamoto *et al.* 2014; Yoon *et al.* 2017; Tan *et al.* 2018). We took 215 lines from this collection that correspond to 100 genes (98 mapped genes and 2 unmapped complementation group) and screened for peroxisomal phenotypes using GFP tagged peroxisomes (Chao *et al.* 2016; Wangler *et al.* 2017a) in conjunction with clonal analysis, allowing generation of homozygous mutant cells within *Drosophila* larval fat body in an otherwise heterozygous animal to bypass early lethality.

It is important to note that X-chromosome contains ∼15% of protein coding genes in *Drosophila*, and there is no correlation between X-linked genes in flies and humans. In this screen, we created homozygous mutant clones in the fat body of developing *Drosophila* larvae in an otherwise heterozygous animal and assayed for changes in the distribution pattern of a peroxisomal reporter, GFP-SKL (Figure 1A). GFP-SKL is a GFP with a C-terminal peroxisomal localization signal, which we have previously shown to be an accurate marker for peroxisomal dynamics (Chao *et al.* 2016; Wangler *et al.* 2017a). We hypothesized that our screen could uncover three major categories of peroxisomal impact: Category A- an increase in the number of peroxisomes, Category B- an increase in the size of peroxisomes, and Category C- decrease or loss of peroxisomes or the marker (Figure 1B-C).

We have previously observed examples of Category B in mutants with peroxisomal fission defects (Chao *et al.* 2016) and Category C in biogenesis defects (Wangler *et al.* 2017a) in *Drosophila*. We screened 215 lethal mutant lines from the already available collection that was mapped to a complementation group or to a gene (Figure 1D, Supplemental Table 1). We considered a hit to be positive when multiple independent co-authors could differentiate a clear difference in the mutant clone compared to the surrounding (heterozygous or homozygous wild-type) cells (Supplementary Figure 1). In total we identified 37 alleles corresponding to 18 genes (Table 1). For these hits, when possible we assayed two or more alleles per gene to confirm a change in GFP-SKL if possible. Some of these mutations led to an inconsistent increase of the peroxisome reporter among the fat body clones even within the same tissue, suggesting that perdurance of the wild type protein or other unmeasured factors may mask a change in peroxisomal dynamics in some clones. Interestingly, an overwhelming majority of the lines fell into Category A (Table 1). Although there are a few known peroxisomal related genes that are located on the *Drosophila* X-chromosome including *Pex5* (homolog of human *PEX5*) and *Mfe2* (homolog of human *HSD17B4*) (Faust *et al.* 2012), these were not part of the collection we screened here (Yamamoto *et al.* 2014). We

Table 2 Human Gene Candidate Analysis

Human Gene	Known Disease in OMIM	pLI	pLI o/e	Missense Z score	Missense Z score o/e	Domino
GSK3A	None	1	0	3.2	0.43	1[A]
BRD4	None	1	0	3.74	0.63	0.997[A]
UPF1	None	1	0.07	5.7	0.41	0.996[A]
L1CAM	# 304100, # 303350, # 307000	1	0.04	2.84	0.66	n/a
NFASC	# 618356	1	0.12	2.59	0.74	0.871[A]
BRD3	None	0.98	0.14	3.76	0.64	0.893[A]
GSK3B	None	0.96	0.14	2.91	0.48	1[A]
SMAD3	# 613795	0.84	0.17	3.47	0.39	0.999[A]
UPF2	None	1	0.03	3.3	0.65	0.723[B]
WDR7	None	1	0.10	2.63	0.75	0.636[B]
BRD2	None	1	0.08	0.5	0.93	0.474[C]
MTPAP	# 613672	1	0.04	1.05	0.84	0.242[D]
PTAR1	None	0.85	0.15	1.51	0.71	0.275[D]
NRCAM	None	0.18	0.24	2.05	0.79	0.191[E]
FH	# 606812 - # 150800	0.09	0.28	1.39	0.77	0.371[D]
HINFP	None	0.03	0.29	1.73	0.73	0.719[B]
PI4KA	# 616531	0	0.36	3.53	0.72	0.589[C]
BRDT	# 617644	0	0.46	0.47	0.94	0.209[D]
CHL1	None	0	0.47	−1.92	1.21	0.152[E]
USP45	None	0	0.74	0.75	0.90	0.168[E]
COQ7	# 616733	0	0.88	−0.42	1.10	0.091[E]
RFT1	# 612015	0	0.77	0.97	0.84	0.063[E]
COQ8B (a.k.a. ADCK4)	# 615573	0	0.78	0.89	0.87	0.056[E]

The human homologs of the X-Pex genes were examined for known Mendelian disease association (OMIM # entries) with genes that are not known to cause disease shown in red (Amberger *et al.* 2015). These are further sorted using data from the public human database gnomAD and the DOMINO scoring system for dominant disease. "pLI" score shows the probability (from 0-1) of a gene having intolerance to loss-of-function variation in the population of individuals represented in gnomAD data. "Missense z-score" show a z-score value for rates of missense variation in a gene. "pLI-o/e" is the observed / expected for loss-of-function variants in a gene, while "Missense o/e" is a similar ratio for missense variants. For DOMINO scores the code shows [A] = "Very likely dominant (0.8-1)", [B] = "Likely dominant (0.6-0.7)", [C] = "Either dominant or recessive (0.4-0.5)", [D] = "Likely recessive (0.2-0.3), [E] = "Very Likely recessive (0-0.1)"

note that none of the genes we isolated in this screen encode proteins that have been localized to peroxisome but are instead suggested to be involved in a wide variety of cellular processes (Table 1).

In the majority of the lines screened, there was no difference between the appearance of the peroxisomal (GFP-SKL) marker in the mutant clone *vs.* the surrounding sister cells, similar to the pattern seen in FRT19A controls (Figure 2A-A"). In one line, we observed enlargement and increased number of peroxisomes qualitatively (Figure 2B-B"). This hit, *fs(1)h*, produced a Category A and B phenotype, and encodes a protein that has been reported to be involved in regulating proper expression of homeotic genes involved in pattern formation, such as *Ultrabithorax* (Florence and Faller 2008). In other hits, including *Rbcn-3B*, *Coq8*, and *Usp16-45*, we observed a Category A phenotype with more GFP-SKL punctae in the clones (Figure 2 C-E", Supplementary Figure 1). Some of the hits from this screen have been studied in other biological contexts. *Rbcn-3B*, which produced a Category A phenotype, encodes a protein involved in Notch signaling during oogenesis through its role in endocytic trafficking and lysosomal function (Yan *et al.* 2009). *Coq8*, which produced a Category A phenotype, encodes a mitochondrial inner membrane protein that is predicted to be involved in electron transport (Zhu *et al.* 2017) In contrast, many of the genes we identified, including *Usp16-45* which is predicted to encode an ubiquitin specific protease, have not been extensively characterized *in vivo* and it is likely that these proteins also have additional functions as most genes are pleiotropic (Wangler *et al.* 2017b). We also quantified the number and the area of peroxisomes in the *Usp16-45* and *fs(1)h* clones (Figure 3A-B). We see a significant increase in the number of peroxisomes per µm^2 in the clones of both *Usp16-45* and *fs(1)h* compared to surrounding cells consistent with the Category A phenotype (Figure 3C). This increase in the number of peroxisomes and changes

in size led to the increase in the GFP-SKL signal (Figure 3D). These results indicate that screening for the GFP-SKL peroxisomal marker is an effective method to identify new genes that can impact peroxisome dynamics or morphology.

Human homologs of genes identified in the fly peroxisome screen are candidates for human disease

In our previous studies from the *X*-screen collection we showed that the process of identification and screening for lethal mutations in *Drosophila* enriches for human Mendelian disease genes (Yamamoto *et al.* 2014). Indeed, our subsequent studies on this collection continued to yield novel human disease genes (Deal and Yamamoto 2019). We therefore wanted to assess this potential for genes from our peroxisome screen within public human databases.

First, we determined the human homolog of each fly gene with the highest predicted score using the Drosophila RNAi Screening center Integrative Ortholog Prediction Tool (DIOPT) tool (Hu *et al.* 2011; Gray *et al.* 2015) (Supplemental Table 2). We also utilized the MARR-VEL tool, MARRVEL allows for simultaneous display of public human genomic data and model organism phenotypes and conservation (Wang *et al.* 2017). For the eighteen fly genes that we considered hits, we found that sixteen of the eighteen had human homologs (88.9%). Thirteen out of sixteen fly genes had a single human homolog while remaining three of the genes had multiple human homologs (Table 1). We also wanted to know if any of the genes were already linked to a human single gene disorder. We examined the twenty human homologs from the X-Pex screen in a Mendelian disease using the Online Mendelian Inheritance in Man (OMIM) database (Amberger *et al.* 2015). Of the twenty-three human genes, nine were listed in relation to at least one Mendelian phenotype where the gene is causative for a
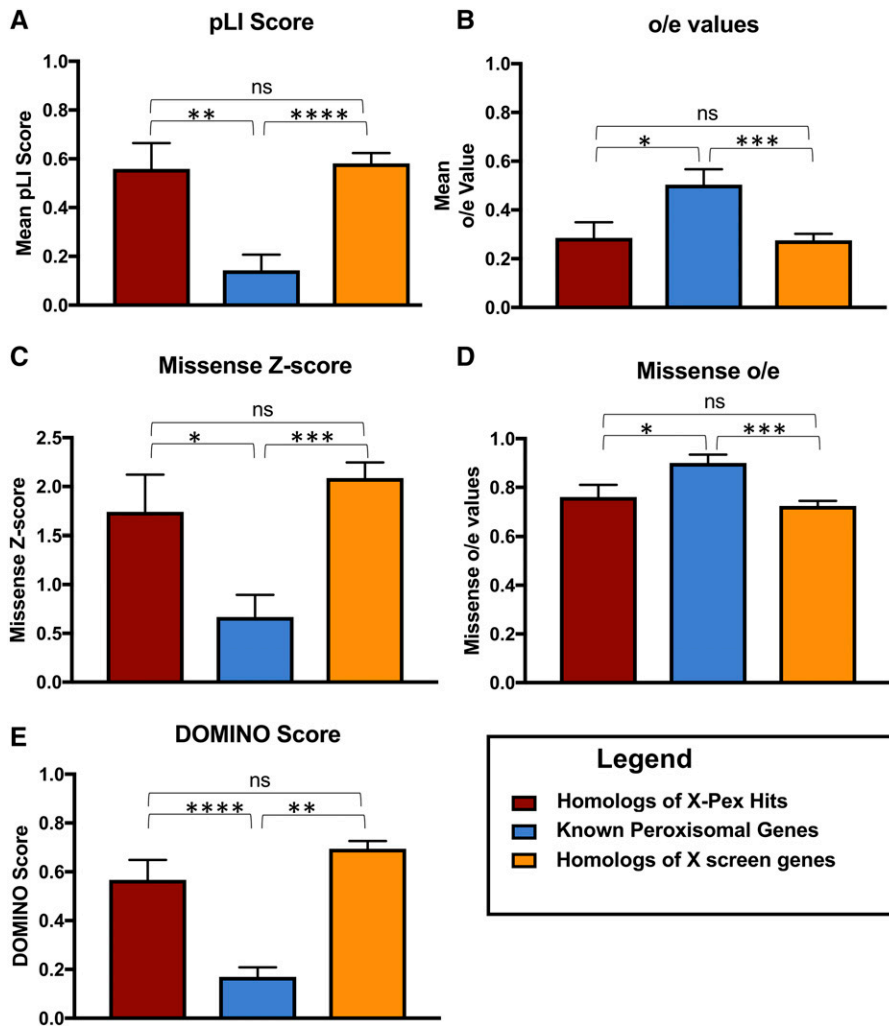
Figure 4 Comparison of known human peroxisomal disease genes to the new X-Pex candidates. A. The Probability of Loss of Function intolerance score (pLi) calculated from human data from the gnomAD database (Lek *et al.* 2016). The X-Pex genes displayed a mean pLi score of 0.55 ± 0.11, n = 20, while the known peroxisomal disease genes had a mean pLi of 0.14 ± 0.06, n = 25, which was statistically significant (P = 0.0016) **. This was also compared to all the homologs of the X screen genes. B. The observed over expected (o/e) loss of function scores calculated from public human data from the gnomAD database. The X-Pex genes had a mean o/e score of 0.29 ± 0.06, n = 20, while the known peroxisomal disease genes had an o/e score of 0.50 ± 0.06, n = 25, which was statistically significant (P = 0.0218)*. This was also compared to all the homologs of the X screen genes. C. The missense constrain z-score calculated from public human data from the gnomAD database. The X-Pex genes had mean missense constrain z-scores of 2.16 ± 0.34, n = 20, while the known peroxisomal genes had z-scores of 0.67± 0.23, n = 25, which was statistically significant (P = 0.0005)***. This was also compared to all the homologs of the X screen genes. D. Missense constraint o/e scores calculated from public human data from the gnomAD database. The X-Pex genes had a mean o/e for missense variants of 0.73 ± 0.04, n = 20, compared to the known peroxisomal disease genes o/e score of 0.90 ± 0.03, n = 25, also statistically significant (P = 0.0025)**. This was also compared to all the homologs of the X screen genes. E. DOMINO scores calculated for the gene sets. The X-Pex gene set had a DOMINO score of 0.53 ± 0.08, n = 20, while the known peroxisomal disease genes had a mean DOMINO score of 0.17 ± 0.04, n = 24, and the difference was statistically significant (P < 0.0001)***. This was also compared to all the homologs of the X screen genes.

described disease (Supplemental Table 3). Fourteen human genes have no known single gene disorder (Table 2 and Supplementary Table 3). In our previous work in this collection, we observed that by screening for lethality, we enrich for essential genes in fly that are homologs of disease genes in humans (Wangler *et al.* 2015). We therefore hypothesized that these fourteen human genes from our screen, not currently associated with human disease could be considered good candidates for undiagnosed cases.

One way to assess whether these genes could be good candidates for undiagnosed disease is to examine whether damaging or deleterious variants in the genes occur in the population at large. If variants in this gene have a deleterious effect prior to reproduction these variants will be selected against in the general population. In order to explore evidence for this we examined public human genomic databases, we were looking for evidence of selective constraint, or lack of damaging variation indicating that gene is under selection in humans, in all twenty-three human genes identified by our screen. To do this we examined the gnomAD database which is a large genome and exome aggregation largely selected for healthy or adult-onset disease cases (Lek *et al.* 2016). For each gene we examined the constraint metrics or evidence that damaging variants in the gene are absent from these

"control" individuals (Table 2, Supplemental Table 3). This type of information would point to a gene being under strong selection twelve of the twenty-three genes had evidence of being under selection. We noted that twelve genes had "observed over expected" (o/e) numbers for loss-of-function variants at 0.2 (20%) or less indicating fewer loss-of function alleles than anticipated by chance. This data suggests that there may be some selection against loss of function alleles for approximately half of the genes possibly due to a haploinsufficient mechanisms. In summary, the genes from our screen could be considered good candidates for undiagnosed cases as they exhibit selective constraint in healthy individuals.

The constraint metrics of the gnomAD dataset are in fact most valuable for showing selective constraint for heterozygous alleles, such as *de novo* mutations or dominant inherited disorders, particularly with early onset or an impact on reproduction (Lek *et al.* 2016). This is because it is much easier to observe a single damaging allele than homozygous or compound heterozygous. We therefore wondered if the X-Pex gene set could be considered good candidates for dominant disease and we examined this through two strategies: 1) comparing the gene metrics to known recessive peroxisomal genes and 2) use of the DOMINO tool for predicting dominant disorders.

In the first approach we compared these same characteristics of these genes to two other sets of genes, first we compared the X-Pex gene set to all the other homologs of the genes that we screened. We also compared to a group of twenty-five well known human peroxisomal disease genes that encode proteins in the peroxisome biogenesis machinery and enzymes involved in very-long-chain fatty acid oxidation, plasmalogen synthesis and reactive oxygen species (Supplemental Table 4). Comparing these three gene sets we found that the human homologs of the essential fly genes had significantly higher probability of loss of function intolerance (Figure 4A-B), and higher missense constraint (Figure 4C-D) in the human databases compared to the known peroxisomal disease genes. This was consistent across both genes that were positive in the peroxisome secondary screen as well as negative. These intolerance scores apply more to dominant disorders than autosomal recessive. Consistent with that, all known *PEX* gene-related Peroxisome biogenesis disorders are autosomal recessive (Braverman *et al.* 2016). (Supplemental Table 4). We therefore hypothesized that the selection of lethals in our original fly screen pointed us to a set of human genes that are more likely to underlie dominant disease.

In order to test this, we used the DOMINO tool to assess the probability of dominant disease for each gene (Quinodoz *et al.* 2017). The DOMINO score, indicating the likelihood of dominant disorders also differed between the X-Pex gene set and the known peroxisomal disease gene set (Figure 4E). The X-Pex gene set had a higher DOMINO score, while the known peroxisomal disease genes had lower DOMINO scores, thus more likelihood of relating to recessive disease and the difference was statistically significant ($P < 0.0001$). As noted, for the known peroxisomal genes this is indeed the case, as twenty-two of the twenty-five genes are disease genes for autosomal recessive disorders (Supplemental Table 4).

With this data we predict that some X-Pex genes could underlie dominant phenotypes, we sought evidence for dominant phenotypes related to alleles in the set of X-Pex genes using public databases of *de novo* events from individuals with disease and controls (Turner *et al.* 2017). This dataset primarily focuses on neurodevelopmental phenotypes and the *de novo* events from diverse cohorts. Strikingly, we observed suggestive results for six genes from the X-Pex gene set with high DOMINO scores (*GSK3A*, *BRD4*, *UPF1*, *BRD3*, *GSK3B* and *SMAD3*) and at least one individual with developmental delay, Autism, or congenital heart disease. These cases are not definitively linked to these loci and are noted to have a missense *de novo* event. Interestingly no missense *de novo* events are observed in control individuals for these genes (Supplemental Table 5).

Whether these genes are ultimately good candidates remains to be explored in undiagnosed cases as these database searches do not definitely link the specific gene with disease. It is also not known whether the peroxisomal phenotype that was observed in our screen would be conserved in humans. Peroxisomes are not routinely examined in clinical samples, so this data provides a key starting point for these genes of interest. Even a secondary impact on peroxisomes without direct interaction could aid in exploring these candidate genes further.

Taken together we propose the X-Pex provides a good list of candidate genes, in particular, *de novo* events in these genes from patients with neurodevelopmental phenotypes should be explored. Considering that peroxisomal disease classically relates to autosomal recessive conditions, the X-Pex gene list may provide an entry point to study the role of *de novo* events in genes that impact peroxisomes that have been missed in previous screens. This study therefore provides additional support for the use of forward genetic screens in model organisms in the study and identification of human disease genes (Yamamoto *et al.* 2014; Wangler *et al.* 2015).

LITERATURE CITED

Amberger, J. S., C. A. Bocchini, F. Schiettecatte, A. F. Scott, and A. Hamosh, 2015 OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. Nucleic Acids Res. 43: D789–D798. https://doi.org/10.1093/nar/gku1205

Baron, M. N., C. M. Klinger, R. A. Rachubinski, and A. J. Simmonds, 2016 A Systematic Cell-Based Analysis of Localization of Predicted Drosophila Peroxisomal Proteins. Traffic 17: 536–553. https://doi.org/10.1111/tra.12384

Braschi, B., P. Denny, K. Gray, T. Jones, R. Seal *et al.*, 2019 Genenames.org: the HGNC and VGNC resources in 2019. Nucleic Acids Res. 47: D786–D792. https://doi.org/10.1093/nar/gky930

Braverman, N. E., G. V. Raymond, W. B. Rizzo, A. B. Moser, M. E. Wilkinson *et al.*, 2016 Peroxisome biogenesis disorders in the Zellweger spectrum: An overview of current diagnosis, clinical manifestations, and treatment guidelines. Mol. Genet. Metab. 117: 313–321. https://doi.org/10.1016/j.ymgme.2015.12.009

Chao, Y.-H., L. A. Robak, F. Xia, M. K. Koenig, A. Adesina *et al.*, 2016 Missense variants in the middle domain of DNM1L in cases of infantile encephalopathy alter peroxisomes and mitochondria when assayed in Drosophila. Hum. Mol. Genet. 25: 1846–1856. https://doi.org/10.1093/hmg/ddw059

Cook, M. S., C. Cazin, M. Amoyel, S. Yamamoto, E. Bach *et al.*, 2017 Neutral Competition for Drosophila Follicle and Cyst Stem Cell Niches Requires Vesicle Trafficking Genes. Genetics 206: 1417–1428. https://doi.org/10.1534/genetics.117.201202

Deal, S. L., and S. Yamamoto, 2019 Unraveling Novel Mechanisms of Neurodegeneration Through a Large-Scale Forward Genetic Screen in Drosophila. Front. Genet. 9: 700. https://doi.org/10.3389/fgene.2018.00700

Fang, X., J. Zhou, W. Liu, X. Duan, U. Gala *et al.*, 2016 Dynamin Regulates Autophagy by Modulating Lysosomal Function. J. Genet. Genomics 43: 77–86. https://doi.org/10.1016/j.jgg.2015.10.005

Faust, J. E., A. Manisundaram, P. T. Ivanova, S. B. Milne, J. B. Summerville *et al.*, 2014 Peroxisomes are required for lipid metabolism and muscle function in Drosophila melanogaster. PLoS One 9: e100213. https://doi.org/10.1371/journal.pone.0100213

Faust, J. E., A. Verma, C. Peng, and J. A. McNew, 2012 An inventory of peroxisomal proteins and pathways in Drosophila melanogaster. Traffic 13: 1378–1392. https://doi.org/10.1111/j.1600-0854.2012.01393.x

Florence, B. L., and D. V. Faller, 2008 Drosophila female sterile (1) homeotic is a multifunctional transcriptional regulator that is modulated by Ras signaling. Dev. Dyn. 237: 554–564. https://doi.org/10.1002/dvdy.21432

Fujiki, Y., K. Okumoto, S. Mukai, M. Honsho, and S. Tamura, 2014 Peroxisome biogenesis in mammalian cells. Front. Physiol. 5: 307. https://doi.org/10.3389/fphys.2014.00307

Gray, K. A., B. Yates, R. L. Seal, M. W. Wright, and E. A. Bruford, 2015 Genenames.org: the HGNC resources in 2015. Nucleic Acids Res. 43: D1079–D1085. https://doi.org/10.1093/nar/gku1071

Haelterman, N. A., L. Jiang, Y. Li, V. Bayat, H. Sandoval *et al.*, 2014 Large-scale identification of chemically induced mutations in Drosophila melanogaster. Genome Res. 24: 1707–1718. https://doi.org/10.1101/gr.174615.114

Hu, Y., I. Flockhart, A. Vinayagam, C. Bergwitz, B. Berger *et al.*, 2011 An integrative approach to ortholog prediction for disease-focused and other functional studies. BMC Bioinformatics 12: 357. https://doi.org/10.1186/1471-2105-12-357

Kao, Y. T., K. L. Gonzalez, and B. Bartel, 2018 Peroxisome Function, Biogenesis, and Dynamics in Plants. Plant Physiol. 176: 162–177. https://doi.org/10.1104/pp.17.01050

Klouwer, F. C., K. Berendse, S. Ferdinandusse, R. J. Wanders, M. Engelen et al., 2015 Zellweger spectrum disorders: clinical overview and management approach. Orphanet J. Rare Dis. 10: 151. https://doi.org/10.1186/s13023-015-0368-9

Lek, M., K. J. Karczewski, E. V. Minikel, K. E. Samocha, E. Banks et al., 2016 Analysis of protein-coding genetic variation in 60,706 humans. Nature 536: 285–291. https://doi.org/10.1038/nature19057

Mast, F. D., R. A. Rachubinski, and J. D. Aitchison, 2015 Signaling dynamics and peroxisomes. Curr. Opin. Cell Biol. 35: 131–136. https://doi.org/10.1016/j.ceb.2015.05.002

Nakayama, M., H. Sato, T. Okuda, N. Fujisawa, N. Kono et al., 2011 Drosophila carrying pex3 or pex16 mutations are models of Zellweger syndrome that reflect its symptoms associated with the absence of peroxisomes. PLoS One 6: e22984. https://doi.org/10.1371/journal.pone.0022984

Quinodoz, M., B. Royer-Bertrand, K. Cisarova, S. A. Di Gioia, A. Superti-Furga et al., 2017 DOMINO: Using Machine Learning to Predict Genes Associated with Dominant Disorders. Am. J. Hum. Genet. 101: 623–629. https://doi.org/10.1016/j.ajhg.2017.09.001

Subramani, S., 1998 Components involved in peroxisome import, biogenesis, proliferation, turnover, and movement. Physiol. Rev. 78: 171–188. https://doi.org/10.1152/physrev.1998.78.1.171

Tan, K. L., N. A. Haelterman, C. S. Kwartler, E. S. Regalado, P. T. Lee et al., 2018 Ari-1 Regulates Myonuclear Organization Together with Parkin and Is Associated with Aortic Aneurysms. Dev Cell 45: 226–244. https://doi.org/10.1016/j.devcel.2018.03.020

Turner, T. N., Q. Yi, N. Krumm, J. Huddleston, K. Hoekzema et al., 2017 denovo-db: a compendium of human de novo variants. Nucleic Acids Res. 45: D804–D811. https://doi.org/10.1093/nar/gkw865

UniProt Consortium, 2019 UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 47: D506–D515. https://doi.org/10.1093/nar/gky1049

Wanders, R. J., 2014 Metabolic functions of peroxisomes in health and disease. Biochimie 98: 36–44. https://doi.org/10.1016/j.biochi.2013.08.022

Wanders, R. J., and H. R. Waterham, 2005 Peroxisomal disorders I: biochemistry and genetics of peroxisome biogenesis disorders. Clin. Genet. 67: 107–133. https://doi.org/10.1111/j.1399-0004.2004.00329.x

Wang, J., R. Al-Ouran, Y. Hu, S. Y. Kim, Y. W. Wan et al., 2017 MARRVEL: Integration of Human and Model Organism Genetic Resources to Facilitate Functional Annotation of the Human Genome. Am. J. Hum. Genet. 100: 843–853. https://doi.org/10.1016/j.ajhg.2017.04.010

Wangler, M. F., Y. H. Chao, V. Bayat, N. Giagtzoglou, A. B. Shinde et al., 2017a Peroxisomal biogenesis is genetically and biochemically linked to carbohydrate metabolism in Drosophila and mouse. PLoS Genet. 13: e1006825. https://doi.org/10.1371/journal.pgen.1006825

Wangler, M. F., S. Yamamoto, and H. J. Bellen, 2015 Fruit flies in biomedical research. Genetics 199: 639–653. https://doi.org/10.1534/genetics.114.171785

Wangler, M. F., S. Yamamoto, H. T. Chao, J. E. Posey, M. Westerfield et al., 2017b Model Organisms Facilitate Rare Disease Diagnosis and Therapeutic Research. Genetics 207: 9–27. https://doi.org/10.1534/genetics.117.203067

Yamamoto, S., V. Bayat, H. J. Bellen, and C. Tan, 2013 Protein phosphatase 1ss limits ring canal constriction during Drosophila germline cyst formation. PLoS One 8: e70502. https://doi.org/10.1371/journal.pone.0070502

Yamamoto, S., M. Jaiswal, W. L. Charng, T. Gambin, E. Karaca et al., 2014 A drosophila genetic resource of mutants to study mechanisms underlying human genetic diseases. Cell 159: 200–214. https://doi.org/10.1016/j.cell.2014.09.002

Yan, Y., N. Denef, and T. Schupbach, 2009 The vacuolar proton pump, V-ATPase, is required for notch signaling and endosomal trafficking in Drosophila. Dev. Cell 17: 387–402. https://doi.org/10.1016/j.devcel.2009.07.001

Yofe, I., K. Soliman, S. G. Chuartzman, B. Morgan, U. Weill et al., 2017 Pex35 is a regulator of peroxisome abundance. J. Cell Sci. 130: 791–804. https://doi.org/10.1242/jcs.187914

Yoon, W. H., H. Sandoval, S. Nagarkar-Jaiswal, M. Jaiswal, S. Yamamoto et al., 2017 Loss of Nardilysin, a Mitochondrial Co-chaperone for alpha-Ketoglutarate Dehydrogenase, Promotes mTORC1 Activation and Neurodegeneration. Neuron 93: 115–131. https://doi.org/10.1016/j.neuron.2016.11.038

Zhu, J. Y., Y. Fu, A. Richman, Z. Zhao, P. E. Ray et al., 2017 A Personalized Model of COQ2 Nephropathy Rescued by the Wild-Type COQ2 Allele or Dietary Coenzyme Q10 Supplementation. J. Am. Soc. Nephrol. 28: 2607–2617. https://doi.org/10.1681/ASN.2016060626

Communicating editor: E. Lecuyer