OXFORD

Human Molecular Genetics, 2022, Vol. 00, 00, 1–10
https://doi.org/10.1093/hmg/ddac203
Advance access publication date 22 August 2022
Review Article

From bugs to bedside: functional annotation of human genetic variation for neurological disorders using invertebrate models

Melanie Mew [1], Kim A. Caldwell [1,2,3,4] and Guy A. Caldwell [1,3,4,*]

[1]Department of Biological Sciences, The University of Alabama, Tuscaloosa, AL 35487, USA
[2]Alabama Research Institute on Aging, The University of Alabama, Tuscaloosa, AL 35487, USA
[3]Center for Convergent Bioscience and Medicine, The University of Alabama, Tuscaloosa, AL 35487, USA
[4]Departments of Neurobiology and Neurology, Center for Neurodegeneration and Experimental Therapeutics, Nathan Shock Center of Excellence for Research in the Basic Biology of Aging, Heersink School of Medicine, University of Alabama at Birmingham, Birmingham, AL 35294, USA
*To whom correspondence should be addressed at: Department of Biological Sciences, The University of Alabama, Box 870344, Tuscaloosa, AL 35487-0344, USA; Tel: 1-205-348-1780; Fax: 1-205-348-1786; Email: gcaldwel@ua.edu

Abstract

The exponential accumulation of DNA sequencing data has opened new avenues for discovering the causative roles of single-nucleotide polymorphisms (SNPs) in neurological diseases. The opportunities emerging from this are staggering, yet only as good as our abilities to glean insights from this surplus of information. Whereas computational biology continues to improve with respect to predictions and molecular modeling, the differences between *in silico* and *in vivo* analysis remain substantial. Invertebrate *in vivo* model systems represent technically advanced, experimentally mature, high-throughput, efficient and cost-effective resources for investigating a disease. With a decades-long track record of enabling investigators to discern function from DNA, fly (*Drosophila*) and worm (*Caenorhabditis elegans*) models have never been better poised to serve as living engines of discovery. Both of these animals have already proven useful in the classification of genetic variants as either pathogenic or benign across a range of neurodevelopmental and neurodegenerative disorders—including autism spectrum disorders, ciliopathies, amyotrophic lateral sclerosis, Alzheimer's and Parkinson's disease. Pathogenic SNPs typically display distinctive phenotypes in functional assays when compared with null alleles and frequently lead to protein products with gain-of-function or partial loss-of-function properties that contribute to neurological disease pathogenesis. The utility of invertebrates is logically limited by overt differences in anatomical and physiological characteristics, and also the evolutionary distance in genome structure. Nevertheless, functional annotation of disease-SNPs using invertebrate models can expedite the process of assigning cellular and organismal consequences to mutations, ascertain insights into mechanisms of action, and accelerate therapeutic target discovery and drug development for neurological conditions.

Introduction

The genome of every individual contains around four million single-nucleotide polymorphisms (SNPs), about 10 000 of which may alter the amino acid sequence and resulting structure of a protein (1). Deciphering how these mutations contribute to genetic conditions is a challenge with wide-ranging implications for diagnosing, prognosing and developing treatments for disease. An expanding plethora of genetic associations to disease litter the biomedical research landscape. Diseases of the brain and nervous system in particular represent an exponentially increasing challenge; a near-complete lack of disease-modifying treatments for these disorders highlights the unmet need and urgency required for more rapid discovery and advancement. This calls for a combination of traditional approaches and innovative strategies that take full advantage of the available experimental options for discovery. Closing the gap between genomic linkage and causality is an unequivocally daunting proposition, and nothing less than a comprehensive assault will suffice to meet this societal burden.

The vast majority of human SNPs are currently classified as Variants of Unknown Significance (VUS), meaning they are potentially deleterious, but insufficient data exist to establish pathogenicity or benignity (2). Whereas a variety of *in silico* algorithms have been developed to predict the most likely effects of exonic SNPs, these programs largely base their results on sequence homology, evolutionary conservation of affected amino acids and, in the case of missense mutations, the biochemical properties of the wild type (WT) versus resulting residue (3). Investigations of the accuracy of bioinformatic algorithms for breast cancer-associated mutations show that they are not reliable enough to be clinically actionable, as they generate unacceptable rates of false-positive predictions of pathogenicity (4,5). On the other hand, compensated pathogenic deviations represent at least 4% of missense mutations involved in genetic disease and, by nature, frequently receive false-negative predictions from *in silico* programs (6). Furthermore, the accuracy of *in silico* prediction algorithms varies across races. African genomes, for example, are genetically diverse but underrepresented in genomics datasets, leading to a substantial bias in predictive software (7).

Evaluating the contributions of SNPs to health and disease is more urgent now than ever given the explosion of DNA sequencing data in recent years. The current cost of sequencing a human genome is around $1000 (8), and databases such as the Genome

Received: July 29, 2022. **Revised:** August 11, 2022. **Accepted:** August 17, 2022

Aggregation Database and UK Biobank now feature the sequences of hundreds of thousands of exomes, as well as tens to hundreds of thousands of genomes (9,10). Additionally, DNA sequencing for newborns may soon be the standard of care, given the potential for early identification of disease risk and subsequent intervention (11). Predicting disease-risk based on genetic information requires an accurate assessment of genetic mutations. Owing to the inconsistency and inaccuracy of current bioinformatic algorithms, studies in living systems are a necessary step towards evaluating the consequences of specific SNPs. Functional studies in well-established disease models that demonstrate deleteriousness, or a lack thereof, represent a powerful means for classifying VUS as either benign or pathogenic (12).

Invertebrate model systems have a substantial and largely unexploited potential to generate actionable data on genetic mutations in a rapid, high-throughput and cost-effective manner. Mammalian studies typically take many years, and these animal populations are expensive to maintain. In addition, mammalian models are more likely to possess paralogs that are functionally redundant and compensate for a deficit associated with a single mutated gene. Therefore, insights at the level of fundamental cellular and genetic processes are more readily parsed in invertebrate systems. Most disease research to date has focused on models with a null allele, gene knockdown or complementation-based strategies, which while important for uncovering disease etiology, fail to identify the contribution of specific genetic variants to pathology. Efforts to understand the functional impact of individual SNPs within human populations are expanding, and both fly (*Drosophila*) and nematode (*C. elegans*) models have already been employed in the successful characterization of specific disease-associated SNPs sourced from sequence analyses of patient DNA samples. Of course, challenges and limitations to the applicability of invertebrates for functional annotation of human variation exist (i.e. limited sequence conservation in non-coding DNA regions). Nevertheless, a variety of experimental strategies and technical approaches have been successfully applied towards this end for several disorders of the brain and nervous system, and this trend is likely to continue (Fig. 1).

Functional Annotation of Neurological Disease-Associated SNPs

Autism spectrum disorders

Autism spectrum disorders (ASD) are a set of complex neurodevelopmental disorders characterized by challenges in communication, social interaction and/or repetitive behavior (13). Twin studies indicate that genetics play a significant role in developing ASD, as monozygotic twin studies report concordance rates of 36–90% (14). Aside from rare cases of familial mutations strongly associated with ASD, the genetic underpinnings of ASD are largely unknown. Many genes and mutations have been associated with ASD through genome-wide association studies (GWAS), which identify genes associated with disease in a hypothesis-free manner by comparing genetic information in affected versus unaffected groups. Validation of the functional involvement of these genes using *in vivo* models can then firmly establish and elucidate the genetic foundations of disease. One key finding of ASD genetics research has been that affected individuals tend to harbor a greater-than-expected number of *de novo* mutations (DNMs) in certain genes (15–17), and this is thought to constitute about 30% of ASD causation (17–20).

Marcogliese *et al*. (21) used *Drosophila* to explore the genetic foundations of ASD by generating a total of 79 ASD-associated DNMs across 74 genes in human cDNAs. These mutant cDNAs were expressed in flies via the GAL4-*UAS* system, a method that allows for the expression of any gene of interest (including fly, human, etc.) in the exact spatiotemporal pattern of an endogenous *Drosophila* gene. The models used in this research expressed WT or mutated human genes under the orthologous promoters of fly genes in a loss-of-function (LoF) background. This allowed for the functional evaluation of human genes in a *Drosophila* model without interference from the endogenous fly ortholog. The study design used an unbiased strategy, as variants were not pre-screened with *in silico* programs. Mutations of interest were only eliminated from the study if the necessary fly lines could not be generated. This screen identified a total of 30 mutations that altered protein function as compared with controls. These mutants displayed a range of developmental and behavioral phenotypes, thereby correlating human DNMs associated with ASD to dysfunctional protein activity *in vivo*. Interestingly, there was no relationship between *in silico* pathogenicity predictions (completed after the functional assays) and the likelihood that a given variant would affect the protein function.

As a further step towards functional annotation of disease-associated SNPs, Marcogliese *et al*. (21) identified a cohort of individuals, some with and some without ASD, who have an uncharacterized neurodevelopmental disorder and harbor mutations in *GLRA2*, an X-linked gene encoding a glycine receptor subunit. Two additional rare missense variants in the *GLRA2* gene were generated in flies; one increased the function of this protein, whereas the other behaved as an LoF allele. Thus, Marcogliese *et al*. (21) functionally tied rare SNPs to ASD and used their humanized *Drosophila* model to link *GLRA2* mutations to a neurodevelopmental disorder.

Caenorhabditis elegans models have also been used to investigate ASD-associated missense mutations. Wong *et al*. (22) examined a list of nearly 2000 mutations across hundreds of human genes associated with ASD and prioritized targets for experimental evaluation that either contained multiple variants or functioned in more than one set of cellular processes. They ultimately modeled 20 missense mutations identified within 11 genes involved in pathways such as synaptic function, gene regulation, neuronal signaling and cytoskeleton function. Mutations were knocked-in to corresponding *C. elegans* orthologs using CRISPR/Cas9 gene editing. To evaluate the effects of these SNPs, phenotypic changes in morphology, locomotion and fecundity were examined. Ultimately, this study established that 14 out of 20 (70%) of the modelled mutations affected the protein function, and ∼30% of the predicted-pathogenic mutations were most likely benign. Compared with null alleles, most of the missense mutations produced less severe phenotypes, underscoring the strength of this experimental strategy to enable a more refined analysis of the effects of mutation on the protein function.

Through large-scale phenomics, McDiarmid *et al*. (23) screened ∼100 orthologs of ASD-associated genes in *C. elegans* for learning, tactile sensitivity and locomotor phenotypes using an automated worm-tracking program that monitored 26 independent phenotypes simultaneously. For genes whose null mutations generated measurable phenotypes, CRISPR/Cas9 was used to introduce rare mutations found in ASD patients at the conserved orthologous position in the *C. elegans* genome. This deep phenotypic investigation found that each mutation affected one or more phenotype to a significant but lesser degree than in the null mutant. Most mutations led to developmental, learning or growth defects. The

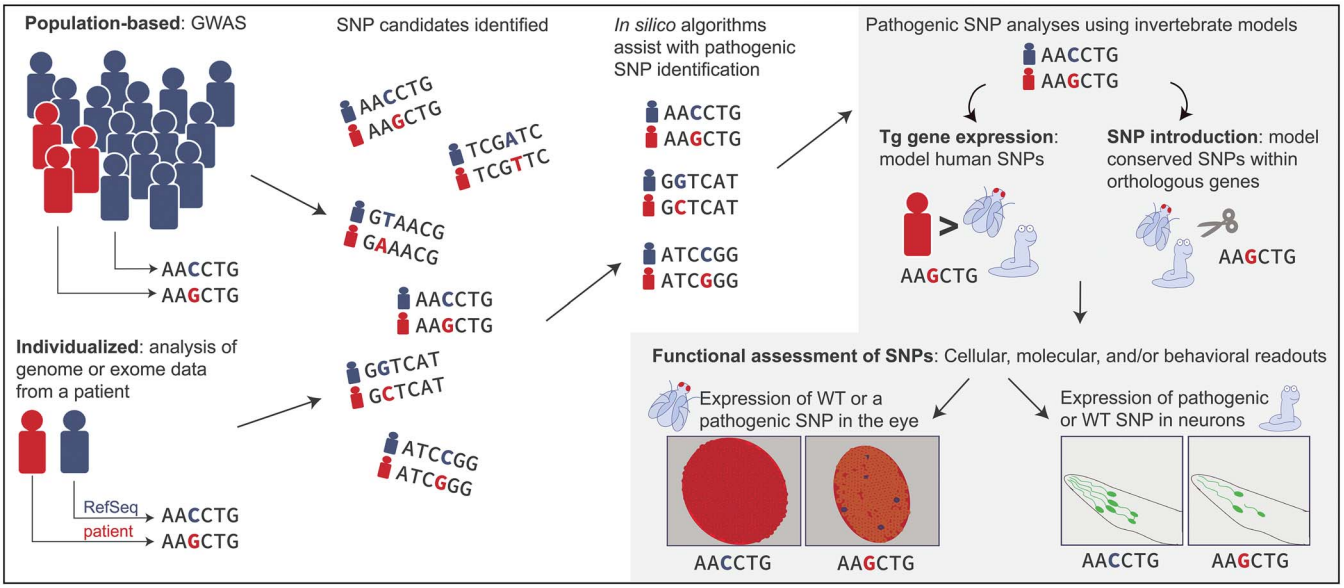

Figure 1. Scheme for the functional assessment of human disease SNPs within invertebrate models. GWAS use population-based metrics to discover disease-associated SNPs. Moreover, genome or exome sequencing data from individual patients can also be analyzed to identify putative disease-causing mutations. Following identification of SNP candidates from humans, *in silico* algorithms can optionally be used to screen for pathogenic mutations (with accuracy rates of ~70%); see Table 1 for example programs. SNPs of interest are then modelled in flies or worms using transgenic expression of human genes (which permits the investigation of any human mutation). Alternately, an SNP can be introduced using CRISPR editing at the conserved position of the orthologous gene (which facilitates the investigation of human mutations occurring at evolutionarily conserved sites). In this manner, human mutations can be functionally annotated via cellular, molecular and behavioral assays using gene-edited invertebrate strains. Examples displayed here represent ocular degeneration in *Drosophila* and dopaminergic neurodegeneration in *C. elegans*.

phenomics analyses performed in this study detailed the functional roles of previously uncharacterized genes and identified an overlooked avenue of potential treatment for long-term learning deficits in ASD individuals. Augmenting neuroligin expression in adult worms was found to rescue impaired habituation phenotypes, indicating that learning challenges may be treatable even once neurodevelopment is complete. Overall, this research established an experimental paradigm through which an array of phenotypic assays served to characterize individual SNPs and reveal the processes impacted by genomic variants.

Ciliopathies

Rare diseases (classified in the USA as those that affect fewer than 200 000 individuals) often receive much less funding and resources than common ones, leaving patients diagnosed with these disorders waiting for discoveries that could greatly improve their lives—including ones that we already have the tools to achieve (24). Given their cost-effectiveness and efficiency, invertebrate models represent attractive preclinical opportunities for investigating the genetic causes and possible treatment options for rare conditions such as ciliopathies. About 35 ciliopathies have been identified, and they collectively affect ~1 in 2000 individuals. These conditions arise from defects in cilia formation and typically affect neurological development and function, in addition to renal, hepatic and other processes (25). *Caenorhabditis elegans* have successfully been used to functionally annotate missense mutations implicated as the potential cause of ciliopathies. The *mksr-2* gene is named for Meckel syndrome—a devastating primary ciliopathy characterized by renal cysts and occipital encephalocele—but is also associated with Joubert syndrome (JBTS), which similarly presents with significant brain malformation (26). In 2021, Lange *et al.* (27) researched two missense variants in human *B9D2* (the *mksr-2* homolog) revealed by DNA sequencing of a JBTS patient. These mutations were generated at the homologous positions in worms using CRISPR/Cas9, and the resulting strains were crossed to obtain compound heterozygous worms that reflected the biallelic genotype of the specific JBTS patient. Assays were performed on these genome-edited worms to assess differences in the localization of MKSR-2, as well as cilia structure, positioning and function. Both modelled mutations recapitulated features of ciliopathies, functioning as hypomorphic alleles when homozygous. In addition, compound heterozygous worms displayed a reduced ability to uptake a lipophilic dye and limited avoidance of an osmotic gradient—two behaviors that rely on proper ciliary function. Ultimately, the two VUS found in a JBTS patient were determined to be pathogenic, demonstrating the utility of *C. elegans* models to provide direct functional annotation of SNPs found in patients.

In a later study, Lange *et al.* (28) explored the effects of various VUS in *TMEM67*, a ciliopathy-associated gene encoding a protein involved in cilia formation. CRISPR/Cas 9 was again used to introduce homozygous missense mutations in the *mks-3* gene, the *C. elegans* ortholog of *TMEM67*, and phenotypic characterization was performed as described above for *mksr-2*. These assays revealed that, out of eight predicted-pathogenic VUS examined, three were benign and five likely contributed to disease. This research illustrates the precise and efficient application of genome editing in *C. elegans* to rapidly ascribe functional significance to VUS in patients. It also exemplifies the importance of functional annotation to screen for pathogenic mutations, as algorithmic programs remain imperfect predictors of variant impact *in vivo*.

Morbidoni *et al.* (29) sought to further extend the degree to which individuals with a ciliopathy can attain a molecular diagnosis for their disease by characterizing the functional consequences of two missense mutations in the *TOGARAM1* gene that were hypothesized to cause a primary ciliopathy when inherited in a compound heterozygous manner. The mutations were

generated in *che-12*, the *C. elegans* ortholog of *TOGRAM1*, using CRISPR/Cas9. While the resulting worms performed similarly to WT animals in an NaCl-gradient chemotaxis assay, they were less able to uptake a lipophilic dye, and their sensory neurons featured shorter and phenotypically abnormal cilia. The results of this study identified the causative role of missense *TOGRAM1* variants in a primary ciliopathy, providing closure to the affected family whose genetics were investigated in this study and establishing two formerly VUS as pathogenic.

Other neurodevelopmental disorders

Similar to ciliopathies, mutations in key mitochondrial genes may produce pleiotropic effects, potentially causing the simultaneous onset of multiple neurological complications. Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) present concurrently in some cases, with mutations in the inner-mitochondrial membrane component gene *CHCHD10* being implicated in this phenomenon. Baek *et al.* (30) generated disease-associated SNPs in both codon-optimized human and *Drosophila* orthologs of the *CHCHD10* gene. They then used an eye development-specific promoter to drive overexpression of each *CHCHD10* version and compared the resulting phenotypes associated with mutated versus WT genes. The overexpression of either mutated form of *CHCHD10* induced a dominant pattern of cellular degeneration in fly eyes, though the endogenous *Drosophila* form was more toxic. A deeper investigation revealed that the missense mutation in *CHCHD10* contributes to mitochondrial dysfunction via the aberrant translocation of TDP-43—a feature shared with other subtypes of ALS-FTD—and that this is mediated by the mitophagy regulatory gene product, PINK1. This suggests that PINK1 may be a putative drug target for treatment of ALS-FTD. *In vivo* assays also revealed that having a single copy of WT *CHCHD10* may protect against disease, thereby indicating that an increase in the WT CHCHD10 activity may represent a viable treatment strategy for ALS-FTD. Overall, this research uses a *Drosophila* model to provide a clear illustration of disease mechanisms associated with mutated *CHCHD10* and ALS-FTD, which informs individuals of the risk of developing and potentially passing on this disease and illuminates a path for drug development.

The *GNAO1* gene encodes a component of a G-protein signal-transducing protein complex, and reduced function of this protein is associated with primary neurological disorders such as encephalopathy and dyskinesia. Di Rocco *et al.* (31) functionally annotated two human *GNAO1* variants in *C. elegans* by mutating its worm ortholog, *goa-1*, via CRISPR/Cas9. These variants generated hypomorphic alleles, and worms containing them displayed hyperactivity and uncoordinated movement. Aldicarb-induced paralysis assays found that this was likely mediated by excess acetylcholine released by ventral cord motor neurons. Mutated worms were then used to identify possible treatments for *GNAO1*-associated dyskinesia, and caffeine was found to improve motor impairments by inhibiting a putative adenosine receptor. This work exemplifies how the evaluation of disease-associated missense SNPs in *C. elegans* can rapidly inform subsequent drug discovery. Wang *et al.* (32) also examined the effects of three additional *GNAO1* variants (*G42R*, *G203R* and *R209C*) in *C. elegans* by mutating *goa-1* using CRISPR/Cas9. Locomotor assays revealed that these mutated worms exhibited hyperactivity, indicating limited functionality of the mutated proteins, and effectively classifying each variant studied as pathogenic. Importantly, this *in vivo* functional strategy using *C. elegans* revealed that these mutations result in a dominant negative LoF effect for GOA-1, a result that was corroborated for *GNAO1* by parallel experimental confirmation in mice (for *G42R*), resolving a controversy that had arisen from differential outcomes obtained from *in vitro* assays (32).

Neurodegenerative disorders

A significant challenge facing the biomedical research community is how to evaluate and treat neurodegenerative disorders. GWAS have already identified many key genetic drivers of neurodegenerative diseases such as Parkinson's disease (PD) (33) and Alzheimer's disease (AD) (34). Nevertheless, heritable contributors to PD and AD etiology remain unresolved, both in their extent of impact and function. Whereas characterization of disease-associated SNPs represents an end goal of functional annotation, invertebrate systems also enable prescreening of larger disease-associated gene sets for LoF or null phenotypes using RNA interference (RNAi). This approach can narrow down and prioritize putative targets of interest based on the phenotype, which reduces the time to discovery.

Shulman *et al.* (35) used RNAi to screen 84 fly orthologs of 67 genes associated with AD through GWAS. By knocking down expression of these genes in conjunction with overexpression of mutated human tau in the developing eye, they identified nine genes that can modify tau pathology. WT flies experienced normal eye development, whereas those expressing mutant tau showed evidence of ocular degeneration. Knockdown of these nine targets reduced the degenerative effects of tau expression, indicating that each of these genes helps mediate the neurotoxicity of tau. Subsequent investigations have further validated several of these genes as contributive to AD pathology. A *Drosophila* gene, *cindr*, encodes a homolog of the human scaffolding protein, CD2-Associated Protein (CD2AP), which is found to colocalize with tau protein in neuronal inclusions of AD patient brain samples (36). Cindr is required for synaptic vesicle release and recycling and regulates the proteasome and neuronal calcium levels (37). *FERMT2* has been shown to be downregulated in AD patients and plays a role in axon guidance, synaptic connections and long-term signal potentiation, in addition to regulating the metabolism of amyloid-precursor protein (38). A SNP in the *PTPRD* gene is significantly associated with the accumulation of neurofibrillary tangles in older individuals with AD (39). Additionally, the Ser/Thr kinase MAST4 participates in a cascade to promote cell survival following exposure to an oxidized cholesterol product but is downregulated in brain samples from AD patients (40). Overall, ocular neurodegeneration assays in *Drosophila* represent a widely applicable strategy for identifying disease-relevant genes and exploring how pathogenic mutations in one protein interact with others to cause disease (Fig. 1).

Another significant contributor to AD risk is the *APOE* gene, for which there are three alleles in the human population encoding slight variants of Apolipoprotein E, a cholesterol transport protein. The *APOEε2* allele is somewhat protective against AD, whereas *APOEε3* is neutral, and individuals with one or two *APOEε4* alleles have an increased risk of developing AD. Griffin *et al.* (41) generated a suite of transgenic *C. elegans* strains expressing the neurotoxic human amyloid-beta peptide [Aβ (1–42)] in combination with each of the different *APOE* alleles in glutamatergic neurons, a primary cell type affected in AD. These worm models were used to decipher potential molecular mechanisms of APOE contribution or protection from AD pathology. When Aβ peptide was co-expressed with distinct human *APOE* variants, the *ε2* allele resulted in the least glutamatergic neurodegeneration, followed by *ε3* and then *ε4*—recapitulating the overall trend of impact observed clinically in AD. Treatment with a drug to increase

cytosolic calcium ion levels reduced neurodegeneration in animals expressing Aβ and *APOE*ε4, but not ε2 or ε3, indicating that *APOE* alleles differentially affect intracellular calcium levels to modulate neuropathology. This research used *C. elegans* to distinguish between functional consequences associated with different *APOE* alleles on neurodegeneration (41) via directed expression of two major drivers of AD, Aβ and *APOE*—neither of which have a strict homolog in the worm genome. Notably, these same transgenic strains were utilized in a drug discovery effort where small molecule inducers of mitophagy were identified from a library of naturally occurring compounds using a workflow involving machine learning (42). Two molecules, kaempferol and rhapontigenin, were both found to prevent glutamatergic neuron loss in transgenic Aβ/APOE4 worms, and also ameliorated Aβ- and tau-induced toxicity in a series of other cellular and animal AD models. Additionally, the induction of mitophagy also attenuated memory loss in behavioral assays using *C. elegans*, as well as in the widely used 3xTg AD mouse model that expresses mutant forms of the Aβ-precursor protein, tau and presenilin (42).

Whereas GWAS offer a valuable, hypothesis-free approach to identifying genetic contributors to disease, limitations of sample size can preclude the discovery of potential disease-associated genes and/or SNPs owing to underpowered statistical analyses. Therefore, introducing hypothesis-based analysis by focusing statistical tests on specific pathways or genetic networks has the potential to increase the power of GWAS data analysis and further identify meaningful genetic relationships to disease. BridGE (Bridging Gene sets with Epistasis) is a GWAS analysis method to evaluate statistically significant enrichment in the concordance of hits among a specified set of related genes (43,44). This method was used by Hallacli *et al*. (45) to evaluate the relevance of SNPs in the human alpha-synuclein locus, *SNCA*, and a defined set of RNA processing-body (P body) genes. They identified 76 significant signals specifically associated with the PD cohort and validated them via RNAi knockdown experiments in *Drosophila*. The outcomes of this applied, hypothesis-based strategy enabled the discovery of correlations between target genes and locomotor deficits in an established PD gene (*SNCA*), in addition to the identification of unexpected genetic modifiers (P body genes), thereby broadening understanding of potential therapeutic targets for treating PD. The implication of P body components as disease effectors of PD highlights the likelihood of similar RNA-associated factors influencing other neurodegenerative diseases.

Transcriptional profiling via RNA sequencing (RNA-seq) has become a canonical strategy for evaluating differential gene expression as a principal readout for functional correlations to SNPs. Splice site mutations represent a significant source of phenotypic variation but have been largely excluded from functional characterization owing to the complex calculations required for their analysis. A new strategy accounts for the functional impact of splicing events as key modifiers of disease through a combined analysis of splicing and expression data. Recently, this approach was used to implicate genetic factors in AD and schizophrenia that would have been missed through RNA-seq or traditional GWAS alone (46).

Recent research has also identified disease-modifying functions of human genetic variants using *C. elegans* models of neurodegeneration. *VPS41* encodes a conserved protein necessary for lysosomal fusion events and regulated secretion of neuropeptides (47). VPS41 was initially identified as exhibiting neuroprotective activity using *C. elegans*, where it was determined to attenuate two major pathological hallmarks of PD: alpha-synuclein aggregation and dopamine neuron degeneration (48). After 12 years, as human genomic data acquisition matured, van der Welle *et al*. (49) reported a correlation between *VPS41* mutations identified across three patients with a neurodegenerative disorder inclusive of ataxia and dystonia symptoms. Accompanying functional studies on patient-derived SNPs in *C. elegans* demonstrated that the expression of at least one copy of normal *VPS41* retained neuroprotective activity, whereas transgenic co-expression of mutated human *VPS41* alleles designed to recapitulate a compound heterozygous state eliminated neuroprotection and increased dopaminergic neurodegeneration in older worms (Fig. 2). These results demonstrate the power of a strategy whereby combined use of invertebrate and human genetic investigation defined a new class of movement disorder (50). Moreover, this approach illustrates the utility of *C. elegans* both as a primary vehicle for disease gene discovery as well as for the functional analysis of genetic information gleaned directly from patients. Subsequent efforts using cell cultures and/or mammalian models are now informed with respect to underlying mechanistic aspects of disease etiology that can accelerate targeted drug discovery.

Challenges and Limitations

Although the numerous strengths and advantages of invertebrate genetic models are well established, the limitations of these systems are not insignificant. Aside from differences in nervous system complexity, anatomy and physiology, one drawback to invertebrate model systems is that many genes and amino acids residues in orthologous proteins are not strictly conserved between humans and flies or worms. Although homology between species is typically higher with respect to genes having functional significance to processes involved in disease (given their essentiality for the survival of living organisms), there are far more disease-contributing SNPs than can be modeled in orthologous genes of invertebrate animals. Using humanized, codon-optimized models that express human orthologs in endogenous LoF backgrounds can sometimes overcome this limitation. When possible, however, investigating conserved SNPs in endogenous proteins is beneficial because it avoids introducing intron disruption or isoform imbalance as confounding factors (51,52).

Another limitation, not unique to invertebrate disease modeling, is that use of GWAS datasets as a source of gene candidates for functional studies can introduce biases because of small differences in collection and processing of samples (53). Therefore, judicious analysis of GWAS results is a critical prerequisite to functional studies. Additionally, variant selection schemes and computational processes for *in vivo* studies can pose a challenge to the thorough evaluation of SNPs found in disease-associated genes. The selection of variants that are predicted to be deleterious by *in silico* algorithms may lead to disease-relevant SNPs being overlooked for functional annotation. In the prior ciliopathy example, Lange *et al*. (28) found that two mutations predicted to be benign by computational methods actually impacted protein function *in vivo*. In such cases, experimentally repeating or revisiting variant investigation using worms or flies can be far less financially burdensome, and more forgiving, when mistakes are inevitably made.

Conclusions and Future Directions

The advantages of using invertebrate model systems to investigate the contribution of individual SNPs to disease pathogenesis warrant serious consideration in preclinical application. We now

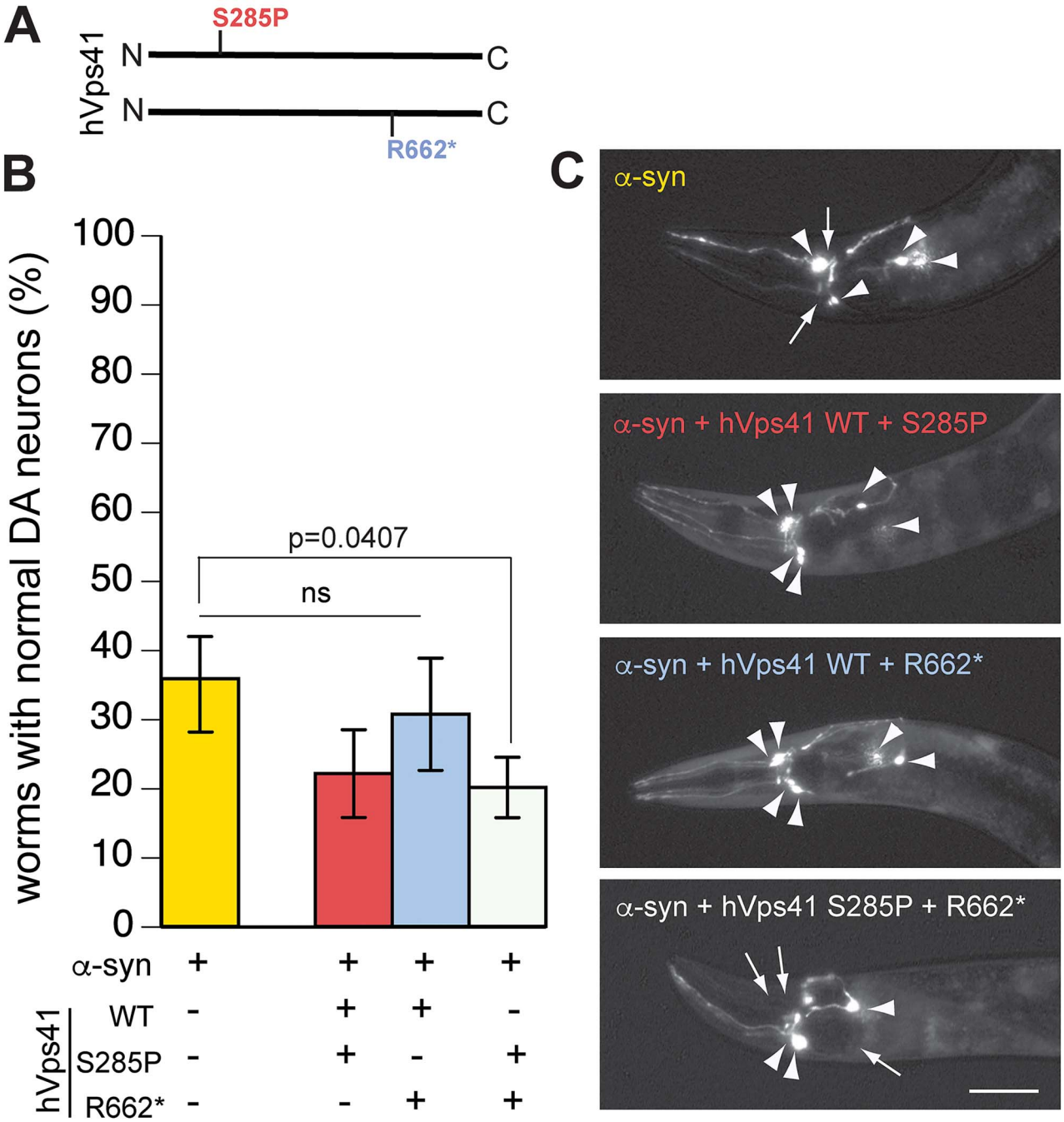

Figure 2. Neurodegeneration caused by transgenic expression of human SNPs in *C. elegans*. SNPs in the human *VPS41* gene, encoding a conserved endolysosomal fusion and trafficking protein (hVPS41), cause a newly defined class of movement disorder that includes symptoms of neurodegeneration, a phenotype revealed initially through functional analysis using transgenic *C. elegans* strains (49). The genetic lesions in patients occur in a compound heterozygous orientation and include the two exemplar mutations depicted in the line diagram (**A**). To examine the impact of *VPS41* patient-derived SNPs on neurodegeneration, transgenic worms expressing human *VPS41* variants exclusively in the *C. elegans* dopamine neurons were crossed to animals overexpressing human alpha-synuclein in the same cells. The resulting strains, which also express GFP to illuminate the dopamine neurons, were scored for the evidence of neurodegeneration (**B**, **C**). Approximately 35% of transgenic worms overexpressing alpha-synuclein alone display the normal complement of six anterior dopamine neurons (the remaining 65% exhibit a loss of neuronal processes and cell bodies). At the same point in lifespan, co-expression of WT hVPS41 did not significantly alter the amount of alpha-synuclein-dependent neurodegeneration (not shown). Similarly, if either one of the two *VPS41* SNPs was co-expressed with WT *VPS41* in the dopamine neurons (i.e. WT + S285P or WT + R662*—stop codon), no significant change in alpha-synuclein-induced neurodegeneration was observed. Only co-expression of both human *VPS41* SNPs (S285P and R662*) resulted in more significant dopaminergic neurodegeneration. One-way analysis of variance with Tukey post-hoc test; day 10, post-hatching analysis. Arrowheads indicate intact dopamine neurons; arrows signify a degenerating neuron. Scale bar = 20 μm.

Table 1. Resources available for functional annotation research with fly and worm models

Program/site	Purpose	Description
WormAtlas https://www.wormatlas.org/	Behavioral and anatomical database for nematodes (including *C. elegans*)	Provides images, descriptions and links to external resources for nematode anatomy and behavior
Wormbase https://wormbase.org/	Genetics and anatomy database for *C. elegans* and other nematodes	Provides detailed information about *C. elegans* genetics, strains, phenotypes and biology, as well as links to additional resources
CeNGEN (*C. elegans* Neuronal Gene Expression Map & Network) https://www.cengen.org/	Resource for *C. elegans* nervous system gene expression data	Provides gene expression data for the complete nervous system of *C. elegans*
Flybase https://flybase.org/	*Drosophila* genetics and biology database	Provides detailed information about *Drosophila* genetics and biology, as well as links to additional resources
Fruit Fly Brain Observatory https://www.fruitflybrain.org/#/	Resource for *Drosophila* nervous system information	Integrated resource for *Drosophila* neurobiology research and information
Genome Aggregation Database (gnomAD) https://gnomad.broadinstitute.org/	Genome/exome sequence database	Provides summary data from WGS and WES projects
PolyPhen-2 http://genetics.bwh.harvard.edu/pph2/	*In silico* prediction algorithm	Predicts SNP conservation and pathogenicity
SIFT (Sorting Intolerant From Tolerant) https://sift.bii.a-star.edu.sg/	*In silico* prediction algorithm	Predicts SNP pathogenicity based on residue conservation and biochemistry
MutationTaster https://www.mutationtaster.org/	*In silico* prediction algorithm	Predicts SNP pathogenicity based on conservation, biochemistry, protein length and potential effects on splicing and non-coding regions, etc.
Mutation Assessor http://mutationassessor.org/r3/	*In silico* prediction algorithm	Predicts SNP pathogenicity based on amino acid conservation, biochemistry and 3D structure
FATHMM (Functional Analysis Through Hidden Markov Models) http://fathmm.biocompute.org.uk/	*In silico* prediction algorithm	Predict SNPs pathogenicity using conservation, biochemistry and Markov models
dbNSFP (Database for Nonsynonymous SNPs Functional Predictions) http://database.liulab.science/dbNSFP	Combines the prediction scores of abovementioned *in silico* algorithms into a single score	Provides a single pathogenicity score from multiple *in silico* programs; also considers the likelihood ratio test
CADD (Combined Annotation Dependent Depletion) https://cadd.gs.washington.edu/	*In silico* prediction algorithm	Predicts SNP pathogenicity using machine learning, conservation, epigenetics and predictions
REVEL (Rare Exome Variant Ensemble Learner) https://sites.google.com/site/revelgenomics/	Combines prediction scores of 13 algorithms into a single score	One pathogenicity score from multiple programs

have the tools to fully detail the exome and even genome of any individual for a reasonable price and within a rapid timeframe. The primary impediment to elucidating disease mechanisms associated with specific mutations is the investigation of such changes in living model systems, which itself poses a frequently time-consuming and expensive task. These challenges can be mitigated with the use of *Drosophila* and *C. elegans* as bio-processors of genetic information into translational outputs through which bona fide functional outcomes enable consequence to be discerned from otherwise superficial associations. The wide array of anatomical maps, genomic information, mature database resources and molecular genetic tools available for these organisms, taken with inherently quick generation times, manageable lifespans and affordable maintenance render them ideal candidates for high throughput and detailed assays of putative disease-causing mutations (Table 1). Leveraging these models allows us to gain valuable insights into how certain SNPs participate in disease pathology. Across different models, missense mutations tend to cause milder and/or fewer phenotypes compared with null mutants (22). Therefore, missense mutations possess functional distinctions that might be missed through common means of investigation but can be parsed using *in vivo* invertebrate models.

Functional annotation of disease SNPs using *in vivo* models is a necessity, as *in silico* analyses will require validation for the foreseeable future. Notably, invertebrate models have the added bioethical advantages that the research community seeks to instill under the banner of the 3Rs (*Replace, Reduce, Refine*) that advocates a responsible use of rodent and primate models, as well as a concerted shift towards less sentient organisms (54). Microscopic worms with a complete nervous system comprised of ~300 neurons fulfill any such aspirational criteria, especially since *C. elegans* research in areas like neurodegeneration has already proven to reproducibly yield results with translational value (55). Similarly, experiments with *Drosophila* will unequivocally continue to elucidate complex genetic, epigenetic and gene-by-environment interactions that contribute to disease. Having already enabled a century of countless pioneering discoveries, fruit flies will likely continue to lead the way as a primary alternative to mammals for generating *in vivo* personalized models, including those that facilitate combinational analysis of genetic factors and environmental exposures (56).

As invertebrate SNP modeling becomes more mainstream, the genetic resources and phenotypic information accrued will hasten functional evaluation of SNP combinations that more accurately reflect the genetic context of patients, and more fully elucidate how SNPs affect mechanisms of disease. For example, using *C. elegans*, Jean *et al.* (57) explored how second-site variants may further modify disease susceptibility, specifically examining SNPs in *zyg-1/PLK4*, a gene linked to microcephaly, primordial dwarfism and chorioretinopathy. Whole-genome sequencing of each mutagenized strain generated throughout the study revealed that the effects of deleterious mutations may be mitigated by other variants within the same gene, and that the broader genomic context is also important for determining phenotypic outcomes. This research encourages a paradigm-shift towards the investigation of any given mutation across varying genetic backgrounds to more accurately reflect the complex biological interactions that occur in different individuals, bolstering the prospect of using invertebrates to investigate combinatorial consequences of distinct mutations within a single disease-linked gene. Once mutations are moved from the status of VUS to the status of pathogenic, they become legitimate targets for the development of drugs and treatments. Additionally, the models used to establish pathogenicity of SNPs can also be employed for drug screens to potentially treat the diseases caused by them.

The examples covered in this review share an approach to understanding disease that embraces the complexity and diversity of mechanisms of neuropathogenesis. It should be noted, however, that we highlight examples of disease-SNPs in exonic regions, whereas the overwhelming majority of human genomic variation lies within regulatory sequences and other non-coding regions of the genome. SNPs in non-coding DNA also represent the main class of mutations typically identified in GWAS and can both contribute to, or cause, disease (58). Since genomic variation in non-coding regions is primarily associated with changes in gene expression, the collective outcomes of the ENCODE and ModENCODE consortia to map transcriptional regulatory sequences in mammals and invertebrates, respectively, provide an extensive resource for hypothesis building and experimentation (59). Importantly, the evolutionary distance and divergence between invertebrate and human non-coding DNA is not absolute. This is exemplified by sequences termed conserved non-coding elements (CNEs) that are retained universally among distant taxa metazoans and can even exhibit stronger sequence similarity than some coding sequences across different species (60). CNEs have thus far been demonstrated to have developmental roles and are implicated in epigenetic regulation, including chromatin remodeling and RNA editing. Significantly, polymorphisms in these highly conserved elements have also been linked to neurodevelopmental disorders and cancers (61). Bioinformatic resources, such as Functional Identification of Non-coding Sequences Using Random Forests (FINSURF), have recently been developed to facilitate identification of likely pathogenic mutations located outside of protein-coding regions of the human genome (62). CNEs across species share common genomic structural arrangements in syntenic blocks, indicating a tightly constrained evolution of their putative *cis*-regulatory role in transcription (63). Although significant distinctions between invertebrate and vertebrate CNEs still exist, the annotation of variation in conserved regions of human non-coding DNA will benefit from the numerous tools and depth of understanding that both *C. elegans* and *Drosophila* provide as preeminent models for investigating developmental biology and epigenetic phenomena.

Predictions of disease risk along the continuum of penetrance that exists for most neurologic disorders are confounded by the unpredictable factors of epigenetics and, of course, age (64); the tragic prevalence of neurodegenerative diseases negates the luxury of time for advancing understanding and implementing therapeutic innovations. Technical advances notwithstanding, the ability to provide a correct diagnosis for patients, with precision and consistency, is the ultimate goal of personalized medicine. This idyllic standard will only be achieved through the systematic investigation of the functional consequences that result from thousands of genomic variants. Although this unequivocally represents a formidable challenge for human geneticists, experimental researchers, clinicians and bioinformaticians, it is an undertaking that will come closer to fruition—one worm or fly at a time.

Acknowledgements

We would like to thank all the scientists whose fine research we have highlighted here, with sincere apologies to others that we have omitted, either inadvertently or owing to space constraints. We are grateful to Kevin Bowling for his advice and astute suggestions. Special thanks to Dan Geschwind and Helen Johnson for the opportunity to contribute our article in time for inclusion in this exciting collection of timely topical reviews.

Conflict of Interest statement. None declared.

Funding

National Institutes of Health (R15NS106460 to K.A.C., R15NS104857 to G.A.C.).

References

1. Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A. and Abecasis, G.R. (2015) A global reference for human genetic variation. *Nature*, **526**, 68–74.
2. Cooper, G.M. (2015) Parlez-vous VUS? *Genome Res.*, **25**, 1423–1426.
3. Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K. and Liu, X. (2015) Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.*, **24**, 2125–2137.
4. Ernst, C., Hahnen, E., Engel, C., Nothnagel, M., Weber, J., Schmutzler, R.K. and Hauke, J. (2018) Performance of *in silico* prediction tools for the classification of rare *BRCA1/2* missense variants in clinical diagnostics. *BMC Med. Genet.*, **11**, 35.
5. Voskanian, A., Katsonis, P., Lichtarge, O., Pejaver, V., Radivojac, P., Mooney, S.D., Capriotti, E., Bromberg, Y., Wang, Y., Miller, M. *et al.* (2019) Assessing the performance of in silico methods for predicting the pathogenicity of variants in the gene CHEK2, among Hispanic females with breast cancer. *Hum. Mutat.*, **40**, 1612–1622.
6. Azevedo, L., Mort, M., Costa, A.C., Silva, R.M., Quelhas, D., Amorim, A. and Cooper, D.N. (2016) Improving the in silico assessment of pathogenicity for compensated variants. *Eur. J. Hum. Genet.*, **25**, 2–7.
7. Bope, C.D., Chimusa, E.R., Nembaware, V., Mazandu, G.K., de Vries, J. and Wonkam, A. (2019) Dissecting *in silico* mutation prediction of variants in African genomes: challenges and perspectives. *Front. Genet.*, **10**, 601.
8. Wetterstrand, K.A. (Accessed 24 July 2022) *DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)*. National Human Genome Research Institute, Bethesda, MD, USA.

https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost.
9. Koch, L. (2020) Exploring human genomic diversity with gnomAD. *Nat. Rev. Genet.*, **21**, 448.
10. Halldorsson, B.V., Eggertsson, H.P., Moore, K.H.S., Hauswedell, H., Eiriksson, O., Ulfarsson, M.O., Palsson, G., Hardarson, M.T., Oddsson, A., Jensson, B.O. *et al.* (2022) The sequences of 150,119 genomes in the UK biobank. *Nature*, **607**, 732–740.
11. Holm, I.A., Agrawal, P.B., Ceyhan-Birsoy, O., Christensen, K.D., Fayer, S., Frankel, L.A., Genetti, C.A., Krier, J.B., LaMay, R.C., Levy, H.L. *et al.* (2018) The BabySeq project: implementing genomic sequencing in newborns. *BMC Pediatr.*, **18**, 225.
12. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E. *et al.* (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.*, **17**, 405–424.
13. Fakhoury, M. (2015) Autistic spectrum disorders: a review of clinical features, theories and diagnosis. *Int. J. Dev. Neurosci.*, **43**, 70–77.
14. Ronald, A. and Hoekstra, R.A. (2011) Autism spectrum disorders and autistic traits: a decade of new twin studies. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, **156B**, 255–274.
15. Coe, B.P., Stessman, H.A.F., Sulovari, A., Geisheker, M.R., Bakken, T.E., Lake, A.M., Dougherty, J.D., Lein, E.S., Hormozdiari, F., Bernier, R.A. and Eichler, E.E. (2019) Neurodevelopmental disease genes implicated by de novo mutation and copy number variation morbidity. *Nat. Genet.*, **51**, 106–116.
16. Fischbach, G.D. and Lord, C. (2010) The Simons simplex collection: a resource for identification of autism genetic risk factors. *Neuron*, **68**, 192–195.
17. Iossifov, I., O'Roak, B.J., Sanders, S.J., Ronemus, M., Krumm, N., Levy, D., Stessman, H.A., Witherspoon, K.T., Vives, L., Patterson, K.E. *et al.* (2014) The contribution of de novo coding mutations to autism spectrum disorder. *Nature*, **515**, 216–221.
18. Satterstrom, F.K., Kosmicki, J.A., Wang, J., Breen, M.S., De Rubeis, S., An, J.Y., Peng, M., Collins, R., Grove, J., Klei, L. *et al.* (2020) Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell*, **180**, 568–584.e23.
19. Takata, A., Miyake, N., Tsurusaki, Y., Fukai, R., Miyatake, S., Koshimizu, E., Kushima, I., Okada, T., Morikawa, M., Uno, Y. *et al.* (2018) Integrative analyses of de novo mutations provide deeper biological insights into autism spectrum disorder. *Cell Rep.*, **22**, 734–747.
20. Yuen, R.K.C., Merico, D., Bookman, M., Howe, J.L., Thiruvahindrapuram, B., Patel, R.V., Whitney, J., Deflaux, N., Bingham, J., Wang, Z. *et al.* (2017) Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat. Neurosci.*, **20**, 602–611.
21. Marcogliese, P.C., Deal, S.L., Andrews, J., Harnish, J.M., Bhavana, V.H., Graves, H.K., Jangam, S., Luo, X., Liu, N., Bei, D. *et al.* (2022) *Drosophila* functional screening of de novo variants in autism uncovers damaging variants and facilitates discovery of rare neurodevelopmental diseases. *Cell Rep.*, **38**, 110517.
22. Wong, W.R., Brugman, K.I., Maher, S., Oh, J.Y., Howe, K., Kato, M. and Sternberg, P.W. (2019) Autism-associated missense genetic variants impact locomotion and neurodevelopment in *Caenorhabditis elegans*. *Hum. Mol. Genet.*, **28**, 2271–2281.
23. McDiarmid, T.A., Belmadani, M., Liang, J., Meili, F., Mathews, E.A., Mullen, G.P., Hendi, A., Wong, W.R., Rand, J.B., Mizumoto, K. *et al.* (2019) Systematic phenomics analysis of autism-associated genes reveals parallel networks underlying reversible impairments in habituation. *Proc. Natl. Acad. Sci.*, **117**, 656–667.
24. Kaufmann, P., Pariser, A.R. and Austin, C. (2018) From scientific discovery to treatments for rare diseases – the view from the National Center for Advancing Translational Sciences – Office of Rare Diseases Research. *Orphanet J. Rare Dis.*, **13**, 196.
25. Kagan, K.O., Dufke, A. and Gembruch, U. (2017) Renal cystic disease and associated ciliopathies. *Curr. Opin. Obstet. Gynecol.*, **29**, 85–94.
26. Parisi, M.A. (2009) Clinical and molecular features of Joubert syndrome and related disorders. *Am. J. Med. Genet. C Semin. Med. Genet.*, **151C**, 326–340.
27. Lange, K.I., Tsiropoulou, S., Kucharska, K. and Blacque, O.E. (2021) Interpreting the pathogenicity of Joubert syndrome missense variants in *Caenorhabditis elegans*. *Dis. Model. Mech.*, **14**, dmm046631.
28. Lange, K.I., Best, S., Tsiropoulou, S., Berry, I., Johnson, C.A. and Blacque, O.E. (2022) Interpreting ciliopathy-associated missense variants of uncertain significance (VUS) in *Caenorhabditis elegans*. *Hum. Mol. Genet.*, **31**, 1574–1587.
29. Morbidoni, V., Agolini, E., Slep, K.C., Pannone, L., Zuccarello, D., Cassina, M., Grosso, E., Gai, G., Salviati, L., Dallapiccola, B. *et al.* (2021) Biallelic mutations in the TOGARAM1 gene cause a novel primary ciliopathy. *J. Med. Genet.*, **58**, 526–533.
30. Baek, M., Choe, Y.J., Bannwarth, S., Kim, J., Maitra, S., Dorn, G.W., II, Taylor, J.P., Paquis-Flucklinger, V. and Kim, N.C. (2021) TDP-43 and PINK1 mediate CHCHD10^{S59L} mutation-induced defects in *Drosophila* and in vitro. *Nat. Commun.*, **12**, 1924.
31. Di Rocco, M., Galosi, S., Lanza, E., Tosato, F., Caprini, D., Folli, V., Friedman, J., Bocchinfuso, G., Martire, A., Di Schiavi, E. *et al.* (2022) Martinelli S. *Caenorhabditis elegans* provides an efficient drug screening platform for GNAO1-related disorders and highlights the potential role of caffeine in controlling dyskinesia. *Hum. Mol. Genet.*, **31**, 929–941.
32. Wang, D., Dao, M., Muntean, B.S., Giles, A.C., Martemyanov, K.A. and Grill, B. (2022) Genetic modeling of GNAO1 disorder delineates mechanisms of Gαo dysfunction. *Hum. Mol. Genet.*, **31**, 510–522.
33. Grenn, F.P., Kim, J.J., Makarious, M.B., Iwaki, H., Illarionova, A., Brolin, K., Kluss, J.H., Schumacher-Schuh, A.F., Leonard, H., Faghri, K. *et al.* (2020) The Parkinson's disease genome-wide association study locus browser. *Mov. Disord.*, **35**, 2056–2067.
34. Wightman, D.P., Jansen, I.E., Savage, J.E., Shadrin, A.A., Bahrami, S., Holland, D., Rongve, A., Børte, S., Winsvold, B.S., Drange, O.K. *et al.* (2021) A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat. Genet.*, **53**, 1276–1282.
35. Shulman, J.M., Imboywa, S., Giagtzoglou, N., Powers, M.P., Hu, Y., Devenport, D., Chipendo, P., Chibnik, L.B., Diamond, A., Perrimon, N. *et al.* (2014) Functional screening in *Drosophila* identifies Alzheimer's disease susceptibility genes and implicates tau-mediated mechanisms. *Hum. Mol. Genet.*, **23**, 870–877.
36. Camacho, J., Rábano, A., Marazuela, P., Bonaterra-Pastra, A., Serna, G., Moliné, T., Cajal, S.R.Y., Martínez-Sáez, E. and Hernández-Guillamon, M. (2022) Association of CD2AP neuronal deposits with Braak neurofibrillary stage in Alzheimer's disease. *Brain Pathol.*, **32**, e13016.
37. Ojelade, S.A., Lee, T.V., Giagtzoglou, N., Yu, L., Ugur, B., Li, Y., Duraine, L., Zuo, Z., Petyuk, V., De Jager, P.L. *et al.* (2019) Cindr, the *Drosophila* homolog of the CD2AP Alzheimer's disease risk gene, is required for synaptic transmission and proteostasis. *Cell Rep.*, **28**, 1799–1813.e5.

38. Eysert, F., Coulon, A., Boscher, E., Vreulx, A.C., Flaig, A., Mendes, T., Hughes, S., Grenier-Boley, B., Hanoulle, X., Demiautte, F. *et al.* (2021) Alzheimer's genetic risk factor FERMT2 (Kindlin-2) controls axonal growth and synaptic plasticity in an APP-dependent manner. *Mol. Psychiatry*, **26**, 5592–5607.
39. Chibnik, L.B., White, C.C., Mukherjee, S., Raj, T., Yu, L., Larson, E.B., Montine, T.J., Keene, C.D., Sonnen, J., Schneider, J.A. *et al.* (2018) Susceptibility to neurofibrillary tangles: role of the PTPRD locus and limited pleiotropy with other neuropathologies. *Mol. Psychiatry*, **23**, 1521–1529.
40. Gongol, B., Marin, T.L., Jeppson, J.D., Mayagoitia, K., Shin, S., Sanchez, N., Kirsch, W.M., Vinters, H.V., Wilson, C.G., Ghribi, O. and Soriano, S. (2017) Cellular hormetic response to 27-hydroxycholesterol promotes neuroprotection through AICD induction of MAST4 abundance and kinase activity. *Sci. Rep.*, **7**, 13898.
41. Griffin, E.F., Scopel, S.E., Stephen, C.A., Holzhauer, A.C., Vaji, M.A., Tuckey, R.A., Berkowitz, L.A., Caldwell, K.A. and Caldwell, G.A. (2019) ApoE-associated modulation of neuroprotection from Aβ-mediated neurodegeneration in transgenic. *Dis. Model. Mech.*, **12**, dmm037218.
42. Xie, C., Zhuang, X.X., Niu, Z., Ai, R., Lautrup, S., Zheng, S., Jiang, Y., Han, R., Gupta, T.S., Cao, S. *et al.* (2022) Amelioration of Alzheimer's disease pathology by mitophagy inducers identified via machine learning and a cross-species workflow. *Nat. Biomed. Eng.*, **6**, 76–93.
43. Fang, G., Wang, W., Paunic, V., Heydari, H., Costanzo, M., Liu, X., Liu, X., VanderSluis, B., Oately, B., Steinbach, M. *et al.* (2019) Discovering genetic interactions bridging pathways in genome-wide association studies. *Nat. Commun.*, **10**, 4274.
44. Wang, W., Xu, Z.Z., Costanzo, M., Boone, C., Lange, C.A. and Myers, C.L. (2017) Pathway-based discovery of genetic interactions in breast cancer. *PLoS Genet.*, **13**, e1006973.
45. Hallacli, E., Kayatekin, C., Nazeen, S., Wang, X.H., Sheinkopf, Z., Sathyakumar, S., Sarkar, S., Jiang, X., Dong, X., Di Maio, R. *et al.* (2022) The Parkinson's disease protein alpha-synuclein is a modulator of processing bodies and mRNA stability. *Cell*, **185**, 2035–2056.e33.
46. Ji, Y., Wei, Q., Chen, R., Wang, Q., Tao, R. and Li, B. (2022) Integration of multidimensional splicing data and GWAS summary statistics for risk gene discovery. *PLoS Genet.*, **18**, e1009814.
47. Asensio, C.S., Sirkis, D.W., Maas, J.W., Jr., Egami, K., To, T.L., Brodsky, F.M., Shu, X., Cheng, Y. and Edwards, R.H. (2013) Self-assembly of VPS41 promotes sorting required for biogenesis of the regulated secretory pathway. *Dev. Cell*, **27**, 425–437.
48. Hamamichi, S., Rivas, R.N., Knight, A.L., Cao, S., Caldwell, K.A. and Caldwell, G.A. (2008) Hypothesis-based RNAi screening identifies neuroprotective genes in a Parkinson's disease model. *Proc. Natl. Acad. Sci.*, **105**, 728–733.
49. van der Welle, R.E.N., Jobling, R., Burns, C., Sanza, P., van der Beek, J.A., Fasano, A., Chen, L., Zwartkruis, F.J., Zwakenberg, S., Griffin, E.F. *et al.* (2021) Neurodegenerative VPS41 variants inhibit HOPS function and mTORC1-dependent TFEB/TFE3 regulation. *EMBO Mol. Med.*, **13**, 13258.
50. Monfrini, E., Zech, M., Steel, D., Kurian, M.A., Winkelmann, J. and Di Fonzo, A. (2021) HOPS-associated neurological disorders (HOPSANDs): linking endolysosomal dysfunction to the pathogenesis of dystonia. *Brain*, **144**, 2610–2615.
51. Reble, E., Dineen, A. and Barr, C.L. (2018) The contribution of alternative splicing to genetic risk for psychiatric disorders. *Genes Brain Behav.*, **17**, e12430.
52. Robison, A.J. (2014) Emerging role of CaMKII in neuropsychiatric disease. *Trends Neurosci.*, **37**, 653–662.
53. Wojcik, G.L., Murphy, J., Edelson, J.L., Gignoux, C.R., Ioannidis, A.G., Manning, A., Rivas, M.A., Buyske, S. and Hendricks, A.E. (2022) Opportunities and challenges for the use of common controls in sequencing studies. *Nat. Rev. Genet.*, https://doi.org/10.1038/s41576-022-00487-4.
54. Zevnik, B., Jerchow, B. and Buch, T. (2022) 3R measures in facilities for the production of genetically modified rodents. *Lab. Anim.*, **51**, 162–177.
55. Gaeta, A.L., Caldwell, K.A. and Caldwell, G.A. (2019) Found in translation: the utility of *C. elegans* alpha-synuclein models of Parkinson's disease. *Brain Sci.*, **9**, 73.
56. Sarkar, S. and Feany, M.B. (2021) Precision medicine on the fly: using *Drosophila* to decipher gene-environment interactions in Parkinson's disease. *Toxicol. Sci.*, **182**, 159–167.
57. Jean, F., Stasiuk, S., Maroilley, T., Diao, C., Galbraith, A. and Tarailo-Graovac, M. (2021) Whole genome sequencing facilitates intragenic variant interpretation following modifier screening in *C. elegans*. *BMC Genom.*, **22**, 820.
58. Zhang, F. and Lupski, J.R. (2015) Non-coding genetic variants in human disease. *Hum. Mol. Genet.*, **24**, R102–R110.
59. The ENCODE Project Consortium (2020) Perspectives on ENCODE. *Nature*, **583**, 693–698.
60. Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S. and Haussler, D. (2004) Ultraconserved elements in the human genome. *Science*, **304**, 1321–1325.
61. Seridi, L., Ryu, T. and Ravasi, T. (2014) Dynamic epigenetic control of highly conserved noncoding elements. *PLoS One*, **9**, e109326.
62. Moyon, L., Berthelot, C., Louis, A., Nguyen, N.T.T. and Roest Crollius, H. (2022) Classification of non-coding variants with high pathogenic impact. *PLoS Genet.*, **18**, e1010191.
63. Polychronopoulos, D., King, J.W.D., Nash, A.J., Tan, G. and Lenhard, B. (2017) Conserved non-coding elements: developmental gene regulation meets genome organization. *Nucleic Acids Res.*, **45**, 12611–12624.
64. Perrone, F., Cacace, R., van der Zee, J. and Van Broeckhoven, C. (2021) Emerging genetic complexity and rare genetic variants in neurodegenerative brain diseases. *Genome Med.*, **13**, 59.