

Invertebrate Model Organisms as a Platform to Investigate Rare Human Neurological Diseases

Ji-Hye Lee^{1,2,3*}

¹Department of Oral Pathology & Life Science in Dentistry, School of Dentistry, ²Dental Life Science Institute, ³Periodontal Disease Signaling Network Research Center, Pusan National University, Yangsan 50612, Korea

Patients suffering from rare human diseases often go through a painful journey for finding a definite molecular diagnosis prerequisite of appropriate cures. With a novel variant isolated from a single patient, determination of its pathogenicity to end such “diagnostic odyssey” requires multi-step processes involving experts in diverse areas of interest, including clinicians, bioinformaticians and research scientists. Recent efforts in building large-scale genomic databases and *in silico* prediction platforms have facilitated identification of potentially pathogenic variants causative of rare human diseases of a Mendelian basis. However, the functional significance of individual variants remains elusive in many cases, thus requiring incorporation of versatile and rapid model organism (MO)-based platforms for functional analyses. In this review, the current scope of rare disease research is briefly discussed. In addition, an overview of invertebrate MOs for their key features relevant to rare neurological diseases is provided, with the characteristics of two representative invertebrate MOs, *Drosophila melanogaster* and *Caenorhabditis elegans*, as well as the challenges against them. Finally, recently developed research networks integrating these MOs in collaborative research are portrayed with an array of bioinformatical analyses embedded. A comprehensive survey of MO-based research activities provided in this review will help us to design a well-structured analysis of candidate genes or potentially pathogenic variants for their roles in rare neurological diseases in future.

Key words: Rare diseases, Nervous system diseases, Invertebrates, *Drosophila melanogaster*, *Caenorhabditis elegans*, Disease models

INTRODUCTION

The number of genes linked to rare Mendelian genetic diseases has been estimated around 4,000 out of a total of approximately 30,000 genes in human genome [1-3]. It is clearly an underestimate of genetic contributions to rare diseases, as there are nearly 6,000 to 13,000 more genes to be identified for their roles in disease pathology [4]. Of approximately 6,100 unique rare diseases curated in Orphanet (www.orpha.net), those of genetic origins could make up to as much as 72%. A recent report has estimated the point prevalence, representing the population burden, less than

1/1,000,000 for the majority of these rare diseases [5]. Based upon this estimation, the overall population prevalence of rare diseases ranges from 3.5% to nearly 6%, affecting as many as 446 million people worldwide [5], with neurological illnesses as the most prevalent category [6]. While relatively low in their prevalence in the general population, the psychological and social burden of these rare diseases cannot be easily ignored, considering the cost and effort to find a definite molecular diagnosis that could take 4.8 years or longer on average [7, 8]. Such a long journey, or so-called “the diagnostic odyssey”, can be further complicated with the uncertain nature of pathogenicity associated with genetic variations in individual patients. The lack of our knowledge in functional characteristics of these variations can be devastating, given that it is virtually impossible to develop therapeutic options without prior knowledge of their physiological consequences.

The current effort to tackle rare human diseases has been facilitated by the widely used genome sequencing technologies,

Submitted January 22, 2022, Revised February 7, 2022,
Accepted February 7, 2022

*To whom correspondence should be addressed.
TEL: 82-51-510-8259, FAX: 82-51-510-8249
e-mail: jihyelee@pusan.ac.kr

including whole-genome and whole-exome sequencing. However, the rate of success in identifying genetic causes with this methodological approach is rather limited, reaching less than 30% among patients referred for diagnosis [9, 10]. Furthermore, the ultimate quest to the pathogenicity of individual genetic variants remains largely unresolved even in approximately 8% of cases with identified genes [11]. While recent advances in *in silico* platforms have helped us to predict the pathogenic profile of variants, the accuracy of their prediction can vary among different algorithms, with approximately 80% of agreement on individual pathogenicity and up to 11% of incorrect prediction [12]. As an alternative to *in silico* prediction paradigms, multiple putative variants with similar phenotypes associated can be assembled together, which may lead to identification of a gene responsible for a unique rare disease. However, such approach could be cost-ineffective with a relatively slower rate of diagnostic decisions and still require validation of the pathogenicity with functional analyses.

In this review, we provide a brief overview of invertebrate model organism (MO)-based approaches to overcome limitations of the current research efforts and to recapitulate the core values of two representative models, *Caenorhabditis elegans* and *Drosophila melanogaster*, in our understanding of rare neurobiological diseases. In addition, examples of recent advances in bioinformatical and network-based activities will be addressed in detail to describe how they are transforming research paradigms. Finally, with some challenges to invertebrate MOs, future perspectives of these MO-based research will be discussed to refine our strategies in ending a painful journey of patients who seek to find an answer for their undiagnosed neurological illnesses.

THE VALUE OF INVERTEBRATE MODEL ORGANISMS IN STUDYING RARE NEUROBIOLOGICAL DISEASES

Invertebrate MOs have helped us to broaden our understanding of biological phenomena across phyla for decades, mostly attributed to the significant degree of conservation observed in molecular mechanisms underlying major biological processes essential for cellular functions. For instance, discoveries in *Caenorhabditis elegans* (*C. elegans*) have delineated basic principles of RNA interference [13] and apoptosis [14]. Similarly, the early findings in *Drosophila melanogaster* (*D. melanogaster*) have provided critical information about the innate immunity and organ growth regulation [15-17]. While a degree of sequence identity at the level of DNA may remain relatively low, the presence of interchangeable orthologs sharing the core functions among different species further strengthens the value of these invertebrate MO-based analyses of human diseases. In addition, a single gene in an

invertebrate model often corresponds to multiple human or vertebrate counterparts, or “co-orthologs”. Such reduced redundancy in invertebrates could be beneficial in uncovering the biological significance of human gene products with multiple subtypes or isoforms when their roles in biology remain elusive. Another cumbersome bottleneck in vertebrate animal model studies often lies in their ineffectiveness in cost and time management. Simple invertebrate models such as *C. elegans* and *D. melanogaster* are advantageous in this respect with relatively short developmental time and reduced cost of breeding. With the ease of genetic manipulations of their genomes, these invertebrate models allow extensive exploration of their phenotypes in significantly reduced time spans. In addition, well-established tools for efficient phenotyping in these organisms are widely available to facilitate investigation of genotype-phenotype relationships crucial for understanding the pathophysiology of rare diseases. With these concepts in mind, the important features of two representative invertebrate models, *C. elegans* and *D. melanogaster*, will be illustrated below, along with some exemplary experimental approaches adopted in *D. melanogaster* in great detail.

The *Caenorhabditis elegans* model

C. elegans is the first MO with its genome fully sequenced in 1998 [18]. A recent genomic analysis has revealed a significant degree of similarity between human and *C. elegans* genomes, with more than 50% of genomes shared while apparent genetic simplicity reflected in relatively low numbers of paralogs [19]. As stated above, a short generation time of four days allows rapid functional characterization of potentially pathogenic variations discovered in the gene of interest [20]. Furthermore, recent advances in genetic manipulation using the CRISPR-Cas9 gene editing technology have facilitated precise manipulation of genes at an individual nucleotide level [21-24]. Another key feature integral to *C. elegans*-based study is well-established knowledgebase and resources that are publicly available, including WormBase (<https://www.wormbase.org>) and *Caenorhabditis* Genetics Center (<https://cgc.umn.edu>), to facilitate integration of novel discoveries with the existing knowledge of disease pathology.

Recent achievements in *C. elegans* study have indeed significantly contributed to our understanding of rare human diseases. For instance, a missense mutation in *Sodium Leak Channel, Non-Selective* (*NALCN*) identified in a single child was successfully modeled in *C. elegans* [25]. This child was clinically presented with intellectual disability and ataxia as well as congenital arthrogryposis. With the conserved residue of human *NALCN* (R1181Q) identified in two *C. elegans* homologs, *nca-1* and *nca-2* [26], a orthologous substitution of the conserved residue was introduced

in *nca-1*, followed by phenotypic characterization of mutants. Importantly, the mutants harboring this variant displayed abnormal locomotive behaviors recapitulating the gain-of-function (GOF) phenotypes previously described in other *nca-1* mutant variants, thus confirming the GOF pathogenicity of a human *NALCN* variant [25]. A successful implementation of *C. elegans* model in rare disease research has also been demonstrated with the help of CRISPR-Cas9 gene editing technology. Two missense variants of *B9D2* (P74S and G155S) identified in patients with ciliopathies due to Joubert syndrome [27] were successfully modeled in *C. elegans* by gene editing in homologous *mksr-2* with over 60% of sequence identity in gene products. Introduction of presumed pathogenic mutations in *C. elegans* resulted in severely disrupted ciliary function and structure [28]. This study further demonstrates the value of *C. elegans* model in deciphering the functional consequences of potentially pathogenic variations causative of rare human diseases.

The *Drosophila melanogaster* model

Since its early contribution by Thomas H. Morgan in 1910s [29, 30], *D. melanogaster* has been at the heart of numerous breakthroughs indispensable for our current understanding of biological phenomena. As in worms, flies have a short developmental time around 10–14 days as well as a relatively short life span generally not exceeding three months. Together with a short generation time, a large number of progenies are produced per each reproduction cycle, thus making it an attractive *in vivo* model to perform large-scale screening experiments. In addition, a rich body of knowledgebase and experimental resources are widely available to the research community, including a solid database (FlyBase, <https://flybase.org>), public stock centers (BDSC, <https://bdsc.indiana.edu>; VDRC, <https://stockcenter.vdrc.at>), and a cDNA clone collection (DGRC, <https://dgrc.bio.indiana.edu>). With its genome completely sequenced in 2000 [31], the degree of conservation between human and fly genomes is significant for coding genes (~65%), with the similarity even greater than 80% for the disease-associated human genes [32, 33]. In addition to genome-wide similarity, highly sophisticated genetic tools are widely available to manipulate genes of interest in a spatio-temporally controlled manner, thus facilitating a functional analysis of human variants with unknown pathogenicity in appropriate cellular or molecular contexts [34–37].

Experimental strategies to investigate rare human diseases using *D. melanogaster* can be divided into four different categories in general. 1) First, an attempt can be made to humanize flies by replacement of a fly gene with a human ortholog or homolog in its wild-type (WT) or variant form. Once the fly loss-of-function

(LOF) mutants are generated and characterized, their phenotypes following introduction of a human ortholog are analyzed in great detail to see if the human counterpart in either a WT or a variant form can functionally rescue the LOF mutants [30]. For instance, a recent report on modeling of Parkinson's disease has successfully demonstrated that L-DOPA-induced dyskinesia often observed in human patients could be phenotypically studied in *Drosophila* larval models by introducing a specific human variant of α -synuclein (A53T) and by monitoring dyskinetic behaviors following an administration of a high concentration of L-DOPA [38]. 2) When a variation in the gene of interest occurs in a conserved region between human and fly genomes, the pathogenicity of a specific variant can be directly investigated in flies by introduction of orthologous variants to an endogenous fly gene via CRISPR-Cas9-based gene editing [39, 40]. The degrees of rescue are compared between LOF mutants harboring either a WT or a variant fly cDNA to analyze the pathogenicity of each variant. 3) If a rapid assessment of a rare variant is required, one can force expression of human or fly cDNAs in a tissue-specific manner in WT animals. This will create an environment in which both endogenous fly gene products and exogenous fly or human counterparts are co-expressed, thus providing an opportunity to reveal a unique property of a variant, including a dominant-negative effect. 4) Finally, functional characterization of fly LOF phenotypes can also provide a clue to understand human variants of certain categories that are likely linked to a complete loss of gene functions, including non-sense, frameshift and deletion mutations. Such mutations in vertebrate models may lead to the lethality of an affected animal, thus making it impossible to study the pathogenicity of each variant and relevant molecular mechanisms. However, a knock-out of fly orthologs in a tissue-specific manner may offer a unique opportunity to study functional characteristics of a gene of interest. Together with the aforementioned intrinsic advantages, versatile applications and adaptations of these experimental paradigms further strengthen the value of *D. melanogaster* model in studying rare human diseases.

BIOINFORMATICAL INTEGRATION OF MULTI-OMICS DATABASES AND MODEL ORGANISM-BASED RESEARCH

In the post-genome era, discoveries of genetic variants in the human genome further emphasize the role of bioinformatical analyses in deciphering the pathogenicity of each variant associated with rare human diseases of a definitive genetic origin. In line with this idea, large-scale sequencing centers have been established to facilitate the effective identification of potentially pathogenic disease variants, applying whole-genome and whole-exome sequenc-

ing to individuals or cohorts with suspected Mendelian diseases. For instance, the Centers for Mendelian Genomics was established in 2012, with a financial support from NIH in US [1]. A recent report from this institution in collaboration with investigators in 36 countries included an analysis of over 18,000 samples representing approximately 1,050 Mendelian phenotypes, which led to identification of nearly 1,000 genes associated with disorders, including 375 genes that have never been mapped to human diseases [2]. Identification of candidate variants from such large-scale sequencing approaches often requires initial validation of their pathogenicity with bioinformatical and computational means. For this purpose, a number of *in silico* prediction algorithms have been developed, aiding to refine the scope of subsequent functional analyses using MOs. In addition, recently established platforms such as CADD (Combined Annotation Dependent Depletion) and REVEL (Rare Exome Variant Ensemble Learner) incorporate machine learning technologies with previously developed algorithms to improve the power of prediction [41, 42]. Briefly, REVEL was established by incorporating a total of 18 pathogenicity prediction scores (8 conservation scores and 10 functional scores) from 13 tools and by introducing additional training sessions with recently unraveled diseases and rare missense variants that have not been previously provided for initial prediction training. In this way, extracted REVEL ensemble scores could provide an improved prediction value for the pathogenicity of missense and rare variants [41]. Similarly, CADD integrates multiple metrics with diverse information types, such as conservation and the nature of variants as missense changes, thus providing quantitative scores (“C-scores”) for predicting the deleteriousness of diverse arrays of genomic variations and for prioritizing potentially causative variants in research and clinical activities (<https://cadd.gs.washington.edu/>) [42].

This step of pathogenicity prediction can be further aided with bioinformatical analyses of genetic and physical interaction databases that provide valuable information about interactions between a gene or protein of interest and others previously associated with a specific Mendelian disease. Databases generated from large-scale proteomics screens in diverse species are included in this type of analysis, along with text-mining strategies for relevant publications based in MO studies. Recently, a few bioinformatical tools have been developed to optimize such analysis, including STRING (Search Tool for Recurring Instances of Neighboring Genes) and MIST (Molecular Interaction Search Tool). STRING provides a database for established and predicted genetic and protein interactions based upon systematic co-expression patterns, shared signals across multiple genomes, inferred interaction knowledge from different organisms and text-mining analysis

of the scientific literature [43]. Similarly, the MIST platform integrates previously identified genetic and protein interactions from humans and multiple MOs, including yeast, worm, fly, zebrafish, frog, rat and mouse, to predict interactions inferred from “interlogs”, i.e. interactions between orthologous genes or proteins in different organisms [44]. A more comprehensive analysis of both sequence- and interaction-based datasets could further facilitate research efforts to delineate the pathogenicity of a rare variant potentially causative of undiagnosed human disease phenotypes.

Aside from the aforementioned large-scale screening efforts, discoveries of pathogenic variants often stem from isolate cases first identified by clinicians in practice. A translational approach from this individual discovery to MO-based research usually begins with integration of curated information across organisms. This step requires extensive literature and database search to see if a specific gene or variant of interest has been previously implicated in human diseases. The components of this initial analysis include, but not limited to, 1) the presence of prior reports of similar disease phenotypes, 2) the allele frequency of a specific variant in the general vs. disease population, 3) the availability of orthologs or homologs of implicated human genes in MOs, 4) their functional characteristics and expression profiles, 5) the site of variants in the gene structure, i.e. whether it is located in functional domains or not, and 6) the conservation of amino acids affected in variants across phyla. Manual curation and analysis of such diverse information is a near-impossible task, considering the sheer number of databases subject to analysis, thus requiring development of efficient tools to integrate multiple human and MO databases. One such tool available is MARRVEL (Model organism Aggregated Resources for Rare Variant ExpLoration) [45-47] (see also Fig. 1 for its integration in MO-based research networks described below). The current scope of MARRVEL encloses information obtained from 115,000 control cohorts and 12.3 million variants, genotype-phenotype relationships described for 6.96 million cases, and over 20,000 Gene Ontology terms corresponding to nearly 236,000 pairs of human genes and MO orthologs or homologs (MARRVEL v2.0). The human genetic databases incorporated in the MARRVEL analysis include ExAc [48], Geno2MP (NHGRI/NHLBI University of Washington-Center for Mendelian Genomics (UW-CMG)), ClinVar [49], DGV [50], DECIPHER [51] and OMIM (<https://omim.org/>). In addition, multiple MO databases are incorporated in the process of information curation, including SGD [52], PomBase [53], WormBase [54], FlyBase [55], ZFIN [56], MGI [57] and RGD [58]. The MARRVEL platform allows users to identify orthologs of a human gene of interest in MOs using DRSC Integrative Ortholog Prediction Tool (DIOPT) [32]. Following identification of available orthologs, it also provides com-

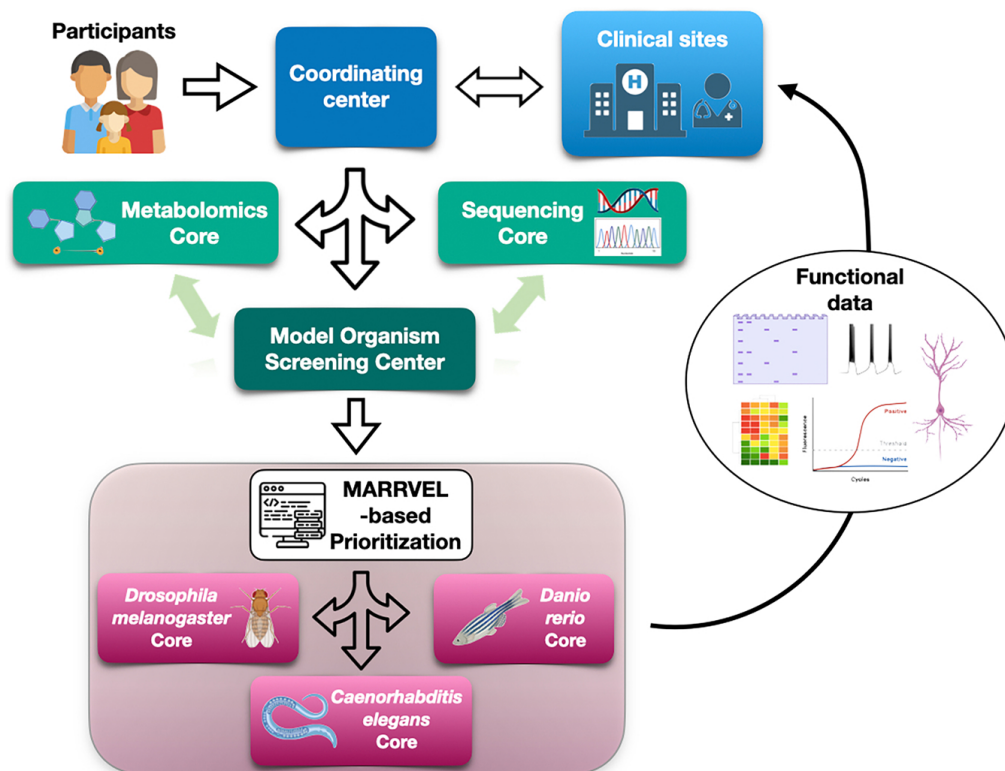


Fig. 1. The workflow of the Undiagnosed Disease Network. The Undiagnosed Disease Network (UDN) consists of 12 Clinical Sites, Coordinating Center, Metabolomics Core, Sequencing Core, and three Model Organism Screening Centers (MOSCs). In MOSCs, there are *Drosophila melanogaster* Core, *Caenorhabditis elegans* Core and two *Danio rerio* Cores. Individual projects are initiated with participant applications submitted to Coordinating Center. The accepted applications are then forwarded to one of the clinical sites in US. The detailed information of clinical phenotypes is then distributed to Metabolomics Core, Sequencing Core and MOSCs for further analyses. Once submitted to MOSCs, bioinformatical tools such as MARRVEL are utilized to prioritize the candidate genes or variants and then assigned to appropriate MO cores for functional studies. The results from these MO-based analyses are then communicated again with experts in clinical sites.

parison data at the level of protein, specifically concerning amino acid sequence alignment, annotation of functional domains and the presence of conserved residues.

While bioinformatical toolkits clearly help scientists to narrow the scope of putative targets of their research, there are some cautions to be made in interpreting the data generated with these means. First of all, the possibility of generating pseudo-positive and pseudo-negative results from *in silico* prediction should not be ignored. Indeed, a recent report on such prediction tools raised an issue of limited validity of prediction especially with clinical datasets [59]. In the case of interaction-based extrapolation of databases, the spatio-temporal context of each interaction needs to be carefully considered for proper assembly of information. Furthermore, functional significance of amino acid substitutions at relatively less critical positions or domains should still not be ignored and awaits further investigation in functional studies. Finally, the general assumption in database-derived assembly of information lies in the accuracy of the previously curated knowledge

of genotype-phenotype relationships available in public databases. Such assumption should be granted upon the quality control of individual experimental studies incorporated, thus requiring us to handle such information with caution.

COLLABORATIVE NETWORK-BASED APPROACHES TO STUDY RARE NEUROLOGICAL DISEASES USING INVERTEBRATE MODEL ORGANISMS

Cost-ineffectiveness of conventional approaches to identify pathogenic variants as well as limitations of *in silico* prediction and relevant bioinformatical tools strongly urge incorporation of MO-based functional studies for affirmative validation of potentially causative variants in a large scale. However, the successful implementation of MO-based studies is only granted with well-coordinated and extensively collaborative activities among independent teams dedicated to analyses of different rare disease features, engaging clinicians, bioinformatics specialists and research

Table 1. Contributions of *Drosophila melanogaster* and *Caenorhabditis elegans* studies conducted through UDN to identification of pathogenic variants responsible for rare neurological disorders

Model organism studied	Human gene of interest	Neurological disease or phenotype associated	Year reported
<i>D. melanogaster</i>	<i>EBF3</i>	Hypotonia, ataxia, delayed development	2017 [62]
<i>D. melanogaster</i>	<i>CACNA1A</i>	Infantile developmental delay, ataxia	2017 [64]
<i>D. melanogaster</i>	<i>IRF2BRL</i>	Severe neurodevelopmental regression, hypotonia, ataxia, seizures, abnormal motor behaviors	2018 [65]
<i>D. melanogaster</i>	<i>WDR37</i>	Neurooculocardio-genitourinary syndrome	2019 [66]
<i>D. melanogaster</i> , <i>D. rerio</i>	<i>BICRA (GLTSCR1)</i>	SWI/SNF-related intellectual disability disorder	2020 [67]
<i>D. melanogaster</i>	<i>TOMM70</i>	Hypotonia, dystonia, ataxia, white matter abnormalities	2020 [68]
<i>D. melanogaster</i>	<i>CDK19</i>	Epileptic encephalopathy, hypotonia, general developmental delay	2020 [69]
<i>D. melanogaster</i>	<i>ACOX1</i>	Glial loss (Schwann cell loss)	2020 [70]
<i>D. melanogaster</i>	<i>TNPO2</i>	General developmental delay, neurologic deficits	2021 [71]
<i>D. melanogaster</i> , <i>D. rerio</i>	<i>GDF11</i>	Craniofacial and vertebral abnormalities, neurological deficits	2021 [72]
<i>D. melanogaster</i>	<i>RNF2 (RING2)</i>	Intellectual disability, seizures, behavioral abnormalities	2021 [73]
<i>C. elegans</i>	<i>NBEA</i>	Neurodevelopmental delay, early childhood epilepsy	2021 [74]

The list of *D. melanogaster* and *C. elegans*-based studies conducted is shown for rare neurological disorders investigated through the Undiagnosed Disease Network during the phases I and II. The citation for each report is provided along with the year published.

scientists. Recent efforts in research fields indeed reflect adoption of such idea in collaborative network-based approaches. Here I briefly discuss the key features and achievements of two representative research networks established with a great emphasis on MO-based research activities.

Undiagnosed Disease Network

The Undiagnosed Disease Network (UDN, <https://undiagnosed.hms.harvard.edu>) was initiated from a NIH-funded project to establish a system that would integrate clinical and research activities in deciphering the molecular basis of rare human diseases [60, 61]. It consists of an interrelating network of centers and cores that encompass expert researches in clinical medicine, bioinformatics and MO-based genetics. The network includes 12 clinical sites, a single coordinating center, sequencing and metabolomics cores, and three Model Organisms Screening Centers (MOSCs) along with a central biorepository. The three MOSCs comprises *C. elegans* and *D. melanogaster* cores as well as two zebrafish cores. Together with the bioinformatics team embedded in its core, the MOSC plays a crucial role in facilitating rare human disease studies affiliated with UDN. The general workflow through UDN involves multi-step and interactive collaborations among clinicians, bioinformaticians and MO scientists. Briefly, research applications are submitted to the coordinating center, some of which are forwarded to clinical sites for further evaluation specifically for disease phenotypes and genetic variations of probands and their family members. Results of clinical evaluation are then forwarded back to its coordinating center as well as to MOSCs and sequencing and metabolomics cores for further studies. The candidate genes selected for MO-based studies are prioritized and assigned to appropriate MO cores

for functional analyses. The research outcomes following this step are again communicated with experts in clinical sites (Fig. 1). As a result of this extensive network-based collaboration, a total of 2,184 participant applications were accepted for study, resulting in 528 cases diagnosed and 362 variants evaluated with MOs. Among the evaluated applications, a total of 133 genes were assigned for functional analyses in MOSC cores, some of which were directly linked to diagnostic advancement (<https://undiagnosed.hms.harvard.edu>).

D. melanogaster and *C. elegans* have had a significant contribution to the current success of UDN as two major cores of MOSCs in the network. As of January 2022, a total of 21 functional *in vivo* studies have been reported for rare neurological disorders with an affiliation of UDN (<https://undiagnosed.hms.harvard.edu>). Among these reports, 12 of them included a systematic analysis of rare genetic variants using flies and worms (Table 1), followed by 5 zebrafish-based and 6 mouse-based studies. Indeed, the first report demonstrating effective integration of MOSCs involved a collaborative effort among clinicians, bioinformatics experts and a MO research group studying *D. melanogaster* [62]. The network-based analysis of a genetic variant in this case was initiated from a report of a single child with a neurodevelopmental disorder carrying a *de novo* variant in *Early B-cell Factor 3 (EBF3)*. Following an initial report of the variant, two more cases were identified, with the same amino acid of *EBF3* gene product affected. These initial clinical and sequencing results led to a subsequent analysis in *Drosophila*. The function of a fly ortholog of human *EBF3*, *knot*, was characterized using fly LOF mutants as well as rescue lines with human cDNA in either a WT or a variant form. A failure of genetic rescue with human *EBF3* carrying patient-derived missense muta-

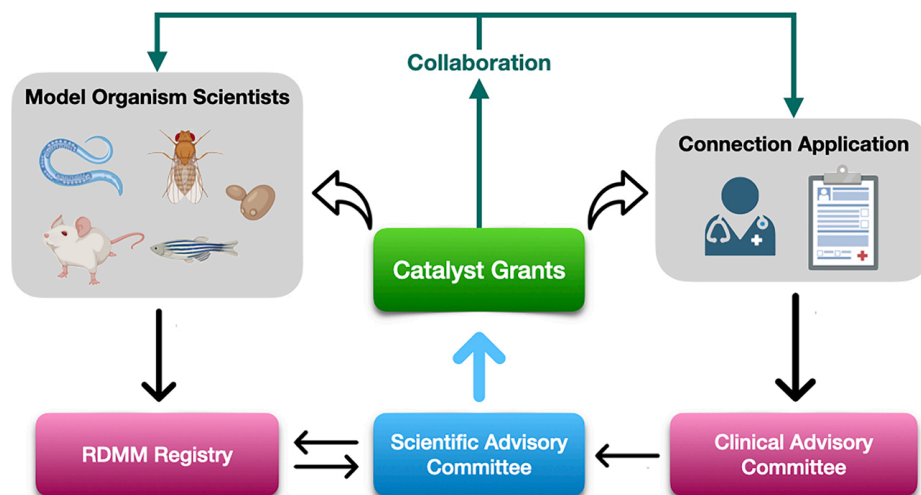


Fig. 2. The workflow of the Rare Diseases Models and Mechanisms network. As a MO-based research network established initially in Canada, the Rare Diseases Models and Mechanisms (RDMM) network is structured in a committee basis. A project in the RDMM network is initiated with a submission of a “Connection Application” from clinicians who identify patients with rare genetic diseases. Once reviewed by the Clinical Advisory Committee, the approved applications are forwarded to the Scientific Advisory Committee for their search of matching MO scientists enrolled in the RDMM Registry. The MO scientists recruited by the Scientific Advisory Committee are then requested to submit a “Model Organism Proposal Application” to the committee. Following an approval of this proposal, MO scientists receive a catalyst fund to initiate an immediate collaborative research with clinicians who submitted the “Connection Application”.

tions confirmed the pathogenicity of these *de novo* variants isolated from just three patients [62]. Together with other examples of coherent integration of MOs in the UDN-affiliated activities (Table 1), these results consistently portray a crucial role of these two MOs in successful identification of pathogenic variants causative of rare and undiagnosed neurological diseases in the interactive network-based research environment. While initially established in US, the horizon of UDN has extended to include participants and researchers around the globe to facilitate international collaboration for rare human diseases, resulting in the formation of the Undiagnosed Disease Network International [63] (<https://www.udninternational.org/>).

Rare Diseases Models and Mechanisms

The Rare Diseases Models and Mechanisms (RDMM) network (<http://www.rare-diseases-catalyst-network.ca/>) was established in 2014 to facilitate collaborations between clinicians and MO scientists in Canada [75]. Unlike UDN described above, its core structure is committee-based to identify and catalyze the connections between clinicians and scientists and supplemented with the Canadian RDMM Registry for its data curation. The overall project flow starts with a submission of a “Connection Application” concerning novel variants of interest to the advisory committee, which then evaluates the proposal based on the following aspects: 1) the quality of genetic data for diseases, 2) the severity of a dis-

ease reported and the need for medical intervention to affected patients, 3) a possibility of developing therapeutic options, 4) the population burden on either a specific or the general population and 5) the novelty of a biological pathway implicated in a disease of interest [75]. Once the application is approved by the advisory committee, the next decision is made by the scientific advisory committee to match it to MO scientists enrolled in the Canadian RDMM Registry. If matched, experts in MO research are invited to submit a “Model Organism Proposal Application” to be approved by the committee as well as the clinician who submits the “Connection Application.” Following an approval of this proposal, MO scientists would receive a catalyst fund to initiate an immediate collaborative research (Fig. 2).

The scope of RDMM network encloses the vast majority of both clinical and research communities in Canada. For instance, the RDMM aims to engage the clinical community dealing with rare diseases through established projects, including FORGE Canada Consortium [76] and the Treatable Intellectual Disability Endeavor protocol in British Columbia (TIDE BC) [77]. In conjunction with this end, the network also engages the MO research community by offering genetic knowledgebase for MO-based research. As a part of this effort, a total of nearly 12,200 genes are classified into three categories. This dataset includes 1) tier 1 for the genes of which functions were directly investigated with MOs in previous studies, 2) tier 2 for those awaiting immediate investigation with

MOs and 3) tier 3 genes with some degree of relevance inferred from their previous studies based upon Gene Ontology terms [75]. The RDMM network has received 135 Connection Applications, in addition to direct submissions of 116 understudied candidate genes suggested from FORGE, Care4Rare and UDN. Among these initial submissions, a total of 105 functional studies were funded, including 85 MO catalyst projects connecting clinicians and MO experts in mouse, fly, zebrafish, worm, yeast and some protozoa [75]. The contribution of invertebrate MOs to recent achievements of the RDMM network is significant in that 17 out of 105 studies are built around *D. melanogaster* and *C. elegans* systems [75]. Importantly, two recent reports for rare neurological diseases have exemplified successful implementation of these invertebrate models in providing functional data to support genotype-phenotype correlation [78] and to validate novel gene discoveries [25]. While initially based in Canada, the RDMM network continues to expand its horizon to the globe, leading to establishment of other regional networks with a similar structural design, including RDMM Europe (<https://solve-rd.eu/rdmm-europe/>), Australian Functional Genomics Network (<https://www.functionalgenomics.org.au/>) and J-RDMM in Japan (<https://j-rdmm.org/>).

CHALLENGES TO INVERTEBRATE MODEL ORGANISM-BASED APPROACHES IN STUDYING RARE HUMAN DISEASES

When conducting functional experiments to study rare human diseases in MOs, it is possible that the phenotypes described in MOs may have no clear relationship to human phenotypes reported from individual patients. This class of complications can be further divided into two different scenarios. First, the phenotypes observed in MOs may not be fully rescued by introduction of human cDNA. In this case, human protein orthologous to MO counterparts may not be functional or cause lethality in the MO background. For instance, a human protein induced for rescue may not form a functional complex with endogenous proteins, or induce toxicity in a mode similar to GOF alleles. In addition, the effects of WT and variant forms of human gene products may not differ from each other on restoring endogenous functions of a MO gene in the LOF background. In such case, an alternative investigation in different MOs should be further sought to rule out the potential pathogenicity of variants.

It should also be noted that MOs carrying a variant of interest may display phenotypes that are relatively unrelated to human diseases, but still allowing to be quantitatively scored. These phenotypes, often referred to as “phenologs”, should be carefully examined, as they may provide valuable insights into the basic biology of human gene of interest albeit a lack of phenotypic homology.

The significance of a “phenolog” is mostly based upon the general principle of conservation observed in key cellular signaling pathways. Even with nearly interchangeable components of a signaling pathway, disrupted endogenous activity of these components may manifest completely different scorable phenotypes among different species. For instance, mutations in *Drosophila notch* result in dysmorphic features in wings, legs and eyes while similar perturbations in human genome are mostly associated with congenital diseases non-homologous to *Drosophila* phenotypes [79]. The careful evaluation of “phenologs” can thus still provide an important biological clue to molecular links between rare human diseases and novel variants with no or less information available for their functional significance.

The pathophysiology of certain human diseases may require organ- or tissue-specific contexts for the experimental studies. In that case, functional alterations need to be modeled in human- or vertebrate-specific organs, thus prohibiting invertebrate MOs from maximizing its usage for intended studies. A zebrafish model can be an appropriate alternative as it possesses virtually all of the same organs as humans. In addition to this context, it is also possible that a human gene of interest may not have orthologs or homologs in invertebrate MOs. Therefore, it is not likely that these models present direct functional evidence to validate the pathogenicity of human variants. Nevertheless, the data obtained from invertebrate MOs can still provide basic knowledge about the function of novel variants or genes of interest in the context of conserved core signaling pathways, thus acting as *in vivo* indicators of perturbed cellular functions.

FUTURE PERSPECTIVES OF INVERTEBRATE MODEL ORGANISM-BASED RESEARCH FOR RARE NEUROLOGICAL DISEASES

As evidenced by recent success stories from network-based approaches such as UDN and RDMM, efficient modeling and molecular diagnosis of rare neurological disorders requires well-coordinated collaborations among clinicians responsible for identification of patients carrying rare genetic variants, bioinformatics specialists analyzing multi-omics and literature-based datasets and research scientists specialized in different MOs. Therefore, recruitment of sufficient numbers of experts in each field as well as experienced research coordinators should be granted to facilitate a systematic analysis of currently available information, to design and employ solid experimental paradigms, and to collectively interpret the findings, in order to provide an appropriate molecular diagnosis for neurological disorders of an unknown nature.

With increasing numbers of genomic data available for analysis,

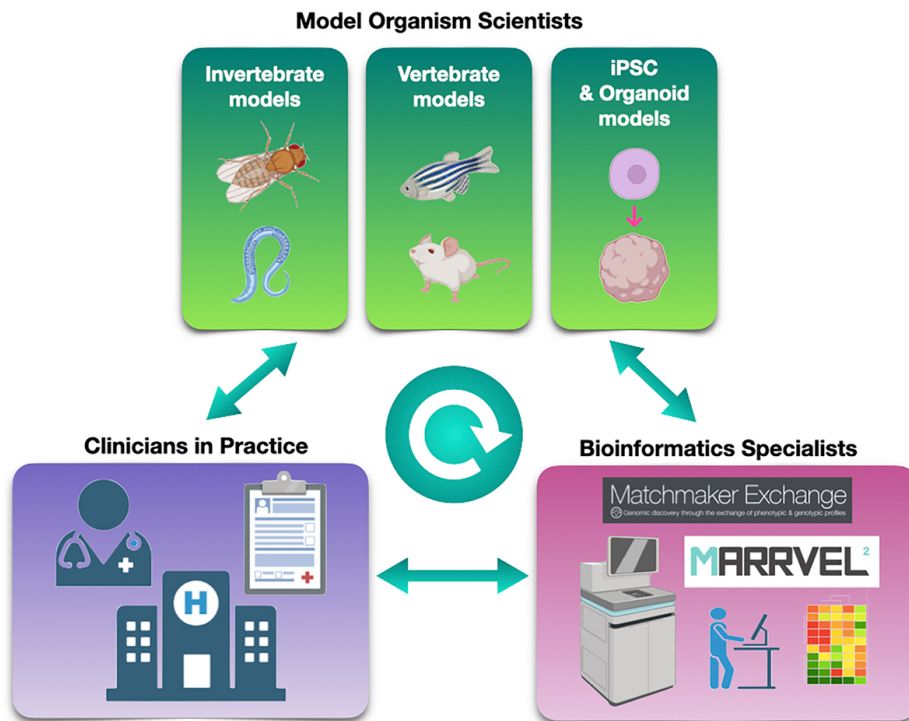


Fig. 3. Three critical legs to high-performance model organism-based studies of rare neurological diseases. For successful identification of potentially pathogenic variants causative of rare neurological diseases, an interactive collaboration is mandatory among clinicians responsible for initial identification of patients, bioinformatics specialists handling multi-omics and knowledgebase datasets and research scientists with expertise in model organisms or systems, including invertebrate and vertebrate animal models as well as iPSC-organoid systems.

an inter-database match has become a critical step in identification of patient pools with similar genotype-phenotype relationships. Besides, an initial clinical identification of rare human diseases often comes with just a single patient carrying an uncharacterized variant. Therefore, it is necessary to recruit more cases with similar genotype-phenotype relationships to enhance the statistical power of subsequent analyses. Several tools or platforms have been recently established, including Matchmaker Exchange (MME, <https://www.matchmakerexchange.org/>) [80-82] that incorporates multiple databases such as GeneMatcher [83], Phenome Central [84], DECIPHER [51, 85], MyGene2 [86], AGHA Patient Archive (<https://mme.australiangenomics.org.au>), PatientMatcher (<https://github.com/Clinical-genomics/patientmatcher/>), seqr (<https://seqr.broadinstitute.org/matchmaker/matchbox>), RD-Connect GPAP (<https://platform.rd-connect.eu/>), and IRUD [87]. So far, MME provides the most extensive coverage of inter-database comparisons for genotype-phenotype relationships, employed in at least 40 reports of gene discoveries from nearly 190,000 participants or cases (<https://www.matchmakerexchange.org/>). With a genomic analysis of patients with undiagnosed neurological phenotypes becoming a widely available diagnostic tool in clinics,

accurate assessment of inter-database information will be an indispensable prerequisite of MO-based approaches before initiation of model-based functional studies.

While speedy and effective analyses can be performed in invertebrate MOs, one can still take advantage of other vertebrate MOs, including zebrafish and mouse, to fill the gaps left behind. Interactive collaborative efforts with the mouse model-based research groups such as International Mouse Phenotyping Consortium (<https://www.mousephenotype.org/>) [88] and Knockout Mouse Phenotyping Program (KOMP2, <https://commonfund.nih.gov/komp2>) will further facilitate the molecular diagnosis of rare neurological disorders (Fig. 3). In case of context-specific modeling of human diseases, invertebrate MOs may not be ranked as the choice of studies, as it needs to re-create a patient-specific condition. This can be alternatively built with the use of patient-derived induced pluripotent stem cells and subsequent generation of context-specific cell types and organoids [89, 90] (Fig. 3).

The most of our efforts in studying rare genetic diseases with a Mendelian basis has been restricted to potentially pathogenic variants within the coding regions. However, with more frequent applications of whole-genome sequencing technology, it is likely

that we discover a significant fraction of variants residing in non-coding regions, including microRNAs, long non-coding RNAs and repetitive DNA sequences. The degree of conservation tends to be significantly low in regulatory genomic regions between human and invertebrate MO genomes. In addition, potentially pathogenic variants occurring in these regions of human genome may cause noticeable changes in expression profiles of other genes. With limited degrees of conservation, invertebrate MOs can be alternatively used to analyze the effect of altered gene expression portraited in human patients on disease phenotypes [91].

CONCLUSION

With the help of technical advances, the rate of novel gene or variant discoveries responsible for rare neurological diseases have continued to increase in the last few decades. Recent reports have estimated that there are nearly 6,000 to 13,000 more genes to be studied for their roles in Mendelian genetic disorders, many of which functions likely affect the integrity of the human nervous system [4, 6]. Among these candidates, a significant fraction ranging from 4,000 to 10,000 is expected to be novel discoveries [4], thus still representing formidable tasks remaining ahead. With an increasing rate of genotype-based identification of novel variants linked to rare neurological diseases, concerted team efforts among clinicians, bioinformaticians and MO scientists will be in greater demand (Fig. 3). These expert communities providing interactive support to one another will serve as cornerstones to help patients with undiagnosed rare neurological disorders to end their painstaking journey in finding answers for a molecular diagnosis and cures.

ACKNOWLEDGEMENTS

This work was supported by the Financial Supporting Project of Long-term Overseas Dispatch of PNU's Tenure-track Faculty, 2019 to Ji-Hye Lee.

REFERENCES

- Bamshad MJ, Shendure JA, Valle D, Hamosh A, Lupski JR, Gibbs RA, Boerwinkle E, Lifton RP, Gerstein M, Gunel M, Mane S, Nickerson DA; Centers for Mendelian Genomics (2012) The centers for Mendelian genomics: a new large-scale initiative to identify the genes underlying rare Mendelian conditions. *Am J Med Genet A* 158A:1523-1525.
- Chong JX, Buckingham KJ, Jhangiani SN, Boehm C, Sobreira N, Smith JD, Harrell TM, McMillin MJ, Wiszniewski W, Gambin T, Coban Akdemir ZH, Doheny K, Scott AF, Avramopoulos D, Chakravarti A, Hoover-Fong J, Mathews D, Witmer PD, Ling H, Hetrick K, Watkins L, Patterson KE, Reinier F, Blue E, Muzny D, Kircher M, Bilguvar K, López-Giráldez F, Sutton VR, Tabor HK, Leal SM, Gunel M, Mane S, Gibbs RA, Boerwinkle E, Hamosh A, Shendure J, Lupski JR, Lifton RP, Valle D, Nickerson DA; Centers for Mendelian Genomics, Bamshad MJ (2015) The genetic basis of Mendelian phenotypes: discoveries, challenges, and opportunities. *Am J Hum Genet* 97:199-215.
- Rath A, Olry A, Dhombres F, Brandt MM, Urbero B, Ayme S (2012) Representation of rare diseases in health information systems: the Orphanet approach to serve a wide range of end users. *Hum Mutat* 33:803-808.
- Bamshad MJ, Nickerson DA, Chong JX (2019) Mendelian gene discovery: fast and furious with no end in sight. *Am J Hum Genet* 105:448-455.
- Nguengang Wakap S, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau V, Murphy D, Le Cam Y, Rath A (2020) Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *Eur J Hum Genet* 28:165-173.
- Osinalde N, Duarri A, Ramirez J, Barrio R, Perez de Nanclares G, Mayor U (2019) Impaired proteostasis in rare neurological diseases. *Semin Cell Dev Biol* 93:164-177.
- Hartin SN, Means JC, Alaimo JT, Younger ST (2020) Expediting rare disease diagnosis: a call to bridge the gap between clinical and functional genomics. *Mol Med* 26:117.
- Blöß S, Klemann C, Rother AK, Mehmecke S, Schumacher U, Mücke U, Mücke M, Stieber C, Klawonn F, Kortum X, Lechner W, Grigull L (2017) Diagnostic needs for rare diseases and shared prediagnostic phenomena: results of a German-wide expert Delphi survey. *PLoS One* 12:e0172532.
- Lee H, Deignan JL, Dorrani N, Strom SP, Kantarci S, Quintero-Rivera F, Das K, Toy T, Harry B, Yourshaw M, Fox M, Fogel BL, Martinez-Agosto JA, Wong DA, Chang VY, Shieh PB, Palmer CG, Dipple KM, Grody WW, Vilain E, Nelson SF (2014) Clinical exome sequencing for genetic identification of rare Mendelian disorders. *JAMA* 312:1880-1887.
- Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, Ward P, Braxton A, Wang M, Buhay C, Veeraraghavan N, Hawes A, Chiang T, Leduc M, Beuten J, Zhang J, He W, Scull J, Willis A, Landsverk M, Craigen WJ, Bekheirnia MR, Stray-Pedersen A, Liu P, Wen S, Alcaraz W, Cui H, Walkiewicz M, Reid J, Bainbridge M, Patel A, Boerwinkle E, Beaudet AL, Lupski JR, Plon SE, Gibbs RA, Eng CM (2014) Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA* 312:1870-1879.

11. Farwell Hagman KD, Shinde DN, Mroske C, Smith E, Radtke K, Shahmirzadi L, El-Khechen D, Powis Z, Chao EC, Alcaraz WA, Helbig KL, Sajan SA, Rossi M, Lu HM, Huether R, Li S, Wu S, Nuñez ME, Tang S (2017) Candidate-gene criteria for clinical reporting: diagnostic exome sequencing identifies altered candidate genes among 8% of patients with undiagnosed diseases. *Genet Med* 19:224-235.
12. Ghosh R, Oak N, Plon SE (2017) Evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. *Genome Biol* 18:225.
13. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391:806-811.
14. Ellis HM, Horvitz HR (1986) Genetic control of programmed cell death in the nematode *C. elegans*. *Cell* 44:817-829.
15. Lemaitre B, Nicolas E, Michaut L, Reichhart JM, Hoffmann JA (1996) The dorsoventral regulatory gene cassette *spätzle/Toll/cactus* controls the potent antifungal response in *Drosophila* adults. *Cell* 86:973-983.
16. Halder G, Johnson RL (2011) Hippo signaling: growth control and beyond. *Development* 138:9-22.
17. Markopoulou K, Artavanis-Tsakonas S (1989) The expression of the neurogenic locus Notch during the postembryonic development of *Drosophila melanogaster* and its relationship to mitotic activity. *J Neurogenet* 6:11-26.
18. *C. elegans* Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282:2012-2018.
19. Kim W, Underwood RS, Greenwald I, Shaye DD (2018) OrthoList 2: a new comparative genomic analysis of human and *caenorhabditis elegans* genes. *Genetics* 210:445-461.
20. Kropp PA, Bauer R, Zafra I, Graham C, Golden A (2021) *Caenorhabditis elegans* for rare disease modeling and drug discovery: strategies and strengths. *Dis Model Mech* 14:dmm049010.
21. Paix A, Folkmann A, Rasoloson D, Seydoux G (2015) High efficiency, homology-directed genome editing in *Caenorhabditis elegans* using CRISPR-Cas9 ribonucleoprotein complexes. *Genetics* 201:47-54.
22. Paix A, Schmidt H, Seydoux G (2016) Cas9-assisted recombineering in *C. elegans*: genome editing using in vivo assembly of linear DNAs. *Nucleic Acids Res* 44:e128.
23. Paix A, Folkmann A, Goldman DH, Kulaga H, Grzelak MJ, Rasoloson D, Paidemarry S, Green R, Reed RR, Seydoux G (2017) Precision genome editing using synthesis-dependent repair of Cas9-induced DNA breaks. *Proc Natl Acad Sci U S A* 114:E10745-E10754.
24. Paix A, Folkmann A, Seydoux G (2017) Precision genome editing using CRISPR-Cas9 and linear repair templates in *C. elegans*. *Methods* 121-122:86-93.
25. Aoyagi K, Rossignol E, Hamdan FF, Mulcahy B, Xie L, Nagamatsu S, Rouleau GA, Zhen M, Michaud JL (2015) A gain-of-function mutation in NALCN in a child with intellectual disability, ataxia, and arthrogryposis. *Hum Mutat* 36:753-757.
26. Topalidou I, Chen PA, Cooper K, Watanabe S, Jorgensen EM, Ailion M (2017) The NCA-1 and NCA-2 ion channels function downstream of Gq and Rho to regulate locomotion in *Caenorhabditis elegans*. *Genetics* 206:265-282.
27. Bachmann-Gagescu R, Dempsey JC, Phelps IG, O'Roak BJ, Knutzen DM, Rue TC, Ishak GE, Isabella CR, Gorden N, Adkins J, Boyle EA, de Lacy N, O'Day D, Alswaid A, Ramadevi A R, Lingappa L, Lourenço C, Martorell L, Garcia-Cazorla À, Ozyürek H, Haliloğlu G, Tuysuz B, Topçu M; University of Washington Center for Mendelian Genomics, Chance P, Parisi MA, Glass IA, Shendure J, Doherty D (2015) Joubert syndrome: a model for untangling recessive disorders with extreme genetic heterogeneity. *J Med Genet* 52:514-522.
28. Lange KI, Tsiropoulou S, Kucharska K, Blacque OE (2021) Interpreting the pathogenicity of Joubert syndrome missense variants in *Caenorhabditis elegans*. *Dis Model Mech* 14:dmm046631.
29. Morgan TH (1910) Sex limited inheritance in *Drosophila*. *Science* 32:120-122.
30. Bellen HJ, Yamamoto S (2015) Morgan's legacy: fruit flies and the functional annotation of conserved genes. *Cell* 163:12-14.
31. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q, Chen LX, Brandon RC, Rogers YH, Blazej RG, Champe M, Pfeiffer BD, Wan KH, Doyle C, Baxter EG, Helt G, Nelson CR, Gabor GL, Abril JF, Agbayani A, An HJ, Andrews-Pfannkoch C, Baldwin D, Ballew RM, Basu A, Baxendale J, Bayraktaroglu L, Beasley EM, Beeson KY, Benos PV, Berman BP, Bhandari D, Bolshakov S, Borkova D, Botchan MR, Bouck J, Brokstein P, Brottier P, Burtis KC, Busam DA, Butler H, Cadieu E, Center A, Chandra I, Cherry JM, Cawley S, Dahlke C, Davenport LB, Davies P, de Pablos B, Delcher A, Deng Z, Mays AD, Dew I, Dietz SM, Dodson K, Doup LE, Downes M, Dugan-Rocha S, Dunkov BC, Dunn P, Durbin KJ, Evangelista CC, Ferraz C, Ferreira S, Fleischmann W, Fosler C, Gabrielian AE, Garg NS, Gelbart WM, Glasser K, Glodek A, Gong F, Gorrell JH, Gu Z, Guan P, Harris M, Harris NL, Harvey D, Heiman TJ, Hernandez

- JR, Houck J, Hostin D, Houston KA, Howland TJ, Wei MH, Ibegwam C, Jalali M, Kalush F, Karpen GH, Ke Z, Kennison JA, Ketchum KA, Kimmel BE, Kodira CD, Kraft C, Kravitz S, Kulp D, Lai Z, Lasko P, Lei Y, Levitsky AA, Li J, Li Z, Liang Y, Lin X, Liu X, Mattei B, McIntosh TC, McLeod MP, McPherson D, Merkulov G, Milshina NV, Mobarry C, Morris J, Moshrefi A, Mount SM, Moy M, Murphy B, Murphy L, Muzny DM, Nelson DL, Nelson DR, Nelson KA, Nixon K, Nusskern DR, Pacleb JM, Palazzolo M, Pittman GS, Pan S, Pollard J, Puri V, Reese MG, Reinert K, Remington K, Saunders RD, Scheeler F, Shen H, Shue BC, Sidén-Kiamos I, Simpson M, Skupski MP, Smith T, Spier E, Spradling AC, Stapleton M, Strong R, Sun E, Svirskas R, Tector C, Turner R, Venter E, Wang AH, Wang X, Wang ZY, Wassarman DA, Weinstock GM, Weissenbach J, Williams SM, Woodage T, Worley KC, Wu D, Yang S, Yao QA, Ye J, Yeh RF, Zaveri JS, Zhan M, Zhang G, Zhao Q, Zheng L, Zheng XH, Zhong FN, Zhong W, Zhou X, Zhu S, Zhu X, Smith HO, Gibbs RA, Myers EW, Rubin GM, Venter JC (2000) The genome sequence of *Drosophila melanogaster*. *Science* 287:2185-2195.
32. Hu Y, Flockhart I, Vinayagam A, Bergwitz C, Berger B, Perrimon N, Mohr SE (2011) An integrative approach to ortholog prediction for disease-focused and other functional studies. *BMC Bioinformatics* 12:357.
33. Reiter LT, Potocki L, Chien S, Gribskov M, Bier E (2001) A systematic analysis of human disease-associated gene sequences in *Drosophila melanogaster*. *Genome Res* 11:1114-1125.
34. Bellen HJ, Wangler MF, Yamamoto S (2019) The fruit fly at the interface of diagnosis and pathogenic mechanisms of rare and common human diseases. *Hum Mol Genet* 28:R207-R214.
35. Harnish JM, Deal SL, Chao HT, Wangler MF, Yamamoto S (2019) In vivo functional study of disease-associated rare human variants using *Drosophila*. *J Vis Exp* 150:e59658.
36. Kanca O, Bellen HJ, Schnorrer F (2017) Gene tagging strategies to assess protein expression, localization, and function in *Drosophila*. *Genetics* 207:389-412.
37. Kanca O, Zirin J, Garcia-Marques J, Knight SM, Yang-Zhou D, Amador G, Chung H, Zuo Z, Ma L, He Y, Lin WW, Fang Y, Ge M, Yamamoto S, Schulze KL, Hu Y, Spradling AC, Mohr SE, Perrimon N, Bellen HJ (2019) An efficient CRISPR-based strategy to insert small and large fragments of DNA using short homology arms. *Elife* 8:e51539.
38. Blosser JA, Podolsky E, Lee D (2020) L-DOPA-induced dyskinesia in a genetic *Drosophila* model of Parkinson's disease. *Exp Neurol* 29:273-284.
39. Li-Kroeger D, Kanca O, Lee PT, Cowan S, Lee MT, Jaiswal M, Salazar JL, He Y, Zuo Z, Bellen HJ (2018) An expanded toolkit for gene tagging based on MiMIC and scarless CRISPR tagging in *Drosophila*. *Elife* 7:e38709.
40. Gratz SJ, Rubinstein CD, Harrison MM, Wildonger J, O'Connor-Giles KM (2015) CRISPR-Cas9 genome editing in *Drosophila*. *Curr Protoc Mol Biol* 111:31.2.1-31.2.20.
41. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, Musolf A, Li Q, Holzinger E, Karyadi D, Cannon-Albright LA, Teerlink CC, Stanford JL, Isaacs WB, Xu J, Cooney KA, Lange EM, Schleutker J, Carpten JD, Powell IJ, Cussenot O, Cancel-Tassin G, Giles GG, MacInnis RJ, Maier C, Hsieh CL, Wiklund F, Catalona WJ, Foulkes WD, Mandal D, Eeles RA, Kote-Jarai Z, Bustamante CD, Schaid DJ, Hastie T, Ostrander EA, Bailey-Wilson JE, Radivojac P, Thibodeau SN, Whittemore AS, Sieh W (2016) REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *Am J Hum Genet* 99:877-885.
42. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M (2019) CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* 47:D886-D894.
43. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ, von Mering C (2017) The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* 45:D362-D368.
44. Hu Y, Vinayagam A, Nand A, Comjean A, Chung V, Hao T, Mohr SE, Perrimon N (2018) Molecular interaction search tool (MIST): an integrated resource for mining gene and protein interaction data. *Nucleic Acids Res* 46:D567-D574.
45. Wang J, Al-Ouran R, Hu Y, Kim SY, Wan YW, Wangler MF, Yamamoto S, Chao HT, Comjean A, Mohr SE, UDN, Perrimon N, Liu Z, Bellen HJ (2017) MARRVEL: integration of human and model organism genetic resources to facilitate functional annotation of the human genome. *Am J Hum Genet* 100:843-853.
46. Wang J, Liu Z, Bellen HJ, Yamamoto S (2019) Navigating MARRVEL, a web-based tool that integrates human genomics and model organism genetics information. *J Vis Exp* 150:e59542.
47. Wang J, Mao D, Fazal F, Kim SY, Yamamoto S, Bellen H, Liu Z (2019) Using MARRVEL v1.2 for bioinformatics analysis of human genes and variant pathogenicity. *Curr Protoc Bioinformatics* 67:e85.
48. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings

- BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won HH, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG; Exome Aggregation Consortium (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536:285-291.
49. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J, Jang W, Katz K, Ovetsky M, Riley G, Sethi A, Tully R, Villamarin-Salomon R, Rubinstein W, Maglott DR (2016) ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* 44:D862-D868.
 50. MacDonald JR, Ziman R, Yuen RK, Feuk L, Scherer SW (2014) The database of genomic variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res* 42(Database issue):D986-D992.
 51. Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, Van Vooren S, Moreau Y, Pettett RM, Carter NP (2009) DECIPHER: database of chromosomal imbalance and phenotype in humans using ensembl resources. *Am J Hum Genet* 84:524-533.
 52. Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR, Fisk DG, Hirschman JE, Hitz BC, Karra K, Krieger CJ, Miyasato SR, Nash RS, Park J, Skrzypek MS, Simison M, Weng S, Wong ED (2012) *Saccharomyces* genome database: the genomics resource of budding yeast. *Nucleic Acids Res* 40(Database issue):D700-D705.
 53. McDowall MD, Harris MA, Lock A, Rutherford K, Staines DM, Bähler J, Kersey PJ, Oliver SG, Wood V (2015) PomBase 2015: updates to the fission yeast database. *Nucleic Acids Res* 43(Database issue):D656-D661.
 54. Howe KL, Bolt BJ, Cain S, Chan J, Chen WJ, Davis P, Done J, Down T, Gao S, Grove C, Harris TW, Kishore R, Lee R, Lomax J, Li Y, Muller HM, Nakamura C, Nuin P, Paulini M, Raciti D, Schindelman G, Stanley E, Tuli MA, Van Auken K, Wang D, Wang X, Williams G, Wright A, Yook K, Berriman M, Kersey P, Schedl T, Stein L, Sternberg PW (2016) WormBase 2016: expanding to enable helminth genomic research. *Nucleic Acids Res* 44:D774-D780.
 55. Attrill H, Falls K, Goodman JL, Millburn GH, Antonazzo G, Rey AJ, Marygold SJ; FlyBase Consortium (2016) FlyBase: establishing a gene group resource for *Drosophila melanogaster*. *Nucleic Acids Res* 44:D786-D792.
 56. Ruzicka L, Bradford YM, Frazer K, Howe DG, Paddock H, Ramachandran S, Singer A, Toro S, Van Slyke CE, Eagle AE, Fashena D, Kalita P, Knight J, Mani P, Martin R, Moxon SA, Pich C, Schaper K, Shao X, Westerfield M (2015) ZFIN, the zebrafish model organism database: updates and new directions. *Genesis* 53:498-509.
 57. Eppig JT, Blake JA, Bult CJ, Kadin JA, Richardson JE; Mouse Genome Database Group (2015) The mouse genome database (MGD): facilitating mouse as a model for human biology and disease. *Nucleic Acids Res* 43:D726-D736.
 58. Shimoyama M, De Pons J, Hayman GT, Lauderkind SJ, Liu W, Nigam R, Petri V, Smith JR, Tutaj M, Wang SJ, Worthey E, Dwinell M, Jacob H (2015) The rat genome database 2015: genomic, phenotypic and environmental variations and disease. *Nucleic Acids Res* 43(Database issue):D743-D750.
 59. Gunning AC, Fryer V, Fasham J, Crosby AH, Ellard S, Baple EL, Wright CF (2021) Assessing performance of pathogenicity predictors using clinically relevant variant datasets. *J Med Genet* 58:547-555.
 60. Gahl WA, Mulvihill JJ, Toro C, Markello TC, Wise AL, Ramoni RB, Adams DR, Tiftt CJ; UDN (2016) The NIH undiagnosed diseases program and network: applications to modern medicine. *Mol Genet Metab* 117:393-400.
 61. Ramoni RB, Mulvihill JJ, Adams DR, Allard P, Ashley EA, Bernstein JA, Gahl WA, Hamid R, Loscalzo J, McCray AT, Shashi V, Tiftt CJ; Undiagnosed Diseases Network, Wise AL (2017) The undiagnosed diseases network: accelerating discovery about health and disease. *Am J Hum Genet* 100:185-192.
 62. Chao HT, Davids M, Burke E, Pappas JG, Rosenfeld JA, McCarty AJ, Davis T, Wolfe L, Toro C, Tiftt C, Xia F, Stong N, Johnson TK, Warr CG; Undiagnosed Diseases Network, Yamamoto S, Adams DR, Markello TC, Gahl WA, Bellen HJ, Wangler MF, Malicdan MCV (2017) A syndromic neurodevelopmental disorder caused by de novo variants in EBF3. *Am J Hum Genet* 100:128-137.
 63. Taruscio D, Groft SC, Cederroth H, Melegh B, Lasko P, Kosaki K, Baynam G, McCray A, Gahl WA (2015) Undiagnosed diseases network international (UDNI): white paper for global actions to meet patient needs. *Mol Genet Metab* 116:223-225.
 64. Luo X, Rosenfeld JA, Yamamoto S, Harel T, Zuo Z, Hall M,

- Wierenga KJ, Pastore MT, Bartholomew D, Delgado MR, Rotenberg J, Lewis RA, Emrick L, Bacino CA, Eldomery MK, Coban Akdemir Z, Xia F, Yang Y, Lalani SR, Lotze T, Lupski JR, Lee B, Bellen HJ, Wangler MF; Members of the UDN (2017) Clinically severe CACNA1A alleles affect synaptic function and neurodegeneration differentially. *PLoS Genet* 13:e1006905.
65. Marcogliese PC, Shashi V, Spillmann RC, Stong N, Rosenfeld JA, Koenig MK, Martínez-Agosto JA, Herzog M, Chen AH, Dickson PI, Lin HJ, Vera MU, Salamon N, Graham JM Jr, Ortiz D, Infante E, Steyaert W, Dermaut B, Poppe B, Chung HL, Zuo Z, Lee PT, Kanca O, Xia F, Yang Y, Smith EC, Jasien J, Kanasagra S, Spiridigliozzi G, El-Dairi M, Lark R, Riley K, Koerber DD, Golden-Grant K; Program for Undiagnosed Diseases (UD-PrOZA); Undiagnosed Diseases Network, Yamamoto S, Wangler ME, Mirzaa G, Hemelsoet D, Lee B, Nelson SF, Goldstein DB, Bellen HJ, Pena LDM (2018) IRF2BPL is associated with neurological phenotypes. *Am J Hum Genet* 103:245-260.
66. Kanca O, Andrews JC, Lee PT, Patel C, Braddock SR, Slavotinek AM, Cohen JS, Gubbels CS, Aldinger KA, Williams J, Indaram M, Fatemi A, Yu TW, Agrawal PB, Vezina G, Simons C, Crawford J, Lau CC; Undiagnosed Diseases Network, Chung WK, Markello TC, Dobyns WB, Adams DR, Gahl WA, Wangler MF, Yamamoto S, Bellen HJ, Malicdan MCV (2019) De novo variants in WDR37 are associated with epilepsy, colobomas, dysmorphism, developmental delay, intellectual disability, and cerebellar hypoplasia. *Am J Hum Genet* 105:413-424.
67. Barish S, Barakat TS, Michel BC, Mashtalir N, Phillips JB, Valencia AM, Ugur B, Wegner J, Scott TM, Bostwick B; Undiagnosed Diseases Network, Murdock DR, Dai H, Perenthaler E, Nikoncuk A, van Slegtenhorst M, Brooks AS, Keren B, Nava C, Mignot C, Douglas J, Rodan L, Nowak C, Ellard S, Stals K, Lynch SA, Faoucher M, Lesca G, Edery P, Engleman KL, Zhou D, Thiffault I, Herriges J, Gass J, Louie RJ, Stolerman E, Washington C, Vetrini F, Otsubo A, Pratt VM, Conboy E, Treat K, Shannon N, Camacho J, Wakeling E, Yuan B, Chen CA, Rosenfeld JA, Westerfield M, Wangler M, Yamamoto S, Kadoch C, Scott DA, Bellen HJ (2020) BICRA, a SWI/SNF complex member, is associated with BAF-disorder related phenotypes in humans and model organisms. *Am J Hum Genet* 107:1096-1112.
68. Dutta D, Briere LC, Kanca O, Marcogliese PC, Walker MA, High FA, Vanderver A, Krier J, Carmichael N, Callahan C, Taft RJ, Simons C, Helman G, Network UD, Wangler MF, Yamamoto S, Sweetser DA, Bellen HJ (2020) De novo mutations in TOMM70, a receptor of the mitochondrial import translocase, cause neurological impairment. *Hum Mol Genet* 29:1568-1579.
69. Chung HL, Mao X, Wang H, Park YJ, Marcogliese PC, Rosenfeld JA, Burrage LC, Liu P, Murdock DR, Yamamoto S, Wangler MF; Undiagnosed Diseases Network, Chao HT, Long H, Feng L, Bacino CA, Bellen HJ, Xiao B (2020) De novo variants in CDK19 are associated with a syndrome involving intellectual disability and epileptic encephalopathy. *Am J Hum Genet* 106:717-725.
70. Chung HL, Wangler MF, Marcogliese PC, Jo J, Ravenscroft TA, Zuo Z, Duraine L, Sadeghzadeh S, Li-Kroeger D, Schmidt RE, Pestronk A, Rosenfeld JA, Burrage L, Herndon MJ, Chen S; Members of Undiagnosed Diseases Network, Shillington A, Vawter-Lee M, Hopkin R, Rodriguez-Smith J, Henrickson M, Lee B, Moser AB, Jones RO, Watkins P, Yoo T, Mar S, Choi M, Bucelli RC, Yamamoto S, Lee HK, Prada CE, Chae JH, Vogel TP, Bellen HJ (2020) Loss- or gain-of-function mutations in ACOX1 cause axonal loss via different mechanisms. *Neuron* 106:589-606.e6.
71. Goodman LD, Cope H, Nil Z, Ravenscroft TA, Charng WL, Lu S, Tien AC, Pfundt R, Koolen DA, Haaxma CA, Veenstra-Knol HE, Wassink-Ruiter JSK, Wevers MR, Jones M, Walsh LE, Klee VH, Theunis M, Legius E, Steel D, Barwick KES, Kurian MA, Mohammad SS, Dale RC, Terhal PA, van Binsbergen E, Kirmse B, Robinette B, Cogné B, Isidor B, Grebe TA, Kulch P, Hainline BE, Sapp K, Morava E, Klee EW, Macke EL, Trapane P, Spencer C, Si Y, Begtrup A, Moulton MJ, Dutta D, Kanca O; Undiagnosed Diseases Network, Wangler MF, Yamamoto S, Bellen HJ, Tan QK (2021) TNPO2 variants associate with human developmental delays, neurologic deficits, and dysmorphic features and alter TNPO2 activity in *Drosophila*. *Am J Hum Genet* 108:1669-1691.
72. Ravenscroft TA, Phillips JB, Fieg E, Bajikar SS, Peirce J, Wegner J, Luna AA, Fox EJ, Yan YL, Rosenfeld JA, Zirin J, Kanca O; Undiagnosed Diseases Network, Benke PJ, Cameron ES, Strehlow V, Platzer K, Jamra RA, Klöckner C, Osmond M, Licata T, Rojas S, Dyment D, Chong JSC, Lincoln S, Stoler JM, Postlethwait JH, Wangler MF, Yamamoto S, Krier J, Westerfield M, Bellen HJ (2021) Heterozygous loss-of-function variants significantly expand the phenotypes associated with loss of GDF11. *Genet Med* 23:1889-1900.
73. Luo X, Schoch K, Jangam SV, Bhavana VH, Graves HK, Kanasagra S, Jasien JM, Stong N, Keren B, Mignot C, Ravelli C; Undiagnosed Diseases Network, Bellen HJ, Wangler MF, Shashi V, Yamamoto S (2021) Rare deleterious de novo missense variants in Rnf2/Ring2 are associated with a neurodevelopmental

- tal disorder with unique clinical features. *Hum Mol Genet* 30:1283-1292.
74. Boulín T, Itani O, El Mouridi S, Leclercq-Blondel A, Gendrel M, Macnamara E, Soldatos A, Murphy JL, Gorman MP, Lindsey A, Shimada S, Turner D, Silverman GA, Baldrige D; Undiagnosed Diseases Network, Malicdan MC, Schedl T, Pak SC (2021) Functional analysis of a de novo variant in the neurodevelopment and generalized epilepsy disease gene NBEA. *Mol Genet Metab* 134:195-202.
75. Boycott KM, Campeau PM, Howley HE, Pavlidis P, Rogic S, Oriel C, Berman JN, Hamilton RM, Hicks GG, Lipshitz HD, Masson JY, Shoubridge EA, Junker A, Leroux MR, McMaster CR, Michaud JL, Turvey SE, Dyment D, Innes AM, van Karnebeek CD, Lehman A, Cohn RD, MacDonald IM, Rachubinski RA, Frosk P, Vandersteen A, Wozniak RW, Pena IA, Wen XY, Lacaze-Masmonteil T, Rankin C, Hieter P (2020) The Canadian rare diseases models and mechanisms (RDMM) network: connecting understudied genes to model organisms. *Am J Hum Genet* 106:143-152.
76. Beaulieu CL, Majewski J, Schwartzentruber J, Samuels ME, Fernandez BA, Bernier FP, Brudno M, Knoppers B, Marcadier J, Dyment D, Adam S, Bulman DE, Jones SJ, Avard D, Nguyen MT, Rousseau F, Marshall C, Wintle RF, Shen Y, Scherer SW; FORGE Canada Consortium, Friedman JM, Michaud JL, Boycott KM (2014) FORGE Canada Consortium: outcomes of a 2-year national rare-disease gene-discovery project. *Am J Hum Genet* 94:809-817.
77. van Karnebeek CD, Stockler-Ipsiroglu S (2014) Early identification of treatable inborn errors of metabolism in children with intellectual disability: the treatable intellectual disability endeavor protocol in British Columbia. *Paediatr Child Health* 19:469-471.
78. Farhan SMK, Nixon KCJ, Everest M, Edwards TN, Long S, Segal D, Knip MJ, Arts HH, Chakrabarti R, Wang J, Robinson JF, Lee D, Mirsattari SM, Rupar CA, Siu VM; FORGE Canada Consortium, Poulter MO, Hegele RA, Kramer JM (2017) Identification of a novel synaptic protein, TMT3C3, involved in periventricular nodular heterotopia with intellectual disability and epilepsy. *Hum Mol Genet* 26:4278-4289.
79. Salazar JL, Yamamoto S (2018) Integration of *Drosophila* and human genetics to understand notch signaling related diseases. *Adv Exp Med Biol* 1066:141-185.
80. Buske OJ, Schiettecatte F, Hutton B, Dumitriu S, Misyura A, Huang L, Hartley T, Girdea M, Sobreira N, Mungall C, Brudno M (2015) The matchmaker exchange API: automating patient matching through the exchange of structured phenotypic and genotypic profiles. *Hum Mutat* 36:922-927.
81. Philippakis AA, Azzariti DR, Beltran S, Brookes AJ, Brownstein CA, Brudno M, Brunner HG, Buske OJ, Carey K, Doll C, Dumitriu S, Dyke SO, den Dunnen JT, Firth HV, Gibbs RA, Girdea M, Gonzalez M, Haendel MA, Hamosh A, Holm IA, Huang L, Hurles ME, Hutton B, Krier JB, Misyura A, Mungall CJ, Paschall J, Paten B, Robinson PN, Schiettecatte F, Sobreira NL, Swaminathan GJ, Taschner PE, Terry SE, Washington NL, Züchner S, Boycott KM, Rehm HL (2015) The matchmaker exchange: a platform for rare disease gene discovery. *Hum Mutat* 36:915-921.
82. Azzariti DR, Hamosh A (2020) Genomic data sharing for novel Mendelian disease gene discovery: the matchmaker exchange. *Annu Rev Genomics Hum Genet* 21:305-326.
83. Sobreira N, Schiettecatte F, Valle D, Hamosh A (2015) GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum Mutat* 36:928-930.
84. Buske OJ, Girdea M, Dumitriu S, Gallinger B, Hartley T, Trang H, Misyura A, Friedman T, Beaulieu C, Bone WP, Links AE, Washington NL, Haendel MA, Robinson PN, Boerkoel CF, Adams D, Gahl WA, Boycott KM, Brudno M (2015) PhenomeCentral: a portal for phenotypic and genotypic matchmaking of patients with rare genetic diseases. *Hum Mutat* 36:931-940.
85. Chatzimichali EA, Brent S, Hutton B, Perrett D, Wright CE, Bevan AP, Hurles ME, Firth HV, Swaminathan GJ (2015) Facilitating collaboration in rare genetic disorders through effective matchmaking in DECIPHER. *Hum Mutat* 36:941-949.
86. Chong JX, Yu JH, Lorentzen P, Park KM, Jamal SM, Tabor HK, Rauch A, Saenz MS, Boltshauser E, Patterson KE, Nickerson DA, Bamshad MJ (2016) Gene discovery for Mendelian conditions via social networking: de novo variants in KDM1A cause developmental delay and distinctive facial features. *Genet Med* 18:788-795.
87. Adachi T, Kawamura K, Furusawa Y, Nishizaki Y, Imanishi N, Umehara S, Izumi K, Suematsu M (2017) Japan's initiative on rare and undiagnosed diseases (IRUD): towards an end to the diagnostic odyssey. *Eur J Hum Genet* 25:1025-1028.
88. Cacheiro P, Haendel MA, Smedley D; International Mouse Phenotyping Consortium and the Monarch Initiative (2019) New models for human disease from the International Mouse Phenotyping Consortium. *Mamm Genome* 30:143-150.
89. Volpato V, Webber C (2020) Addressing variability in iPSC-derived models of human disease: guidelines to promote reproducibility. *Dis Model Mech* 13:dmm042317.
90. Schutgens F, Clevers H (2020) Human organoids: tools for

-
- understanding biology and treating diseases. *Annu Rev Pathol* 15:211-234.
91. Wangler MF, Hu Y, Shulman JM (2017) *Drosophila* and genome-wide association studies: a review and resource for the functional dissection of human complex traits. *Dis Model Mech* 10:77-88.