

Model Organisms Facilitate Rare Disease Diagnosis and Therapeutic Research

Michael F. Wangler,^{*,†,‡,§,1} Shinya Yamamoto,^{*,‡,§,***,1} Hsiao-Tuan Chao,^{†,‡,††} Jennifer E. Posey,^{*} Monte Westerfield,^{**} John Postlethwait,^{**} Members of the Undiagnosed Diseases Network (UDN),²

Philip Hieter,^{§§} Kym M. Boycott,^{***} Philippe M. Campeau,^{†††} and Hugo J. Bellen^{*,‡,§,***,†††,3}

^{*}Department of Molecular and Human Genetics, [†]Department of Pediatrics, [§]Program in Developmental Biology, ^{**}Department of Neuroscience, ^{††}Department of Pediatrics, Section of Child Neurology, and ^{†††}Howard Hughes Medical Institute, Baylor College of Medicine (BCM), Houston, Texas 77030, [‡]Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, Texas 77030, ^{‡‡}Institute of Neuroscience, University of Oregon, Eugene, Oregon 97403, ^{§§}Michael Smith Laboratories, University of British Columbia, Vancouver, British Columbia V6T 1Z4C, Canada, ^{***}Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ontario K1H 8L1, Canada, ^{††††}Department of Pediatrics, University of Montreal, Quebec H3T 1C5, Canada

ORCID IDs: 0000-0001-5245-5910 (M.F.W.); 0000-0003-2172-8036 (S.Y.); 0000-0003-4814-6765 (J.E.P.); 0000-0002-5476-2137 (J.P.); 0000-0003-4186-8052 (K.M.B.); 0000-0001-9713-7107 (P.M.C.); 0000-0001-5992-5989 (H.J.B.)

ABSTRACT Efforts to identify the genetic underpinnings of rare undiagnosed diseases increasingly involve the use of next-generation sequencing and comparative genomic hybridization methods. These efforts are limited by a lack of knowledge regarding gene function, and an inability to predict the impact of genetic variation on the encoded protein function. Diagnostic challenges posed by undiagnosed diseases have solutions in model organism research, which provides a wealth of detailed biological information. Model organism geneticists are by necessity experts in particular genes, gene families, specific organs, and biological functions. Here, we review the current state of research into undiagnosed diseases, highlighting large efforts in North America and internationally, including the Undiagnosed Diseases Network (UDN) (Supplemental Material, File S1) and UDN International (UDNI), the Centers for Mendelian Genomics (CMG), and the Canadian Rare Diseases Models and Mechanisms Network (RDMM). We discuss how merging human genetics with model organism research guides experimental studies to solve these medical mysteries, gain new insights into disease pathogenesis, and uncover new therapeutic strategies.

KEYWORDS functional genomics; *Drosophila*; zebrafish; human; genetic diseases; whole-exome sequencing; diagnostics

DISEASE diagnosis, one of the primary goals of medicine, is a process to discover the origin or nature of disease in a particular patient. Without an accurate diagnosis, we can neither identify the cause nor design an effective treatment strategy to suppress or ameliorate the condition. Many patients with rare conditions undergo a “diagnostic odyssey,” a long and frustrating journey to obtain an accurate diagnosis. Several organizations have estimated that ~80% of these undiagnosed diseases have a genetic origin (IOM 2010).

Advanced technology is increasingly important for diagnosis and has been transformative for rare Mendelian disorders (Gonzaga-Jauregui *et al.* 2012). Whole-exome sequencing (WES) has emerged as an effective diagnostic tool for patients who have undergone comparative genomic hybridization methods to rule out chromosomal abnormalities (Need *et al.* 2012; Yang *et al.* 2013, 2014; Lee *et al.* 2014; Retterer *et al.* 2016). However, sequencing also uncovers many rare variants for which the functional impact is not known (Coventry *et al.* 2010; Lupski *et al.* 2011), and interpretation can be difficult as healthy individuals often carry alleles that would be deleterious or lethal when homozygous (Gao *et al.* 2015). Model organism studies can be a pivotal resource for understanding these variants (Yamamoto *et al.* 2014). In many cases, studying genes simultaneously in humans and model organisms provides additional insight into the cause of a particular disease,

Copyright © 2017 by the Genetics Society of America

doi: <https://doi.org/10.1534/genetics.117.203067>

Manuscript received April 14, 2017; accepted for publication July 6, 2017.

Available freely online through the author-supported open access option.

¹These authors contributed equally to this work.

²A full list of members is provided as Supplementary Material.

³Corresponding author: Baylor College of Medicine, 1250 Moursund St., Suite 1125, Houston, TX 77030. E-mail: hbellen@bcm.edu

thereby contributing to an understanding of the pathogenic processes.

Studying human genomic variants in model organisms is facilitated by collaboration between clinicians and model organism researchers (Hieter and Boycott 2014; McGurk *et al.* 2015; Bonini and Berger 2017; Chao *et al.* 2017b; Manolio *et al.* 2017). We highlighted the importance of model organisms, particularly the fruit fly *Drosophila melanogaster*, in the context of human biology (Wangler *et al.* 2015), and described how model organism researchers are becoming integral to disease research. The questions we raised about the future of model organism research, and subsequent discussion at [The Allied Genetics Conference in 2016](#), which included the [Director of the National Institutes of Health \(NIH\)](#), made clear the priority that model organism research should have in biomedical research.

In this article, we show how model organism research can be integral to unraveling the molecular mechanisms of disease. We begin this discussion with illustrative examples that document the value model organisms have brought to human disease research, including Mendelian diseases and cancer. We highlight three approaches for engaging model organism laboratories with the expanding list of human genes and variants requiring functional annotation. We primarily focus on two model organisms—*Drosophila* and zebrafish (Figure 1)—and highlight large-scale efforts in the United States (US) and Canada that integrate model organisms into human disease diagnosis and therapy development. Projects such as the [Undiagnosed Diseases Network \(UDN\)](#) with its centralized Model Organisms Screening Center (MOSC) (Gahl *et al.* 2016), the [Centers for Mendelian Genomics \(CMG\)](#) (Chong *et al.* 2015) and the [Canadian Rare Diseases Models and Mechanisms Network \(RDMM\)](#) provide exciting opportunities for model organism researchers, human geneticists, and clinicians to collaborate to determine the potential pathogenicity of disease-linked rare variants, gain deeper pathophysiological understanding of diseases, and discover potential treatments and therapies for patients and families. Lessons learned from these complementary projects will likely lead to the establishment of an efficient framework for facilitating the diagnosis and study of rare diseases.

Evolutionary Conservation of Genes and Pathways: Orthologs and Phenologs

Biologists often focus on differences among organisms. Indeed, the diversity of life is staggering. Our first impulse is to assume that this diversity arises from different developmental mechanisms, and that these processes must rely on different genes, signaling pathways, and molecular events. For example, it is difficult to comprehend that a fruit fly or a fish could share many genes and signaling pathways with humans. They differ in so many morphological features, and their physiology seems so different, that we tend to imagine there are no or very few commonalities. It was therefore surprising when it became apparent in the 1980s and 1990s that numerous genes

are conserved across species and are used repeatedly in many different contexts.

In the past 25 years it became apparent that many biological processes are evolutionarily conserved, and that discoveries in yeast, worms, and flies have a direct relevance to vertebrate and human biology. The 2016 Nobel Prize in Physiology or Medicine was awarded to Yoshinori Ohsumi for his pioneering genetic studies in yeast that led to the discovery and mechanistic understanding of autophagy, a fundamental process involved in cancer, neurological disorders, and other diseases (Takeshige *et al.* 1992; Tsukada and Ohsumi 1993; Mizushima *et al.* 1998). Similar major discoveries with yeast include the identification of core mediators of the cell cycle (Hartwell *et al.* 1970, 1974; Nurse and Bissett 1981; Balter and Vogel 2001) and mechanistic understanding of vesicle trafficking (Novick *et al.* 1980) (2001 and 2013 Nobel Prizes, respectively). Other insights arose from the study of invertebrate animals, including the discovery of molecular mechanisms of learning and memory in the sea slug *Aplysia* (Kandel 2001) (2000 Nobel Prize), RNA interference (Fire *et al.* 1998) and apoptotic pathways (Ellis and Horvitz 1986) in the nematode worm *Caenorhabditis elegans* (2006 and 2002 Nobel Prizes, respectively), and identification of conserved developmental genes and pathways (Lewis 1978; Nusslein-Volhard and Wieschaus 1980) and innate immunity mechanisms (Lemaitre *et al.* 1996) in *Drosophila* (1995 and 2011 Nobel Prizes, respectively) had major impacts on biomedical research. Basic knowledge obtained from model organism research is now being translated to mammalian contexts with remarkable speed. For example, control of organ growth and size was shown to be regulated by genes of the Hippo signaling pathway in flies in 2003 (Halder and Johnson 2011). It took only a few years to show that this pathway is conserved in humans and that it plays a critical role in organ growth, cancer, and regeneration (Saucedo and Edgar 2007).

Conservation of genes and processes extends to numerous physiological pathways. For example, genes encoding proteins that play important roles in synaptic transmission, action potential propagation, and many other forms of neuronal communication or muscle function are highly conserved and play similar roles in worms, flies, and mammals (Bellen *et al.* 2010). Similarly, the following sections illustrate elegant examples of what can be learned about human biology from invertebrates and fish. Obviously, not all genes are evolutionarily conserved from invertebrates to human; ~35% of human genes have no obvious orthologs in flies (Wangler *et al.* 2015), but can nevertheless be studied in vertebrates like zebrafish and mice. Indeed, zebrafish allow more robust phenotypic comparisons than flies for organs such as the heart, gut, liver, pancreas, and kidney, and have emerged as a key model system to study processes such as hematopoiesis and neural crest development (Boatman *et al.* 2013; Mayor and Theveneau 2013).

One consideration when selecting an organism for disease studies is its overall genomic architecture. After the divergence of the human and fly lineages, two rounds of whole

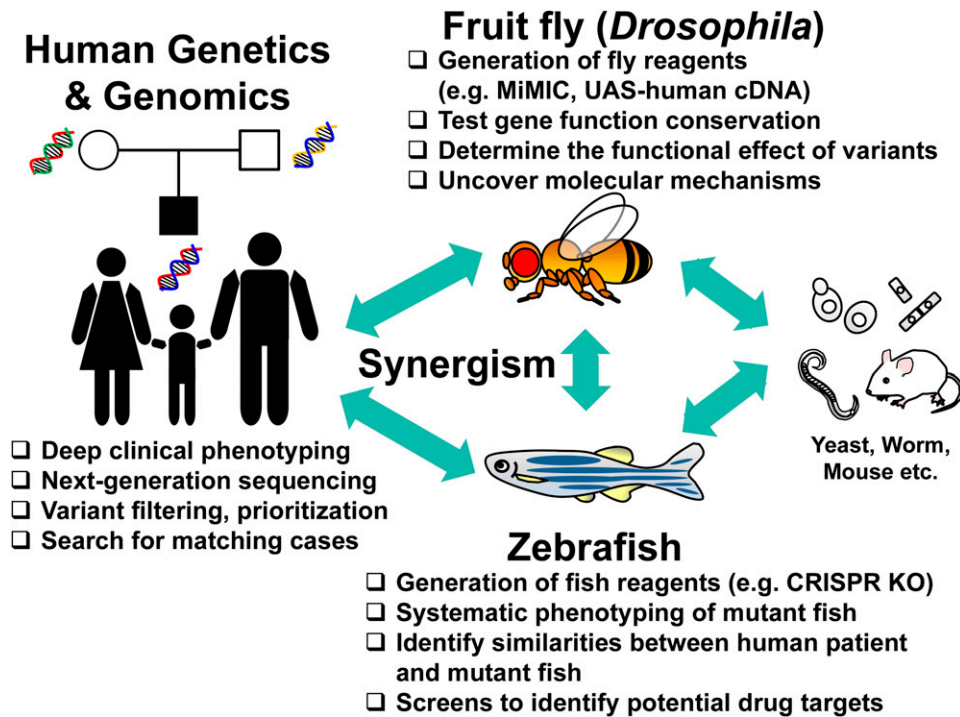


Figure 1 Collaborations among clinicians, human geneticists and model organism researchers facilitate diagnosis and studies of undiagnosed conditions. Candidate causative genes and variants identified from a patient with an undiagnosed disease can be explored in a number of genetic model organisms. Using state-of-the-art genome engineering technologies in these model systems, one can assess whether the variants of interest lead to functional consequences *in vivo*, and obtain phenotypic information that may directly or indirectly relate to the patient's condition. Integration of biological information from multiple species can be complementary or/and confirmatory.

genome duplication likely occurred in the last common ancestor to all vertebrates (Holland *et al.* 1994; Dehal and Boore 2005). As a consequence, humans often have two to four co-orthologs of individual fly genes: for example, four *HOX* gene clusters in humans vs. one in flies. According to the principle of subfunctionalization (Force *et al.* 1999), the human co-orthologs likely share among them ancestral functions represented by the single-copy fly genes. Likewise, teleosts (majority of bony fish) experienced an additional genome duplication that provided zebrafish with two copies of ~25% of human genes (Howe *et al.* 2013). The direct orthology of zebrafish genes to human orthologs is usually unambiguous (Braasch *et al.* 2016), which facilitates connection of zebrafish experimental findings to human biology.

Flies and fish have tissues and organs that are specific to their phyla. One may wonder if wings of flies or dorsal fins in fish can teach us something that is directly relevant to human development or physiology. Mutations that affect the development of wings in flies pioneered the characterization of key developmental pathways such as the Notch (mutants have a “notch” in the wing due to loss of distal wing tissue) and Wingless (Wnt) signaling pathways. Interestingly, pathogenic variants in many genes in the Wnt pathway in humans cause Robinow syndrome—a severe skeletal dysplasia with short stature, short limbs and digits, dysmorphic facial features, and a variety of other problems including cardiac anomalies (MIM# 180700, 268310, 616331, and 616894). Pathogenic variants in *WNT5A* (a homolog of *Drosophila Wnt5*) (Person *et al.* 2010), its receptor *ROR2* (Afzal *et al.* 2000; van Bokhoven *et al.* 2000), as well as mutations in signal mediators *DVL1* (Bunn *et al.* 2015; White *et al.* 2015) and *DVL3* (White *et al.* 2016) (homologs of the fly *dishevelled* gene) cause

Robinow syndrome. Thus, although flies do not have an endoskeleton like in vertebrates, conserved genetic machinery that regulates wing patterning in *Drosophila* is responsible for skeletal and craniofacial defects in humans. This parsimony arises because evolutionarily conserved molecular pathways are reused in different cellular contexts and tissues.

These observations led to the concept of “phenologs” or “orthologous phenotypes” (McGary *et al.* 2011). Orthologous genes often cause different phenotypes in different species, yet the proteins encoded by these genes have similar molecular functions. To discover phenologs, one can cluster genes known to cause a human disease and determine the phenotypes caused by mutations in the corresponding genes in model organisms. For example, human genes associated with breast cancer overlap with *C. elegans* genes that mutate to cause a high frequency of male offspring. While the phenotypes in humans (breast cancer) and worms (increased male offspring) do not appear related, they derive from common underlying genetic defects (subset of genes involved in DNA damage response). Once these nonobvious relationships are documented, then additional genetic links can be made between models and humans (McGary *et al.* 2011). For genes that are evolutionarily conserved, simple model organisms with short generation times and amenable to inexpensive and efficient genetic manipulations can yield rapid insights into basic biological functions through detailed *in vivo* studies (Lehner 2013).

Finally, when using one species to understand another, a key experiment is to determine how interchangeable genes are between species if one wishes to assess whether a human variant may be pathogenic. Several studies performed in yeast indicate that many yeast genes (33–47%, depending on the

study and assay) can be replaced by their human counterparts (Hamza *et al.* 2015; Kachroo *et al.* 2015). Many human genes can also significantly or fully rescue fly mutant phenotypes, allowing tests of the effects of human variants on protein function *in vivo*. Although this is based on a biased sample of ~25 *Drosophila* genes, it suggests that this approach can be productive. One of the most spectacular examples of a cross-species gene rescue is the *atonal* gene, associated with deafness, blindness, and loss of proprioception in flies as well as in mammals (Jarman *et al.* 1993). The fly *atonal* gene fully rescues loss of the mouse *Atoh1* gene (Wang *et al.* 2002). Similarly, the mouse *Atoh1* gene rescues almost all phenotypes associated with the loss of *atonal* in flies (Wang *et al.* 2002). These results are remarkable considering that *atonal* and *Atoh1* share only a short stretch of sequence similarity in the DNA binding domain, and most orthology prediction programs do not identify these genes as strong ortholog candidates. We anticipate that, in many cases, testing human genes in model organisms will provide an experimentally productive platform for analysis of gene function, variant pathogenicity, and disease pathogenesis, even when tested in very different biological contexts.

Model Organism Studies Can Facilitate Diagnosis and Treatment

Drosophila, TRP channels and a spectrum of Mendelian disorders

One of the earliest examples of a *Drosophila* mutant that informed studies of an extensive gene family implicated in numerous human disorders comes from the fly *transient receptor potential* (*trp*) gene. Studies in the 1960s identified a nonphototoxic fly mutant with a distinct electroretinogram phenotype (Cosens and Manning 1969). Subsequent studies revealed that the affected gene encodes a pore-forming cation channel that is the founding member of a large, diverse family of evolutionarily conserved proteins known as TRP channels (Montell 2005). TRP channel family members have a weak voltage-sensing transmembrane domain, a selectivity filter, and diverse N- and C-terminal domains that provide versatile activation mechanisms. TRP family members encode unique channels that respond to light, sound, chemicals, temperature, pressure, or tactile stimuli, and can integrate multiple signals. The human genome contains 28 TRP channel family members (Gray *et al.* 2015), 11 of which are implicated in Mendelian disorders. These disorders have diverse clinical presentations, different inheritance patterns, and affect distinct tissues (Table 1). Insights from *Drosophila* led to an understanding of TRP channel function, which laid the groundwork for defining the cause of these disorders as resulting from gain-of-function (GOF), loss-of-function (LOF) or modulation (alteration)-of-function mechanisms. The most extreme disease is caused by *TRPV4* (Dai *et al.* 2010)—a gene that underlies a broad spectrum of skeletal and nervous system phenotypes. An autosomal dominant condition *brachyolmia type 3* (MIM# 113500), which is characterized by short trunk, short stature, and scoliosis, as well as a series of

more severe skeletal dysplasias, is due to GOF missense mutations in *TRPV4* that causes the channels to be activated by stimuli to which they would not normally respond (Nishimura *et al.* 2012). Other variants in *TRPV4* cause *congenital distal spinal muscular atrophy* (MIM# 600175) and *Charcot-Marie-Tooth disease type 2C* (MIM# 606071) (Deng *et al.* 2010; Landouere *et al.* 2010; Nilius and Owsianik 2010). Heterozygous mutations leading to these neurological phenotypes appear to have a complex impact on channel function and appear as GOF or LOF in different assays (Auer-Grumbach *et al.* 2010; Deng *et al.* 2010; Landouere *et al.* 2010). These functional insights would likely be identified more slowly without the mechanistic understanding of the channels provided by studies in *Drosophila*.

Zebrafish and melanoma

Research on zebrafish complements research on flies because zebrafish share vertebrate-specific features with human, such as similar organ structures. Zebrafish offer a number of experimental advantages for investigation of human disease mechanisms and therapeutic strategies (Phillips and Westerfield 2014). Chief among these are the ease of genetic manipulation, the ability to replace zebrafish genes with human genes, sensitive phenotypic analyses, and the capacity to conduct high-throughput screens of small molecules for potential therapeutics.

Examples discussed here involve human melanomas, which are genetically diverse cancers. This genetic heterogeneity makes it difficult to discover which gene mutations are primary drivers of oncogenesis, and which are critical modifiers that promote metastatic disease. A valine-to-glutamic acid mutation at position 600 in the human *BRAF* gene is the most frequent mutation driving human melanomas, but mutations in other genes are required for metastasis (Pollock *et al.* 2003). In searches for “second hit” loci in *BRAF*^{V600E} melanomas, a region of chromosome 1q21 was identified as a key driver of metastasis, but the presence of >50 genes in the identified interval made it difficult to determine the primary driver (Lin *et al.* 2008). Testing genes in the corresponding zebrafish chromosomal interval revealed a single gene, *SET domain, bifurcated 1* (*SETDB1*) that cooperates with *BRAF*^{V600E} to drive melanoma formation and growth (Ceol *et al.* 2011). The 1q21 interval was subsequently linked to familial melanoma in human (Macgregor *et al.* 2011), establishing *SETDB1* as a major human melanoma oncogene. A high-throughput chemical screen of these zebrafish with ~2000 substances identified inhibitors of dihydroorotate dehydrogenase (DHODH), such as the anti-inflammatory drug leflunomide, as suppressors of neural crest development and melanoma formation (White *et al.* 2011). This work in zebrafish led to [Phase I/II clinical trials of leflunomide in combination with a previously studied BRAF-inhibitor for treatment of melanoma](#). The ability to create sensitized zebrafish lines with human genes, coupled with the ability to screen thousands of compounds for their capacity to rescue disease phenotypes, illustrates the power of zebrafish studies for illuminating human disease pathogenesis and revealing novel drug targets.

Table 1 Select Mendelian disorders from TRP channel family members

Gene Name	Inheritance	Disorder (MIM#, name)	Affected Tissues	Proposed Genetic Mechanism
<i>MCOLN1</i>	AR	#252650, Mucopolidosis IV	Neurologic, ocular, gastrointestinal	LOF
<i>TRPM1</i>	AR	#613216, Night blindness, congenital stationary (complete), 1C	Photoreceptors	LOF
<i>TRPM6</i>	AR	#602014, Hypomagnesemia 1, intestinal	Gastrointestinal	LOF
<i>PKD2</i>	AD	#613095, Polycystic kidney disease 2	Renal	LOF (haploinsufficiency)
<i>TRPA1</i>	AD	#615040, Episodic pain syndrome, familial	Peripheral nervous system	GOF
<i>TRPC3</i>	AD	#616410, ?Spinocerebellar ataxia 41	Central nervous system	Toxic GOF
<i>TRPC6</i>	AD	#603965, Glomerulosclerosis, focal segmental, 2	Renal	GOF
<i>TRPM4</i>	AD	#604559, Progressive familial heart block, type IB	Cardiac	GOF
<i>TRPM7</i>	AD	#105500, Amyotrophic lateral sclerosis-parkinsonism/dementia complex, susceptibility	Central nervous system	Modulation of function?
<i>TRPV3</i>	AD	#616400, ?palmoplantar keratoderma	Skin	GOF
<i>TRPV4</i>	AD	#614594, Olmsted syndrome Two spectra of Mendelian disorders- skeletal and neurologic (see text)	Skeletal in one group, neurologic in another group	Complex (see text)

AR, Autosomal recessive; AD, autosomal dominant; LOF, loss of function; GOF, gain of function.

Model Organisms and Medical Diagnosis

These examples of fly and fish research illustrate the often under-recognized foundational contributions to human health that model organism research provides. Rapid advances in medical technology have expanded the need to incorporate these insights into the practice of medicine. Next-generation sequencing, particularly WES, transformed the diagnosis of rare disease (Gonzaga-Jauregui *et al.* 2012; Yang *et al.* 2014; Retterer *et al.* 2016). Furthermore, the coupling of genomic data with phenotypes extracted from medical electronic health records enables large-scale analysis of the clinical impact of rare variants (Abul-Husn *et al.* 2016; Dewey *et al.* 2016). However, this new sequencing technology presents challenges. Each personal genome contains hundreds of rare and unique variants that can have dramatic functional significance (Coventry *et al.* 2010), but most are without obvious effect. Their sheer number can make interpretation difficult, leaving the majority of individuals with apparent Mendelian diseases undiagnosed (Need *et al.* 2012; Yang *et al.* 2014; Posey *et al.* 2016; Retterer *et al.* 2016). Merging results from many genomes has significantly increased the power of sequence analysis, but, at the same time, has dramatically increased the need to study the impact of gene variants on protein function. We argued for an important role of model organism research studies in this effort (Bellen and Yamamoto 2015). Here, we present several examples of the early success of this approach.

Centers for Mendelian Genomics

The study of Mendelian disorders has yielded many significant insights into human biology, starting with the recognition of *alkaptonuria* (MIM# 203500) as an example of Mendel's laws operating in humans (Garrod 1902, 1923). The availability of WES propelled novel disease gene and variant discovery, leading to in-depth cataloging of the function of thousands of genes in the human genome, and providing insights into the mechanisms of both rare Mendelian disease and more common, complex traits (Albert *et al.* 2007; Levy *et al.* 2007; Wheeler *et al.* 2008; Bainbridge *et al.* 2010; Lupski *et al.* 2010; Blair *et al.* 2013; Wu *et al.* 2015). Nonetheless, the diagnostic rate (% of cases in which a pathogenic mutation in a known disease gene was identified) of clinical WES remains at 25–30% (Yang *et al.* 2013, 2014; Lee *et al.* 2014; Farwell *et al.* 2015; Retterer *et al.* 2016; Eldomery *et al.* 2017; Wenger *et al.* 2017). The diagnostic yield varies depending on age and how the patients were selected, with diagnostic rates as high as 57% in infants in a tertiary pediatric center vs. 17.5% for adults (Posey *et al.* 2016; Stark *et al.* 2016). A review of [Online Mendelian Inheritance in Man \(OMIM\)](#) reveals 8487 human disease phenotypes with suspected Mendelian basis, of which 6017 (70.9%) have a known molecular basis.

For the past 6 years, the National Institutes of Health (NIH) has funded several sequencing centers to use next-generation

sequencing techniques, primarily whole-exome sequencing, to identify and describe as many Mendelian disorders as possible (Bamshad *et al.* 2012). The first 4 years of funding to three independent CMGs (Baylor Hopkins CMG, University of Washington CMG, and Yale CMG) produced discoveries of 956 genes, including 375 not previously associated with human disease (Chong *et al.* 2015). The overarching approach of the CMGs is to identify individuals, families, and cohorts with conditions that have a high probability of being Mendelian, and to apply WES to identify potentially causative disease genes and variants. The CMGs benefitted significantly from development of several computational tools that enable disease gene and variant identification, phenotype-driven analysis, and identification of clinicians and researchers studying the same gene or phenotype (Table 2). One tool essential for the development of these and other studies is *GeneMatcher*. By allowing researchers to identify each other based on their interest in a gene, *GeneMatcher* is an active collaboration tool in human genetics. Importantly, this tool also allows model organism researchers to identify human geneticists and clinicians interested in collaborations (Sobreira *et al.* 2015b).

The potential power of integrating model organism studies into interrogations of the functions of candidate disease genes and variants identified by the CMG is illustrated by the analysis of human homologs of a set of 153 *Drosophila* genes required for nervous system development or function (Yamamoto *et al.* 2014). For ~30% of these, a human disease had already been associated with the homologs of the *Drosophila* gene. Interrogation of exome variant data from human homologs of the 153 *Drosophila* genes led to molecular diagnoses in six families. In one example variants in a known disease gene for early onset retinal disease (*CRX*), caused a later onset retinal phenotype not previously seen with *CRX* variants. Such a finding is often referred to as a “phenotypic expansion” of a known disease-linked gene. In another example, a novel disease gene whose biological function was poorly characterized (*ANKLE2*) was discovered to be linked to *autosomal recessive primary microcephaly 16* (MIM# 616681). Within 2 years of the publication of the screen of these 153 fly genes, an additional ~15% of the genes have become linked to Mendelian diseases. A key insight from this study led to the realization that essential genes in *Drosophila* with more than one human homolog are eight times more likely to be associated with Mendelian diseases, allowing the prioritization of disease candidate genes (Yamamoto *et al.* 2014).

There are now several additional examples of novel disease-gene discoveries by the CMG that benefited from collaborative functional studies in *Drosophila*. These include: *metabolic encephalomyopathic crises, recurrent, with rhabdomyolysis, cardiac arrhythmias, and neurodegeneration* (MIM# 616878) and *TANGO2* (Bard *et al.* 2006; Kremer *et al.* 2016; Lalani *et al.* 2016); *cerebellar atrophy, visual impairment, and psychomotor*

Table 2 Bioinformatic tools and databases developed by the Centers for Mendelian Genomics

Function	Name	Description	Interface for Fly Researchers	References
Variant identification	HMZDelFinder	Identification of homozygous and hemizygous deletions from exome variant data		Gambin <i>et al.</i> (2017)
Variant identification	DNMfinder	Identification of <i>de novo</i> variants from trio-WES data		Eldomey <i>et al.</i> (2017)
Phenotype database/analysis	PhenoDB https://www.mendeliangenomics.org/	Searchable database of clinical phenotypes for individuals and their relatives, phenotypes reported by expert clinicians	Rapid identification of patients with phenotypes of potential interest for further study	Hamosh <i>et al.</i> (2013), Sobreira <i>et al.</i> (2015a)
Phenotype database/analysis	OMIMExplorer https://omimexplorer.research.bcm.edu/	Prioritization of exome variants based on clinical phenotype data		James <i>et al.</i> (2016)
Phenotype database/analysis	Geno2MP http://geno2mp.gs.washington.edu/Geno2MP	Searchable database of genotype and phenotype data	Gene and variant-based searching allows identification of related human phenotypes	Chong <i>et al.</i> (2016a)
Gene and/or phenotype matching	GeneMatcher https://genematcher.org/	Identification of collaborators studying the same gene and/or phenotype, functions as a MME node	Identification of physicians and/or scientists who may have study patients with a relevant genotype or phenotype	Sobreira <i>et al.</i> (2015a,b)
Gene and/or phenotype matching	MyGene2 http://www.mygene2.org/	Phenotype and genotype database that includes patient-entered data, supports data matching, functions as a MME node	Identification of physicians and/or scientists who may have study patients with a relevant genotype or phenotype	Chong <i>et al.</i> (2016b)

retardation (MIM# 616875) and *EMC1* (Harel *et al.* 2016a); *Harel-Yoon syndrome* (MIM# 617183) and *ATAD3A* (Harel *et al.* 2016b); *lissencephaly 6 with microcephaly* (MIM# 616212) and *KATNB1* (Mishra-Gorur *et al.* 2014); *steroid-resistant nephrotic syndrome* (MIM# 615573) and *ADCK4* (Ashraf *et al.* 2013); and *NRD1* and *OGDHL* as novel disease genes associated with nervous system phenotypes (Yamamoto *et al.* 2014; Yoon *et al.* 2017). A key characteristic of these efforts is the bidirectional impact of insights: candidate genes identified from sequencing human patients can be studied with reverse genetics in flies, and orthologs of essential *Drosophila* genes can be examined in human genomic data. These bidirectional approaches provide a synergistic effect (Figure 1) (Wangler *et al.* 2015).

Studies of other model organisms also support novel disease gene discovery efforts. For example, the discovery of an association between variation in *FAT1* and a human phenotype of steroid-resistant nephropathy with neurologic involvement benefitted from studies in zebrafish and mice, demonstrating that loss of *Fat1* leads to abnormal podocyte and brain development (Ciani *et al.* 2003; Skouloudaki *et al.* 2009; Gee *et al.* 2016). Both organisms also contributed to the identification of additional genes associated with abnormal brain structure or function, such as *CLP1* in *pontocerebellar hypoplasia* (MIM# 615803) (Karaca *et al.* 2014; Schaffer *et al.* 2014), *KATNB1* in *lissencephaly 6 with microcephaly* (MIM# 616212) (Mishra-Gorur *et al.* 2014), and *FMN2* in *nonsyndromic intellectual disability* (MIM# 616193) (Law *et al.* 2014). Both organisms also proved useful in the study of non-neurologic phenotypes, such as the role of *ZSWIM6* in *acromelic frontonasal dysostosis* (MIM# 603671) (Smith *et al.* 2014), *CAV1*, and *neonatal lipodystrophy* (Garg *et al.* 2015), *WDPCP* (MIM# 217085 and 615992) and *INTU* in *ciliopathy syndromes* (Toriyama *et al.* 2016), and *FOXE* in *thoracic aortic aneurysms and dissections* (Kuang *et al.* 2016).

Undiagnosed diseases program and network

Patients with a rare disease often remain undiagnosed or misdiagnosed for years (Markello *et al.* 2011). The “diagnostic odyssey” experienced by many of these patients is an emotional and financial burden to patients, families, and the health-care system. Addressing this problem was part of the motivation for the establishment of the Undiagnosed Diseases Program (UDP), a NIH intramural program led by the NHGRI (National Human Genome Research Institute) (Gahl *et al.* 2012). The UDP model is unique for several reasons. First, the patient is the applicant and the patient’s participation drives the process, which is in contrast to most other clinical studies in which referring physicians are the primary applicants. Second, the study starts with a broad and in-depth clinical evaluation. Third, this model of systematic and tailored clinical phenotyping for single, unique cases is followed by genomic assessments using the latest technologies. By accepting applications from individuals with diverse undiagnosed conditions, sorting through

thousands of applications to identify patients with strong objective findings, performing extensive phenotyping and clinical testing on accepted patients at the NIH Clinical Center, and performing WES in the majority of the cases, the UDP is able to provide diagnoses for patients and families who have exhausted traditional diagnostic modalities (Tift and Adams 2014). Successes include the description of new disease genes, such as *NT5E* in arterial calcifications (St Hilaire *et al.* 2011). The genetic definition of this rare condition implicated adenosine metabolism in more common cases of vascular pathology (Markello *et al.* 2011). In addition, the effort has uncovered cases of known conditions that were missed because of misleading laboratory data or unusual phenotypic features that obscured the diagnosis (Gahl *et al.* 2012), and revealed “phenotypic expansion” for disease genes such as *AFG3L2* (Pierson *et al.* 2011). These efforts are valuable for the medical care of patients, and provide a better understanding of the unanticipated functions of genes.

The successes of the UDP continued after expansion to the UDN (Gahl *et al.* 2016; Ramoni *et al.* 2017). The UDN follows a similar strategy as the UDP but is decentralized and involves seven Clinical Sites [Stanford Medicine, UCLA School of Medicine, Baylor College of Medicine (BCM) with Texas Children’s Hospital, Vanderbilt University Medical Center, Duke Medicine with Columbia University Medical Center, Harvard Teaching Hospitals (Boston Children’s Hospital, Brigham and Women’s Hospital, Massachusetts General Hospital) and NIH], two Sequencing Cores (BCM, HudsonAlpha with Illumina), one Model Organisms Screening Center (MOSC, BCM with University of Oregon), one Metabolomics Core (Pacific Northwest National Laboratory with Oregon Health & Science University), and one Coordination Center (Harvard Medical School) (Figure 2). Recent discoveries by the UDN provided key insights into phenotypic expansion of *NR5A1* in human sex determination (Bashamboo *et al.* 2016), *ASXL2* and *Shashi-Pena syndrome* (MIM# 617190) (Shashi *et al.* 2016), and *EBF3* and *hypotoia, ataxia and delayed development syndrome* (MIM# 617330) (Chao *et al.* 2017a), a gene discovery that was also made simultaneously by the CMG and the Deciphering Developmental Disorders Study (DDDS) (Harms *et al.* 2017; Sleven *et al.* 2017). Use of WES, and, more recently, whole-genome sequencing (WGS) by the UDP and UDN are key in providing diagnostic results for patients who would have remained undiagnosed.

Given its approach as a network providing in-depth clinical evaluation for the most difficult to diagnose cases, the UDN faces several challenges. For example, genomic discovery in single patients is dramatically hampered by the lack of statistical power. In this context, the need for the UDN to functionally validate candidate gene variants is paramount. Therefore, extensive phenotyping (performed by the Clinical Sites and Metabolomics Core), precise genotyping (performed by the Sequencing Cores), functional exploration of candidate genes and variants (performed by

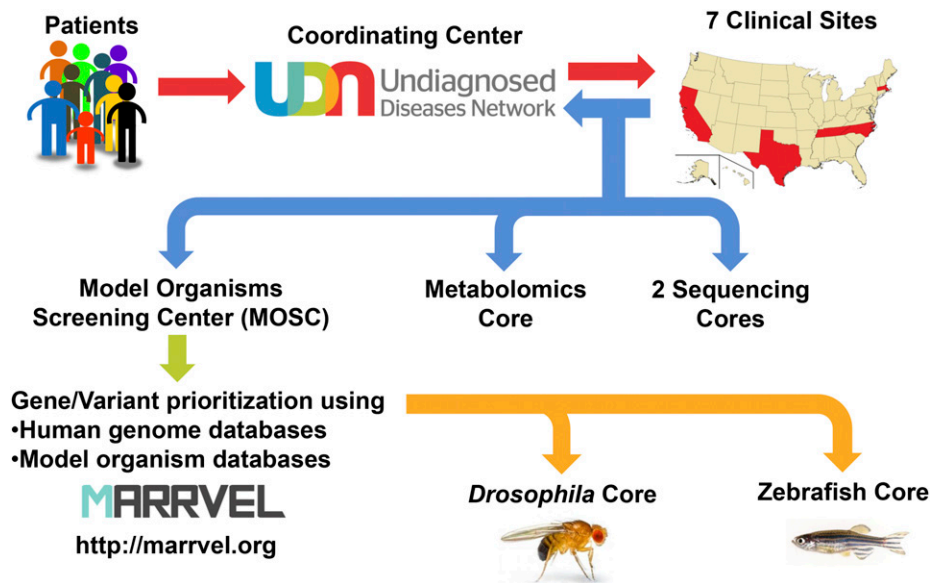


Figure 2 The workflow of the UDN and MOSC. Patients with undiagnosed conditions apply to the UDN primarily through a website (UDN Gateway) that is hosted by the Coordinating Center. Application forms and past medical records are then screened by a case review committee to identify cases with objective findings. Once a patient is accepted, she/he will receive a clinical workup in one of the Clinical Sites. For most cases, WES or WGS are performed on the patient and immediate relatives by one of the two Sequencing Cores. In addition, untargeted metabolomics may be performed on patient samples by the Metabolomics Core. By combining the phenotype and genotype information, some cases can be solved without further investigation. If a diagnosis is not made, the clinical site submits candidate gene/variant information to the MOSC together with a brief description of the patient's condition. The MOSC first performs a database search using the MARRVEL tool

to aggregate existing information on the human gene/variant and its model organism orthologs. In addition, matchmaking with patients in other disease cohorts are attempted through collaborations to identify other individuals with similar genotype and phenotype. Once a variant is considered to be a high priority candidate, experiments to assess gene and variant function are designed by the MOSC investigators and pursued in the *Drosophila* Core or in the Zebrafish Core.

the MOSC) and integration of this information (performed by the Coordination Center) are all key aspects of such a patient-centered effort.

The overall strategy, work flow and some of the key features of the UDN MOSC are highlighted in Figure 2. The MOSC receives variants submitted by the UDN Clinical Sites. Each Clinical Site integrates genotype and extensive phenotype information to prioritize candidate variants. The MOSC investigators communicate closely with the primary UDN physicians who evaluated the patient, and this dialogue continues throughout the process of investigation. The first step of the MOSC pipeline is to utilize databases to prioritize or filter candidate genomic variants. This quality control step is crucial for the MOSC to confirm that the submitted variant has a high chance of being pathogenic prior to designing and performing the actual experiments in model organisms. Public human exome and genome sequences of individuals who lack severe early-onset Mendelian disorders, such as those in the ExAC (Exome Aggregation Consortium) browser (Lek *et al.* 2016), are cross-referenced to the genes and variants from UDN patients. The rationale is that, if a similar genotype is found in people without the disease phenotype, it is unlikely that variant of interest in a candidate gene is responsible for the patient's phenotype. However, one needs to be cautious that this filtration may not be possible for adult onset diseases if the variant in the population genomics databases are from younger individuals that may go on to develop the same disease at a later time point in life. Next, additional human databases with Mendelian disorders, including WES data from the CMG, are studied to identify candidate gene or variant matches. Using this

strategy of database integration, the UDN MOSC recently aided in the diagnosis of a previously undiagnosed developmental syndrome (Schoch *et al.* 2017). From the WES data of a patient in the UDN with a sporadic phenotype of infantile epilepsy, cataracts, and developmental delay, the Duke Clinical Site identified a *de novo* missense variant in the *nucleus accumbens associated 1 (NACCI)* gene as one of several candidates for the child's disorder. Very little gene and protein function information was available to aid in determining whether the *NACCI* variant might be disease-causing. Working with the MOSC, a resulting collaboration between the UDN and the CMG, posting of the gene on GeneMatcher, and identification of additional cases in clinical laboratories including Baylor Genetics and UCLA Clinical Genomics Center, demonstrated a clear role for this *de novo* variant in the patient's phenotype. The combined data of these efforts led from the initial " $n = 1$ " to identification of a total of seven individuals from seven independent families, all with the same *de novo* missense variant in *NACCI* and presenting with the same disease phenotypes (Schoch *et al.* 2017). In this case, the presence of a unique variant and a unique phenotype made collaboration and matchmaking between studies and centers the key to discovery, and model organism studies were not necessary to prove the causality of this *de novo* variant. Nevertheless, this example clearly illustrates how the prescreening step prior to experimentation in model organisms can be productive. In addition, having the ability to interface with physicians and other scientists, through GeneMatcher and other matchmaking efforts, will allow model organism researchers to contribute to the diagnosis prior to the actual model organism work.

After searching for additional cases, the MOSC uses model organism databases including yeast, worm, fly, fish, and rodents to understand the extent to which the gene is conserved, where and when the gene is expressed, what the loss of function phenotypes are, whether the site of the candidate human pathogenic variant is conserved within the orthologous proteins of model organisms, and what tools and reagents have already been generated to study the gene. A recently developed web-based tool named **MARRVEL (Model organism Aggregated Resources for Rare Variant ExpLoration)** automates most of these searches and allows anyone to quickly retrieve the relevant information in a simple comprehensive format (Wang *et al.* 2017).

Next, genes are assigned to the *Drosophila* or Zebrafish Core facilities, where the genes and variants from the UDN patients are studied for their functional impact, to provide a better understanding of the relationships among the gene, variant, and disease phenotypes (Figure 1). Genes that are conserved in *Drosophila* (~65% of human genes) are assigned to the *Drosophila* Core, and the remaining (~35%) genes are assigned to the Zebrafish Core. Occasionally, variants identified in diseases that affect vertebrate specific organ systems such as bone and neural crest derived structures are also assigned to the fish core.

In the *Drosophila* Core, experiments are tailored to each gene and variant using diverse genetic strategies. One strategy depends on the versatile engineered transposable element, MiMIC (*Minos*-mediated integration cassette) (Venken *et al.* 2011), which allows “humanization” of *Drosophila* genes (Bellen and Yamamoto 2015) (Figure 3). This pipeline allows for: (1) generation of strong LOF alleles in the fly gene of interest; (2) expression of the yeast GAL4 transcription factor driven by endogenous enhancers of the gene of interest; (3) functional replacement of the fly gene with the reference (wild-type) human gene to test for rescue of the fly mutant phenotype with the orthologous reference human cDNA expressed under control of a GAL4 expressed in the proper spatial and temporal pattern; and (4) variant function analysis through expression of the human disease allele variant in the same fly line. Other strategies are also used because different reagents (*e.g.*, well characterized LOF alleles, experimentally verified RNAi lines) may be available or rescue with human cDNA may not always be successful (Harel *et al.* 2016b).

In the Zebrafish Core, CRISPR/Cas9 gene editing technology is used to target mutations into the orthologous gene and to LOF alleles in regions of the gene encoding critical functions, such as DNA binding domains, protein interaction domains, or enzymatic catalytic sites. The Zebrafish Core injects CRISPR guide RNAs (gRNAs) and Cas9 mRNA or protein into one-cell zebrafish embryos, where gene editing occurs during early cleavage divisions of the embryo. This results in large clones of mutant cells, several of which almost invariably enter the germline. To increase throughput, the Core combines multiple gRNAs targeting different genes in the same injection (Shah *et al.* 2015). This multiplexing

strategy reduces the number of animals to be raised because each individual carries mutations in multiple genes. After using PCR to genotype adults developed from injected embryos, recovering mutant alleles after outcrosses, and separating mutant lines in subsequent generations, the Zebrafish Core applies a wide range of assays to characterize mutant phenotypes. Phenotypic features found in the zebrafish mutants are compared to the patients’ clinical features, and investigations to study the underlying pathobiology are initiated. Subsequent rescue experiments with wild-type and variant cDNAs or mRNAs can validate the pathogenicity of the variants.

Through these strategies, the UDN MOSC recently aided the identification of novel cases and the functional validation of *de novo* variants in *EBF3* using *Drosophila* (Chao *et al.* 2017a). Starting with a case from the UDP of a child with a neurodevelopmental disorder with a *de novo* variant of unknown significance in *EBF3* (*Early B-Cell Factor 3*), the MOSC identified two additional patients with *de novo* variants altering the same amino acid of the *EBF3* protein through collaborative database searches. In addition, using a MiMIC insertion line in the fly *EBF3* ortholog *knot*—a gene that was well-studied in the context of neural development—the MOSC showed that a reference human *EBF3* cDNA can rescue the lethality of the fly *knot* LOF mutation, showing that human *EBF3* can functionally replace the *Drosophila* ortholog. The missense *de novo* variants found in the patient abolished the ability of human *EBF3* to rescue mutant flies (Chao *et al.* 2017a). Combined with *in vitro* studies, this work provided clear evidence that the *EBF3 de novo* variants are pathogenic in the three cases. Both statistical human genetic evidence and the fly functional analysis support this conclusion (Chao *et al.* 2017a). In parallel, two other groups simultaneously provided additional cases of *EBF3 de novo* variants (Harms *et al.* 2017; Slevén *et al.* 2017). This study is the first to demonstrate the effectiveness of the MOSC pipeline in variant validation for the UDN.

The Canadian RDMM network

In Canada, a series of rare disease gene discovery projects, including FORGE (Finding of Rare Disease Genes) (Beaulieu *et al.* 2014), Care4Rare (Sawyer *et al.* 2016), IGNITE (Identifying Genes and Novel Therapeutics to Enhance Treatment), Omics2TreatID (Tarailo-Graovac *et al.* 2016), and TIDE BC (Sayson *et al.* 2015; Tarailo-Graovac *et al.* 2016), led to the identification of >300 disease genes over the past 5 years. A complementary national project was established to expedite collaboration between basic scientists and clinicians for model organism-based functional studies of rare disease gene variants, and for the development of new therapeutic strategies using model systems. The RDMM connects Canada’s disease gene discovery projects with the Canadian model organism communities of yeast, *C. elegans*, *Drosophila*, zebrafish, and mouse (Figure 4). Collaborations are established between clinicians and basic researchers as

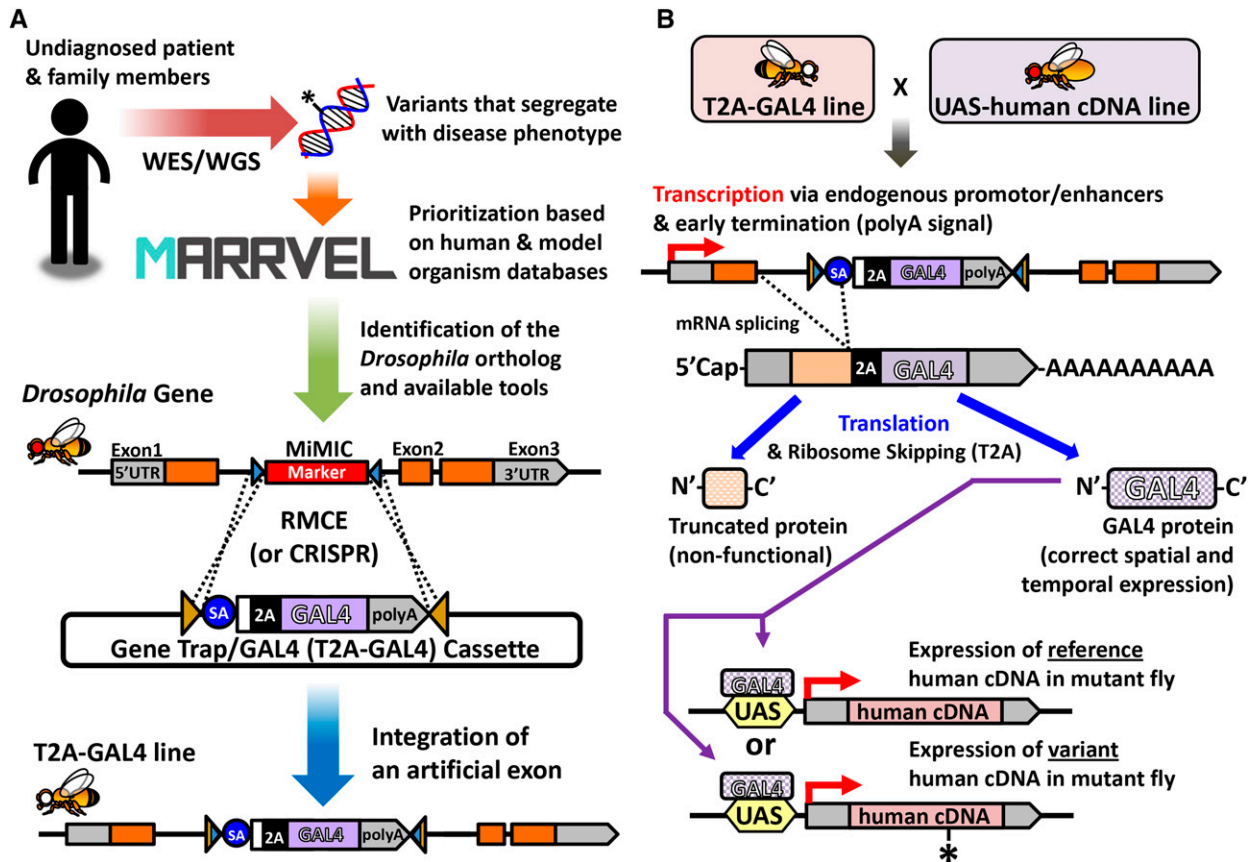


Figure 3 Strategy to “humanize” a *Drosophila* gene to assess functional consequences of a novel variant. (A) For most genes, the *Drosophila* Core of the MOSC performs functional studies of a patient variant by humanizing the orthologous gene in the fly. First, a fly gene that is most likely to be the ortholog of the human gene is identified using the MARRVEL tool. MARRVEL also provides a link to the FlyBase page that displays known biological function, transcriptomics and proteomics data, mutant phenotypes and available resources for the *Drosophila* gene of interest. If a coding intronic MiMIC is available, the *Drosophila* Core uses this as an entry point to study the gene. Through recombinase-mediated cassette exchange (RMCE), an artificial exon is integrated that functions as a gene trap, creating a strong LOF allele. This artificial exon contains a T2A ribosomal skipping sequence and a coding sequence for the GAL4 transcriptional activator (T2A-GAL4). (B) By crossing the T2A-GAL4 strain to a transgenic fly that carries a UAS-human cDNA construct (together with a deficiency of the locus or an independent mutant allele of the fly gene, data not shown), the fly gene can be humanized. When the gene of interest is transcribed, the splice acceptor (SA) in the artificial exon splices into the upstream exon. Since a transcription termination sequence (polyA) is present at the 3' end of this artificial exon, the transcript is terminated, and the remaining portion of the fly gene is not transcribed. When this transcript is translated, a truncated protein that is usually nonfunctional is made together with a GAL4 protein. GAL4 is expressed in the same spatial and temporal pattern as the fly gene, allowing expression of the corresponding human cDNA under the control of the UAS element (GAL4 target sequence). By comparing the ability of the reference (wild type) and variant (mutant) to rescue the fly mutant phenotype, one can assess whether the variant of interest impacts protein function.

soon as possible after the initial identification of candidate genes, and a rapid response seed grant is awarded to support immediate functional experiments—a unique feature of the RDMM. The anticipated outcomes are: (1) elucidation of gene function and functional validation of human genetic variants that cause disease; (2) high impact publications of disease gene discoveries through inclusion of functional data; (3) potential rationale for treatment (e.g., identification of candidate drug targets) via knowledge of a disease gene pathway; and (4) establishment of longer term collaborations between basic scientists and clinicians that will lead to basic or applied research.

Using the RDMM infrastructure, a clinician or an investigator in a genetics laboratory who identifies a new candidate gene can submit a “connect application” (Figure 4). Connect

applications can also be submitted by a clinical researcher who wants to propose a model organism project for a known disease gene but for which pathophysiological understanding or treatment options are lacking. The Clinical Advisory Committee evaluates these applications within 2 weeks, and, if approved, the network registry is searched for potential orthologous gene matches that identify a model organism researcher who is an expert on the gene of interest. To generate this registry, all investigators working with model organisms in Canada were encouraged to establish an RDMM account and submit genes they designate as tier 1, 2, or 3 depending on the researcher’s level of expertise and availability of models in their laboratory. Selected investigators are invited to write a short “catalyst grant proposal application” on how the researcher would assess the functional

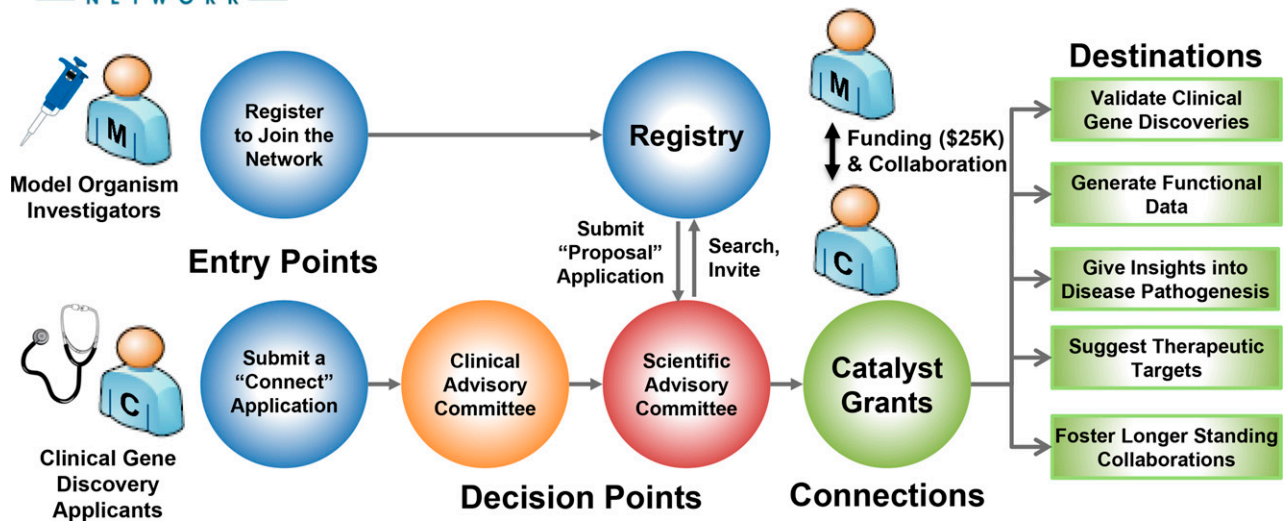


Figure 4 The workflow of the Canadian RDMM Network. RDMM connects Canada's disease gene discovery projects with the Canadian model organism researchers. Investigators that work with yeast, *C. elegans*, *Drosophila*, zebrafish, or mouse are encouraged to join the network. Upon registration, the investigator provides a list of genes or genetic pathways in which they are experts. In parallel, a physician or a human geneticist submits a "connect application" for cases that they wish to find a model organism collaborator for. If the case is approved by the Clinical Advisory Committee, the Scientific Advisory Committee performs a search of the model organism registry and identifies an investigator that specializes in the orthologous gene. Upon matchmaking, the model organism investigator and the physician/human geneticist discuss a working plan and submit a proposal to the Scientific Advisory Committee. If the case is approved, funding is provided (1 year, Can\$25,000) to generate a disease model and study the candidate gene/variant. The long-term goal of this process is to connect clinicians and basic researchers to establish a collaborative network across the country to facilitate rare disease research.

consequences of mutations. The intention is to rapidly identify researchers who have a history and track record of studying the function of orthologs of rare disease-linked genes. Ideally, the selected individual is poised to immediately conduct functional analyses that will provide new biological insight into a human disease gene. The model organism investigator is given a year to generate a model and provide a report with the collaborating clinician. The goal of this process is not only to validate candidate genes, investigate known disease genes, and generate models to test possible drugs, but also to facilitate links between clinicians and basic researchers to create collaborative networks that will lead to larger projects.

To date, over 400 model organism researchers have registered in the database, entered over 5800 genes of interest, and 46 catalyst grants were funded since January 2015. RDMM-funded projects already led to identification of *NANS deficiency* (MIM# 610442) (van Karnebeek *et al.* 2016)—a condition with skeletal dysplasia and developmental delay caused by insufficient sialic acid synthesis. Knockdown of the orthologous zebrafish *nansa* gene, encoding sialic acid synthase, in embryos resulted in abnormal skeletal development similar to the patients' phenotype. This phenotype in zebrafish mutants was partially rescued by the addition of

sialic acid to the culture water, opening the door to potential treatment with oral sialic acid in human patients. Another example is the identification and characterization of a dominant mutation in *NALCN*, which encodes a cation channel, causing intellectual disability, ataxia, and arthrogryposis (Aoyagi *et al.* 2015)—a gene previously associated only with a recessive condition with developmental delay and hypotonia (MIM# 615419). Introduction of the analogous variant into the *C. elegans* homolog *nca-1* induced a dominant GOF neurological phenotype, providing evidence that the disease-linked variant likely has a similar impact on the human *NALCN* protein.

Facilitating international collaboration through rare and undiagnosed disease research consortia

A number of institutes and organizations around the world are actively engaged in undiagnosed rare diseases research: Japan [Initiative on Rare and Undiagnosed Diseases (IRUD)]; Italy (Telethon Undiagnosed Diseases Program); and Australia and New Zealand [Undiagnosed Diseases Program—Australia and New Zealand (Baynam *et al.* 2017)]. Recent establishment of large international consortia, such as the Undiagnosed Diseases Network International (UDNI) and the International Rare Diseases Research Consortium (IRDIRC), is beginning to

facilitate interactions and information exchange among undiagnosed disease researchers worldwide. Patient advocate organizations such as [National Organization for Rare Disorders \(NORD\)](#) and [European Organization for Rare Diseases \(EURORDIS\)](#) collaborate with these consortia to facilitate diagnoses and research of rare diseases. In the absence of functional data, much of the current efforts are focused around identifying a second patient with overlapping phenotypes and genotypes to try to solve the more difficult “ $n = 1$ ” cases. Considering that functional information of genes and variants will be necessary for most new disease genes and variants of unknown significance that are discovered through these efforts, we foresee that an international network of model organism researchers will boost the pace of disease gene discovery and downstream mechanistic studies necessary for understanding pathogenesis.

The Future of Integrating Human Genomics and Model Organisms

The drive to assess the function of human genes based on model organism research will continue to increase. As our knowledge deepens, understanding the role of human genes will greatly benefit from this knowledge base. Several general themes are emerging that will pave the way forward.

Pleiotropy and undiagnosed disease

Pleiotropy is the phenomenon of mutations in the same gene causing multiple phenotypes. This can be due to a gene being used in different tissues or at different stages of development, or encoding a protein that has more than one molecular function. Insights from human genome sequencing efforts provided several examples of pleiotropy. For instance, sequencing often uncovers a known disease gene in a patient with undiagnosed disease, and, in retrospect, the variant is diagnostic. However, the gene might not have been considered initially because the patient did not precisely match what was previously understood about the clinical phenotype of the disease; in retrospect, the patient’s symptoms extended beyond what were previously recognized as the core clinical features. This realization of “phenotypic expansion” is a product of limited clinical descriptions and the inherent pleiotropy of many human genes (Chong *et al.* 2015). Phenotypic expansion often continues long after the gene is first associated with a disease. Pleiotropy in human disease can also manifest as a single gene associated with multiple distinct diseases. For example, pleiotropic effects are evident for the *TRPV4* locus with different mutations causing distinct diseases, as described earlier. Therefore, the key role of model organisms to facilitate understanding of the many functions of a pleiotropic gene cannot be overstated.

Noncoding variants and model organisms

For Mendelian disease, WES is a successful clinically available diagnostic tool. However, WES only targets coding regions, and noncoding genetic variants can also play an important

role in human disease (Zhang and Lupski 2015). For example, *microRNA-140* (*miR-140*) regulates palate formation in zebrafish (Eberhart *et al.* 2008), and a single nucleotide polymorphism (SNP, rs7205289) within the human *miR-140* gene, which decreases miR-140 processing, was found to be strongly associated with nonsyndromic cleft palate in human patients (Li *et al.* 2010). While some annotated noncoding but transcribed elements, such as miRNAs and long noncoding RNAs (lncRNAs) could be added to the target for sequencing, the identification of causative sequence variants in regulatory DNAs is difficult due to our incomplete understanding of the logic of human gene regulatory networks. Furthermore as additional WGS data become available, more noncoding variants will be identified. Understanding the consequences of these variants will continue to be a very significant challenge. Although there are examples of pathogenic noncoding copy number variants (CNVs) upstream of disease loci (McCarroll *et al.* 2008; Zhang *et al.* 2010), the current view is that the vast majority of disease-causing mutations from rare Mendelian disorders relate to coding or splicing variants. Interestingly, most GWAS (Genome Wide Association Studies) loci associated with common disorders are genetic variants in noncoding or regulatory regions (Edwards *et al.* 2013). Given the overall lack of conservation between noncoding regions in humans and model organisms, it is challenging to study noncoding variants, although some strategies are emerging in vertebrate model organisms (Zhang and Lupski 2015). In cases where the noncoding region is not conserved, model organism strategies to understand the consequences of changes in gene expression may offer the best insights. For example, if the variant can be shown to affect gene regulation in human cells or tissues (*e.g.*, gene expression data from RNA-seq in patient samples), then gene knock-down or overexpression in model organisms may be productive (Wangler *et al.* 2017). However, as the field elucidates more noncoding variants associated with diseases, then more direct ways to study these regulatory variants may be needed.

Human database integration and matchmaking

Genetic diagnoses have greatly benefitted from information in numerous human databases. Efforts to use human genomics for diagnosis depend on comparing candidate variants to other variants at the locus found in other databases. The success of efforts, such as [Deciphering-Developmental-Disorders-Study](#) (2015), [ExAC](#) and its expansion project [gnomAD](#) ([Genome Aggregation Database](#)) (Lek *et al.* 2016), [Geno2MP](#) (Chong *et al.* 2016a), and the [DiscovEHR Collaboration](#) (Abul-Husn *et al.* 2016; Dewey *et al.* 2016), in creating publically available datasets is a result of these diagnostic needs. Use of integrated match-making programs that comprise the [Matchmaker Exchange \(MME\)](#) (Philippakis *et al.* 2015) allows the identification of patients with similar phenotypes and genotypes across different platforms. In Matchmaker Exchange, databases such as [GeneMatcher](#) (Sobreira *et al.* 2015b), [PhenomeCentral](#) (Buske *et al.* 2015), [DECIPHER](#)

(Databases of genomic variation and Phenotype in Humans using Ensembl Resources) (Chatzimichali *et al.* 2015), MyGene2 (Chong *et al.* 2016b), *matchbox/seqr*, and Australian Genomics Health Alliance (AGHA) Patient Archive are connected to one another, thereby supporting queries of data from >20,000 unrelated patients, and facilitating prioritization of genes and variants.

Model organism resources and infrastructure

Maintaining and expanding the resources and infrastructure to support model organism research will be essential for the future of both undiagnosed and diagnosed disease research, for our ability to identify molecular mechanisms quickly and efficiently, and to develop potential therapies for human disease (Table 3). We strongly feel this should be embraced and well supported by the NIH and government agencies in other countries that fund biomedical research. Indeed, rapid analyses of variants of unknown significance and deeper mechanistic studies often depend upon the availability of reagents in model organisms. If these resources are not available and need to be developed, researchers will be less likely to embark on such studies. Similarly, the organism-specific databases, including *Saccharomyces* Genome Database (SGD, budding yeast), PomBase (fission yeast), WormBase (*C. elegans*), FlyBase (*Drosophila*), The Zebrafish Information Network (ZFIN, zebrafish), Mouse Genome Informatics (MGI, mouse), and Rat Genome Database (RGD, rat), are critical to mine, organize, and curate biological information required to prioritize candidate disease genes. Although databases and mining tools integrating information from multiple key model organism databases and corresponding human genes are being developed by the Alliance of Genome Resources, the Monarch Initiative, and other efforts such as MARRVEL and Gene2Function, specific model organism databases will continue to play a critical role because they curate and annotate organism-specific datasets that are seminal to their respective research communities. For example, FlyBase (Bilder and Irvine 2017) and ZFIN (Howe *et al.* 2017) are queried tens of millions of times a year, mostly by members of the fly and zebrafish communities, respectively, which are estimated to consist of >6000 researchers each. Hence, both an integrated databases and model organism-specific databases are required to facilitate the synergy between the medical genetics and model organism research communities.

In addition to these digital resources, the UDN, RDMM and similar projects in other countries will rely on publically distributed reagents and strains to facilitate gene and variant annotation. For example, a public collection of human cDNA libraries being assembled by the Mammalian Gene Collection (MGC) consortium (Temple *et al.* 2009). Transgenic yeast (Kachroo *et al.* 2015), *Drosophila*, and zebrafish (Davis *et al.* 2014) strains that allow expression of these human cDNAs *in vivo* in model organisms will provide easily accessible and valuable “off-the-shelf” resources to support and encourage the use of model organisms for functional analyses

of human variants. In addition, a collection of gene knock-in (Nagarkar-Jaiswal *et al.* 2015) or BAC/phosmid transgenic lines (Sarov *et al.* 2016) that tag most proteins in the genome with epitopes or fluorescent proteins would allow investigators to assess expression, subcellular distribution, and function of genes quickly by integrating cell biological, biochemical, and proteomic approaches. For example, GFP-tagged strains can also be used to conditionally knockdown the transcript or protein of interest in a reversible and temporally and spatially controlled manner (Nagarkar-Jaiswal *et al.* 2015). Stock centers, including the Bloomington *Drosophila* Stock Center (BDSC), *Caenorhabditis* Genetics Center (CGC), Zebrafish International Resource Center (ZIRC), Mutant Mouse Resource and Research Centers (MMRRC), and Mouse Mutant Resource (MMR) that assist in the maintenance of these and other useful mutant and transgenic resources depend upon continued support by the NIH, international research community, and end-users. By combining these resources, and the genotype to phenotype information obtained from human patients, one can more rapidly study the LOF phenotype, test whether the molecular function of the human and model organism genes are conserved, determine whether a variant of unknown significance is functional, study the expression pattern and subcellular localization of the protein of interest, identify physical interaction partners, and understand the function of the gene in the organ system and cell type of interest (Bellen and Yamamoto 2015). Maintaining and expanding the resources of model organism communities will provide invaluable information to expand our ability to diagnose and dissect the molecular mechanisms underlying human diseases.

Finally, nonmodel organisms may also provide key insights into disease pathogenesis (Russell *et al.* 2017). Indeed, some animals adapted to a specific environment show a phenotype that would be lethal or detrimental in humans. These include osteopenia and profound anemia in Antarctic icefish or eye loss and metabolic syndromes in cavefish (Albertson *et al.* 2009). These organisms can be considered as “evolutionary mutant models,” which may help us better understand similar disease conditions in human. These organisms may also provide critical clues about key molecular mechanisms that could be targeted for therapeutic interventions.

Collaboration and teamwork

Teamwork is paramount to drive model organism research in undiagnosed disease. Collaboration between clinicians who identify potential genes of interest and biologists who study model organisms to assess the function of homologous genes forms a first layer. In the CMG model, the clinicians and researchers seek collaboration based on the large number of candidate variants identified. Collaborative efforts are established with model organism researchers on a case-by-case basis. Although this strategy has been very productive, it may be difficult to expand outside large genomic centers. This justifies the need for approaches like the RDMM and UDN. In

Table 3 Online resources cited

Comments from the NIH Leadership on Model Organism Research Support Francis Collins at TAGC 2016 on YouTube NIH Extramural Nexus I Open Mike I A Look at Trends in NIH's Model Organism Research Support.	https://www.youtube.com/watch?v=f9FXNU1YWQ0 https://mexus.od.nih.gov/all/2016/07/14/a-look-at-trends-in-nih-model-organism-research-support/
Rare and undiagnosed disease research consortia and organizations	
CMG-The Centers for Mendelian Genomics	http://mendelian.org/
DDDS-Deciphering Developmental Disorders Study	https://www.ddduk.org/
DiscovEHR collaboration	http://www.discovershare.com/
EURORDIS-The European Organization for Rare Diseases	http://www.eurordis.org/
IGNITE project	http://igniteproject.ca/
IRD/IRC-The International Rare Diseases Research Consortium	http://www.amed.go.jp/en/program/IRUD/
IRUD-The Initiative on Rare and Undiagnosed Diseases	https://rarediseases.org/
NORD-National Organization for Rare Disorders	http://www.rare-diseases-catalyst-network.ca/
RDMM-The Rare Diseases: Models and Mechanisms Network	http://www.telethon.it/node/49281
Telethon Undiagnosed Diseases Program	https://undiagnosed.hms.harvard.edu/
UDN-The Undiagnosed Diseases Network	https://undiagnosed.hms.harvard.edu/research/model-organisms/
UDN I MOSC-Model Organisms Screening Center	http://www.udhinternational.org/
UDNI-Undiagnosed Diseases Network International	
Human genomic databases, tools and resources	
Australian Genomics Health Alliance Patient Archive	https://mme.australiangenomics.org.au/
ClinicalTrials.gov/Leflunomide+Vemurafenib in V600 Mutant Met. Melanoma	https://clinicaltrials.gov/ct2/show/study/NCT01611675
DECIPHER-Database of genomic variation and Phenotype in Humans using Ensembl Resources	https://decipher.sanger.ac.uk/
ExAC (Exome Aggregation Consortium) browser	http://exac.broadinstitute.org/
GeneMatcher	https://genematcher.org/
Geno2MP	http://geno2mp.gs.washington.edu/Geno2MP/
gnomAD-Genome Aggregation Database	http://gnomad.broadinstitute.org/
<i>matchbox/seqr</i> at Broad Institute	https://seqr.broadinstitute.org/
Matchmaker Exchange	http://www.matchmakerexchange.org/
MyGene2	https://mygene2.org/MyGene2/
OMIM-Online Mendelian Inheritance in Man	https://www.omim.org/
PhenomeCentral	https://phenomecentral.org/
Cross-species mining tools and resources	
Alliance of Genome Resources	http://www.alliancegenome.org/
DIOP-DRSC Integrative Ortholog Prediction Tool	http://www.flyrnai.org/diopt
Gene2Function	http://www.gene2function.org/
MARVEL-Model organism Aggregated Resources for Rare Variant Exploration	http://marvel.org/
Monarch Initiative	https://monarchinitiative.org/
Model organism databases, tools and resources	
BDSC-Bloomington Drosophila Stock Center	http://fly.bio.indiana.edu/
BDSC I Human proteins under the control of UAS	http://fly.bio.indiana.edu/Browse/uas/uas_hsap.php
CGC-Caenorhabditis Genetics Center	https://cgc.umn.edu/
FlyBase	http://flybase.org/
MGC-Mammalian Gene Collection	https://genecollections.nci.nih.gov/MGC/
MGH-Mouse Genome Informatics	http://www.informatics.jax.org/
MMR-Mouse Mutant Resource	https://www.jax.org/research-and-faculty/tools/mouse-mutant-resource
MMRRC-Mutant Mouse Resource and Research Centers	https://www.mmrc.org/
PomBase	https://www.pombase.org/

(continued)

Table 3, continued

RGD-Rat Genome Database	http://rgd.mcw.edu/
SGD-Saccharomyces Genome Database	http://www.yeastgenome.org/
WormBase	http://www.wormbase.org/
ZFIN-The Zebrafish Information Network	http://zfin.org/
ZIRC-Zebrafish International Resource Center	https://zebrafish.org/

the UDN model, novel human variants are provided to a central entity, the MOSC, which is responsible for rapid assessment of the variants. The advantages are that the MOSC follows a standard approach in defined model organisms so that each variant is subject to the same analytical process. This tactic allows coordinated and systematic processing of numerous variants simultaneously. However, this process may not take full advantage of the rich expertise of investigators in particular areas. In the RDMM model variants are distributed, and the advantage of this organization is that novel variants are studied by experts with strong interest in a particular gene, tissue, or pathway. However, the methods and mechanistic depth for each variant may differ significantly because a centralized organization allows for a more efficient workflow and better resource allocation.

It is possible that a model combining the centralized (UDN-MOSC) approach and the distributive (RDMM) approach may ultimately be the best way to study candidate variants in model organisms. A potential workflow would be first to engage a primary center where genome biologists check variants against other human genomic datasets, chose the ideal model organism, select the best technologies, efficiently and economically generate the reagents, and rapidly test candidate variants. The acquired information and reagents could then be distributed to expert colleagues who wish to pursue the biology of the gene and disease in depth. Structuring multi-organism teams in the face of rapidly expanding lists of candidate variants and genes is a challenge that requires clinicians, human geneticists, bioinformaticians, and model organism investigators to work together. This collaboration allows rapid translation of knowledge obtained in one species to another and avoids duplication of effort. Integrating and acknowledging the strengths of each model organism is required to accelerate discoveries that identify pathogenic mechanisms and lead to development of novel therapeutic strategies.

Acknowledgments

We thank James R. Lupski, Paolo Moretti, Loren Del Mar Pena, Joy D. Cogan, and Jasper Rine for valuable input and critical reading of the manuscript. This work was supported in part by National Institutes of Health (NIH) grants U54NS093793 (H.J.B., M.F.W., S.Y., M.W., and J.P.), R24OD022005 (H.J.B.), R01GM067858 (H.J.B.), HD22486 (M.W.) GM114438 (M.W.), Alzheimer's Association (NIRH-15-364099 to S.Y.), and Simons Foundation Autism Research Initiative (368479 to M.F.W. and S.Y.). H.J.B. is an investigator of the Howard Hughes Medical Institute (HHMI) and received support from the Robert A. and Renee E. Belfer Family Foundation, the Huffington Foundation, and Target ALS. H.-T.C. is supported by an American Academy of Neurology (AAN) Neuroscience Research Scholarship and the CNCDP-K12 Fellowship from the National Institute of Neurological Disorders and Stroke (1K12NS098482). J.E.P. was supported by the Ting Tsung and Wei Fong Chao Foundation and National Human

Genome Research Institute (NHGRI) grant K08 HG008986. The Rare Diseases Models and Mechanisms (RDMM) Network is funded by the Canadian Institutes of Health Research grant RCN-137793. P.M.C. is funded in part by clinician scientist awards from the Canadian Institutes of Health Research (CIHR) (RN-315908) and the FRQS (30647).

Literature Cited

- Abul-Husn, N. S., K. Manickam, L. K. Jones, E. A. Wright, D. N. Hartzel *et al.*, 2016 Genetic identification of familial hypercholesterolemia within a single U.S. health care system. *Science* 354: aaf7000.
- Afzal, A. R., A. Rajab, C. D. Fenske, M. Oldridge, N. Elanko *et al.*, 2000 Recessive Robinow syndrome, allelic to dominant brachydactyly type B, is caused by mutation of ROR2. *Nat. Genet.* 25: 419–422.
- Albert, T. J., M. N. Molla, D. M. Muzny, L. Nazareth, D. Wheeler *et al.*, 2007 Direct selection of human genomic loci by microarray hybridization. *Nat. Methods* 4: 903–905.
- Albertson, R. C., W. Cresko, H. W. Detrich, III, and J. H. Postlethwait, 2009 Evolutionary mutant models for human disease. *Trends Genet.* 25: 74–81.
- Aoyagi, K., E. Rossignol, F. F. Hamdan, B. Mulcahy, L. Xie *et al.*, 2015 A gain-of-function mutation in NALCN in a child with intellectual disability, ataxia, and arthrogryposis. *Hum. Mutat.* 36: 753–757.
- Ashraf, S., H. Y. Gee, S. Woerner, L. X. Xie, V. Vega-Warner *et al.*, 2013 ADCK4 mutations promote steroid-resistant nephrotic syndrome through CoQ10 biosynthesis disruption. *J. Clin. Invest.* 123: 5179–5189.
- Auer-Grumbach, M., A. Olschewski, L. Papic, H. Kremer, M. E. McEntagart *et al.*, 2010 Alterations in the ankyrin domain of TRPV4 cause congenital distal SMA, scapuloperoneal SMA and HMSN2C. *Nat. Genet.* 42: 160–164.
- Bainbridge, M. N., M. Wang, D. L. Burgess, C. Kovar, M. J. Rodesch *et al.*, 2010 Whole exome capture in solution with 3 Gbp of data. *Genome Biol.* 11: R62.
- Balter, M., and G. Vogel, 2001 Nobel prize in physiology or medicine. Cycling toward Stockholm. *Science* 294: 502–503.
- Bamshad, M. J., J. A. Shendure, D. Valle, A. Hamosh, J. R. Lupski *et al.*, 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.
- Bard, F., L. Casano, A. Mallabiabarrena, E. Wallace, K. Saito *et al.*, 2006 Functional genomics reveals genes involved in protein secretion and Golgi organization. *Nature* 439: 604–607.
- Bashamboo, A., P. A. Donohoue, E. Vilain, S. Rojo, P. Calvel *et al.*, 2016 A recurrent p.Arg92Trp variant in steroidogenic factor-1 (NR5A1) can act as a molecular switch in human sex development. *Hum. Mol. Genet.* 25: 3446–3453.
- Baynam, G., S. Broley, A. Bauskis, N. Pachter, F. McKenzie *et al.*, 2017 Initiating an undiagnosed diseases program in the Western Australian public health system. *Orphanet J. Rare Dis.* 12: 83.
- Beaulieu, C. L., J. Majewski, J. Schwartzentruber, M. E. Samuels, B. A. Fernandez *et al.*, 2014 FORGE Canada Consortium: outcomes of a 2-year national rare-disease gene-discovery project. *Am. J. Hum. Genet.* 94: 809–817.
- Bellen, H. J., and S. Yamamoto, 2015 Morgan's legacy: fruit flies and the functional annotation of conserved genes. *Cell* 163: 12–14.
- Bellen, H. J., C. Tong, and H. Tsuda, 2010 100 years of *Drosophila* research and its impact on vertebrate neuroscience: a history lesson for the future. *Nat. Rev. Neurosci.* 11: 514–522.
- Bilder, D., and K. D. Irvine, 2017 Taking stock of the *Drosophila* research ecosystem. *Genetics* 206: 1227–1236.
- Blair, D. R., C. S. Lyttle, J. M. Mortensen, C. F. Bearden, A. B. Jensen *et al.*, 2013 A nondegenerate code of deleterious variants in Mendelian loci contributes to complex disease risk. *Cell* 155: 70–80.
- Boatman, S., F. Barrett, S. Satishchandran, L. Jing, I. Shestopalov *et al.*, 2013 Assaying hematopoiesis using zebrafish. *Blood Cells Mol. Dis.* 51: 271–276.
- Bonini, N. M., and S. L. Berger, 2017 The sustained impact of model organisms in genetics and epigenetics. *Genetics* 205: 1–4.
- Braasch, I., A. R. Gehrke, J. J. Smith, K. Kawasaki, T. Manousaki *et al.*, 2016 The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat. Genet.* 48: 427–437.
- Bunn, K. J., P. Daniel, H. S. Rosken, A. C. O'Neill, S. R. Cameron-Christie *et al.*, 2015 Mutations in DVL1 cause an osteosclerotic form of Robinow syndrome. *Am. J. Hum. Genet.* 96: 623–630.
- Buske, O. J., M. Girdea, S. Dumitriu, B. Gallinger, T. Hartley *et al.*, 2015 PhenomeCentral: a portal for phenotypic and genotypic matchmaking of patients with rare genetic diseases. *Hum. Mutat.* 36: 931–940.
- Ceol, C. J., Y. Houvras, J. Jane-Valbuena, S. Bilodeau, D. A. Orlando *et al.*, 2011 The histone methyltransferase SETDB1 is recurrently amplified in melanoma and accelerates its onset. *Nature* 471: 513–517.
- Chao, H. T., M. Davids, E. Burke, J. G. Pappas, J. A. Rosenfeld *et al.*, 2017a A syndromic neurodevelopmental disorder caused by de novo variants in EBF3. *Am. J. Hum. Genet.* 100: 128–137.
- Chao, H. T., L. Liu, and H. J. Bellen, 2017b Building dialogues between clinical and biomedical research through cross-species collaborations. *Semin. Cell Dev. Biol.* pii: S1084-9521(17)30260-4.
- Chatzimichali, E. A., S. Brent, B. Hutton, D. Perrett, C. F. Wright *et al.*, 2015 Facilitating collaboration in rare genetic disorders through effective matchmaking in DECIPHER. *Hum. Mutat.* 36: 941–949.
- Chong, J. X., K. J. Buckingham, S. N. Jhangiani, C. Boehm, N. Sobreira *et al.*, 2015 The genetic basis of Mendelian phenotypes: discoveries, challenges, and opportunities. *Am. J. Hum. Genet.* 97: 199–215.
- Chong, J. X., V. Caputo, I. G. Phelps, L. Stella, L. Worgan *et al.*, 2016a Recessive inactivating mutations in TBCK, encoding a Rab GTPase-activating protein, cause severe infantile syndromic encephalopathy. *Am. J. Hum. Genet.* 98: 772–781.
- Chong, J. X., J. H. Yu, P. Lorentzen, K. M. Park, S. M. Jamal *et al.*, 2016b 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.
- Ciani, L., A. Patel, N. D. Allen, and C. French-Constant, 2003 Mice lacking the giant protocadherin mFAT1 exhibit renal slit junction abnormalities and a partially penetrant cyclopia and anophthalmia phenotype. *Mol. Cell. Biol.* 23: 3575–3582.
- Cosens, D. J., and A. Manning, 1969 Abnormal electroretinogram from a *Drosophila* mutant. *Nature* 224: 285–287.
- Coventry, A., L. M. Bull-Otterson, X. Liu, A. G. Clark, T. J. Maxwell *et al.*, 2010 Deep resequencing reveals excess rare recent variants consistent with explosive population growth. *Nat. Commun.* 1: 131.
- Dai, J., T. J. Cho, S. Unger, E. Lausch, G. Nishimura *et al.*, 2010 TRPV4-pathway, a novel channelopathy affecting diverse systems. *J. Hum. Genet.* 55: 400–402.
- Davis, E. E., S. Frangakis, and N. Katsanis, 2014 Interpreting human genetic variation with in vivo zebrafish assays. *Biochim. Biophys. Acta* 1842: 1960–1970.
- Deciphering-Developmental-Disorders-Study, 2015 Large-scale discovery of novel genetic causes of developmental disorders. *Nature* 519: 223–228.

- Dehal, P., and J. L. Boore, 2005 Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* 3: e314.
- Deng, H. X., C. J. Klein, J. Yan, Y. Shi, Y. Wu *et al.*, 2010 Scapuloperoneal spinal muscular atrophy and CMT2C are allelic disorders caused by alterations in TRPV4. *Nat. Genet.* 42: 165–169.
- Dewey, F. E., M. F. Murray, J. D. Overton, L. Habegger, J. B. Leader *et al.*, 2016 Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. *Science* 354: aaf6814.
- Eberhart, J. K., X. He, M. E. Swartz, Y. L. Yan, H. Song *et al.*, 2008 MicroRNA Mirn140 modulates Pdgf signaling during palatogenesis. *Nat. Genet.* 40: 290–298.
- Edwards, S. L., J. Beesley, J. D. French, and A. M. Dunning, 2013 Beyond GWASs: illuminating the dark road from association to function. *Am. J. Hum. Genet.* 93: 779–797.
- Eldomery, M. K., Z. Coban-Akdemir, T. Harel, J. A. Rosenfeld, T. Gambin *et al.*, 2017 Lessons learned from additional research analyses of unsolved clinical exome cases. *Genome Med.* 9: 26.
- Ellis, H. M., and H. R. Horvitz, 1986 Genetic control of programmed cell death in the nematode *C. elegans*. *Cell* 44: 817–829.
- Farwell, K. D., L. Shahmirzadi, D. El-Khechen, Z. Powis, E. C. Chao *et al.*, 2015 Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. *Genet. Med.* 17: 578–586.
- Fire, A., S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver *et al.*, 1998 Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391: 806–811.
- Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan *et al.*, 1999 Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151: 1531–1545.
- Gahl, W. A., T. C. Markello, C. Toro, K. F. Fajardo, M. Sincan *et al.*, 2012 The National Institutes of Health Undiagnosed Diseases Program: insights into rare diseases. *Genet. Med.* 14: 51–59.
- Gahl, W. A., J. J. Mulvihill, C. Toro, T. C. Markello, A. L. Wise *et al.*, 2016 The NIH undiagnosed diseases program and network: applications to modern medicine. *Mol. Genet. Metab.* 117: 393–400.
- Gambin, T., Z. C. Akdemir, B. Yuan, S. Gu, T. Chiang *et al.*, 2017 Homozygous and hemizygous CNV detection from exome sequencing data in a Mendelian disease cohort. *Nucleic Acids Res.* 45: 1633–1648.
- Gao, Z., D. Waggoner, M. Stephens, C. Ober, and M. Przeworski, 2015 An estimate of the average number of recessive lethal mutations carried by humans. *Genetics* 199: 1243–1254.
- Garg, A., M. Kircher, M. Del Campo, R. S. Amato, A. K. Agarwal *et al.*, 2015 Whole exome sequencing identifies de novo heterozygous CAV1 mutations associated with a novel neonatal onset lipodystrophy syndrome. *Am. J. Med. Genet. A.* 167A: 1796–1806.
- Garrod, A. E., 1902 The incidence of alkaptonuria: a study in chemical individuality. *Lancet* ii: 1616–1620.
- Garrod, A. E., 1923 *Inborn Errors of Metabolism*. Henry Frowde and Hodder & Stoughton, London.
- Gee, H. Y., C. E. Sadowski, P. K. Aggarwal, J. D. Porath, T. A. Yakulov *et al.*, 2016 FAT1 mutations cause a glomerulotubular nephropathy. *Nat. Commun.* 7: 10822.
- Gonzaga-Jauregui, C., J. R. Lupski, and R. A. Gibbs, 2012 Human genome sequencing in health and disease. *Annu. Rev. Med.* 63: 35–61.
- Gray, K. A., B. Yates, R. L. Seal, M. W. Wright, and E. A. Bruford, 2015 Genenames.org: the HGNC resources in 2015. *Nucleic Acids Res.* 43: D1079–D1085.
- Halder, G., and R. L. Johnson, 2011 Hippo signaling: growth control and beyond. *Development* 138: 9–22.
- Hamosh, A., N. Sobreira, J. Hoover-Fong, V. R. Sutton, C. Boehm *et al.*, 2013 PhenoDB: a new web-based tool for the collection, storage, and analysis of phenotypic features. *Hum. Mutat.* 34: 566–571.
- Hamza, A., E. Tammpere, M. Kofoed, C. Keong, J. Chiang *et al.*, 2015 Complementation of yeast genes with human genes as an experimental platform for functional testing of human genetic variants. *Genetics* 201: 1263–1274.
- Harel, T., G. Yesil, Y. Bayram, Z. Coban-Akdemir, W. L. Charnig *et al.*, 2016a Monoallelic and biallelic variants in EMCI identified in individuals with global developmental delay, hypotonia, scoliosis, and cerebellar atrophy. *Am. J. Hum. Genet.* 98: 562–570.
- Harel, T., W. H. Yoon, C. Garone, S. Gu, Z. Coban-Akdemir *et al.*, 2016b Recurrent de novo and biallelic variation of ATAD3A, encoding a mitochondrial membrane protein, results in distinct neurological syndromes. *Am. J. Hum. Genet.* 99: 831–845.
- Harms, F. L., K. M. Girisha, A. A. Hardigan, F. Kortum, A. Shukla *et al.*, 2017 Mutations in EBF3 disturb transcriptional profiles and cause intellectual disability, ataxia, and facial dysmorphism. *Am. J. Hum. Genet.* 100: 117–127.
- Hartwell, L. H., J. Culotti, and B. Reid, 1970 Genetic control of the cell-division cycle in yeast. I. Detection of mutants. *Proc. Natl. Acad. Sci. USA* 66: 352–359.
- Hartwell, L. H., J. Culotti, J. R. Pringle, and B. J. Reid, 1974 Genetic control of the cell division cycle in yeast. *Science* 183: 46–51.
- Hieter, P., and K. M. Boycott, 2014 Understanding rare disease pathogenesis: a grand challenge for model organisms. *Genetics* 198: 443–445.
- Holland, P. W., J. Garcia-Fernandez, N. A. Williams, and A. Sidow, 1994 Gene duplications and the origins of vertebrate development. *Dev. Suppl.* 120: 125–133.
- Howe, D. G., Y. M. Bradford, A. Eagle, D. Fashena, K. Frazer *et al.*, 2017 The zebrafish model organism database: new support for human disease models, mutation details, gene expression phenotypes and searching. *Nucleic Acids Res.* 45: D758–D768.
- Howe, K., M. D. Clark, C. F. Torroja, J. Torrance, C. Berthelot *et al.*, 2013 The zebrafish reference genome sequence and its relationship to the human genome. *Nature* 496: 498–503.
- IOM, 2010 *Rare Diseases and Orphan Products: Accelerating Research and Development*. National Academies Press, Washington, DC.
- James, R. A., I. M. Campbell, E. S. Chen, P. M. Boone, M. A. Rao *et al.*, 2016 A visual and curatorial approach to clinical variant prioritization and disease gene discovery in genome-wide diagnostics. *Genome Med.* 8: 13.
- Jarman, A. P., Y. Grau, L. Y. Jan, and Y. N. Jan, 1993 atonal is a proneural gene that directs chordotonal organ formation in the *Drosophila* peripheral nervous system. *Cell* 73: 1307–1321.
- Kachroo, A. H., J. M. Laurent, C. M. Yellman, A. G. Meyer, C. O. Wilke *et al.*, 2015 Evolution. Systematic humanization of yeast genes reveals conserved functions and genetic modularity. *Science* 348: 921–925.
- Kandel, E. R., 2001 The molecular biology of memory storage: a dialogue between genes and synapses. *Science* 294: 1030–1038.
- Karaca, E., S. Weitzer, D. Pehlivan, H. Shiraiishi, T. Gogakos *et al.*, 2014 Human CLP1 mutations alter tRNA biogenesis, affecting both peripheral and central nervous system function. *Cell* 157: 636–650.
- Kremer, L. S., F. Distelmaier, B. Alhaddad, M. Hempel, A. Iuso *et al.*, 2016 Bi-allelic truncating mutations in TANGO2 cause infancy-onset recurrent metabolic crises with encephalocardiomyopathy. *Am. J. Hum. Genet.* 98: 358–362.
- Kuang, S. Q., O. Medina-Martinez, D. C. Guo, L. Gong, E. S. Regalado *et al.*, 2016 FOXE3 mutations predispose to thoracic aortic aneurysms and dissections. *J. Clin. Invest.* 126: 948–961.
- Lalani, S. R., P. Liu, J. A. Rosenfeld, L. B. Watkin, T. Chiang *et al.*, 2016 Recurrent muscle weakness with rhabdomyolysis, metabolic crises, and cardiac arrhythmia due to bi-allelic TANGO2 mutations. *Am. J. Hum. Genet.* 98: 347–357.

- Landoure, G., A. A. Zdebik, T. L. Martinez, B. G. Burnett, H. C. Stanescu *et al.*, 2010 Mutations in TRPV4 cause Charcot-Marie-Tooth disease type 2C. *Nat. Genet.* 42: 170–174.
- Law, R., T. Dixon-Salazar, J. Jerber, N. Cai, A. A. Abbasi *et al.*, 2014 Biallelic truncating mutations in FMN2, encoding the actin-regulatory protein Formin 2, cause nonsyndromic autosomal-recessive intellectual disability. *Am. J. Hum. Genet.* 95: 721–728.
- Lee, H., J. L. Deignan, N. Dorrani, S. P. Strom, S. Kantarci *et al.*, 2014 Clinical exome sequencing for genetic identification of rare Mendelian disorders. *JAMA* 312: 1880–1887.
- Lehner, B., 2013 Genotype to phenotype: lessons from model organisms for human genetics. *Nat. Rev. Genet.* 14: 168–178.
- Lek, M., K. J. Karczewski, E. V. Minikel, K. E. Samocha, E. Banks *et al.*, 2016 Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536: 285–291.
- Lemaitre, B., E. Nicolas, L. Michaut, J. M. Reichhart, and J. A. Hoffmann, 1996 The dorsoventral regulatory gene cassette spatzle/Toll/cactus controls the potent antifungal response in *Drosophila* adults. *Cell* 86: 973–983.
- Levy, S., G. Sutton, P. C. Ng, L. Feuk, A. L. Halpern *et al.*, 2007 The diploid genome sequence of an individual human. *PLoS Biol.* 5: e254.
- Lewis, E. B., 1978 A gene complex controlling segmentation in *Drosophila*. *Nature* 276: 565–570.
- Li, L., T. Meng, Z. Jia, G. Zhu, and B. Shi, 2010 Single nucleotide polymorphism associated with nonsyndromic cleft palate influences the processing of miR-140. *Am. J. Med. Genet. A.* 152A: 856–862.
- Lin, W. M., A. C. Baker, R. Beroukhim, W. Winckler, W. Feng *et al.*, 2008 Modeling genomic diversity and tumor dependency in malignant melanoma. *Cancer Res.* 68: 664–673.
- Lupski, J. R., J. G. Reid, C. Gonzaga-Jauregui, D. Rio Deiros, D. C. Chen *et al.*, 2010 Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N. Engl. J. Med.* 362: 1181–1191.
- Lupski, J. R., J. W. Belmont, E. Boerwinkle, and R. A. Gibbs, 2011 Clan genomics and the complex architecture of human disease. *Cell* 147: 32–43.
- Macgregor, S., G. W. Montgomery, J. Z. Liu, Z. Z. Zhao, A. K. Henders *et al.*, 2011 Genome-wide association study identifies a new melanoma susceptibility locus at 1q21.3. *Nat. Genet.* 43: 1114–1118.
- Manolio, T. A., D. M. Fowler, L. M. Starita, M. A. Haendel, D. G. MacArthur *et al.*, 2017 Bedside back to bench: building bridges between basic and clinical genomic research. *Cell* 169: 6–12.
- Markello, T. C., L. K. Pak, C. St Hilaire, H. Dorward, S. G. Ziegler *et al.*, 2011 Vascular pathology of medial arterial calcifications in NT5E deficiency: implications for the role of adenosine in pseudoxanthoma elasticum. *Mol. Genet. Metab.* 103: 44–50.
- Mayor, R., and E. Theveneau, 2013 The neural crest. *Development* 140: 2247–2251.
- McCarroll, S. A., A. Huett, P. Kuballa, S. D. Chileski, A. Landry *et al.*, 2008 Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. *Nat. Genet.* 40: 1107–1112.
- McGary, K. L., T. J. Park, J. O. Woods, H. J. Cha, J. B. Wallingford *et al.*, 2011 Systematic discovery of nonobvious human disease models through orthologous phenotypes. *Proc. Natl. Acad. Sci. USA* 107: 6544–6549.
- McGurk, L., A. Berson, and N. M. Bonini, 2015 *Drosophila* as an in vivo model for human neurodegenerative disease. *Genetics* 201: 377–402.
- Mishra-Gorur, K., A. O. Caglayan, A. E. Schaffer, C. Chabu, O. Henegariu *et al.*, 2014 Mutations in KATNB1 cause complex cerebral malformations by disrupting asymmetrically dividing neural progenitors. *Neuron* 84: 1226–1239.
- Mizushima, N., T. Noda, T. Yoshimori, Y. Tanaka, T. Ishii *et al.*, 1998 A protein conjugation system essential for autophagy. *Nature* 395: 395–398.
- Montell, C., 2005 *Drosophila* TRP channels. *Pflugers Arch.* 451: 19–28.
- Nagarkar-Jaiswal, S., P. T. Lee, M. E. Campbell, K. Chen, S. Anguiano-Zarate *et al.*, 2015 A library of MiMICs allows tagging of genes and reversible, spatial and temporal knockdown of proteins in *Drosophila*. *Elife* 4: e05338.
- Need, A. C., V. Shashi, Y. Hitomi, K. Schoch, K. V. Shianna *et al.*, 2012 Clinical application of exome sequencing in undiagnosed genetic conditions. *J. Med. Genet.* 49: 353–361.
- Nilius, B., and G. Owsianik, 2010 Channelopathies converge on TRPV4. *Nat. Genet.* 42: 98–100.
- Nishimura, G., E. Lausch, R. Savarirayan, M. Shiba, J. Spranger *et al.*, 2012 TRPV4-associated skeletal dysplasias. *Am. J. Med. Genet. C. Semin. Med. Genet.* 160C: 190–204.
- Novick, P., C. Field, and R. Schekman, 1980 Identification of 23 complementation groups required for post-translational events in the yeast secretory pathway. *Cell* 21: 205–215.
- Nurse, P., and Y. Bissett, 1981 Gene required in G1 for commitment to cell cycle and in G2 for control of mitosis in fission yeast. *Nature* 292: 558–560.
- Nusslein-Volhard, C., and E. Wieschaus, 1980 Mutations affecting segment number and polarity in *Drosophila*. *Nature* 287: 795–801.
- Person, A. D., S. Beiraghi, C. M. Sieben, S. Hermanson, A. N. Neumann *et al.*, 2010 WNT5A mutations in patients with autosomal dominant Robinow syndrome. *Dev. Dyn.* 239: 327–337.
- Philippakis, A. A., D. R. Azzariti, S. Beltran, A. J. Brookes, C. A. Brownstein *et al.*, 2015 The matchmaker exchange: a platform for rare disease gene discovery. *Hum. Mutat.* 36: 915–921.
- Phillips, J. B., and M. Westerfield, 2014 Zebrafish models in translational research: tipping the scales toward advancements in human health. *Dis. Model. Mech.* 7: 739–743.
- Pierson, T. M., D. Adams, F. Bonn, P. Martinelli, P. F. Cherukuri *et al.*, 2011 Whole-exome sequencing identifies homozygous AFG3L2 mutations in a spastic ataxia-neuropathy syndrome linked to mitochondrial m-AAA proteases. *PLoS Genet.* 7: e1002325.
- Pollock, P. M., U. L. Harper, K. S. Hansen, L. M. Yudt, M. Stark *et al.*, 2003 High frequency of BRAF mutations in nevi. *Nat. Genet.* 33: 19–20.
- Posey, J. E., J. A. Rosenfeld, R. A. James, M. Bainbridge, Z. Niu *et al.*, 2016 Molecular diagnostic experience of whole-exome sequencing in adult patients. *Genet. Med.* 18: 678–685.
- Ramoni, R. B., J. J. Mulvihill, D. R. Adams, P. Allard, E. A. Ashley *et al.*, 2017 The undiagnosed diseases network: accelerating discovery about health and disease. *Am. J. Hum. Genet.* 100: 185–192.
- Retterer, K., J. Juusola, M. T. Cho, P. Vitazka, F. Millan *et al.*, 2016 Clinical application of whole-exome sequencing across clinical indications. *Genet. Med.* 18: 696–704.
- Russell, J. J., J. A. Theriot, P. Sood, W. F. Marshall, L. F. Landweber *et al.*, 2017 Non-model model organisms. *BMC Biol.* 15: 55.
- Sarov, M., C. Barz, H. Jambor, M. Y. Hein, C. Schmied *et al.*, 2016 A genome-wide resource for the analysis of protein localisation in *Drosophila*. *Elife* 5: e12068.
- Saucedo, L. J., and B. A. Edgar, 2007 Filling out the Hippo pathway. *Nat. Rev. Mol. Cell Biol.* 8: 613–621.
- Sawyer, S. L., T. Hartley, D. A. Dymont, C. L. Beaulieu, J. Schwartzentruber *et al.*, 2016 Utility of whole-exome sequencing for those near the end of the diagnostic odyssey: time to address gaps in care. *Clin. Genet.* 89: 275–284.
- Sayson, B., M. A. Popurs, M. Lafek, R. Berkow, S. Stockler-Ipsiroglu *et al.*, 2015 Retrospective analysis supports algorithm as efficient diagnostic approach to treatable intellectual developmental disabilities. *Mol. Genet. Metab.* 115: 1–9.

- Schaffer, A. E., V. R. Eggen, A. O. Caglayan, M. S. Reuter, E. Scott *et al.*, 2014 CLP1 founder mutation links tRNA splicing and maturation to cerebellar development and neurodegeneration. *Cell* 157: 651–663.
- Schoch, K., L. Meng, S. Szeling, D. R. Bearden, A. Stray-Pedersen *et al.*, 2017 A recurrent de novo variant in NACC1 causes a syndrome characterized by infantile epilepsy, cataracts, and profound developmental delay. *Am. J. Hum. Genet.* 100: 343–351.
- Shah, A. N., C. F. Davey, A. C. Whitebitch, A. C. Miller, and C. B. Moens, 2015 Rapid reverse genetic screening using CRISPR in zebrafish. *Nat. Methods* 12: 535–540.
- Shashi, V., L. D. Pena, K. Kim, B. Burton, M. Hempel *et al.*, 2016 De novo truncating variants in ASXL2 are associated with a unique and recognizable clinical phenotype. *Am. J. Hum. Genet.* 99: 991–999.
- Skouloudaki, K., M. Puetz, M. Simons, J. R. Courbard, C. Boehlke *et al.*, 2009 Scribble participates in Hippo signaling and is required for normal zebrafish pronephros development. *Proc. Natl. Acad. Sci. USA* 106: 8579–8584.
- Sleven, H., S. J. Welsh, J. Yu, M. E. Churchill, C. F. Wright *et al.*, 2017 De novo mutations in EBF3 cause a neurodevelopmental syndrome. *Am. J. Hum. Genet.* 100: 138–150.
- Smith, J. D., A. V. Hing, C. M. Clarke, N. M. Johnson, F. A. Perez *et al.*, 2014 Exome sequencing identifies a recurrent de novo ZSWIM6 mutation associated with acromelic frontonasal dysostosis. *Am. J. Hum. Genet.* 95: 235–240.
- Sobreira, N., F. Schiettecatte, C. Boehm, D. Valle, and A. Hamosh, 2015a New tools for Mendelian disease gene identification: PhenoDB variant analysis module; and GeneMatcher, a web-based tool for linking investigators with an interest in the same gene. *Hum. Mutat.* 36: 425–431.
- Sobreira, N., F. Schiettecatte, D. Valle, and A. Hamosh, 2015b GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum. Mutat.* 36: 928–930.
- St Hilaire, C., S. G. Ziegler, T. C. Markello, A. Brusco, C. Groden *et al.*, 2011 NT5E mutations and arterial calcifications. *N. Engl. J. Med.* 364: 432–442.
- Stark, Z., T. Y. Tan, B. Chong, G. R. Brett, P. Yap *et al.*, 2016 A prospective evaluation of whole-exome sequencing as a first-tier molecular test in infants with suspected monogenic disorders. *Genet. Med.* 18: 1090–1096.
- Takehige, K., M. Baba, S. Tsuboi, T. Noda, and Y. Ohsumi, 1992 Autophagy in yeast demonstrated with proteinase-deficient mutants and conditions for its induction. *J. Cell Biol.* 119: 301–311.
- Tarailo-Graovac, M., C. Shyr, C. J. Ross, G. A. Horvath, R. Salvarinova *et al.*, 2016 Exome sequencing and the management of neurometabolic disorders. *N. Engl. J. Med.* 374: 2246–2255.
- Temple, G., D. S. Gerhard, R. Rasooly, E. A. Feingold, P. J. Good *et al.*, 2009 The completion of the Mammalian Gene Collection (MGC). *Genome Res.* 19: 2324–2333.
- Tift, C. J., and D. R. Adams, 2014 The National Institutes of Health undiagnosed diseases program. *Curr. Opin. Pediatr.* 26: 626–633.
- Toriyama, M., C. Lee, S. P. Taylor, I. Duran, D. H. Cohn *et al.*, 2016 The ciliopathy-associated CPLANE proteins direct basal body recruitment of intraflagellar transport machinery. *Nat. Genet.* 48: 648–656.
- Tsukada, M., and Y. Ohsumi, 1993 Isolation and characterization of autophagy-defective mutants of *Saccharomyces cerevisiae*. *FEBS Lett.* 333: 169–174.
- van Bokhoven, H., J. Celli, H. Kayserili, E. van Beusekom, S. Balci *et al.*, 2000 Mutation of the gene encoding the ROR2 tyrosine kinase causes autosomal recessive Robinow syndrome. *Nat. Genet.* 25: 423–426.
- van Karnebeek, C. D., L. Bonafe, X. Y. Wen, M. Tarailo-Graovac, S. Balzano *et al.*, 2016 NANS-mediated synthesis of sialic acid is required for brain and skeletal development. *Nat. Genet.* 48: 777–784.
- Venken, K. J., K. L. Schulze, N. A. Haelterman, H. Pan, Y. He *et al.*, 2011 MiMIC: a highly versatile transposon insertion resource for engineering *Drosophila melanogaster* genes. *Nat. Methods* 8: 737–743.
- Wang, J., R. Al-Ouran, Y. Hu, S. Y. Kim, Y. W. Wan *et al.*, 2017 MARRVEL: integration of human and model organism genetic resources to facilitate functional annotation of the human genome. *Am. J. Hum. Genet.* 100: 843–853.
- Wang, V. Y., B. A. Hassan, H. J. Bellen, and H. Y. Zoghbi, 2002 *Drosophila* atonal fully rescues the phenotype of Math1 null mice: new functions evolve in new cellular contexts. *Curr. Biol.* 12: 1611–1616.
- Wangler, M. F., S. Yamamoto, and H. J. Bellen, 2015 Fruit flies in biomedical research. *Genetics* 199: 639–653.
- Wangler, M. F., Y. Hu, and J. M. Shulman, 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.
- Wenger, A. M., H. Guturu, J. A. Bernstein, and G. Bejerano, 2017 Systematic reanalysis of clinical exome data yields additional diagnoses: implications for providers. *Genet. Med.* 19: 209–214.
- Wheeler, D. A., M. Srinivasan, M. Egholm, Y. Shen, L. Chen *et al.*, 2008 The complete genome of an individual by massively parallel DNA sequencing. *Nature* 452: 872–876.
- White, J., J. F. Mazzeu, A. Hoischen, S. N. Jhangiani, T. Gambin *et al.*, 2015 DVL1 frameshift mutations clustering in the penultimate exon cause autosomal-dominant Robinow syndrome. *Am. J. Hum. Genet.* 96: 612–622.
- White, J. J., J. F. Mazzeu, A. Hoischen, Y. Bayram, M. Withers *et al.*, 2016 DVL3 alleles resulting in a -1 frameshift of the last exon mediate autosomal-dominant Robinow syndrome. *Am. J. Hum. Genet.* 98: 553–561.
- White, R. M., J. Cech, S. Ratanasirintrao, C. Y. Lin, P. B. Rahl *et al.*, 2011 DHODH modulates transcriptional elongation in the neural crest and melanoma. *Nature* 471: 518–522.
- Wu, N., X. Ming, J. Xiao, Z. Wu, X. Chen *et al.*, 2015 TBX6 null variants and a common hypomorphic allele in congenital scoliosis. *N. Engl. J. Med.* 372: 341–350.
- Yamamoto, S., M. Jaiswal, W. L. Charng, T. Gambin, E. Karaca *et al.*, 2014 A *Drosophila* genetic resource of mutants to study mechanisms underlying human genetic diseases. *Cell* 159: 200–214.
- Yang, Y., D. M. Muzny, J. G. Reid, M. N. Bainbridge, A. Willis *et al.*, 2013 Clinical whole-exome sequencing for the diagnosis of Mendelian disorders. *N. Engl. J. Med.* 369: 1502–1511.
- Yang, Y., D. M. Muzny, F. Xia, Z. Niu, R. Person *et al.*, 2014 Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA* 312: 1870–1879.
- Yoon, W. H., H. Sandoval, S. Nagarkar-Jaiswal, M. Jaiswal, S. Yamamoto *et al.*, 2017 Loss of Nardilysin, a mitochondrial co-chaperone for alpha-ketoglutarate dehydrogenase, promotes mTORC1 activation and neurodegeneration. *Neuron* 93: 115–131.
- Zhang, F., and J. R. Lupski, 2015 Non-coding genetic variants in human disease. *Hum. Mol. Genet.* 24: R102–R110.
- Zhang, F., P. Seeman, P. Liu, M. A. Weterman, C. Gonzaga-Jauregui *et al.*, 2010 Mechanisms for nonrecurrent genomic rearrangements associated with CMT1A or HNPP: rare CNVs as a cause for missing heritability. *Am. J. Hum. Genet.* 86: 892–903.

Communicating editor: M. Johnston