Drosophila functional screening of *de novo* variants in autism uncovers deleterious variants and facilitates discovery of rare neurodevelopmental diseases

Paul C Marcogliese[1,2,25], Samantha L Deal[1,2,3,25], Jonathan Andrews[1,2,25], J Michael Harnish[1,2,3,25], V Hemanjani Bhavana[1,2], Hillary K Graves[1,2], Sharayu Jangam[1,2], Xi Luo[1,2,4], Ning Liu[1,2], Danqing Bei[1,2], Yu-Hsin Chao[1,2], Brooke Hull[1,2], Pei-Tseng Lee[1,2], Hongling Pan[1,2], Colleen M. Longley[2,3], Hsiao-Tuan Chao[1,2,5,6,7], Hyunglok Chung[1,2], Nele A Haelterman[1,2], Oguz Kanca[1,2], Sathiya N Manivannan[1,2], Linda Z Rossetti[1], Amanda Gerard[1,8], Eva Maria Christina Schwaibold[9], Renzo Guerrini[10], Annalisa Vetro[10], Eleina England[11], Chaya N Murali[1,8], Tahsin Stefan Barakat[12], Marieke F van Dooren[12], Martina Wilke[12], Marjon van Slegtenhorst[12], Gaetan Lesca[13], Isabelle Sabatier[14], Nicolas Chatron[13], Catherine A Brownstein[15,16], Jill A Madden[15,16], Pankaj B Agrawal[15,16,17,18], Roberto Keller[19], Lisa Pavinato[20,21], Alfredo Brusco[20,22], Jill A Rosenfeld[1,23], Ronit Marom[1,8], Michael F Wangler[1,2,5,24]*, Shinya Yamamoto[1,2,3,6,24]*

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

[2] Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital (TCH), Houston, TX, 77030, USA

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

[4] Department of Pediatrics, Division of Hematology/Oncology, BCM, Houston, TX, 77030, USA

[5] Department of Pediatrics, Division of Neurology and Developmental Neuroscience, BCM, Houston, TX, 77030, USA

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

[7] McNair Medical Institute, The Robert and Janice McNair Foundation, Houston, TX, 77030, USA

[8] TCH, Houston, TX, 77030, USA

[9] Institute of Human Genetics, Heidelberg University, Heidelberg, Germany

[10] Neuroscience Department, Children's Hospital Meyer-University of Florence, Florence, Italy

[11] The Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

[12] Department of Clinical Genetics, Erasmus MC University Medical Center, Rotterdam, The Netherlands

[13] Department of Medical Genetics, Lyon University Hospital, Université Claude Bernard Lyon 1, Lyon, France

[14] Department of Pediatric Neurology, Lyon University Hospitals, Lyon, France

[15] Division of Genetics and Genomics, Boston Children's Hospital, MA, 02115, USA

[16] The Manton Center for Orphan Disease Research, Boston Children's Hospital, MA, 02115, USA

[17] Division of Newborn Medicine, Boston Children's Hospital, 02115, MA USA

[18] Department of Pediatrics, Harvard Medical School, MA, 02115, USA

[19] Adult Autism Center, Mental Health Department, Health Unit ASL Città di Torino, Turin, Italy

[20] Department of Medical Sciences, University of Torino, Turin, Italy

[21] Institute of Human Genetics and Center for Molecular Medicine Cologne, University of Cologne, Cologne, Germany

[22] Medical Genetics Unit, Città della Salute e della Scienza, University Hospital, Turin, Italy

[23] Baylor Genetics Laboratories, Houston, TX, 77021, USA

[24] Development, Disease Models & Therapeutics Graduate Program, BCM, Houston, TX, 77030, USA

[25] These authors contributed equally to this study

*co-corresponding authors: M.F.W.: mw147467@bcm.edu (+1-832-824-8716), S.Y.: yamamoto@bcm.edu (+1-832-824-8119)

Keywords: autism spectrum disorders, functional genomics, missense variants, rare genetic diseases, undiagnosed diseases, *Drosophila melanogaster*

Abstract:

Individuals with autism spectrum disorders (ASD) exhibit an increased burden of *de novo* variants in a broadening range of genes. We functionally tested the effects of ASD missense variants using *Drosophila* through 'humanization' rescue and overexpression-based strategies. We studied 79 ASD variants in 74 genes identified in the Simons Simplex Collection and found 38% of them caused functional alterations. Moreover, we identified *GLRA2* as the cause of a spectrum of neurodevelopmental phenotypes beyond ASD in eight previously undiagnosed subjects. Functional characterization of variants in ASD candidate genes point to conserved neurobiological mechanisms and facilitates gene discovery for rare neurodevelopmental diseases.

Introduction:

Autism spectrum disorder (ASD) is a complex neurodevelopmental disorder with impairments in social interaction, communication and restricted interests or repetitive behaviors[1]. Individuals affected by ASD exhibit a higher burden of *de novo* mutations (DNMs), particularly in low-functioning cases, in an expanding list of genes[2–4]. The genetic burden of DNMs in ASD patients has been estimated to account for ~30% of disease causation[4–9]. While these studies have implicated hundreds of genes in ASD pathogenesis, which of these genes and variants causally contribute to this disease remains unknown. Missense DNMs in particular present a unique challenge because most genes lack established functional assays. *Drosophila melanogaster* is a genetically tractable system that is widely used to study human diseases[10–12]. Here, we integrate a number of state-of-the-art technologies in the fly field to establish an *in vivo* pipeline to effectively study the functional impact of DNMs identified in a large cohort of ASD patients.

Results

Prioritization of ASD variants to study in *Drosophila*

We prioritized genes with coding DNMs identified in ASD probands from the Simons Simplex Collection (SSC)[4] that were conserved in *Drosophila*. Because of the lack of a clear pathogenicity, we primarily focused on missense variants, and a few variants that were in-frame indels or truncating variants in single-exon genes (Fig.1, Supplementary Table 1). We then selected variants in genes that correspond to fly genes with intronic MiMIC elements, a versatile transposon that allows sophisticated genomic manipulations[13,14]. By converting the original MiMICs into T2A-GAL4 (TG4) lines via

recombinase-mediated cassette exchange[15–17], we generated a collection of fly lines that behaves as loss-of-function (LoF) alleles that simultaneously produce a GAL4 transactivator in the exact temporal and spatial pattern as the gene of interest[18]. Using this strategy, we successfully generated 109 TG4 lines that correspond to 128 SSC genes (some fly genes correspond to multiple human genes with variants in SSC). In parallel, we also generated 106 UAS human reference transgenic (Ref-Tg) and 88 SSC-DNM transgenic (SSC-Tg) strains (Fig. 1e-g, Supplementary Tables 2 and 3). These lines can be used in combination with TG4 lines to 'humanize' *Drosophila* genes, or can be crossed to ubiquitous or tissue-specific drivers to ectopically overexpress reference or variant human proteins[10] (Fig. 1h).

Humanization of essential *Drosophila* gene reveals loss-of-function ASD variants

We identified 47 lethal TG4 mutants that correspond to 60 human ASD candidate genes for which both reference and variant human cDNA transgenic fly lines were successfully established (Supplementary Fig. 1, Supplementary Table 4). To assess whether the human homolog can replace the corresponding fly genes, we determined whether UAS-Ref-Tg can rescue the lethality of lethal TG4 mutants. We assessed rescue at four temperatures (18°C, 22°C, 25°C, 29°C) as the GAL4/UAS system is temperature dependent[14]. We found that lethality was suppressed in 17 of 37 genes tested (46%; Fig. 2a, Supplementary Table 5). We next tested whether SSC-DNMs have functional consequences by comparing the rescue efficiency of UAS-Ref-Tg and UAS-SSC-Tg. We observed significant functional difference in the ability for SSC variants to rescue lethality for *ABL2*, *CAT*, *CHST2*, *TRPM6* (2 variants), and *TRIP12* (Fig. 2b-d). For *ABL2* and *CAT*,

we further found that humanized flies carrying the SSC-DNM had significantly decreased lifespan compared to reference animals (Fig. 2e, f). Overall, we found that 32% (6/19) of SSC-DNMs functionally differed from the reference *in vivo*, all behaving as LoF alleles.

To assess whether the fly homologs of human ASD candidate genes from SSC are expressed in the central nervous system (CNS), we crossed each TG4 line to UAS-nlsGFP (green fluorescent protein with a nuclear localization signal) and performed co-staining with neuronal (Elav) and glia (Repo) nuclear markers (Fig. 2g). All five genes associated with deleterious LoF DNMs were expressed in the adult brain (Fig. 2h) as well as in the developing (third instar larval) CNS (Supplementary Fig. 2). All five genes were found predominantly in neurons in adult brains, however, *Cat* was found to be enriched in glia in larval CNS.

Overexpression assays revealed loss-of-function ASD variants in genes corresponding to essential fly genes

We complemented our rescue-based assay by overexpressing Ref-Tg and SSC-Tg in a wild-type background. We found 18 human genes caused lethality in >70% of animals when overexpressed using a ubiquitous (*tub-GAL4*) driver. Of these, seven variants in seven genes (*GRK4*, *ITGA8*, *IRF2BPL*, *MINK1*, *NPFFR2*, *PDK2*, and *TSC2*) behaved as LoF alleles because they failed to reduce the expected viability to the extent of the corresponding reference allele (Fig. 3a, b). Expression of ten human genes caused morphological phenotypes in the wing when driven using a wing-specific (*nub-GAL4*) driver. Three variants in two genes (*IRF2BPL* and *ATP2B2*) were identified as LoF alleles based on this driver and assay. Overexpression of reference *IRF2BPL* in the developing

wing caused lethality, whereas neither the *IRF2BPL*^{F30L} missense nor the *IRF2BPL*^{N701Tfs66*} frameshift (note that *IRF2BPL* is a single-exon gene) variants cause a phenotype (Fig. 3c). Ectopic expression of *ATP2B2* in the developing wing disc causes a curled wing phenotype while the *ATP2B2*^{T818M} variant fails to do so (Fig. 3d). Expression of four human genes caused morphological phenotypes in the eye when driven using an eye-specific (*GMR-GAL4*) driver, but none of the corresponding SSC-DNMs altered these defects (Fig. 3a, Supplementary Table 5). Of 37 Ref-Tgs tested, 18 (corresponding to 21 out of 42 SSC-Tgs tested) failed to produce scorable phenotypes, and none of the SSC-Tgs behaved as Gain-of-Function (GoF) alleles. Therefore, we conclude that 43% (9/21) of the SSC-Tgs tested with an overexpression strategy behaved as LoF alleles.

Imaging and cell-type expression analysis revealed that all eight fly genes that correspond to LoF SSC-DNMs are expressed in the adult brain and larval CNS [see [19] for *pits* (*IRF2BPL*) expression that is expressed widely in most neurons]. *msn* (*MINK1*, *TNIK*) is enriched in glia. *if* (*ITGA8*) is not detected in either neurons or glia but revealed a pattern reminiscent of cells in *pars intercerebralis*, a neuroendocrine organ analogous to the mammalian hypothalamus[20] (Fig. 3e arrow, Supplementary Fig. 3).

Loss- and Gain-of-Function ASD variants identified through behavior analysis on humanized flies for viable TG4 lines

While we were able to test the function of 37 human genes based on rescue of lethality and overexpression phenotypes (Figs. 2 and 3), 62 TG4 lines corresponding to 70 SSC-ASD candidate genes were viable and did not exhibit any obvious morphological phenotypes. To determine if any of these viable mutants display a behavioral phenotype

that can be used for variant function assessment, we performed courtship assays[21,22] (Fig. 4a). Courtship involves a complex set of neurological components involving sensory input, processing and motor output[22]. We measured the amount of time a male fly spent performing wing extensions (courtship) and copulating. We also measured the time flies spent moving within the test chamber to assess their locomotion. Finally, we also tracked grooming, a stereotypic behavior in flies that involves a complex neurocircuit[23]. Of 21 viable TG4 lines analyzed, we found that 15 display behavioral alterations (Fig. 4c-f and Supplementary Fig. 4a-d).

Of 15 viable TG4 lines with quantitative behavior defects, we were able to humanize eight of them using Ref-Tgs and SSC-Tg to assess variant function. We found five SSC-DNMs that showed functional alterations from the reference allele in at least one of four behavioral paradigms (Fig. 4b). Two variants (*GLRA2*^{N136S} and *KCND3*^{R86P}) behaved as LoF alleles. Humanized reference *GLRA2* flies failed to copulate at all, while the humanized *GLRA2*^{N136S} flies were capable of copulating within the trial period similar to the TG4 mutant alone (Fig. 4d). Humanized *KCND3*^{R86P} flies displayed increased movement and decreased grooming behavior when compared to the reference flies (Fig. 4e, f). Interestingly, three variants (*KDM2A*^{R449K}, *ALDH1L1*^{N900H} and *USP30*^{P200S}) behaved as GoF alleles (e.g. hypermorph, antimorphic, neomorph). Humanized *KDM2A* reference flies had a trend for increased time spent copulating compared to TG4 mutant alone, however *KDM2A*^{R449K} flies showed significant increase in time spend copulating (Fig. 4d). Humanized *ALDH1L1*^{N900H} fly displayed a significant reduction in courtship and an increase in grooming behavior when compared to the humanized reference fly or the

TG4 mutant alone (Fig. 4c, f). Humanized *USP30*^{P200S} variant flies displayed decreased grooming behavior when compared to humanized reference flies (Fig. 4f).

Finally, we determined the CNS expression of TG4 lines corresponding to all 8 lines we were able to humanize. Surprisingly, we only detected expression of 4/8 genes in the larval and adult CNS (Supplementary Fig. 4e, Fig. 4g). In sum, over half (15/21) of the viable TG4 mutants assayed show some behavioral alteration compared to a commonly used control (*Canton-S)* strain (shown in red lines in Fig. 4c-f and Supplementary Fig. 4a-d), and 63% (5/8) SSC-DNMs act functionally different from reference using quantitative behavioral measurements in flies: two behaved as LoF alleles whereas three behaved as GoF alleles. This is in contrast to the humanization-based functional studies performed on essential gene homologs that only revealed LoF alleles.

Overexpression assays revealed ASD variants with diverse functional consequences in genes that correspond to viable TG4 lines

Next, to complement the functional studies based on rescue-based experiments performed on viable TG4 mutants (Fig. 4), we overexpressed Ref-Tg and SSC-Tgs in a wild-type background as we did for transgenes corresponding to essential fly genes (Fig. 3). We found that overexpression of 12 variants displayed functional differences whereas five variants acted in a similar fashion to the reference transgene (Fig. 5a). Another 12 reference and their respective DNMs failed to produce any phenotype in the paradigms we tested.

Three variants (*GLRA2*^{N136S}, *KCND3*^{R86P} and *EPHA1*^{V567I}) behaved as LoF alleles when assessed with one or multiple drivers. *GLRA2*^{N136S} and *KCND3*^{R86P} abolished the activity of the reference transgenes to cause lethality when overexpressed with a ubiquitous driver (Fig. 5b). In the wing, *KCND3*^{R86P} also failed to produce a severe notching phenotype that is observed by expression of reference transgene (Fig. 5e). Also in the wing, expression of reference *EPHA1* caused a wing size reduction and wing margin serration, yet the *EPHA1*^{V567I} variant still caused serrated wings but wings were normal in size (Fig. 5e).

Seven variants (*ACE*^{Y818C}, *GPC5*^{M133T}, *MYH9*^{R1571Q}, *PC*^{P1024R}, *SLC23A1*^{L465M}, *HTR1D*^{T99N}, *BAIAP2L1*^{A481V}) behaved as GoF alleles. Flies overexpressing mutant forms of *ACE*, *GPC5*, *MYH9*, *PC*, and *SLC23A1* exhibited enhanced lethality when compared to reference protein (Fig. 5c). GoF nature of *MYH9*^{R1571Q} was also seen based on a wing size-based assay (Fig. 5e). *HTR1D*^{T99N} displayed consistent stronger phenotypes compared to reference when expressed in the eye or the wing, resulting in eye size reduction and absent wing phenotype, respectively (Fig. 5d, e). *BAIAP2L1*^{A481V} caused a smaller, more crumpled wing phenotype compared to its reference allele (Fig. 5e).

Intriguingly, *EPHB1*^{V916M} and *MAP4K1*^{M725T} exhibited conflicting results in eye and wing, thus could not be categorized as a simple LoF or GoF variant. While overexpression of reference *EPHB1* or *MAP4K1* in the eye causes eye size reduction phenotype, SSC variant forms of either gene result in normal eyes, indicating they behave as LoF alleles in this tissue. However, the same variant transgenes for these two genes expressed in the wing result in blistered or crumpled wings, respectively, that are phenotypically

stronger than the reference alleles (Fig. 5d, e), indicating they behave as GoF alleles in this tissue.

Imaging analysis revealed that 7/11 corresponding fly genes (one fly gene *Eph* corresponds to both *EPHA1* and *EPHB1*) were expressed in the larval and adult CNS. Most genes are enriched in neuronal subpopulations, with the exception of *Eph*, which is enriched in glia in the larval CNS (Fig. 4g, h and 5f, Supplementary Fig. 5). In summary, 71% (12/17) of SSC-DNMs impact gene function: three acted as LoF, seven acted as GoF, and two gave complex results. Similar to results obtained from humanization experiments in corresponding viable TG4 lines (Fig. 4), this is in contrast to functional studies performed on essential gene homologs that only revealed LoF alleles (Fig. 3).

Thirty deleterious SSC variants identified by merging all functional data

Considering all data from both rescue-based and overexpression strategies of human genes corresponding to homologs of both essential and non-essential fly genes in our study, we found 29 missense and 1 frameshift SSC-DNMs displaying functional differences when compared to their respective reference allele (1 variant for 26 genes, 2 variants for 2 genes) (Table 1, Supplementary Table 5). Approximately 53% (30/57) of SSC-DNMs tested exhibited functional differences compared to the reference. Intriguingly in our study, we only found GoF variants for genes corresponding to viable TG4 fly mutants (Supplementary Fig. 6a). Our dual approach (rescue-based combined with overexpression-based studies) was complimentary as *GRK4*, *NPFFR2*, and *PDK2* SSC-DNMs were found as LoF variants when overexpressed ubiquitously, yet these variants

were able to rescue lethality in a similar manner to their respective reference alleles (Fig. 3, Supplementary Table 5).

When we informatically surveyed the genes and variants with functional consequences identified through our screen in comparison to other genes included in our study (variants with comparable function or those lacking a phenotype to assess) using the MARRVEL tool[24], we did not find any significant differences in gene level metrics such as pLI (probability of Loss of function Intolerance)[25], LOEUF [Loss of function observed/expected (o/e) upper bound fraction], missense o/e (observed/expected)[26], pathogenicity prediction scores based on several *in silico* algorithms including SIFT[27], PolyPhen-2[28], and CADD[29], or absence/presence of identical variant in gnomAD[25] (Supplementary Tables 5 and 6, Supplementary Fig. 6b-af). By analyzing Gene Ontology (GO) by PantherDB[30], ASD candidate genes from SSC with deleterious variants *in vivo* were compared to all protein coding genes. We found significant enrichment for genes with GO terms for 'synapse (cellular component)' and 'ATP binding (function)' (Supplementary Fig. 7). Finally, we systematically imaged the expression pattern of all TG4 lines generated through in our study to document their expression in the adult and larval CNS in neurons or glia as a resource for the community (Supplementary Figs. 8-9).

Rare LoF and GoF variants in *GLRA2* cause X-linked neurodevelopmental disorders with overlapping phenotypes in males and females, respectively

Many genes implicated in ASD are associated with neurodevelopmental disorders beyond autism[31,32]. Therefore we asked if genes with disruptive SSC-DNMs could be used to discover new disease-causing genes and disorders by identifying patients with

rare potentially deleterious variants that have been unrecognized[33–35]. Out of 28 genes in which we identified damaging SSC-DNMs, eight are associated with Mendelian diseases with neurological presentations in OMIM[36] (Table 1). For one of these genes (*IRF2BPL*), we recently reported *de novo* truncating variants in *IRF2BPL* as the cause of a novel severe neurodevelopmental disorder with abnormal movements, loss of speech and seizures in collaboration with the Undiagnosed Diseases Network[19,35]. Here, we report identification of *GLRA2* as a cause of novel neurodevelopmental syndromes with overlapping features such as developmental delay (DD), intellectual disability (ID), ASD and epilepsy identified through re-analysis of clinical exome sequencing data and GeneMatcher[33] (Figure 1h).

We identified rare variants in *GLRA2* in eight unrelated subjects with or without autistic features. In addition to developmental and cognitive delay of variable severity, 3/8 subjects have microcephaly, 4/8 subjects have a history of epilepsy and 6/8 subjects have ocular manifestations, including congenital nystagmus that improved with age in 3 of them (Table 2). *GLRA2* (*Glycine Receptor Alpha 2*) is an X-linked gene that encodes a subunit of a glycine-gated chloride channel[37]. All five female subjects harbored DNMs including a recurring $GLRA2^{T296M}$ variant found in 4/5 female subjects. This variant was also identified in a female subject in previous large-scale developmental disorder study[38]. The three male subjects had maternally inherited variants, and the mothers are not symptomatic. The CADD scores for all three male subjects are predicted to be damaging. A maternally-inherited microdeletion of *GLRA2* has also been found in a single male patient with ASD[39], indicating that hemizygous LoF allele of this gene in males may cause ASD. Indeed, we determined that the $GLRA2^{N136S}$ variant present in the SSC in a male

subject acts as a LoF allele based on our behavioral assay on humanized flies (Fig. 4d) as well as through overexpression studies (Fig. 5b).

To further understand the functional consequences of variants found in our *GLRA2* cohort, we generated additional transgenic flies to assay the function of p.T296M (found in female subjects) and p.R252C (found in a male subject) variants (Fig. 6a, Supplementary Fig. 10a). By overexpressing variant GLRA2 using a ubiquitous driver, we found that *GLRA2*^{R252C} behaves as a LoF allele (Fig. 6b), similar to *GLRA2*^{N136S} (Fig. 5). In contrast, this assay did not distinguish *GLRA2*^{T296M} from the reference (Fig. 6b). Given the recurrent nature of this variant, as well as structural prediction that the residue has a potential role in obstruction of the ion pore in the closed conformation (Supplementary Fig. 10g)[40,41], we further tested this allele using additional GAL4 drivers. Using a *pnr-GAL4* that is expressed in the dorsocentral stripe in the notum, we found that *GLRA2*^{T296M}, but not the reference or other *GLRA2* variant tested, causes lethality when expressed at high levels (Supplementary Fig. 10b). When we expressed *GLRA2*^{T296M} at lower levels by manipulating the temperature, we found that this variant induces the formation of melanized nodules in the thorax, a phenotype that we never observed by expressing the reference and other *GLRA2* variants assessed (Fig. 6c, d, Supplementary Fig. 10b).

To further examine the functional consequences of overexpression of reference and variant *GLRA2* in the nervous system, we performed electroretinogram (ERG) recordings on the fly eye expressing human *GLRA2* using two distinct drivers. Pan-neuronal driver (*nSyb-GAL4*)[42] allows one to express *GLRA2* cDNA in both pre-synaptic photoreceptors and post-synaptic neurons in the nervous system. Using this driver, we found a significant increase in amplitudes of "*OFF*" transients with *GLRA2*^{T296M} (Fig. 6e,

f, Supplementary Fig. 10c, d), indicating an increase in synaptic transmission[43], supporting the finding in the notum that p.T296M behaves as a GoF allele. Interestingly, when we limited the expression of *GLRA2* to the pre-synaptic photoreceptors using *Rh1-GAL4*[44], we did not observe any functional difference between $GLRA2^{T296M}$ and the reference allele. However, with this driver, we were able to discern that both $GLRA2^{R252C}$ seen in Subject 6 and $GLRA2^{N136S}$ found in an SSC subject behave as LoF alleles based on observing a decrease in amplitude of *"OFF"* transients, indicating a decrease in synaptic transmission (Fig. 6 g, h, Supplementary Fig. 10e, f). Hence, we have identified a cohort of subjects with deleterious variants in an X-linked gene *GLRA2* and shown that a recurrent missense DNM in females acts as a GoF allele, whereas alleles in male subjects behave as LoF alleles.

Discussion

In this study, we generated >300 fly strains that allow functional studies of human variants and homologous fly genes *in vivo*. These reagents can be used to study many coding variants that are being identified through next-generation sequencing efforts in human genomics in diverse disease cohorts beyond ASD. Our screen elucidated 30 SSC variants with functional differences compared to reference, which was over half (~53%) of the genes in which we were able to perform a comparative functional assay.

Our screen was part of a larger effort to characterize the functional consequences of missense *de novo* changes from the SSC dataset using different strategies. One approach was based on a proteomics by performing yeast-two-hybrid assays on 109 SSC-DNMs found in patients, showing 20% protein-protein interactions that are found in

reference proteins are disrupted by variants[45]. Another group reported that ~70% of 37 SSC-DNMs knocked-in to homologous *C. elegans* genes caused scorable phenotypes[46]. These studies are complementary to each other because while some variants have been identified as deleterious by more than one approach (e.g. *GLRA2*^{N136S} identified in both worm and fly screens), others are uniquely identified in one study, some of which could be due to technical limitations. For example, our approach of utilizing human cDNA transgenes allowed us to test variant function regardless of residue conservation in *Drosophila*, which were not tested in *C. elegans* due to lack of conservation. Of the 29 disruptive missense SSC-DNMs identified through our study, 14 affected residues that were conserved in flies and 10 in worms.

To take a rather unbiased approach, our gene prioritization was based on gene-level conservation and tool availability (e.g. intronic MiMIC lines, full length human cDNA) rather than based on gene level constrains and variant level prediction scores. Hence, our study subset, although somewhat limited, can be considered as a random sample of ASD-implicated genes and variants. Interestingly, we could not find any significant difference in pathogenicity prediction for disruptive variants *in vivo*. Out of the 29 missense SSC-DNMs that had functional consequences in our assays, four were not predicted as damaging variants (CADD<10), nine had moderate scores (CADD: 10-20), and 16 were predicted to be disruptive (CADD>20). Understanding how variants that are not predicted to be damaging based on state-of-the-art informatics programs impact protein function may provide guidance to improve the accuracy of *in silico* tools.

To study functional consequences of SSC-DNMs, we took two conceptually different approaches (humanization-based rescue strategy and overexpression-based

strategy). Indeed, the two approaches were complementary as only two variants (*GLRA2*^{N136S} and *KCND3*^{R86P}) were detected in both screens, showing consistent LoF effects using both approaches. Two disruptive SSC-DNMs, *EPHB1*^{V916M} *and MAP4K1*^{M725T}, behaved as complex alleles, displaying discordant phenotypes in the eye and wing. This suggests these variants may behave in a context-dependent manner, acting as a GoF allele in one tissue while behaving as a LoF allele in another. Thus, one functional assay may not be enough to reveal the full nature of pathogenic mechanisms, and some disease-associated variants may act differently in different tissues or cell types.

Starting from a single *de novo* hemizygous missense variant that we identified as a LoF allele in *GLRA2* (p.N136S), we identified a cohort with overlapping neurodevelopmental phenotypes carrying LoF or GoF variants in this gene. Interestingly, female subjects harbored DNMs and male subjects carried maternally inherited missense variants in this X-linked gene which undergoes random X-inactivation in females but not males[47]. The X-linked status of *GLRA2* may mean that variants causing reduced *GLRA2* activity lead to disease in males but can be tolerated in heterozygous females. This is evident from non-symptomatic mothers of male patients who had maternally inherited alleles (Subjects 6-8). In contrast, GoF mutations in this channel could be overrepresented in females since hyperactivation of this channel may cause neurological defects[48]. While the exact mechanism of how the p.T296M variant affects GLRA2 function remains unclear, the presence of melanized nodules in flies expressing this variant are indicative of an innate immune response[49], potentially a result of leaky ion channel function[50]. Fittingly, our structural data revealed that the p.T296M is adjacent to a critical

residue that is likely to be important for keeping the ion pore in a conformationally closed state (Supplementary Fig. 10g).

In summary, we have utilized a model organism-based *in vivo* functional genomics approach to study the functional consequence of rare genetic events in a common neurological disorder, ASD. In addition to garnering variant functional data for ASD subjects in the SSC, we leveraged this information to identify and document a novel rare neurological condition. Such bi-directional communication and collaboration between bench scientists and clinicians greatly facilitate the functional studies of human variants found in common diseases such as ASD, and can also lead to novel discoveries that have impact on rare disease research.

Acknowledgments

We express our appreciation to the patients and families for their participation in this study. We thank Pradnya Bhadane, Nora Duran, Mark Durham, Shelley Gibson, Yuchun He, Mei-Chu Huang, Matthew Lagarde, Wen-Wen Lin, and Karen Schulze for technical or administrative assistance. We thank Hugo Bellen for insightful scientific discussions valuable comments on this manuscript. We sincerely thank the late Kenneth Scott for access to many of the human cDNAs used in this study. This work was primarily supported by a Simons Foundation Autism Research Initiative Functional Screen Award (368479) to M.F.W. and S.Y. Generation of human cDNA transgenic lines were in part supported by NIH/ORIP (R24OD022005). Confocal microscopy at BCM is supported in part by NIH/NICHD (U54HD083092) to the Intellectual and Developmental Disabilities Research Center (IDDRC) Neurovisualization Core. P.C.M. is supported by CIHR (MFE-

164712) and the Stand by Eli Foundation. J.A. is supported by NIH/NINDS (F32NS110174). R.M. is supported by NIH/NIGMS (T32GM07526-43), and through BCM Chao physician-scientist award. C.M.L. is part of the BCM Medical Scientist Training Program and McNair MD/PhD Student Scholars supported by the McNair Medical Institute at the Robert and Janice McNair Foundation and funded by NIH F30 Award (F30MH118804). H.T.C. received funding support by the American Academy of Neurology and CNCDP-K12. T.S.B. is supported by the Netherlands Organization for Scientific Research (ZonMW Veni, Grant 91617021), an Erasmus MC Fellowship 2017 and Erasmus MC Human Disease Model Award 2018. The DECODE-EE project (Health Research Call 2018, Tuscany Region) provided research funding to R.G. and A.V. A.B. and L.P. received funding specifically appointed to Department of Medical Sciences from the Italian Ministry for Education, University and Research (Ministero dell'Istruzione, dell'Università e della Ricerca—MIUR) under the programme 'Dipartimenti di Eccellenza 2018–2022' Project code D15D18000410001.

Author Contributions

M.F.W. and S.Y. conceived, designed the project. P.C.M., J.A., S.L.D., and J.M.H. designed and conducted most fly experiments and analyzed the data. V.H.B. and Y.C. performed cloning and mutagenesis. H.K.G. performed cloning and coordination. S.J. performed immunostaining and confocal microscopy. X.L. performed structural analysis and generated reagents. N.L. D.B., Y.C., P.L., B.H., H.P., C.B., H.T.C., H.C., N.A.H., O.K., and S.N.M contributed to reagent generation and some fly experiments. R.M., A.G., E.M.C.S., R.G., A.V., C.N.M., T.S.B., M.R.V., M.W., M.V., G.L., I.S., N.C., J.A.M., P.B.A.

R.K., L.P., and A.B. reported and described *GLRA2* subjects. L.Z.R. and J.A.R. aided in subject match-making. P.C.M., J.A., S.L.D., J.M.H., R.M., M.F.W. and S.Y. wrote the manuscript.

Declaration of Interests

The Department of Molecular and Human Genetics at Baylor College of Medicine receives revenue from clinical genetic testing completed at Baylor Genetics Laboratories.

Materials & Methods

Additional gene/variant prioritization details:

We prioritized genes and coding de novo missense mutations (DNMs) identified in the Simons Simplex Collection (SSC) from the initial study using exome sequencing[4]. In this cohort, 1,708 ASD proband-specific *de novo* missense or in-frame indels were identified through WES (Fig. 1a), corresponding to 1,519 unique human genes. Of these, 920 fly genes corresponding to 1,032 human genes were identified. 487 human genes had no or weak ortholog candidates in *Drosophila* based on multiple ortholog prediction algorithms scores (cut off: DIOPT < 4/16)[51] (DIOPT Version 8 accessed January 2020) (Supplementary Table 1). By overlapping these 920 *Drosophila* genes with available fly lines containing MiMIC transposons within coding introns that permit targeting of all annotated protein isoforms (1,732 insertions)[13,14], we identified reagents for 122 fly genes corresponding to 143 human genes and 179 ASD proband variants from the SSC.

Of the 122 fly genes we selected to work on, we were able to successfully generate 109 TG4 lines through RMCE of MiMIC elements through genetic crosses (see below for

specific methods)[15]. To generate UAS human reference transgenic (Ref-Tg) and human SSC candidate variant transgenic (SSC-Tg), we obtained human ORF (open reading frame) collections from the Mammalian Gene Collection[52] or commercial sources. We generated SSC-Tg cDNAs by site-directed mutagenesis protocols (see below for specific methods). The reference and mutagenized ORF were shuttled into to the pUASg.HA-attB destination vector (Fig. 1f)[53]. Transgenes were integrated into a precise location in the fly genome using ϕC31 transgenesis technology. Out of the 143 human genes and 179 SSC-DNMs of interest that we attempted to generate, we were successful in generating 194 UAS-cDNA (106 Ref-Tgs; 88 SSC-Tgs) flies (Fig. 1g, Supplementary Tables 2 and 3).

We were able to make a complete set of TG4, UAS-Ref-Tgs and UAS-SSC-Tgs lines for 65 fly genes corresponding to 74 human genes and 79 variants (some fly genes correspond to multiple human genes, and multiple SSC variants are found for a small subset of human genes), which were critical to test the variant function using a rescue-based humanization strategy. For an additional 42 human genes we generated 42 TG4 mutants. For 32 human genes we generated Ref-Tgs and for 9 genes we generated SSC-Tgs. In summary, 303 Drosophila stocks were generated for this project as a resource for the community, and these stocks are available from the Bloomington *Drosophila* Stock Center or in the process of being transferred and registered at BDSC.

Generation of TG4 lines:

All *TG4* alleles in this study were generated by ϕC31-mediated recombination-mediated cassette exchange of MiMIC (Minos mediated integration cassette) insertion lines[13,14,17]. Conversion of the original MiMIC element was performed via genetic by crossing *UAS-*

2xEGFP, hs-Cre,vas-dϕC31, Trojan T2A-GAL4 triplet flies to each MiMIC strain and following a crossing scheme[15]. 73 TG4 lines were described previously but not extensively characterized[16], while 35 lines were generated specifically for this study.

Generation of UAS-human cDNA lines:

The majority of reference human cDNA clones were obtained in either pDONR221 or pDONR223 donor vectors. The LR clonase II (ThermoFisher) enzyme was used to shuttle ORFs into the p.UASg-HA.attB destination vector via Gateway™ cloning. Some ORFs that were not Gateway compatible were obtained from additional sources (Supplementary Table 2), amplified with flanking *attB* sites and cloned into pDONR223 plasmid using BP clonase II (ThermoFisher). Sequence-verified variants were generated in the DONR vectors by either site-directed mutagenesis (SDM) via or High-Throughput Mutagenesis (HiTM) as previously described[54]. SDM was performed with primers generated using NEBaseChanger (see Supplementary Table 3) with the Q5® mutagenesis kit (NEB). Sequence-verified reference and variant ORFs in the pUASg-HA.attB destination plasmid were microinjected into ~200 embryos in one three *attP* docking sites (*attP86Fb*, *VK00037* or *VK00033*) docking sites by ϕC31 mediated transgenesis[53,55]. The docking site of choice were selected based on the genomic locus of the corresponding fly gene. In principal, VK00037 docking site on the 2nd chromosome was used for human genes that correspond to fly genes on the X, 3rd or 4th chromosome, whereas VK00033 or attP86Fb docking site on the 3rd chromosome was used for human genes that correspond to fly genes on the 2nd chromosome.

Fly husbandry:

Unless otherwise noted, all flies used in experiments were grown in a temperature and humidity-controlled incubator at 25°C and 50% humidity on a 12-hour light/dark cycle. Some experiments were conducted at different temperatures that are specifically indicated in the text and figures. Stocks were reared on standard fly food (water, yeast, soy flour, cornmeal, agar, corn syrup, and propionic acid) at room temperature (~22°C) and routinely maintained.

Fly stocks used that were not generated here:

tub-GAL4 (*y^1 w^*; P{w[+mC]=tubP-GAL4}LL7/TM3, Sb^1 Ser^1*) BDSC_5138, *GMR-GAL4* (*w^*; P{w[+mC]=GAL4-ninaE.GMR}12*) BDSC_1104, *nub-GAL4* (*P{GawB}nubbin-AC-62*)[56], *nSyb-GAL4* (*y^1 w^*; P{nSyb-GAL4.S}3*) BDSC_51635, *Rh1-GAL4* (*P{ry[+t7.2]=rh1-GAL4}3, ry[506])* BDSC_8691, *pnr-GAL4* (*y^1 w^{1118}; P{w[+mW.hs]=GawB}pnr[MD237]/TM3, P{w[+mC]=UAS-y.C}MC2, Ser^1*) BDSC_3039, *UAS-LacZ* (*w^*; P{w[+mC]=UAS-lacZ.Exel}2*) BDSC _8529, UAS-nlsGFP (*w^{1118}; P{w[+mC]=UAS-GFP.nls}14*) BDSC_4775.

Electroretinograms (ERG):

ERG recordings on adult flies were performed on *nSyb-GAL4*[42] and *Rh1-GAL4*[44] driven *UAS-GLRA2* at 5 days post-eclosion raised at 25°C in 12h light/12h dark cycle as previously described[57] using LabChart software (AD instruments). 4-10 flies were examined for each genotype. Recording was repeated at least 3 times per fly. Quantification and statistical analysis was performed using ANOVA followed by Bonferroni's multiple comparison test using Prism 8.0.

Complementation test of lethality in TG4 lines:

Out of the 109 TG4 mutants generated, 64 TG4 mutants were homozygous lethal. Because lethality can be caused by disruption of the gene of interest or due to second site lethal mutations carried on the same chromosome, we performed complementation test using standard methodology. For genes on the 2nd and 3rd chromosome, female heterozygous TG4 lines balanced with either *SM6a* or *TM3, Sb, Ser*, respectively, were crossed with male flies carrying a corresponding deficiency (Df) that covers the gene of interest (see Supplementary table 4). Three independent crosses were set at 25°C for each TG4 line and we determined if any *TG4* flies survived to the adult stage *in trans* with their corresponding Df (*TG4/Df*). If viable, a second Df line covering the same gene was used to validate this finding to make sure the complementation is not due to some problematic Df lines. If TG4 was viable over two independent Df lines, we ascribed the lethality to a second site mutation on the *TG4* chromosome. *TG4* that remained lethal in trans with a Df line are be considered to be disrupting an essential gene in flies. For five genes on the X-chromosome of the fly, complementation was performed by first rescuing hemizygous TG4 males with a duplication (Dp) line obtained from BDSC (see Supplementary table 4), and crossing these rescued flies to female TG4/FM7 flies. If TG4/Y; Dp/+ lines were viable, we ascribed the lethality of TG4 to the gene of interest. All Df and Dp lines were obtained from BDSC, and the specific stock used in our analysis are listed in Supplementary Table 4.

Through this experiment, we found 64 *TG4* mutant lines that were homozygous lethal, and 47 remained lethal when *in trans* with a corresponding deficiency line (Supplementary Fig. 1a, Supplementary Table 4). The 47 essential genes in *D. mel* corresponded to 60 SSC related human genes (Supplementary Fig. 1b). The lethality of 17 *TG4* lines

corresponding to 18 human genes were due to a second site lethal mutation, potentially present in the original MiMIC line, or introduced during RMCE which has been reported previously[14]. These TG4 lines together with viable TG4 lines are likely associated with non-essential genes in *Drosophila*.

Rescue of lethality in TG4 lines by UAS-human cDNA transgenes:

In order to assess the ability of human reference or SSC variant cDNAs to rescue lethality observed in *TG4* mutants in essential genes, we first double balanced all Df lines that fail to complement a lethal TG4 line with UAS-reference or variant cDNA lines. For genes on the 2nd chromosome, we generated *Df/CyO; UAS-cDNA/(TM3, Sb, Ser)* stocks. For genes on the 3rd chromosome, we generated *UAS-cDNA/(CyO); Df/TM3, Sb, Ser.* Heterozygous *TG4/Balancer* females were crossed to double balanced Df/Balancer, UAS-human cDNA males at multiple temperatures (18°C, 22°C, 25°C, 29°C) to determine rescue of lethality to adult stage. A minimum of two independent crosses were conducted at each temperature. For the five genes on the X-chromosome of the fly, we attempted rescue by crossing female *TG4/FM7* flies to UAS-cDNA/(SM6a) males to generate hemizygous TG4 males that expresses human cDNA (TG4/Y; UAS-cDNA/+) to test their viability. Statistical analysis was performed using Two-way ANOVA followed by Sidak's multiple comparison test across temperature and genotype.

Lifespan assays:

For *Drosophila* lifespan, newly eclosed flies were separated by genotype and sex and incubated at 25°C. Flies were transferred into a fresh vial every two days and survival was determined once a day. 11-49 flies were tested per group. Statistical analysis was performed using Log-rank (Mantel-Cox) test.

Behavioral assays:

Of 48 *TG4* mutants that were viable when *in trans* with a corresponding deficiency, 17 lines exhibited lethality in homozygous states, indicating the presence of a second site lethal mutation. Out of 31 *TG4* mutants that were homozygous viable, we prioritized to study 21 *TG4* mutants based on reagent availability. Courtship assay was performed as previously described[22]. *TG4* lines were backcrossed to *Canton-S* strain in order to eliminate known courtship deficiencies present in the y^1 w^* background, which all TG4 lines are initially generated on. Collection of socially naïve adults was performed by isolating pupae in 16 x 100 polystyrene vials containing approximately 1 ml of fly food. After eclosion, flies were anesthetized briefly with CO_2 to ensure they were healthy and lacking wing damage. Anesthetized flies were returned to their vials and allowed 24 hours to recover before testing. Courtship assays were performed in a 6 well acrylic plate with 40mm circular wells, with a depth of 3mm and a slope of 11 degrees, as per the chamber design in Simon et al, 2010[58]. One Canton-S virgin female (6-10 days post-eclosion), and one *TG4* mutant male fly (3-5 days post-eclosion) with or without UAS-human cDNAs were simultaneously introduced into the chamber via aspiration. Recordings were taken using a Basler 1920UM, 1.9MP, 165FPS, USB3 Monochromatic camera using the BASLER Pylon module, with an adjusted capturer rate of 33 fps (frames per second). Conversion of captured images into a movie file was performed via a custom MatLab script, and tracking of flies in the movie was performed using the Caltech Flytracker[59]. Machine learning assessment of courtship was performed using JAABA[60] using classifiers that scored at 95% or higher accuracy during ground-truthing trials. At least 10 animals were tested per genotype. Analysis of data was performed using Excel (Microsoft)

and Prism (GraphPad). A ROUT (Q=1%) test was performed in Prism to identify outliers. Determination of significance in behavior tests was performed using the Kruskal-Wallis one-way analysis of variance and the Dunn's multiple comparison test. P-values of 0.05 or less were considered significant.

Overexpression assays to assess lethality and morphological phenotypes:

To detect any differences in the phenotypes induced by overexpression of reference and variant human cDNA in order to assess variant function, we crossed UAS-human cDNAs with reference or variant alleles to ubiquitous (*tub-GAL4*)[61], wing (*nub-GAL4*)[56] or eye (*GMR-GAL4*)[62] specific drivers. In the ubiquitous expression screen, 3-4 virgin females of *tub-GAL4/TM3 Sb* flies were crossed to 2-4 males of the UAS-cDNA reference and variant at 25°C. After 3-4 days, the parents were transferred into new vials, and the new vial was placed at 29°C while the old vial was kept at 25°C, allowing us to test two temperatures simultaneously. The parents were discarded after 3-5 days. Flies were collected after most of the pupae eclosed. The total number of flies were counted and scored with the genotype of interest (i.e. *tub-GAL4>UAS-cDNA*) as well as all other genotypes, (i.e. genotypes with balancers). A minimum of 10 flies were scored per experiment, though for the majority of crosses 50-100 flies were scored in this primary analysis. Viability was calculated by taking the % of observed/expected based on Mendelian ratio, and any UAS-cDNA with survival less than 70% was recorded as having scorable phenotype (lethal or semi-lethal). All of lines showing a phenotype at 29°C also showed phenotypes at 25°C, so subsequent experiments were performed at 25°C. To validate our hits, we performed the same viability assay, except each UAS-cDNA was tested at least three times to statistically validate that there is a difference between

reference and variant. In addition, two independent UAS-cDNA transgenic lines established from the same construct were tested for each reference and variant. A variant was considered to have functional consequence (true hit) if both transgenic lines showed the same phenotype. In the cases where the difference is rather minor (e.g. <20% difference between survival), this was considered within the variation of the experiment paradigm, and the variant phenotype was documented. Functional study using wing or eye drivers were performed using similar strategies, but morphological phenotypes were scored instead of lethality.

Imaging of adult fly morphology:

Drosophila eyes, wings and nota (dorsal thorax) were imaged after flies were frozen at -20°C for at least 24 hours. Wings for some flies were dissected in 70% EtOH and mounted onto slides for imaging. Images were obtained with the Leica MZ16 stereomicroscope equipped with Optronics MicroFire Camera and Image Pro Plus 7.0 software to extend the depth-of-field for Z-stack images.

Expression analysis of TG4 lines in larval and adult brains:

All *TG4* lines are crossed with UAS nlsGFP (3rd chromosome) at room temperature. The brains of GFP positive third instar larvae and 3-5 days old adult flies were dissected in 1X phosphate-buffered saline (PBS). Adult brains were fixed immediately in 4% paraformaldehyde (PFA) and incubated at 4°C overnight (o/n) on a shaker. Next day these brains were post-fixed with 4% PFA with 2% Triton-X in PBS (PBST), kept in a vacuum container for an hour to get rid of the air from the tracheal tissue also make the tissue more permissive. Fixative was replaced every 10 minutes during this post-fixation step. Larval brains were fixed for 50 minutes on a rotator at room temperature. After

thorough washing with PBS with 0.2% Triton (PBTX) both adult and larval brains were incubated with primary antibodies overnight (o/n) at 4°C on shaker. The sample were extensively washed with 0.2% PBTX before secondary antibodies were applied at room temperature for 2 hours. Samples were thoroughly washed with PBST and mounted on a glass slide using Vectasheild (Vector Labs, H-1000-10). Primary antibodies used: Mouse anti-repo (DSHB: 8D12) 1:50, Rat anti-elav (DSHB: 7E8A10) 1:100, Goat anti-GFP (ABCAM: ab6662) 1:500. Secondary antibodies used: Anti-mouse-647 (Jackson ImmunoResearch: 715-605-151) 1:250, Anti-rat-Cy3 (Jackson ImmunoResearch: 712-165-153) 1:500. The samples were scanned using a laser confocal microscope (Zeiss LSM 880), and images were processed using ZEN (Zeiss) and Imaris (Oxford Instruments) software.

GeneOntology analysis:

Gene Ontology (GO) analysis was determined based on the PANTHER (Protein Analysis Through Evolutionary Relationships) system (http://www.pantherdb.org; date last accessed October 31, 2020[30]. Statistical analysis was performed by using the default PANTHER Overrepresentation Test (Released 20200728), Annotation Version and Release Date: GO Ontology database DOI: 10.5281/zenodo.4033054 Released 2020-09-10 which used the Fisher's Exact test with a false discovery rate $p<0.05$.

Patient recruitment and consent:

Affected individuals were investigated by their referring physicians at local sites. Prior to research studies, informed consent was obtained according to the institutional review boards (IRB) and ethnics committees of each institution. Individuals who were ascertained in diagnostic testing procedures (and/or their legal guardians) gave clinical

written informed consent for testing, and their permission for inclusion of their anonymized data in this cohort series. This was obtained using standard forms at each local site by the responsible referring physicians.

Exome sequencing and identification of GLRA2 variants:

Subjects 1, 2 and 4 had clinical exome sequencing at GeneDx (Gaithersburg, MD, United States), at the Praxis für Humangenetik Tubingen (Tubingen, Germany), and at Baylor Genetics (Houston, TX, United States), respectively. Subject 3 WES was performed at the Meyer Children's Hospital, University of Florence, in the context of the DESIRE program and as previously described[63]. Briefly, the SureSelectXT Clinical Research Exome kit (Agilent Technologies, Santa Clara, CA) was used for library preparation and target enrichment, and paired-end sequencing was performed using Illumina sequencer (NextSeq550, Illumina, San Diego, CA, USA) to obtain an average coverage of above 80x, with 97.6% of target bases covered at least 10x. Reads were aligned to the GRCh37/hg19 human genome reference assembly by the BWA software package, and the GATK suite was used for base quality score recalibration, realignment of insertion/deletions (InDels), and variant calling[64,65]. Variant annotation and filtering pipeline included available software (VarSeq, Golden Helix, Inc v1.4.6), focusing on non-synonymous/splice site variants with minor allele frequency (MAF) lower than 0.01 in the GnomAD database[26] (http://gnomad.broadinstitute.org/), an internal healthy control database and pre-computed genomic variants score from dbNSFP[66]. Subject 5 had exome sequencing at Lyon Universiy Hospital (Lyon, France). The SeqCap EZ Medexome kit (Roche, Pleasanton, CA, USA) was used for library preparation and target enrichment before paired-end sequencing using an Illumina instrument (NextSeq500,

Illulina, San Diego, CA, USA). A mean depth of coverage of 133x was obtained with 99.0% of target bases covered at least 10x. Reads were aligned to the GRCh37/hg19 human genome reference assembly by the BWA software package, and the GATK suite was used for base quality score recalibration, realignment of insertion/deletions (InDels), and variant calling[64,65]. Variant annotation was performed with SnpEFF and filtering pipeline focused on non-synonymous/splice site variants with minor allele frequency (MAF) lower than 0.01 in the GnomAD database[26] (http://gnomad.broadinstitute.org/).

Subject 6 WES was performed at the Erasmus MC as previously described[67]. In brief, exome-coding DNA was captured with the Agilent SureSelect Clinical Research Exome (CRE) kit (v2). Sequencing was performed on an Illumina HiSeq 4000 platform with 150-bp paired-end reads. Reads were aligned to hg19 using BWA (BWA-MEM v0.7.13) and variants were called using the GATK Haplotype Caller[64] v3.7 (https://www.broadinstitute.org/gatk/). Detected variants were annotated, filtered and prioritized using the Bench lab NGS v5.0.2 platform (Agilent technologies). Subject 7 WES and data processing were performed by the Genomics Platorm at the Broad Institute of MIT and Harvard with an Illumina Nextera or Twist exome capture (~38 Mb target), and sequenced (150 bp paired reads) to cover >80% of targets at 20x and a mean target coverage of >100x. WES data was processed through a pipeline based on Picard and mapping done using the BWA aligner to the human genome build 38. Variants were called using Genome Analysis Toolkit (GATK) HaplotypeCaller package version 3.5[64] (https://www.broadinstitute.org/gatk/). Subject 8 WES was performed in collaboration with the Autism Sequencing Consortium (ASC) at the Broad Institute on Illumina HiSeq sequencers using the Illumina Nextera exome capture kit. Exome sequencing data was

processed through a pipeline based on Picard and mapping done using the BWA aligner to the human genome build 37 (hg19). Variants were called using Genome Analysis Toolkit (GATK) HaplotypeCaller package version 3.4[64] (https://www.broadinstitute.org/gatk/). Variant call accuracy was estimated using the GATK Variant Quality Score Recalibration (VQSR) approach. High-quality variants with an effect on the coding sequence or affecting splice site regions were filtered against public databases (dbSNP150 and gnomAD V.2.0) to retain (i) private and clinically associated variants; and (ii) annotated variants with an unknown frequency or having minor allele frequency <0.1%, and occurring with a frequency <2% in an in-house database including frequency data from > 1,500 population-matched WES. The functional impact of variants was analyzed by CADD V.1.3, Mendelian Clinically Applicable Pathogenicity V.1.0[29,68], and using InterVar V.0.1.6 to obtain clinical interpretation according to American College of Medical Genetics and Genomics/Association for Molecular Pathology 2015 guidelines[69]. GeneMatcher[33] (https://genematcher.org/) assisted in the recruitment of Subjects 2, 3 and 5-8.

Western blot:

Five heads of *nSyb-GAL4 UAS-GLRA2* reference and variant flies aged for 5 days post eclosion were lysed in 30µL NETN buffer (50mM Tris pH 7.5, 150mM NaCl, 0.5% NP-40, 1 mM EDTA) with an electric douncer for 10 seconds for three times on ice. 30µL of 2x Laemmli Sample Buffer (Bio-Rad) with 10% 2-mercaptoethanol was added to the lysis and incubated on ice for 10 min. Samples were boiled at 95°C and spun at 14,000 RPM for 5 minutes at 4 °C. The soluble fraction was loaded onto a standard SDS-PAGE gel. PVDF (polyvinylidene difluoride) membrane activated for 1 minute with 100% methanol.

After running and wet transfer, the membrane was blocked in 5% skim milk for 1 hour. The membrane was incubated (overnight, shaking, at 4°C) with mouse anti-HA (HA.11, 1:1,000, 901501, BioLegend) and mouse anti-Actin (C4) (1:50,000, MAB1501, EMD Millipore) primary antibodies in 3% BSA (bovine serum albumin), followed by 10 minute washes (3 times) with 1% Triton-X in Tris-buffered saline (TBST). We incubated this with goat anti-mouse HRP-conjugated (1:15000, 115-035-146, Jackson ImmunoResearch) secondary antibody in skim milk. The membrane was washed three times with 1% TBST and detected with Western Lightning™ Chemiluminescence Reagent Plus (perkinelmerNEL104001EA) ECL solution using the Bio-Rad ChemiDoc MP imaging system.

Structural biological analysis of GLRA2 patient variants:

Protein residues that corresponds to GLRA2 patient variants were mapped onto the crystal protein structure of GLRA1 protein in Protein Data Bank (PBD, ID: 4X5T)[70] using the PyMOL (https://pymol.org/)[71] because GLRA1 and GLRA2 are highly homologous proteins (85% similarity, 78% identity and 3% gaps) based on DIOPT[51].

Image generation:

Cartoon images in Fig. 1h, Supplementary Fig. 1a, and Fig. 2g were generated with BioRender.com.

Extended Supplemental Data – GLRA2 subject case histories

Subject 1 is an 8-year-old female with global developmental and cognitive delay. Pregnancy was naturally conceived and uncomplicated, other than decreased fetal movements noted by the mother. She was delivered at term (39 weeks gestational age) via c/section due to breech presentation. Birth weight was 3,600 grams. Neonatal period

was uneventful. There were no feeding difficulties and her growth remained within the normal limits. She was delayed with all her milestones but most significantly for speech (walked at 18 months, first words at 24 months and combined words to sentences at 4-5 years of age, scribbled with a crayon at 3.5 years). She was diagnosed with mixed expressive-receptive speech delay and received speech therapy, occupational therapy and physical therapy interventions. In school she exhibits learning problems, inattention and is below her grade level. She has a modified curriculum and is receiving resources in reading and math. There is no history of developmental regression or seizures. The medical history is otherwise significant for nystagmus that was first noted in infancy and improved with age, as well as myopia and astigmatism requiring corrective glasses. Family ethnicity is Hispanic and the family history was non-contributory. The patient had a normal brain MRI (magnetic resonance imaging) at 6 months of age. EEG (electroencephalography) at 4 years of age showed a slow and poorly formed background, indicative of mild encephalopathy, but did not detect epileptiform activity. Genetic testing included: mitochondrial DNA sequencing which detected a pathogenic variant m.13042 G>A though at heteroplasmy level of 1.9%, however this was felt unlikely to explain the phenotype. CMA (chromosomal microarray) was negative. Trio whole exome sequencing (WES) detected a *de novo*, heterozygous variant of unknown clinical significance in *GLRA2*, c.887C>T, p.Thr296Met (NC_000023.10: g.14627284C>T). This variant is absent in gnomAD. More recent clinical reanalysis of exome data did not detect any other candidates that may explain the phenotype.

Subject 2 is a 6-year-old female with epilepsy, developmental delay (DD), mild intellectual disability (ID) and autism spectrum disorder (ASD). Pregnancy was uncomplicated and she was delivered at term (41 weeks gestational age) via vaginal delivery with vacuum extraction. The neonatal period was uneventful. At the age of 6 months, she developed a severe epileptic encephalopathy with myoclonic seizures. Seizure control was achieved with medications, and she has been seizure-free without medications since the age of about two years old. Delayed psychomotor development was noted, most significantly for her speech with a mixed expressive-receptive speech delay (non-verbal). Her ability to concentrate is poor and she displays mood swings. The medical history is otherwise significant for nystagmus that was first noted in infancy (6 weeks old) and improved with age, and sleep disturbance. She has mild microcephaly [< 1st centile: -2.84 standard deviation (SD)] and mild bilateral cutaneous 3rd-4th syndactyly, with no other congenital anomalies. Family ethnicity is European (German/Italian) and the family history is significant for a maternal aunt that had epilepsy in adulthood but her cognitive development was normal. Brain MRI showed delayed myelination at 7 months old and a small arachnoid cyst. EEG was abnormal for bilateral synchronized, sometimes high amplitude spike/polyspike-waves-complexes, and bitemporo-occipital hints for severe functional defects with epileptic potentials. Chromosomal analysis, Angelman syndrome methylation study, epilepsy next generation sequencing (NGS) gene panel and *MECP2* sequencing were negative. Trio WES detected a *de novo*, heterozygous variant of unknown clinical significance in *GLRA2*, c.887C>T, p.Thr296Met (NC_000023.10: g.14627284C>T). This variant is absent in gnomAD.

Subject 3 is a 5 year 6 months old female with DD, microcephaly, abnormal eye movements and ataxic gait. Pregnancy was uncomplicated and she was born at term via c/section. Abnormal eye movements were noticed two weeks after birth, during hospitalization due to a lower respiratory tract infection. At the age of 6 months, clinical examination revealed mildly delayed developmental milestones and erratic conjugate eye movements akin to opsoclonus. At age 4 years OFC (occipitofrontal circumference) was 43 cm (< 1st centile: -4.28 SD) and ophthalmological evaluation revealed alternating exotropia, for which patching therapy was initiated. Language was limited to a few words and neuropsychological evaluation documented moderate developmental delay (Bayley-III). The patient could walk unsupported with ataxic gait. At age 5 years 6 months, erratic eye movements were considerably reduced and she could walk independently but her expressive language was still limited to a few words, with delayed receptive speech and nonverbal communicative skills. Family ethnicity is European and the family history is unremarkable. Brain MRI at 6 months of age showed mild cortical atrophy with thinning of the corpus callosum. EEG, while awake and asleep, laboratory and metabolic investigations were unremarkable. Array-CGH (comparative genomic hybridization) highlighted a maternally inherited 3q25.32 duplication (chr3:157746089-158324659, hg19) that was interpreted as likely benign. Trio WES detected a *de novo* heterozygous variant in *GLRA2*, c.887C>T, p.Thr296Met (NC_000023.10: g.14627284C>T). This variant is absent in gnomAD. In addition, it detected a *de novo* variant in *CACNA1B*, c.5381C>T, p.Thr1794Met (NC_000009.11:g.141000212C>T), which is a variant of unknown significance in a gene that is linked to an autosomal recessive condition (Neurodevelopmental disorder with seizures and nonepileptic hyperkinetic movements,

MIM #618497). Failure to identify a second allele in this gene reduces the likelihood that this variant is responsible for this patient's phenotype.

Subject 4 was a female infant with seizures and severe developmental delay who passed away at 7 months of age secondary to complications of COVID-19 infection. Pregnancy was uneventful and she was born at term (40 weeks gestational age). She was noted to have focal seizures at 2-3 weeks of age, and was diagnosed with infantile spasms when she was 5 months old. At 6 months of age she was not reaching for objects, not sitting up and only making high-pitched sounds. She had borderline microcephaly with dysmorphic features including midface retrusion, apparent hypotelorism, deep set eyes, thick eyebrows, downturned corners of the mouth, and wide-spaced nipples. Family ethnicity is Hispanic and the family history was unremarkable. She had normal plasma and CSF (cerebrospinal fluid) lactate, pipecolic acid and piperideine-6-carboxylate, ammonia, urine organic acids, plasma amino acids, acylcarnitine profile, and CSF amino acids. An Epilepsy gene panel was non-diagnostic. Trio WES detected a *de novo* heterozygous variant in *GLRA2*, c.887C>T, p.Thr296Met (NC_000023.10: g.14627284C>T). This variant is absent in gnomAD.

Subject 5 is a 6 years and 7 months old female with a history of infantile spasms, epilepsy and intellectual disability. She was born at term and first presented with infantile spasms at 3 months of age. This evolved to atonic and tonic-clonic seizures as she grew up. She was delayed with all milestones (walked at 4.5 years old and remains non-verbal). She had nystagmus that improved with age and strabismus. The medical history is otherwise

significant for hyperactivity, inattention and sleep disturbance. Her ethnicity is African (Senegal). Brain MRI at 3 years of age showed cortical and white matter atrophy, including vermian atrophy. EEG showed hypsarrythmia at onset and she had normal interictal EEG afterwards. She had normal SNP (single nucleotide polymorphism) array, negative targeted epilepsy panel and negative metabolic lab results. Trio WES identified a *de novo* heterozygous variant in *GLRA2*, c.140T>C, p.Phe47Ser (NC_000023.10: g.14550432T>C). This variant is absent in gnomAD.

Subject 6 is an 11-month-old male with hypotonia, DD and dysmorphic craniofacial features. Pregnancy was uncomplicated, he was delivered at term (38 and 3/7 weeks gestational age) and the neonatal period was uneventful. Soon after birth dysmorphic features were noted, including an elongated face, high anterior hairline, epicanthal folds, downslanting palpebral fissures and a bulbous nose. Growth remains within the normal limits. His medical history is otherwise significant for obstructive sleep apnea and strabismus. Family ethnicity is European (Dutch) and the family history is significant for the maternal grandfather who has not further specified unexplained neurological complaints, and which could not be further investigated. Investigations for metabolic disorders, Fragile X syndrome and a SNP-array were normal. Trio WES identified a rare variant in *GLRA2*, c.754C>T, p.Arg252Cys (NC_000023.10: g.14627151C>T), which was inherited from mother. No other possible disease explaining variant was identified. The mother displayed skewed X chromosome inactivation (82% on two measurements). The variant was absent in the maternal uncle and the maternal grandmother, but was inherited from the maternal grandfather, who was not available for clinical investigations. His level

of functioning remains unknown. This variant is present in one heterozygous female in gnomAD.

Subject 7 is a 7-year-old male with epilepsy, DD with regression, and ASD. Pregnancy was uncomplicated. He was born full term via uncomplicated delivery, and his early development was as expected. He was speaking in sentences at 2.5 years old when he started having generalized tonic-clonic seizures. He developed staring spells, ataxia, and an increased frequency of myoclonic jerks, which around the age of 6 years old were occurring 20 times per day on average, with 5-6 atonic seizures per day each lasting less than 30 seconds. Following seizure onset he experienced developmental regression. At 3 years of age he was diagnosed with ASD. At 6 years of age his vocabulary was about 20 words, with gains in development lost following significant seizures. His ethnicity is European, and the family history is significant for a younger brother with ASD, although he has not presented with seizures. Neither mother nor father have a history of seizures or delays. At age three, EEG depicted generalized slowing and generalized epileptiform discharges associated with myoclonic jerks. MRI showed minimal increased T2 signal intensity on the occipital lobes that was thought to be within normal limits. Genetics testing for Fragile X syndrome, Prader-Willi/Angelman syndromes, and congenital disorders of glycosylation were normal. Additional tests, including plasma amino acids, lysosomal enzymes, and cerebral creatine deficiency were also normal. Microarray reported a maternally inherited 1p33 deletion of unknown significance (48,688,391-49,922,153). The patient was enrolled to The Manton Center for Orphan Disease Gene Discovery Core protocol. Trio WES discovered a maternally inherited variant in *GLRA2*,

c.862G>A, p. Ala288Thr (NC_000023.10: g.14627259G>A). This variant is absent in gnomAD.

Subject 8 is a 35-year-old male with a history of DD, learning disabilities and ASD. Pregnancy was uncomplicated and he was born at term (40 weeks gestational age). Since early childhood he showed slow movement and difficulties in motor coordination. He walked and said his first words at 24 months, and first sentences at age 3 years of age. In school learning disabilities were noted, including difficulties in writing, reading, praxias, temporal orientation, calculation, drawing, and visuo-spatial organization. He graduated high school and continued to higher education, though he did not complete a degree. Neuropsychiatric assessment in adulthood was consistent with ASD and social and cognitive deficits. There is no history of seizures. The medical history is otherwise significant for environmental allergies, myopia and astigmatism. The ethnicity is European, and the family history is unremarkable, except for a maternal grandmother with Alzheimer's dementia. The patient had a normal brain MRI at 29 years old. Trio WES identified a maternally inherited variant in *GLRA2*, c.1186C>A, p. Pro396Thr (NC_000023.10: g.14748434C>A). This variant is present in 3 heterozygous females and 1 hemizygous male in gnomAD.

References

1. Association, A. P. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. (American Psychiatric Pub, 2013).

2. Fischbach, G. D. & Lord, C. The Simons Simplex Collection: a resource for identification of autism genetic risk factors. *Neuron* **68**, 192–195 (2010).

3. Coe, B. P. *et al.* Neurodevelopmental disease genes implicated by de novo mutation and copy number variation morbidity. *Nature genetics* **51**, 106–116 (2019).

4. Iossifov, I. *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216–221 (2014).

5. Sanders, S. J. *et al.* De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012).

6. Rubeis, S. D. *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014).

7. Yuen, R. K. C. *et al.* Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nature neuroscience* **20**, 602–611 (2017).

8. Takata, A. *et al.* Integrative Analyses of De Novo Mutations Provide Deeper Biological Insights into Autism Spectrum Disorder. *Cell reports* **22**, 734–747 (2018).

9. Satterstrom, F. K. *et al.* Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**, 568-584.e23 (2020).

10. Bellen, H. J., Wangler, M. F. & Yamamoto, S. The fruit fly at the interface of diagnosis and pathogenic mechanisms of rare and common human diseases. *Human molecular genetics* **28**, R207–R214 (2019).

11. Marcogliese, P. C. & Wangler, M. F. *Drosophila as a Model for Human Diseases*. vol. 23 (American Cancer Society, 2001).

12. Link, N. & Bellen, H. J. Using Drosophila to drive the diagnosis and understand the mechanisms of rare human diseases. *Development (Cambridge, England)* **147**, dev191411 (2020).

13. Venken, K. J. T. *et al.* MiMIC: a highly versatile transposon insertion resource for engineering Drosophila melanogaster genes. *Nature methods* **8**, 737–743 (2011).

14. Nagarkar-Jaiswal, S. *et al.* A library of MiMICs allows tagging of genes and reversible, spatial and temporal knockdown of proteins in Drosophila. *eLife* **4**, 2743 (2015).

15. Diao, F. *et al.* Plug-and-play genetic access to drosophila cell types using exchangeable exon cassettes. *Cell reports* **10**, 1410–1421 (2015).

16. Lee, P.-T. *et al.* A gene-specific T2A-GAL4 library for Drosophila. *eLife* **7**, 1377 (2018).

17. Gnerer, J. P., Venken, K. J. T. & Dierick, H. A. Gene-specific cell labeling using MiMIC transposons. *Nucleic acids research* **43**, e56–e56 (2015).

18. Tang, W. *et al.* Faithful expression of multiple proteins via 2A-peptide self-processing: a versatile and reliable method for manipulating brain circuits. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **29**, 8621–8629 (2009).

19. Marcogliese, P. C. *et al.* IRF2BPL Is Associated with Neurological Phenotypes. *American journal of human genetics* **103**, 245–260 (2018).

20. Velasco, B. de *et al.* Specification and development of the pars intercerebralis and pars lateralis, neuroendocrine command centers in the Drosophila brain. *Developmental biology* **302**, 309–323 (2007).

21. Kanca, O. *et al.* De Novo Variants in WDR37 Are Associated with Epilepsy, Colobomas, Dysmorphism, Developmental Delay, Intellectual Disability, and Cerebellar Hypoplasia. *American journal of human genetics* **105**, 413–424 (2019).

22. Guo, H. *et al.* Disruptive mutations in TANC2 define a neurodevelopmental syndrome associated with psychiatric disorders. *Nature communications* **10**, 4679–17 (2019).

23. Seeds, A. M. *et al.* A suppression hierarchy among competing motor programs drives sequential grooming in Drosophila. *eLife* **3**, e02951 (2014).

24. Wang, J. *et al.* MARRVEL: Integration of Human and Model Organism Genetic Resources to Facilitate Functional Annotation of the Human Genome. *Am J Hum Genetics* **100**, 843–853 (2017).

25. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).

26. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).

27. Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M. & Ng, P. C. SIFT missense predictions for genomes. *Nature protocols* **11**, 1–9 (2016).

28. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nature methods* **7**, 248–249 (2010).

29. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nature genetics* **46**, 310–315 (2014).

30. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics* **25**, 25–29 (2000).

31. Levitt, P. & Campbell, D. B. The genetic and neurobiologic compass points toward common signaling dysfunctions in autism spectrum disorders. *The Journal of clinical investigation* **119**, 747–754 (2009).

32. Sullivan, P. F. & Geschwind, D. H. Defining the Genetic, Genomic, Cellular, and Diagnostic Architectures of Psychiatric Disorders. *Cell* **177**, 162–183 (2019).

33. Sobreira, N., Schiettecatte, F., Valle, D. & Hamosh, A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Human mutation* **36**, 928–930 (2015).

34. Chong, J. X. *et al.* The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. *American journal of human genetics* **97**, 199–215 (2015).

35. Gahl, W. A. *et al.* The NIH Undiagnosed Diseases Program and Network: Applications to modern medicine. *Molecular genetics and metabolism* **117**, 393–400 (2016).

36. Amberger, J. S., Bocchini, C. A., Scott, A. F. & Hamosh, A. OMIM.org: leveraging knowledge across phenotype-gene relationships. *Nucleic acids research* **47**, D1038–D1043 (2019).

37. Zeilhofer, H. U., Acuña, M. A., Gingras, J. & Yévenes, G. E. Glycine receptors and glycine transporters: targets for novel analgesics? *Cellular and molecular life sciences : CMLS* **75**, 447–465 (2018).

38. Study, D. D. D. Prevalence and architecture of de novo mutations in developmental disorders. *Nature* **542**, 433–438 (2017).

39. Pinto, D. *et al.* Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466**, 368–372 (2010).

40. Du, J., Lü, W., Wu, S., Cheng, Y. & Gouaux, E. Glycine receptor mechanism elucidated by electron cryo-microscopy. *Nature* **526**, 224–229 (2015).

41. Moraga-Cid, G. *et al.* Allosteric and hyperekplexic mutant phenotypes investigated on an α1 glycine receptor transmembrane structure. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 2865–2870 (2015).

42. Pauli, A. *et al.* Cell-type-specific TEV protease cleavage reveals cohesin functions in Drosophila neurons. *Developmental cell* **14**, 239–251 (2008).

43. Deal, S. L. & Yamamoto, S. Unraveling Novel Mechanisms of Neurodegeneration Through a Large-Scale Forward Genetic Screen in Drosophila. *Frontiers in genetics* **9**, 700 (2018).

44. Xiong, B. *et al.* Crag is a GEF for Rab11 required for rhodopsin trafficking and maintenance of adult photoreceptor cells. *PLoS biology* **10**, e1001438 (2012).

45. Chen, S. *et al.* An interactome perturbation framework prioritizes damaging missense mutations for developmental disorders. *Nature genetics* **50**, 1032–1040 (2018).

46. Wong, W.-R. *et al.* Autism-associated missense genetic variants impact locomotion and neurodevelopment in Caenorhabditis elegans. *Human molecular genetics* **28**, 2271–2281 (2019).

47. Barakat, T. S. & Gribnau, J. X chromosome inactivation in the cycle of life. *Development* **139**, 2085–2089 (2012).

48. Zhang, Y., Ho, T. N. T., Harvey, R. J., Lynch, J. W. & Keramidas, A. Structure-Function Analysis of the GlyR α2 Subunit Autism Mutation p.R323L Reveals a Gain-of-Function. *Front Mol Neurosci* **10**, 158 (2017).

49. Dudzic, J. P., Hanson, M. A., Iatsenko, I., Kondo, S. & Lemaitre, B. More Than Black or White: Melanization and Toll Share Regulatory Serine Proteases in Drosophila. *Cell reports* **27**, 1050-1061.e3 (2019).

50. Feske, S., Wulff, H. & Skolnik, E. Y. Ion channels in innate and adaptive immunity. *Annual review of immunology* **33**, 291–353 (2015).

51. Hu, Y. *et al.* An integrative approach to ortholog prediction for disease-focused and other functional studies. *BMC bioinformatics* **12**, 357–16 (2011).

52. Team, M. P. *et al.* The completion of the Mammalian Gene Collection (MGC). *Genome research* **19**, 2324–2333 (2009).

53. Bischof, J., Maeda, R. K., Hediger, M., Karch, F. & Basler, K. An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 3312–3317 (2007).

54. Tsang, Y. H. *et al.* Functional annotation of rare gene aberration drivers of pancreatic cancer. *Nature communications* **7**, 10500–11 (2016).

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

56. Calleja, M. *et al.* Generation of medial and lateral dorsal body domains by the pannier gene of Drosophila. *Development (Cambridge, England)* **127**, 3971–3980 (2000).

57. Verstreken, P. *et al.* Synaptojanin is recruited by endophilin to promote synaptic vesicle uncoating. *Neuron* **40**, 733–748 (2003).

58. Simon, J. C. & Dickinson, M. H. A new chamber for studying the behavior of Drosophila. *PloS one* **5**, e8793 (2010).

59. Eyjolfsdottir, E. *et al.* Detecting Social Actions of Fruit Flies. in vol. 8690 772–787 (2014).

60. Kabra, M., Robie, A. A., Rivera-Alba, M., Branson, S. & Branson, K. JAABA: interactive machine learning for automatic annotation of animal behavior. *Nature methods* **10**, 64–67 (2013).

61. Guelman, S. *et al.* The essential gene wda encodes a WD40 repeat subunit of Drosophila SAGA required for histone H3 acetylation. *Molecular and cellular biology* **26**, 7178–7189 (2006).

62. Mehta, S. Q. *et al.* Mutations in Drosophila sec15 reveal a function in neuronal targeting for a subset of exocyst components. *Neuron* **46**, 219–232 (2005).

63. Vetro, A. *et al.* Early infantile epileptic-dyskinetic encephalopathy due to biallelic PIGP mutations. *Neurology Genetics* **6**, e387 (2020).

64. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297–1303 (2010).

65. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**, 491–498 (2011).

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

67. Perenthaler, E. *et al.* Loss of UGP2 in brain leads to a severe epileptic encephalopathy, emphasizing that bi-allelic isoform-specific start-loss mutations of essential genes can cause genetic diseases. *Acta Neuropathol* **139**, 415–442 (2020).

68. Jagadeesh, K. A. *et al.* M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nat Genet* **48**, 1581–1586 (2016).

69. Li, Q. & Wang, K. InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. *Am J Hum Genetics* **100**, 267–280 (2017).

70. Burley, S. K. *et al.* RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic acids research* **47**, D464–D474 (2019).

71. Yuan, S., Chan, H. C. S., Filipek, S. & Vogel, H. PyMOL and Inkscape Bridge the Data and the Data Visualization. *Structure (London, England : 1993)* **24**, 2041–2042 (2016).

Fig. 1: Gene & variant prioritization, resource generation and screening outline.

a, Criteria to prioritize ASD candidate genes and variants for this study. **b-d**, Gene level constraints from control individuals (gnomAD). **e**, Schematic depicting generation and effect of *TG4* lines on gene function. **f**, Schematic illustrating generation of UAS-human cDNA constructs. **g**, Total number of *Drosophila* reagents generated for this study. **h**, Screening paradigms using both humanization and overexpression strategies to assess SSC-DNM function.

Fig. 1:

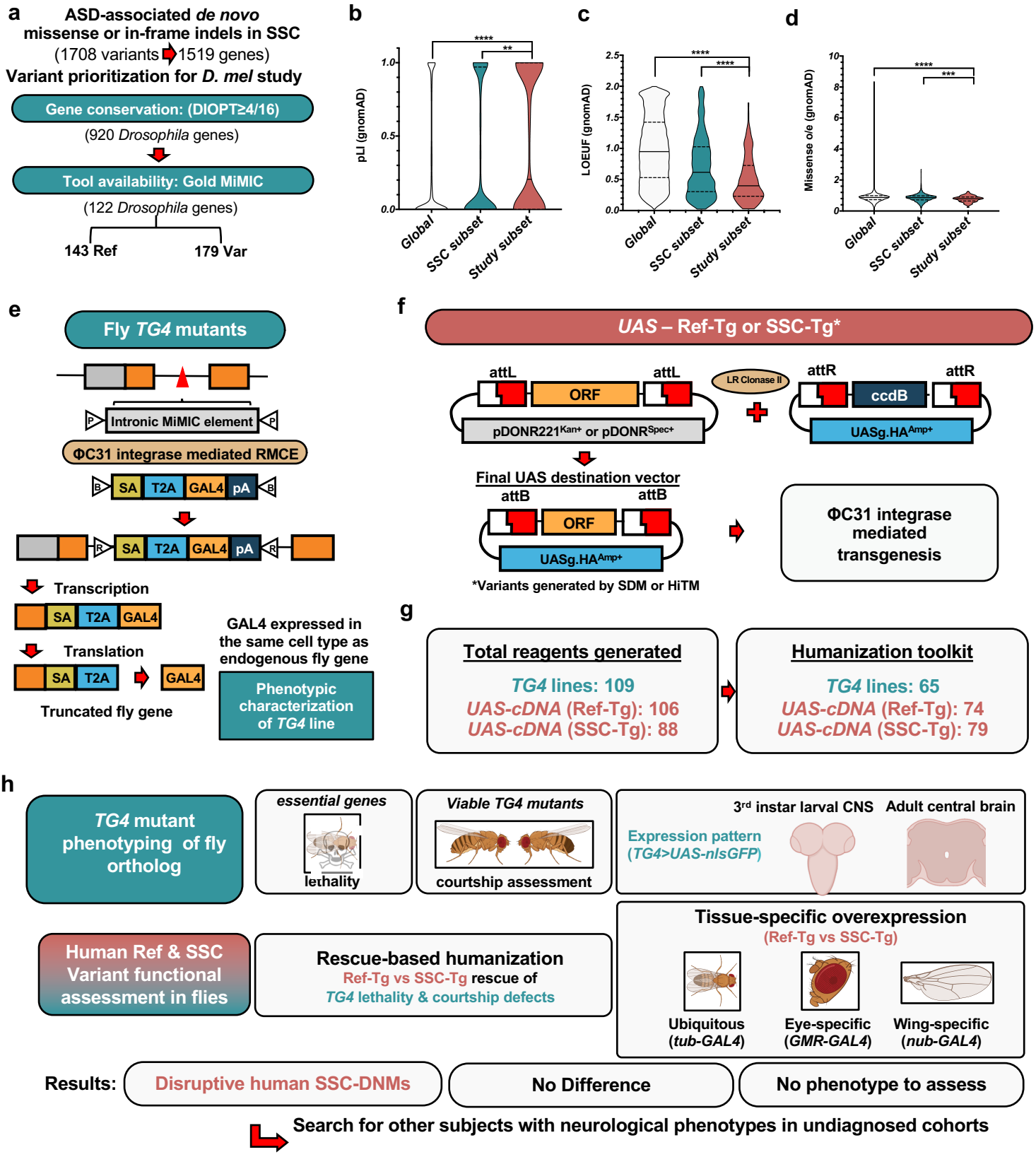

Fig. 2: Assessment of SSC-DNM function through humanization of essential fly genes.

a, Rescue of lethality to adult stage by *TG4* driven UAS-reference human cDNA and subsequent comparison of reference and variant cDNA. **b-d**, Observed/expected Mendelian ratios for rescue of humanized *TG4* mutants across different temperatures. 3 independent crosses where set per genotype and n>50 flies were quantified for each cross. Statistical analyses were performed by ANOVA followed by Sidak's multiple comparisons test. **e-f**, Lifespan analysis of humanized *TG4* lines at 25°C. Survival comparisons obtained by Log-rank (Mantel-Cox) test. **g-h**, Cartoon and confocal images showing expression pattern of UAS-nlsGFP driven by *TG4* (green), and co-staining corresponding to neurons (Elav, magenta) and glia (Repo, cyan). Colocalization of GFP and each cell specific marker (white). Scale bar = 25 μm. Dotted magenta lines outline of the brain. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

Fig. 2:

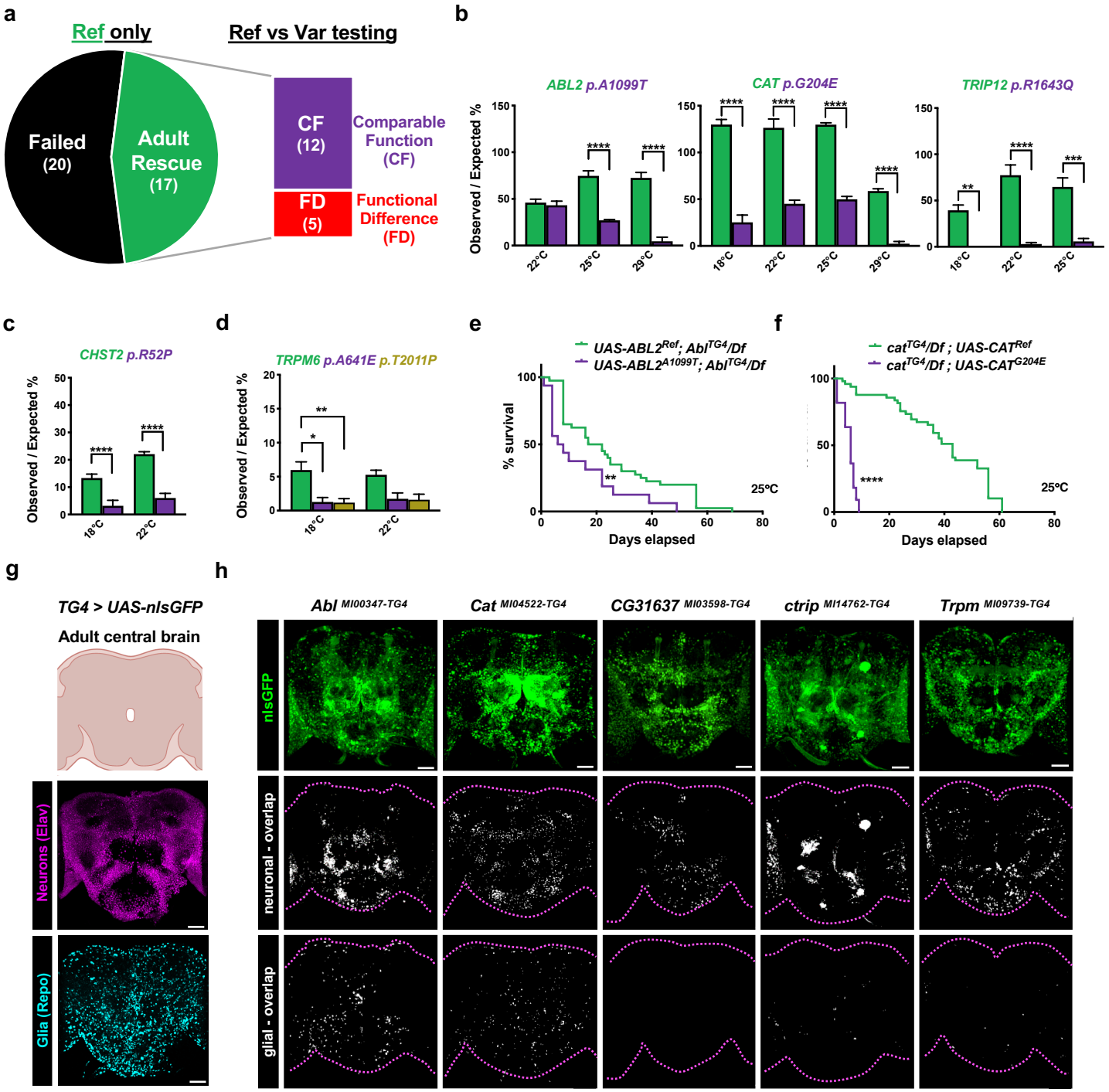

Fig. 3: Variant assessment by overexpression of reference and SSC-DNM corresponding to essential fly genes.

a, Phenotypes observed upon overexpressing the reference and variant cDNAs using a ubiquitous driver (*tub-GAL4*) at 25°C, an eye-specific driver (*GMR-GAL4*) at 29°C, or a wing-specific driver (*nub-GAL4*) at 25°C. Black denotes if there was no phenotype (NP), purple if there was a comparable phenotype (CP), or red if there was a functional difference (FD). **b**, Quantification of viability upon overexpression of reference or variant human cDNAs using a ubiquitous driver for genes where the variants showed a functional difference. Minimum of 3 independent crosses were set with two independent UAS-transgenic lines. 50-100 flies (a minimum of 10 if overexpression caused survival defects) were scored. Statistical analyses were performed by unpaired t-test. **c-d**, Images of fly wings depicting morphological phenotypes using the wing-specific driver. *IRF2BPL and ATP2B2* crosses were performed at 25°C. **e**, Expression analysis of UAS-nlsGFP driven by *TG4* (green) relative to neuronal (Elav) and glial (Repo) markers in the adult brain. The white arrow on the top panel indicates cells of the *pars intercerebralis.* White signal in the lower two rows display overlap between nlsGFP and the neuronal or glial marker. Scale bar = 25 μm. Dotted magenta lines outline the brain. *p<0.05, ***p<0.001, ****p<0.0001.

Fig. 3:

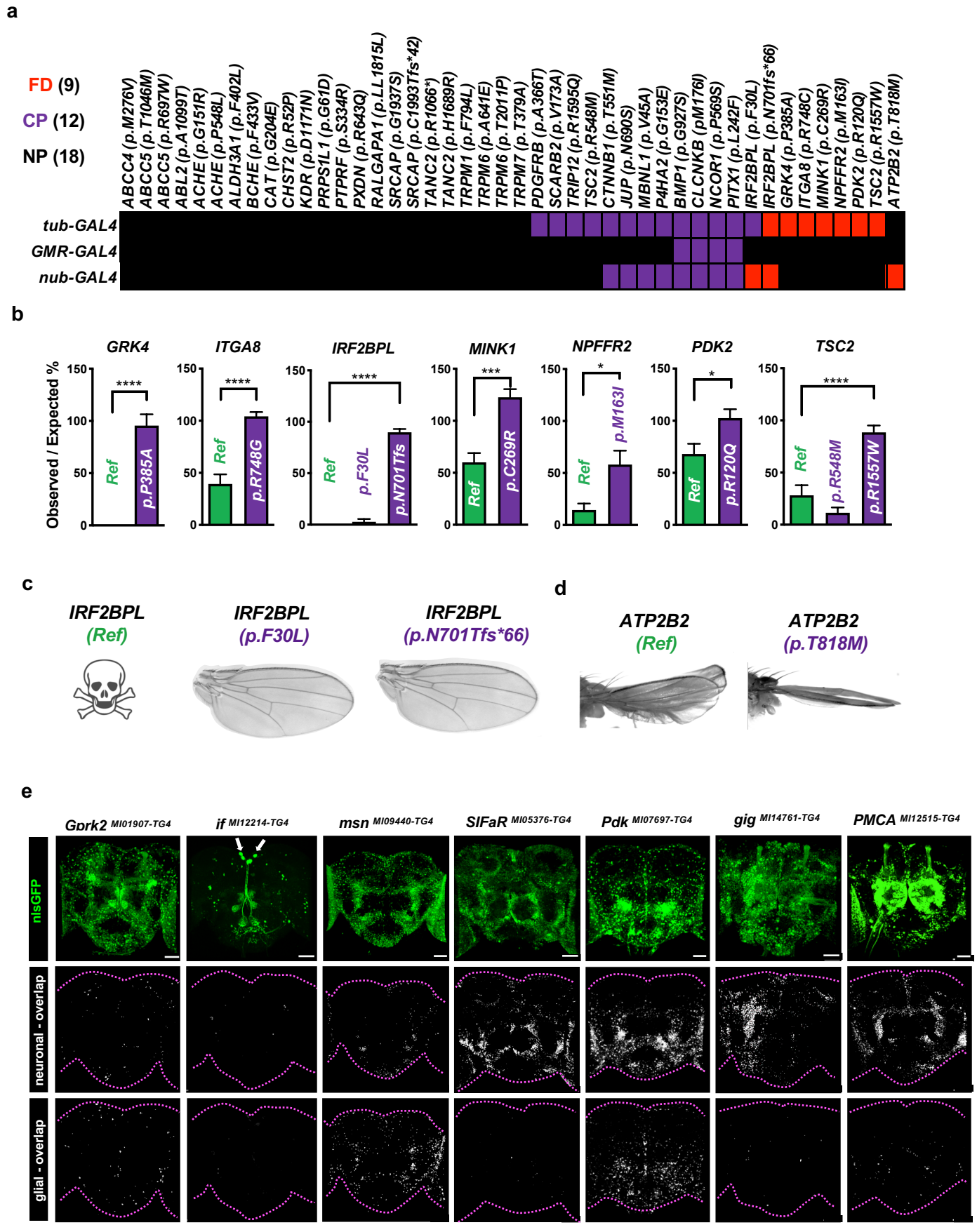

Fig. 4: Assessment of SSC-DNM function through humanization of viable TG4 lines and behavioral analysis.

a, Analysis pipeline used to evaluate *Drosophila* behavior. **b**, SSC-DNMs in which variants display significant differences in time spent performing a specific behavior (courtship, copulation, movement, or grooming) when compared to reference humanized flies. **c-f**, The number of frames male flies spent performing courtship (single-wing extensions), copulating, moving within the chamber, or grooming during a 30-minute test period. The red line represents the average number of frames a *Canton-S* (control) male spends performing the same task. n=10-40 flies were used per genotype. Statistical analysis performed by Kruskal-Wallis one-way analysis of variance and the Dunn's multiple comparison test. **g**, Representative images demonstrating the expression pattern of UAS-nlsGFPdriven by *TG4* in the adult brain co-stained with neuronal (Elav) and glial (Repo) markers. Scale bar = 25 μm. White signal in the two bottom rows display overlap between nlsGFP and the neuronal or glial marker. *p<0.05, ***p<0.001, ****p<0.0001, ns (not significant).

Fig. 4:

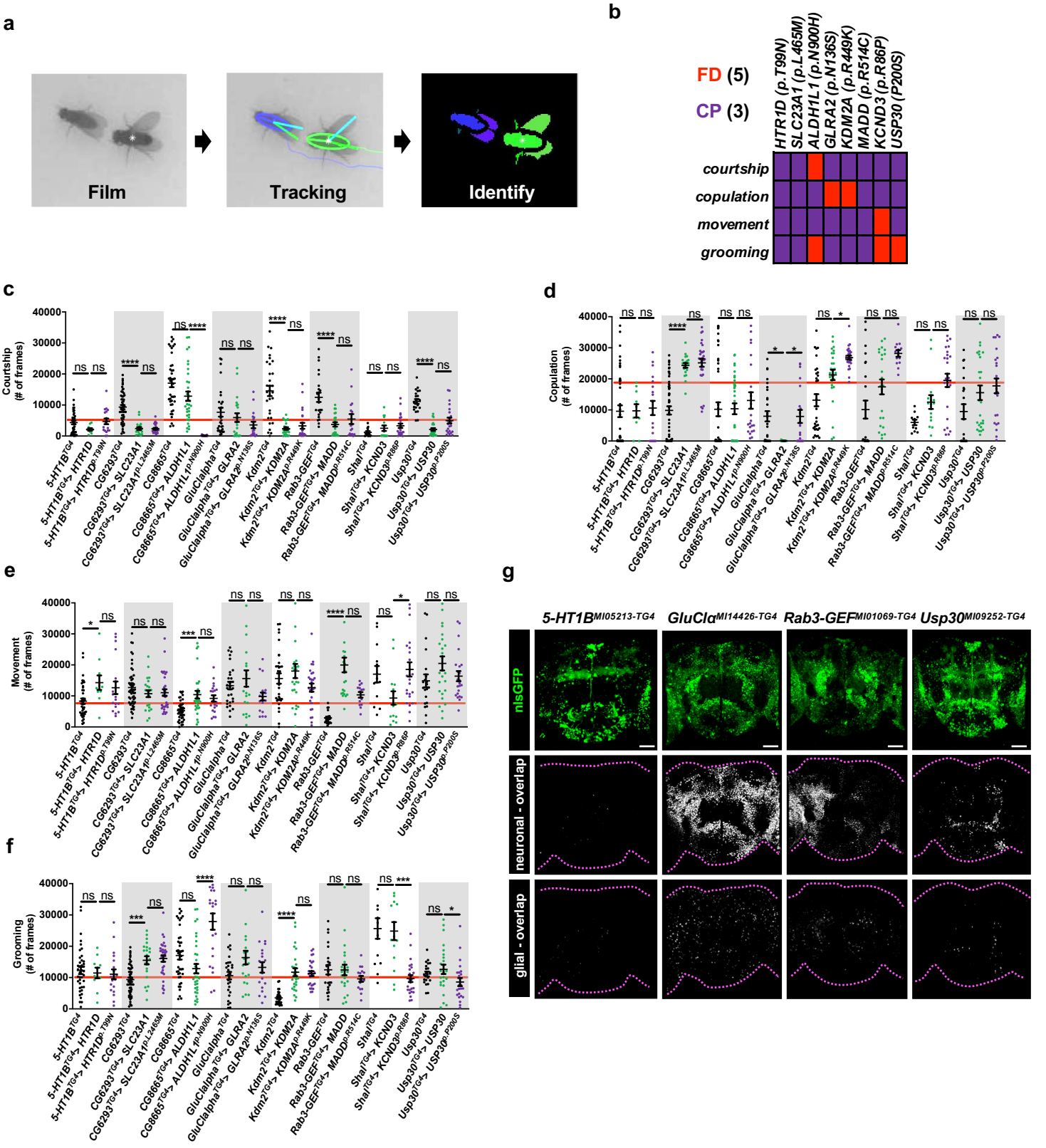

Fig. 5: Variant assessment by overexpression of reference and SSC-DNM corresponding to viable TG4 lines.

a, Phenotypes observed upon overexpressing the reference and variant cDNAs using a ubiquitous driver (*tub-GAL4*) at 25°C, an eye-specific driver (*GMR-GAL4*) at 29°C, and a wing-specific driver (*nub-GAL4*) at 25°C. Black denotes if there was no phenotype (NP), purple if there was a comparable phenotype (CP), or red if there was a functional difference (FD). **b-c**, Quantification of viability upon overexpression the reference or variant SSC-DNMs using a ubiquitous driver (*tub-GAL4*) for genes where the variants showed a functional difference. Minimum of 3 independent crosses were set with two independent UAS-transgenic lines. 50-100 flies (a minimum of 10 if overexpression caused survival defects) were scored. Statistical analyses were performed by unpaired t-test. **d-e**, Representation of optical sections of eyes and wings for variants with a functional difference using *GMR-GAL4* and *nub-GAL4*, respectively. **f**, Representative images demonstrating the expression pattern of UAS-nlsGFP driven by *TG4* in the adult brain co-stained with neuronal (Elav) and glial (Repo) markers. White signal in the bottom two rows display overlap between GFP and the neuronal or glial marker. Scale bar = 25 µm. **p<0.01, ***p<0.001, ****p<0.0001.

Fig. 5:

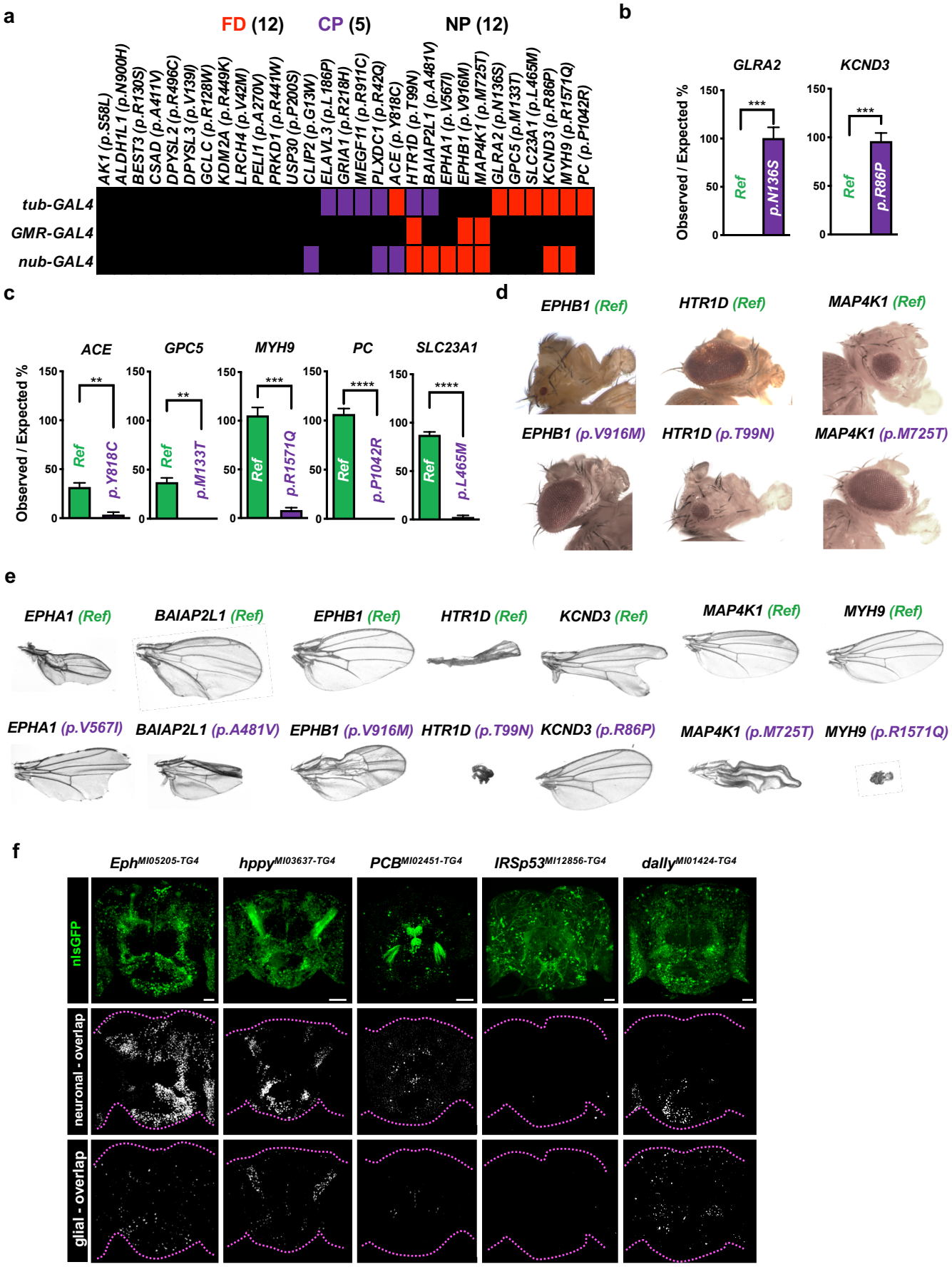

Fig. 6: *GLRA2*^{T296M} found in female patients acts as a GoF allele while *GLRA2*^{R252C} *and GLRA2*^{N136S} found in male patients behave as LoF alleles.

a, Schematic diagram of domain structure of GLRA2 and the relative positions of subject variants functionally assessed in *Drosophila*. **b**, Mendelian ratios upon overexpression the *GLRA2* reference or variant human cDNAs using a ubiquitous driver (*tub-GAL4*). **c-d**, Representative images and quantification of melanized nodules formed on the notum of flies expressing *GLRA2*^{T296M} driven by a dorsocentral throax-specific (*pnr-GAL4*) driver at 25°C. **e-h**, Representative traces of ERG and quantification of "*OFF*"-transient amplitude (blue bracket) in animals expressing GLRA2 pan-neuronally (both pre-synaptic photoreceptors and post-synaptic laminar neurons, *nSyb-GAL4*) or only in the pre-synaptic photoreceptors (*Rh1-GAL4*).

Fig. 6:

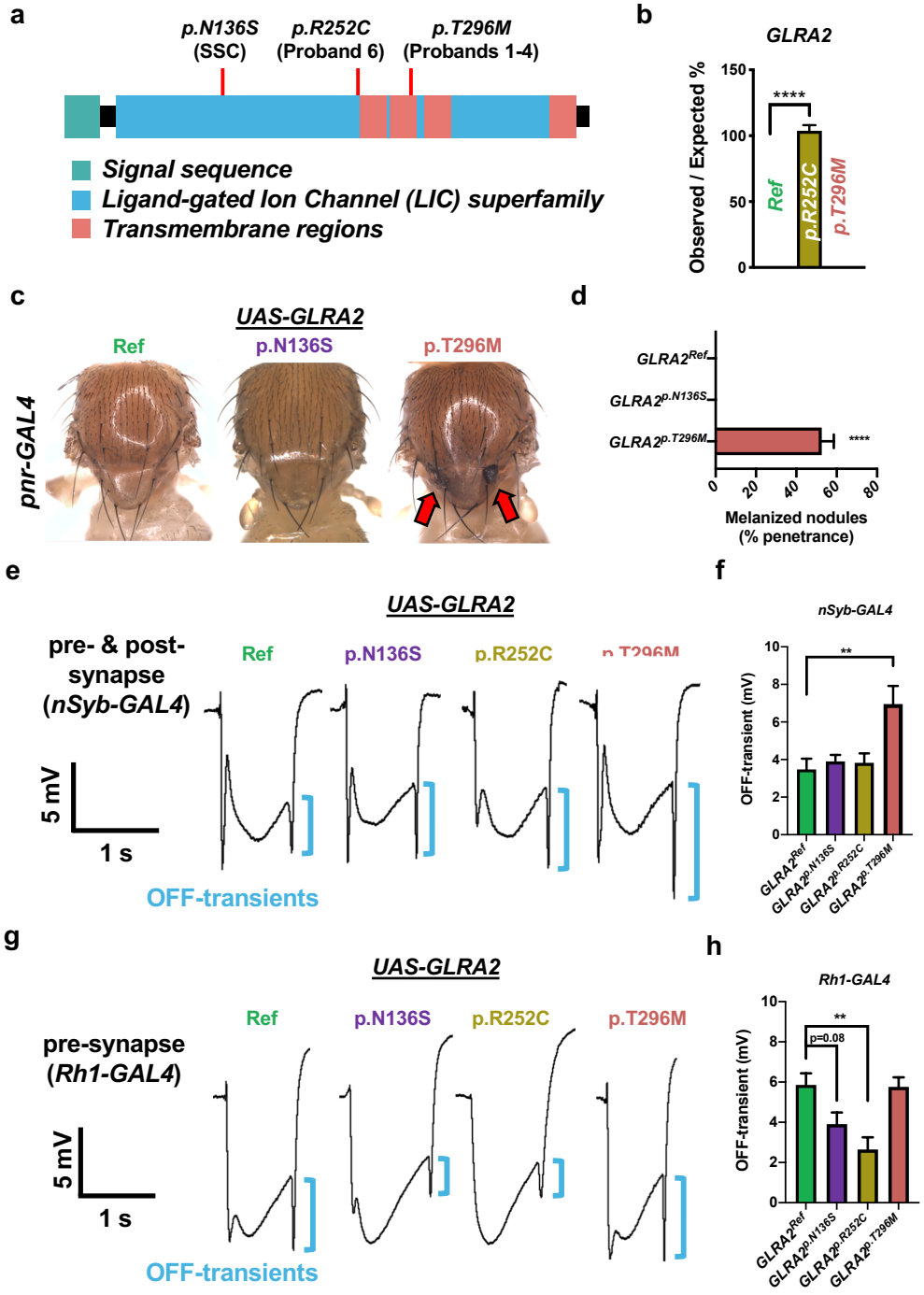

Table 1: Identification of 30 SSC-DNMs with functional consequences.

List of all human genes and corresponding SSC variants determined to have a functional difference across all assays in this study.

H. sap gene	pLI (LOEUF)	Missense O/E	OMIM disease	SSC variant	CADD	*D. mel gene*	TG4 lethality	Functional assay	SSC-CV consequence
ABL2	0 (0.58)	0.81	-	p.A1099T	28.2	*Abl*	Yes	RB	LoF
ACE	0 (1.08)	1.07	267430 (AR)	p.Y818C	7.5	*Ance*	No	OE	GoF
ALDH1L1	0 (0.78)	0.93	-	p.N900H	9.8	*CG8665*	No	RB	GoF
ATP2B2	1 (0.15)	0.54	601386 (AR)	p.T818M	33.0	*PMCA*	Yes	OE	LoF
BAIAP2L1	0 (0.65)	0.90	-	p.A481V	17.8	*IRSp53*	No	OE	GoF
CAT	0 (1.05)	1.01	614097 (AR)	p.G204E	28.1	*Cat*	Yes	RB	LoF
CHST2	0.02 (0.81)	0.66	-	p.R52P	12.8	*CG31637*	Yes	RB	LoF
EPHA1	0 (1.04)	0.97	-	p.V567I	1.3	*Eph*	No	OE	LoF
EPHB1	1 (0.26)	0.73	-	p.V916M	34.0	*Eph*	No	OE	Complex
GLRA2	0.97 (0.30)	0.43	-	p.N136S	25.1	*GluClα*	No	RB, OE	LoF
GPC5	0 (1.08)	1.09	-	p.M133T	24.3	*dally*	No*	OE	GoF
GRK4	0 (1.09)	1.09	-	p.P385A	26.0	*Gprk2*	Yes	OE	LoF
HTR1D	0 (1.30)	0.98	-	p.T99N	19.4	*5-HT1B*	No	OE	GoF
IRF2BPL	0.84 (0.41)	0.90	618088 (AD)	p.F30L p.N701fs	24.8 -	*Pits*	Yes	OE OE	LoF LoF
ITGA8	0 (0.66)	1.03	191830 (AR)	p.R748C	35.0	*if*	Yes	OE	LoF
KCND3	0.99 (0.28)	0.48	607346 (AD)	p.R86P	32.0	*Shal*	No	RB, OE	LoF
KDM2A	1 (0.04)	0.43	-	p.R449K	5.7	*Kdm2*	No	RB	GoF
MAP4K1	0.99 (0.29)	0.59	-	p.M725T	21.3	*hppy*	No	OE	Complex
MINK1	1 (0.13)	0.60	-	p.C269R	26.8	*msn*	Yes	OE	LoF
MYH9	1 (0.09)	0.71	603622 (AD)	p.R1571Q	35.0	*Mhc*	No*	OE	GoF
NPFFR2	0 (1.13)	1.20	-	p.M163I	13.3	*SIFaR*	Yes	OE	LoF
PC	0.01 (0.43)	0.69	266150 (AR)	p.P1042R	24.6	*PCB*	No	OE	GoF
PDK2	0 (0.92)	0.63	-	p.R120Q	25.3	*Pdk*	Yes	OE	LoF
SLC23A1	0.02 (0.54)	0.71	-	p.L465M	17.9	*CG6293*	No	OE	GoF
TRIP12	1 (0.06)	0.60	617752 (AD)	p.R1643Q	36.0	*ctrip*	Yes	RB	LoF
TRPM6	0 (0.45)	0.86	602014 (AR)	p.T2011P p.A641E	12.2 17.8	*Trpm*	Yes	RB RB	LoF LoF
TSC2	1 (0.07)	1.03	613254 (AD)	p.R1557W	16.0	*gig*	Yes	OE	LoF
USP30	0 (0.66)	0.76	-	p.P200S	14.7	*Usp30*	No	RB	LoF

Nervous system disease (OMIM) *known lethal mutants

Table 2: Salient features of subjects with *GLRA2* variants.

Abbreviations are as follows: MRI, magnetic resonance imaging; EEG, electroencephalography, CADD, Combined Annotation Dependent Depletion.

Subject	1	2	3	4	5	6	7	8
***GLRA2* Variant (hg19, NM_001118886.2)**	c.887C>T, p.Thr296Met	c.887C>T, p.Thr296Met	c.887C>T, p.Thr296Met	c.887C>T, p.Thr296Met	c.140T>C, p.Phe47Ser	c.754C>T, p.Arg252Cys	c.862G>A, p. Ala288Thr	c.1186C>A, p. Pro396Thr
Inheritance	*De novo*	*De novo*	*De novo*	*De novo*	*De Novo*	Maternal	Maternal	Maternal
CADD Score	27	27	27	27	27.8	31	27.2	20.9
Gender	Female	Female	Female	Female	Female	Male	Male	Male
Age at most recent evaluation (years)	6.7	6.5	5.5	0.5	6.7	0.9	7	34
Developmental delay/intellectual disability	Yes	Yes	Yes	Yes	Yes	Yes	Yes, with regression	Yes
Hypotonia/incoordination	No	No	Yes, ataxic gait	Yes	No	Yes	Yes, ataxia	Yes
Autism spectrum disorder	No	Yes	No	N/A	No	N/A	Yes	Yes
Inattention/hyperactivity	Yes	Yes	No	N/A	Yes	N/A	No	No
Sleep disturbance	No	Yes	No	No	Yes	Yes	No	No
Microcephaly	No	Yes	Yes	Yes	No	No	No	No
Ocular features	Myopia, astigmatism and nystagmus (improved with age)	Nystagmus (improved with age)	Alternating exotropia, borderline opsoclonus	None	Strabismus, nystagmus (improved with age)	Strabismus	Myopia	Myopia, astigmatism
Epilepsy	No	Yes	No	Yes	Yes	No	Yes	No
EEG findings	Slow background suggestive of mild encephalopathy	Bilateral synchronized high amplitude spikes, Epileptic potentials	Normal	Slow background, infantile spasms, multifocal spikes during sleep	Infantile spasms, then normal interictal EEG	Not performed	Generalized slowing and generalized epileptiform discharges associated with myoclonic jerks	Not performed
Brain MRI findings	Normal	Delayed myelination, a small arachnoid cyst	Mild cortical atrophy, thinning of corpus collosum	Normal	Cortical and white matter atrophy, including vermian atrophy	Not performed	Minimally increased T2 signal intensity on the occipital lobes	Normal