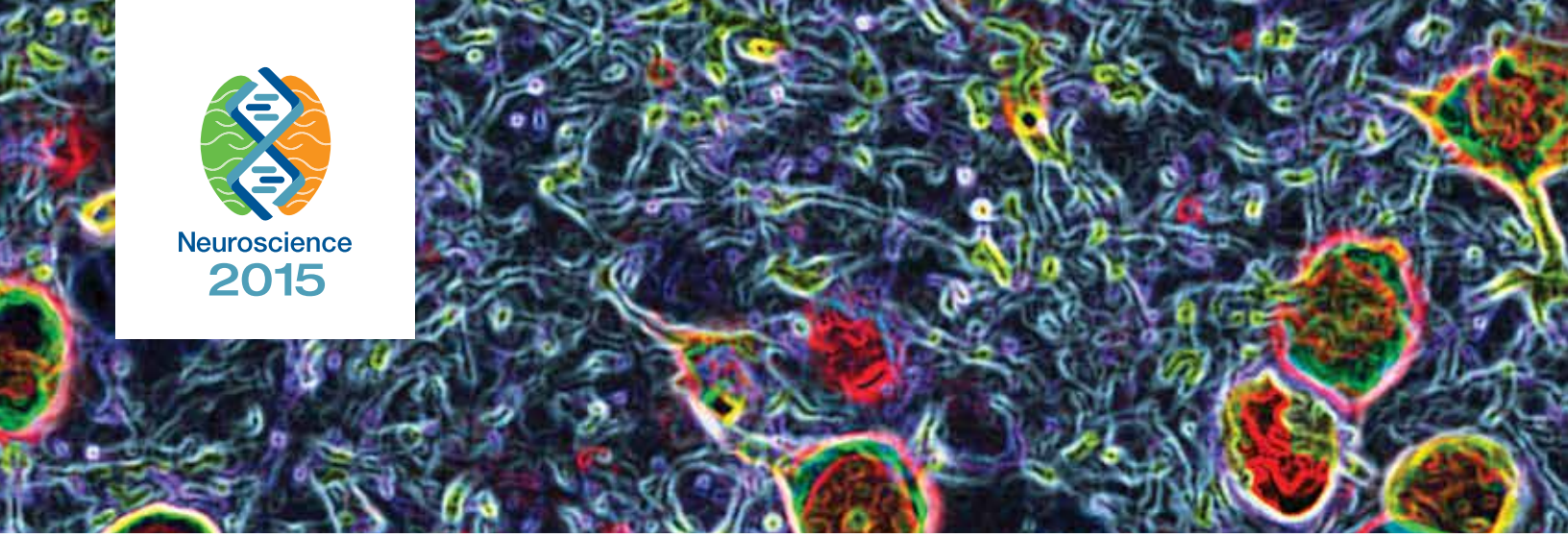
Neuroscience 2015

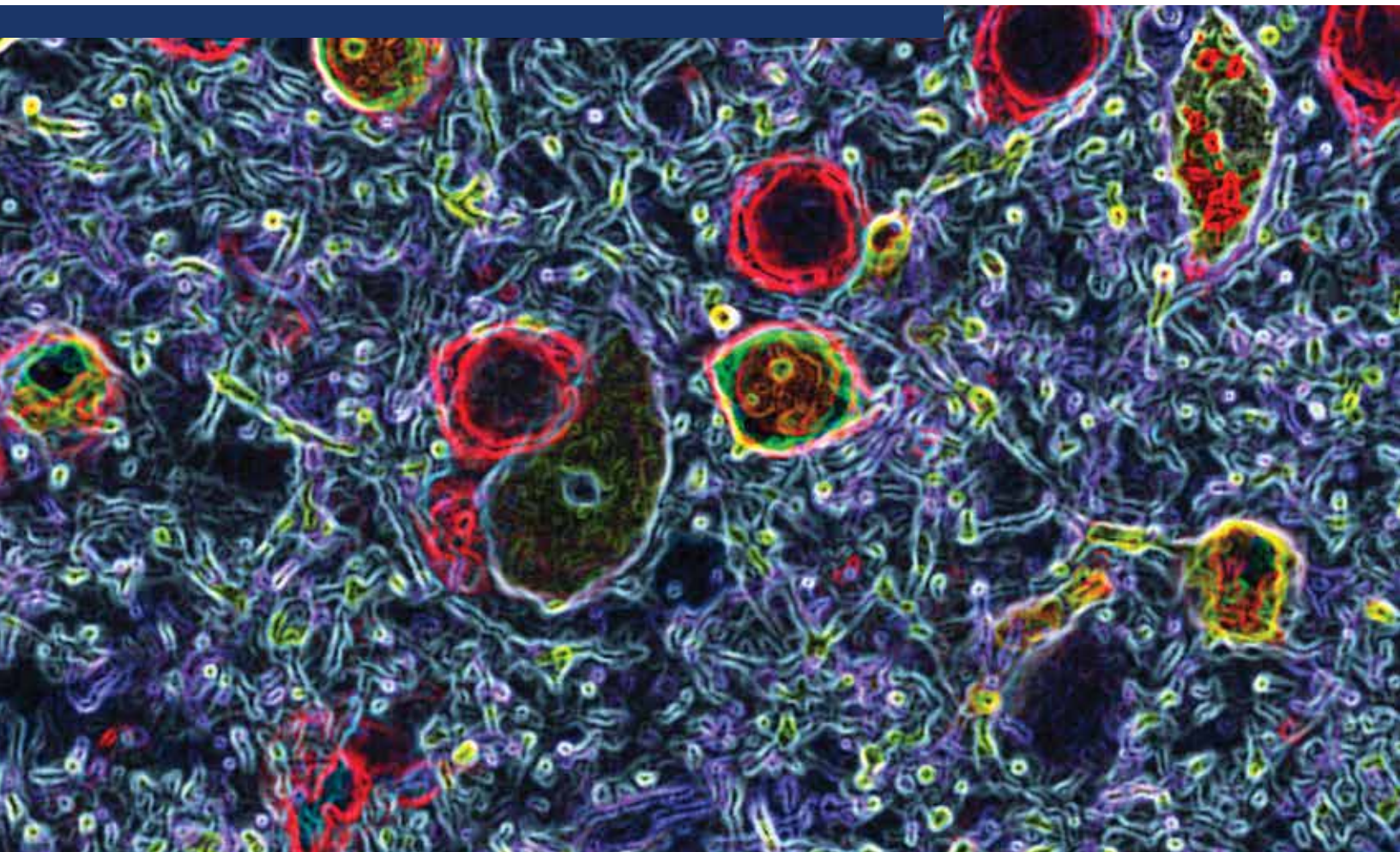


SHORT COURSE II

The Impact of Human Genetics and Genomics in Neurobiology

Organized by Nicholas Katsanis, PhD

SfN Society for Neuroscience



Short Course II

The Impact of Human Genetics and Genomics in Neurobiology

Organized by Nicholas Katsanis, PhD

SOCIETY *for* NEUROSCIENCE

Please cite articles using the model:
[AUTHOR'S LAST NAME, AUTHOR'S FIRST & MIDDLE INITIALS] (2015)
[CHAPTER TITLE] In: The Impact of Human Genetics and Genomics in Neurobiology.
(Katsanis N, ed) pp. [xx-xx]. Chicago, IL: Society for Neuroscience.

Neuroscience 2015

SHORT COURSE 2

The Impact of Human Genetics and Genomics in Neurobiology: From Disease Discovery to Fundamental Mechanisms (and Back)

Organized by: Nicholas Katsanis, PhD
Friday, October 16, 2015
8:30 a.m.–6 p.m.
Location: McCormick Place | Room: S406A | Chicago, IL

TIME	TALK TITLE	SPEAKER	
8–8:30 a.m.	**Check-In**		
8:30–8:40 a.m.	Opening Remarks	Nicholas Katsanis, PhD | Duke University	
8:40–9:30 a.m.	The Role of the MHC in Schizophrenia	Steven McCarroll, PhD | Harvard Medical School	
9:30–10:20 a.m.	Leveraging Genome Data in Psychiatric Illness	Benjamin Neale, PhD | Broad Institute	
10:20–10:50 a.m.	**Morning Break**		
10:50–11:40 a.m.	Computational Approaches to Annotating Neurodevelopment and Neurodegenerative Disease Genomes	Shamil Sunyaev, PhD | Harvard Medical School	
11:40–12:30 p.m.	Genetics-Based Therapeutics in Huntington's Disease	Albert La Spada, MD, PhD | University of California, San Diego	
12:30–1:30 p.m.	**Lunch**		
1:30–2:20 p.m.	Genetics of Alzheimer's Disease	Alison Goate, DPhil | Icahn School of Medicine at Mount Sinai	
2:20–3:10 p.m.	Functional Dissection of Neurodevelopmental Disorders	Nicolas Katsanis, PhD | Duke University	
3:10–4 p.m.	Summary, Discussion, Breakout Guide		
4–4:15 p.m.	**Afternoon Break**		

AFTERNOON BREAKOUT SESSIONS I PARTICIPANTS SELECT DISCUSSION GROUPS AT 4:15 P.M. AND 5:15 P.M.

TIME	BREAKOUT SESSIONS	SPEAKERS	ROOM
4:15–5 p.m.	Group 1: Using ExAC and other population-based databases to assess exome data	Benjamin Neale, PhD | Broad Institute	S403B
	Group 2: Designing assays in zebrafish: considerations for feasibility, throughput and accuracy	Erica Davis, PhD | Duke University	S405A
	Group 3: The Special Case of Copy Number Variants in Neuropsychiatric Disorders	Nicholas Katsanis, PhD | Duke University	S405B
5–5:15 p.m.	**Afternoon Break**		
5:15–6 p.m.	**Repeat sessions above. Select a second discussion group.**		

Table of Contents

Introduction

Medical genetics and genomics are at an inflection point. They are being driven in part by the evolution of high-throughput sequencing technologies, and in part by the needs and desires of diverse research and clinical communities to coalesce in order to solve previously intractable problems. In the realm of rare genetic disorders, the advent of whole exome/whole genome sequencing at a cost that is accessible to individual investigators has precipitated a torrent of disease-gene discoveries. This work has also highlighted the urgency of improving the ability to interpret genomic data and the acute necessity of sharing data and tools across the globe. Similarly, for complex genetic disorders, the need to power studies to dissect the genetic causality of polygenic traits has become a major driving force behind the aggregation of multinational consortia and the sharing of inexpensively produced genotype data from thousands of individuals.

Neurodevelopmental and neurodegenerative disorders have been at the forefront of both genetics and genomics research. From the identification of hundreds of genes mutated in rare neurodevelopmental disorders to what is arguably the first genome-wide association for a neurodegenerative trait (the loss of photoreceptors in age-related macular degeneration), the impact of deleterious genetic and genomic variation on the development and maintenance of the nervous system has been measured extensively. In this course, we will focus on the lessons learned from these experiences, discuss emergent trends, and formulate the next grand challenges that will likely occupy the interface between state-of-the-art genetics and cellular and behavioral neuroscience.

At the core of this course, we will use clinical exemplars that include major psychiatric conditions such as schizophrenia, Alzheimer disease, and Huntington disease to highlight how the journey from genomics to disease mechanism(s) to potential drug discovery is still a challenging one. We will explore how computational and biological assays can be deployed to answer questions about genetic causality and pathomechanism, and we will discuss what the tools of the future need to look like to serve the needs of both the clinical and the research community. In parallel, through dedicated afternoon sessions, we will go into greater depth about the strengths and limitations of current population-based methods for dissecting pathomechanisms. We will also review the state of the art in the deployment of functional assays for establishing the physiological relevance of candidate loci to disease pathology, and *in vivo* complementation assays for determining variant pathogenicity. Lastly, we will discuss in greater depth the functional dissection of larger genomic lesions, copy number variants, which are now understood to represent a significant, potent driver of neurodevelopmental traits.

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

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.

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).

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

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).

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

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

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,

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

NOTES

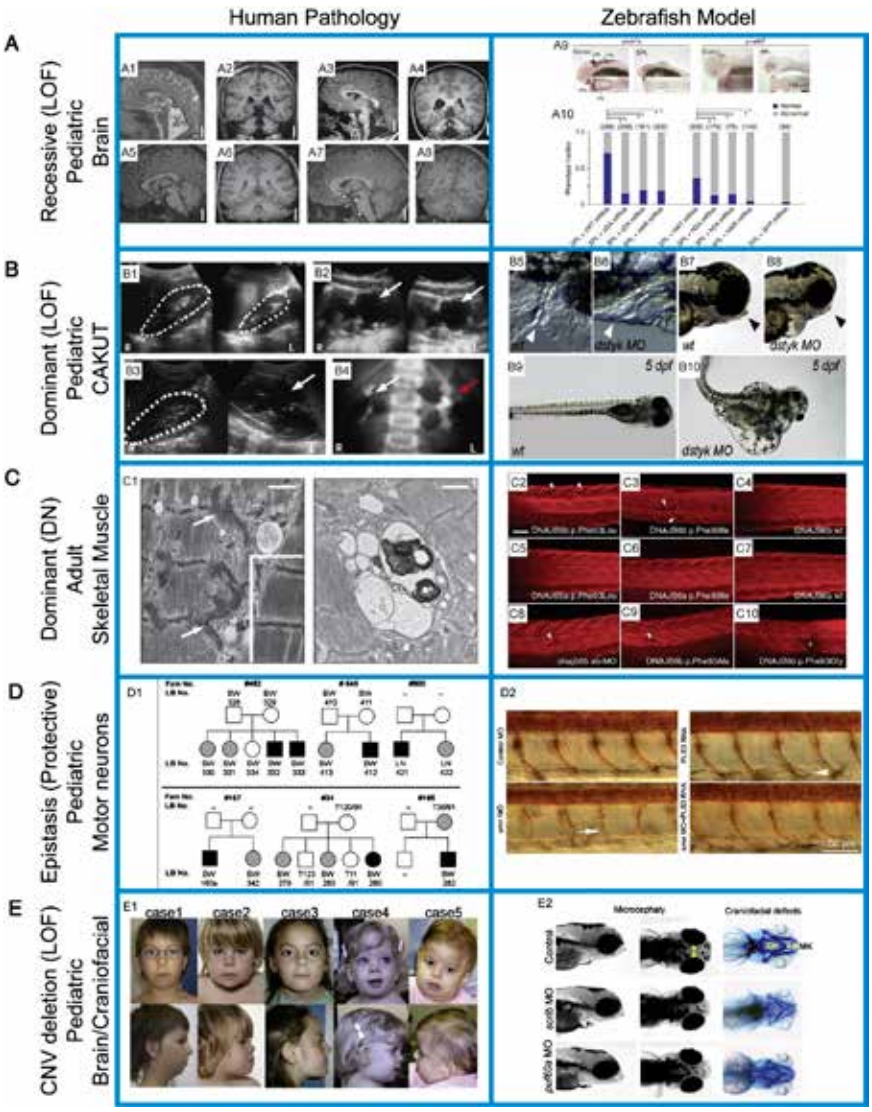


Figure 1. Zebrafish models of human genetic disease. *A*, Mutations in pontocerebellar hypoplasia caused by *EXOSC3* and *in vivo* complementation studies in zebrafish recapitulate the brain phenotypes observed in patients and demonstrate that missense mutations are functional null variants. Left panel, 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.

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. 1C) (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

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

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.

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.

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.

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.

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.

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.

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.

Alzheimer's Disease Genetics: From the Bench to the Clinic

Celeste M. Karch, PhD,[1] Carlos Cruchaga, PhD,[1] and Alison Goate, DPhil[2]

[1]Department of Psychiatry and Hope Center for Neurological Disorders
Washington University School of Medicine
St. Louis, Missouri

[2]Department of Neuroscience
Icahn School of Medicine at Mount Sinai
New York, New York

NOTES

Introduction

Alzheimer's disease (AD) is defined clinically by a gradual decline in memory and other cognitive functions, and neuropathologically by gross atrophy of the brain and the accumulation of extracellular amyloid plaques and intracellular neurofibrillary tangles. Genetic, biochemical, and neuropathological data suggest that aggregation of β-amyloid (Aβ) is central to initiating AD pathogenesis (Hardy and Selkoe, 2002). Aβ is a proteolytic fragment of the amyloid precursor protein (APP), generated as a result of sequential cleavage by β- and γ-secretases. Deposition of extracellular amyloid plaques is followed by the accumulation of neurofibrillary tangles in neuronal cell bodies and associated processes. Neurofibrillary tangles are composed of hyperphosphorylated tau aggregates. The presence of neurofibrillary tangles in AD brains is strongly correlated with neuronal dysfunction and disease progression (Holtzman et al., 2011). The amyloid cascade hypothesis posits that changes in APP and/or Aβ homeostasis lead to the aggregation of Aβ and deposition in plaques and that these events are sufficient to initiate the cascade of pathological and clinical changes associated with AD, including the aggregation of tau protein in neurofibrillary tangles (Hardy, 1997).

Linkage Studies

Dominantly inherited mutations in *β-amyloid precursor protein (APP)*, *presenilin 1 (PSEN1)*, and *presenilin 2 (PSEN2)* cause early-onset AD. These genes, as well as *APOE*, were identified through genetic linkage studies in families.

APP

APP encodes a ubiquitously expressed type 1 transmembrane protein. The majority of APP is proteolyzed by α- and γ-secretases, leading to cleavage of APP within the Aβ domain. The result is nonpathogenic fragments: sAPPα and α-C-terminal fragment (CTF). Alternatively, APP can be cleaved through sequential proteolytic cleavage by β- and γ-secretases to generate Aβ peptides: sAPPβ, and β-CTF. Cell surface APP is internalized, allowing Aβ to be generated in the endocytic pathway and secreted into the extracellular space. Dominant mutations in *APP* account for approximately 14% of early-onset autosomal dominant cases of AD, and more than 30 mutations have been described (Alzheimer Disease & Frontotemporal Dementia Mutation Database, n.d.). Two recessive *APP* mutations, A673V and E693Δ, also reportedly cause early-onset AD. The majority of mutations in APP cluster in the region that is adjacent to or within the Aβ domain.

Mutations in *APP* have revealed many important aspects of the molecular mechanisms underlying AD pathogenesis. The Swedish mutation (KM670/671NL) increases plasma Aβ levels by twofold to threefold by altering β-secretase cleavage efficiency. Duplications of *APP* and the surrounding sequence are also associated with early-onset AD. Families carrying these duplications exhibit classic AD neuropathology and cerebral amyloid angiopathy. Additionally, individuals with Down syndrome, which results from trisomy of chromosome 21, develop AD neuropathology. Individuals with partial trisomy of chromosome 21, which does not include the *APP* gene, fail to develop AD neuropathology. Thus, excess Aβ production is sufficient to cause AD. Several *APP* mutations cluster at or after the C-terminal portion of the Aβ domain. These mutations alter γ-secretase function, leading to a shift in APP processing that increases the highly amyloidogenic Aβ42 fragment while reducing the Aβ40 fragment. The result is an altered Aβ42/Aβ40 ratio without a change in total Aβ levels (Bergmans and De Strooper, 2010). Because Aβ42 is more prone to aggregate than Aβ40, these findings suggest that Aβ aggregation is a critical component of AD pathogenesis. In contrast, mutations such as the Arctic mutation (E693G) and the Dutch mutation (E693Q) that occur within the Aβ domain likely increase the aggregation rate of the mutant peptide. Individuals carrying these mutations develop hereditary cerebral hemorrhage with amyloidosis, which is characterized by predominant vascular Aβ deposition with diffuse plaques in the parenchymal tissue. These mutations provide further evidence that Aβ aggregation is a critical process in AD pathogenesis. Genetic changes that lead to altered APP processing and Aβ accumulation may produce variable neurological and neurovascular phenotypes.

PSEN1 and *PSEN2*

PSEN1 and *PSEN2* are critical components of the γ-secretase complex, which cleaves APP into Aβ fragments, and localize in the endoplasmic reticulum and Golgi apparatus. As many as 185 dominant, pathogenic mutations have been identified in *PSEN1*, accounting for approximately 80% of autosomal dominant AD (ADAD) cases; 13 pathogenic mutations have been identified in *PSEN2*, accounting for approximately 5% of ADAD cases (Alzheimer Disease & Frontotemporal

NOTES

Dementia Mutation Database, n.d.). *PSEN1* and *PSEN2* mutations are distributed throughout the protein, with some clustering occurring in the transmembrane domains (Guerreiro et al., 2010). Dominantly inherited mutations in *PSEN1*, as well as *APP*, have also been identified in late-onset AD cases with a strong family history of disease. These families may carry additional genetic variants that delay age at onset of the normally fully penetrant disease mutation. Mutations in *PSEN1* and *PSEN2* alter the proteolytic site preference of γ-secretase. Whereas wild-type γ-secretase cleaves APP to generate more Aβ40 than Aβ42, mutations in the γ-secretase components lead to enhanced production of the Aβ42 fragment. In turn, because Aβ42 is highly aggregation-prone, the increase in the Aβ42 fragment is likely sufficient to cause AD.

Apolipoprotein E

Apolipoprotein E (APOE) encodes a pleiotropic glycoprotein that is highly expressed in liver, brain, and macrophages, where it plays a role in mobilization and redistribution of cholesterol. APOE has also been implicated in neuronal growth and repair, nerve regeneration, immune response, and activation of lipolytic enzymes (Mahley and Rall, 2000). *APOE* occurs as three isoforms that differ at two amino acid residues (112 and 158): *APOEε2*, *APOEε3*, and *APOEε4*. APOEε3 is the most common *APOE* isoform, occurring in approximately 72% of the population. Family-based methods originally identified a genetic linkage between AD and the region of chromosome 19 that contains the *APOE* gene. *APOEε4* increases risk in familial and sporadic early-onset and late-onset AD, increasing risk threefold for heterozygous carriers and increasing risk 8-fold to 10-fold for homozygous carriers (Farrer et al., 1997). *APOEε4* also has a dose-dependent effect on age at onset. Interestingly, *APOEε2* decreases the risk for late-onset AD and delays age at onset (Corder et al., 1994). APOE binds to Aβ, influencing the clearance of soluble Aβ and Aβ aggregation (Castellano et al., 2011). APOE4 binds to Aβ more rapidly than does APOE3, resulting in accelerated fibril formation. APOE also regulates Aβ metabolism indirectly by interacting with low-density lipoprotein receptor-related protein 1 receptors. *In vivo*, APOE influences the amount and structure of intraparenchymal Aβ deposits in an isoform-dependent manner. Thus, the major risk gene associated with AD likely influences Aβ metabolism as a mechanism of pathogenicity.

Genome-Wide Association Studies

Only 50% of individuals with AD carry an *APOEε4* allele, and only approximately 2% carry a pathogenic mutation, suggesting that other genetic factors must contribute to risk for the disease. The first genome-wide association study (GWAS) for AD to use thousands of AD cases and elderly nondemented controls successfully generated replicable associations for several new genetic risk factors, including *clusterin (CLU)*, *phosphatidylinositol-binding clathrin assembly protein (PICALM)*, *complement receptor 1 (CR1)*, *and bridging integrator protein 1 (BIN1)* (Harold et al., 2009; Lambert et al., 2009; Seshadri et al., 2010). Two subsequent studies that each included more than 8000 cases and a similar number of controls identified four additional loci with genome-wide significant evidence for association ($<5 \times 10^{-8}$) (Hollingworth et al., 2011; Naj et al., 2011) (Fig. 1).

Recently, a meta-analysis of GWAS data from 74,046 individuals from four large consortia confirmed these associations and reported 11 new susceptibility loci for AD (Lambert et al., 2013) (Fig. 1). For most of these loci, the specific functional variants and genes remain to be identified. These studies employed samples of individuals of European descent, but two additional GWASs examined African American and Asian (Korean and Japanese) case-control series. In addition to *APOE*, the most significant finding in the African-American study of 5896 cases and controls was of single nucleotide polymorphisms (SNPs) in *ABCA7*, which reached genome-wide significance and exhibited a somewhat larger effect than in European samples (Reitz et al., 2013). The study in Japanese and Korean subjects used a multistage strategy to identify a novel locus, *SORL1*, which was replicated in a large European American cohort (Lambert et al., 2013; Miyashita et al., 2013). Together, these studies demonstrate that some of these loci show similar effects across populations (e.g., *BIN1*), while others appear to have a bigger impact in some populations (e.g., *ABCA7*). As larger datasets become available for other populations, a more complete picture of the population-specific loci and those that are shared across populations will more fully resolve.

The common SNPs identified in these GWASs alter risk by 10–15%, suggesting that the effect of these risk alleles is much smaller than that of *APOEε4*, unless they are tagging rare alleles of larger effect. For the most part, the functional alleles responsible for each of these associations remain to be determined.

NOTES

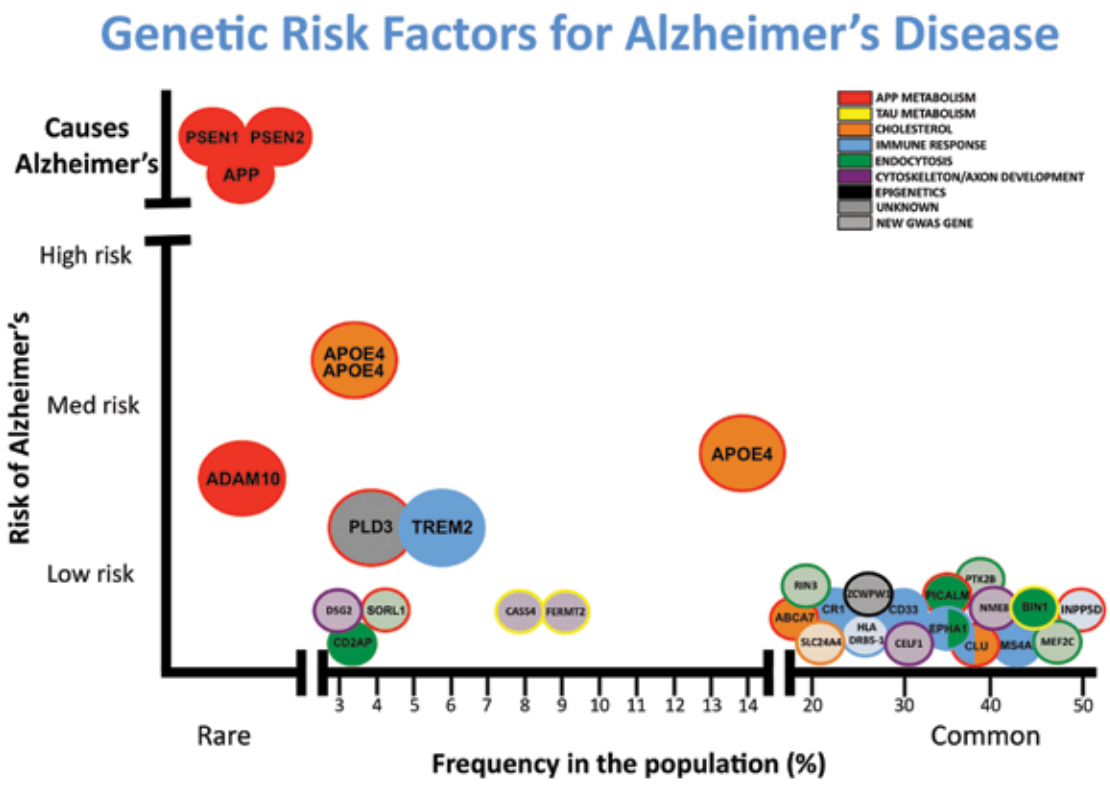


Figure 1. Cell-type expression of AD risk genes may influence AD pathogenesis. The major cell types present in the brain are depicted. Genes are listed in the cell type in which they are most highly expressed. Note: GWAS loci that contain multiple genes were excluded from this figure because it remains unclear which gene is responsible for the signal (*CR1, BIN1, PICALM, HLA, SLC24A4, DSG2, ZCWPW1, CELF1, FERMT2, or CASS4*). Source for genetic information: http://web.stanford.edu/group/barres_lab/brain_rnaseq.html.

For some of the GWAS loci, several genes are found within the associated region, any one of which may contain the functional risk variant. However, many of these AD risk loci have putative functions in the immune system *(CLU, CR1, ABCA7, CD33,* and *EPHA1)*; four are involved in processes at the cell membrane, including endocytosis (*PICALM, BIN1, CD33,* and *CD2AP)*; and three are involved in lipid biology *(APOE, CLU,* and *ABCA7)* (Fig. 1). Although some of the new genes appear to be involved in Aβ metabolism (e.g., *CLU* and *PICALM*), the fact that others likely influence inflammation, endocytosis, and lipid biology suggests that targeting specific components of these pathways may lead to novel directions for drug discovery and treatment.

Sequencing Studies

The most recent GWAS for late-onset AD combined several large datasets and identified more than 20 loci associated with AD risk. However, each of these new loci account for only a small percentage of the genetic variance that contributes to AD susceptibility. Thus, a large proportion of the genetic heritability for AD has not been explained by GWASs. It has been hypothesized that low-frequency (minor allele frequency [MAF] 1–5%) and rare variants (MAF < 1%) might explain the missing genetic heritability. As was the case with linkage studies and GWASs, initial sequencing efforts focused on candidate genes in small populations until the arrival of next-generation sequencing technology, which allowed for low-cost, high-throughput sequencing and ushered in the era of whole-exome and whole-genome sequencing. Even with these new advances, the first studies using this technology also focused on the pathogenic AD genes. Subsequent genome-wide studies have led to the identification of novel variants and genes.

One powerful approach to identifying AD risk variants with large effect sizes has been to study families with a strong history of late-onset AD. Studies sponsored

NOTES

by the National Institutes of Mental Health and the National Institute on Aging have invested in the collection and characterization of large family datasets with at least two late-onset AD cases. Large sequencing studies in these cohorts have shown that mutations in *APP, PSEN1*, and *PSEN2* also cause AD in 1.5–2% of these cases (Cruchaga et al., 2012). In addition to identifying known and likely pathogenic mutations in *APP* and *PSEN1*, these studies have revealed variants that may modify AD risk. For example, the *PSEN2* variants R62H and R71W were found to be associated with a lower age at onset, between 5 and 7 years earlier than in noncarriers (Benitez et al., 2013). *PSEN1 E318G* (rs17125721) was previously classified as nonpathogenic because it does not segregate according to disease status in some families, but analyses in large case-control series have demonstrated that in *APOEε4* carriers, *PSEN1-E318G* is associated with a 10-fold increased risk of developing AD and an earlier age at onset of AD (Benitez et al., 2013). Together, these results suggest that some variants in *PSEN1* and *PSEN2* are risk factors for AD rather than fully penetrant, causative mutations.

Protective APP Variants

It is well known that *APOEε4* is associated with increased AD risk and that *APOEε2* decreases AD risk, demonstrating that different variants within the same gene can have opposing effects on disease risk. However, most genetic studies of AD have focused on identifying variants that increase AD risk. Jonsson et al. reported that rs63750847 (*APP A673T*) was associated with reduced risk of AD (odds ratio = 0.23) in the Icelandic population (Jonsson et al., 2012). Interestingly, this mutation occurs at the same amino acid residue as a recessive mutation reported to cause early-onset AD (*APP A673V*). This mutation is near the proteolytic cleavage site of BACE1 (position 2 in the Aβ peptide) and results in impaired BACE1 cleavage of APP in the A673T carriers and reduction of Aβ40 and Aβ42 *in vitro* (Jonsson et al., 2012). This study provides a proof of principle for the hypothesis that reducing the β-cleavage of APP may protect against AD.

New AD Risk Genes

TREM2

Using whole-exome sequencing and whole-genome sequencing strategies, two groups simultaneously reported a low-frequency variant (*TREM2* R47H; rs75932628) in triggering receptor expressed on myeloid cells 2 protein (*TREM2*) that was associated with a twofold to threefold increase in AD risk (Guerreiro et al., 2013; Jonsson et al., 2013). Guerreiro et al. described several rare variants in *TREM2* that were observed more frequently in cases than in controls, suggesting that there may be multiple rare variants in *TREM2* that increase risk for AD. This theory is supported by data from several subsequent studies in European and African American populations. Some studies also suggest that *TREM2* R47H could be associated with Parkinson's disease, frontotemporal dementia, and amyotrophic lateral sclerosis, but this association remains controversial. In previous studies, rare homozygous loss-of-function mutations in *TREM2* were associated with an autosomal recessive form of early-onset dementia: polycystic lipomembranous osteodysplasia with sclerosing leukoencephalopathy (Paloneva et al., 2003). Subsequent studies found *TREM2* mutations in three patients with frontotemporal-like dementia without any bone-associated symptoms (Guerreiro et al., 2013). These observations highlight a recurring theme in neurodegenerative diseases: Heterozygous and homozygous mutations in the same gene can lead to clinically distinct disorders.

TREM2 is a type 1 transmembrane receptor protein expressed on myeloid cells, including microglia, monocyte-derived dendritic cells, osteoclasts, and bone marrow–derived macrophages. In the brain, TREM2 is expressed primarily on microglia and has been shown to control two signaling pathways: regulation of phagocytosis and suppression of inflammation reactivity. TREM2 deficiency and haploinsufficiency in a mouse model of β-amyloid deposition augment Aβ accumulation owing to a dysfunctional response of microglia, which fail to cluster around Aβ plaques and become apoptotic (Wang et al., 2015). This study also demonstrated that TREM2 senses a broad array of anionic and zwitterionic lipids known to associate with fibrillar Aβ in lipid membranes and to be exposed on the surface of damaged neurons. Remarkably, the R47H risk factor for AD impairs TREM2 detection of lipid ligands. Thus, TREM2 detects damage-associated lipid patterns associated with neurodegeneration, sustaining the microglial response to Aβ accumulation. The identification of coding variants in *TREM2* that increase the risk for AD supports the role of the immune response and inflammation in AD pathogenesis. Interestingly, coding variants in other genes in the *TREM* gene family may also modify AD disease risk, as is the case for a common protective variant in *TREML2* (Benitez et al., 2014).

NOTES

PLD3, ABCA7, UNC5C, AKAP9, and *SORL1*

Applying a family-based design to identify AD risk genes has revealed several rare, missense variants in *phospholipase D3 (PLD3)*, ATP-binding cassette subfamily 1, member 7 (*ABCA7*), Netrin receptor gene (*UNC5C*), A kinase anchor protein 9 (*AKAP9*), and sortilin-related receptor, L (DLR class) A repeats (*SORL1*) in individual studies. *ABCA7* and *SORL1* had previously been implicated by GWAS, which suggests that both common and rare variants of these genes modulate risk. Several loss-of-function alleles have been reported in *ABCA7* that are found at higher frequency in AD cases compared with controls, implicating *ABCA7* function in protection against developing AD. For *PLD3, UNC5C,* and *AKAP9*, additional studies in large cohorts will be necessary to confirm whether or not these genes influence the risk for AD.

Disease Mechanisms Implicated by Studies of Genetic Risk Factors for AD

Early-onset AD mutations in *APP, PSEN1,* and *PSEN2* lead to altered production or ratios of Aβ isoforms in the brain, while APOE influences Aβ clearance and aggregation; both observations support the hypothesis that Aβ levels are critical for disease pathogenesis (Hardy, 1997). GWASs have now identified polymorphisms in or near more than 20 genes that are associated with AD risk. The identification of common variants that have small effects on AD risk has created a broader picture of the processes and pathways involved in AD risk, including lipid metabolism, the inflammatory response, and endocytosis (Fig. 1). Whole-genome and exome-sequencing studies have also identified risk alleles in *TREM2, ABCA7, UNC5C, SORL1,* and *PLD3*. The identification of rare variants in the population that have moderate-to-large effects on AD risk will be most valuable for identifying pathways that are central to AD pathogenesis.

Future Directions for Genetic Studies

Substantial progress has been made during the past five years toward understanding the genetic architecture of AD because of the technological advances in genotyping and sequencing that have made it feasible to genotype or sequence thousands of individuals. GWASs have now identified more than 20 loci that influence risk for AD. It is clear from these studies that, with the exception of *APOE*, common risk alleles for AD have modest effects on risk individually but point to a number of important disease pathways. In other disorders in which even larger samples have been genotyped, more risk loci have been identified, suggesting that there is some value to continuing to increase the number of samples with GWAS data. Whereas very large sample sizes have been examined for people of European descent, this is not true for other populations. Large GWASs in these populations may identify genes not detected in Europeans. In addition, other AD-related phenotypes, including rate of progression of AD, rate of cognitive decline, and age at onset of AD, have yet to be examined in the largest datasets.

An alternative approach to increasing sample size is to use gene-based or network-based analyses to identify the genes that, while not of genome-wide significance, are overrepresented among SNPs, which have low *p* values in GWASs. This approach has identified additional significant genes that further implicate immune regulation, energy metabolism, and protein degradation in AD risk (Escott-Price et al., 2014). Another approach has compared GWAS data across related disorders to identify common underlying genes that demonstrate a genetic link between cardiovascular disease and AD risk. A remaining challenge for the field is to follow up the GWAS loci to identify the specific functional alleles and mechanisms underlying the GWAS signals. We anticipate that similar success will be observed as large-scale sequencing projects begin to yield results. Early studies in small discovery datasets have already identified several promising genes that modify AD risk. So far, these studies have focused largely on whole-exome sequencing and have identified coding variations that raise or lower the risk for AD by as much as twofold to threefold. This has the advantage of making functional follow-up of the genes identified through sequencing substantially easier that those identified in GWAS.

Several groups are performing whole-exome or whole-genome sequencing in unrelated AD cases and controls, and in families with multiple members affected by AD. It is anticipated that whole-exome or whole-genome sequencing will be available on more than 20,000 well-characterized samples before the end of 2015. Both the GWAS data and the sequencing data will be made available through public repositories such as dbGaP (the database of

NOTES

Genotypes and Phenotypes, available at http://www.ncbi.nlm.nih.gov/gap). Combining these data with other large datasets, such as transcriptomics from human brain tissue, immune cells, or induced pluripotent stem cell–derived neurons and glia, will provide further insight into the disease pathways, and regulatory nodes within these pathways, that may provide druggable targets for future therapies. An early application of this integrative network–based approach used brain tissue from more than 1600 AD cases and nondemented individuals to identify an immune-specific and microglia-specific module that is dominated by genes involved in pathogen phagocytosis (Forabosco et al., 2013; Zhang et al., 2013). This module contains *TREM2* (identified by whole-exome sequencing) but shows that *TYROBP* (aka DAP12, which binds to TREM2) is a key regulator of the network and is upregulated in AD brains. Mouse microglia overexpressing intact or truncated TYROBP reveal expression changes that significantly overlap with the human brain TYROBP network. This causal network structure was able to predict responses to gene perturbations, and thus presents a useful framework for testing models of disease mechanisms underlying AD.

Acknowledgments

C.M.K. is supported by a Career Development Award from the National Institutes of Health (NIH) (Grant K01 AG046374); A.M.G. received support from the Ronald M. Loeb Center for Alzheimer's Disease and the NIH (Grant AG035083); and C.C. received support from the NIH (Grant AG044546). This paper was adapted from Karch CM, Cruchaga C, Goate AM (2014) Alzheimer's disease genetics: from the bench to the clinic. Neuron 83:11–26.

References

Alzheimer Disease & Frontotemporal Dementia Mutation Database (n.d.). Available at http://www.molgen.vib-ua.be/ADMutations

Benitez BA, Jin, SC, Guerreiro R, Graham R, Lord J, Harold D, Sims R, Lambert JC, Gibbs JR, Bras J, Sassi C, Harari O, Bertelsen S, Lupton MK, Powell J, Bellenguez C, Brown K, Medway C, Haddick PC, van der Brug MP, et al. (2014) Missense variant in *TREML2* protects against Alzheimer's disease. Neurobiol Aging 35:1510 e1519–1526.

Benitez BA, Karch CM, Cai Y, Jin SC, Cooper B, Carrell D, Bertelsen S, Chibnik L, Schneider JA, Bennett DA; Alzheimer's Disease Neuroimaging Initiative; Genetic and Environmental Risk for Alzheimer's Disease Consortium GERAD, Fagan AM, Holtzman D, Morris JC, Goate AM, Cruchaga C (2013) The *PSEN1*, p.E318G variant increases the risk of Alzheimer's disease in APOE-epsilon4 carriers. PLoS Genet 9:e1003685.

Bergmans BA, De Strooper B (2010) Gamma-secretases: from cell biology to therapeutic strategies. Lancet Neurol 9:215–226.

Castellano JM, Kim J, Stewart FR, Jiang H, DeMattos RB, Patterson BW, Fagan AM, Morris JC, Mawuenyega KG, Cruchaga C, Goate AM, Bales KR, Paul SM, Bateman RJ, Holtzman DM (2011) Human apoE isoforms differentially regulate brain amyloid-beta peptide clearance. Sci Transl Med 3:89ra57.

Corder EH, Saunders AM, Risch NJ, Strittmatter WJ, Schmechel DE, Gaskell PC Jr, Rimmler JB, Locke PA, Conneally PM, Schmader KE, et al. (1994) Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. Nat Genet 7:180–184.

Cruchaga C, Haller G, Chakraverty S, Mayo K, Vallania FL, Mitra RD, Faber K, Williamson J, Bird T, Diaz-Arrastia R, Foroud TM, Boeve BF, Graff-Radford NR, St Jean P, Lawson M, Ehm MG, Mayeux R, Goate AM; NIA-LOAD/NCRAD Family Study Consortium (2012) Rare variants in *APP, PSEN1* and *PSEN2* increase risk for AD in late-onset Alzheimer's disease families. PLoS One 7:e31039.

Escott-Price V, Bellenguez C, Wang LS, Choi SH, Harold D, Jones L, Holmans P, Gerrish A, Vedernikov A, Richards A, DeStefano AL, Lambert JC, Ibrahim-Verbaas CA, Naj AC, Sims R, Jun G, Bis JC, Beecham GW, Grenier-Boley B, Russo G, et al. (2014) Gene-wide analysis detects two new susceptibility genes for Alzheimer's disease. PloS One 9:e94661.

Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, Myers RH, Pericak-Vance MA, Risch N, van Duijn CM (1997) Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. JAMA 278:1349–1356.

Forabosco P, Ramasamy A, Trabzuni D, Walker R, Smith C, Bras J, Levine AP, Hardy J, Pocock JM, Guerreiro R, Weale ME, Ryten M (2013) Insights into TREM2 biology by network analysis of human brain gene expression data. Neurobiol Aging 34:2699–2714.

NOTES

Guerreiro RJ, Baquero M, Blesa R, Boada M, Brás JM, Bullido MJ, Calado A, Crook R, Ferreira C, Frank A, Gómez-Isla T, Hernández I, Lleó A, Machado A, Martínez-Lage P, Masdeu J, Molina-Porcel L, Molinuevo JL, Pastor P, Pérez-Tur J, et al. (2010). Genetic screening of Alzheimer's disease genes in Iberian and African samples yields novel mutations in presenilins and *APP*. Neurobiol Aging 31:725–731.

Guerreiro R, Wojtas A, Bras J, Carrasquillo M, Rogaeva E, Majounie E, Cruchaga C, Sassi C, Kauwe JS, Younkin S, Hazrati L, Collinge J, Pocock J, Lashley T, Williams J, Lambert JC, Amouyel P, Goate A, Rademakers R, Morgan K, (2013) *TREM2* variants in Alzheimer's disease. N Engl J Med 368:117–127.

Hardy J (1997) The Alzheimer family of diseases: many etiologies, one pathogenesis? Proc Natl Acad Sci USA 94:2095–2097.

Hardy J, Selkoe DJ (2002) The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. Science 297:353–356.

Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, Pahwa JS, Moskvina V, Dowzell K, Williams A, Jones N, Thomas C, Stretton A, Morgan AR, Lovestone S, Powell J, Proitsi P, Lupton MK, Brayne C, Rubinsztein DC, et al. (2009) Genome-wide association study identifies variants at *CLU* and *PICALM* associated with Alzheimer's disease. Nat Genet 41:1088–1093.

Hollingworth P, Harold D, Sims R, Gerrish A, Lambert JC, Carrasquillo MM, Abraham R, Hamshere ML, Pahwa JS, Moskvina V, Dowzell K, Jones N, Stretton A, Thomas C, Richards A, Ivanov D, Widdowson C, Chapman J, Lovestone S, Powell J, et al. (2011) Common variants at *ABCA7*, *MS4A6A/MS4A4E*, *EPHA1*, *CD33* and *CD2AP* are associated with Alzheimer's disease. Nat Genet 43:429–435.

Holtzman DM, Morris JC, Goate AM (2011) Alzheimer's disease: the challenge of the second century. Sci Transl Med 3:77sr71.

Jonsson T, Atwal JK, Steinberg S, Snaedal J, Jonsson PV, Bjornsson S, Stefansson H, Sulem P, Gudbjartsson D, Maloney J, Hoyte K, Gustafson A, Liu Y, Lu Y, Bhangale T, Graham RR, Huttenlocher J, Bjornsdottir G, Andreassen OA, Jönsson EG, et al. (2012) A mutation in *APP* protects against Alzheimer's disease and age-related cognitive decline. Nature 488:96–99.

Jonsson T, Stefansson H, Steinberg S, Jonsdottir I, Jonsson PV, Snaedal J, Bjornsson S, Huttenlocher J, Levey AI, Lah JJ, Rujescu D, Hampel H, Giegling I, Andreassen OA, Engedal K, Ulstein I, Djurovic S, Ibrahim-Verbaas C, Hofman A, Ikram MA, et al. (2013) Variant of *TREM2* associated with the risk of Alzheimer's disease. N Engl J Med 368:107–116.

Lambert JC, Heath S, Even G, Campion D, Sleegers K, Hiltunen M, Combarros O, Zelenika D, Bullido MJ, Tavernier B, Letenneur L, Bettens K, Berr C, Pasquier F, Fiévet N, Barberger-Gateau P, Engelborghs S, De Deyn P, Mateo I, Franck A, et al. (2009) Genome-wide association study identifies variants at *CLU* and *CR1* associated with Alzheimer's disease. Nat Genet 41:1094–1099.

Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, DeStafano AL, Bis JC, Beecham GW, Grenier-Boley B, Russo G, Thorton-Wells TA, Jones N, Smith AV, Chouraki V, Thomas C, Ikram MA, Zelenika D, Vardarajan BN, Kamatani Y, et al. (2013) Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet 45:1452–1458.

Mahley RW, Rall SC Jr (2000) Apolipoprotein E: far more than a lipid transport protein. Annu Rev Genomics Hum Genet 1:507–537.

Miyashita A, Koike A, Jun G, Wang LS, Takahashi S, Matsubara E, Kawarabayashi T, Shoji M, Tomita N, Arai H, Asada T, Harigaya Y, Ikeda M, Amari M, Hanyu H, Higuchi S, Ikeuchi T, Nishizawa M, Suga M, Kawase Y, et al. (2013) *SORL1* is genetically associated with late-onset Alzheimer's disease in Japanese, Koreans and Caucasians. PLoS One 8:e58618.

Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buros J, Gallins PJ, Buxbaum JD, Jarvik GP, Crane PK, Larson EB, Bird TD, Boeve BF, Graff-Radford NR, De Jager PL, Evans D, Schneider JA, Carrasquillo MM, Ertekin-Taner N, Younkin SG, et al. (2011) Common variants at *MS4A4/MS4A6E*, *CD2AP*, *CD33* and *EPHA1* are associated with late-onset Alzheimer's disease. Nat Genet 43:436–441.

Paloneva J, Mandelin J, Kiialainen A, Bohling T, Prudlo J, Hakola P, Haltia M, Konttinen YT, Peltonen L (2003) DAP12/TREM2 deficiency results in impaired osteoclast differentiation and osteoporotic features. J Exp Med 198:669–675.

NOTES

Reitz C, Tosto G, Vardarajan B, Rogaeva E, Ghani M, Rogers RS, Conrad C, Haines JL, Pericak-Vance MA, Fallin MD, Foroud T, Farrer LA, Schellenberg GD, George-Hyslop PS, Mayeux R; Alzheimer's Disease Genetics Consortium (ADGC) (2013) Independent and epistatic effects of variants in VPS10-d receptors on Alzheimer disease risk and processing of the amyloid precursor protein (APP). Transl Psychiatry 3:e256.

Seshadri S, Fitzpatrick AL, Ikram MA, DeStefano AL, Gudnason V, Boada M, Bis JC, Smith AV, Carassquillo MM, Lambert JC, Harold D, Schrijvers EM, Ramirez-Lorca R, Debette S, Longstreth WT Jr, Janssens AC, Pankratz VS, Dartigues JF, Hollingworth P, Aspelund T, et al. (2010) Genome-wide analysis of genetic loci associated with Alzheimer disease. JAMA 303:1832–1840.

Wang Y, Cella M, Mallinson K, Ulrich JD, Young KL, Robinette ML, Gilfillan S, Krishnan GM, Sudhakar S, Zinselmeyer BH, Holtzman DM, Cirrito JR, Colonna M (2015) TREM2 lipid sensing sustains the microglial response in an Alzheimer's disease model. Cell 160:1061–1071.

Zhang B, Gaiteri C, Bodea LG, Wang Z, McElwee J, Podtelezhnikov AA, Zhang C, Xie T, Tran L, Dobrin R, Fluder E, Clurman B, Melquist S, Narayanan M, Suver C, Shah H, Mahajan M, Gillis T, Mysore J, MacDonald ME, et al. (2013) Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. Cell 153:707–720.

Transcription Dysregulation of the PGC-1α Pathway in Huntington's Disease Pathogenesis: From Metabolic Derangement to Neurodegeneration

Albert R. La Spada, MD, PhD

Departments of Cellular and Molecular Medicine,
Neurosciences, and Pediatrics
Division of Biological Sciences, the Institute for Genomic Medicine
Sanford Consortium for Regenerative Medicine
University of California, San Diego
La Jolla, California

NOTES

Clinical Description and Molecular Genetics of Huntington's Disease

Huntington's disease (HD) is an autosomal dominant neurodegenerative disorder characterized by motor and cognitive impairment (Nance, 1997). The motor abnormality stems from dysfunction of the involuntary movement control region of the midbrain known as the striatum, and is manifested as a hallmark feature of uncontrollable dance-like movements ("chorea"). In HD, significant degeneration and atrophy occur in the striatum and cerebral cortex, while cerebellar, thalamic, and spinal cord neuron populations are spared (Ross et al., 1997). Neuroanatomical studies have revealed that GABAergic medium spiny neurons (MSNs) in the striatum are selectively vulnerable in HD, while medium aspiny and large cholinergic striatal neurons are preserved (Ferrante et al., 1985; Graveland et al., 1985). HD is relentlessly progressive, and after pronounced cognitive decline, patients succumb to the disease usually 15 to 25 years after disease onset. In 1993, a CAG triplet repeat expansion mutation in the coding region of the *huntingtin (htt)* gene was identified as the cause of HD (MacDonald et al., 1993). As observed for other polyglutamine (polyQ) repeat diseases, polyQ-htt tracts that exceed a certain length threshold (~37 repeats) adopt a novel pathogenic conformation, yielding conformers that are resistant to normal protein turnover, culminating in cell toxicity and neurodegeneration (La Spada and Taylor, 2010).

PGC-1α links Transcription Interference With Mitochondrial Abnormalities in HD

Neurons in the brain have enormous demands for continued mitochondrial production of high-energy phosphate-bonded compounds. In 1993, Beal and colleagues reported that chronic administration of a mitochondrial toxin, 3-nitropropionic acid, resulted in a selective loss of medium spiny neurons in the striatum (Beal et al., 1993). This provocative finding, which was corroborated by numerous studies in HD cell culture models, mice, and human patients (Lin and Beal, 2006), suggested that mitochondrial dysfunction may underlie HD pathogenesis and account for cell-type specificity in this disorder. At the same time, the necessity of nuclear localization of htt for HD disease pathogenesis highlighted nuclear pathology as a key step in the neurotoxicity cascade (Saudou et al., 1998). An extensive body of literature then emerged, suggesting that N-terminal fragments of mutant htt protein interfere with gene transcription in HD (Riley and Orr, 2006).

HD Transgenic Mice Display Profound Thermoregulatory Defects

Because neurological deficits in HD are gradually progressive, considerable emphasis has been placed on identifying objective and reproducible measures of disease onset and progression (i.e., "biomarkers") to improve the predictive value of therapeutic trials. We therefore chose to evaluate several metabolic parameters, including body temperature, in a commonly used model of HD, the *N171-82Q* transgenic mouse (Schilling et al., 1999). When we monitored body temperature, we found that HD mice developed progressive hypothermia compared with their wild-type (WT) counterparts (Figs. *1A, B*). As motor symptoms and weight loss progressed, some HD *N171-82Q* mice displayed profoundly deranged thermoregulation, with body temperatures dropping to 27°C or less. HD mice with temperatures below 30°C were not within hours or minutes of death, as they remained mobile and alive for at least another 48 h, often considerably longer. In light of this striking hypothermia phenotype, we reasoned that HD *N171-82Q* mice might not be capable of maintaining body temperature in the face of a 4°C cold challenge—a process known as "adaptive thermogenesis" (Lowell and Spiegelman, 2000).

After obtaining HD *N171-18Q* transgenic mice to control for htt protein overexpression, we established three cohorts of mice: HD 82Q, HD 18Q, and WT. Individual mice were placed at 4°C for up to 9 h, and body temperatures were recorded at 1-h intervals. Although control mice were able to maintain normal thermoregulation, HD transgenic mice displayed significant reductions in body temperature during the cold challenge, even when presymptomatic for baseline hypothermia (Figs. 1*C, D*). In rodents, brown adipose tissue (BAT) is the principal tissue that mediates adaptive thermogenesis, and is distinguished from white fat by its high degree of vascularization and mitochondrial density (Wang et al., 2005). Hematoxylin and eosin staining of BAT from HD mice revealed marked abnormalities, including reductions in cell density and nuclei number (Figs. 1*E, F*). Indeed, the BAT of HD mice appeared like white fat in histology sections, suggesting that the thermogenesis defect likely involves abnormalities in BAT composition and function. Importantly, reverse transcriptase PCR (RT-PCR) analysis indicated that the mutant *htt* transgene is expressed in BAT (Weydt et al., 2006).

NOTES

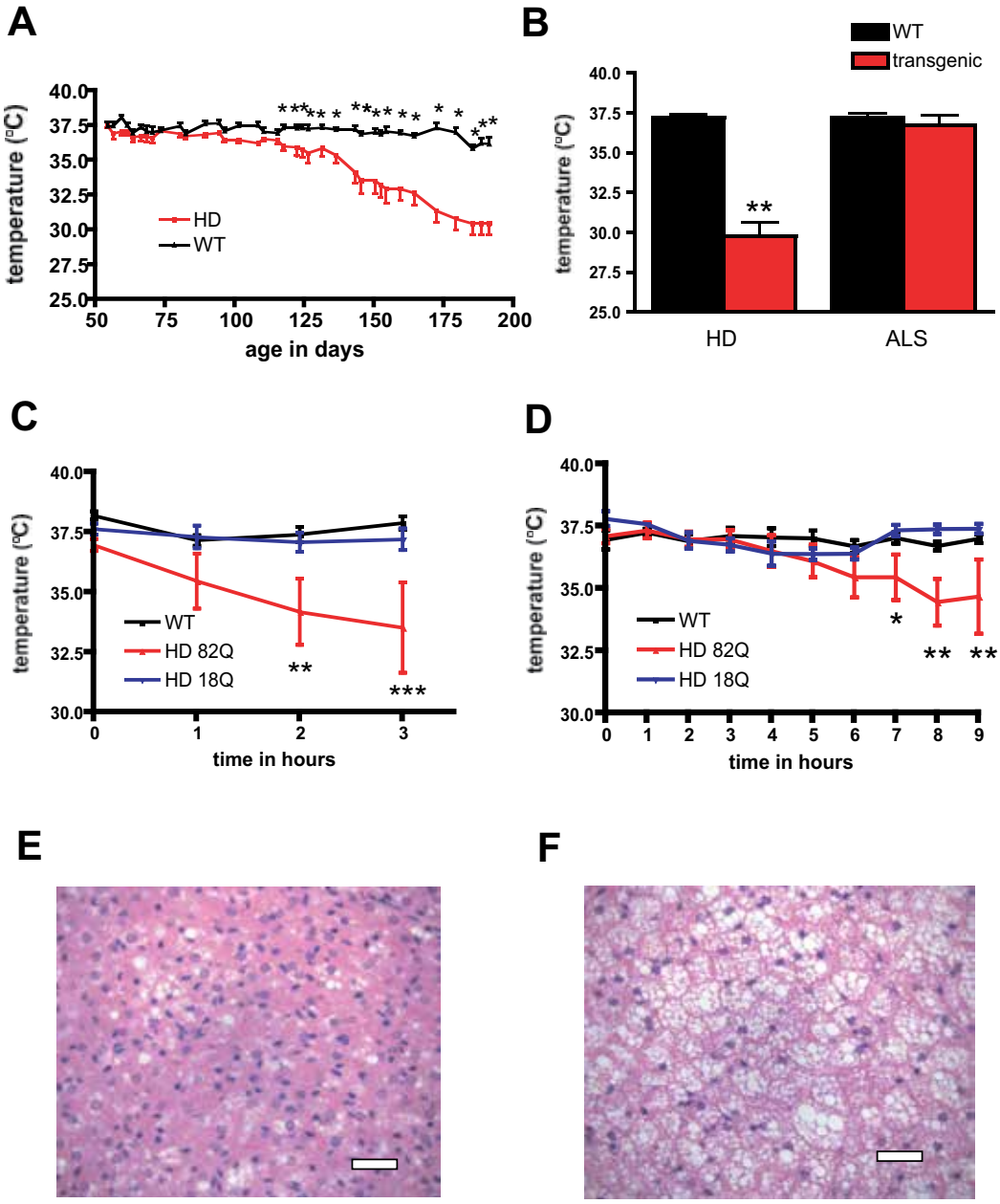


Figure 1. HD mice display a temperature regulation defect and abnormal BAT. ***A***, Body temperature in HD N171-82Q male mice (red line) and nontransgenic controls (WT, black line). By 120 d, HD mice display a marked reduction in body temperature (p = 0.027), and this reduction progresses with time. By 150 d, the difference in body temperature is extreme (HD: ~33°C; WT: ~37°C; p = 0.003). ***B***, End-stage HD mice exhibit profound hypothermia (n = 6; $p < 0.001$ by t test), while end-stage SOD1 G93A ALS (amyotrophic lateral sclerosis) mice do not (n = 5; p = 0.45 by t test). ***C***, HD transgenic mice display an adaptive thermogenesis defect. Twenty-week-old mice were placed at 4°C, and body temperatures were recorded. By 1 h, HD 82Q mice displayed a significant reduction in body temperature (**$p < 0.01$) that worsened with time (***$p < 0.001$). ***D***, Younger HD mice displayed a thermoregulatory defect by 7 h into the cold challenge (*$p < 0.05$) that progressively worsened (**$p < 0.01$). ***E–F***, Twenty-week-old HD mice and WT controls were euthanized, and infrascapular BAT samples were stained. While WT BAT appears normal (***E***), BAT from HD 82Q mice is markedly abnormal (***F***), showing decreased cellular content (note fewer nuclei) and marked accumulation of large lipid droplets. Scale bar, 20 μm. Modified with permission from Weydt et al. (2006), their Fig. 1.

NOTES

The PGC-1α–UCP-1 Circuit Is Disrupted in the BAT of HD Transgenic Mice

In mammals, after cold is sensed by the hypothalamus, an increase in sympathetic tone in the periphery ensues. In rodents, BAT is the target of this increased sympathetic output. PPAR-gamma coactivator-1-alpha (PGC-1α) is a transcription coactivator whose expression in BAT is dramatically upregulated in response to β-adrenergic stimulation (Puigserver et al., 1998). The principal effector of adaptive thermogenesis in BAT is uncoupling protein 1 (UCP-1), whose expression is restricted to mitochondria of BAT (Puigserver and Spiegelman, 2003). To determine whether PGC-1α transactivation of UCP-1 is normal in HD mice, we dissected infrascapular BAT after cold challenge, isolated total RNA, and measured *PGC-1α* and *UCP-1* transcripts. We observed marked upregulation of *PGC-1α* in the BAT of cold-challenged control and HD mice (Fig. 2*A*). This result indicates that hypothalamic sensing of temperature change, elevation of sympathetic tone, and β-adrenergic stimulation of *PGC-1α* in BAT are intact. Detection of *c-fos* upregulation in the ventromedial hypothalamic nucleus of cold-challenged HD mice confirmed hypothalamus activation in response to cold (Weydt et al., 2006). Despite preservation of hypothalamus-mediated β-adrenergic stimulation of *PGC-1α* in BAT, cold-challenged HD mice failed to upregulate *UCP-1* messenger RNA (mRNA) (Fig. 2*B*). In addition, cold-challenged levels of UCP-1 protein were decreased in the BAT of HD mice (Fig. 2*C*). These results suggest that interference with PGC-1α coactivation of UCP-1 in BAT accounts for the adaptive thermogenesis defect in HD.

To further investigate this hypothesis, 3T3-L1 preadipocyte cells were transfected with UCP-1 promoter–reporter constructs along with mutant or normal htt in the presence or absence of PGC-1α. While baseline transactivation levels were similar, polyQ-htt repressed stimulation of UCP-1 promoter activity; importantly, mutant htt repression of *UCP-1* transcription could be overcome by coexpression of PGC-1α (Fig. 2*D*). Because preadipocyte cells are not committed to BAT differentiation, we established primary brown adipocyte cultures from HD *N171-82Q* and control mice. Upon norepinephrine (NE) stimulation, primary brown adipocytes from HD mice and nontransgenic controls displayed comparable *PGC-1α* induction (Weydt et al., 2006); however, *UCP-1* induction was significantly blunted in adipocytes expressing polyQ-expanded htt (Fig. 2*E*). Failure of *UCP-1* induction was confirmed at the protein level (Weydt et al., 2006).

Evidence for PGC-1α Transcription Interference in Mouse Striatum

To determine whether PGC-1α function was compromised in the striatum of HD *N171-82Q* transgenic mice, we isolated striatal RNAs and measured the expression level of PGC-1α target genes whose protein products mediate oxidative metabolism in mitochondria (Mootha et al., 2003; Leone et al., 2005). In 20-week-old HD mice, there was a significant reduction in the expression of such mitochondrial genes (Fig. 3*A*). These findings support a role for PGC-1α transcription interference in the degeneration of the striatum in HD.

Human HD Patients Display PGC-1α Transcription Interference in the Striatum

PGC-1α transcription abnormalities in the brain and periphery of the *N171-82Q* HD model led us to ask: Do HD patients display PGC-1α transcription interference in the striatum? To address this question, we analyzed caudate nucleus microarray expression data obtained from a large cohort of human HD patients and matched controls (Hodges et al., 2006). We selected 26 genes known to rely on PGC-1α coactivator function for their expression (Mootha et al., 2003; Leone et al., 2005), and using the gcrma application from the Bioconductor open-source software program (www.bioconductor.org) (Bolstad et al., 2003; Gentleman et al., 2004), we noted significant reductions in 24 of these 26 PGC-1α target genes (Fig. 3*B*) (Weydt et al., 2006). The presence of significant expression reductions in 35 of 46 probes (corresponding to the 26 PGC-1α target genes) from the Affymetrix GeneChip Human Genome U133 Array Set (HG-U133A/B, consisting of 45,000 probes) (Affymetrix, Santa Clara, CA) is highly unlikely to occur by chance ($p < 0.0001$; χ^2). Thus, these results strongly support PGC-1α transcription interference in the striatum of presymptomatic and early-stage HD patients.

To validate these findings, we obtained striatal RNAs from a subset of these cases and performed real-time RT-PCR analysis. We confirmed significant reductions in mitochondrial PGC-1α target genes in the human HD sample set (Fig. 3C). To control for the validation analysis, we included the glial

NOTES

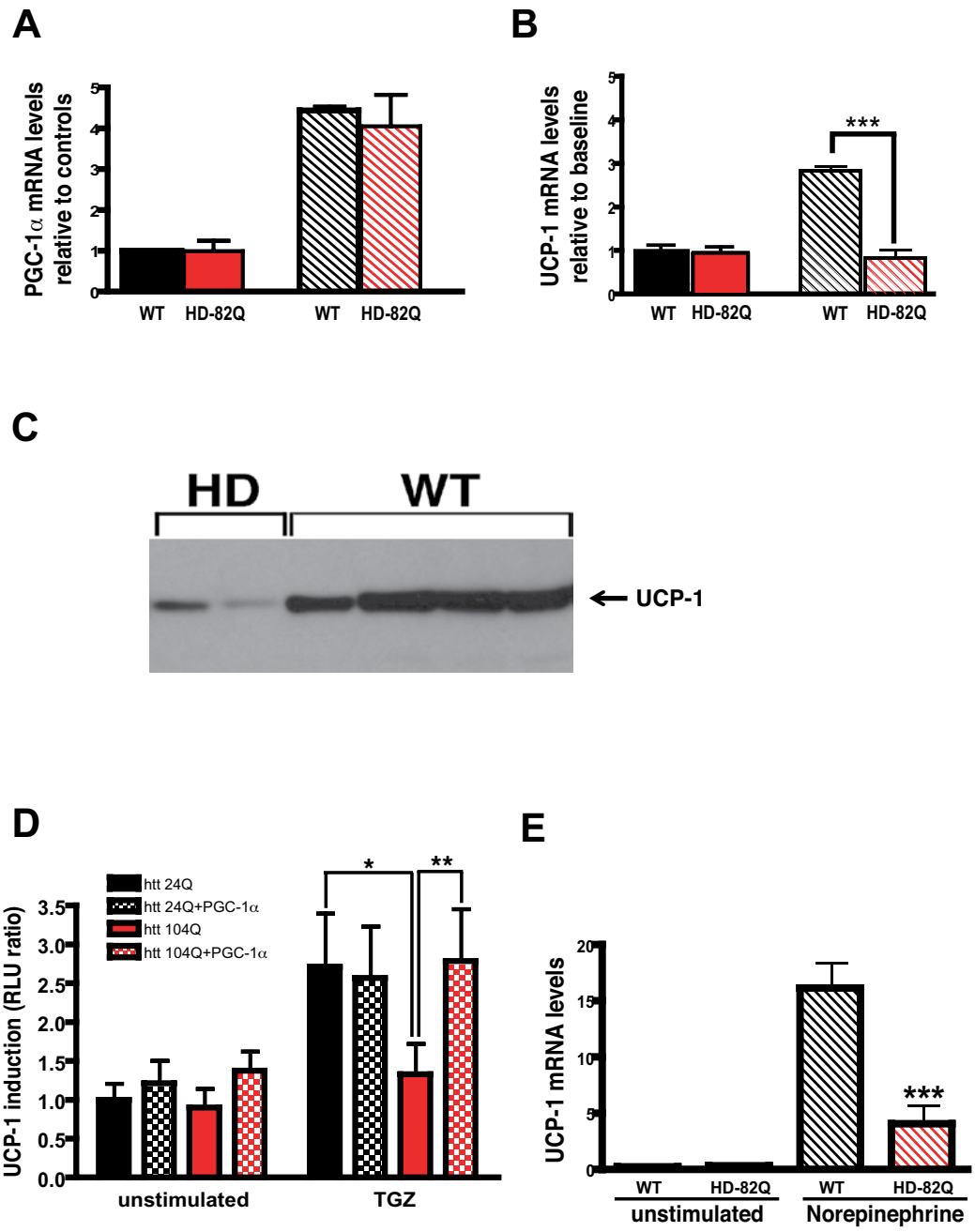


Figure 2. HD mice display PGC-1α transcription interference and BAT abnormality. ***A***, HD mice induce PGC-1α upon cold challenge. Here we see quantitative PCR analysis of BAT RNAs, from HD 82Q (red) and WT mice (black) housed at room temperature (solid bars) or in the cold for 3 h (striped bars). Upregulation of PGC-1α is similar ($p = 0.77$; *t* test). ***B***, HD mice fail to induce UCP-1 RNA during cold challenge. WT mice increase UCP-1 by approximately threefold, but HD 82Q mice display no induction (***$p < 0.001$). ***C***, HD transgenic mice fail to upregulate UCP-1 protein expression during cold challenge. Western blot analysis of brown fat UCP-1 protein levels in cold-challenged HD N171-82Q mice (HD) and nontransgenic littermate controls (WT) indicates upregulation of UCP-1 in WT mice at the end of the 3-h cold challenge compared with HD mice. Staining of SDS-PAGE gels prior to transfer confirmed equivalent loading of brown fat protein samples (not shown). ***D***, polyQ-Htt transcription interference of PGC-1α. 3T3-L1 preadipocytes were cotransfected with PPARγ, retinoid X receptor–alpha (RXRα), UCP-1 promoter reporter, Renilla luciferase–cytomegalovirus (pRL-CMV), exon1/2 htt 24Q or 104Q, and PGC-1α. Htt 104Q suppresses the UCP-1 promoter reporter in cells treated with troglitazone (TGZ) (*$p = 0.011$). Transfection of PGC-1α rescues htt 104Q repression (**$p < 0.005$). ***E***, HD brown adipocytes do not respond to NE. NE yields upregulation of UCP-1 in WT adipocytes, but UCP-1 induction is blunted in adipocytes expressing htt 82Q (***$p < 0.001$ by t test). RLU, Relative luciferase units. Modified with permission from Weydt et al. (2006), their Figs. 2*A*–*D*.

NOTES

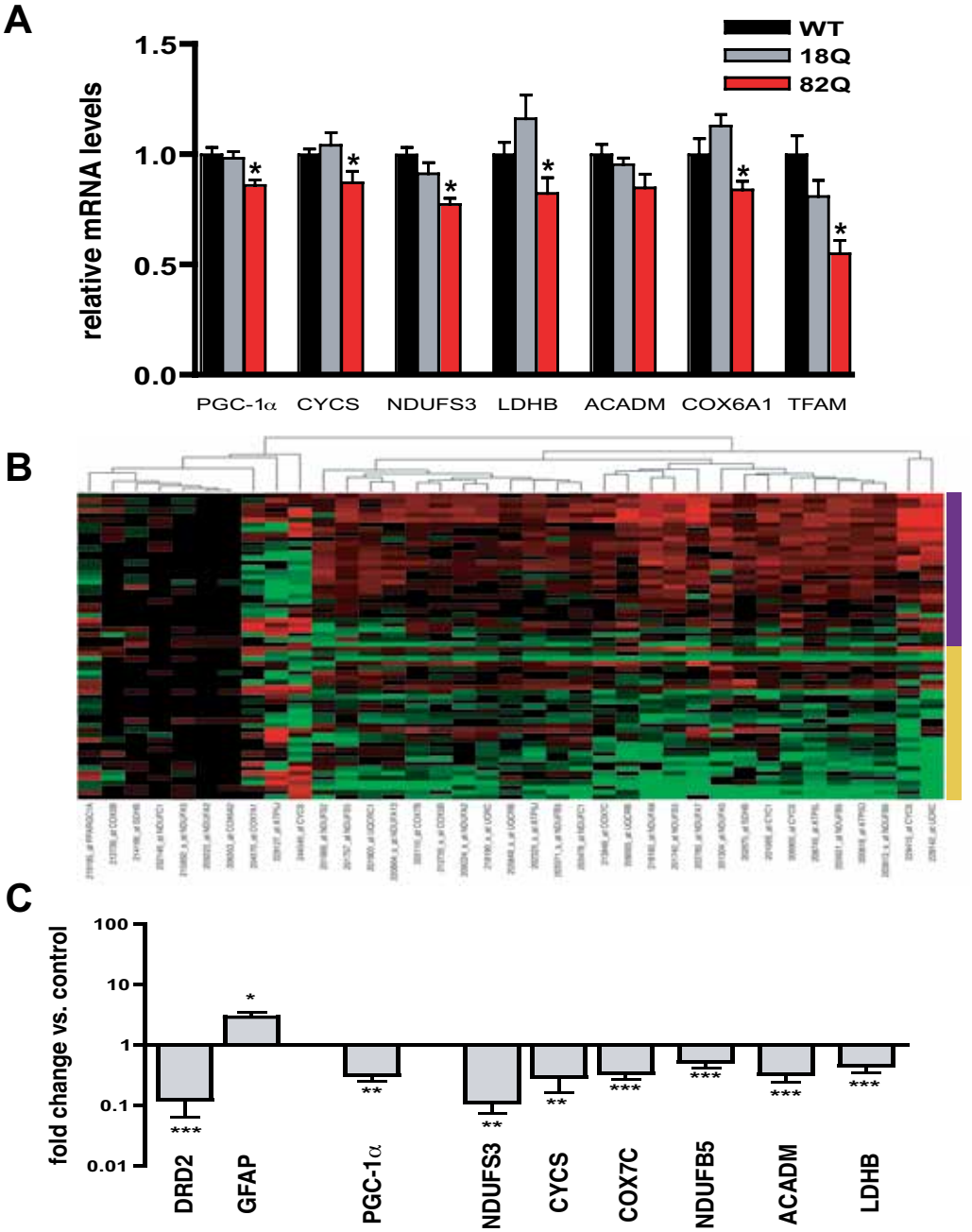


Figure 3. PGC-1α transcription interference in HD striatum. ***A***, RT-PCR analysis of striatal RNAs from sets (n = 5/group) of 20-week-old HD 82Q mice (red), 18Q mice (gray), and WT mice (black) indicates that RNA levels for PGC-1α and six of its mitochondrial target genes are reduced in HD brain. (All HD 82Q: *$p < 0.05$, except p = 0.17 for ACADM). ***B***, Microarray expression analysis of PGC-1α–regulated genes in human caudate. Here we see a heat map comparing the caudate nucleus expression of 26 PGC-1α target genes for 32 Grade 0–2 HD patients (gold bar) and 32 controls (purple bar). Most PGC-1α target genes are downregulated in HD patients. ***C***, Confirmation of microarray data. We obtained striatal RNA samples for HD patients and controls and measured RNA expression levels for six PGC-1α targets, PGC-1α, and two control genes (*GFAP* and *DRD2*). We thus confirmed significant reductions in the expression of PGC-1α targets and detected reduced PGC-1α in human HD striatum from early-grade patients. Statistical comparisons were performed using the t test (*$p < 0.05$; **$p < 0.005$; ***$p < 0.0005$). Modified with permission from Weydt et al. (2006), their Fig. 4.

NOTES

fibrillary acidic protein *(GFAP)* and dopamine D2 receptor *(DRD2)* genes in this experiment, as GFAP is known to be significantly upregulated and DRD2 is known to be significantly downregulated in expression studies of HD caudate (Luthi-Carter et al., 2000; Hodges et al., 2006).

PGC-1α Impairment Links Transcription Interference With HD Mitochondrial Dysfunction

Based on our findings (Weydt et al., 2006) and the work of another group published independently (Cui et al., 2006), a new model for HD pathogenesis has emerged (McGill and Beal, 2006; Ross and Thompson, 2006; Greenamyre, 2007). According to this model, mitochondrial dysfunction in HD is postulated to result directly from PGC-1α transcription interference.

Induction of PGC-1α Expression Rescues Neurological Phenotypes in HD Transgenic Mice

To determine whether increased expression of PGC-1α can ameliorate HD, we developed a system to induce PGC-1α expression in transgenic mice by obtaining a tet-responsive element (TRE)–PGC-1α transgenic line (Russell et al., 2004). We also derived a line of *Rosa26-rtTA* mice (Belteki et al., 2005) and crossed *Rosa26-rtTA* males with *TRE-PGC-1α* females to generate *Rosa26-rtTA–TRE-PGC-1α* bigenic mice. When *Rosa26-rtTA–TRE-PGC-1α* bigenic mice receive doxycycline (Dox), the rtTA becomes activated and should promote the expression of *PGC-1α*.

To validate our induction system, we derived *Rosa26-rtTA–TRE-PGC-1α* bigenic mice. These mice were fed Dox for 6 weeks beginning at weaning, and we then observed marked induction of *PGC-1α* (Tsunemi et al., 2012). To test whether restoring PGC-1α function can ameliorate neurological disease in HD, we crossed HD *N171-82Q* mice with inducible PGC-1α bigenic mice, utilizing a breeding scheme that yielded three different cohorts: triple transgenic mice, HD mice (no *rtTA* or *PGC-1α* transgenes), and non-HD controls. We subjected the cohorts to behavioral testing, and noted that expression of PGC-1α at levels consistent with prior induction significantly improved forepaw grip strength, gait, and performance on the ledge test in HD mice (Tsunemi et al., 2012). PGC-1α expression also enabled HD triple transgenic mice to perform comparably with control mice on the rotarod.

PGC-1α Prevents Huntingtin Protein Aggregation and Rescues HD Neurodegeneration

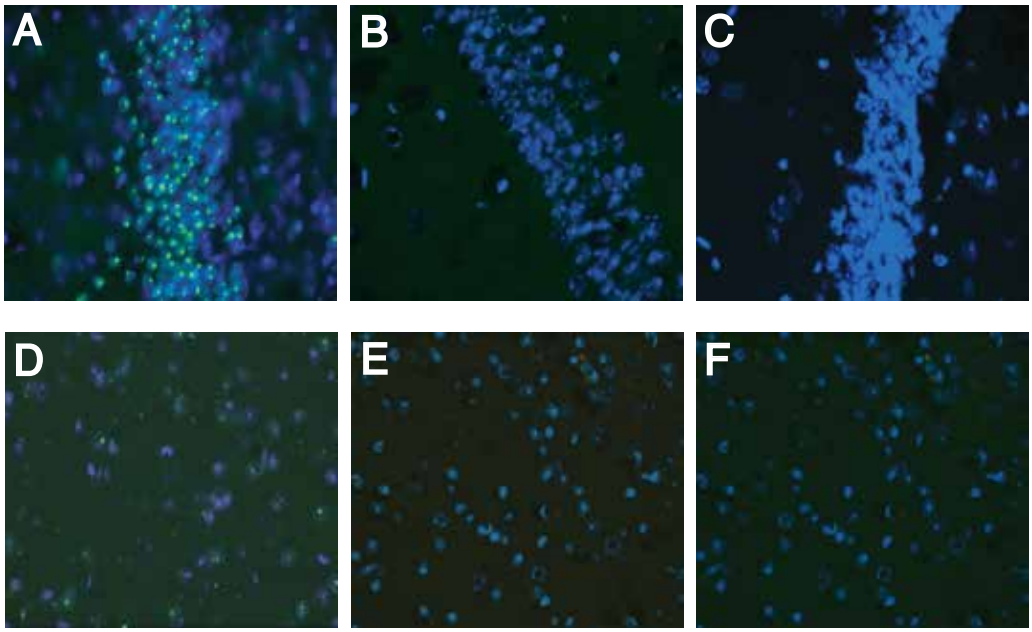
The formation of protein aggregates, visible at the light microscope level, is an established pathological hallmark of HD. Although aggregates are not the toxic species, their production requires misfolded htt; hence, their elimination correlates with marked reductions in pathogenic htt protein (Rubinsztein, 2006). When we examined the brains of 18-week-old HD mice induced to express PGC-1α, we observed a dramatic reduction in htt protein aggregation in hippocampus and cortex (Figs. 4*A–F*). Quantification of neurons containing htt protein aggregates confirmed this observation and demonstrated that induction of PGC-1α in triple transgenic mice is required for this outcome (Fig. 4G). We also noted a significant reduction in htt protein aggregation in the striatum, though there were fewer cells with aggregates there.

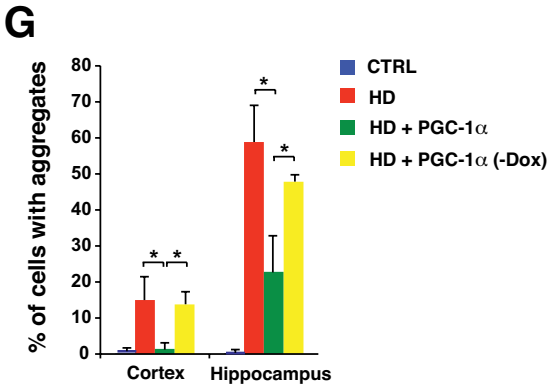
Filter trap assay is a widely used method for measuring SDS-insoluble misfolded proteins (Muchowski et al., 2002), and a variety of antibodies are available for detecting different amyloidogenic species, including 1C2 (polyQ tracts), A11 (prefibrillar oligomers), and OC (fibrils) (Kayed et al., 2007). Using these antibodies, we performed filter trap assays on protein lysates isolated from the striatum of HD, triple transgenic, and control mice, and noted obvious reductions in SDS-insoluble htt, oligomeric htt, and fibrillar htt in HD mice induced to express PGC-1α (Fig. 4H). Observed reductions in insoluble htt species could not be attributed to an effect of PGC-1α on HD transgene expression because quantitative RT-PCR analysis had revealed similar levels of *htt* transgene mRNA in HD transgenic mice expressing PGC-1α and in HD mice lacking both transgenes (Tsunemi et al., 2012). To determine whether improved behavior and reduced htt aggregation in HD triple transgenic mice were accompanied by an amelioration of neurodegeneration, we completed a stereological assessment of the striatum, and found that induction of PGC-1α significantly increased striatal volume and neuron number (Tsunemi et al., 2012).

PPARδ Is a Potent Regulator of Mitochondrial and Metabolic Function

The peroxisome proliferator–activated receptors (PPARs) are a family of nuclear, ligand-activated receptors that transcriptionally regulate important metabolic and physiological functions. Three PPARs

NOTES





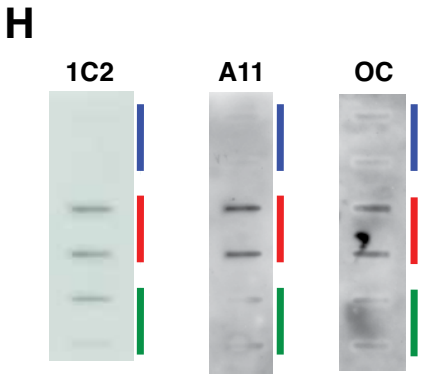


Figure 4. PGC-1α expression prevents htt aggregate formation. ***A–F***, Sections from the frontal cortex (***A–C***) and hippocampus CA3 region (***D–F***) of 18-week-old HD mice (***A, D***), non-HD littermate controls (***B, E***), and HD mice induced to express PGC-1α (***C, F***). Anti-htt antibody EM48 (green); DAPI (blue). ***G***, Quantification of htt aggregate formation in 18-week-old HD mice (*$p < 0.05$). ***H***, Filter trap assays were performed using different antibodies that detect alternative misfolded species of htt protein. 1C2 (polyQ tracts), A11 (oligomers), and OC (fibrils) each reveal a reduction in SDS-insoluble htt protein for 18-week-old HD transgenic mice expressing PGC-1α (green), compared with HD mice lacking the PGC-1α transgene (red). Non-HD controls do not exhibit appreciable levels of SDS-insoluble, oligomeric, or fibrillar htt protein (blue). Error bars indicate SD. Modified with permission from Tsunemi et al. (2012), their Figs. 2*A–H*.

NOTES

have been identified thus far: PPARα, PPARδ, and PPARγ (Berger and Moller, 2002). PPARα is expressed primarily in the liver, heart, and muscle and plays a key role in regulating fatty-acid breakdown. PPARγ regulates adipogenesis and thermoregulation in mammals, and is thus highly expressed in adipose tissue. PPARδ is widely expressed and regulates fatty-acid catabolism. Whereas all three PPARs interact with PGC-1α to promote oxidative metabolism and metabolic activity, overexpression of constitutively active PPARδ in skeletal muscle in transgenic mice dramatically favors a shift in muscle fibers to an oxidative metabolic status, thereby vastly improving exercise performance, even in untrained mice (Wang et al., 2004). A greatly enhanced oxidative metabolic shift can also be achieved in the skeletal muscle of WT mice treated with the PPARδ agonist GW501516, when combined with exercise (Narkar et al., 2008).

PPARδ CNS Expression and Neural Function: Is PPARδ Involved in HD?

While the role of PPARδ in CNS is yet to be explored, quantitative Western blot analysis indicates that PPARδ protein is expressed more highly in brain than in muscle, by at least twofold (Girroir et al., 2008). As PPARδ exhibits the highest expression of all three PPARs in the brain, one group has reported that PPARδ agonist treatment can ameliorate ischemic brain injury and reduce MPTP-induced striatal dopamine depletion in rodents (Iwashita et al., 2007). Interestingly, PPARα and PPARγ agonists have been shown to be therapeutically beneficial in the MPTP rodent model of Parkinson's disease (Dehmer et al., 2004; Kreisler et al., 2007). However, because PPARδ can be partially activated by these agonists, neuroprotection in these studies may actually result from PPARδ activation. In any event, there is good reason to believe that PPARδ could be involved in normal neural function and disease and to expect that PPARδ dysfunction could be contributing to HD pathogenesis.

Acknowledgments

This paper was adapted from Weydt et al., 2006, Thermoregulatory and metabolic defects in Huntington's disease transgenic mice implicate PGC-1alpha in Huntington's disease neurodegeneration, Cell Metab 4:349–362; and Tsunemi et al., 2012, PGC-1alpha rescues Huntington's disease proteotoxicity by preventing oxidative stress and promoting TFEB function, Sci Transl Med 4:142ra197.

References

Beal MF, Brouillet E, Jenkins BG, Ferrante RJ, Kowall NW, Miller JM, Storey E, Srivastava R, Rosen BR, Hyman BT (1993) Neurochemical and histologic characterization of striatal excitotoxic lesions produced by the mitochondrial toxin 3-nitropropionic acid. J Neurosci 13:4181–4192.

Belteki G, Haigh J, Kabacs N, Haigh K, Sison K, Costantini F, Whitsett J, Quaggin SE, Nagy A (2005) Conditional and inducible transgene expression in mice through the combinatorial use of Cre-mediated recombination and tetracycline induction. Nucleic Acids Res 33:e51.

Berger J, Moller DE (2002) The mechanisms of action of PPARs. Annu Rev Med 53:409–435.

Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19:185–193.

Cui L, Jeong H, Borovecki F, Parkhurst CN, Tanese N, Krainc D (2006) Transcriptional repression of PGC-1alpha by mutant *huntingtin* leads to mitochondrial dysfunction and neurodegeneration. Cell 127:59–69.

Dehmer T, Heneka MT, Sastre M, Dichgans J, Schulz JB (2004) Protection by pioglitazone in the MPTP model of Parkinson's disease correlates with I kappa B alpha induction and block of NF kappa B and iNOS activation. J Neurochem 88:494–501.

Ferrante RJ, Kowall NW, Beal MF, Richardson EP, Jr Bird ED, Martin JB (1985) Selective sparing of a class of striatal neurons in Huntington's disease. Science 230:561–563.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5:R80.

Girroir EE, Hollingshead HE, He P, Zhu B, Perdew GH, Peters JM (2008) Quantitative expression patterns of peroxisome proliferator-activated receptor-beta/delta (PPARbeta/delta) protein in mice. Biochem Biophys Res Commun 371:456–461.

Graveland GA, Williams RS, DiFiglia M (1985) Evidence for degenerative and regenerative changes in neostriatal spiny neurons in Huntington's disease. Science 227:770–773.

NOTES

Greenamyre JT (2007) Huntington's disease—making connections. N Engl J Med 356:518–520.

Hodges A, Strand AD, Aragaki AK, Kuhn A, Sengstag T, Hughes G, Elliston LA, Hartog C, Goldstein DR, Thu D, Hollingsworth ZR, Collin F, Synek B, Holmans PA, Young AB, Wexler NS, Delorenzi M, Kooperberg C, Augood SJ, Faull RL, et al. (2006) Regional and cellular gene expression changes in human Huntington's disease brain. Hum Mol Genet 15:965–977.

Iwashita A, Muramatsu Y, Yamazaki T, Muramoto M, Kita Y, Yamazaki S, Mihara K, Moriguchi A, Matsuoka N (2007) Neuroprotective efficacy of the peroxisome proliferator–activated receptor delta-selective agonists *in vitro* and *in vivo*. J Pharmacol Exp Ther 320:1087–1096.

Kayed R, Head E, Sarsoza F, Saing T, Cotman CW, Necula M, Margol L, Wu J, Breydo L, Thompson JL, Rasool S, Gurlo T, Butler P, Glabe CG (2007) Fibril specific, conformation dependent antibodies recognize a generic epitope common to amyloid fibrils and fibrillar oligomers that is absent in prefibrillar oligomers. Mol Neurodegener 2:18.

Kreisler A, Gele P, Wiart JF, Lhermitte M, Destee A, Bordet R (2007) Lipid-lowering drugs in the *MPTP* mouse model of Parkinson's disease: fenofibrate has a neuroprotective effect, whereas bezafibrate and HMG-CoA reductase inhibitors do not. Brain Res 1135:77–84.

La Spada AR, Taylor JP (2010) Repeat expansion disease: progress and puzzles in disease pathogenesis. Nat Rev Genet 11:247–258.

Leone TC, Lehman JJ, Finck BN, Schaeffer PJ, Wende AR, Boudina S, Courtois M, Wozniak DF, Sambandam N, Bernal-Mizrachi C, Chen Z, Holloszy JO, Medeiros DM, Schmidt RE, Saffitz JE, Abel ED, Semenkovich CF, Kelly DP (2005) PGC-1alpha deficiency causes multi-system energy metabolic derangements: muscle dysfunction, abnormal weight control and hepatic steatosis. PLoS Biol 3:e101.

Lin MT, Beal MF (2006) Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. Nature 443:787–795.

Lowell BB, Spiegelman BM (2000) Towards a molecular understanding of adaptive thermogenesis. Nature 404:652–660.

Luthi-Carter R, Strand A, Peters NL, Solano SM, Hollingsworth ZR, Menon AS, Frey AS, Spektor BS, Penney EB, Schilling G, Ross CA, Borchelt DR, Tapscott SJ, Young AB, Cha JH, Olson JM (2000) Decreased expression of striatal signaling genes in a mouse model of Huntington's disease. Hum Mol Genet 9:1259–1271.

MacDonald ME, Ambrose CM, Duyao MP, Myers RH, Lin C, Srinidhi L, Barnes G, Taylor SA, James M, Groot N, MacFarlane H, Jenkins B, Anderson MA, Wexler NS, Gusella JF, Bates GP, Baxendale S, Hummerich H, Kirby S, North M, et al. (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72:971–983.

McGill JK, Beal MF (2006) PGC-1alpha, a new therapeutic target in Huntington's disease? Cell 127:465–468.

Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstråle M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, et al. (2003) PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet 34:267–273.

Muchowski PJ, Ning K, D'Souza-Schorey C, Fields S (2002) Requirement of an intact microtubule cytoskeleton for aggregation and inclusion body formation by a mutant huntingtin fragment. Proc Natl Acad Sci USA 99:727–732.

Nance MA (1997) Clinical aspects of CAG repeat diseases. Brain Pathol 7:881–900.

Narkar VA, Downes M, Yu RT, Embler E, Wang YX, Banayo E, Mihaylova MM, Nelson MC, Zou Y, Juguilon H, Kang H, Shaw RJ, Evans RM (2008) AMPK and PPARdelta agonists are exercise mimetics. Cell 134:405–415.

Puigserver P, Spiegelman BM (2003) Peroxisome proliferator–activated receptor-gamma coactivator 1 alpha (PGC-1 alpha): transcriptional coactivator and metabolic regulator. Endocr Rev 24:78–90.

Puigserver P, Wu Z, Park CW, Graves R, Wright M, Spiegelman BM (1998) A cold-inducible coactivator of nuclear receptors linked to adaptive thermogenesis. Cell 92:829–839.

Riley BE, Orr HT (2006) Polyglutamine neurodegenerative diseases and regulation of transcription: assembling the puzzle. Genes Dev 20:2183–2192.

NOTES

Ross CA, Thompson LM (2006) Transcription meets metabolism in neurodegeneration. Nat Med 12:1239–1241.

Ross CA, Becher MW, Colomer V, Engelender S, Wood JD, Sharp AH (1997) Huntington's disease and dentatorubral-pallidoluysian atrophy: proteins, pathogenesis and pathology. Brain Pathol 7:1003–1016.

Rubinsztein DC (2006) The roles of intracellular protein-degradation pathways in neurodegeneration. Nature 443:780–786.

Russell LK, Mansfield CM, Lehman JJ, Kovacs A, Courtois M, Saffitz JE, Medeiros DM, Valencik ML, McDonald JA, Kelly DP (2004) Cardiac-specific induction of the transcriptional coactivator peroxisome proliferator-activated receptor gamma coactivator-1alpha promotes mitochondrial biogenesis and reversible cardiomyopathy in a developmental stage-dependent manner. Circ Res 94:525–533.

Saudou F, Finkbeiner S, Devys D, Greenberg ME (1998) Huntingtin acts in the nucleus to induce apoptosis but death does not correlate with the formation of intranuclear inclusions. Cell 95:55–66.

Schilling G, Becher MW, Sharp AH, Jinnah HA, Duan K, Kotzuk JA, Slunt HH, Ratovitski T, Cooper JK, Jenkins NA, Copeland NG, Price DL, Ross CA, Borchelt DR (1999) Intranuclear inclusions and neuritic aggregates in transgenic mice expressing a mutant N-terminal fragment of huntingtin. Hum Mol Genet 8:397–407.

Tsunemi T, Ashe TD, Morrison BE, Soriano KR, Au J, Roque RA, Lazarowski ER, Damian VA, Masliah E, La Spada AR (2012) PGC-1alpha rescues Huntington's disease proteotoxicity by preventing oxidative stress and promoting TFEB function. Sci Transl Med 4:142ra197.

Wang L, Liu J, Saha P, Huang J, Chan L, Spiegelman B, Moore DD (2005) The orphan nuclear receptor SHP regulates PGC-1alpha expression and energy production in brown adipocytes. Cell Metab 2:227–238.

Wang YX, Zhang CL, Yu RT, Cho HK, Nelson MC, Bayuga-Ocampo CR, Ham J, Kang H, Evans RM (2004) Regulation of muscle fiber type and running endurance by PPARdelta. PLoS Biol 2:e294.

Weydt P, Pineda VV, Torrence AE, Libby RT, Satterfield TF, Lazarowski ER, Gilbert ML, Morton GJ, Bammler TK, Strand AD, Cui L, Beyer RP, Easley CN, Smith AC, Krainc D, Luquet S, Sweet IR, Schwartz MW, La Spada AR (2006) Thermoregulatory and metabolic defects in Huntington's disease transgenic mice implicate PGC-1alpha in Huntington's disease neurodegeneration. Cell Metab 4:349–362.

Genetic Analysis of Schizophrenia and Bipolar Disorder Reveals Polygenicity But Also Suggests New Directions for Molecular Interrogation

Benjamin M. Neale, PhD[1,2] and Pamela Sklar, MD, PhD[3]

[1]Analytical and Translational Genetics Unit
Psychiatric and Neurodevelopmental Genetics Unit
Department of Medicine, Massachusetts General Hospital
Harvard Medical School
Boston, Massachusetts

[2]Stanley Center for Psychiatric Research
Program in Medical and Population Genetics
Broad Institute of Harvard and MIT
Cambridge, Massachusetts

[3]Division of Psychiatric Genomics
Department of Psychiatry,
Icahn School of Medicine at Mount Sinai
New York, New York

NOTES

Introduction

Schizophrenia and bipolar disorder are among the most debilitating psychiatric illnesses and represent a tremendous public health burden. Nearly a century ago, Emil Kraepelin delineated two forms of mental illness as "dementia praecox" (now called schizophrenia) and "manic-depressive illness" (now called bipolar disorder). This dichotomy has been one of the founding principles of modern Western psychiatry, but in recent years, analysis of new genetic data is leading to its reexamination. In defining "dementia praecox," Kraepelin reviewed its apparently familial nature, noting a high degree of sibling sharing as well as parent offspring sharing, albeit to a lesser extent (Kraepelin, 1919). Thus, even a century ago, researchers were observing inheritance in families for severe mental illness.

These early views of schizophrenia and bipolar disorder as inherited disorders have been refined by almost one hundred years of twin and family studies. These studies demonstrated substantial heritability (the proportion of disease liability due to genetic factors) for both schizophrenia (estimates ranging from 60% to 90%) and bipolar disorder (estimates ranging from 60% to 80%) and showed that these disorders can co-occur in families (Berrettini, 2000; McGuffin et al., 2003; Sullivan et al., 2003; Lichtenstein et al., 2009). The strong and consistent evidence for high heritability suggested that disease genes might be identified using genetic approaches. Linkage, which is one of the earliest methods of genetic analysis, works by scanning the genome in search of regions that are shared by family members who are affected by the disease under study. This development spurred many linkage studies to search for schizophrenia and bipolar disorder regions and the genes within them (Levinson et al., 2003; Lewis et al., 2003; and Segurado et al., 2003). However, linkage is effective only when there is limited "locus heterogeneity" (e.g., if the genetic variation that influences schizophrenia is restricted to a few regions of the genome) or when there are large pedigrees with a nearly Mendelian cause (i.e., a genetic variant is nearly sufficient to cause disease in all affected members of the pedigree). To date, for schizophrenia and bipolar disorder, linkage has met with no clear success despite the meta-analysis of thousands of samples. These findings indicate that risk variants are not fully causal and that many regions in the genome are likely relevant to both schizophrenia and bipolar disorder.

The past decade of human genetics research has seen a staggering technological revolution in our ability to gather information about the genome. The SNP Consortium (Sachidanandam et al., 2001; Thorisson and Stein, 2003) and International HapMap project (International HapMap Consortium, 2003, 2005, 2007) created a catalog of common DNA variations and characterized the genome-wide patterns of linkage disequilibrium (LD). LD is a term used to describe the correlation between genetic variations, i.e., two variants are said to be in LD if the genotypes correlate. The HapMap project, in particular, was a central community-wide resource that made genotype data available for individuals from multiple ethnicities, providing information on allele frequency and LD across different continental populations. These large-scale evaluations of genetic variation also laid the groundwork for genome-wide association studies (GWAS) by providing a sufficiently comprehensive set of genetic markers to effectively test genome-wide.

The most pervasive type of genetic marker is the single nucleotide polymorphism (SNP), which is a single base change in the DNA sequence. GWAS enables a systematic and unbiased population-based evaluation of individual DNA variants for association with disease across the genome. In GWAS, the classes of genetic variations best explored have been those that are common in the population (common genetic variants, typically with minor allele frequency > 1–5%, depending on the study), and this technique has been successfully applied across a wide range of complex traits. The array technology used for GWAS also revealed copy number variation (large deletions or duplications of DNA from individual chromosomes), which has been shown to confer risk of several psychiatric illnesses including autism, schizophrenia, bipolar disorder, and attention deficit hyperactivity disorder (GAIN Collaborative Research Group et al., 2007; International Schizophrenia Consortium, 2008; Williams et al., 2010; Levy et al., 2011; Sanders et al., 2011; Malhotra and Sebat, 2012; Ramos-Quiroga et al., 2014; Rees et al., 2014; Stefansson et al., 2014; Szatkiewicz et al., 2014). Beyond GWAS, next-generation sequencing technologies have been developed that enable the discovery of rare (< 1% minor allele frequency) and private variation (effectively specific to individuals or families), thus extending the application of association tests into this range.

In general, the results from the current GWAS and sequencing studies of schizophrenia and bipolar disorder clearly show that a great many genetic variants influence the risk of schizophrenia and

NOTES

bipolar disorder. Both common and rare variants contribute to the genetic architecture of these diseases, and most effect sizes are small to modest, necessitating large-scale genetic studies to robustly identify novel risk factors. The identification of these genetic variants and the interpretation of the biological consequences of said variants will form the basis for new insights into the pathogenic processes that underlie schizophrenia and bipolar disorder.

Many Common DNA Variants Play a Large Role in Schizophrenia and Bipolar Disorder

The first GWAS of common variants in schizophrenia and bipolar disorder identified only a small handful of genome-wide significant loci (GAIN Collaborative Research Group et al., 2007; O'Donovan et al., 2008; International Schizophrenia Consortium et al., 2009). This limited initial success was driven mainly by the small effect size each variant is likely to have (< 1.5-fold influence on risk) and the comparatively modest sample sizes (< 10,000 individuals). In spite of a relative paucity of strongly associated loci (compared with some other traits, like those of Crohn's disease [Barrett et al., 2008] or age-related macular degeneration [Klein et al., 2005]), one of the first large-scale GWAS of schizophrenia demonstrated a myriad of DNA variants whose effects are too small to detect individually, but when summed together clearly contributed to schizophrenia risk (International Schizophrenia Consortium et al., 2009). This GWAS provided the first molecular evidence that schizophrenia is highly polygenic, and subsequently, it has been widely validated in many independent samples.

The many risk variants, when summed in this manner, essentially form a polygenic risk score (PRS). That is, based on a person's genotype, it is possible to count the number of risk alleles he or she has and to use that count to predict risk for diseases such as schizophrenia. This prediction can be improved by weighting each variant's contribution to the score based on the strength of association between that allele and the disease outcome, so that alleles that show larger effect sizes are counted more heavily in creating the score. Based on part of the early schizophrenia GWAS sample, such PRSs were then used to predict the risk of disease in an independent subsample. This procedure consistently demonstrates a minimal predictive ability but strong evidence for the combined role of common variation. Ensuring that there is no overlap between the discovery sample (the sample used to estimate the genetic effects across the genome) and the testing sample (the sample used to evaluate the predictive validity of the PRS) is essential to avoid false-positive associations and overestimation of the predictive validity of such a score. The PRS from the early GWAS explained only a small amount of the variance in liability to becoming ill with schizophrenia. At present, the PRS score is not sufficiently specific for clinical use. One of the predictions from the work suggested that increasing sample size would increase the explanatory power of such scores because it would more accurately estimate the true effect size of these genetic variants.

Not only did the schizophrenia PRS predict schizophrenia, but perhaps more surprisingly, it also showed predictive ability for bipolar disorder, implying that these two disorders share many more genetic risk factors than had been expected. In order to quantitate how much genetic overlap exists between schizophrenia and bipolar disorder (and four other majors forms of psychopathology), genome-wide complex trait analysis (GCTA) was performed (Yang et al., 2011). The basic principle behind GCTA is that if a trait has a genetic component, then people who are more phenotypically similar (e.g., both have schizophrenia) will tend to be more genetically similar (i.e., they will tend to share risk alleles for the disease in question). This framework enables not only the assessment of evidence for heritability from common DNA markers across the genome (rather than using twin and family relationships to tease out genetic contribution to disease) but also a multivariate approach for comparing different diseases. Consistent with the polygenic prediction work, when schizophrenia and bipolar disorder were analyzed jointly using GCTA, there was a substantial overlap in their genetic basis, with a genetic correlation of ~0.7 (Cross-Disorder Group of the Psychiatric Genomics Consortium et al., 2013). This genetic overlap highlights that there may be biological pathways that do not respect the traditional clinical boundaries but rather confer risk of both schizophrenia and bipolar disorder.

In the past few years, there has been a massive increase in the sample sizes for schizophrenia GWAS owing to the efforts of the Psychiatric Genomics Consortium and others (Ripke et al., 2013). As sample sizes have increased, the number of identified loci has increased as expected, most recently with 108 risk loci identified at genome-wide significance in a sample of more than 36,000 cases (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). Furthermore, as predicted, the specificity of the PRS also improved, in line with the initial PRS work, though it remains of limited

NOTES

clinical utility. As with other complex traits, the effect size of the genetic variants in these regions is quite modest (odds ratios ≈ 1.2). In most loci, the strongest associated variant is noncoding, consistent with the underlying causal alleles having a regulatory impact on disease. Furthermore, many genome-wide significant loci harbor multiple genes, any of which could be driving the association. A number of these loci contain genes that code for proteins supporting several prior biological hypotheses, such as *DRD2* (the D2 subtype of the dopamine receptor, thought to be the antipsychotic drug target); *GRM3*, *GRIN2A*, *SRR*, and *GRIA1* (involved in glutamatergic neurotransmission and synaptic plasticity); and *CACNA1C*, *CACNB2*, and *CACNA1I* (calcium-channel signaling). Because the majority of risk variants are found in noncoding regions, the precise biological mechanisms will be harder to uncover; however, many will be regulatory in nature.

At a global level, a series of analyses examining different biological annotations has been performed to evaluate whether further insights could be gleaned from these results. One of the most important sources of genomic annotations is the ENCODE/Roadmap Epigenomics Project (www.roadmapepigenomics.org) (ENCODE Project Consortium, 2011, 2012). This international collaborative effort is focused on measuring different functional and regulatory features of the genome. These features range from chromatin state (i.e., how exposed DNA is in the cell, which can be assayed through DNase I hypersensitivity) to histone modification (which relates to how DNA is bound and packaged in the cell) to DNA methylation (which may play a role in epigenetic regulation of gene expression, as a mechanism for silencing genes). These annotations were leveraged in the analysis of the most recent schizophrenia GWAS to demonstrate excess association in regions containing neuronal enhancers, as well as immune enhancers, an association that persists even after controlling for neuronal enhancers (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). Similarly, GWAS loci in schizophrenia are enriched for variants that influence gene expression (expression quantitative trait loci [eQTLs]), some of which are located in cis-regulatory elements, including promoters (Richards et al., 2012). These global analyses suggest that information on functional annotations can be used to prioritize genes in a risk locus for biological follow-up. One key example has been *CACNA1C*, which encodes the alpha subunit of the L-type calcium channel. In *CACNA1C*, the schizophrenia risk variant is associated with transcriptional regulation in the brain and is positioned within an enhancer sequence that physically interacts through chromosome loops with the promoter region of the gene (Roussos et al., 2014). This example demonstrates one paradigm for moving from DNA variant to biology, although many others are expected to emerge.

Available GWAS of bipolar disorder have used smaller sample sizes and identified approximately 10 genome-wide significant loci (Ferreira et al., 2008; Sklar et al., 2008; Chen et al., 2013; Green et al., 2013; Mendenhall et al., 2013). Explorations of the overlap of bipolar disorder–associated loci demonstrate that many, although not all, have shared effects in schizophrenia.

Rare *De Novo* and Inherited Variation

Copy number variants (CNVs) were the first rare variants found to be associated with psychiatric illness because they were large, easily detectable, and thus amenable to a wide variety of microarray technologies, including GWAS. A large body of work has shown that the rates of inherited and *de novo* (newly arising) CNVs are elevated in schizophrenia and, to a lesser extent, bipolar disorder. Many CNVs confer high risks (2.1–49.5), but none are determinative, and several genomic regions are frequent targets (International Schizophrenia Consortium, 2008; Malhotra and Sebat, 2012; Rees et al., 2014). Notably, these CNVs can produce a wide variety of neuropsychiatric phenotypes (most commonly, autism spectrum disorder, learning disability, and epilepsy) and are enriched for genes involved in synaptic processes and neuronal development (Kirov et al., 2012).

Beyond the GWAS of schizophrenia and bipolar disorder, the advent of next-generation sequencing technologies has enabled the assessment of rare and *de novo* variation for novel risk factors at the single-base level. The analysis of rare variation is in many ways more challenging than GWAS. For rare variation, case control study designs are one of the primary approaches, much in the same manner as GWAS. In contrast to GWAS, a direct test of each variant is effectively impossible because the number of copies of any allele is quite small. To overcome the limited power of testing individual rare variants, it is necessary to sensibly group them together to identify risk factors (Li and Leal, 2008; Price et al., 2010; Neale et al., 2011; Li et al., 2013). For the coding region, grouping variation is comparatively straightforward, as the gene is a natural analytic unit. Grouping together genetic variation outside the coding region is more challenging as our ability to functionally annotate noncoding genetic variation

NOTES

remains limited compared with the coding region, and currently appreciated functional units are typically too small to encompass a sufficient number of variants to perform a test (Zuk et al., 2014).

Thus far, the majority of exome sequencing completed for psychiatric illness has been for schizophrenia. The largest effort to date is an exome sequencing project of 2536 schizophrenia cases and 2543 controls in Sweden (Purcell et al., 2014). Notably, no single gene or rare variant has been associated with the disease, beyond chance expectation. Perhaps unexpectedly, however, a high degree of polygenicity was also observed for rare variants, meaning that rare variants scattered across a large number of genes likely influence the risk of schizophrenia. In many ways, this work echoes the earlier GWAS findings, with the implication that expanded sample sizes may yield significant loci and variants as they have for GWAS. Even though no single gene was identified, mutations predicted to disrupt gene function were found in sets of genes implicated in the CNV and GWAS studies described above. These sets included voltage-gated calcium channels and the signaling complex formed by the activity-regulated cytoskeleton-associated scaffold protein of the postsynaptic density.

The other major way to discover rare variation that increases the risk of disease is to search for *de novo* mutations by sequencing both parents and an affected child (also termed a proband), seeking to identify newly mutated DNA variants found only in the child. This approach has been successfully applied to severe, single-gene phenotypes found in intellectual disability and Kabuki syndrome to map novel risk genes (Ng et al,. 2010; Veltman and Brunner, 2012). The motivations for searching for *de novo* mutations that influence risk are based on the low background rate of such mutations (approximately one per offspring in the coding region) and the reduced fecundity that has been observed for individuals diagnosed with schizophrenia (Power et al., 2013).

More than 900 schizophrenia patients and their parents' exomes or coding regions have been sequenced to identify *de novo* mutations that might exert a strong influence on the risk of disease (Girard et al., 2011; Xu et al., 2012; Gulsuner et al., 2013; Fromer et al,. 2014; McCarthy et al., 2014). Even with this number of trios, however, few examples exist in which *de novo* mutations occur more than once in the same gene. In contrast to intellectual disability and autism, for which definitive genes have been identified (the discovery of which was driven by highly penetrant alleles), the results of the work on schizophrenia suggest few such strong-acting alleles. However, there is evidence for an enrichment of mutations in sets of genes, consistent with the model that the set of *de novo* mutations identified represents a mixture of variants that confer risk as well as those that are simply background events (Fromer et al., 2014). To distinguish the risk-conferring genes, here too it will be necessary to further expand the sample size.

Convergence on Pathways and Implications for Neurobiology

For schizophrenia and bipolar disorder, strong evidence favors a genetic component to risk. Overall, emerging evidence favors some shared genetic risk across the allele frequency spectrum, from rare to common, although the overlap is far from complete (Fig. 1). As described earlier, rare and common variants are enriched in several synaptic components, including calcium-channel subunits and postsynaptic elements. In fact, multiple independent lines of genetic data point to voltage-gated calcium channels. For example, common variation in the pore-forming alpha subunit *CACNA1C* is significantly associated with both schizophrenia and bipolar disorder, whereas variation in *CACNB2* is significantly associated with schizophrenia and at 10^{-4} in bipolar disorder (Psychiatric GWAS Consortium Bipolar Disorder Working Group, 2011; Ripke et al., 2013). As a set, voltage-gated calcium-channel genes were found to be enriched in rarer variants in the Swedish exome study (Purcell et al., 2014). Furthermore, in the severe Mendelian disorder Timothy syndrome, which results from single-base mutations in *CACNA1C* (the subunit most associated with schizophrenia), cases often present with autism spectrum phenotypes (Splawski et al., 2005). Similarly, multiple independent lines of genetic evidence converge on postsynaptic components and glutamate signaling. In the Psychiatric Genomics Consortium schizophrenia GWAS, genomic loci containing several NMDA and AMPA receptor subunits were associated, and in the *de novo* CNV and sequencing studies, postsynaptic density components such as *DLG1* and *DLG2* were found to be enriched.

Tying these observations together, we know that calcium-mediated signaling has an important role in neuronal differentiation by regulating axonal growth and guidance, and this process is also controlled by glutamate signaling (Rosenberg and Spitzer, 2011). Thus, future investigations should explore the effects of disease variants in these groups of genes on neuronal development and the inefficient neuronal circuitry observed in schizophrenia. Intriguing evidence is also beginning to converge on several

NOTES

neurodevelopmental genes, such as *KCTD13*, the gene encoding the polymerase delta-interacting protein 1. Although previously not a strong biological candidate, this gene is found in a region with significant common variants associated with schizophrenia and lies in a schizophrenia and autism–associated duplication on chromosome 16p11.2. Remarkably, the overexpression or reduction of *KCTD13* mRNA in zebrafish produces significant changes in head size, and *Kctd13* knockdown in the embryonic mouse brain decreases neurogenesis (Golzio et al., 2012). A further intriguing example is the fragile X mental retardation protein, FMRP, encoded by the *FMR1* gene, that regulates translation and is needed at synapses for normal glutamate receptor signaling and neurogenesis (Callan and Zarnescu, 2011) as well as being a common cause of mental impairment. Rare disruptive mutations in FMR1 such as nonsense, essential splice site, or frameshift mutations are enriched in schizophrenia cases.

A natural question for genetics is the extent to which this work adds to our understanding of disease, as the emerging biological themes are relevant to major systems that have previously been implicated. However, prior to the genetics work, calcium-channel signaling and abnormalities were not heavily investigated as a pathogenic mechanism in schizophrenia or bipolar disorder. Similarly, *KCTD13* has emerged as a novel candidate for follow-up functional characterization. Furthermore, genetic risk is consistent with these dysfunctions playing an etiological role in schizophrenia and bipolar disorder, aiding in the resolution of whether these dysfunctions are pathogenic or sequelae of the disease process. These genetic data signal a sea change, in that there are now multiple avenues of statistically confident genetic observations implicating specific biological processes. An important limitation of these studies is, of course, that they have not yet identified the precise alleles and, in some cases, the precise genes to target in the future. Although most evidence currently involves schizophrenia, we expect similar findings to emerge for bipolar disorder as sample sizes increase. The individual genetic effects acting on these diseases can broadly be described as spanning the allele frequency spectrum, with generally modest effect sizes, and strongly suggest multiple complex biological processes that are relevant to disease and have relevance for downstream neurobiological experiments. Across the population, many different combinations of alleles—some rare and some common—will contribute to the ultimate phenotype.

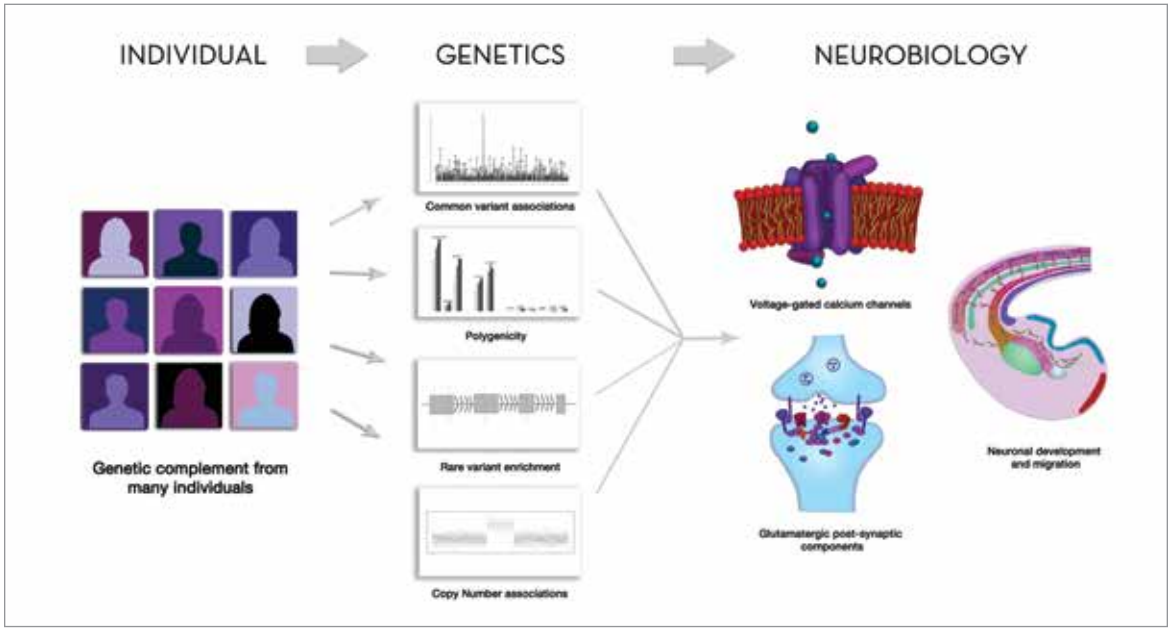


Figure 1. Moving from individuals to genetics to neurobiology: convergence of systems and pathways. Combining genetic information in many forms from large numbers of individuals can be used to identify specific genomic elements that contribute to disease. When integrated, these elements point to abnormalities in voltage-gated calcium channels, postsynaptic proteins, and neurodevelopmental molecules. Illustrations of common variant associations and polygenicity were adapted from International Schizophrenia Consortium et al. (2009), their Figs. 1 and 2, and the illustration of neuronal development and migration was adapted from Marin et al. (2010), their Fig. 5.

NOTES

Next Steps

The genetic analysis of schizophrenia, bipolar disorder, and other mental illnesses is yielding robustly significant genetic variants. These first unequivocal findings demonstrate that large-scale, unbiased screens of genetics can identify novel risk factors. However, these results are only the beginning of a long journey toward understanding the underlying biological processes involved in such diseases.

One prevailing paradigm for biological investigation proceeds in a gene-by-gene fashion, often necessitated by the difficulty and cost of the experiments. Unfortunately, this paradigm generally dictates that only variants conferring strong risk for a disease can be investigated. For schizophrenia, there are no such variants; rather, there is a plethora of variants that confer more moderate risk that need to be investigated. Thus, we will need to adapt methods that allow multiple genes and variants to be studied simultaneously in a more global, unbiased manner. The GWAS results also suggest a substantial role for regulatory variation in the pathogenesis of disease. To gain insight into how these regulatory variants influence risk, we will need to produce comprehensive maps of genomic gene expression and regulatory regions, such as enhancers and promoters in human brain tissue as well as in individual human neuronal subtypes. These efforts are gaining traction in several consortium projects, such as the CommonMind Consortium (commonmind.org), the Lieber Institute for Brain Development (www.libd.org), and the PsychENCODE project (psychencode.org). Given the limited availability of brain tissue, induced pluripotent stem cell–derived neuronal cell lines may provide another important resource for characterizing gene expression and regulatory regions (Brennand et al., 2011). Furthermore, these neuronal cell lines may form the basis for small-molecule screens to aid in the development of novel therapeutics. Integrative approaches that focus on developing biological networks from diverse sets of data can help focus attention on key biological drivers (Schadt and Bjorkegren, 2012). There are, or eventually will be, large-scale catalogues of gene expression, proteomics, protein interaction, drug interactions, and other data that will set the course for a more integrative biological approach.

Understanding how genetic variation influences gene regulation across developmental time points and in response to environmental stimuli is one of the key challenges for translating genetic discoveries into actionable biological hypotheses that can power a new round of therapeutic development. Fortunately, the current project of identifying genetic loci through GWAS and sequencing is moving forward at a rapid pace. This will lead to many more high-confidence loci that will more precisely pinpoint the most productive avenues for follow-up.

Acknowledgment

This paper was excerpted from a previously published article: Neale BM, Sklar P (2015) Genetic analysis of schizophrenia and bipolar disorder reveals polygenicity but also suggests new directions for molecular interrogation. Curr Opin Neurobiol 30:131–138.

References

Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, Brant SR, Silverberg MS, Taylor KD, Barmada MM, Bitton A, Dassopoulos T, Datta LW, Green T, Griffiths AM, Kistner EO, Murtha MT, Regueiro MD, Rotter JI, Schumm LP, et al. (2008) Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet 40:955–962.

Berrettini WH (2000) Are schizophrenic and bipolar disorders related? A review of family and molecular studies. Biol Psychiatry 48:531–538.

Brennand KJ, Simone A, Jou J, Gelboin-Burkhart C, Tran N, Sangar S, Li Y, Mu Y, Chen G, Yu D, McCarthy S, Sebat J, Gage FH (2011) Modelling schizophrenia using human induced pluripotent stem cells. Nature 473:221–225.

Callan MA, Zarnescu DC (2011) Heads-up: new roles for the fragile X mental retardation protein in neural stem and progenitor cells. Genesis 49:424–440.

Chen DT, Jiang X, Akula N, Shugart YY, Wendland JR, Steele CJ, Kassem L, Park JH, Chatterjee N, Jamain S, Cheng A, Leboyer M, Muglia P, Schulze TG, Cichon S, Nöthen MM, Rietschel M; Bipolar Genome Study (BiGS) Consortium, McMahon FJ, Farmer A, et al. (2013) Genome-wide association study meta-analysis of European and Asian-ancestry samples identifies three novel loci associated with bipolar disorder. Mol Psychiatry 18:195–205.

NOTES

Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, Perlis RH, Mowry BJ, Thapar A, Goddard ME, Witte JS, Absher D, Agartz I, Akil H, Amin F, Andreassen OA, Anjorin A, Anney R, Anttila V, Arking DE, et al. (2013) Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet 45:984–994.

ENCODE Project Consortium (2011) A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol 9:e1001046.

ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74.

Ferreira MA, O'Donovan MC, Meng YA, Jones IR, Ruderfer DM, Jones L, Fan J, Kirov G, Perlis RH, Green EK, Smoller JW, Grozeva D, Stone J, Nikolov I, Chambert K, Hamshere ML, Nimgaonkar VL, Moskvina V, Thase ME, Caesar S, et al. (2008) Collaborative genome-wide association analysis supports a role for *ANK3* and *CACNA1C* in bipolar disorder. Nat Genet 40:1056–1058.

Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P, Georgieva L, Rees E, Palta P, Ruderfer DM, Carrera N, Humphreys I, Johnson JS, Roussos P, Barker DD, Banks E, Milanova V, Grant SG, Hannon E, Rose SA, et al. (2014) *De novo* mutations in schizophrenia implicate synaptic networks. Nature 506:179–184.

GAIN Collaborative Research Group, Manolio TA, Rodriguez LL, Brooks L, Abecasis G, Collaborative Association Study of Psoriasis, Ballinger D, Daly M, Donnelly P, Faraone SV, International Multi-Center ADHD Genetics Project, Frazer K, Gabriel S, Gejman P; Molecular Genetics of Schizophrenia Collaboration, Guttmacher A, Harris EL, Insel T, Kelsoe JR; Bipolar Genome Study, et al. (2007) New models of collaboration in genome-wide association studies: the Genetic Association Information Network. Nat Genet 39:1045–1051.

Girard SL, Gauthier J, Noreau A, Xiong L, Zhou S, Jouan L, Dionne-Laporte A, Spiegelman D, Henrion E, Diallo O, Thibodeau P, Bachand I, Bao JY, Tong AH, Lin CH, Millet B, Jaafari N, Joober R, Dion PA, Lok S, et al. (2011) Increased exonic *de novo* mutation rate in individuals with schizophrenia. Nat Genet 43:860–863.

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.

Green EK, Hamshere M, Forty L, Gordon-Smith K, Fraser C, Russell E, Grozeva D, Kirov G, Holmans P, Moran JL, Shaun Purcell, Pamela Sklar, Michael J Owen, Michael C O'Donovan, Lisa Jones, Wellcome Trust Case Control Consortium, Ian R Jones, Nick Craddock (2013) Replication of bipolar disorder susceptibility alleles and identification of two novel genome-wide significant associations in a new bipolar disorder case-control sample. Mol Psychiatry 18:1302–1307.

Gulsuner S, Walsh T, Watts AC, Lee MK, Thornton AM, Casadei S, Rippey C, Shahin H, Consortium on the Genetics of Schizophrenia (COGS), PAARTNERS Study Group, Nimgaonkar VL, Go RC, Savage RM, Swerdlow NR, Gur RE, Braff DL, King MC, McClellan JM (2013) Spatial and temporal mapping of *de novo* mutations in schizophrenia to a fetal prefrontal cortical network. Cell 154:518–529.

International HapMap Consortium (2003) The International HapMap Project. Nature 426:789–796.

International HapMap Consortium (2005) A haplotype map of the human genome. Nature 437:1299–1320.

International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, et al. (2007) A second generation human haplotype map of over 3.1 million SNPs. Nature 449:851–861.

International Schizophrenia Consortium (2008) Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature 455:237–241.

International Schizophrenia Consortium, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460:748–752.

NOTES

Kirov G, Pocklington AJ, Holmans P, Ivanov D, Ikeda M, Ruderfer D, Moran J, Chambert K, Toncheva D, Georgieva L, Grozeva D, Fjodorova M, Wollerton R, Rees E, Nikolov I, van de Lagemaat LN, Bayés A, Fernandez E, Olason PI, Böttcher Y, et al. (2012) *De novo* CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. Mol Psychiatry 17:142–153.

Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J (2005) Complement factor H polymorphism in age-related macular degeneration. Science 308:385–389.

Kraepelin E, Barclay RM, Robertson GM (1919) Dementia præcox and paraphrenia. Edinburgh, UK: E & S Livingstone.

Levinson DF, Levinson MD, Segurado R, Lewis CM (2003) Genome scan meta-analysis of schizophrenia and bipolar disorder, part I: Methods and power analysis. Am J Hum Genet 73:17–33.

Levy D, Ronemus M, Yamrom B, Lee YH, Leotta A, Kendall J, Marks S, Lakshmi B, Pai D, Ye K, Buja A, Krieger A, Yoon S, Troge J, Rodgers L, Iossifov I, Wigler M (2011) Rare *de novo* and transmitted copy-number variation in autistic spectrum disorders. Neuron 70:886–897.

Lewis CM, Levinson DF, Wise LH, DeLisi LE, Straub RE, Hovatta I, Williams NM, Schwab SG, Pulver AE, Faraone SV, Brzustowicz LM, Kaufmann CA, Garver DL, Gurling HM, Lindholm E, Coon H, Moises HW, Byerley W, Shaw SH, Mesen A, et al. (2003) Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. Am J Hum Genet 73:34–48.

Li B, Leal SM (2008) Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. Am J Hum Genet 83:311–321.

Li B, Liu DJ, Leal SM (2013) Identifying rare variants associated with complex traits via sequencing. Curr Protoc Hum Genet Chapter 1:Unit 1.26.

Lichtenstein P, Yip BH, Bjork C, Pawitan Y, Cannon TD, Sullivan PF, Hultman CM (2009) Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. Lancet 373:234–239.

Malhotra D, Sebat J (2012) CNVs: harbingers of a rare variant revolution in psychiatric genetics. Cell 148:1223–1241.

Marin O, Valiente M, Ge X, Tsai LH (2010) Guiding neuronal cell migrations. Cold Spring Harb Perspect Biol 2:a001834.

McCarthy SE, Gillis J, Kramer M, Lihm J, Yoon S, Berstein Y, Mistry M, Pavlidis P, Solomon R, Ghiban E, Antoniou E, Kelleher E, O'Brien C, Donohoe G, Gill M, Morris DW, McCombie WR, Corvin A (2014) *De novo* mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. Mol Psychiatry 19:652–658.

McGuffin P, Rijsdijk F, Andrew M, Sham P, Katz R, Cardno A (2003) The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. Arch Gen Psychiatry 60:497–502.

Mendenhall EM, Williamson KE, Reyon D, Zou JY, Ram O, Joung JK, Bernstein BE (2013) Locus-specific editing of histone modifications at endogenous enhancers. Nat Biotechnol 31:1133–1136.

Neale BM, Rivas MA, Voight BF, Altshuler D, Devlin B, Orho-Melander M, Kathiresan S, Purcell SM, Roeder K, Daly MJ (2011) Testing for an unusual distribution of rare variants. PLoS Genet 7:e1001322.

Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, Lee C, Turner EH, Smith JD, Rieder MJ, Yoshiura K, Matsumoto N, Ohta T, Niikawa N, Nickerson DA, Bamshad MJ, et al. (2010) Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet 42:790–793.

O'Donovan MC, Craddock N, Norton N, Williams H, Peirce T, Moskvina V, Nikolov I, Hamshere M, Carroll L, Georgieva L, Dwyer S, Holmans P, Marchini JL, Spencer CC, Howie B, Leung HT, Hartmann AM, Möller HJ, Morris DW, Shi Y, et al. (2008) Identification of loci associated with schizophrenia by genome-wide association and follow-up. Nat Genet 40:1053–1055.

Power RA, Kyaga S, Uher R, MacCabe JH, Langstrom N, Landen M, McGuffin P, Lewis CM, Lichtenstein P, Svensson AC (2013) Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings. JAMA Psychiatry 70:22–30.

Price AL, Kryukov GV, de Bakker PI, Purcell SM, Staples J, Wei LJ, Sunyaev SR (2010) Pooled association tests for rare variants in exon-resequencing studies. Am J Hum Genet 86:832–838.

NOTES

Psychiatric GWAS Consortium Bipolar Disorder Working Group (2011) Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nat Genet 43:977–983.

Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P, O'Dushlaine C, Chambert K, Bergen SE, Kähler A, Duncan L, Stahl E, Genovese G, Fernández E, Collins MO, Komiyama NH, Choudhary JS, Magnusson PK, Banks E, Shakir K, et al. (2014) A polygenic burden of rare disruptive mutations in schizophrenia. Nature 506:185–190.

Ramos-Quiroga JA, Sanchez-Mora C, Casas M, Garcia-Martinez I, Bosch R, Nogueira M, Corrales M, Palomar G, Vidal R, Coll-Tane M, Bayés M, Cormand B, Ribasés M (2014) Genome-wide copy number variation analysis in adult attention-deficit and hyperactivity disorder. J Psychiatr Res 49:60–67.

Rees E, Walters JT, Chambert KD, O'Dushlaine C, Szatkiewicz J, Richards AL, Georgieva L, Mahoney-Davies G, Legge SE, Moran JL, Genovese G, Levinson D, Morris DW, Cormican P, Kendler KS, O'Neill FA, Riley B, Gill M, Corvin A; Wellcome Trust Case Control Consortium, et al. (2014) CNV analysis in a large schizophrenia sample implicates deletions at 16p12.1 and SLC1A1 and duplications at 1p36.33 and CGNL1. Hum Mol Genet 23:1669–1676.

Richards AL, Jones L, Moskvina V, Kirov G, Gejman PV, Levinson DF, Sanders AR, Molecular Genetics of Schizophrenia Collaboration, International Schizophrenia Consortium (ISC), Purcell S, Visscher PM, Craddock N, Owen MJ, Holmans P, O'Donovan MC (2012) Schizophrenia susceptibility alleles are enriched for alleles that affect gene expression in adult human brain. Mol Psychiatry 17:193–201.

Ripke S, O'Dushlaine C, Chambert K, Moran JL, Kahler AK, Akterin S, Bergen SE, Collins AL, Crowley JJ, Fromer M, Kim Y, Lee SH, Magnusson PK, Sanchez N, Stahl EA, Williams S, Wray NR, Xia K, Bettella F, Borglum AD, et al. (2013) Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nat Genet 45:1150–1159.

Rosenberg SS, Spitzer NC (2011) Calcium signaling in neuronal development. Cold Spring Harb Perspect Biol 3:a004259.

Roussos P, Mitchell AC, Voloudakis G, Fullard JF, Pothula VM, Tsang J, Stahl EA, Georgakopoulos A, Ruderfer DM, Charney A, Okada Y, Siminovitch KA, Worthington J, Padyukov L, Klareskog L, Gregersen PK, Plenge RM, Raychaudhuri S, Fromer M, Purcell SM, et al. (2014) A role for noncoding variant in schizophrenia. Cell Rep 9:1417–1429.

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, et al. (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409:928–933.

Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, Chu SH, Moreau MP, Gupta AR, Thomson SA, Mason CE, Bilguvar K, Celestino-Soper PB, Choi M, Crawford EL, Davis L, Wright NR, Dhodapkar RM, DiCola M, DiLullo NM, et al. (2011) Multiple recurrent *de novo* CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron 70:863–885.

Schadt EE, Bjorkegren JL (2012) NEW: network-enabled wisdom in biology, medicine, and health care. Sci Transl Med 4:115rv111.

Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014) Biological insights from 108 schizophrenia-associated genetic loci. Nature 511:421–427.

Segurado R, Detera-Wadleigh SD, Levinson DF, Lewis CM, Gill M, Nurnberger JI, Jr., Craddock N, DePaulo JR, Baron M, Gershon ES, Ekholm J, Cichon S, Turecki G, Claes S, Kelsoe JR, Schofield PR, Badenhop RF, Morissette J, Coon H, Blackwood D, et al. (2003) Genome scan meta-analysis of schizophrenia and bipolar disorder, part III: Bipolar disorder. Am J Hum Genet 73:49–62.

Sklar P, Smoller JW, Fan J, Ferreira MA, Perlis RH, Chambert K, Nimgaonkar VL, McQueen MB, Faraone SV, Kirby A, PIW de Bakker, MN Ogdie, ME Thase, GS Sachs, K Todd-Brown, SB Gabriel, C Sougnez, C Gates, B Blumenstiel, M Defelice, et al. (2008) Whole-genome association study of bipolar disorder. Mol Psychiatry 13:558–569.

NOTES

Splawski I, Timothy KW, Decher N, Kumar P, Sachse FB, Beggs AH, Sanguinetti MC, Keating MT (2005) Severe arrhythmia disorder caused by cardiac L-type calcium channel mutations. Proc Natl Acad Sci USA 102:8089–8096; discussion 8086–8088.

Stefansson H, Meyer-Lindenberg A, Steinberg S, Magnusdottir B, Morgen K, Arnarsdottir S, Bjornsdottir G, Walters GB, Jonsdottir GA, Doyle OM, Tost H, Grimm O, Kristjansdottir S, Snorrason H, Davidsdottir SR, Gudmundsson LJ, Jonsson GF, Stefansdottir B, Helgadottir I, Haraldsson M, et al. (2014) CNVs conferring risk of autism or schizophrenia affect cognition in controls. Nature 505:361–366.

Sullivan PF, Kendler KS, Neale MC (2003) Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Arch Gen Psychiatry 60:1187–1192.

Szatkiewicz JP, O'Dushlaine C, Chen G, Chambert K, Moran JL, Neale BM, Fromer M, Ruderfer D, Akterin S, Bergen SE, Kähler A, Magnusson PK, Kim Y, Crowley JJ, Rees E, Kirov G, O'Donovan MC, Owen MJ, Walters J, Scolnick E, et al. (2014) Copy number variation in schizophrenia in Sweden. Mol Psychiatry 19:762–773.

Thorisson GA, Stein LD (2003) The SNP Consortium website: past, present and future. Nucleic Acids Res 31:124–127.

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

Williams NM, Zaharieva I, Martin A, Langley K, Mantripragada K, Fossdal R, Stefansson H, Stefansson K, Magnusson P, Gudmundsson OO, Gustafsson O, Holmans P, Owen MJ, O'Donovan M, Thapar A (2010) Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. Lancet 376:1401–1408.

Xu B, Ionita-Laza I, Roos JL, Boone B, Woodrick S, Sun Y, Levy S, Gogos JA, Karayiorgou M (2012) *De novo* gene mutations highlight patterns of genetic and neural complexity in schizophrenia. Nat Genet 44:1365–1369.

Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM, de Andrade M, Feenstra B, Feingold E, Hayes MG, Hill WG, Landi MT, Alonso A, Lettre G, Lin P, Ling H, Lowe W, Mathias RA, Melbye M, Pugh E, et al. (2011) Genome partitioning of genetic variation for complex traits using common SNPs. Nat Genet 43:519–525.

Zuk O, Schaffner SF, Samocha K, Do R, Hechter E, Kathiresan S, Daly MJ, Neale BM, Sunyaev SR, Lander ES (2014) Searching for missing heritability: designing rare variant association studies. Proc Natl Acad Sci USA 111:E455–E464.

Inferring Causality and Functional Significance of Human Coding DNA Variants

Shamil R. Sunyaev, PhD

Genetics Division
Brigham & Women's Hospital
Harvard Medical School
Boston, Massachusetts

NOTES

Introduction

Sequencing technology enables the complete characterization of human genetic variation. Statistical genetics studies identify numerous loci linked or associated to phenotypes of direct medical interest. The major remaining challenge is to characterize functionally significant alleles that are causally implicated in the genetic basis of human traits. Here, I review three sources of evidence for the functional significance of human DNA variants in protein-coding genes. These include (1) statistical genetics considerations such as cosegregation with the phenotype, allele frequency in unaffected controls and recurrence; (2) *in vitro* functional assays and model organism experiments; and (3) computational methods for predicting the functional effect of amino acid substitutions. In spite of many successes of recent studies, functional characterization of human allelic variants remains problematic.

Large-scale sequencing projects have revealed the landscape of human genetic variation (1000 Genomes Project Consortium, 2010; Tennessen et al., 2012). Linkage and association studies identified a large number of loci involved in various human phenotypes. In spite of this spectacular progress, characterization of functionally significant human alleles causally involved in phenotypes (i.e., directly contributing to the biology of phenotypes) remains challenging.

The problem of establishing a causal relationship between a phenotype and a specific sequence variant arises at multiple levels (Table 1). It spans both Mendelian and complex trait genetics, even though many aspects of the problem and approaches to address them are different.

In the simplest case of a Mendelian monogenic trait unequivocally linked to a particular gene, the problem is in distinguishing between benign and pathogenic alleles in this gene. This creates a major bottleneck in clinical genetic diagnostics (Plon et al., 2008). Many allelic variants observed in genes of diagnostic importance remain classified as variants of unknown significance (VUSs).

For Mendelian phenotypes with unknown genetic background, sequencing studies now provide a powerful way to identify causal genes. Briefly, the strategy involves finding a gene where all or most patients carry functional variants that are not observed in multiple unaffected controls (Ng et al., 2010). Usually, all coding nonsynonymous variants and variants disrupting canonic splice sites are considered functional and other variants are ignored. Although this strategy generated many successes, it lacks power if sample sizes are small (only two or three patients available) or in the case of oligogenic phenotypes. Knowledge of functional significance of allelic variants would greatly empower sequencing studies aiming at mapping genes underlying Mendelian disorders.

Remarkable progress in sequencing technology now allows detecting *de novo* mutations using parent–child trio sequencing (Roach et al., 2010). This approach has been successfully applied to a number of Mendelian traits and to complex psychiatric phenotypes such as autism and schizophrenia (Roach et al., 2010; Xu et al., 2011; Neale et al., 2012; O'Roak et al., 2012; Sanders et al., 2012). Relatively small numbers of *de novo* mutations facilitate the analysis. On average, humans carry on the order of 100 *de novo* point mutations with only few (on average 1) of them coding (Nachman and Crowell, 2000; Kondrashov, 2003; Kong et al., 2012; Sun et al., 2012). However, these mutations are typically unique to individual patients. Therefore, it is impossible to use statistical approaches to infer their involvement in phenotypes in case of whole-exome or whole-genome sequencing experiments.

Naturally, the magnitude of the problem is amplified when considering variants involved in complex traits. Genome-wide association studies (GWAS) identified a multitude of common SNPs (single nucleotide polymorphisms) associated with human complex traits. However, most of these SNPs are not causal and simply tag causal alleles due to linkage disequilibrium (LD). LD greatly facilitates mapping but equally complicates pinpointing causal variants by statistical means because association signals of many variants are confounded. In many cases, even the identity of a causal gene, rather than a specific allele, is not known. The problem is exacerbated because most of GWAS peaks are in noncoding regions. Moreover, it is possible that multiple causal variants give rise to a single GWAS peak. A number of sequencing projects aiming at finding causal variants underlying GWAS peaks are ongoing. The dominant hypothesis is that the variants responsible for the observed associations are common. Scenarios where associations of common SNPs are caused by low-frequency variants or even by multiple rare variants have also been proposed (Dickson et al., 2010), although subsequent work suggested that such scenarios do not explain many GWAS peaks (Wray et al., 2011).

NOTES

Table 1. Importance of the functional analysis in various types of human genetics studies

Analysis of Mendelian traits				Analysis of rare variants in complex traits			
Interpretation of variants in previously mapped genes		Mapping genes by whole-genome/whole-exome sequencing		Analysis of rare variants in candidate genes		Mapping genes by whole-genome/whole-exome sequencing	
Uncharacterized variants not known to be *de novo*	*De novo* mutations	Segregating variants	*De novo* mutations	Rare variants	*De novo* mutations	Rare variants	*De novo* mutations
Analysis of the functional effect and causality is essential	Usually regarded as sufficient evidence of functionality	Functional analysis is not essential to map genes but can potentially increase power	Functional analysis is essential for isolated mutations; recurrence may provide a statistical argument in favor of functionality	Functional analysis was shown to increase power	Likely a sufficient evidence of functionality	Functional analysis was hypothesized to increase power	Functional analysis is essential for isolated mutations; recurrence may provide a statistical argument in favor of functionality

Examples of the functional characterization of variants underlying GWAS signals are still rare. One early example includes demonstration that a common variant creating a transcription factor binding site for the CCAAT/enhancer-binding protein alters the hepatic expression of the *SORT1* gene. This variant explains the corresponding GWAS signal for association with LDL-cholesterol (Musunuru et al., 2010). Fine-mapping studies have been reported recently for strong association signals within the human leukocyte antigen region (International HIV Controllers Study et al., 2010; Raychaudhuri et al., 2012). For common noncoding variants, analysis of intermediate molecular phenotypes related to transcriptional regulation such as mRNA expression (Stranger et al., 2007) and chromatin accessibility (McDaniell et al., 2010; Degner et al., 2012; Maurano et al., 2012) offers a potential way forward. These early studies on functional effects of common noncoding variants are outside of scope of this review.

A number of successful candidate gene-sequencing studies discovered associations of multiple rare coding variants with complex phenotypes (Cohen et al., 2004; Ahituv et al., 2007; Ji et al., 2008; Romeo et al., 2009; Johansen et al., 2010; Momozawa et al., 2011; Rivas et al., 2011; Bonnefond et al., 2012; Jordan et al., 2012; Kiezun et al., 2012). Ongoing whole-exome sequencing studies attempt an unbiased search for genes harboring multiple rare variants collectively associated with complex traits (Price et al., 2010). In the simplest form, this analysis detects an excess of rare coding variants in cases versus controls. The association signal is provided by functional variants, whereas neutral alleles are a source of noise masking the association signal. Again, functional significance of individual rare variants cannot be inferred by statistical means. In contrast to common variants, LD does not confound the signal. However, the association test for individual rare variants lacks statistical power, given that they are observed a handful of times (or even once) in the sample. The ability to discriminate between functional and neutral alleles would dramatically increase the potential of sequencing studies focusing on rare variants in complex traits. Several published studies demonstrated that highlighting functional variants using experimental (Romeo et al., 2009; Bonnefond et al., 2012) or computational approaches (Ahituv et al., 2007; Ji et al., 2008; MacArthur et al., 2012) increases the power of these studies.

Understanding the functional significance of human alleles is also of great importance for evolutionary and population genetics. Accurate inference of functional consequences of human DNA variants would help characterizing the role of natural selection in shaping population genetic variation (Tennessen et al., 2012).

Overall, medical genetics is interested in finding "pathogenic" mutations that causally influence traits of medical interest. Population genetics focuses on "deleterious" alleles that evolve under purifying selection. In contrast, functional analysis is focused

NOTES

on the "damaging" effect on molecular function. The rationale for this approach is that the effects on phenotypes and fitness must be mediated by the effects on molecular function, even though the converse is not necessarily true. The existence of many common loss-of-function variants in humans (MacArthur et al., 2012) and events of adaptive pseudogenization (Wang et al., 2006) clearly show that damaging alleles may be neutral or beneficial rather than deleterious. It is also feasible that most of human alleles that are subject to purifying selection have no detectable effects on medically relevant phenotypes in the current environment. However, most studies implicitly assume the strong relationship between the effects on molecular function, fitness, and phenotypes. For example, many computational methods for predicting the functional effects of human alleles are based on the inference of purifying selection from comparative genomics data.

Here, I review current strategies to infer causality and functional significance of human protein–coding DNA variants, including variants involved in Mendelian human traits and rare coding variants involved in complex phenotypes.

Inferring the Functional Significance of Missense Mutations Involved in Mendelian Phenotypes

As noted earlier, the problem of assigning functional significance to variants involved in Mendelian phenotypes arises both in the context of gene discovery and in the context of interpreting VUSs in known genes. The overwhelming majority of sequence variants causing Mendelian traits are coding. Among coding variants, "missense" changes are the most difficult to interpret (most of synonymous changes are benign, and most of nonsense or splice-site changes are damaging). Three potential strategies to infer causality and functional significance could be employed: (1) the strategy based on statistical genetics, (2) *in vitro* or *in vivo* experimental analysis, and (3) computational predictions based on evolutionary and structural considerations.

Statistical arguments

In some cases, purely statistical arguments can be employed in favor of the causal relationship between DNA variants and Mendelian traits. Importantly, the arguments discussed below are specific to Mendelian genetics and, in the most part, cannot be applied to variants underlying complex phenotypes.

Analysis of cosegregation of the DNA variant with the phenotype is probably the most accurate method for establishing causality by statistical means. However, at least five informative meioses are needed to support causality (Jordan et al., 2011), and sufficiently large pedigrees are usually unavailable. In addition, segregation analysis may be misleading if more than one rare variant is present in the locus and cosegregate with the phenotype.

Another important consideration is the analysis of allele frequency in unaffected controls. This analysis has been dramatically facilitated by large-scale sequencing efforts such as the 1000 Genomes Project (1000 Genomes Project Consortium, 2010) and Exome Sequencing Project (ESP) (Tennessen et al., 2012). Presence in healthy controls at appreciable frequency may reveal whether the allelic variant is a benign polymorphism segregating in the population, which will exclude the possibility that this variant is involved in the disease phenotype with high penetrance (this approach is obviously noninformative for variance of incomplete penetrance unless larger case–control study is pursued). Although it is easy to infer that the variant is benign (or, at least, not of high penetrance) if it is seen in a number of unaffected individuals, it is much less clear if its absence in multiple controls may serve as a strong support for the pathogenicity. Most importantly, for some genes such as *BRCA1* and *BRCA2*, the number of sequenced cases vastly exceeds the number of sequenced controls, making the analysis of allele frequency in unaffected controls noninformative. Next, differences in global and even local ancestry may complicate conclusions because many rare variants are specific to individual human populations. Also, ESP contains data on individuals with various diseases, so not all sequenced individuals should be automatically assumed to be unaffected.

Even in the simplest possible case of a variant observed in a single patient with a dominant phenotype absent in a panel of ideally ancestry-matched control subjects, the number of control subjects should be very large.

At the first glance, population genetics supports the use of moderate numbers of controls. Under the standard model of a constant size population with no natural selection, the chance that a variant observed in a patient will not be seen in n normal controls is $1/(n + 1)$ (Mitchell et al., 2005) This suggests that if the variant is not found in 100 controls, then the chance for the mutation to have no phenotypic effect is < 1%. Therefore, absence in a moderate number of controls would support pathogenicity. The following

NOTES

factors suggest that this is a stark underestimate for human populations: (1) human population growth, which has resulted in an excess of rare alleles, (2) selection against moderately deleterious alleles, and (3) human migrations, which have resulted in rare alleles not seen in multiple controls (Sunyaev et al., 2000; Marth et al., 2004; Williamson et al., 2005; Kryukov et al., 2007, 2009; Boyko et al., 2008; Li et al., 2010). As seen from Figure 1, a more complex population genetics model incorporating population growth and natural selection (Kryukov et al., 2009) but not migration predicts that there is > 1% chance that a benign variant observed in a single patient would not be detected in as many as 10,000 controls. Taking into account the effects of migration would likely make this number even higher. Therefore, the sole observation of the absence in multiple unaffected controls is insufficient to convincingly imply functional significance of a sequence variant.

In some cases, the evidence for pathogenicity of specific mutations can be provided by the observation of recurrence. For example, independent occurrence of two exactly same mutations has been observed in Baraitser–Winter syndrome (Rivière et al., 2012a). Three different mutations in the same codon have been reported in the analysis of the Myhre syndrome (Le Goff et al., 2011), strongly suggesting the functional importance of this particular amino acid position.

A growing number of publications (Heinzen et al., 2012; Rivière et al., 2012a, b; Van Houdt et al., 2012) report *de novo* mutations as evident from parent–child trio sequencing. The observation of *de novo* mutation in a gene known to be involved in the phenotype (i.e., a gene under an independently reported linkage peak or a gene with multiple *de novo* mutations in other families) is highly informative

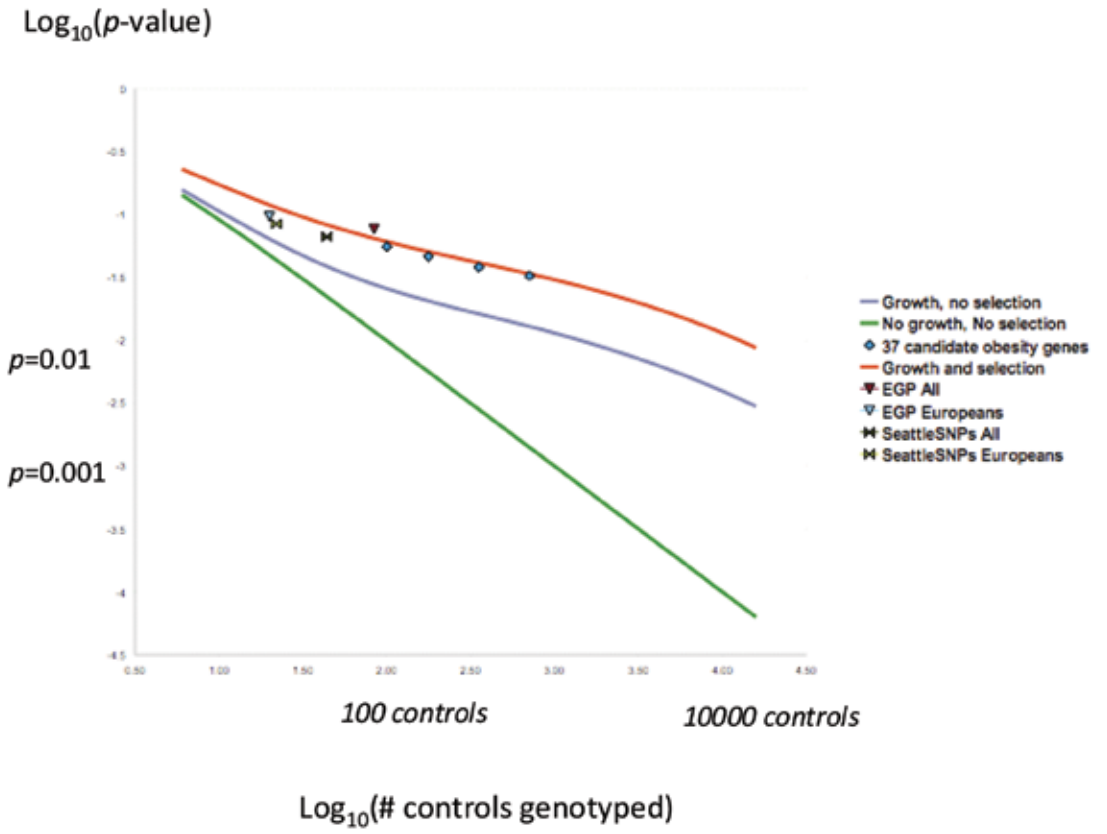


Figure 1. The probability that a nonpathogenic variant observed in a single patient would not be observed in multiple controls. Log–log scale plot is shown for theoretical model assuming constant population and no natural selection (Ahituv et al., 2007) (green line); a population genetic model assuming recent population growth and no natural selection (Kryukov et al., 2009) (blue line); and a population genetic model that incorporates both population growth and natural selection (Kryukov et al., 2009) (red line). Results of theoretical models are shown together with estimates based on real data obtained by resampling from three available systematic resequencing datasets: the Environmental Genome Project dataset, the Seattle SNP dataset (Livingston et al., 2004), and the Obesity Sequencing Study dataset (Ahituv et al., 2007).

NOTES

about the functional significance of the mutation. Indeed, the rate of point mutations in humans is on the order of 10^{-8} per nucleotide per generation and approximately 10^{-5} per protein-coding gene per generation (Nachman and Crowell, 2000; Kondrashov, 2003; Roach et al., 2010; Kong et al., 2012; Sun et al., 2012). Therefore, it is unlikely that a *de novo* mutation unrelated to the phenotype is observed in a known gene. The situation is different, however, in the analysis of whole-exome or whole-genome sequencing without the knowledge of causal genes. Although *de novo* mutations can be considered excellent candidates, especially for dominantly inherited traits, independent functional validation is usually required.

Experimental evidence

Direct experimental functional analysis is a highly laborious but a highly convincing method to study the effect of human allelic variants. Experimental approaches include the analysis of protein expression and localization, *in vitro* functional assays, and genetic manipulation on model organisms. The enthusiasm for direct experimental methods should be accompanied by a cautionary note that specific aspects of molecular function analyzed using *in vitro* assays in some cases may be unrelated to the phenotype, and the effects of mutations on model organisms sometimes may be uninformative about the human condition.

In many cases, missense mutations result in changes of protein expression and localization. Some recent studies relied on immunostaining to assess effects of individual human alleles (Boileau et al., 2012; Wortman et al., 2012). Testing other aspects of protein function requires development of specific functional assays. Phosphorylation assays can be applied for proteins involved in signaling. A recent study of implicated mutations in tyrosine kinase domain of the colony-stimulating factor 1 receptor (*CSF1R*) in hereditary diffuse leukoencephalopathy serves as an example (Rademakers et al., 2011). Autophosphorylation of CSF1R after stimulation with the colony-stimulating factor 1 (CSF1) was used to assay the function of human mutations. Phosphorylation of downstream targets was also examined in the study that identified mutations in *AKT3*, *PIK3R2*, and *PIK3CA*, causing a spectrum of related megalencephaly syndromes (Rivière et al., 2012b).

Changes in protein–protein interactions can be used to detect the effect of mutations on proteins involved in complexes. *In vitro* protein aggregation assay was used to test for the function of the co-chaperone DNAJB6 that was shown to cause limb-girdle muscular dystrophy (Sarparanta et al., 2012).

Functional assay to test lipid metabolism in incubated keranocytes was used in a recent study that linked *PNPLA1* to congenital ichthyosis (Grall et al., 2012). The same study used differentiation assay.

In some cases, mapping mutations on protein three-dimensional structure may provide a key insight into the functional mechanisms. For example, structural localization of *KLHL3* mutations causing familial hyperkalemic hypertension shows spatial clustering that helped to generate a biological hypothesis (Louis-Dit-Picard et al., 2012).

Model organisms amenable to genetic manipulation provide a possibility to test the phenotypic rather than molecular consequences of human allelic variants. The mammalian mouse model has been a model of choice for years to test the phenotypic effect of human genes. However, testing allelic series in the mouse is highly laborious. Therefore, zebrafish is being increasingly used to test the effect of human mutations because this vertebrate species is a powerful genetic model and a convenient system to screen for phenotypes. The approach involves knocking down the fish ortholog of the human gene and assaying the phenotypic effect. Next, if injecting human wild-type mRNA results in a phenotypic rescue, individual alleles can be tested for the potential to rescue the phenotype. Last, coinjection of wild-type and mutant mRNAs provides a test for dominant negative effects. Recent examples of the successful application of this approach include the analysis of mutations in co-chaperone DNAJB6 (Sarparanta et al., 2012) and mutations in the RNA exosome component causing pontocerebellar hypoplasia and spinal motor neuron degeneration (Wan et al., 2012). The zebrafish model was employed with the great success in characterizing multiple variants in several genes involved in ciliopathies (Zaghloul et al., 2011).

In some cases, much more distant model organisms appear helpful in interpreting human mutations. For example, a yeast system was successfully used to functionally characterize 84 human variants observed in patients with cystathionine-β-synthase deficiency (Mayfield et al., 2012).

Interestingly, dog is another species helping to establish the relationship between human mutations and phenotypes (Grall et al., 2012).

NOTES

Computational predictions

Additional supporting evidence for the functional significance of missense mutations can be provided by computational prediction algorithms. At this time, the accuracy of computational predictions is about 75–80% (Hicks et al., 2011), with the accuracy estimates dependent on datasets or databases that are used to define pathogenic and benign variants. Thus, the computational analysis is less informative than direct experimental evidence. However, given that the computational methods do not involve any additional labor and cost and can be applied to any gene, many studies rely, at least in part, on the application of computational methods. The accuracy of the methods can be higher for highly confident predictions (Jordan et al., 2010). If accompanied with rigorous accuracy estimation on a disease-specific dataset, bioinformatically derived prediction information may assist clinical decision-making. Current American College of Medical Genetics (ACMG) and International Agency for Research on Cancer (IARC) recommendations endorse the application of computational methods in genetic diagnostics but only in combination with other criteria (Plon et al., 2008; Richards et al., 2008). As more protein sequences and structures accompanied by training data (known disease-causing mutations and neutral polymorphisms) are available, the classification accuracy will improve. At the same time, principle difficulties in improving the accuracy of prediction methods remain and are discussed at the end of this section.

Despite the variety of approaches that exist (Table 2), these methods all rely heavily on two fundamental observations. First, the regions of proteins that are critical to function evolve under long-term negative selection; thus, when the sequence of a human protein is aligned for comparison to its homologs from other species, these sites will display only specific patterns of amino acid residue variation or complete conservation. The analysis of phylogenetic information in the form of multiple sequence alignment is a powerful source of information about the spectrum of residues allowed at a particular position of the protein of interest (Chasman et al., 2001; Ng et al., 2001; Sunyaev et al., 2001). Second, most pathogenic mutations affect protein stability (Yue et al., 2005; Potapov et al., 2009). In general, the prediction techniques based on protein spatial structure can be applied only if the structure has been resolved for the query protein or its close homolog, which is true only for a minor fraction of human proteins. However, even for the proteins with known spatial structure, structure-based methods work best only in addition to phylogeny-based approaches and provide only a slight increase in the accuracy of the methods (Kumar et al., 2009; Adzhubei et al., 2010).

Although existing methods all rely on evolutionary pattern and, sometimes, protein structure, they differ in algorithmic details. For example, SIFT (Sorting Intolerant from Tolerant) (Ng et al., 2001) and PolyPhen-2 (Adzhubei et al., 2010) estimate the probability that the mutant amino acid would fit the amino acid position given the observed substitution pattern. MAPP (Multivariate Analysis of Protein Polymorphism) (Stone and Sidow, 2005) analyzes conservation of physicochemical properties of amino acids, and LRT (Chun and Fay, 2009) and GERP (Cooper et al., 2010) estimate selective constraint. The methods based on multiple features also differ in the machine learning algorithms they employ. For example, MutationTaster (Schwarz et al., 2010) and PolyPhen-2 (Adzhubei et al. 2010) rely on the naive Bayes classifier, and SNAP (Bromberg et al., 2008) utilizes a neural network. Although different methods use essentially the same information, surprisingly, the methods are commonly discordant. This can be explained only in part by different threshold settings. This observation motivated the development of "umbrella" methods that combine predictions made by different algorithms such as Condel (González-Pérez and López-Bigas, 2011).

The accuracy of the methods could be potentially improved if the scope of the methods were narrower, specifically focused on a single phenotype and a group of genes involved in this phenotype. Such methods employ gene-specific training datasets, gene phylogeny, protein features, and classification rules optimized for a particular set of genes involved in a specific disease. Recently developed methods include a method focused on the *BRCA1* gene, involved in risk of breast and ovarian cancer (Karchin et al., 2007), and a method focused on genes encoding proteins of the heart sarcomere involved in hypertrophic cardiomyopathy (Jordan et al., 2011).

Two important basic effects hamper further development of new prediction methods of higher accuracy. First, the existing approaches may have intrinsic difficulties differentiating between mutations of large effect, important for genetic diagnostics, and slightly deleterious sequence variants in phylogenetically conserved positions, whose existence in genomes of apparently healthy humans is confirmed by numerous resequencing studies. Second, it was shown that human disease mutations are occasionally observed as wild-type alleles in vertebrate orthologs (Kondrashov et al., 2002). Most likely, this is the result of epistatic interactions.

NOTES

Table 2. A selection of online tools for predicting the functional effect of protein coding variants

AlignGVGD	Conservation of physicochemical properties	agvgd.iarc.fr
Condel	Prediction method based on combining other methods	bg.upf.edu/condel
MAPP (Multivariate Analysis of Protein Polymorphism)	Conservation of physicochemical properties	mendel.stanford.edu/sidowlab/downloads/MAPP
MutationTaster	Bayes classifier over multiple sequence features and conservation	mutationtaster.org
PMut	Evolutionary and structural features combined using a machine learning method	mmb2.pcb.ub.es:8080/PMut
PolyPhen-2	Evolutionary and structural features combined using naive Bayes classifier	genetics.bwh.harvard.edu/pph2
SIFT (Sorting Intolerant from Tolerant)	Evolutionary method based on position-specific scoring matrix	sift-dna.org
SNAP	Several evolutionary and structural features combined using a neural network	www.rostlab.org/services/snap
SNPs3D	Combination of a phylogenetic and a structural method; uses support vector machine	www.snps3d.org

Compensatory sequence changes enable amino acid changes corresponding to disease mutations in humans to be benign in a different genetic background. Current prediction methods analyze substitution patterns at individual positions and do not account for epistatic interactions. Compensatory changes should be taken into account to substantially increase the accuracy of computational approaches.

Functional Analysis of Rare Nonsynonymous Variants Involved in Complex Phenotypes

The analysis of complex traits presents a different set of issues. In this review, I limit the discussion to rare nonsynonymous variants in complex traits and leave out the discussion of functional effects of common variants identified by GWAS.

There is a growing interest in the role of rare variants in human complex traits. This interest, combined with the availability of next-generation sequencing technology, propels ongoing whole-exome sequencing studies (Do et al., 2012). For individual very rare variants, the phenotypic effect cannot be identified by the association test. Cosegregation is noninformative about variants involved in complex traits. The existing statistical approaches analyze rare variants collectively, grouping them by gene or pathway (Kiezun et al., 2012). In this approach, the statistical signal provided by functionally significant variants is frequently masked by noise due to benign alleles included in the same statistical test. Candidate gene–based studies showed that focusing on functionally significant alleles can increase statistical signal and, hence, the power to detect an association between the presence of rare variants and complex traits. The signal of association of rare variants in melatonin receptor 1B (*MTNR1B*) with type 2 diabetes increases only if variants that affect melatonin binding are considered (Bonnefond et al., 2012). *In vitro* experiments also helped to increase statistical signal of association in *ANGPL* genes with triglycerides (Romeo et al., 2009).

Statistical power of exome-sequencing studies is expected to be relatively low (Kryukov et al., 2009), so knowledge of functional variants would potentially help identify genes harboring rare variants associated with complex traits. Using experimental approaches at the whole-exome scale is not feasible. Some studies argued that computational methods for predicting the functional effect of human nonsynonymous alleles might be used to increase the power of sequencing studies (Ahituv et al., 2007; Price et al., 2010). Some statistical methods allow for weighting alleles based on potential functional effects. Likely, most sequencing studies would employ both tests weighted with predicted functional significance and tests grouping all nonsynonymous variants, disregarding predicted effect on function.

NOTES

Conclusion

Assigning functional significance to human alleles and inferring the causal relationship between DNA variants and phenotypes remains the central issue in human genetics. The most efficient way forward would combine statistical genetics considerations, *in vivo* and *in vitro* experimental studies, and computational approaches. Low throughput of current experimental methods and insufficient accuracy of computational predictions should be addressed to confidently annotate massive data on human genetic variation from the functional perspective.

An additional issue raising the problem to even a greater level of complexity is that, in many cases, the same functional variant can have different phenotypic consequences varying in both expressivity and penetrance depending on other genetic and environmental factors.

Acknowledgments

I am grateful to Gregory Kryukov for the analysis presented in Figure 1. I also acknowledge support from National Institutes of Health grants R01MH084676 and R01GM078598.

References

1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing. Nature 467:1061–1073.

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR (2010) A method and server for predicting damaging missense mutations. Nat Methods 7:248–249.

Ahituv N, Kavaslar N, Schackwitz W, Ustaszewska A, Martin J, Hebert S, Doelle H, Ersoy B, Kryukov G, Schmidt S, Yosef N, Ruppin E, Sharan R, Vaisse C, Sunyaev S, Dent R, Cohen J, McPherson R, Pennacchio LA (2007) Medical sequencing at the extremes of human body mass. Am J Hum Genet 80:779–791.

Boileau C, Guo DC, Hanna N, Regalado ES, Detaint D, Gong L, Varret M, Prakash SK, Li AH, d'Indy H, Braverman AC, Grandchamp B, Kwartler CS, Gouya L, Santos-Cortez RL, Abifadel M, Leal SM, Muti C, Shendure J, Gross MS, et al. (2012) *TGFB2* mutations cause familial thoracic aortic aneurysms and dissections associated with mild systemic features of Marfan syndrome. Nat Genet 44:916–921.

Bonnefond A, Clément N, Fawcett K, Yengo L, Vaillant E, Guillaume JL, Dechaume A, Payne F, Roussel R, Czernichow S, Hercberg S, Hadjadj S, Balkau B, Marre M, Lantieri O, Langenberg C, Bouatia-Naji N; Meta-Analysis of Glucose and Insulin-Related Traits Consortium (MAGIC), Charpentier G, Vaxillaire M, et al. (2012) Rare *MTNR1B* variants impairing melatonin receptor 1B function contribute to type 2 diabetes. Nat Genet 44:297–301.

Boyko AR, Williamson SH, Indap AR, Degenhardt JD, Hernandez RD, Lohmueller KE, Adams MD, Schmidt S, Sninsky JJ, Sunyaev SR, White TJ, Nielsen R, Clark AG, Bustamante CD (2008) Assessing the evolutionary impact of amino acid mutations in the human genome. PLoS Genet 4:e1000083.

Bromberg Y, Yachdav G, Rost B (2008) SNAP predicts effect of mutations on protein function. Bioinformatics 24:2397–2398.

Chasman D, Adams RM (2001). Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation. J Mol Biol 307:683–706.

Chun S, Fay JC (2009) Identification of deleterious mutations within three human genomes. Genome Res 19:1553–1561.

Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH (2004) Multiple rare alleles contribute to low plasma levels of HDL cholesterol. Science 305:869–872.

Cooper GM, Goode DL, Ng SB, Sidow A, Bamshad MJ, Shendure J, Nickerson DA (2010) Single-nucleotide evolutionary constraint scores highlight disease-causing mutations. Nat Methods 7:250–251.

Degner JF, Pai AA, Pique-Regi R, Veyrieras JB, Gaffney DJ, Pickrell JK, De Leon S, Michelini K, Lewellen N, Crawford GE, Stephens M, Gilad Y, Pritchard JK (2012) DNase I sensitivity QTLs are a major determinant of human expression variation. Nature 482:390–394.

Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB (2010) Rare variants create synthetic genome-wide associations. PLoS Biol 8:e1000294.

Do R, Kathiresan S, Abecasis G (2012) Exome sequencing and complex disease: practical aspects of rare variant association studies. Hum Mol Genet 21:R1–R9.

NOTES

González-Pérez A, López-Bigas N (2011) Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. Am J Hum Genet 88:440–449.

Grall A, Guaguère E, Planchais S, Grond S, Bourrat E, Hausser I, Hitte C, Le Gallo M, Derbois C, Kim GJ, Lagoutte L, Degorce-Rubiales F, Radner FP, Thomas A, Küry S, Bensignor E, Fontaine J, Pin D, Zimmermann R, Zechner R, et al. (2012) *PNPLA1* mutations cause autosomal recessive congenital ichthyosis in golden retriever 4s and humans. Nat Genet 44:140–147.

Heinzen EL, Swoboda KJ, Hitomi Y, Gurrieri F, Nicole S, de Vries B, Tiziano FD, Fontaine B, Walley NM, Heavin S, Panagiotakaki E; European Alternating Hemiplegia of Childhood (AHC) Genetics Consortium; Biobanca e Registro Clinico per l'Emiplegia Alternante (I.B.AHC) Consortium; European Network for Research on Alternating Hemiplegia (ENRAH) for Small and Medium-sized Enterpriese (SMEs) Consortium, Fiori S, Abiusi E, Di Pietro L, Sweney MT, Newcomb TM, Viollet L, et al. (2012) De novo mutations in *ATP1A3* cause alternating hemiplegia of childhood. Nat Genet 44:1030–1034.

Hicks S, Wheeler DA, Plon SE, Kimmel M (2011) Prediction of missense mutation functionality depends on both the algorithm and sequence alignment employed. Hum Mutat 32:661–668.

International HIV Controllers Study, Pereyra F, Jia X, McLaren PJ, Telenti A, de Bakker PI, Walker BD, Ripke S, Brumme CJ, Pulit SL, Carrington M, Kadie CM, Carlson JM, Heckerman D, Graham RR, Plenge RM, Deeks SG, Gianniny L, Crawford G, Sullivan J, et al. (2010) The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science 330:1551–1557.

Ji W, Foo JN, O'Roak BJ, Zhao H, Larson MG, Simon DB, Newton-Cheh C, State MW, Levy D, Lifton RP (2008) Rare independent mutations in renal salt handling genes contribute to blood pressure variation. Nat Genet 40:592–599.

Johansen CT, Wang J, Lanktree MB, McIntyre AD, Ban MR, Martins RA, Kennedy BA, Hassell RG, Visser ME, Schwartz SM, Voight BF, Elosua R, Salomaa V, O'Donnell CJ, Dallinga-Thie GM, Anand SS, Yusuf S, Huff MW, Kathiresan S, Hegele RA (2010) Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. Nat Genet 42:684–687.

Jordan CT, Cao L, Roberson ED, Duan S, Helms C.A, Nair RP, Duffin KC, Stuart PE, Goldgar D, Hayashi G, Olfson EH, Feng BJ, Pullinger CR, Kane JP, Wise CA, Goldbach-Mansky R, Lowes MA, Peddle L, Chandran V, Liao W, et al. (2012) Rare and common variants in *CARD14*, encoding an epidermal regulator of NF-kappaB, in psoriasis. Am J Hum Genet 90:796–808.

Jordan DM, Ramensky VE, Sunyaev SR (2010) Human allelic variation: perspective from protein function, structure, and evolution. Curr Opin Struct Biol 20:342–350.

Jordan DM, Kiezun A, Baxter SM, Agarwala V, Green RC, Murray MF, Pugh T, Lebo MS, Rehm HL, Funke BH, Sunyaev SR (2011) Development and validation of a computational method for assessment of missense variants in hypertrophic cardiomyopathy. Am J Hum Genet 88:183–192.

Karchin R, Monteiro AN, Tavtigian SV, Carvalho MA, Sali A (2007) Functional impact of missense variants in *BRCA1* predicted by supervised learning. PLoS Comput Biol 3:e26.

Kiezun A, Garimella K, Do R, Stitziel NO, Neale BM, McLaren PJ, Gupta N, Sklar P, Sullivan PF, Moran JL (2012) Exome sequencing and the genetic basis of complex traits. Nat Genet 44:623–630.

Kondrashov AS (2003) Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. Hum Mutat 21:12–27.

Kondrashov AS, Sunyaev S, Kondrashov FA (2002) Dobzhansky–Muller incompatibilities in protein evolution. Proc Natl Acad Sci USA 99:14878–14883.

Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson SA, Sigurdsson A, Jonasdottir A, Jonasdottir A, Wong WS, Sigurdsson G, Walters GB, Steinberg S, Helgason H, Thorleifsson G, Gudbjartsson DF, Helgason A, Magnusson OT, Thorsteinsdottir U, et al. (2012) Rate of *de novo* mutations and the importance of father's age to disease risk. Nature 488:471–475.

Kryukov GV, Pennacchio LA, Sunyaev SR (2007) Most rare missense alleles are deleterious in humans: Implications for complex disease and association studies. Am J Hum Genet 80:727–739.

Kryukov, GV, Shpunt A, Stamatoyannopoulos JA, Sunyaev SR (2009) Power of deep, all-exon resequencing for discovery of human trait genes. Proc Natl Acad Sci USA 106:3871–3876.

NOTES

Kumar S, Suleski MP, Markov GJ, Lawrence S, Marco A, Filipski AJ (2009) Positional conservation and amino acids shape the correct diagnosis and population frequencies of benign and damaging personal amino acid mutations. Genome Res 19:1562–1569.

Le Goff C, Mahaut C, Abhyankar A, Le Goff W, Serre V, Afenjar A, Destrée A, di Rocco M, Héron D, Jacquemont S, Marlin S, Simon M, Tolmie J, Verloes A, Casanova JL, Munnich A, Cormier-Daire V (2011) Mutations at a single codon in Mad homology 2 domain of SMAD4 cause Myhre syndrome. Nat Genet 44:85–88.

Li Y, Vinckenbosch N, Tian G, Huerta-Sanchez E, Jiang T, Jiang H, Albrechtsen A, Andersen G, Cao H, Korneliussen T, Grarup N, Guo Y, Hellman I, Jin X, Li Q, Liu J, Liu X, Sparsø T, Tang M, Wu H, et al. (2010) Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. Nat Genet 42:969–972.

Livingston RJ, von Niederhausern A, Jegga AG, Crawford DC, Carlson CS, Rieder MJ, Gowrisankar S, Aronow BJ, Weiss RB, Nickerson DA (2004) Pattern of sequence variation across 213 environmental response genes. Genome Res 14:1821–1831.

Louis-Dit-Picard H, Barc J, Trujillano D, Miserey-Lenkei S, Bouatia-Naji N, Pylypenko O, Beaurain G, Bonnefond A, Sand O, Simian C, Vidal-Petiot E, Soukaseum C, Mandet C, Broux F, Chabre O, Delahousse M, Esnault V, Fiquet B, Houillier P, Bagnis CI, et al. (2012) *KLHL3* mutations cause familial hyperkalemic hypertension by impairing ion transport in the distal nephron. Nat Genet 44:456–460.

MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, Jostins L, Habegger L, Pickrell JK, Montgomery SB, Albers CA, Zhang ZD, Conrad DF, Lunter G, Zheng H, Ayub Q, DePristo MA, Banks E, Hu M, Handsaker RE, et al. (2012) A systematic survey of loss-of-function variants in human protein–coding genes. Science 335:823–828.

Marth GT, Czabarka E, Murvai J, Sherry ST (2004) The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. Genetics 166:351–372.

Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J (2012) Systematic localization of common disease-associated variation in regulatory DNA. Science 337:1190–1195.

Mayfield JA, Davies MW, Dimster-Denk D, Pleskac N, McCarthy S, Boydston EA, Fink L, Lin XX, Narain AS, Meighan M, Rine J (2012) Surrogate genetics and metabolic profiling for characterization of human disease alleles. Genetics 190:1309–1323.

McDaniell R, Lee BK, Song L, Liu Z, Boyle AP, Erdos MR, Scott LJ, Morken MA, Kucera KS, Battenhouse A, Keefe D, Collins FS, Willard HF, Lieb JD, Furey TS, Crawford GE, Iyer VR, Birney E (2010) Heritable individual-specific and allele-specific chromatin signatures in humans. Science 328:235–239.

Mitchell AA, Chakravarti A, Cutler DJ (2005) On the probability that a novel variant is a disease-causing mutation. Genome Res 15:960–966.

Momozawa Y, Mni M, Nakamura K, Coppieters W, Almer S, Amininejad L, Cleynen I, Colombel JF, de Rijk P, Dewit O, Finkel Y, Gassull MA, Goossens D, Laukens D, Lémann M, Libioulle C, O'Morain C, Reenaers C, Rutgeerts P, Tysk C, et al. (2011) Resequencing of positional candidates identifies low frequency *IL23R* coding variants protecting against inflammatory bowel disease. Nat Genet 43:43–47.

Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, Li X, Li H, Kuperwasser N, Ruda VM, Pirruccello JP, Muchmore B, Prokunina-Olsson L, Hall JL, Schadt EE, Morales CR, Lund-Katz S, Phillips MC, Wong J, Cantley W, et al. (2010) From noncoding variant to phenotype via *SORT1* at the 1p13 cholesterol locus. Nature 466:714–719.

Nachman MW, Crowell SL (2000) Estimate of the mutation rate per nucleotide in humans. Genetics 156:297–304.

Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, Lin CF, Stevens C, Wang LS, Makarov V, Polak P, Yoon S, Maguire J, Crawford EL, Campbell NG, Geller ET, Valladares O, Schafer C, Liu H, Zhao T, et al. (2012) Patterns and rates of exonic *de novo* mutations in autism spectrum disorders. Nature 485:242–245.

NOTES

Ng PC, Henikoff S (2001) Predicting deleterious amino acid substitutions. Genome Res 11:863–874.

Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ (2010) Exome sequencing identifies the cause of a Mendelian disorder. Nat Genet 42:30–35.

O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, Levy R, Ko A, Lee C, Smith JD, Turner EH, Stanaway IB, Vernot B, Malig M, Baker C, Reilly B, Akey JM, Borenstein E, Rieder MJ, Nickerson DA, et al. (2012) Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. Nature 485:246–50.

Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS, Hogervorst FB, Hoogerbrugge N, Spurdle AB, Tavtigian SV; IARC Unclassified Genetic Variants Working Group (2008) Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. Hum Mutat 29:1282–1291.

Potapov V, Cohen M, Schreiber G (2009) Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. Protein Eng Des Sel 22:553–560.

Price AL, Kryukov GV, de Bakker PI, Purcell SM, Staples J, Wei LJ, Sunyaev SR (2010) Pooled association tests for rare variants in exon-resequencing studies. Am J Hum Genet 86:832–838.

Rademakers R, Baker M, Nicholson AM, Rutherford NJ, Finch N, Soto-Ortolaza A, Lash J, Wider C, Wojtas A, DeJesus-Hernandez M, Adamson J, Kouri N, Sundal C, Shuster EA, Aasly J, MacKenzie J, Roeber S, Kretzschmar HA, Boeve BF, Knopman DS, et al. Mutations in the colony stimulating factor 1 receptor (*CSF1R*) gene cause hereditary diffuse leukoencephalopathy with spheroids. Nat Genet 44:200–205.

Raychaudhuri S, Sandor C, Stahl EA, Freudenberg J, Lee HS, Jia X, Alfredsson L, Padyukov L, Klareskog L, Worthington J, Siminovitch KA, Bae SC, Plenge RM, Gregersen PK, de Bakker PI (2012) Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat Genet 44:291–296.

Richards CS, Bale S, Bellissimo DB, Das S, Grody WW, Hegde MR, Lyon E, Ward BE, Molecular Subcommittee of the ACMG Laboratory Quality Assurance Committee (2008) ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. Genet Med 10:294–300.

Rivas MA, Beaudoin M, Gardet A, Stevens C, Sharma Y, Zhang CK, Boucher G, Ripke S, Ellinghaus D, Burtt N, Fennell T, Kirby A, Latiano A, Goyette P, Green T, Halfvarson J, Haritunians T, Korn JM, Kuruvilla F, Lagacé C, Neale B, et al. (2011) Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. Nat Genet 43:1066–1073.

Rivière JB, Mirzaa GM, O'Roak BJ, Beddaoui M, Alcantara D, Conway RL, St-Onge J, Schwartzentruber JA, Gripp KW, Nikkel SM, Worthylake T, Sullivan CT, Ward TR, Butler HE, Kramer NA, Albrecht B, Armour CM, Armstrong L, Caluseriu O, Cytrynbaum C, et al. (2012a) *De novo* germline and postzygotic mutations in *AKT3*, *PIK3R2* and *PIK3CA* cause a spectrum of related megalencephaly syndromes. Nat Genet 44:934–940.

Rivière JB, van Bon BW, Hoischen A, Kholmanskikh SS, O'Roak BJ, Gilissen C, Gijsen S, Sullivan CT, Christian SL, Abdul-Rahman OA, Atkin JF, Chassaing N, Drouin-Garraud V, Fry AE, Fryns JP, Gripp KW, Kempers M, Kleefstra T, Mancini GM, Nowaczyk MJ, et al. (2012b) *De novo* mutations in the actin genes *ACTB* and *ACTG1* cause Baraitser–Winter syndrome. Nat Genet 44:440–444.

Roach JC, Glusman G, Smit AF, Huff C.D, Hubley R, Shannon PT, Rowen L, Pant KP, Goodman N, Bamshad M, Shendure J, Drmanac R, Jorde LB, Hood L, Galas DJ (2010) Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science 328:67–639.

Romeo S, Yin W, Kozlitina J, Pennacchio LA, Boerwinkle E, Hobbs HH, Cohen JC (2009) Rare loss-of-function mutations in *ANGPTL* family members contribute to plasma triglyceride levels in humans. J Clin Invest 119:70–79.

NOTES

Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, Ercan-Sencicek AG, DiLullo NM, Parikshak NN, Stein JL, Walker MF, Ober GT, Teran NA, Song Y, El-Fishawy P, Murtha RC, Choi M, Overton JD, Bjornson RD, Carriero NJ, et al. (2012) *De novo* mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 485:237–241.

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.

Schwarz JM, Rödelsperger C, Schuelke M, Seelow D (2010) MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods 7:575–576.

Stone EA, Sidow A (2005) Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. Genome Res 15:978–986.

Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C, Ingle CE, Dunning M, Flicek P, Koller D, Montgomery S, Tavaré S, Deloukas P, Dermitzakis E (2007) Population genomics of human gene expression. Nat Genet 39:1217–1224.

Sun JX, Helgason A, Masson G, Ebenesersdóttir SS, Li H, Mallick S, Gnerre S, Patterson N, Kong A, Reich D, Stefansson K (2012) A direct characterization of human mutation based on microsatellites. Nat Genet 44:1161–1165.

Sunyaev SR, Ramensky VE, Koch I, Lathe W 3d, Kondrashov AS, Bork P (2001) Prediction of deleterious human alleles. Hum Mol Genet 15:591–597.

Sunyaev SR, Lathe WC 3d, Ramensky VE, Bork P (2000) SNP frequencies in human genes an excess of rare alleles and differing modes of selection. Trends Genet 16:335–337.

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, et al. (2012) Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science 337:64–69.

Van Houdt JK, Nowakowska BA, Sousa SB, van Schaik BD, Seuntjens E, Avonce N, Sifrim A, Abdul-Rahman OA, van den Boogaard MJ, Bottani A, Castori M, Cormier-Daire V, Deardorff MA, Filges I, Fryer A, Fryns JP, Gana S, Garavelli L, Gillessen-Kaesbach G, Hall BD, et al. (2012) Heterozygous missense mutations in *SMARCA2* cause Nicolaides-Baraitser syndrome. Nat Genet 44:445–449.

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, et al. (2012) Mutations in the RNA exosome component gene *EXOSC3* cause pontocerebellar hypoplasia and spinal motor neuron degeneration. Nat Genet 44:704–708.

Wang X, Grus WE, Zhang J (2006) Gene losses during human origins. PLoS Biol 4:e52.

Williamson SH, Hernandez R, Feldel-Alon A, Zhe L, Nielsen R, Bustamante CD (2005) Simultaneous inference of selection and population growth from patterns of variation in the human genome. Proc Natl Acad Sci USA 102:7882–7887.

Wortmann SB, Vaz FM, Gardeitchik T, Vissers LE, Renkema GH, Schuurs-Hoeijmakers JH, Kulik W, Lammens M, Christin C, Kluijtmans LA, Rodenburg RJ, Nijtmans LG, Grünewald A, Klein C, Gerhold JM, Kozicz T, van Hasselt PM, Harakalova M, Kloosterman W, Barić I, et al. (2012) Mutations in the phospholipid remodeling gene *SERAC1* impair mitochondrial function and intracellular cholesterol trafficking and cause dystonia and deafness. Nat Genet 44:797–802.

Wray NR, Purcell, SM, Visscher PM (2011) Synthetic associations created by rare variants do not explain most GWAS results. PLoS Biol 9:e1000579.

Xu B, Roos JL, Dexheimer P, Boone B, Plummer B, Levy S, Gogos JA, Karayiorgou M (2011) Exome sequencing supports a *de novo* mutational paradigm for schizophrenia. Nat Genet 43:864–868.

Yue P, Li Z, Moult J (2005) Loss of protein structure stability as a major causative factor in monogenic disease. J Mol Biol 353:459–473.

Zaghloul NA, Katsanis N (2011) Zebrafish assays of ciliopathies. Methods Cell Biol 105:257–272.