A *Drosophila* Genetic Resource of Mutants to Study Mechanisms Underlying Human Genetic Diseases

Shinya Yamamoto,[1,2,3,24] Manish Jaiswal,[2,4,24] Wu-Lin Charng,[1,2] Tomasz Gambin,[2,5] Ender Karaca,[2] Ghayda Mirzaa,[6,7] Wojciech Wiszniewski,[2,8] Hector Sandoval,[2] Nele A. Haelterman,[1] Bo Xiong,[1] Ke Zhang,[9] Vafa Bayat,[1] Gabriela David,[1] Tongchao Li,[1] Kuchuan Chen,[1] Upasana Gala,[1] Tamar Harel,[2,8] Davut Pehlivan,[2] Samantha Penney,[2,8] Lisenka E.L.M. Vissers,[10] Joep de Ligt,[10] Shalini N. Jhangiani,[11] Yajing Xie,[12] Stephen H. Tsang,[12,13] Yesim Parman,[14] Merve Sivaci,[15] Esra Battaloglu,[15] Donna Muzny,[2,11] Ying-Wooi Wan,[3,16] Zhandong Liu,[3,17] Alexander T. Lin-Moore,[2] Robin D. Clark,[18] Cynthia J. Curry,[19,20] Nichole Link,[2] Karen L. Schulze,[2,4] Eric Boerwinkle,[11,21] William B. Dobyns,[6,7,22] Rando Allikmets,[12,13] Richard A. Gibbs,[2,11] Rui Chen,[1,2,11] James R. Lupski,[2,8,11] Michael F. Wangler,[2,8,*] and Hugo J. Bellen[1,2,3,4,9,23,*]

[1]Program in Developmental Biology, Baylor College of Medicine (BCM), Houston, TX 77030, USA
[2]Department of Molecular and Human Genetics, BCM, Houston, TX 77030, USA
[3]Jan and Dan Duncan Neurological Research Institute, Houston, TX 77030, USA
[4]Howard Hughes Medical Institute, Houston, TX 77030, USA
[5]Institute of Computer Science, Warsaw University of Technology, 00-661 Warsaw, Poland
[6]Department of Pediatrics, University of Washington, Seattle, WA 98195, USA
[7]Center for Integrative Brain Research, Seattle Children's Research Institute, Seattle, WA 98101, USA
[8]Texas Children's Hospital, Houston, TX 77030, USA
[9]Program in Structural and Computational Biology and Molecular Biophysics, BCM, Houston, TX 77030, USA
[10]Department of Human Genetics, Radboudumc, PO Box 9101, 6500 HB, Nijmegen, The Netherlands
[11]Human Genome Sequencing Center, BCM, Houston, TX 77030, USA
[12]Department of Ophthalmology, Columbia University College of Physicians and Surgeons, New York, NY 10032, USA
[13]Department of Pathology and Cell Biology, Columbia University College of Physicians and Surgeons, New York, NY 10032, USA
[14]Neurology Department and Neuropathology Laboratory, Istanbul University Medical School, Istanbul 34390, Turkey
[15]Department of Molecular Biology and Genetics, Bogazici University, Istanbul 34342, Turkey
[16]Department of Obstetrics and Gynecology, BCM, Houston, TX 77030, USA
[17]Department of Pediatrics, BCM, Houston, TX 77030, USA
[18]Division of Medical Genetics, Department of Pediatrics, Loma Linda University Medical Center, Loma Linda, CA 92354, USA
[19]Department of Pediatrics, University of California San Francisco, San Francisco, CA 94143, USA
[20]Genetic Medicine Central California, Fresno, CA 93701, USA
[21]Human Genetics Center, University of Texas, Health Science Center, Houston, TX 77030, USA
[22]Department of Neurology, University of Washington, Seattle WA 98195, USA
[23]Department of Neuroscience, BCM, Houston, TX 77030, USA
[24]Co-first author
*Correspondence: michael.wangler@bcm.edu (M.F.W.), hbellen@bcm.edu (H.J.B.)
 http://dx.doi.org/10.1016/j.cell.2014.09.002

SUMMARY

Invertebrate model systems are powerful tools for studying human disease owing to their genetic tractability and ease of screening. We conducted a mosaic genetic screen of lethal mutations on the *Drosophila* X chromosome to identify genes required for the development, function, and maintenance of the nervous system. We identified 165 genes, most of whose function has not been studied in vivo. In parallel, we investigated rare variant alleles in 1,929 human exomes from families with unsolved Mendelian disease. Genes that are essential in flies and have multiple human homologs were found to be likely to be associated with human diseases. Merging the human data sets with the fly genes allowed us to identify disease-associated mutations in six families and to provide insights into microcephaly associated with brain dysgenesis. This bidirectional synergism between fly genetics and human genomics facilitates the functional annotation of evolutionarily conserved genes involved in human health.

INTRODUCTION

Unbiased genetic chemical mutagenesis screens in flies have led to the discovery of the vast majority of genes in developmental signaling pathways (Nüsslein-Volhard and Wieschaus, 1980). Most genes important to these pathways have now been shown to function as oncogenes or tumor suppressors

(Pastor-Pareja and Xu, 2013). Similarly, in some areas of neurobiology, genetic screens in flies have led to the discovery of genes important to nervous system function including TRP channels, potassium channels, and pathways that affect diurnal rhythmicity. Subsequent studies have identified many diseases that are associated with mutations or deletions of human homologs (Bellen et al., 2010). However, our molecular understanding of neurological disorders such as neurodegenerative disease has mostly relied on reverse genetics (Lu and Vogel, 2009). Although some genes required for neuronal maintenance have been identified from genetic screens for viable mutations that exhibit shortened life span, electroretinogram defects, abnormal phototaxis, and retinal histology defects, or temperature-sensitive paralysis, no large-scale systematic screens to directly probe neurodegeneration have been carried out, (reviewed in Jaiswal et al., 2012). In addition, because of lethal phenotypes, the role of numerous essential genes in neuronal maintenance is not known. We therefore implemented a genetic mosaic screen to identify essential genes required for neuronal maintenance on the X chromosome.

One major limitation in chemical mutagenesis screens has been the inability to systematically identify an abundance of causative mutations. However, with the advent of numerous mapping tools and whole-genome sequencing (WGS), it should be possible to identify hundreds of causative mutations from a single mutagenesis experiment in which a multitude of phenotypes are scored in parallel for each mutation.

In humans, the study of Mendelian traits has led to the discovery of thousands of disease genes. Currently, identification of rare disease-causing mutations is rapidly evolving because whole-exome sequencing (WES) technologies are driving the process (Bainbridge et al., 2011a; Lupski et al., 2013). However, the capability to detect rare variants in personal genomes has provided a diagnostic challenge. Traditionally, the identification of causative or associated genetic variation has relied on gene identification in families or patient cohorts followed by genetic studies in model organisms to define the function of the gene in vivo. Several studies have made use of phenotypic information in Drosophila to identify genes associated with human diseases or traits (Bayat et al., 2012; Neely et al., 2010). However, the large number of variants detected by WES with poorly defined phenotypic consequences makes it challenging to tie a specific variant/gene to a given disease phenotype. Yet, these rare variants have a strong contribution to disease (Lupski et al., 2011). The interpretation of such genome-wide variation is hindered by our lack of understanding of gene function for the majority of annotated genes in the human genome.

We identified mutations in 165 genes, most of which have not been characterized previously in vivo. We provide data that suggest this gene set can be utilized as a resource to study numerous disease-causing genes. In addition, we present data that there is a fundamental difference between ethyl methanesulfonate (EMS) screens and RNAi screens. Moreover, we show that fly genes with more than one homolog are much more likely to be associated with human genetic disorders. Finally, we demonstrate that merging data sets—genes identified in the fly screen and rare variant alleles in the human homologs in families with Mendelian disease—can assist in human disease gene discovery and provide biological insights into disease mechanisms.

RESULTS

A Mosaic Genetic Screen on the X Chromosome

To isolate mutations in essential genes that are required for proper development, function, and maintenance of the Drosophila nervous system, we performed an F3 adult mosaic screen on an isogenic (iso) y w FRT19A X chromosome (Figure 1 and Figures S1 and S2 available online). We mutagenized males using a low concentration of ethyl methane-sulfonate (EMS), established 31,530 mutagenized stocks, and identified 5,857 stocks that carry recessive lethal mutations. To identify a broad spectrum of mutations and isolate genes that affect multiple biological processes, we screened for numerous phenotypes that affect the nervous system. We also screened for seemingly unrelated phenotypes, such as wing and pigmentation defects. Genes that affect wing veins and notching have been shown to play roles in critical pathways that affect numerous organs, including the nervous system. To assess phenotypes in the tissues of interest, we induced mitotic clones in the thorax and wing with Ultrabithorax-flippase (Ubx-FLP) (Jafar-Nejad et al., 2005) and in the eye with eyeless-flippase (ey-FLP) (Newsome et al., 2000). We did not pursue mutations that caused cell lethality or showed no/minor phenotypes (Figure 1A). While these genes are clearly important, they are difficult to study and these mutants were not kept. We selected 2,083 lethal lines with interesting phenotypes for further characterization (Figures 1A and 1B).

In the Ubx-FLP screen, we assessed the number and size of mechanosensory organs (bristles) on the fly cuticle to identify genes required for neural development (Figures 1C and 1D and S2A–S2C) (Charng et al., 2014). We also screened for alterations in the color of bristles and cuticle to permit identification of genes involved in dopamine synthesis, secretion, metabolism, or melanization (Yamamoto and Seto, 2014) (Figure S2D). In addition, we selected mutations that affect wing morphogenesis to isolate genes that regulate core signaling pathways, including Notch, Wnt, Hedgehog, and BMP/TGF-β (Bier, 2005) (Figures S2E–S2J). Indeed, these pathways have been implicated in synaptic plasticity and neuronal maintenance in both fly and vertebrate nervous systems. In the ey-FLP screen, we assessed morphological defects in the eye and head to isolate genes involved in neuronal patterning, specification, and differentiation (Figures S2K–S2O). Moreover, we screened for mutations that cause glossy eye patches (Figure S2P) or mutations that cause a head overgrowth (Figures S2Q–S2S). Glossy eye phenotypes are associated with mitochondrial mutations (Liao et al., 2006), while head overgrowth is linked to genes in Hippo signaling, TOR signaling, intracellular trafficking, and cell polarity/adhesion, and these pathways are implicated in disorders such as autism, intellectual disability, and neurodegenerative diseases (Emoto, 2012; Saksena and Emr, 2009).

To isolate mutations that affect neuronal development, function, and maintenance in the visual system, we recorded electroretinograms (ERGs) in mutant eye clones in 3- to 4-week-old flies (Figures 1E–1I). By analyzing the on and off transients of ERGs
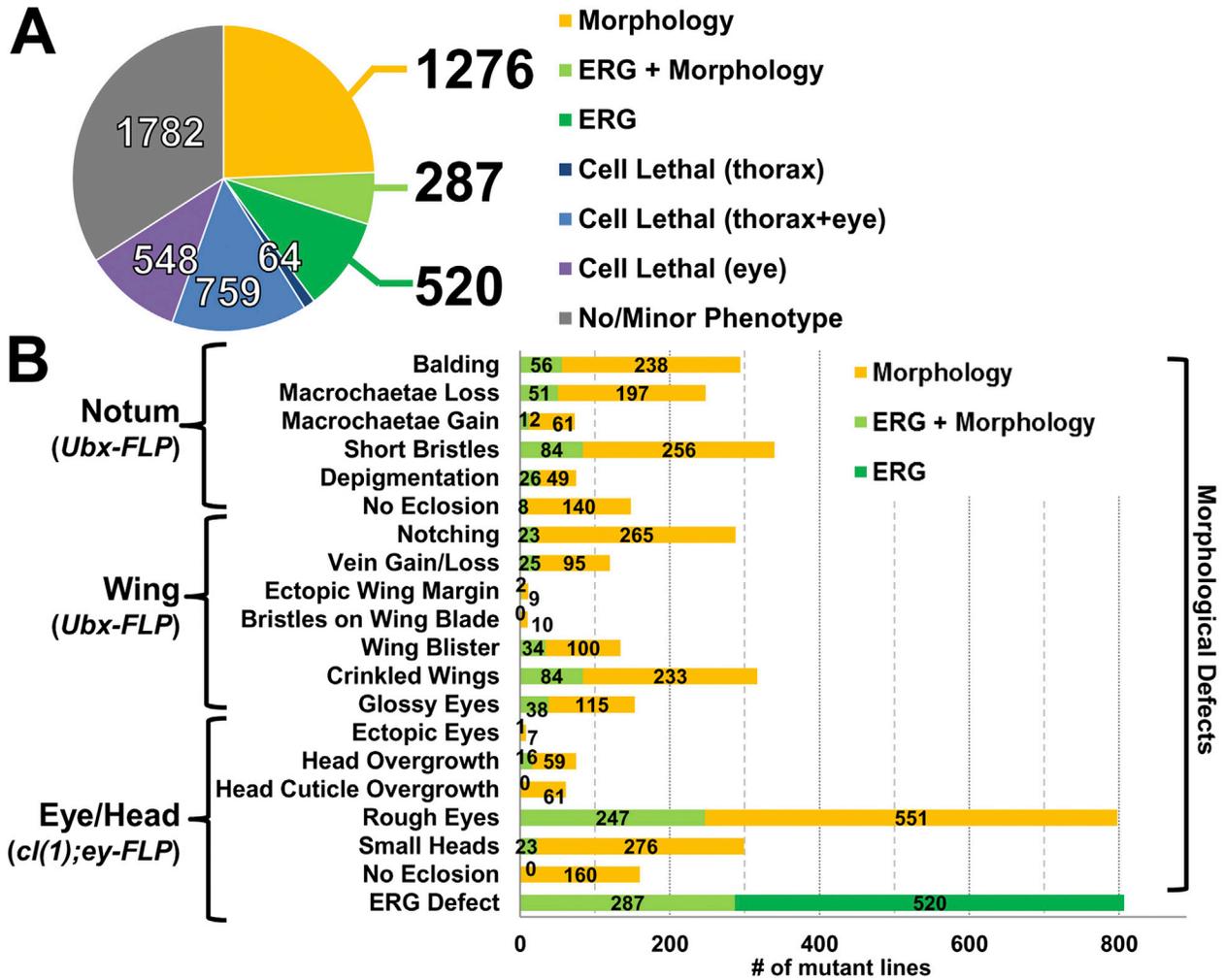
(legend on next page)

(Figure 1H), one can assess photoreceptor synaptic activity and axon guidance. A loss or reduction in the amplitude of depolarization (Figure 1G) is typically associated with genes that play a role in phototransduction, loss of which typically causes retinal degeneration (Wang and Montell, 2007). To identify mutations that cause a progressive demise of neurons, we screened young and old animals for ERG defects (Figures 1F and 1I). Ultrastructural defects in the photoreceptor terminals of young and old flies were also examined in some mutants with strong ERG phenotypes (Figures 1J–1M). Based on both the morphology screen and the ERG screen, we attempted to map 1,918 mutations (Figures S1 and S3).

Mutation Identification

On the X chromosome, complementation testing requires a genomic duplication on another chromosome to rescue male lethality. We selected 21 large ($\sim$0.5 Mb to $\sim$2 Mb) duplications that cover $\sim$95% of the X chromosome (Cook et al., 2010), crossed them into the mutant backgrounds, and rescued the lethality of 1,385 mutations (Figure S3). This permitted mapping of the lethality to 26 cytological intervals of the X chromosome. Complementation tests between mutants with similar phenotypes rescued by the same duplication allowed us to establish complementation groups. We grouped 450 mutations into 109 multiple allele complementation groups. The remaining 935 mutant strains include single alleles and a large number of mutations not yet assigned to complementation groups. To map the genes, we first performed deficiency mapping and Sanger sequencing. This allowed identification of the locus for 63 complementation groups. For the remaining groups and single alleles, we performed WGS (Haelterman et al., 2014) and rescued the phenotypes with molecularly defined $\sim$80 kb P[acman] duplications (Venken et al., 2010). By using both approaches, we were able to map 614 mutations to 165 genes, including 81 loci that have not been characterized in vivo (Tables 1 and S1) and are predicted to be involved in many diverse processes based on gene ontology analysis (Figures S2T and S2U).

Chemical Mutagenesis versus RNAi Screens

Two of the phenotypes that we screened, bristle development and depigmentation, allow a direct comparison between this screen and a genome-wide RNAi screen (Mummery-Widmer et al., 2009). This RNAi screen covered $\sim$80% of all X chromo-some protein coding genes. Interestingly, only 14% of the genes we identified in the bristle screen were also isolated in the RNAi screen (Figures 2A and 2B). Similarly, only 18% of the genes that we identified from the pigmentation screen were also identified in the RNAi screen (Figures 2C and 2D). Conversely, we did not identify the vast majority of genes that were identified by RNAi. In addition, a comparison of our gene list and those of two RNAi screens for wing margin (Saj et al., 2010) and eye morphological defects (Oortveld et al., 2013), show that these screens also identified very different sets of genes (Figures 2E and 2F). In summary, chemical screens identify a distinctive set of genes when compared to RNAi-based screens.

Links to Human Diseases Based on Online Mendelian Inheritance in Man

We next sought to determine if the 165 genes we identified in flies could enhance the understanding of human disease associated genes. Strikingly, 93% (153) of the fly genes isolated have homologs in humans (Tables 1 and S1; Figure 3A). This is a strong enrichment (χ^2 = 129, p < 0.001) for evolutionarily conserved genes between humans and flies when compared to the whole fly genome as only 48% of all fly genes have human homologs (Figure 3B). Moreover, the human homologs of 31% (48/153) of the identified fly genes have been associated with human disease in Online Mendelian Inheritance in Man (OMIM), 79% (38/48) of which exhibit neurological signs and symptoms (Figure 3A; Table S1). Of the genes that are conserved but not yet associated with Mendelian diseases with neurological symptoms, 65 genes have potential relationships to neurologic diseases (Figure 3A; Table S2). Therefore, the essential genes that we identified in this screen are highly conserved and many of their homologs have already been implicated in human disorders, showing that the screening strategy is effective.

Data analysis revealed a striking difference in the number of genes associated with disease depending on the number of human homologs for each fly gene. Fly genes that have a single human homolog have many fewer disease genes represented in the OMIM database than those that have more than one homolog. There is a 2-fold enrichment (χ^2 = 10.7, p < 0.001) of fly genes with more than one human homolog associated with diseases in the OMIM database compared to fly genes that have

Figure 1. Summary of the *Drosophila* X Chromosome Screen

(A and B) Pie chart (A) and bar graph (B) of phenotypes scored in the screen. The numbers represent mutations in each phenotypic category. Note that one strain may show more than one phenotype in (B).

(C and D) Examples of phenotypes observed in the notum. (C) Clones induced in a wild-type background, clone borders are marked by a white dotted line, (D) example of bristle loss in mutant clones (white arrows) (see Extended Experimental Procedures).

(E–I) Examples of ERG traces from mutant clones in the eye. A typical ERG has an on transient (blue arrows), depolarization (orange line) and an off transient (blue arrow head). ERGs were recorded in young (1- to 3-day-old) and old (3- to 4-week-old) flies for each genotype. (E) ERG of young or aged flies that show no obvious difference. (F) ERGs showing amplitude reduction in aged flies. (G) ERGs showing amplitude and on- and off-transient reduction in both young and aged mutants. (H) ERGs showing no or very small on transient in both young and aged flies. (I) ERGs showing on and off transients that are either absent or very small in aged flies carrying mutant clones in eye.

(J–M) Ultrastructural analysis using transmission electron microscopy (TEM) on young (2-day-old) and aged (3-week-old) mosaic flies. Red arrowheads indicate the rhabdomeres. (J) Young wild-type control eye: regular array of ommatidial structures with seven rhabdomeres surrounded by pigment (glia) cells. (K) Young mutant rhabdomeres showing intact structures. (L) Aged control eye tissue with intact rhabdomeres. (M) Aged mutant eye tissue with a strong degeneration of rhabdomeres.

See also Figures S1, S2, S3.

Cell

Table 1. List of 165 Fly Genes and 259 Corresponding Human Homologs Identified from the Screen

Fly Gene	Human Homologs (*OMIM)	Fly Gene	Human Homologs (*OMIM)	Fly Gene	Human Homologs (*OMIM)
Aats-his	**HARS*, HARS2***	COQ7	COQ7	**para**	**SCN1A*, SCN2A*, SCN3A,**
AP-1γ	AP1G1, AP1G2, **AP4E1***	Crag	DENND4A, DENND4B, DENND4C		**SCN4A*, SCN5A*, SCN7A, SCN8A*, SCN9A*, SCN10A*, SCN11A***
ari-1	ARIH1				
arm	**CTNNB1***	**Cyp4d2**	**CYP4V2***		
Arp2	ACTR2	DAAM	DAAM1, DAAM2	parvin	PARVA, PARVB, PARVG
ATP7	**ATP7A*, ATP7B***	**dlg1**	DLG1, DLG2, **DLG3***, DLG4	pck	CLDN12
baz	PARD3, PARD3B	Dlic	DYNC1LI1, DYNC1LI2	**Pgd**	**PGD***
ben	UBE2N	dor	VPS18	**phl**	ARAF, **BRAF*, RAF1***
br	-	dsh	DVL1, DVL2, DVL3	PI4KIIIalpha	PI4KA
Brms1	BRMS1, BRMS1L	**Dsor1**	**MAP2K1*, MAP2K2***	**por**	**PORCN***
cac	**CACNA1A***, CACNA1B CACNA1E	dwg	MZF1, ZSCAN22	pot	-
		Efr	SLC35B4	PpV	PPP6C
Cap	**SMC3***	egh	-	Prosα4	PSMA7, PSMA8
car	VPS33A	**eIF2B-ε**	**EIF2B5***	Psf3	GINS3
CDC45L	CDC45L	elav	ELAVL1, ELAVL2, ELAVL3, ELAVL4	rap	FZR1
Cdk7	CDK7			Rbcn-3A	DMXL1, DMXL2
CG11409	-	ewg	NRF1	**Rbcn-3B**	WDR7, **WDR72***
CG11417	ESF1	**fh**	**FXN***	**Rbf**	**RB1***, RBL1, RBL2
CG11418	**MTPAP***	flil	FLII	Rhp	RHPN1, RHPN2
CG12125	FAM73A, FAM73B	flw	PPP1CB	RpII215	POLR2A
CG12991	-	fs(1)h	BRD2, BRD3, BRD4, BRDT	RpS5a	RPS5
CG13365	-			Sas10	UTP3
CG14442	ZNF821	Fur2	PCSK5, PCSK6	**schlank**	CERS1, CERS2, **CERS3***, CERS4, CERS5, CERS6
CG14786	**LRPPRC***	Gtp-bp	SRPR		
CG15208	C21orf2	hfw	-	**scu**	**HSD17B10***
CG15896	KIAA0391	Hlc	DDX56	Sec16	SEC16A, SEC16B
CG1597	**MOGS***	**hop**	JAK1, **JAK2*, JAK3*, TYK2***	sgg	GSK3A, GSK3B
CG1703	ABCF1	Hr4	NR6A1	**shi**	DNM1, **DNM2***, DNM3
CG1749	UBA5	Hsc70-3	HSPA5	**sicily**	**NDUFAF6***
CG17776	-	Inx2	-	skpA	SKP1
CG17829	HINFP	kdn	CS	**Smox**	SMAD2, **SMAD3***
CG18624	NDUFB1	l(1)1Bi	MYBBP1A	smr	NCOR1, NCOR2
CG2025	NRD1	l(1)G0156	IDH3A	SNF1A	PRKAA1, PRKAA2
CG2918	HYOU1	l(1)G0222	ANKLE2	sno	SBNO1, SBNO2
CG3011	SHMT1, SHMT2	l(1)G0230	ATP5D	**Sp1**	**SP7***, SP8, SP9
CG3149	**RFT1***	**l(1)G0255**	**FH***	**stim**	**STIM1***, STIM2
CG32649	**ADCK3***, ADCK4	**l(1)G0334**	**PDHA1***, PDHA2	svr	CPD
CG32694	-	**Marf**	MFN1, **MFN2***	tay	AUTS2
CG32795	TMEM120A, TMEM120B	**Mcm6**	**MCM6***	temp	PTAR1
CG34401	ZSWIM8	**mew**	**ITGA3*, ITGA6*, ITGA7***	TH1	NELFCD
CG3446	**NDUFA13***	mRNA-cap	RNGTT	tko	MRPS12
CG3704	GPN1	mRpL38	MRPL38	**trr**	KMT2C, **KMT2D***
CG3857	SMG9	mRpS25	MRPS25	**ubqn**	UBQLN1, **UBQLN2***, UBQLN3, UBQLN4, UBQLNL
CG4078	**RTEL1***	mRpS30	MRPS30		
CG4165	USP16	mst	MSTO1	Upf1	UPF1
CG42237	PLA2G3, PROCA1	mus101	TOPBP1	Upf2	UPF2
CG42593	UBR3	mxc	NPAT	Usf	USF1, USF2

(Continued on next page)

204 Cell *159*, 200–214, September 25, 2014 ©2014 Elsevier Inc.

Table 1. *Continued*

Fly Gene	Human Homologs (*OMIM)	Fly Gene	Human Homologs (*OMIM)	Fly Gene	Human Homologs (*OMIM)
CG42749	-	***Myb***	***MYB****, *MYBL1*, *MYBL2*	*Usp7*	*USP7*
CG4407	*FLAD1*	***mys***	*ITGB1*, ***ITGB2****, ***ITGB4****,	*vnd*	*NKX2-2*, *NKX2-8*
CG4542	***ALG8****		*ITGB5*, *ITGB6*, *ITGB7*, *ITGB8*	*Vps26*	*VPS26A*, *VPS26B*
CG7065	-			*wapl*	*WAPAL*
CG7358	*ZC3H13*	***N***	***NOTCH1****, ***NOTCH2****,	*wds*	*WDR5*, *WDR5B*
CG8184	***HUWE1****		***NOTCH3****, *NOTCH4*	*wus*	*DNAJC22*
CG8636	*EIF3G*	*nej*	***CREBBP****, ***EP300****	*Ykt6*	*YKT6*
CG8949	*WAC*	*Nmd3*	*NMD3*	*Zpr1*	*ZNF259*
CG9650	*BCL11A*, *BCL11B*, *ZNF296*	*nonC*	*SMG1*	*βCop*	*COPB1*
Chc	*CLTC*, *CLTCL1*	***Nrg***	*CHL1*, ***L1CAM****, *NFASC*,	***β-Spec***	***SPTB****, *SPTBN1*, ***SPTBN2****,
CkIIbeta	*CSNK2B*		*NRCAM*		*SPTBN4*
CkIα	*CSNK1A1*, *CSNK1A1L*	*Nup93-1*	*NUP93*	*δCOP*	*ARCN1*
comt	*NSF*	***oc***	***CRX****, *OTX1*, ***OTX2****		

Human genes associated with Mendelian disease are marked with an asterisk and bold type, the corresponding fly gene is shown in bold. See also Tables S1, S5.

only one human homolog, 47% versus 22% (Figure 3C). This prompted us to assess if the bias is conserved for all fly genes. We found that a similar bias holds throughout the genome. Fly genes with more than one human homolog are more likely to be associated with diseases in the OMIM database than those with a single homolog, 40% versus 20% ($\chi^2 = 386$, $p < 0.001$) (Figure 3D and Extended Experimental Procedures). Indeed, 57 fly genes with more than one human homolog account for 100 diseases in the OMIM database (1.75 diseases per fly gene), an 8-fold enrichment when compared to fly genes with a single homolog (0.22 diseases per fly gene) (Figure 3E). This enrichment is not simply due to an absolute increase in the total number of human homologs because evolutionarily conserved genes that have more than one homolog are three times more enriched for OMIM diseases, 0.62 versus 0.22 diseases per human gene (Figure 3E). The difference between 1.75 and 0.62 is due to the number of homologs. Indeed, there are on average ∼3 human homologs for every fly gene that has more than one human homolog (data not shown). These data suggest that evolutionary gene duplication with divergence and further specialization of gene function may allow tolerance of mutation and viability versus lethality.

Since all of the mutations we isolated cause homozygous lethality, we analyzed the correlation between lethality, the number of human homologs, and their links to OMIM diseases for the entire fly genome. The number of essential genes in *Drosophila* has been estimated to be approximately 5,000 (Benos et al., 2001). Currently only ∼2,000 essential genes in FlyBase have transposable elements or EMS/X-ray-induced mutations (Marygold et al., 2013), representing about 40% of all essential fly genes. The proportion of essential genes varies with evolutionary conservation: an estimated 11% of the genes that do not have human homologs are essential, whereas 38% of the genes that have a single human homolog are essential ($\chi^2 = 354$, $p < 0.001$) (Figure 3F). Finally, an estimated 61% of the fly genes with more than one human homolog are essential. These data show that fly genes that have more than one human homolog
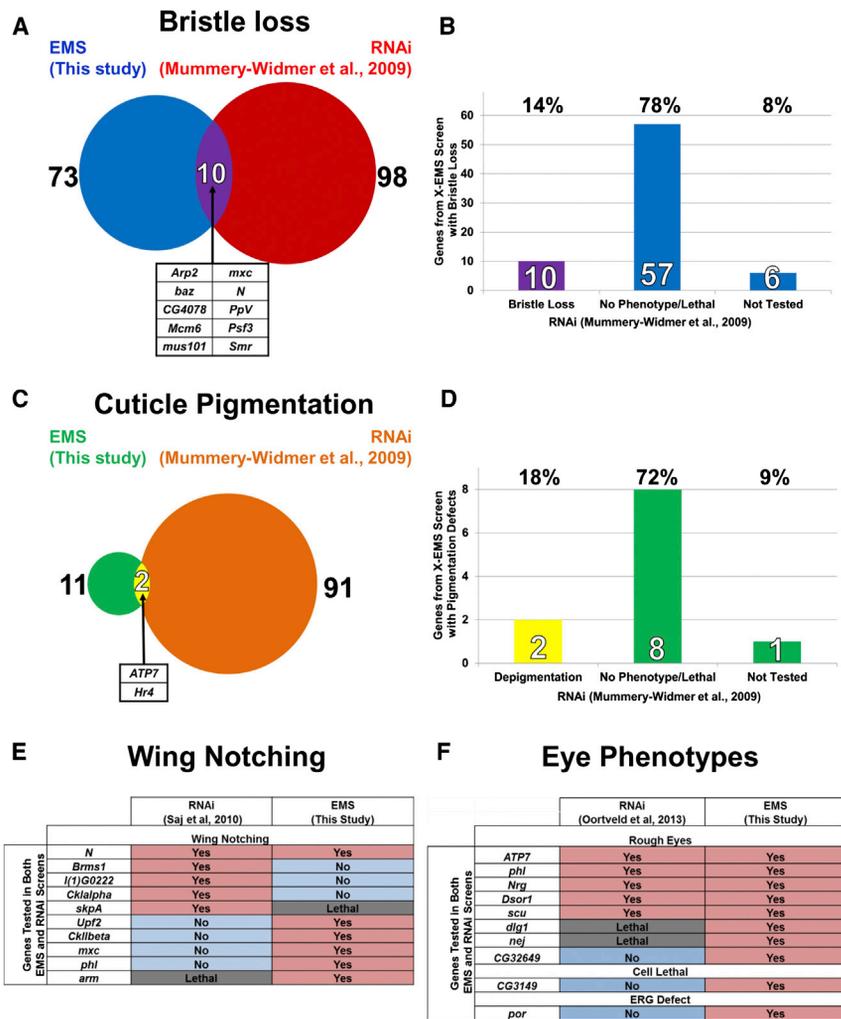
are more likely to cause lethality when mutated. Finally, human homologs of essential genes in *Drosophila* are more likely to be associated with human genetic diseases ($\chi^2 = 88$, $p < 0.001$) (Figure 3G). Therefore, we conclude that genes that are essential in flies and have multiple human homologs are the most likely to be associated with human diseases, potentially due to gene duplication and redundancy.

Combining Fly and Human Mutant Screen Data Sets to Identify Disease Genes

We next utilized the fly gene data set uncovered from the forward genetic screen in combination with a human exome data set to identify new human disease genes. We undertook a systematic search of all the variants in the human homologs of the genes identified from the *Drosophila* screen within WES data generated from undiagnosed cases of Mendelian diseases. This included 1,929 individuals in the Baylor-Hopkins Centers for Mendelian Genomics (BHCMG) (Figure 4).

BHCMG uses next-generation sequencing to discover the genetic basis of as many Mendelian diseases as possible (Bamshad et al., 2012). The study population includes singleton cases with sporadic disease, single families, and when possible, larger cohorts of affected individuals with a range of rare Mendelian phenotypes. A wide range of disorders are under investigation (http://www.mendelian.org/). In general, patients are recruited when a Mendelian disease seems highly likely and all reasonable efforts at a molecular diagnosis have failed. Due to the rare nature of the phenotypes, information from other patients or additional biological information from model organisms is required to fulfill the burden of proof for gene/disease association in such cases. For this reason, our *Drosophila* resource of mutant genes was integrated with our human exome variant and Mendelian phenotype (Hamosh et al., 2013) databases, and the combination approach was used to solve some of the cases.

We analyzed 237 out of the 259 (Table 1) homologs of fly genes identified through the X chromosome screen as they were validated at the time of analysis. We included all 237 genes,

A Bristle loss

EMS (This study) — RNAi (Mummery-Widmer et al., 2009)

73 | 10 | 98

Arp2	mxc
baz	N
CG4078	PpV
Mcm6	Psf3
mus101	Smr

B

14% | 78% | 8%

10 (Bristle Loss) | 57 (No Phenotype/Lethal) | 6 (Not Tested)

Genes from X-EMS Screen with Bristle Loss

RNAi (Mummery-Widmer et al., 2009)

C Cuticle Pigmentation

EMS (This study) — RNAi (Mummery-Widmer et al., 2009)

11 | 2 | 91

ATP7
Hr4

D

18% | 72% | 9%

2 (Depigmentation) | 8 (No Phenotype/Lethal) | 1 (Not Tested)

Genes from X-EMS Screen with Pigmentation Defects

RNAi (Mummery-Widmer et al., 2009)

E Wing Notching

		RNAi (Saj et al, 2010)	EMS (This Study)
		Wing Notching	
Genes Tested in Both EMS and RNAi Screens	N	Yes	Yes
	Brms1	Yes	No
	l(1)G0222	Yes	No
	CkIalpha	Yes	No
	skpA	Yes	Lethal
	Upf2	No	Yes
	CkIIbeta	No	Yes
	mxc	No	Yes
	phl	No	Yes
	arm	Lethal	Yes

F Eye Phenotypes

		RNAi (Oortveld et al, 2013)	EMS (This Study)
		Rough Eyes	
Genes Tested in Both EMS and RNAi Screens	ATP7	Yes	Yes
	phl	Yes	Yes
	Nrg	Yes	Yes
	Dsor1	Yes	Yes
	scu	Yes	Yes
	dlg1	Lethal	Yes
	nej	Lethal	Yes
	CG32649	No	Yes
		Cell Lethal	
	CG3149	No	Yes
		ERG Defect	
	por	No	Yes

Figure 2. Comparison of Results from This EMS Screen and Previous RNAi Screens

(A and B) Venn diagram (A) and bar graph (B) showing overlap between two screens for bristle loss defects. The genes that were identified in the EMS screen were also screened by RNAi (Mummery-Widmer et al., 2009) and 10 caused a bristle loss whereas 57 showed no phenotype or caused lethality.

(C and D) Venn diagram (C) and bar graph (D) showing overlap between two screens for pigmentation defects (this screen and the RNAi screen of Mummery-Widmer).

(E) Comparison of the results of these screens for wing notching defects.

(F) Comparison of the results of these two screens for eye morphological defects.

filled Mendelian expectations of a dominantly inherited disorder in the family under investigation. Interestingly, 22/31 individuals in which the variant met Mendelian expectations had a neurological disease.

As a proof-of-principle, we report six patients/families with mutations in three genes. In addition, we identified 25 other individuals in which the variant in the homolog of the fly gene met Mendelian expectation. Some of these individuals were found to have candidate variants in multiple genes, some had too few living relatives for further study, and for others, studies are ongoing. Therefore, a systematic search of variants within the genes identified in the *Drosophila* screen was able to identify and prioritize a subset of variants with Mendelian inheritance in families that could be studied. Among these, we found examples of known disease genes (*DNM2*), a novel disease association to a known disease gene (*CRX*), and novel candidate genes for disease (*ANKLE2*).

DNM2 and Charcot-Marie-Tooth Neuropathy

Examination of a homolog of *Drosophila shibire (shi)*, the gene that encodes Dynamin, led to a molecular diagnosis for two individuals with heterozygous mutations in *DNM2* (Figures S4A and S4B). Both patients were diagnosed with a distal symmetric polyneuropathy consistent with Charcot-Marie-Tooth disease (CMT) (See Extended Results). Mutations in *DNM2* are associated with CMT Type 2M (OMIM 606482), an axonal form primarily affecting neurons (Figure S4C). Patient 1, the proband in Figure S4A, presented at age 12 with hand tremor, calf cramps, lower limb paresthesias, and difficulty with heel walking. She is a member of a family with three generations of neuropathy (Figure S4A), and the heterozygous G358R variant cosegregated with CMT (Figure S4A). Patient 2, the proband in Figure 4B, (currently 88 years old) presented at age 40 with lower extremity weakness. His nerve conduction studies showed low amplitudes and borderline

regardless of whether they were previously identified to be associated with Mendelian diseases in OMIM, to avoid any bias. We filtered out variants reported as having greater than 1% allele frequency in databases of control individuals (See Extended Experimental Procedures). Under the assumption of a recessive model data set, we included all variants that met these criteria and were homozygous or had two heterozygous variants affecting the same gene. The latter set was not tested for *cis* or *trans* orientation of the variants prior to analysis. A dominant model included heterozygous variants. These were filtered even more stringently for allele frequency such that only variants that had not been observed in the control data sets were studied (Table S3).
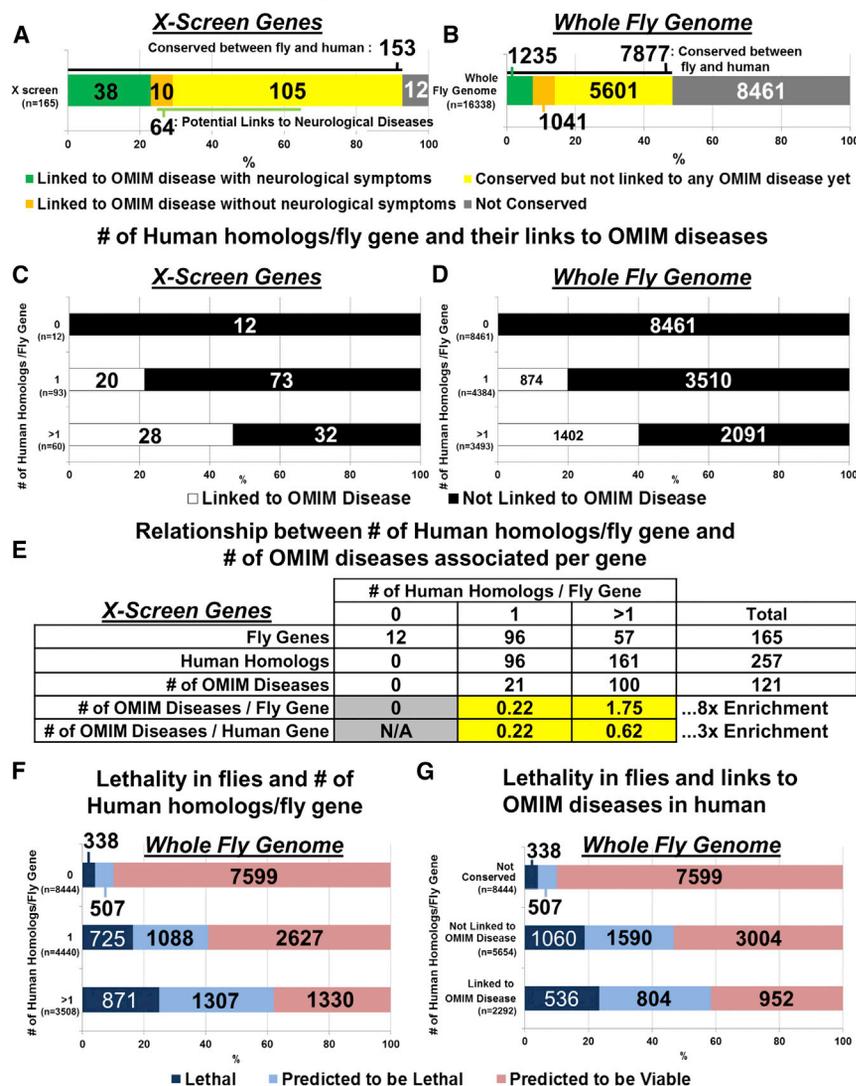
To explore potential associations with disease, we prioritized variants for segregation analysis within families (Figure 4). We performed Sanger sequencing or explored segregation in families for 64 variants in 24 genes within 34 individuals in the recessive data set and found that 15 variants in 8 genes within 10 individuals fulfilled Mendelian expectations for recessive inheritance. Likewise, for the dominant data set, we explored the segregation for 158 variants in 85 genes within 99 individuals. We found 22 variants in 15 genes within 21 individuals that ful-

Conservation of fly genes and their links to OMIM diseases

Figure 3. Essential Fly Genes Associated with More Than One Human Homolog Are More Likely to be Linked to Human Diseases

(A) Classification of genes identified in the screen based on human homologs and associated diseases.

(B) Classification of the whole fly genome according to the same criteria as in (A).

(C and D) Relationship between the number of human homologs per fly gene and their association with human diseases for genes identified in the screen (C) and the whole fly genome (D).

(E) The number of human homologs per fly gene and their enrichment in OMIM associated human diseases.

(F) Relationship between the number of human homologs per fly gene and lethality in flies.

(G) Relationship between genes associated with lethality in flies and OMIM associated human diseases.

See also Table S2.

---

childhood vision loss seen in dominant cone-rod dystrophy, Leber congenital amaurosis, and autosomal dominant retinitis pigmentosa (OMIM 120970, 613829).
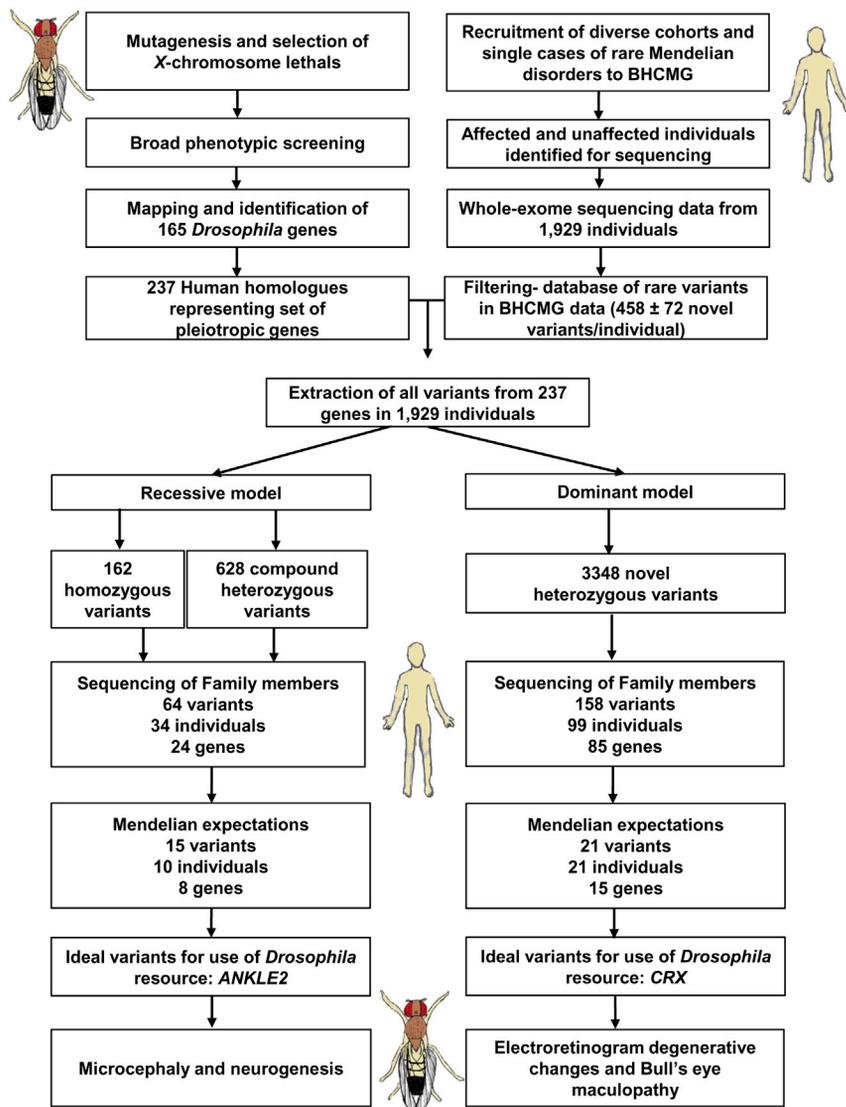
The three cases of bull's eye maculopathy included two individuals with no family history of retinal disease (patients 3 and 4) and one multigenerational pedigree (patient 5 [S150X]) (Figure 5A). The affected individuals in the family of patient 5 developed symptoms at age 50 (range 28–63 years), and three family members with the S150X mutation had minimal symptoms at initial evaluation between the age of 55–60. Despite having near normal vision, ophthalmologic exam in the retina of these individuals revealed advanced bull's eye maculopathy with foveal sparing explaining the modest effect on vision.

Patient 5 exhibits retinal abnormalities (Figure 5B–B'), abnormal autofluorescence in the fundus (Figure 5C–C'), aberrant Optical Coherence Tomography (OCT, Figure 5D–D') and electroretinograms (Figure 5E), all consistent with bull's eye maculopathy. The three new alleles are all encoding predicted truncations of the OTX transcription factor domain (Figure 5F).

Functional analysis of homozygous *oc* mutant clones reveal that the ERGs in young animals are nearly normal (Figure 5G) but defective in 7-day-old flies, including reduced amplitude and loss of on transients (Figure 5G, blue arrows). This suggests that the photoreceptors become impaired over time. In summary, the defects in flies and humans show similarities.

ANKLE2 and Microcephaly

The *Drosophila* screen identified a mutation in *l(1)G0222*, the homolog of *ANKLE2* (dAnkle2) (Table 1). The mutation causes a loss of thoracic bristles and underdevelopment of the sensory organs in clones (Figure 6A). The human WES data identified

slowed velocities. He carries an E341K mutation in *DNM2* (Figure S4D). In addition to *DNM2*, WES revealed a variant in another CMT gene, *LRSAM1* in this patient (Figure S4B). Interestingly, dominant as well as recessive mutations in *LRSAM1* can cause CMT2P (OMIM 614436). Hence, either one or a combination of both genes may cause CMT in this family. While some clinical features of the probands made diagnosis difficult, the phenotypes of these cases were indeed consistent with CMT type 2.

CRX and Bull's Eye Maculopathy

Examination of one of the human homologs of *Drosophila ocelliless* (*oc, CRX* in humans) led to the identification of three cases of bull's eye maculopathy associated with dominant *CRX* alleles. *oc* encodes a homeobox transcription factor that regulates photoreceptor development (Vandendries et al., 1996). Identifying cases of bull's eye maculopathy, a late-onset slowly progressive retinal disorder, with *CRX* alleles was surprising because *CRX* is typically associated with much more severe

Figure 4. Flowchart for Discovery and Functional Studies of Disease Genes Using the _Drosophila_ Resource and Human Exome Data

See also Table S3, Figure S4.

normal (Figures 6G and 6H). A younger sister born a year later had severe microcephaly, spasticity, and similar hyper- and hypopigmented macules over all areas of her body. She died 24 hr after delivery from cardiac failure associated with poor contractility, although the basis for this was not known.

WES data of the proband, his affected sister, and both parents revealed four candidate genes that meet Mendelian expectation and are expressed in the CNS (Table S4). Table S4 shows the variants with their scores from four predictions programs (Liu et al., 2011). _ANKLE2_ was prioritized as a good candidate. To assess if _dAnkle2_ is involved in CNS development, we examined the brains of _Drosophila_ mutant larvae. Brain size in early third instar larval stages is similar to that of controls (Figure S5A). However, later in third larval stage, the brain becomes progressively smaller than control larvae (Figure S5A and Figures 6I and J). To confirm that _dAnkle2_ is an ortholog of human _ANKLE2_, we ubiquitously expressed human _ANKLE2_ in mutant flies and observed rescue of lethality and the small brain phenotype (Figures 6K–6L). These data indicate that _ANKLE2_ is implicated in CNS development and its molecular function is evolutionarily conserved.

variants in _ANKLE2_ in a family with apparent recessive microcephaly (Figures 6B and 6C). The proband, patient 6, has an extreme small head circumference, a low sloping forehead, ptosis, small jaw, multiple hyper- and hypopigmented macules over all areas of his body, and spastic quadriplegia (Figure 6D–6H; Extended Results, "Clinical Case Histories"). During his first year of life, he had unexplained anemia, and glaucoma. At 3 years, he had onset of seizures, and at 5.5 years, his weight was 10.7 kg (−4 SD), length 83.8 cm (−6 SD) and fronto-occipital circumference 38.2 cm (−9 SD).

Brain MRI in the newborn period demonstrated a low forehead, several scalp ruggae, and mildly enlarged extra-axial space with communication between the posterior lateral ventricles and the mesial extra-axial space. Other brain abnormalities included a simplified gyral pattern, mildly thickened cortex, small frontal horns of the lateral ventricles with mildly enlarged posterior horns of the lateral ventricles, and agenesis of the corpus callosum. The brainstem and cerebellum appeared relatively

To explore the cause of the small brain phenotype in _dAnkle2_ mutants, we assessed defects in processes which can cause small brain phenotypes: mitosis, asymmetric cell division, and apoptosis (Rujano et al., 2013). The number of neuroblasts, marked by Miranda (Ceron et al., 2001) is severely reduced in late third instar brain lobes (Figures 6M–6O and S5B and S5C). In the few neuroblasts that undergo division, the spindles are properly oriented toward the polarity axis (Figures S5D and S5E). In addition, centriole duplication, impaired in many primary human microcephaly syndromes (Kaindl et al., 2010), is not affected in _dAnkle2_ mutants (Figures S5F and S5G). Hence, loss of _dAnkle2_ causes a severe reduction in neuroblast number but does not seem to affect asymmetric division or centriole number.

To assess proliferation in the CNS, we induced mitotic clones of _dAnkle2_ in the brain and labeled them with Bromodeoxyuridine (BrdU)(Figures 6P–6R). As shown in Figure 6R, BrdU incorporation is strongly reduced in mutant clones when
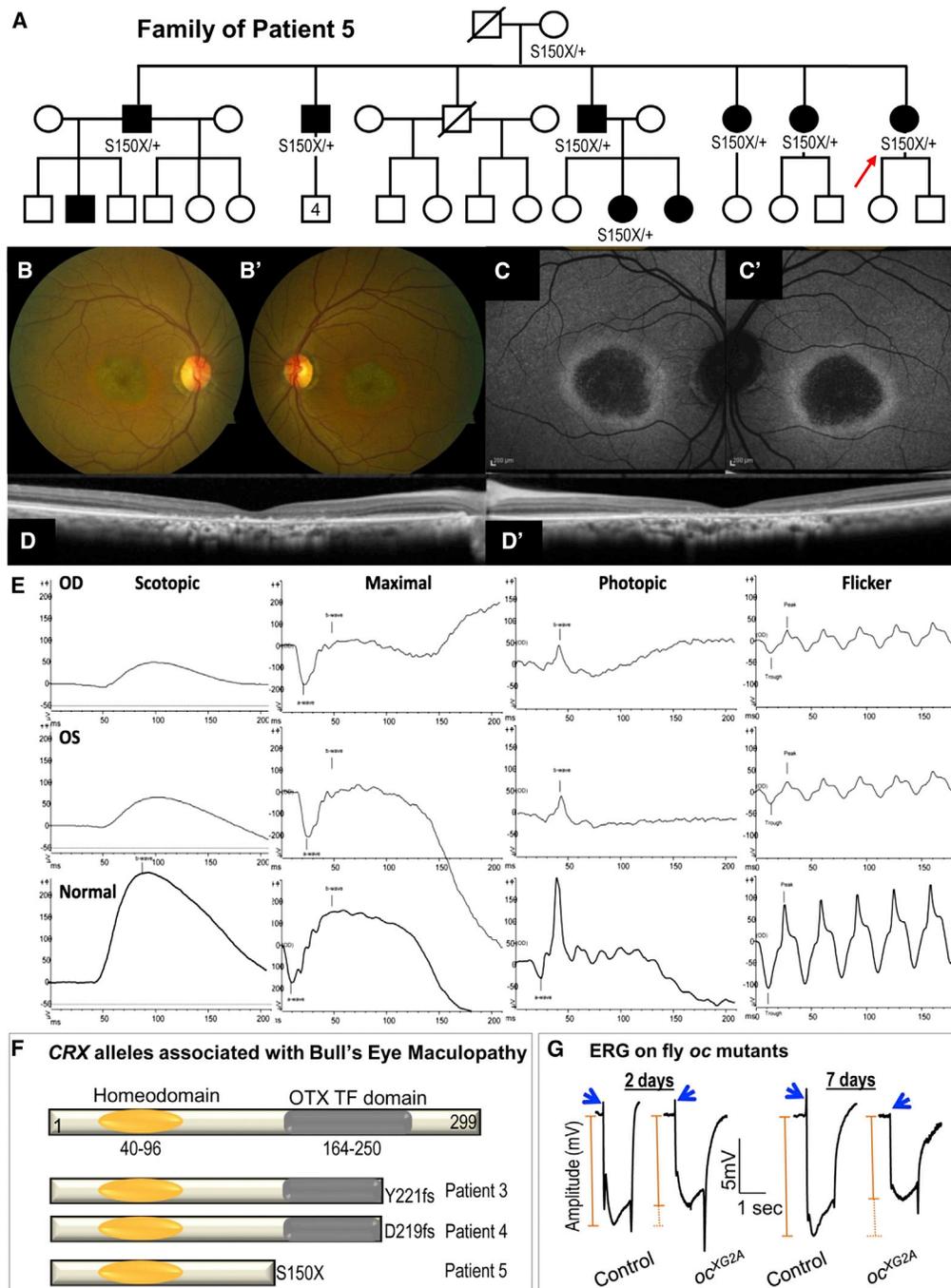
**Figure 5. Mutations in *CRX* Cause Bull's Eye Maculopathy**

(A) Pedigree of the family of patient 5 (red arrow) with multiple individuals with bull's eye maculopathy. The S150X mutation in CRX was identified in eight family members. DNA was not available for family members for whom screening results are not indicated.

(B–D) Clinical phenotypes of patient 5. (B–B') Fundus photography show fine granularity in the outer retina and speckled glistening deposits arranged in a ring around the macula. Peripheral fundi appear unaffected. (C–C') Autofluorescence images reveal a bull's eye phenotype with hypofluorescent macula surrounded by a hyperautofluorescent ring, suggesting a continuously atrophic macular area. (D–D') Optical coherence tomography shows central loss of the outer nuclear layer, ellipsoid line, external limiting membrane, and retinal pigment epithelium atrophy corresponding to area of hypoautofluorescence in (C–C').

(E) ERG of the proband: Electroretinographic traces showed implicit time delay and amplitude reduction in both scotopic and especially photopic responses in keeping with generalized cone-rod dysfunction.

(F) Structure of CRX protein and mutations in patients 3–5.

(G) ERG of control and *oc* mutant clone in 2-day-old and 7-day-old (in light) adult flies. Blue arrows indicate on transient in ERG. On transients are lost in 7-day-old flies. The orange line indicates the amplitude of ERG.
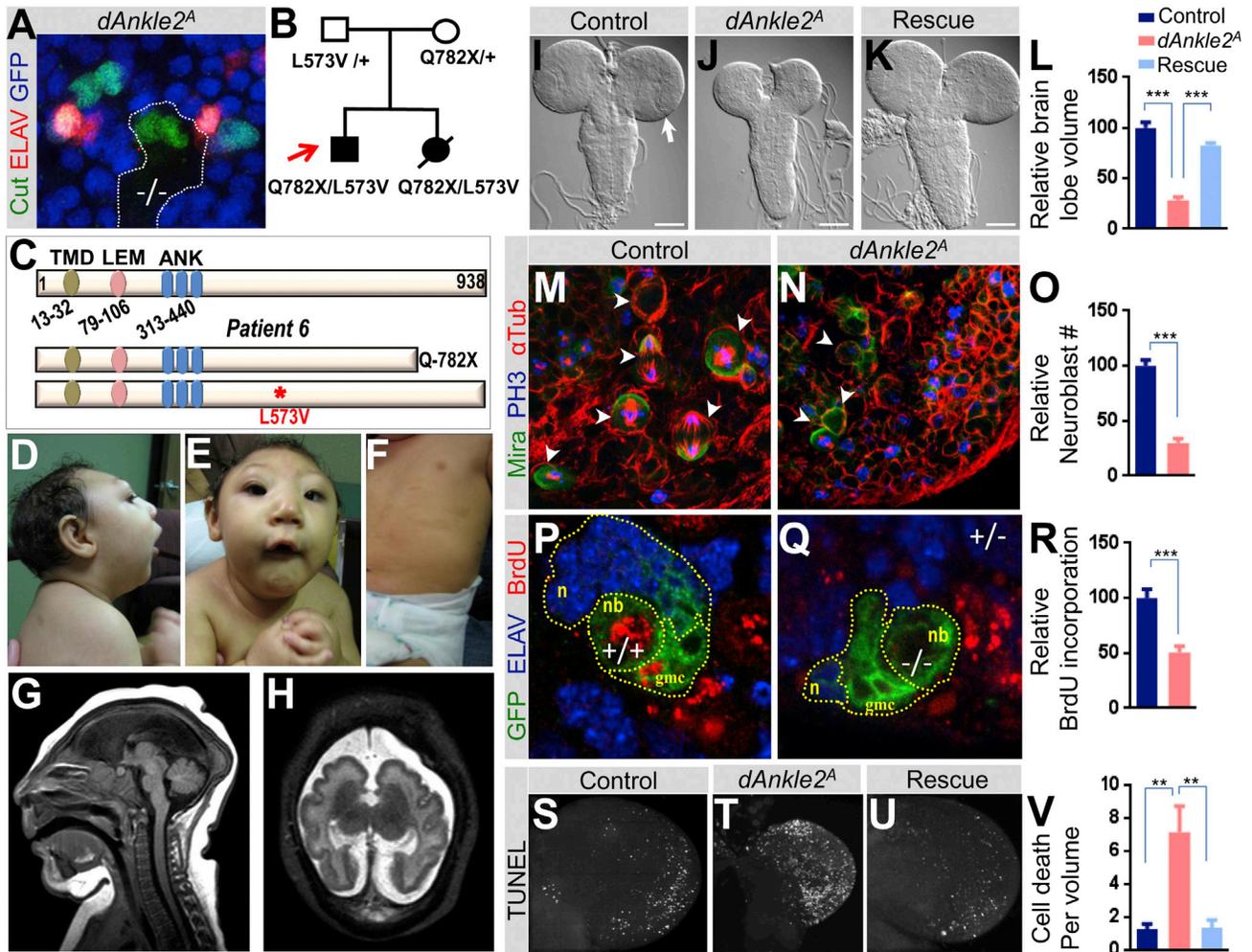
Figure 6. *ANKLE2* and Microcephaly

(A) *dAnkle2* mutant clone of the peripheral nervous system in the thorax of a fly. In wild-type tissue (GFP, shown in blue), sensory organs are comprised of four cells marked by Cut (green), one of which is a neuron marked by ELAV (red). In the mutant clone (−/−, nonblue), the number of cells per sensory organ is reduced to two and does not contain a differentiated neuron.

(B) Pedigree of the family of patient 6 (red arrow) with a severe microcephaly phenotype. Both affected individuals inherited variants from both parents in *ANKLE2*.

(C) Structure of ANKLE2 protein and mutations in patient 6. Abbreviations: transmembrane domain (TMD), LAP2/emerin/MAN1 domain (LEM), ankyrin repeats (ANK).

(D and E) Clinical phenotypes of the proband with a severe sloping forehead, microcephaly, and micrognathia.

(F) Scattered hyperpigmented macules on the trunk.

(G) Sagittal brain MRI of the proband in infancy with severe microcephaly, agenesis of the corpus callosum and a collapsed skull with scalp ruggae.

(H) Axial brain MRI showing polymicrogyria-like cortical brain malformations.

(I–L) Third instar larval brain of (I) control (*y w FRT19Aiso*); scale bar, 100 microns (J) *dAnkle2* mutant, and (K) *dAnkle2* mutant in which the human *ANKLE2* cDNA is ubiquitous expressed (Rescue). Note that brain lobe (arrow in I) size is reduced in *dAnkle2* mutant (J) and the phenotype is rescued by *ANKLE2* expression (K). Relative brain lobe volume of control, *dAnkle2*, and rescue using 3D confocal images is quantified in (L).

(M–O) Larval CNS neuroblasts (arrowheads) in control and *dAnkle2* mutant. Neuroblasts are marked by Miranda (Mira, green), chromosomes in dividing cells are marked by Phospo-Histone3 (PH3, blue), and spindles in dividing cells are marked by α-Tubulin (αTub, red). Relative number of neuroblasts in control and *dAnkle2* is shown in (O).

(P–R) BrdU incorporation (red) in control (P) and *dAnkle2* mutant clones (Q) marked by GFP (green, dotted lines) in larval brains. Differentiated neurons are marked by ELAV (blue). Neuroblast (nb), ganglion mother cells (gmc), and neurons (n) are marked. Quantification of relative BrdU incorporation is shown in (R).

(S–V) TUNEL assay in third instar larval brain lobes of (S) control, (T) *dAnkle2* mutant, and (U) Rescue. Quantification of TUNEL positive cells/volume (cell death) is shown in (V).

In (L, O, R, and V), error bars indicated SEM, *** indicates a p value < 0.001 and ** indicates a p value < 0.01. See also Table S4, Figure S5.

compared to wild-type clones, indicating that cell proliferation is severely impaired. In addition, the mutant clones (Figure 6Q) that contain a neuroblast and its progeny, the ganglion mother cells and neurons, contain many fewer cells than wild-type clones (Figure 6P). Finally, we observe a dramatic increase in apoptotic cells marked by TUNEL in the *dAnkle2* mutant brain lobes

(Figures 6S, 6T, and 6V). This cell death is rescued by the expression of the human cDNA encoding *ANKLE2* (Figures 6U and 6V). Therefore, defects in proliferation and excessive apoptosis are both contributing to the loss of CNS cells in *dAnkle2*.

DISCUSSION

Here we describe the generation of a large set of chemically induced lethal mutations on the *Drosophila* X chromosome that were screened for predominantly neurological phenotypes in adult mosaic flies. The mutations were assigned to complementation groups, mapped, and sequenced to associate as many genes as possible with specific phenotypes. We identified and rescued the lethality associated with mutations in 165 genes using a variety of mapping and sequencing methods. These mutations are available through the Bloomington *Drosophila* Stock Center and provide a valuable resource to study the function of human genes in *Drosophila* especially since 93% of the genes are evolutionarily conserved in human.

This mutant collection contains 21 genes associated with human diseases for which no mutations were previously available. The fly mutants thus enable the study of the basic molecular mechanism of 26 human diseases, including Leigh syndrome (*CG14786/LRPPRC*, *l(1)G0334/PDHA1*, and *sicily/NDUFAF6*), congenital disorders of glycosylation (*CG1597/MOGS*, and *CG3149/RFT1*), Usher syndrome (*Aats-his/HARS*), Friedreich ataxia (*fh/FXN*), and amyotrophic lateral sclerosis (*ubqn/UBQLN2*). Based on the gene list from the *Drosophila* screen, we explored a database of 1,929 human exomes from a Mendelian disease resource of patients with rare diseases. We examined the personal genomes for rare variants of the fly homologs and prioritized a subset of human rare variant alleles for segregation analysis. We report six families with distinct diseases in which the variants segregate and are likely responsible for causing the associated Mendelian disease.

The approach described here provides a valuable resource to study the function of many disease genes in different tissues. We propose that the screen strategy be expanded to the autosomes, and a number of guiding principles should be considered based on this study. First, the use of low concentrations of EMS is important as it minimizes the number of second site lethal and visible mutations (Haelterman et al., 2014). Second, screening for lethal mutations has major advantages as 93% of the isolated genes that are essential for viability are conserved, whereas only 48% of all *Drosophila* genes have evolutionarily conserved human homologs. Third, the isolation of lethal mutations also greatly facilitates genetic mapping. Fourth, screening for many different phenotypes casts a broader net and permits isolation of mutations in many different genes, a strategy that is also used in mice (White et al., 2013). Fifth, analyzing different phenotypes revealed that mutations in the majority of the genes cause more than one phenotype, consistent with extensive pleiotropism.

Comparison of the gene list identified from our EMS screen and several RNAi screens have shown that these approaches reveal very distinct sets of genes. There are multiple reasons that may lead to this difference. For example, since our screen was aimed at identifying mutations that cause lethality, we have not screened for genes that are nonessential. Thus, a number of genes that are nonessential but cause morphological defects are missed in our screen. On the other hand, RNAi may not be efficient or cause off-targeting effects (Green et al., 2014; Mohr, 2014). Regardless of the methods that are being used, rescue experiments and independent validation are critical to determine that the phenotype one observes is due to loss of the gene of interest when performing a genetic screen.

It is interesting to note that from our screen, essential fly genes with two or more homologs in humans have a significantly higher likelihood of being associated with Mendelian diseases than those that only have a single human homolog (Figure 3). This suggests that gene duplications of essential genes and subsequent evolutionary divergence may lead to genes that are partially redundant and more likely to be disease associated. Hence, when analyzing human exomes, it would seem more productive to start with homologs of evolutionarily conserved essential *Drosophila* genes that have two or more human homologs. In addition to these relationships to Mendelian traits, 17% (26/153) of the fly genes that have human homologs have been identified in GWAS (genome-wide association studies) for neurological disorders (Table S5). Hence, the collection of mutations described here may permit us to study genes for complex traits.

We uncovered a genetic basis in a few cases for which the gene was previously known. For example, the study of *DNM2* revealed previously studied phenotypes associated with mutations in the gene (CMT, Figure S4). In another case we observed that mutations in a gene caused unexpected phenotypes. Indeed, we identified three families with bull's eye maculopathy, a condition that is much milder and with a later age of onset than conditions typically associated with *CRX* truncations such as Leber congenital amaurosis (leading to blindness before a year of life) and cone rod dystrophy (a condition with onset in the first or second decade). Interestingly, other truncating alleles have been reported both N- and C-terminally to the OTX transcription factor domain in patients with these severe phenotypes. Therefore, while CRX mutations can produce variable phenotypes (Huang et al., 2012), bull's eye maculopathy has not been associated with deleterious CRX variants. Our data suggest that some symptoms may manifest at older ages, and the phenotypic spectrum of *CRX* mutations includes late-onset mild retinopathy.

We identified deleterious alleles in *ANKLE2* in two individuals in a family affected by severe microcephaly. In flies, we observed severe defects in neuroblast proliferation and excessive apoptosis in the third instar larval brain of *dAnkle2* mutants. This knowledge, combined with the observation that expression of human *ANKLE2* in *dAnkle2* mutants rescues lethality, brain size, and apoptosis, provide strong evidence that *ANKLE2* is responsible for the microcephaly in the family. Moreover, ANKLE2 has been shown to physically and genetically interact with VRK1 in *C. elegans* and vertebrates (Asencio et al., 2012), and loss of fly *VRK1* (also known as *ballchen (ball)* or *nhk-1* in flies) also causes a small brain phenotype in third instar larvae (Cullen et al., 2005). It is therefore interesting to note that mutations in *VRK1* also cause microcephaly in patients (Figure S5H) (Gonzaga-Jauregui et al., 2013).

The pattern of brain abnormalities and microcephaly in our patient with *ANKLE2* mutations is somewhat similar to patients with autosomal recessive *CLP1* mutations. *CLP1* encodes an RNA kinase involved in tRNA splicing (Karaca et al., 2014; Schaffer et al., 2014). The *Clp1* homozygous kinase-dead mouse exhibits microcephaly that worsens with age due to apoptosis. Hence, apoptosis may be a common denominator in these forms of microcephaly.

Phenotypic information of *Drosophila* mutants allows researchers to understand the potential in vivo function of their human homologs. The cases of *oc/CRX* and *dAnkle2/ANKLE2* are examples in which some direct phenotypic comparisons are possible between the fly mutant and human conditions. However, one of the major drawbacks of comparing phenotypes in different species is that a comparison between different tissues and organs is not always obvious. How do we relate wing vein defects or a rough eye with the phenotypes observed in human genetic diseases? Numerous strategies have been outlined by Lehner (Lehner, 2013) and one of the most compelling strategies is based on orthologous phenotypes or phenologs (McGary et al., 2010). Genes tend to work in evolutionarily conserved pathways, allowing the direct transfer from genotype-phenotype relations between species. For example, mutations in a subset of genes that function in mitochondrial quality control cause a high incidence of muscle mitochondrial defects in adult flies and Parkinson disease (PD) in humans (Jaiswal et al., 2012), suggesting that new genes that affect muscle mitochondria in adult flies are good candidates for PD. Indeed, it may well be that phenotypic similarities between fly and man will be the exception rather than the rule. Regardless, we provide evidence that the use of unbiased screens in the fly and the resulting genetic resources will provide opportunities to prioritize human exome variants and to explore the underlying function of these and many other disease-causing genes in vivo.

EXPERIMENTAL PROCEDURES

Fly Strains
The strains used in this study including the mutations and duplications and deletion strains used for mapping are described in Flybase (Marygold et al., 2013) (see also Extended Experimental Procedures).

Isogenization and Mutagenesis
Isogenization of *y w FRT19A* chromosome was performed using standard genetic crosses. Mutagenesis was performed by feeding isogenized *y w FRT19A iso* males with sucrose solution containing a low concentration (7.5–10 mM) of EMS as described (Bökel, 2008). After recovery from mutagenesis, these males were mated en masse with *Df(1)JA27/FM7c Kr > GFP* virgin females for 3 days. In the F1 generation, *y w mut* FRT19A/FM7c Kr > GFP* (mut* indicates the EMS-induced mutation) virgins were collected and 33,887 individual females were crossed with *FM7c Kr > GFP* males to establish independent balanced stocks. A total of 5,859 lines carried lethal mutations and the remaining stocks were discarded.

Complementation and Mapping
Lines that exhibited a strong morphological and/or ERG phenotype were subjected initially to duplication mapping. Subsequently, lines that were rescued by the same duplication and exhibit similar phenotypes were crossed inter se to establish complementation groups based on lethality. Complementation groups were further fine mapped using deficiencies that cover the region of interest.

Gene Identification
When a complementation group was mapped to a small region (~30–300 kb, varies depending on available resources), we searched for publically available lethal mutations that map to the same region using FlyBase (Marygold et al., 2013). We performed complementation tests using >1 mutant allele when possible. For complementation groups that complemented all available lethal mutations in the region, we performed Sanger sequencing using standard methods. To expedite gene identification we also used Illumina-based whole-genome sequencing technology (Haelterman et al., 2014)

Ethics Statement
Informed consent was obtained prior to participation from all subjects or parents of recruited subjects under an Institutional Review Board approved protocol at BCM.

Study Subjects
The analysis of 1,929 exomes from BHCMG described was performed in a database from the WES of over 160 separate phenotypic cohorts. The sequencing data included family-based studies in which both affected and unaffected family members were sequenced, single individuals with unique phenotypes, as well as larger cohorts of up to 50–60 cases with the same phenotype. Selection of subjects was performed by a phenotypic review committee based on the likelihood of the Mendelian inheritance for the disease phenotype.

Whole-Exome Capture, Sequencing and Data Analysis
All of the subjects enrolled in the BHCMG underwent WES using methods previously described (Lupski et al., 2013) (Extended Experimental Procedures). Produced sequence reads were mapped and aligned to the GRCh37 (hg19) human genome reference assembly using the HGSC Mercury analysis pipeline (http://www.tinyurl.com/HGSC-Mercury/). Variants were determined and called using the Atlas2 suite to produce a variant call file (VCF). High-quality variants were annotated using an in-house developed suite of annotation tools (Bainbridge et al., 2011a).

ANKLE2 Construct and Transgenesis
Human *ANKLE2* cDNA was cloned into pUASTattB (Bischof et al., 2007) tagged vectors (N-terminal FLAG) using In-Fusion HD Cloning Kit (Clontech) and vector was linearized with NotI and XhoI. The construct was inserted in VK33 (Venken et al., 2006).

ACCESSION NUMBERS

The dbGAP accession number for the data reported in this paper is phs000711.v1.p1. Additional details are available at http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000711.v1.p1. Biosample IDs for patient 1 (BAB3655), patient 2 (BAB3659), patient 6 (LR06-300a1), and patient 6 family data (LR06-300a2, LR06-300f, LR06-300m) are in Table S6 and data for these individuals is available at this link: http://www.ncbi.nlm.nih.gov/sra?Db=sra&DbFrom=bioproject&Cmd=Link&LinkName=bioproject_sra&LinkReadableName=SRA&ordinalpos=1&IdsFromResult=237879

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Results, Extended Experimental Procedures, five figures, five tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2014.09.002.

AUTHOR CONTRIBUTIONS

Thorax and wing screen: S.Y. and W.L.C. Eye and ERG screen: M.J., B.X., K.Z., and V.B. Fly mapping: S.Y., M.J., W.L.C., B.X., K.Z., V.B., H.S., N.A.H., G.D., T.L., K.C., U.G., A.T.L.-M., K.L.S., and R. Chen. Bioinformatic analysis: S.Y., M.J., M.F.W., Y.-W.W. and Z.L. Whole-exome sequencing: T.G., S.N.J., D.M., E. Boerwinkle, R.A.G. and J.R.L. Human genome analysis S.Y., M.J., W.L.C., T.G., E.K., W.W., L.E.L.M.V., J.d.D., T.H., H.S., N.H., G.D.,

T.L., K.C., U.G., and M.F.W. Clinical data and segregation analysis: E.K., D.P., Y.P., M.S. and E. Battaloglu, Y.X., S.H.T. and R.A. S.P., G.M., R.D. Clark, C.J.C. and W.B.D. Fly experiments on oc mutant: M.J., and dAnkle2 mutant: M.J., N.L., W.L.C. Designed the study and wrote the manuscript: S.Y., M.J., J.R.L., M.F.W. and H.J.B. S.Y. and M.J. contributed equally.

REFERENCES

Asencio, C., Davidson, I.F., Santarella-Mellwig, R., Ly-Hartig, T.B., Mall, M., Wallenfang, M.R., Mattaj, I.W., and Gorjánácz, M. (2012). Coordination of kinase and phosphatase activities by Lem4 enables nuclear envelope reassembly during mitosis. Cell 150, 122–135.

Bainbridge, M.N., Wiszniewski, W., Murdock, D.R., Friedman, J., Gonzaga-Jauregui, C., Newsham, I., Reid, J.G., Fink, J.K., Morgan, M.B., Gingras, M.C., et al. (2011a). Whole-genome sequencing for optimized patient management. Sci. Transl. Med. 3, re3.

Bamshad, M.J., Shendure, J.A., Valle, D., Hamosh, A., Lupski, J.R., Gibbs, R.A., Boerwinkle, E., Lifton, R.P., Gerstein, M., Gunel, M., et al.; Centers for Mendelian Genomics (2012). The Centers for Mendelian Genomics: a new large-scale initiative to identify the genes underlying rare Mendelian conditions. Am. J. Med. Genet. A. 158A, 1523–1525.

Bayat, V., Thiffault, I., Jaiswal, M., Tétreault, M., Donti, T., Sasarman, F., Bernard, G., Demers-Lamarche, J., Dicaire, M.J., Mathieu, J., et al. (2012). Mutations in the mitochondrial methionyl-tRNA synthetase cause a neurodegenerative phenotype in flies and a recessive ataxia (ARSAL) in humans. PLoS Biol. 10, e1001288.

Bellen, H.J., Tong, C., and Tsuda, H. (2010). 100 years of Drosophila research and its impact on vertebrate neuroscience: a history lesson for the future. Nat. Rev. Neurosci. 11, 514–522.

Benos, P.V., Gatt, M.K., Murphy, L., Harris, D., Barrell, B., Ferraz, C., Vidal, S., Brun, C., Demaille, J., Cadieu, E., et al. (2001). From first base: the sequence of the tip of the X chromosome of Drosophila melanogaster, a comparison of two sequencing strategies. Genome Res. 11, 710–730.

Bier, E. (2005). Drosophila, the golden bug, emerges as a tool for human genetics. Nat. Rev. Genet. 6, 9–23.

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

Bökel, C. (2008). EMS screens : from mutagenesis to screening and mapping. Methods Mol. Biol. 420, 119–138.

Ceron, J., González, C., and Tejedor, F.J. (2001). Patterns of cell division and expression of asymmetric cell fate determinants in postembryonic neuroblast lineages of Drosophila. Dev. Biol. 230, 125–138.

Charng, W.L., Yamamoto, S., and Bellen, H.J. (2014). Shared mechanisms between Drosophila peripheral nervous system development and human neurodegenerative diseases. Curr. Opin. Neurobiol. 27C, 158–164.

Cook, R.K., Deal, M.E., Deal, J.A., Garton, R.D., Brown, C.A., Ward, M.E., Andrade, R.S., Spana, E.P., Kaufman, T.C., and Cook, K.R. (2010). A new resource for characterizing X-linked genes in Drosophila melanogaster: systematic coverage and subdivision of the X chromosome with nested, Y-linked duplications. Genetics 186, 1095–1109.

Cullen, C.F., Brittle, A.L., Ito, T., and Ohkura, H. (2005). The conserved kinase NHK-1 is essential for mitotic progression and unifying acentrosomal meiotic spindles in Drosophila melanogaster. J. Cell Biol. 171, 593–602.

Emoto, K. (2012). Signaling mechanisms that coordinate the development and maintenance of dendritic fields. Curr. Opin. Neurobiol. 22, 805–811.

Gonzaga-Jauregui, C., Lotze, T., Jamal, L., Penney, S., Campbell, I.M., Pehlivan, D., Hunter, J.V., Woodbury, S.L., Raymond, G., Adesina, A.M., et al. (2013). Mutations in VRK1 associated with complex motor and sensory axonal neuropathy plus microcephaly. JAMA Neurol. 70, 1491–1498.

Green, E.W., Fedele, G., Giorgini, F., and Kyriacou, C.P. (2014). A Drosophila RNAi collection is subject to dominant phenotypic effects. Nat. Methods 11, 222–223.

Haelterman, N., Jiang, L., Li, S., Bayat, V., Ugur, B., Tan, K.L., Zhang, K., Bei, D., Xiong, B., Charng, W.L., et al. (2014). Large-scale identification of chemically induced mutations in Drosophila melanogaster. Genome Res. 24, 1707–1718.

Hamosh, A., Sobreira, N., Hoover-Fong, J., Sutton, V.R., Boehm, C., Schiettecatte, F., and Valle, D. (2013). PhenoDB: a new web-based tool for the collection, storage, and analysis of phenotypic features. Hum. Mutat. 34, 566–571.

Huang, L., Xiao, X., Li, S., Jia, X., Wang, P., Guo, X., and Zhang, Q. (2012). CRX variants in cone-rod dystrophy and mutation overview. Biochem. Biophys. Res. Commun. 426, 498–503.

Jafar-Nejad, H., Andrews, H.K., Acar, M., Bayat, V., Wirtz-Peitz, F., Mehta, S.Q., Knoblich, J.A., and Bellen, H.J. (2005). Sec15, a component of the exocyst, promotes notch signaling during the asymmetric division of Drosophila sensory organ precursors. Dev. Cell 9, 351–363.

Jaiswal, M., Sandoval, H., Zhang, K., Bayat, V., and Bellen, H.J. (2012). Probing mechanisms that underlie human neurodegenerative diseases in Drosophila. Annu. Rev. Genet. 46, 371–396.

Kaindl, A.M., Passemard, S., Kumar, P., Kraemer, N., Issa, L., Zwirner, A., Gerard, B., Verloes, A., Mani, S., and Gressens, P. (2010). Many roads lead to primary autosomal recessive microcephaly. Prog. Neurobiol. 90, 363–383.

Karaca, E., Weitzer, S., Pehlivan, D., Shiraishi, H., Gogakos, T., Hanada, T., Jhangiani, S.N., Wiszniewski, W., Withers, M., Campbell, I.M., et al.; Baylor Hopkins Center for Mendelian Genomics (2014). Human CLP1 mutations alter tRNA biogenesis, affecting both peripheral and central nervous system function. Cell 157, 636–650.

Lehner, B. (2013). Genotype to phenotype: lessons from model organisms for human genetics. Nat. Rev. Genet. 14, 168–178.

Liao, T.S., Call, G.B., Guptan, P., Cespedes, A., Marshall, J., Yackle, K., Owusu-Ansah, E., Mandal, S., Fang, Q.A., Goodstein, G.L., et al. (2006). An efficient genetic screen in Drosophila to identify nuclear-encoded genes with mitochondrial function. Genetics 174, 525–533.

Liu, X., Jian, X., and Boerwinkle, E. (2011). dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. Hum. Mutat. 32, 894–899.

Lu, B., and Vogel, H. (2009). *Drosophila* models of neurodegenerative diseases. Annu. Rev. Pathol. *4*, 315–342.

Lupski, J.R., Belmont, J.W., Boerwinkle, E., and Gibbs, R.A. (2011). Clan genomics and the complex architecture of human disease. Cell *147*, 32–43.

Lupski, J.R., Gonzaga-Jauregui, C., Yang, Y., Bainbridge, M.N., Jhangiani, S., Buhay, C.J., Kovar, C.L., Wang, M., Hawes, A.C., Reid, J.G., et al. (2013). Exome sequencing resolves apparent incidental findings and reveals further complexity of SH3TC2 variant alleles causing Charcot-Marie-Tooth neuropathy. Genome Med *5*, 57.

Marygold, S.J., Leyland, P.C., Seal, R.L., Goodman, J.L., Thurmond, J., Strelets, V.B., and Wilson, R.J.; FlyBase consortium (2013). FlyBase: improvements to the bibliography. Nucleic Acids Res. *41* (Database issue), D751–D757.

McGary, K.L., Park, T.J., Woods, J.O., Cha, H.J., Wallingford, J.B., and Marcotte, E.M. (2010). Systematic discovery of nonobvious human disease models through orthologous phenotypes. Proc. Natl. Acad. Sci. USA *107*, 6544–6549.

Mohr, S.E. (2014). RNAi screening in *Drosophila* cells and in vivo. Methods *68*, 82–88.

Mummery-Widmer, J.L., Yamazaki, M., Stoeger, T., Novatchkova, M., Bhalerao, S., Chen, D., Dietzl, G., Dickson, B.J., and Knoblich, J.A. (2009). Genome-wide analysis of Notch signalling in *Drosophila* by transgenic RNAi. Nature *458*, 987–992.

Neely, G.G., Hess, A., Costigan, M., Keene, A.C., Goulas, S., Langeslag, M., Griffin, R.S., Belfer, I., Dai, F., Smith, S.B., et al. (2010). A genome-wide *Drosophila* screen for heat nociception identifies $\alpha 2\delta 3$ as an evolutionarily conserved pain gene. Cell *143*, 628–638.

Newsome, T.P., Asling, B., and Dickson, B.J. (2000). Analysis of *Drosophila* photoreceptor axon guidance in eye-specific mosaics. Development *127*, 851–860.

Nüsslein-Volhard, C., and Wieschaus, E. (1980). Mutations affecting segment number and polarity in *Drosophila*. Nature *287*, 795–801.

Oortveld, M.A., Keerthikumar, S., Oti, M., Nijhof, B., Fernandes, A.C., Kochinke, K., Castells-Nobau, A., van Engelen, E., Ellenkamp, T., Eshuis, L., et al. (2013). Human intellectual disability genes form conserved functional modules in *Drosophila*. PLoS Genet. *9*, e1003911.

Pastor-Pareja, J.C., and Xu, T. (2013). Dissecting social cell biology and tumors using *Drosophila* genetics. Annu. Rev. Genet. *47*, 51–74.

Rujano, M.A., Sanchez-Pulido, L., Pennetier, C., le Dez, G., and Basto, R. (2013). The microcephaly protein Asp regulates neuroepithelium morphogenesis by controlling the spatial distribution of myosin II. Nat. Cell Biol. *15*, 1294–1306.

Saj, A., Arziman, Z., Stempfle, D., van Belle, W., Sauder, U., Horn, T., Dürrenberger, M., Paro, R., Boutros, M., and Merdes, G. (2010). A combined ex vivo and in vivo RNAi screen for notch regulators in *Drosophila* reveals an extensive notch interaction network. Dev. Cell *18*, 862–876.

Saksena, S., and Emr, S.D. (2009). ESCRTs and human disease. Biochem. Soc. Trans. *37*, 167–172.

Schaffer, A.E., Eggens, V.R., Caglayan, A.O., Reuter, M.S., Scott, E., Coufal, N.G., Silhavy, J.L., Xue, Y., Kayserili, H., Yasuno, K., et al. (2014). CLP1 founder mutation links tRNA splicing and maturation to cerebellar development and neurodegeneration. Cell *157*, 651–663.

Vandendries, E.R., Johnson, D., and Reinke, R. (1996). orthodenticle is required for photoreceptor cell development in the *Drosophila* eye. Dev. Biol. *173*, 243–255.

Venken, K.J., He, Y., Hoskins, R.A., and Bellen, H.J. (2006). P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in D. melanogaster. Science *314*, 1747–1751.

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

Wang, T., and Montell, C. (2007). Phototransduction and retinal degeneration in *Drosophila*. Pflugers Arch. *454*, 821–847.

White, J.K., Gerdin, A.K., Karp, N.A., Ryder, E., Buljan, M., Bussell, J.N., Salisbury, J., Clare, S., Ingham, N.J., Podrini, C., et al.; Sanger Institute Mouse Genetics Project (2013). Genome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes. Cell *154*, 452–464.

Yamamoto, S., and Seto, E.S. (2014). Dopamine dynamics and signaling in *Drosophila*: an overview of genes, drugs and behavioral paradigms. Exp. Anim. *63*, 107–119.