Interpreting Human Genetic Variation With *In Vivo* Zebrafish Assays

Erica E. Davis, PhD, Stephan Frangakis, BS, and Nicholas Katsanis, PhD

Center for Human Disease Modeling
Duke University Medical Center
Durham, North Carolina

Society *for* Neuroscience

© 2015 Katsanis

NOTES

Introduction

Rapid advances and cost erosion in exome and genome analysis of patients with both rare and common genetic disorders have accelerated gene discovery and illuminated fundamental biological mechanisms. The thrill of discovery has been accompanied, however, by the sobering appreciation that human genomes are burdened with a large number of rare and ultra-rare variants, thereby posing a significant challenge in dissecting both the effect of such alleles on protein function and the biological relevance of these events to patient pathology. In an effort to develop model systems that are able to generate surrogates of human pathologies, a powerful suite of tools has been developed in zebrafish, taking advantage of the relatively small (compared with invertebrate models) evolutionary distance of that genome to humans, the orthology of several organs and signaling processes, and the suitability of this organism for medium- and high-throughput phenotypic screening. Here we will review the use of this model organism in dissecting human genetic disorders; we will highlight how diverse strategies have informed disease causality and genetic architecture; and we will discuss relative strengths and limitations of these approaches in the context of medical genome sequencing.

Challenges and Opportunities From Human Genetics

Major inflections in genomic advances have always been accompanied by accelerated discovery of lesions associated with human pathologies. The development of the first karyotype led rapidly to the discovery of syndromes of polyploidy (Lejeune et al., 1959), while the then nascent technologies of genome mapping, cloning, and sequencing yielded early insights into rare disease pathogenesis (Collins, 1992). As the field progressed, molecular cytogenetics at the sub-Mb and ultimately kb-level resolution revealed the high contribution of copy number variants (CNVs) to both rare and common human genetic disorders (Golzio and Katsanis, 2013), while, most recently, whole-exome and whole-genome sequencing (WES/WGS) has hyperaccelerated disease gene discovery both in historical cohorts and in the real-time clinical setting (Katsanis and Katsanis, 2013).

Amid the euphoria of discovery and the acutely increased expectations from patients and their physicians that the application of genomics can accelerate diagnosis and focus treatment options, the sobering realization has also emerged that each individual human genome is burdened with a large number of rare and ultra-rare alleles. Considering bona fide pathogenic mutations alone in the average human exome, studies have reported a median of 50–150 nonsense mutations, several in homozygosity, while the abundance of unique single nucleotide variants (SNVs) can be in the low-to-mid 100s (1000 Genomes Project Consortium et al., 2010). Importantly, the number of rare and ultra-rare SNVs has continued to increase proportionately to the number of available exomes and genomes (Tennessen et al., 2012), indicating that we are unlikely to reach saturation of such alleles soon. These observations have generated a significant interpretive problem for disease gene discovery and for clinical genomics, as population-based arguments alone have been unable to dissect the contribution of the majority of these alleles to clinical phenotypes. Computational algorithms that take into consideration a variety of evolutionary, structural, and biophysical properties of proteins have been of some assistance; however, their predictive ability (estimated in the 70–80% range [Castellana and Mazza, 2013]) has remained somewhat limited, mandating that definitive assessment of pathogenicity be carried out through other methods.

Animal studies combine the identification of candidate alleles for human diseases with mutant organisms that recapitulate the human mutation or loss of gene function, and have improved our understanding of the causal link between genetic mutation and phenotypic trait (Aitman et al., 2011). Numerous animal models have been developed to study both monogenic and complex disease. Each model system has its advantages and limitations, such as genetic and anatomic homology to humans, the size of the genetic toolkit, generation time, and cost. Here we will focus on the application of zebrafish in modeling human genetic disease; this organism has gained utility by bridging the gap between the high-throughput abilities of invertebrates and the orthology of structure of mammals (Tables 1, 2). Although not a panacea, the implementation of zebrafish complementation studies (Niederriter et al., 2013)—suppression of the orthologous zebrafish gene and rescue with either a mutant or wild-type human mRNA to determine pathogenicity—in human and medical genomics has facilitated disease gene discovery in both monogenic and complex traits, and has also found application in modeling more intricate (and challenging) genetic lesions that include CNVs and epistatic interactions. We will review the tools available, discuss their possible uses and limitations, and place the current vector of development of this model organism in the context of the ever-expanding generation of patient genomic data and the need for their accurate interpretation.

© 2015 Katsanis

NOTES

Table 1. General attributes and similarities of laboratory organisms used to model human genetic disease

	C. elegans	*D. melanogaster*	*D. rerio*	*M. Musculus*
Percent identity with *Homo sapiens*	43%	61%	70%	80%
Genome size	9.7×10^7 bp	1.3×10^8 bp	1.4×10^9 bp	2.5×10^9 bp
Exome size	28.1 Mb	30.9 Mb	96 Mb	49.6 Mb
Practical attributes				
Husbandry demands	$	$	$	$$$
Cost per animal	$	$	$	$$$
Characterized inbred strains	+	+	+	++++
Outbred laboratory strains	+	+	+++	++
Germline/embryonic cryopreservation	Yes	No	Yes	Yes
Lifespan	2 weeks	0.3 years	2–3 years	1.3–3 years
Generation interval	5.5 days	2 weeks	3 months	6–8 weeks
Number of offspring	300 larva	10–20 eggs	N 200 embryos/ clutch	10–12 pups/ litter
Embryonic development	*ex vivo*	*ex vivo*	*ex vivo*	*in utero*
Molecular biology tools				
Transgenesis*				
Gene targeting*	++++	+++	+	++++
Conditional gene targeting	+	++	+	++++
Transient in vivo assays*	+++	++	++++	+
Allelic series from TILLING*	+++	+++	++++	++
Affordability of large-scale screens**	++++	++++	+++	+
Cell biology tools				
Cell lines and tissue culture	+	++	+	++++
Antibody reagents	+	++	+	++++
In situ probes	+	+++	++++	+++
Disease process				
Birth defects	+	++	++++	++++
Adult-onset	++	+	+	++++
Behavioral	++	++	++	++
Aging	+++	++	++	++
Metabolic	++	++	+++	+++

*Reverse genetics.
**Forward genetic.

Adapted from a table of Lieschke and Currie (2007).

© 2015 Katsanis

NOTES

Animal Models of Human Genetic Disease

A deep understanding of the genetic architecture of human disease, underlying cellular and molecular mechanisms, and the development of therapeutic paradigms is dependent on model organisms that can robustly capture the pathology under investigation. Mammalian models such as the mouse (*Mus musculus*) have historically been attractive platforms by virtue of a high level of genomic sequence homology to humans (> 80%) (Mouse Genome Sequencing Consortium, et al., 2002), highly conserved anatomical and physiological features, and a diverse repertoire of gene-targeting strategies to recapitulate human disease phenotypes (Capecchi, 2005; Devoy et al., 2012). However, in the context of human genomics, time and cost now present significant drawbacks. By contrast, invertebrate models such as the nematode worm (*Caenorhabditis elegans*) or the fruit fly (*Drosophila melanogaster*) offer inexpensive alternatives, and have proven especially powerful for studying orthologous genes of interest through the use of sophisticated gene manipulation strategies (RNA interference, transposable insertion elements, etc.) (Antoshechkin and Sternberg, 2007; Wijshake et al., 2014). Nonetheless, the reduced cost and increased experimental tractability of these models are accompanied by a greater disparity in cellular processes and structures in comparison to humans, in large part owing to a decreased percentage of genes shared between species (43% and 61% for worm and fly versus human, respectively [Lander et al., 2001]). An intermediate model, the zebrafish (*Danio rerio*), has emerged as a strong candidate to achieve the experimental tractability of its invertebrate counterparts, but with the genomic and physiological proximity of a vertebrate for the investigation of human genetic disease.

Zebrafish: An Overview

The zebrafish is a tropical teleost that lives in the fresh waters of Southeast Asia. In the late 1960s, George Streisinger transitioned this common aquarium species to a model for basic research of embryogenesis and organ development because of its "desirable attributes," including a relatively short generation time of three to four months, the ability of mating pairs to generate several hundred embryos that develop rapidly and synchronously ex vivo, and the small size of adult fish (3 cm in length), making them easy to care for (Streisinger et al., 1981). Moreover, embryos are transparent, allowing facile microscopic visualization in the first days of development, with major organ formation occurring 24 h postfertilization. Zebrafish have a diploid genome but their genomic structure differs notably from that of other vertebrates by the major teleost-specific genome duplication that has resulted in subfunctionalization and neofunctionalization of genes (Amores et al., 1998; Postlethwait et al., 1998; Meyer and Schartl, 1999). Importantly, the biomedical research community now has a publicly available, extensively annotated version of the zebrafish genome at its disposal, of which 70% of genes have an identifiable human ortholog (Howe et al., 2013a). Additionally, a vast catalogue of mutants, transgenic reporters, and gene-specific expression data has been generated from over two decades of dedicated *D. rerio* use for "phenotype-driven" forward genetic screens and "gene-driven" reverse genetic approaches. These data are curated in ZFIN (the Zebrafish Model Organism Database), a community-wide resource warehousing genomic information, anatomical atlases, molecular tools, and links to zebrafish publications (www.zfin.org; Howe et al., 2013b).

Forward Genetics: Advances in Vertebrate Developmental Biology

Initial forays into zebrafish research predated the precise knowledge of gene content or location within the zebrafish genome, and were not necessarily motivated by targeted questions of human pathology. Rather, most forward screens were conducted to understand vertebrate embryonic development by (1) introducing random mutations throughout the genome; (2) conducting an informative breeding scheme to generate progeny with homozygous recessive mutations; (3) evaluating animals for a measurable phenotypic readout; and (4) identifying the mutation and gene underscoring the phenotype. Used widely across multiple model organisms, the application of this traditionally laborious strategy in zebrafish has been reviewed extensively elsewhere (Lawson and Wolfe, 2011).

The first zebrafish screens were reported in the 1980s and involved the application of gamma rays to induce recessive lethal mutations. However, this approach resulted in significant chromosomal breaks that rendered mapping to a single locus difficult (Chakrabarti et al., 1983; Streisinger, 1983). Alkylating agents, primarily N-ethyl-N-nitrosourea (ENU), replaced gamma rays as an effective mutagen, and application resulted in discrete genomic mutagenesis in zebrafish germ cells that

© 2015 Katsanis

NOTES

Table 2. Anatomical comparisons between zebrafish and humans

Anatomy	Key similarities
Embryology	• Cleavage, early patterning, gastrulation, somitogenesis, organogenesis are all represented
Skeletal system	• Ossified skeleton comprising cartilage and bone
Muscle	• Axial and appendicular muscle groups • Skeletal, cardiac, and smooth muscle cell types, with similar cellular architecture and machinery • Fast and slow skeletal muscle fibers
Nervous system and behavior	• CNS anatomy: forebrain, midbrain, and hindbrain, including diencephalon, telencephalon, and cerebellum • Peripheral nervous system has motor and sensory components • Enteric and autonomic nervous systems • Specialized sensory organs, eye, olfactory system, and vestibular system are well conserved • Complex behaviors and integrated neural function: memory, conditioned responses and social behaviors (e.g., schooling)
Hematopoietic and lymphoid/ immune systems	• Multiple hematopoietic cell types: erythrocytes, myeloid cells (neutrophils, eosinophils, monocytes and macrophages), T- and B-lymphocytes • Coagulation cascade for hemostasis • Innate and adaptive humoral and cellular immunity
Cardiovascular system	• Multichamber heart with an atrium and ventricle • Circulation within arteries and veins • Separate lymphatic circulation • Cardiac differentiation occurs through similar signaling pathways (e.g., *nkx2.5, bmp2b*) • Similar electrical properties and conduction patterns (SA node, slow atrial conductance, AV node, fast ventricular conductance)
Respiratory system	• Cellular gas exchange • Oxygenation is dependent on circulation and hemoglobin carriage
Gastrointestinal system	• Major organs: liver, exocrine, and endocrine pancreas, gall bladder • Zonal specializations along the length of the absorptive alimentary • Immune cells in lamina propria
Renal and urinary systems	• Glomerular anatomy and function
Reproductive system	• Molecular and embryological biology of germ-cell development • Cellular anatomy of germ-cell organs, the testis, and ovary
Endocrine system	• Most endocrine systems represented, including hypothalamic/hypophyseal axis (glucocorticoids, growth hormone, thyroid hormone, prolactin), parathyroid hormone, insulin, and rennin
Skin and appendages	• Ectodermal derivative • Pigmentation pattern is due to neural crest–derived pigment cells, including melanocytes

Table adapted from Lieschke and Currie (2007).

© 2015 Katsanis

NOTES

Anatomy	Key differences
Embryology	• Rapid • Influence of maternal transcripts • Nonplacental, involves hatching
Skeletal system	• Lack long bone, cancellous bone, and bone marrow • Joints are not weight-bearing
Muscle	• Tail-driven locomotion depends on alternating contraction of myotomal muscle • Appendicular muscle bulk is proportionately small
Nervous system and behavior	• Fish-specific sensory organs, such as the lateral line • Fish behaviors and cognitive function are simplified compared with human behavior • Significant difference in population of dopaminergic neurons (telencephalic vs midbrain) • Some immediate early genes and neuropeptides not conserved in zebrafish
Hematopoietic and lymphoid/ immune systems	• Erythrocytes are nucleated • Possess thrombocytes rather than platelets • Kidney interstitium is the hematopoietic site
Cardiovascular system	• Has left–right distinctions in cardiac anatomy, but does not have separate left–right circulations; that is, the heart has only two chambers • So far, no evidence for secondary heart field derivatives • Lymph nodes have not been described • Embryos are not dependent on functioning CV system for larval development • Atria and ventricles express different myosin heavy chains during development (human hearts only later differentiate between atrial and ventricular mhc) • Heart has high regenerative capacity, even in adult animals
Respiratory system	• Respiration occurs in gills, not lungs • No pulmonary circulation • Endoderm-derived swim bladder (functioning as a variable buoyancy device), which corresponds embryologically but not functionally to the lungs
Gastrointestinal system	• Lack an acidified digestive organ tract • Have an intestinal bulb rather than stomach • Intestinal Paneth cells not present
Renal and urinary systems	• Filtration occurs in anterior and posterior kidneys • Mesonephric rather than metanephric adult kidney • No bladder or prostate gland • No structure in zebrafish homologous to descending or ascending thin limb of nephron in mammals
Reproductive system	• No sex chromosomes • Mechanism of sex determination is uncertain • Fertilization is *ex vivo* (no uterus or the related internal female reproductive organs) • Oocytes are surrounded by a chorion, not the zona pellucida, which must be penetrated by sperm • Nonlactating; no breast equivalent
Endocrine system	• Differences in anatomical distribution of glands, e.g., discrete parathyroid glands do not seem to be present • Prolactin has a primary role in osmoregulation
Skin and appendages	• Lack appendages (hair follicles, sebaceous glands) • Additional pigment cell types: xanthophores and iridophores

© 2015 Katsanis

NOTES

could be mapped to a single gene (~1 mutant per genome evaluated) (Mullins et al., 1994; Solnica-Krezel et al., 1994). This discovery led to large-scale efforts by labs in Tübingen, Germany, and Boston to apply ENU screening to zebrafish. Within two years, their combined efforts led to the characterization of ~4000 embryonic lethal phenotypes; these include gastrulation (Hammerschmidt et al., 1996; Solnica-Krezel et al., 1996); somitogenesis (van Eeden et al., 1996); brain (Brand et al., 1996; Jiang et al., 1996; Schier et al., 1996); cardiovascular (Stainier et al., 1996); and craniofacial development mutants (Neuhauss et al., 1996; Piotrowski et al., 1996; Schilling et al., 1996).

Although forward genetic screens in zebrafish contributed significantly to the fundamental understanding of early embryonic development, the impact on such findings to inherited disease in humans has been sporadic. This modest connection can be attributed to four main reasons. First, such screens are unable to capture alleles that confer a dominant negative (antagonizes the wild-type protein function) or gain-of-function (mutation confers a protein function different from that of wild-type protein) effect. Second, phenotypes must have a measurable phenotypic readout in early embryonic or larval stages, decreasing the possibility of detecting adult-onset or degenerative phenotypes. Third, although such screens were able to uncover discrete gene functions, the odds of generating precisely the same allele by ENU as has been seen in a patient is remote. Finally, this approach is confounded further by the fact that the zebrafish genome underwent a teleost-specific duplication (Amores et al., 1998; Postlethwait et al., 1998; Meyer et al., 1999). Among the genes for which there is an identifiable human ortholog, 47% have a one-to-one orthologous relationship with a human counterpart, while the remainder of zebrafish genes have complicated one-to-many or many-to-one orthology in comparison to the human gene (Howe et al., 2013a). As a result, duplicated gene function may either be (a) retained in both copies, making them functionally redundant; (b) lost in one of the two copies, wherein it becomes a pseudogene; or (c) a novel and divergent gene function is acquired by one of the two copies. Therefore, mutations in only one of two functionally redundant orthologs might not display a phenotype.

Nonetheless, ENU mutants have been successful in drawing anatomical correlates for genes implicated in recessive human disorders that cause anatomical birth defects. For instance, the craniofacial mutant *crusher*^{m299} is caused by a nonsense mutation in *sec23a* (Lang et al., 2006); at the same time as this discovery, *SEC23A* mutations in humans were shown to cause a clinically relevant craniofacial dysmorphology, cranio-lenticulo-sutural dysplasia, bolstering the evidence of causality in both species (Boyadjiev et al., 2006). Importantly, the recent application of WGS (Obholzer et al., 2012), WES (Ryan et al., 2013), and improved mapping strategies (Leshchiner et al., 2012) to zebrafish ENU mutants has enabled the rapid and cost-effective identification of mutations, justifying the continued use of forward genetics to assist with assigning causality in human genetic disease.

Reverse Genetics: From Candidate Causal Gene to Physiologically Relevant Animal Model

Forward genetic screening involves the unbiased examination of phenotypes resulting from mutations in the zebrafish genome. However, the randomness of this approach is hampered by the inability to specifically target every coding gene and/or specific mutations implicated in human pathology. To circumvent this problem, the precise targeting of candidate genes and alleles can be achieved through several methods that have been developed over the past ~15 years.

First, transient gene manipulation can be achieved through the injection of either morpholino (MO) antisense oligonucleotides (suppression) or capped *in vitro* transcribed messenger RNA (mRNA) (ectopic expression) into zebrafish embryos. MOs are stable molecules that consist of a large, nonribose morpholine backbone with the four DNA bases pairing stably with mRNA at either the translation start site (to disrupt protein synthesis) or at intron–exon boundaries (to disrupt mRNA splicing) (Summerton et al., 1997). The use of MOs to confer effective gene knockdown was first shown in zebrafish in 2000; Nasevicius and Ekker recapitulated the developmental phenotypes of five different embryonic lethal mutants and developed models of the human genetic disorders hepatoerythropoietic porphyria and holoprosencephaly through the suppression of *urod* and *shh*, respectively (Nasevicius and Ekker, 2000). Since this report, MOs have been used broadly to study vertebrate development and disease; coinjection of MO and orthologous mRNA has been employed for the systematic functional testing of alleles identified in humans, offering a powerful approach for analysis of variant pathogenicity and direction of effect (Niederriter et al., 2013). Still, this methodology does have notable drawbacks: (1) MO efficacy is limited to ~3–5 d

© 2015 Katsanis

NOTES

(Nasevicius and Ekker, 2000), and similarly, the presence of mRNA is limited to the same embryonic timeframe; (2) with few exceptions (Shestopalov et al., 2007), injected MOs and mRNAs do not confer spatial- or temporal-specific activity; and (3) MOs can give rise to spurious phenotypes resulting from off-target effects (Eisen and Smith, 2008). Even so, the use of this methodology within the appropriate developmental stage, and with the appropriate experimental controls (a) targeting with a splice-blocking MO to demonstrate incorrectly spliced RNA; (b) specific rescue of MO phenotypes with orthologous wild-type mRNA; (c) demonstration of a similar phenotype with multiple MOs targeting the same gene; or (d) comparison with a mutant when possible, and if appropriate (Eisen and Smith, 2008) can allow for the correct interpretation of MO phenotypes to establish relevance to human genetic disease through the recapitulation of loss-of-function or dominant negative effects.

Second, it is possible to readily obtain germline zebrafish mutants for a gene of interest; doing so avoids the phenotypic variability associated with MOs and allows the observation of phenotypes beyond early larval stages. Targeting Induced Local Lesions in Genomes (TILLING) was the first reverse genetic approach to produce germline mutations in a gene of interest. Similar to forward screens, TILLING involves ENU mutagenesis of adult male zebrafish and generation of F1 families. Sperm from F1 males is then cryopreserved while genomic lesions are screened in target genes, typically in early exons or near exonic regions encoding critical protein domains, through PCR amplicon screening (Weinholds et al., 2002). The completion of the zebrafish genome coupled to next-generation sequencing has increased significantly the throughput of the screening aspect of this strategy. TILLING mutants have been identified for > 38% of all known zebrafish protein coding genes (Sanger Institute Zebrafish Genome Project: http://www.sanger.ac.uk/Projects/D_rerio/zmp; Kettleborough et al., 2013); this corresponds to ~60% of orthologous genes associated with a human phenotype in the Online Mendelian Inheritance in Man (OMIM: http://www.omim.org) database. The ongoing TILLING efforts hope to generate a comprehensive resource of putative null or hypomorphic models of human genetic disease; however, it is critical to be cognizant of the possibility that ENU may introduce multiple lesions in the genome. Ideally, multiple mutants with different alleles in the same gene should be phenotypically characterized to ensure that the pathology is specific. The same guidelines are true for retrovirus (Wang et al., 2007) or transposon (Sivasubbu et al., 2007) insertional mutants used in similar reverse genetics approaches.

Both forward ENU screens and TILLING are laborious; alternative approaches have recently expanded the utility of the zebrafish by enabling precise and germline transmittable gene targeting that does not require excessive downstream screening to identify mutations (Wijshake et al., 2014). First, zinc finger nucleases (ZFNs) utilize a zinc finger array to enable target sequence specificity (typically, the early exon of a gene), and a *FokI* endonuclease to guide cleavage and subsequent repair at the target site (Urnov et al., 2010); this was first utilized to target the gol locus (mutation of which results in absence of pigment), *ntl* (a regulator of early embryogenesis), and *kdr* (vascular endothelial growth factor-2 receptor), as visible proof-of-principle phenotypes (Doyon et al., 2008; Meng et al., 2008). Second, transcription activator–like TAL effector nucleases (TALENs) have similarly been optimized to achieve locus-specific genome editing and have been shown to achieve greater specificity of and alteration of target sequences than ZFNs (Bedell et al., 2012). A third, more recent advancement in zebrafish genome editing technology involves clustered, regularly interspaced, short palindromic repeats (CRISPRs), bacterial type II systems that guide RNAs to direct site-specific DNA cleavage by the Cas9 endonuclease (Hwang et al., 2013). Each of ZFN, TALEN, and CRISPR technologies have expanded the molecular toolkit of the zebrafish (for comparisons, see Table 3), accelerating studies of vertebrate development and improving our understanding of analogous phenotypes to human disease. For example, CRISPR/Cas9 was used to edit the *gata5* locus (Chang et al., 2013), and mutant embryos displayed a cardia bifida phenotype mimicking both the *faut*^{m236a} zebrafish mutant (Reiter et al., 1999) and humans with congenital heart defects (Padang et al., 2012; Wei et al., 2013a, b). However, there is currently a relative paucity of reports in which human-driven WES/WGS studies have been followed with the generation of such stable mutants; this is largely the result of the relative newness of the technology and/or the amount of time and labor still required to generate and characterize mutants; we anticipate the landscape of the field to change rapidly in the coming months and years.

Humanizing the Zebrafish to Study Mutations Detected in Humans

Taken together, the zebrafish exemplifies a tractable and physiologically relevant tool to model genetic variation in humans. Each of the forward and reverse genetics tools has limitations, and in particular,

© 2015 Katsanis

NOTES

Table 3. Comparison of different reverse genetics approaches

Methodology	Class	Specificity (targeting sequence)	Off-target effects	Technology adoption time and costs	Ongoing time and cost	Somatic DNA cutting efficiency	Germline efficacy
Morpholinos	Morpholino oligo	25 nt	Low	Low	Low	n/a	n/a
ZFNs	Protein-DNA	18+ bp (pair)	Low	High	Moderate	Low (~2%)	Low
TALENs	Protein-DNA	30+ bp (pair)	Ultra low	Moderate	Low	Moderate to high (~20–50%)	Moderate to high
CRISPR-Cas9	RNA-DNA	12+ bp	Ultra low	Low	Low	Moderate to high (~30–60%)	Moderate to high

Table adapted from Blackburn et al. (2013).

places significant emphasis on the study of loss-of-function effects of single genes, potentially making them an overly simplistic model to investigate oligogenic or even complex traits. In a growing number of instances, however, it has been possible to balance experimental tractability, specificity, and cross-species phenotypic similarity to establish:

(1) Physiological relevance of a gene to a human clinical phenotype;

(2) Allele pathogenicity; and

(3) Direction of allele effect for a vast array of human genetic disorders with diverse models of inheritance, phenotypes, and ages of onset.

Recessive disease

Disorders that segregate under a recessive mode of inheritance, especially congenital or pediatric-onset disorders with an abnormality in a structure with an anatomical counterpart in the developing zebrafish, have achieved widespread use toward demonstrating physiological relevance. This often represents the extent of functional data presented in instances when the human mutations have an unambiguous loss-of-function effect on the protein (nonsense, frameshift, or splice-site). For instance, causal mutations identified in primary ciliary dyskinesia (PCD) are almost exclusively null changes, and transient MO-based studies in zebrafish have shown that proteins of *a priori* unknown function, including *CCDC39*, *ARMC4*, and *ZMYND10*, give rise to left–right asymmetry defects phenotypes found in humans (Merveille et al., 2011; Hjeij et al., 2013; Zariwala et al., 2013). In other recessive disorders, such as pontocerebellar hypoplasia (PCH), the zebrafish has assisted in establishing clinical relevance and allele pathogenicity (Fig. 1A). Wan et al. identified nonsynonymous changes in *EXOSC3*, encoding exosome component 3, following WES of four affected siblings; MO-induced suppression resulted in phenotypes that were relevant to the human clinical features of microcephaly and reduced motility in *exosc3* morphants. Additionally, *in vivo* complementation of *exosc3* MO phenotypes with either zebrafish or human mRNA harboring the missense mutations found in patients failed to rescue the phenotype, indicating that these were loss-of-function alleles (Wan et al., 2012). Even so, transient *in vivo* complementation assays are not applicable to every gene. Human genes with an open reading frame (ORF) larger than ~6 kb are challenging to transcribe *in vitro*, likely explaining why large genes such as *NBEAL2*, encoding neurobeachin-like 2, the novel genetic cause for gray platelet syndrome (ORF of 8.2 kb), were shown to cause a relevant thrombocytopenia phenotype in zebrafish morphants, but the missense mutations identified in patients were not tested (Albers et al., 2011).

Dominant disorders

In contrast with recessive disorders, in which the allele effect is typically loss-of-function, traits that segregate under an autosomal dominant inheritance pattern are the result of either a haploinsufficiency, dominant negative, or gain-of-function mechanism. For some dominant pediatric-onset disorders, the genetic evidence of a heterozygous null variant segregating in a large pedigree with fully penetrant disease is sufficient to suggest that the direction of allele effect is haploinsufficiency, and these predictions have been confirmed in zebrafish through MO-induced gene suppression. For example, dilated cardiomyopathy is caused by nonsense, splice-site, or missense mutations in the gene encoding heat shock

© 2015 Katsanis

NOTES

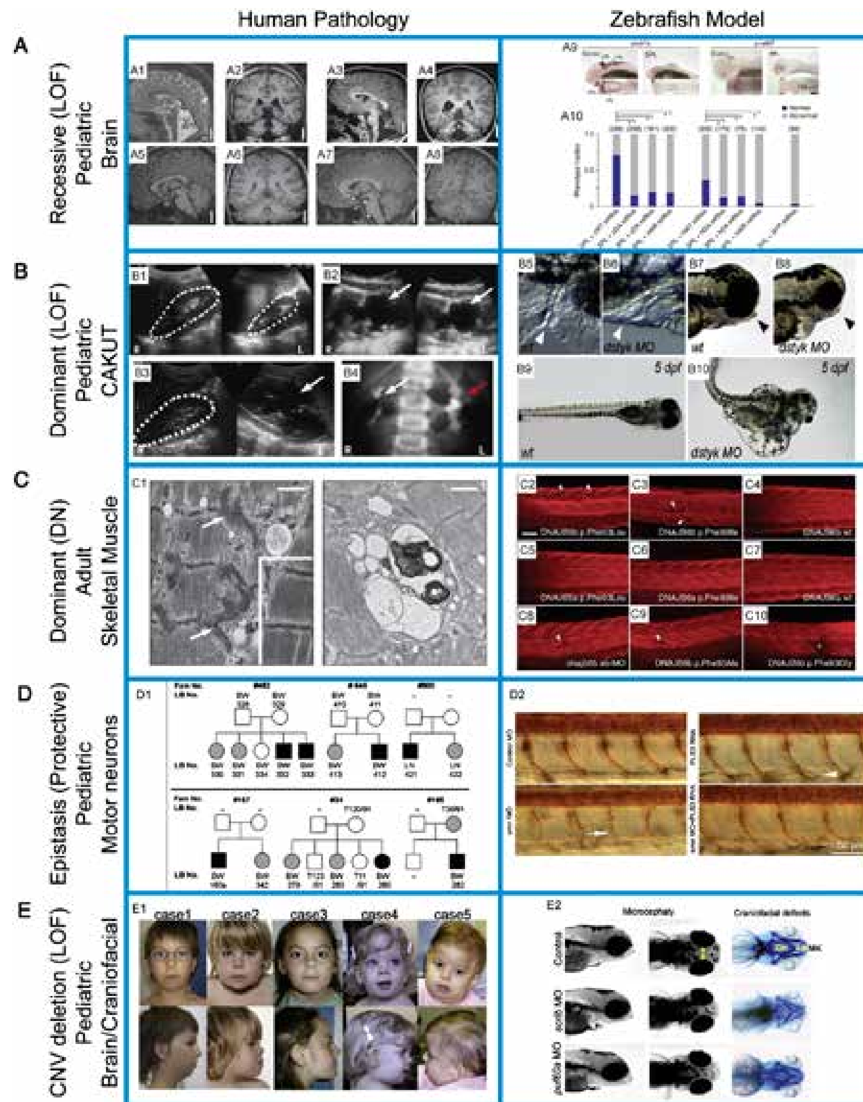

Figure 1. Zebrafish models of human genetic disease. *A*, Mutations in pontocerebellar hypoplasia caused by *EXOSC3* and *in vivo* complementation studies in zebrafish recapitulate the brain phenotypes observed in patients and demonstrate that missense mutations are functional null variants. Left panel, neuroimaging of affected individuals (A1–A4; top row) and control images (A5–A8; bottom row); right panel, whole-mount *in situ* hybridization in-splice blocking MO-injected embryos in lateral view (A9; inset: dorsal view, with rostral to the left) demonstrated diminished expression of dorsal hindbrain progenitor-specific marker *atoh1a* and cerebellar-specific marker *pvalb7* (A10; quantification). Images reproduced from Wan et al. (2012). *B*, Congenital abnormalities of the kidney and urinary tract are caused by haploinsufficiency of *DSTYK*. Left, panels B1–B3, with hypoplasia of the left kidney (Panel B1, kidneys outlined by dashed lines), bilateral hydronephrosis (Panel B2, arrows) caused by ureteropelvic junction obstruction detected at birth, and hydronephrosis only of the left kidney (Panel B3, arrow) caused by ureteropelvic junction obstruction. The intravenous pyelogram in Panel B4 shows blunting of fornices on the right side (white arrow); right, MO-induced knockdown of *dstyk* embryos, live lateral images show absence of the patent pronephric duct opening (arrows). Images reproduced from Sanna-Cherchi et al. (2013). *C*, Adult-onset limb-girdle muscular dystrophy is caused by dominant negative mutations in *DNAJB6*. C1, Transmission electron microscopy showed early disruption of Z-disks (arrows; left) and autophagic pathology (right) in LGMD1D; C2–C10, lateral views of zebrafish embryos 2 d postfertilization subjected to whole-mount immunofluorescence staining of slow myosin heavy-chain display myofiber abnormalities (arrows). Images reproduced from Sarparanta et al. (2012). *D*, PLS3 overexpression exerts a protective effect on *SMN1* deletion to rescue motor neuron defects in spinal muscular atrophy (SMA). D1, pedigrees of SMA-discordant families showing unaffected (gray) and affected (black) *SMN1*-deleted siblings; D2, lateral view of zebrafish embryos injected with control MO, *smn* MO, *PLS3* RNA, and *smn* MO + PLS3 RNA. Motor axons were visualized with *znp1* antibody at 36 h postfertilization and show rescue of *smn* MO with *PLS3* RNA. Images reproduced from Oprea et al. (2008). *E, SCRIB* and *PUF60* suppression drive the multisystemic phenotypes of the 8q24.3 CNV. E1, Photographs of five individuals with the 8q23.4 CNV show craniofacial abnormalities and microcephaly; E2, Lateral and dorsal views of control and *scrib* or *puf60a* MO-injected embryos at 5 dpf show head size and craniofacial defects observed in affected individuals. Images reproduced from Daubert et al. (2013). LOF, loss-of-function; DN, dominant negative.

© 2015 Katsanis

NOTES

protein co-chaperone BCL2-associated athanogene 3 (*BAG3*), and gene suppression results in similar cardiac phenotypes in zebrafish embryos (Norton et al., 2011). Similarly, congenital abnormalities of the kidney and the urinary tract (CAKUT) associated with a loss-of-function splice-site mutation segregating in a dominant pedigree were identified in the dual serine–threonine kinase encoded by *DSTYK* (Fig. 1*B*); the human phenotypes were recapitulated in *dstyk* morphant embryos (Sanna-Cherchi et al., 2013). Transient experiments in zebrafish embryos can also determine allele pathogenicity and capture a dominant negative direction of effect that is isoform-specific. This is exemplified by the transient functional studies of missense mutations in the co-chaperone protein, DNAJB6, associated recently with adult-onset limb girdle muscular dystrophy (Fig. 1*C*) (Sarparanta et al., 2012). Coinjection of mutant DNAJB6 mRNA in the presence of equivalent amounts of wild-type transcript resulted in myofiber defects in zebrafish embryos; injection of increasing amounts wild-type mRNA with a fixed concentration of mutant resulted in a phenotypic rescue, indicating the dominant toxicity of the mutant alleles. Given the nature of these mutations (deleterious in heterozygosity), the use of zebrafish for dissecting dominant disorders will likely remain restricted to transient MO- and mRNA-based studies until the development of more sophisticated conditional gene suppression/expression techniques in zebrafish.

De novo variants

Variants that arise *de novo* as a product of either germline mosaicism or early developmental DNA replication errors are significant contributors to the human mutational burden (Veltman and Brunner, 2012). Similar to variants that underscore autosomal dominant disorders, *de novo* changes may give rise to clinical phenotypes produced from falling below a gene dosage threshold, dominant negative effects, or acquisition of a novel function. As such, an unbiased approach toward dissecting the direction of *de novo* allele effect is critical once physiological relevance has been determined. For instance, transient approaches have been carried out in zebrafish to investigate *de novo* missense mutations in *CACNA1C*, encoding the voltage-gated calcium channel $Ca_v1.2$, in the pathophysiology of Timothy syndrome (TS), a pediatric disorder characterized by cardiac arrhythmias, syndactyly, and craniofacial abnormalities. Ectopic expression of mutant mRNA and suppression of *cacna1c* in zebrafish embryos not only revealed that the mutation confers a gain-of-function effect, but also demonstrated a novel role for $Ca_v1.2$ in the nonexcitable cells of the developing jaw (Ramachandran et al., 2013).

CNVs

Frequently arising *de novo*, copy number variants (CNVs) likewise represent a significant molecular basis for human genetic disease (Inoue and Lupski, 2002). These variations in genomic structure can range in size from a few thousand to millions of base pairs, are not identifiable by conventional chromosomal banding, and can encompass from one to hundreds of genes (Stankiewicz and Lupski, 2010). Although genotype–phenotype correlations among affected individuals with overlapping CNVs can assist in narrowing specific genetic drivers, CNVs have been historically intractable to functional interpretation in animal models, with sparse reports of human CNVs being modeled in the mouse (Lindsay, 2001). Zebrafish models have emerged recently as powerful tools to dissect both recurrent and nonrecurrent CNVs. First, systematic zebrafish modeling of the 29 genes in the recurrent reciprocal 16p11.2 duplication/deletion CNV—associated with a range of neurocognitive defects—found the main driver of the neuroanatomical phenotypes to be *KCTD13*, causing mirrored macrocephaly and microcephaly upon suppression or overexpression in zebrafish, respectively (Golzio et al., 2012). Second, MO-induced suppression of three genes in the 8q24.3 nonrecurrent deletion CNV in zebrafish embryos revealed that the planar cell polarity effector SCRIB, and the splicing factor PUF60 could be linked to distinct aspects of the renal, short stature, coloboma, and cardiac phenotypes observed in five individuals with overlapping microdeletions at this locus (Fig. 1*E*) (Dauber et al., 2013).

Second-site modifiers

The demonstration of second-site phenotype modification in primarily recessive human genetic disease has been fueled by the use of *in vivo* assays in zebrafish. The ciliopathies (disorders underscored by dysfunction of the primary cilium) have been causally linked with more than 50 different loci, can give rise to a constellation of human phenotypes, and have been an ideal system to study epistasis (Davis and Katsanis, 2012). The recent dissection of the genetic architecture of Bardet–Biedl syndrome (BBS), a ciliopathy hallmarked by retinal degeneration, obesity, postaxial polydactyly, renal abnormalities, and intellectual disability (1) informed the pathogenic potential of missense BBS alleles contributing to the disorder (null, hypomorphic, or dominant negative); (2) revealed the surprising contribution of dominant negative alleles in oligogenic pedigrees with BBS; and (3) provided sensitivity (98%) and specificity (82%) metrics for

© 2015 Katsanis

NOTES

the zebrafish *in vivo* complementation assay to predict allele pathogenicity (Zaghloul et al., 2010). Transient zebrafish *in vivo* complementation assays have similarly been used to identify RPGRIP1L A229T as a modulator of retinal endophenotypes (Khanna et al., 2009), RET as a modifier of Hirschsprung phenotypes in BBS (de Pontual et al., 2009), and TTC21B as a frequent contributor to mutational load in ciliopathies (Davis et al., 2011). Second-site modification phenomena are not unique to the ciliopathies; for example, overexpression of plastin 3 (*PLS3*) to mimic the gene expression in unaffected individuals improved the axon length and growth defects associated with *SMN1* deletion in spinal muscular atrophy (SMA). The interaction of these two genes was shown, in part, through modeling of SMA genotype and phenotype correlates in zebrafish embryos (Fig. 1*D*) (Oprea et al., 2008).

Complex traits

Genome-wide association studies (GWAS) alone have been hampered by an inability to connect risk association to genes and underlying mechanism. However, the zebrafish has emerged as a tool to dissect genes at or near loci that confer significant risk for the complex trait under investigation. Rare alleles in GWAS hits have been more straightforward to dissect, since the strategy has been similar to that of Mendelian traits. For example, a combination of *in vitro* analysis of enzyme stability and secretion and vascular integrity in the retina in zebrafish embryos demonstrated a functional role for a rare allele in the gene encoding complement factor I (CFI), thus providing direct evidence for a loss-of-function role of CFI in AMD (van de Ven et al., 2013). Zebrafish have also been used in the absence of candidate coding changes in GWAS-identified loci. In one example, Liu et al. assessed the physiological relevance of candidate genes identified from a GWAS of chronic kidney disease among African American populations; these efforts identified *KCNQ1* as a functionally relevant candidate owing to the glomerular filtration defects observed in *kcnq1* morphants (Liu et al., 2011). Similarly, functional validation of loci associated with platelet count in cohorts of European ancestry yielded 11 novel genes implicated in *D. rerio* blood cell formation (Gieger et al., 2011). Although numerous questions remain regarding the combinatorial effects of GWAS hits, their mechanistic basis for conferring risk, and the physiological relevance of significantly associated sites, both coding and noncoding, the zebrafish offers a tractable tool to begin to dissect existing GWAS data.

Adult-Onset Disease

The majority of zebrafish models discussed so far have been used to understand the role of genes and alleles in pediatric and congenital disorders, in large part because transient MO and mRNA analysis is possible only during development. Nonetheless, given the correct tools and appropriate assays, this model organism is also useful for the study of adult-onset disorders. AMD was one example described above, the utility of the model being extracted from the ability to model vascular integrity in zebrafish embryos, a phenotype relevant to AMD pathology. There are numerous other examples as well. In particular, phenotypic proxies of human neurodegenerative diseases such as schizophrenia, Huntington, Parkinson, and Alzheimer disease have been used to examine the role of various genes in these diseases (Best, 2008; Bandmann and Burton, 2010), although in almost all cases, these studies involved genetic mutants and/or stable transgenes. In one example, stable transgenic zebrafish expressing human 4-repeat Tau showed Tau accumulation within neuronal cell bodies and axons in neurons throughout the adult brain, resembling neurofibrillary tangles (Bai et al., 2007). Some defects, such as behavioral phenotypes, are not immediately observable through anatomic or histological methods, making it necessary to employ more-sensitive methods of analysis. To this extent, assays for memory and learning impairment (Arthur and Levin, 2001) and conditioned avoidance (Wullimann and Mueller, 2004) allow for quantifiable testing of subtle phenotypes in adult fish.

Conclusion: The Road Ahead

In the study of human genetics, animal models have provided insight into genetics and pathophysiology. The efficacy of a model organism always hinges on whether that organism appropriately models the target pathology of humans, and whether the experiments necessary to provide burden of proof are tractable and not cost-prohibitive. In the context of human and medical genomics, we anticipate that the entire spectrum of model organisms will continue to be used. Nonetheless, it is clear that models such as zebrafish and possibly other similar organisms, such as *Medaka* or *Xenopus* that offer transparency, low cost, and the ability to manipulate their genome efficiently will gain a prominent role as the community strives to model thousands of candidate disease–associated genes and alleles.

A key requirement for the widespread use of the zebrafish to determine pathogenicity of alleles

© 2015 Katsanis

NOTES

identified by WES/WGS moving forward will be to (1) improve throughput of physiologically relevant models of human disease (through the rapid generation of *in vivo* complementation assays); and (2) perform automated phenotyping and image analysis. While MOs may still be a tractable option for the evaluation of early developmental phenotypes, the recent ability to generate knock-in mutants (Auer et al., 2014), conditional zebrafish mutants (Maddison et al., 2014), and multiplexed mutants (Jao et al., 2013) holds great promise toward accelerating disease-modeling throughput. Equally important, some platforms for high-throughput phenotypic screening of zebrafish larvae have been developed recently (Pardo-Martin et al., 2013). For example, one system captures hundreds of three-dimensional morphological features with speed and accuracy, clustering quantitative phenotypic signatures so that multiple phenotypes can be detected and classified simultaneously. Nonetheless, this system is limited to bright-field images, rendering marker analysis (by RNA *in situ* or antibody staining) difficult. Given that embryo phenotyping represents the most significant bottleneck in scaling the use of zebrafish in human genomics, there is an acute need to develop additional transgenic reporter lines to assist with visualization of cellular and anatomical structures of interest (Kawakami et al., 2000), as well as imaging and embryo manipulation technologies further.

It is also important to note that zebrafish analysis, like all other genetic and molecular biology tools, has its limitations. MO studies can sometimes generate conflicting data, especially in the context of early developmental phenotypes that are most sensitive to toxic effects. Rescue studies and a minimum of two MOs per gene tested, when possible, are essential to validate findings. Similarly, multiple independent lines from genome editing experiments will need to be studied to ensure that the phenotypes observed are driven by the engineered mutation, not an off-site introduced allele. Moreover, not all human genes and alleles are modelable; some 25–30% of the human transcriptome is not present in zebrafish, while other genes can be difficult to model because of divergent functions or extreme dosage sensitivity (especially transcription factors), rendering them experimentally difficult. Further, for some disorders (e.g., pulmonary fibrosis) there cannot be a credible phenotypic surrogate, and distant surrogates might lead to incorrect conclusions. Finally, most zebrafish studies to date have focused on coding variation; modeling noncoding variation is significantly more taxing yet remains important, not least because regulatory regions are likely to be enriched for alleles that drive GWAS signals (McClellan and King, 2010). This work is possible once we recognize that evolutionary constraints might render some data uninterpretable. For example, testing multiple sequences located within a 50-kb block of the regulatory domain of *IRX3* (certain variants of which are associated with obesity in humans) resulted in transgenic zebrafish with expression in pancreas (Ragvin et al., 2010); knockdown of *irx3* in zebrafish reduced the number of pancreatic beta cells. However, given current designs, it will be difficult to execute such experiments at the throughout required to address the needs of the human genetics community.

Despite these limitations, modeling human variation in zebrafish embryos has been a significant contributor toward dissecting the causality of genes and alleles in human genetic disorders. Moreover, the development of human disease models will serve as a platform for the discovery of novel therapeutic paradigms by employing high-throughput small-molecule screening approaches (Zon and Peterson, 2005; Tan and Zon, 2011). Moving forward, we anticipate that the community, through the combinatorial use of all the tools discussed here, will saturate in the coming years a significant fraction of the morbid human genome. Ultimately, we imagine that such studies will inform the design of improved computational algorithms, probably through the training of thousands of human alleles tested *in vivo*, which will in turn represent the next inflection point in human and medical genomics.

Acknowledgments

This work was supported by grants R01HD04260, R01DK072301 and R01DK075972 to N.K.; R01EY021872 to E.E.D.; P30DK096493, P50DK096415, and funding from the European Union 7th Frameword Programme under Grant Agreement no. 241955, project SYSCILIA (to E.E.D. and N.K.). N.K. is a Distinguished Brumley Professor. This chapter was adapted from an article of the same title: Davis EA, Frangakis, S, Katsanis N (2014) Biochimica et Biophysica Acta 1842:1960–1970.

References

1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA (2010) A map of human genome variation from population-scale sequencing. Nature 467:1061–1073.

© 2015 Katsanis

NOTES

Aitman TJ, Boone C, Churchill GA, Hengartner MO, Mackay TF, Stemple DL (2011) The future of model organisms in human disease research. Nat Rev Genet 12:575–582.

Albers CA, Cvejic A, Favier R, Bouwmans EE, Alessi MC, Bertone P, Jordan G, Kettleborough RN, Kiddle G, Kostadima M, Read RJ, Sipos B, Sivapalaratnam S, Smethurst PA, Stephens J, Voss K, Nurden A, Rendon A, Nurden P, Ouwehand WH (2011) Exome sequencing identifies *NBEAL2* as the causative gene for gray platelet syndrome. Nat Genet 43:735–737.

Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang YL, Westerfield M, Ekker M, Postlethwait JH (1998) Zebrafish hox clusters and vertebrate genome evolution. Science 282:1711–1714.

Antoshechkin I, Sternberg PW (2007) The versatile worm: genetic and genomic resources for *Caenorhabditis elegans* research. Nat Rev Genet 8:518–532.

Arthur D, Levin E (2001) Spatial and non-spatial visual discrimination learning in zebrafish (*Danio rerio*). Anim Cogn 4:125–131.

Auer TO, Duroure K, De Cian A, Concordet JP, Del Bene F (2014) Highly efficient CRISPR/Cas9–mediated knock-in in zebrafish by homology-independent DNA repair. Genome Res 24:142–153.

Bai Q, Garver JA, Hukriede NA, Burton EA (2007) Generation of a transgenic zebrafish model of Tauopathy using a novel promoter element derived from the zebrafish *eno2* gene. Nucleic Acids Res 35:6501–6516.

Bandmann O, Burton EA (2010) Genetic zebrafish models of neurodegenerative diseases. Neurobiol Dis 40:58–65.

Bedell VM, Wang Y, Campbell JM, Poshusta TL, Starker CG, Krug II RG, Tan W, Penheiter SG, Ma AC, Leung AY, Fahrenkrug SC, Carlson DF, Voytas DF, Clark KJ, Essner JJ, Ekker SC (2012) *In vivo* genome editing using a high-efficiency TALEN system. Nature 491:114–118.

Best JDA, Alderton WK (2008) Zebrafish: an *in vivo* model for the study of neurological diseases. Neuropsychiatr Dis Treat 4:567–576.

Blackburn PR, Campbell JM, Clark KJ, Ekker SC (2013) The CRISPR system—keeping zebrafish gene targeting fresh. Zebrafish 10:116–118.

Boyadjiev SA, Fromme JC, Ben J, Chong SS, Nauta C, Hur DJ, Zhang G, Hamamoto S, Schekman R, Ravazzola M, Orci L, Eyaid W (2006) Cranio-lenticulo-sutural dysplasia is caused by a *SEC23A* mutation leading to abnormal endoplasmicreticulum-to-Golgi trafficking. Nat Genet 38:1192–1197.

Brand M, Heisenberg CP, Jiang YJ, Beuchle D, Lun K, Furutani-Seiki M, Granato M, Haffter P, Hammerschmidt M, Kane DA, Kelsh RN, Mullins MC, Odenthal J, van Eeden FJ, Nüsslein-Volhard C (1996) Mutations in zebrafish genes affecting the formation of the boundary between midbrain and hindbrain. Development 123:179–190.

Capecchi MR (2005) Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. Nat Rev Genet 6:507–512.

Castellana S, Mazza T (2013) Congruency in the prediction of pathogenic missense mutations: state-of-the-art Web-based tools. Brief Bioinform 14:448–459.

Chakrabarti S, Streisinger G, Singer F, Walker C (1983) Frequency of gamma-ray induced specific locus and recessive lethal mutations in mature germ cells of the zebrafish, *Brachydanio Rerio*. (1983) Genetics 103:109–123.

Chang N, Sun C, Gao L, Zhu D, Xu X, Zhu X, Xiong JW, Xi JJ (2013) Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. Cell Res 23:465–472.

Collins FS (1992) Positional cloning: Let's not call it reverse anymore. Nat Genet 1:3–6.

Daubert A, Golzio C, Guenot C, Jodelka FM, Kibaek M, Kjaergaard S, Leheup B, Martinet D, Nowaczyk MJ, Rosenfeld JA, Zeesman S, Zunich J, Beckmann JS, Hirschhorn JN, Hastings ML, Jacquemont S, Katsanis N (2013) *SCRIB* and *PUF60* are primary drivers of the multisystemic phenotypes of the 8q24.3 copy-number variant. Am J Hum Genet 93:798–811.

Davis EE, N. Katsanis (2012) The ciliopathies: a transitional model into systems biology of human genetic disease. Curr Opin Genet Dev 22:290–303.

© 2015 Katsanis

NOTES

Davis EE, Zhang Q, Liu Q, Diplas BH, Davey LM, Hartley J, Stoetzel C, Szymanska K, Ramaswami G, Logan CV, Muzny DM, Young AC, Wheeler DA, Cruz P, Morgan M, Lewis LR, Cherukuri P, Maskeri B, Hansen NF, Mullikin JC, et al. (2011) *TTC21B* contributes both causal and modifying alleles across the ciliopathy spectrum. Nat Genet 43:189–196.

de Pontual L, Zaghloul NA, Thomas S, Davis EE, McGaughey DM, Dollfus H, Baumann C, Bessling SL, Babarit C, Pelet A, Gascue C, Beales P, Munnich A, Lyonnet S, Etchevers H, Attie-Bitach T, Badano JL, McCallion AS, Katsanis N, Amiel J (2009) Epistasis between *RET* and *BBS* mutations modulates enteric innervation and causes syndromic Hirschsprung disease. Proc Natl Acad Sci USA 106:13921–13926.

Devoy A, Bunton-Stasyshyn RK, Tybulewicz VL, Smith AJ, Fisher EM (2012) Genomically humanized mice: technologies and promises. Nat Rev Genet 13:14–20.

Doyon Y, McCammon JM, Miller JC, Faraji F, Ngo C, Katibah GE, Amora R, Hocking TD, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Amacher SL (2008) Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. Nat Biotechnol 26:702–708.

Eisen JS, Smith JC (2008) Controlling morpholino experiments: Don't stop making antisense. Development 135:1735–1743.

Gieger C, Radhakrishnan A, Cvejic A, Tang W, Porcu E, Pistis G, Serbanovic-Canic J, Elling U, Goodall AH, Labrune Y, Lopez LM, Mägi R, Meacham S, Okada Y, Pirastu N, Sorice R, Teumer A, Voss K, Zhang W, Ramirez-Solis R, et al. (2011) New gene functions in megakaryopoiesis and platelet formation. Nature 480:201–208.

Golzio C, Willer J, Talkowski ME, Oh EC, Taniguchi Y, Jacquemont S, Reymond A, Sun M, Sawa A, Gusella JF, Kamiya A, Beckmann JS, Katsanis N (2012) *KCTD13* is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. Nature 485:363–367.

Golzio C, Katsanis N (2013) Genetic architecture of reciprocal CNVs. Curr Opin Genet Dev 23:240–248.

Hammerschmidt M, Pelegri F, Mullins MC, Kane DA, Brand M, van Eeden FJ, Furutani-Seiki M, Granato M, Haffter P, Heisenberg CP, Jiang YJ, Kelsh RN, Odenthal J, Warga RM, Nüsslein-Volhard C (1996) Mutations affecting morphogenesis during gastrulation and tail formation in the zebrafish, *Danio rerio*. Development 123:143–151.

Hjeij R, Lindstrand A, Francis R, Zariwala MA, Liu X, Li Y, Damerla R, Dougherty GW, Abouhamed M, Olbrich H, Loges NT, Pennekamp P, Davis EE, Carvalho CM, Pehlivan D, Werner C, Raidt J, Köhler G, Häffner K, Reyes-Mugica M, et al. (2013) *ARMC4* mutations cause primary ciliary dyskinesia with randomization of left/right body asymmetry. Am J Hum Genet 93:357–367.

Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L, McLaren S, Sealy I, Caccamo M, Churcher C, Scott C, Barrett JC, Koch R, Rauch GJ, White S, Chow W, et al. (2013a) The zebrafish reference genome sequence and its relationship to the human genome. Nature 496:498–503.

Howe DG, Bradford YM, Conlin T, Eagle AE, Fashena D, Frazer K, Knight J, Mani P, Martin R, Moxon SA, Paddock H, Pich C, Ramachandran S, Ruef BJ, Ruzicka L, Schaper K, Shao X, Singer A, Sprunger B, Van Slyke CE, Westerfield M (2013b) ZFIN, the Zebrafish Model Organism Database: increased support for mutants and transgenics. Nucleic Acids Res 41:D854–D860.

Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh JR, Joung JK (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol 31:227–229.

Inoue K, Lupski JR (2002) Molecular mechanisms for genomic disorders. Annu Rev Genomics Hum Genet 3:199–242.

Jao LE, Wente SR, Chen W (2013) Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. Proc Natl Acad Sci USA 110:13904–13909.

Jiang YJ, Brand M, Heisenberg CP, Beuchle D, Furutani-Seiki M, Kelsh RN, Warga RM, Granato M, Haffter P, Hammerschmidt M, Kane DA, Mullins MC, Odenthal J, van Eeden FJ, Nüsslein-Volhard C (1996) Mutations affecting neurogenesis and brain morphology in the zebrafish, *Danio rerio*. Development 123:205–216.

Katsanis SH, Katsanis N (2013) Molecular genetic testing and the future of clinical genomics. Nat Rev Genet 14:415–426.

© 2015 Katsanis

NOTES

Kawakami K, Shima A, Kawakami N (2000) Identification of a functional transposase of the *Tol2* element, an *Ac*-like element from the Japanese Medaka fish, and its transposition in the zebrafish germ lineage. Proc Natl Acad Sci USA 97:11403–11408.

Kettleborough RN, Busch-Nentwich EM, Harvey SA, Dooley CM, de Bruijn E, van Eeden F, Sealy I, White RJ, Herd C, Nijman IJ, Fényes F, Mehroke S, Scahill C, Gibbons R, Wali N, Carruthers S, Hall A, Yen J, Cuppen E, Stemple DL (2013) A systematic genome-wide analysis of zebrafish protein-coding gene function. Nature 496:494–497.

Khanna H, Davis EE, Murga-Zamalloa CA, Estrada-Cuzcano A, Lopez I, den Hollander AI, Zonneveld MN, Othman MI, Waseem N, Chakarova CF, Maubaret C, Diaz-Font A, MacDonald I, Muzny DM, Wheeler DA, Morgan M, Lewis LR, Logan CV, Tan PL, Beer MA (2009) A common allele in *RPGRIP1L* is a modifier of retinal degeneration in ciliopathies. Nat Genet 41:739–745.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921.

Lang MR, Lapierre LA, Frotscher M, Goldenring JR, Knapik EW (2006) Secretory COPII coat component Sec23a is essential for craniofacial chondrocyte maturation. Nat Genet 38:1198–1203.

Lawson ND, Wolfe SA (2011) Forward and reverse genetic approaches for the analysis of vertebrate development in the zebrafish. Dev Cell 21:48–64.

Lejeune J, Gautier M, Turpin R (1959) Study of somatic chromosomes from 9 mongoloid children. C R Hebd Seances Acad Sci 248:1721–1722.

Leshchiner I, Alexa K, Kelsey P, Adzhubei I, Austin-Tse CA, Cooney JD, Anderson H, King MJ, Stottmann RW, Ga\rnaas MK, Ha S, Drummond IA, Paw BH, North TE, Beier DR, Goessling W, Sunyaev SR (2012) Mutation mapping and identification by whole-genome sequencing, Genome Res 22:1541–1548.

Lieschke GJ, Currie PD (2007) Animal models of human disease: zebrafish swim into view. Nat Rev Genet 8:353–367.

Lindsay EA (2001) Chromosomal microdeletions: dissecting del22q11 syndrome. Nat Rev Genet 2:858–868.

Liu CT, Garnaas MK, Tin A, Kottgen A, Franceschini N, Peralta CA, de Boer IH, Lu X, Atkinson E, Ding J, Nalls M, Shriner D, Coresh J, Kutlar A, Bibbins-Domingo K, Siscovick D, Akylbekova E, Wyatt S, Astor B, Mychaleckjy J (2011) Genetic association for renal traits among participants of African ancestry reveals new loci for renal function. PLoS Genet 7:e1002264.

Maddison LA, Li M, Chen W (2014) Conditional gene-trap mutagenesis in zebrafish. Methods Mol Biol 1101:393–411.

McClellan J, King MC (2010) Genetic heterogeneity in human disease. Cell 141:210–217.

Meng X, Noyes MB, Zhu LJ, Lawson ND, Wolfe SA (2008) Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. Nat Biotechnol 26:695–701.

Merveille AC, Davis EE, Becker-Heck A, Legendre M, Amirav I, Bataille G, Belmont J, Beydon N, Billen F, Clément A, Clercx C, Coste A, Crosbie R, de Blic J, Deleuze S, Duquesnoy P, Escalier D, Escudier E, Fliegauf M, Horvath J (2011) CCDC39 is required for assembly of inner dynein arms and the dynein regulatory complex and for normal ciliary motility in humans and dogs. Nat Genet 43:72–78.

Meyer A, Schartl M (1999) Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. Curr Opin Cell Biol 11:699–704.

Mouse Genome Sequencing Consortium, Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, et al. (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420:520–562.

Mullins MC, Hammerschmidt M, Haffter P, Nusslein-Volhard C (1994) Large-scale mutagenesis in the zebrafish: in search of genes controlling development in a vertebrate. Curr Biol 4:189–202.

Nasevicius A, Ekker SC (2000) Effective targeted gene 'knockdown' in zebrafish. Nat Genet 26:216–220.

© 2015 Katsanis

NOTES

Neuhauss SC, Solnica-Krezel L, Schier AF, Zwartkruis F, Stemple DL, Malicki J, Abdelilah S, Stainier DY, Driever W (1996) Mutations affecting craniofacial development in zebrafish, Development 123:357–367.

Niederriter AR, Davis EE, Golzio C, Oh EC, Tsai IC, Katsanis N (2013) *In vivo* modeling of the morbid human genome using *Danio rerio*. J Vis Exp (78):e50338.

Norton N, Li D, Rieder MJ, Siegfried JD, Rampersaud E, Züchner S, Mangos S, Gonzalez-Quintana J, Wang L, McGee S, Reiser J, Martin E, Nickerson DA, Hershberger RE (2011) Genome-wide studies of copy number variation and exome sequencing identify rare variants in *BAG3* as a cause of dilated cardiomyopathy. Am J Hum Genet 88:273–282.

Obholzer N, Swinburne IA, Schwab E, Nechiporuk AV, Nicolson T, Megason SG (2012) Rapid positional cloning of zebrafish mutations by linkage and homozygosity mapping using whole-genome sequencing, Development 139:4280–4290.

Oprea GE, Kröber S, McWhorter ML, Rossoll W, Müller S, Krawczak M, Bassell GJ, Beattie CE, Wirth B (2008) Plastin 3 is a protective modifier of autosomal recessive spinal muscular atrophy. Science 320:524–527.

Padang R, Bagnall RD, Richmond DR, Bannon PG, Semsarian C (2012) Rare nonsynonymous variations in the transcriptional activation domains of *GATA5* in bicuspid aortic valve disease. J Mol Cell Cardiol 53:277–281.

Pardo-Martin C, Allalou A, Medina J, Eimon PM, Wählby C, Fatih Yanik M (2013) High-throughput hyperdimensional vertebrate phenotyping. Nat Commun 4:1467.

Piotrowski T, Schilling TF, Brand M, Jiang YJ, Heisenberg CP, Beuchle D, Grandel H, van Eeden FJ, Furutani-Seiki M, Granato M, Haffter P, Hammerschmidt M, Kane DA, Kelsh RN, Mullins MC, Odenthal J, Warga RM, Nüsslein-Volhard C (1996) Jaw and branchial arch mutants in zebrafish II: anterior arches and cartilage differentiation. Development 123:345–356.

Postlethwait JH, Yan YL, Gates MA, Horne S, Amores A, Brownlie A, Donovan A, Egan ES, Force A, Gong Z, Goutel C, Fritz A, Kelsh R, Knapik E, Liao E, Paw B, Ransom D, Singer A, Thomson M, Abduljabbar TS, et al. (1998) Vertebrate genome evolution and the zebrafish gene map. Nat Genet 18:345–349.

Ragvin A, Moro E, Fredman D, Navratilova P, Drivenes Ø, Engström PG, Alonso ME, de la Calle Mustienes E, Gómez Skarmeta JL, Tavares MJ, Casares F, Manzanares M, van Heyningen V, Molven A, Njølstad PR, Argenton F, Lenhard B, Becker TS (2010) Long-range gene regulation links genomic type 2 diabetes and obesity risk regions to *HHEX, SOX4*, and *IRX3*. Proc Natl Acad Sci USA 107:775–780.

Ramachandran KV, Hennessey JA, Barnett AS, Yin X, Stadt HA, Foster E, Shah RA, Yazawa M, Dolmetsch RE, Kirby ML, Pitt GS (2013) Calcium influx through L-type $Ca_V1.2$ $Ca2^+$ channels regulates mandibular development. J Clin Invest 123:1638–1646.

Reiter JF, Alexander J, Rodaway A, Yelon D, Patient R, Holder N, Stainier DY (1999) $Gata_5$ is required for the development of the heart and endoderm in zebrafish. Genes Dev 13:2983–2995.

Ryan S, Willer J, Marjoram L, Bagwell J, Mankiewicz J, Leshchiner I, Goessling W, Bagnat M, Katsanis N (2013) Rapid identification of kidney cyst mutations by whole exome sequencing in zebrafish. Development 140:4445–4451.

Sanna-Cherchi S, Sampogna RV, Papeta N, Burgess KE, Nees SN, Perry BJ, Choi M, Bodria M, Liu Y, Weng PL, Lozanovski VJ, Verbitsky M, Lugani F, Sterken R, Paragas N, Caridi G, Carrea A, Dagnino M, Materna-Kiryluk A, Santamaria G, et al. (2013) Mutations in *DSTYK* and dominant urinary tract malformations. N Engl J Med 369:621–629.

Sarparanta J, Jonson PH, Golzio C, Sandell S, Luque H, Screen M, McDonald K, Stajich JM, Mahjneh I, Vihola A, Raheem O, Penttilä S, Lehtinen S, Huovinen S, Palmio J, Tasca G, Ricci E, Hackman P, Hauser M, Katsanis N, et al. (2012) Mutations affecting the cytoplasmic functions of the co-chaperone DNAJB6 cause limb-girdle muscular dystrophy. Nat Genet 44:450–455.

Schier AF, Neuhauss SC, Harvey M, Malicki J, Solnica-Krezel L, Stainier DY, Zwartkruis F, Abdelilah S, Stemple DL, Rangini Z, Yang H, Driever W (1996) Mutations affecting the development of the embryonic zebrafish brain. Development 123:165–178.

© 2015 Katsanis

NOTES

Schilling TF, Piotrowski T, Grandel H, Brand M, Heisenberg CP, Jiang YJ, Beuchle D, Hammerschmidt M, Kane DA, Mullins MC, van Eeden FJ, Kelsh RN, Furutani-Seiki M, Granato M, Haffter P, Odenthal J, Warga RM, Trowe T, Nüsslein-Volhard C (1996) Jaw and branchial arch mutants in zebrafish I: branchial arches. Development 123:329–344.

Shestopalov IA, Sinha S, Chen JK (2007) Light-controlled gene silencing in zebrafish embryos. Nat Chem Biol 3:650–651.

Sivasubbu S, Balciunas D, Amsterdam A, Ekker SC (2007) Insertional mutagenesis strategies in zebrafish. Genome Biol 8(Suppl. 1):S9.

Solnica-Krezel L, Schier AF, Driever W (1994) Efficient recovery of ENU-induced mutations from the zebrafish germline. Genetics 136:1401–1420.

Solnica-Krezel L, Stemple DL, Mountcastle-Shah E, Rangini Z, Neuhauss SC, Malicki J, Schier AF, Stainier DY, Zwartkruis F, Abdelilah S, Driever W (1996) Mutations affecting cell fates and cellular rearrangements during gastrulation in zebrafish. Development 123:67–80.

Stainier DY, Fouquet B, Chen JN, Warren KS, Weinstein BM, Meiler SE, Mohideen MA, Neuhauss SC, Solnica-Krezel L, Schier AF, Zwartkruis F, Stemple DL, Malicki J, Driever W, Fishman MC (1996) Mutations affecting the formation and function of the cardiovascular system in the zebrafish embryo. Development 123:285–292.

Stankiewicz P, Lupski JR (2010) Structural variation in the human genome and its role in disease. Annu Rev Med 61:437–455.

Streisinger G (1983) Extrapolations from species to species and from various cell types in assessing risks from chemical mutagens. Mutat Res 114:93–105.

Streisinger G, Walker C, Dower N, Knauber D, Singer F (1981) Production of clones of homozygous diploid zebra fish (*Brachydanio rerio*). Nature 291:293–296.

Summerton J, Weller D (1997) Morpholino antisense oligomers: design, preparation, and properties. Antisense Nucleic Acid Drug Dev 7:187–195.

Tan JL, Zon LI (2011) Chemical screening in zebrafish for novel biological and therapeutic discovery. Methods Cell Biol 105:493–516.

Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, McGee S, Do R, Liu X, Jun G, Kang HM, Jordan D, Leal SM, Gabriel S, Rieder MJ, Abecasis G, Altshuler D, Nickerson DA, Boerwinkle E, Sunyaev S (2012) Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science 337:64–69.

Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD (2010) Genome editing with engineered zinc finger nucleases. Nat Rev Genet 11:636–646.

van de Ven JP, Nilsson SC, Tan PL, Buitendijk GH, Ristau T, Mohlin FC, Nabuurs SB, Schoenmaker-Koller FE, Smailhodzic D, Campochiaro PA, Zack DJ, Duvvari MR, Bakker B, Paun CC, Boon CJ, Uitterlinden AG, Liakopoulos S, Klevering BJ, Fauser S, Daha MR, et al. (2013) A functional variant in the *CFI* gene confers a high risk of agerelated macular degeneration. Nat Genet 45:813–817.

van Eeden FJ, Granato M, Schach U, Brand M, Furutani-Seiki M, Haffter P, Hammerschmidt M, Heisenberg CP, Jiang YJ, Kane DA, Kelsh RN, Mullins MC, Odenthal J, Warga RM, Allende ML, Weinberg ES, Nüsslein-Volhard C (1996) Mutations affecting somite formation and patterning in the zebrafish, *Danio rerio*. Development 123:153–164.

Veltman JA, Brunner HG (2012) *De novo* mutations in human genetic disease. Nat Rev Genet 13:565–575.

Wan J, Yourshaw M, Mamsa H, Rudnik-Schöneborn S, Menezes MP, Hong JE, Leong DW, Senderek J, Salman MS, Chitayat D, Seeman P, von Moers A, Graul-Neumann L, Kornberg AJ, Castro-Gago M, Sobrido MJ, Sanefuji M, Shieh PB, Salamon N, Kim RC, (2012) Mutations in the RNA exosome component gene *EXOSC3* cause pontocerebellar hypoplasia and spinal motor neuron degeneration. Nat Genet 44:704–708.

Wang D, Jao LE, Zheng N, Dolan K, Ivey J, Zonies S, Wu X, Wu K, Yang H, Meng Q, Zhu Z, Zhang B, Lin S, Burgess SM (2007) Efficient genome-wide mutagenesis of zebrafish genes by retroviral insertions. Proc Natl Acad Sci USA 104:12428–12433.

Wei D, Bao H, Liu XY, Zhou N, Q Wang, Li RH, Xu YJ, Yang YQ (2013a) $GATA_5$ loss-of-function mutations underlie tetralogy of Fallot. Int J Med Sci 10:34–42.

© 2015 Katsanis

NOTES

Wei D, Bao H, Zhou N, Zheng GF, Liu XY, Yang YQ (2013b) *GATA*$_5$ loss-of-function mutation responsible for the congenital ventriculoseptal defect. Pediatr Cardiol 34:504–511.

Wienholds E, Schulte-Merker S, Walderich B, Plasterk RH (2002) Target-selected inactivation of the zebrafish *rag1* gene. Science 297:99–102.

Wijshake T, Baker DJ, van de Sluis B (2014) Endonucleases: new tools to edit the mouse genome. Biochim Biophys Acta 1842:1942–1950.

Wullimann MF, Mueller T (2004) Teleostean and mammalian forebrains contrasted: evidence from genes to behavior. J Comp Neurol 475:143–162.

Zaghloul NA, Liu Y, Gerdes JM, Gascue C, Oh EC, Leitch CC, Bromberg Y, Binkley J, Leibel RL, Sidow A, Badano JL, Katsanis N (2010) Functional analyses of variants reveal a significant role for dominant negative and common alleles in oligogenic Bardet–Biedl syndrome. Proc Natl Acad Sci USA 107:10602–10607.

Zariwala MA, Gee HY, Kurkowiak M, Al-Mutairi DA, Leigh MW, Hurd TW, Hjeij R, Dell SD, Chaki M, Dougherty GW, Adan M, Spear PC, Esteve-Rudd J, Loges NT, Rosenfeld M, Diaz KA, Olbrich H, Wolf WE, Sheridan E, Batten TF (2013) *ZMYND10* is mutated in primary ciliary dyskinesia and interacts with LRRC6. Am J Hum Genet 93:336–345.

Zon LI, Peterson RT (2005) *In vivo* drug discovery in the zebrafish. Nat Rev Drug Discov 4:35–44.

© 2015 Katsanis