

Progress in the Genetics of Polygenic Brain Disorders: Significant New Challenges for Neurobiology

Steven A. McCarroll^{1,2,*} and Steven E. Hyman^{1,3,*}

¹Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA 02141, USA

²Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

³Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA

*Correspondence: mccarroll@genetics.med.harvard.edu (S.A.M.), seh@harvard.edu (S.E.H.)

<http://dx.doi.org/10.1016/j.neuron.2013.10.046>

Advances in genome analysis, accompanied by the assembly of large patient cohorts, are making possible successful genetic analyses of polygenic brain disorders. If the resulting molecular clues, previously hidden in the genomes of affected individuals, are to yield useful information about pathogenesis and inform the discovery of new treatments, neurobiology will have to rise to many difficult challenges. Here we review the underlying logic of the genetic investigations, describe in more detail progress in schizophrenia and autism, and outline the challenges for neurobiology that lie ahead. We argue that technologies at the disposal of neuroscience are adequately advanced to begin to study the biology of common and devastating polygenic disorders.

Introduction

Genomes encode the key macromolecular building blocks of our cells, RNA, and proteins. In concert with intracellular and extracellular signals, our genomes regulate the times, places, quantities, and cell-type-specific patterns of expression of messenger RNAs (mRNAs) that give rise to proteins and of RNAs with independent functions. These macromolecules, in turn, direct the synthesis and trafficking of essentially all other molecules within cells. Analysis of the completed genome sequences of many organisms, together with biochemistry, physiology, and other disciplines, have made it possible to identify many if not essentially all of the genes that encode components of receptors, ion channels, synaptic proteins, and other molecular complexes of central interest to neurobiology. Increasingly powerful technologies, grounded in genetics and molecular biology, permit neuroscientists to manipulate the genomes of cells and model organisms to understand both normal function of the nervous system and disease processes (Cong et al., 2013; Fenno et al., 2011; Wang et al., 2013). Currently, information derived from genes and genomes provides neuroscientists with molecular clues to the properties of the many thousands of neuronal and glial cell types in the brain, to functional properties of brain circuits, and ultimately to important aspects of cognition, emotion, and behavior.

Recognition of the importance of genetic and genomic information is not meant as an embrace of simplistic genetic determinism, which was discarded by neuroscientists long ago based on the evidence—one obvious refutation is the frequent discordance of monozygotic twins (who share 100% of their DNA sequences) for important phenotypes, including both normal variation and disease phenotypes including schizophrenia and mood disorders. Such discordance, for the most part not yet understood in detail, is grounded in complex interactions of genes with stochastic and environmental factors that influence brain development, maturation, and function. That said, genomes carry enormous biological influence: the remarkable similarities

of basic brain structure and function within species are testimony to the central significance of the genetic blueprint. A recent demonstration that human pluripotent stem cells in vitro (extremely distant from a natural developmental environment) can give rise to cerebral organoids with discrete recognizable brain structures and significant features of a cerebral cortex (Lancaster et al., 2013) serves as a remarkable reminder of the information contained in genomes—even if the resulting organoids are only pale simulacra of a human brain.

Genetic information is particularly important to neurobiologists studying brain disease because the human brain is, both for ethical and practical reasons, generally inviolable. Scientists studying the biology of cancer or immunologic diseases, for example, can have direct access to diseased tissues obtained from surgical specimens or blood. The resulting cells can be examined for somatic mutations, epigenetic marks, patterns of gene expression, and other molecular indicia. In contrast, for the most part, the human brain can only be examined indirectly in life. Thus, when disorders of the CNS have a significant hereditary component of risk, the ability to obtain molecular clues from genetic analysis may create the most effective current opportunities for scientific investigation. The utility of genetic insights is particularly salient in brain disorders that affect evolutionarily recent brain circuits and regions or that for other reasons have been difficult to model in animals. These include common psychiatric disorders such as autism, schizophrenia, bipolar disorder, and major depression as well as late-onset versions of neurodegenerative disorders such as Parkinson's disease and Alzheimer's disease. In the case of the psychiatric disorders, the relative lack of neuropathology that can be analyzed in post-mortem tissue makes genetic information even more valuable as a source of molecular clues to pathogenesis.

Psychiatric disorders have long been recognized to cluster in families even though they do not segregate in simple, Mendelian fashion. Twin and adoption studies demonstrated that familiarity

resulted from heredity, thus suggesting that information about the molecular basis of these serious and disabling disorders is hidden in DNA sequence variation. Two factors have made it possible during the last 5 years to reveal such heretofore hidden information: technological innovation in genome analysis and recognition of the need for scale. Advances in genomics made it possible to prosecute large-scale unbiased genome-wide searches both cheaply—the cost of sequencing DNA has declined approximately one million-fold in the last decade—and accurately. At the same time, a new appreciation of the scale of analysis required to successfully attack heterogeneous, polygenic disorders has led to the examination of tens of thousands of genomes, and thus, finally, to genetic findings that replicate across large studies. For example, large-scale genetic analyses (involving 80,094 individuals, both patients and controls) have now contributed to recognition of 110 loci that influence susceptibility to multiple sclerosis ([International Multiple Sclerosis Genetics Consortium, 2013](#)). Among the psychiatric disorders, genetic analyses have arguably yielded the first substantial, if still early, insights into molecular mechanisms of disease. Such findings across many common brain disorders promise to make the coming 25 years very different from the previous 25, not only with respect to understandings of pathogenesis but also—it is to be hoped—effective therapeutics. Such success will only come to pass, however, if neurobiology rises to the difficult challenge of putting genetics results to work.

How Our Genomes Operate—and How They Do Not

A naive but pervasive view of human genetic variation sees the human genome as an optimized end product of evolution. In this view, a human genome, like a Shakespearean sonnet, is perfectly composed, with a place for everything, and everything in its place. Such a genome, perfected through many rounds of natural selection, brings us a long and disease-free life, unless a new mutation or an unfortunate calamity of environment causes an illness. In fact, analysis of the sequences of thousands of human genomes demonstrates that far from conforming to some uniform model of optimization, our genomes teem with functional variation. The two haploid genomes that we inherit from our parents differ at millions of sites ([Abecasis et al., 2010](#)). Several thousand variants affect the copy number of large, multikilobase genomic segments ([Handsaker et al., 2011](#); [Conrad et al., 2010](#)). Each genome has thousands of variants that affect the expression of nearby genes, with different sets of regulatory variants acting in different tissues ([Nica et al., 2011](#); [Fu et al., 2012](#)). Each diploid human genome has about 100 gene-disrupting variants, from large deletions to single-nucleotide nonsense variants that ablate the functions of specific genes; in any individual, some 20 of these genes may be inactivated in both copies ([MacArthur et al., 2012](#)). Thousands of protein-coding genes harbor missense variants that may influence their function in complex ways ([Abecasis et al., 2010](#)).

The human genome as it exists in real human populations, then, is less a Shakespearean sonnet than a collection of seven billion drafts. In any genome, we can readily find thousands of variants that look functional on paper, many of which may be confirmed as functional by some laboratory assay. However, most apparently functional variants have, at least to date, no

demonstrated association to disease phenotypes when evaluated in large numbers of individuals. In sum, it is easy to find variation, even functional variation, but against this complex background it is very difficult to identify gene variants that contribute to any particular illness phenotype. This challenge notwithstanding, it is clear that the genome is the right place to look for molecular underpinnings of illness. Studies of psychiatric disorders that compared the concordance rates of monozygotic versus dizygotic twin pairs estimate heritability at 0.81 for schizophrenia ([Sullivan et al., 2003](#)), 0.75 for bipolar disorder ([Smoller and Finn, 2003](#)), and 0.80 for autism spectrum disorders ([Ronald and Hoekstra, 2011](#)). Some assumptions inherent in twin studies have been questioned, but recent analytical techniques, which use genome-wide molecular data to derive unbiased estimates of heritability, strongly confirm a significant role for inheritance in shaping risk ([Lee et al., 2012](#); [Yang et al., 2010](#)). One can conclude that insights about the molecular nature of brain illnesses are encoded in the sequences of individual human genomes. The challenge is to find the variants that matter, among the far-larger number of variants that do not. The challenge is heightened given that variants do not act in isolation or on isogenic backgrounds, nor can human developmental environments be held constant as genomes vary.

Mendelian Brain Disorders

Over the past two decades, it has become increasingly straightforward to identify the causal genes for highly penetrant, Mendelian (monogenic) human diseases. Among monogenic brain disorders, significant early discoveries included the identification of CGG repeats within the *FMR1* gene as the cause of Fragile X syndrome ([Fu et al., 1991](#)), identification of the genetic cause of Huntington's disease ([The Huntington's Disease Collaborative Research Group, 1993](#)), and the demonstration that mutations in the *MECP2* gene produced Rett syndrome ([Amir et al., 1999](#)). Identification of these causative genes made it possible to develop a wide range of tools ranging from antibodies to transgenic mice, although successful clinical trials of therapies based on these discoveries have been slow to follow.

One reason for the difficulty in discovering therapeutics is that apparently monogenic disorders are not always as simple to analyze as might initially appear. Affected individuals for any given disorder may have different mutations in the causative gene, which may influence such features as age of onset, disease severity, and treatment response. For example, in Rett syndrome, diverse mutations have been identified in the *MECP2* gene ([Lee et al., 2001](#)). Moreover, some *MECP2* mutations produce not Rett syndrome but other neuropsychiatric symptoms such as obsessive-compulsive disorder and attention-deficit hyperactivity disorder; some individuals with well-diagnosed Rett syndrome do not have *MECP2* mutations at all ([Suter et al., 2013](#)). In addition, mutations in a single causative gene may only be a portal to far greater molecular complexity. Thus, for example, *FMR1*, which is a neuronal polyribosome-associated RNA binding protein, has been shown to affect the translation of 842 mRNA transcripts ([Darnell et al., 2011](#)) each with their own “downstream” biology; many of these individual targets are now being implicated for subtler contributions to complex, polygenic disease.

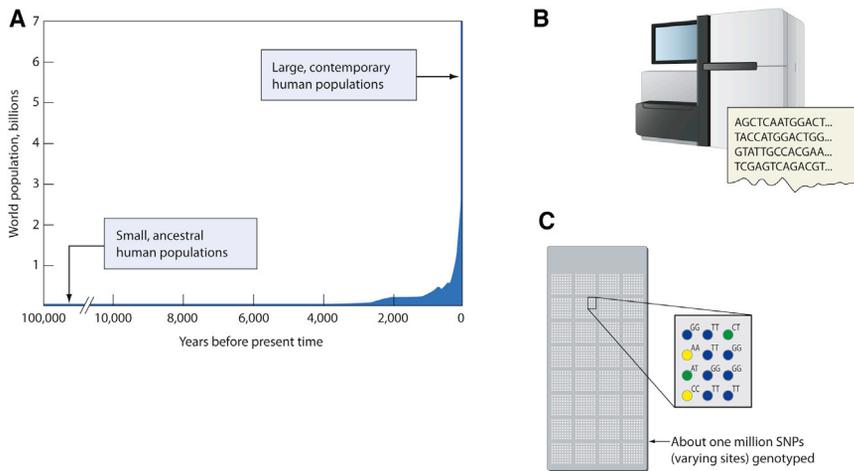


Figure 1. Different Kinds of Genetic Variation and Different Approaches to Genetic Analysis Are Influenced by Human Population History

(A) Humans have an eccentric population history, in which small, ancestral populations rapidly expanded into a population of some seven billion individuals.

(B) Genetic variants that have arisen recently in human history, including new mutations, are generally ascertained by genome sequencing.

(C) Polymorphisms that were already present in the small, ancestral populations can be systematically evaluated for relationship to phenotypes, on thousands of genetic backgrounds, using inexpensive array-based assay platforms.

Polygenicity

The unexpected complexity of many monogenic brain disorders pales in comparison with the emerging complexity of common polygenic brain disorders, a challenge that is only now coming into view. Because severe, highly penetrant mutations often produce marked decrements in reproductive fitness, they tend to be rare. In contrast, many common human illnesses result from the interaction of a large number of genes (polygenicity) in combination with nongenetic risk factors. Moreover, disease phenotypes tend to result from different combinations of genetic (and likely nongenetic) risk factors in different families and individuals. The use of the term “risk factor” rather than “cause” indicates that among polygenic disorders, any individual sequence variant (or environmental factor) acts in a statistical rather than deterministic fashion. No single genetic variant is necessary or sufficient for the disorder phenotype and thus cannot be used to predict phenotype except in a probabilistic manner.

Several important genetic results support polygenic models, for example, in schizophrenia and autism. The first is the finding that large numbers of common variants shape an individual’s disease risk. Statistical geneticists define a “polygene” from large constellations of common alleles that are observed (in one cohort) at slightly higher frequencies in schizophrenic patients than controls. When such alleles are subsequently evaluated in other cohorts, schizophrenic patients are found to carry more such alleles (on average) than control subjects do. (Purcell et al., 2009). Within families, schizophrenic children of unaffected parents also tend to have inherited more than the 50% of such alleles that they would be expected to have inherited by chance from parents heterozygous at the relevant loci (Ruderfer et al., 2011). These results suggest that one important component of genetic risk for a polygenic disorder such as schizophrenia arises from many small genetic nudges, rather than a single, hard shove.

The polygenic model is also supported by rare alleles of larger effect. For example, a substantial minority of autistic patients (about 5%–10%), but only a small fraction of the general population (about 1%), have *de novo* deletions and duplications of large (>500 kb) genomic segments in their genomes. This excess in patients indicates that such “copy number variations” (CNVs) contribute to the patients’ phenotypes. However, these CNVs

are widely distributed across the genome, at more than 100 different loci. Only a few loci show recurrent mutations, and these recurrent mutations account for only about 1%–2% of patients. The statistical distribution of influences across the genome suggests that hundreds of different human genes can mutate to influence autism risk (Sanders et al., 2011).

Perhaps the emerging polygenicity of mental disorders, involving many hundreds, and perhaps more than a thousand, genetic loci, should not have come as a surprise. Cognition, executive function, and emotional regulation are not the result of simple, five-protein metabolic pathways. The assembly of synapses and neuronal circuits involves complex biological processes that recruit thousands of gene products. The implications of polygenicity were not initially recognized, which doomed older linkage studies of psychiatric disorders in the 1980s and 1990s to failure. Small studies in the 1990s and early 2000s that attempted to show association of plausible biological candidate genes with disease suffered a similar fate. They produced equivocal results that ultimately failed the test of replication. Only when recent unbiased, genome-wide studies were adequately sized, did they succeed in distinguishing disease-associated variants from genome-wide statistical fluctuations.

Genome Variation, Rare and Common

To understand how human gene discovery approaches work, one needs to take a brief detour through human population history. Humans have an eccentric population history: although some seven billion humans currently inhabit the earth, we were a far-smaller species only 100,000 years (about four thousand generations) ago, and even a significantly smaller population 150 years (seven to eight generations) ago. The dramatic expansion of human populations from smaller groups of ancestors has profoundly shaped the patterns of variation that exist in human genomes. It also defines some of the key opportunities for discovering the sequence variants that contribute to phenotypes (Figure 1).

Studies of Rare Variation

An intriguing set of genetic variants has arisen in rapidly expanding modern populations, even involving new mutation in the most

recent generation. Based on their recency, these variants are both rare and relatively unfiltered by natural selection; thus, they could in principle include more deleterious mutations. An early view of the contributions of rare variants came from the observations of large CNVs in the genomes of several percent of autism and schizophrenia patients. Such CNVs appear to confer substantial increases in risk. Interestingly, they have proven to be only partially penetrant, increasing risk from a background rate of about 0.5%–1.0% to about 4%–20%. Because most of the recurrent CNVs are large (hundreds of kilobases) and affect the dosage of many genes, it has been difficult to derive actionable neurobiological insight from them.

Many more rare variants exist at the fine-scale of single-nucleotide variants and small indels (insertions or deletions). An important advance in the ability to identify and study rare variants comes from innovations in sequencing technology. Today the protein-coding parts of a patient's genome (the "exome") can be sequenced for well under \$1,000, enabling exome-sequencing studies of hundreds of patients. As in the case of common variants, it is challenging to distinguish the rare variants that contribute to a phenotype, from the background of other rare variation that is present in each genome.

To reduce the background created by the hundreds of protein-altering variants in each genome, one common study design sequences father-mother-proband trios, then focuses on those protein-altering mutations in the proband that are *de novo* mutations, i.e., that were not inherited from either parent. The challenge in this analysis comes from the fact that protein-altering mutations unrelated to disease arise in the general population at an appreciable rate. Disease-predisposing variants are not immediately distinguishable from this background, except to the extent they recur in the same genes in different individuals with the disease under investigation.

To date, the most convincing implication of individual genes has come from studies of congenital and child-onset disorders such as autism, intellectual disability, and pervasive developmental delay. For autism, four large studies of father-mother-offspring trios collectively ascertained *de novo* mutations in more than one thousand autism patients (Sanders et al., 2012; O'Roak et al., 2012a; Neale et al., 2012; Iossifov et al., 2012). Analysis of the trios from these studies, when considered jointly, identified *CHD8* and *SCN2A* as genes harboring recurrent, disruptive mutations in autistic patients. Deeper sequencing of 44 genes in another 2,446 patients also observed recurrent mutations in *DYRK1A*, *GRIN2B*, *TBR1*, *PTEN*, and *TBL1XR1* (O'Roak et al., 2012b). Notably, studies of *de novo* mutations in children with severe intellectual disability identify mutations in some of these same genes (Rauch et al., 2012; de Ligt et al., 2012). *De novo* mutations may make a smaller contribution to teen or adult-onset disorders such as schizophrenia: studies have not yet found statistically convincing levels of recurrent mutations in individual genes, though one study reports a greater-than-chance rate of mutations in cortically expressed genes as a group (Girard et al., 2011; Xu et al., 2012; Gulsuner et al., 2013).

The results of exome sequencing studies support models of significant polygenicity for autism and schizophrenia. Iossifov and colleagues estimate from the statistical distribution of disruptive

mutations across genes that 350–400 autism susceptibility loci exist in the genome—an estimate broadly consistent with estimates from the distribution of *de novo* CNVs (Iossifov et al., 2012; Sanders et al., 2011). Lim et al. (2013) estimate that about 5% of autism cases may also arise from noncomplementing genetic variants at dozens of different loci. Given this polygenicity, larger studies will be required to identify concentrations of mutations in additional genes at a level that is statistically meaningful.

Insights from Common Variation

Another powerful way to connect genome variation with complex phenotypes arises from the genetic variation that was present in the small, ancient populations that subsequently expanded across the world (Figure 1). Such alleles are today found among millions of people on almost every continent, in fact present in almost any large assembly of people. These common variants are finite and modest in number, numbering in the millions (rather than the billions), and can be cataloged by sequencing a few hundred individuals (Abecasis et al., 2012). Targeted molecular assays for these variants are then arrayed onto inexpensive array-based platforms for genome-wide SNP genotyping. This makes it possible to evaluate the reservoir of common polymorphism, more or less systematically, by using such arrays together with statistical techniques that impute the states of untyped variants from the typed ones (Li et al., 2009). The low cost of such platforms (many cost less than \$100 per genome) allows this approach to be brought to bear on many thousands of genomes. This sample size is critical for rigorously measuring the effect of a variant by studying it on thousands of genetic backgrounds and in diverse environments.

Common-variant association studies for psychiatric disorders appeared for many years to be unsuccessful, particularly when compared to the extensive gene discovery from common-variant approaches in autoimmune disease, cardiovascular disease, metabolic disease, stature, body mass, and other phenotypes. With hindsight, the problem was that studies of schizophrenia, the most deeply studied psychiatric disorder to date, were simply underpowered. These studies were initially being pursued at a scale insufficient to find all but the strongest effect (the *HLA* locus) in a disorder as highly polygenic as schizophrenia. As international collaborations and focused research resources have expanded sample size, common-variant association studies are finding far more genetic influences on risk. In schizophrenia, as sample sizes in international meta-analysis by the Psychiatric Genomics Consortium (PGC) reached 10,000 cases in 2011, another five schizophrenia loci were uncovered (Lee et al., 2012). A more recent expansion of sample size to include a large Swedish cohort found 22 loci with genome-wide levels of significance (Ripke et al., 2013). Most excitingly, an ongoing "stage 2" analysis by the PGC, comprising some 35,000 cases and a larger number of controls, is on a path to find 100 or more such influences with high levels of confidence. The genomic segments implicated in common-variant association studies are typically tens of kilobases long. A majority of them span just one to three genes, and about a third point crisply to a single gene. Such studies are nominating a substantial number of individual genes for biological analysis. Indeed, common-variant studies are creating an initial molecular

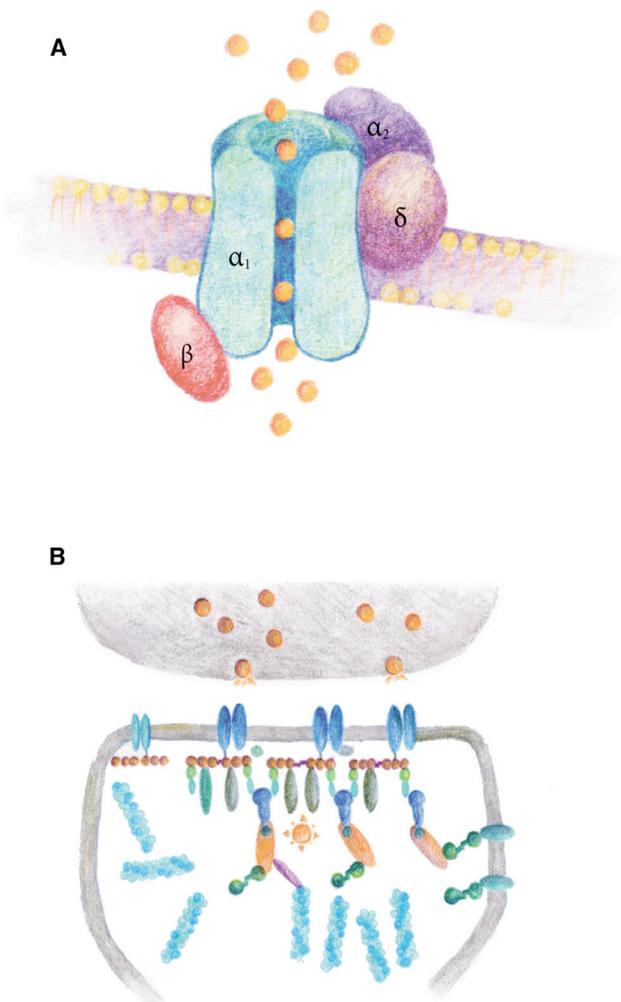


Figure 2. Emerging Genetic Evidence on Complex Brain Disorders Converges on Specific Molecular Complexes

(A) Multiple subunits of voltage-gated calcium channels are among the genome's strongest associations to schizophrenia and bipolar disorder. (B) Genes encoding the postsynaptic components of excitatory synapses are implicated by both rare and common variants in schizophrenia.

"parts list" for schizophrenia and may do so for bipolar disorder and autism when sample sizes catch up to the levels reached in schizophrenia. (Today, they are several times lower—approximately 7,500 and 3,000 cases, respectively, in published studies for bipolar disorder and autism [Psychiatric GWAS Consortium Bipolar Disorder Working Group, 2011; Anney et al., 2010; Weiss et al., 2009].)

Genetic Evidence Converges on Specific Molecular Complexes

Perhaps the most exciting aspect of the emerging schizophrenia genetics results is that constellations of genetic findings are converging on identifiable molecular complexes and pathways.

Voltage-Gated Calcium Channels

Common variants in the genes encoding multiple subunits of voltage-gated calcium channels are strongly implicated in schizophrenia and bipolar disorder (Figure 2). Common polymor-

phisms in the *CACNA1C* gene, which encodes a pore-forming subunit of the channel, are among the strongest associations in both schizophrenia and bipolar disorder (Ferreira et al., 2008; Psychiatric GWAS Consortium Bipolar Disorder Working Group, 2011; Ripke et al., 2013). Common polymorphisms in the *CACNB2* gene, encoding a regulatory subunit of the same channel, are also among the strongest associations for schizophrenia and cross-disorder risk (Lee et al., 2013; Ripke et al., 2013). Genes encoding the full set of potential subunits show a statistically remarkable level of association as a group (Psychiatric GWAS Consortium Bipolar Disorder Working Group, 2011). Together, these results suggest that these channels exhibit a surprising level of functional polymorphism in human populations and that this polymorphism shapes individuals' risk for schizophrenia and bipolar disorder.

Postsynaptic Components of Excitatory Synapses

Several genetic results implicate genes encoding the postsynaptic components of excitatory synapses. The de novo CNVs observed in schizophrenia patients have a strong statistical tendency to affect the genes defined through proteomics as components of the postsynaptic density (Kirov et al., 2012) (Figure 2). In emerging exome sequencing data, these same genes also appear to harbor loss-of-function variants in schizophrenia cases more frequently than in controls and to be enriched for de novo point mutations. Such results are likely to strengthen as the "synaptome" is more completely delineated in future experiments. As sample sizes expand, such results will also begin to implicate genes individually rather than as members of a group.

A Large Role for Regulatory DNA

While the genetic variation implicated in common-variant studies maps to neurobiologically meaningful and related sets of genes, it often maps to what are today the least interpretable components of those genes. Most frequently, the implicated haplotypes—sets of nearby alleles that segregate together—do not overlap with protein-coding sequence but rather reside within large introns and in sequences just upstream of the implicated genes. These results suggest that much of the variation among individuals in a population may arise not from broken proteins but from variation in the quantitative levels or cell-type-specific patterns with which these genes are expressed.

In organisms from plants to mammals, experiments on natural variation in traits within species have often suggested a large role for variation outside of the protein-coding sequences of genes. Variation in the regulatory parts of genomes allows nature to experiment with the place, time, quantity, and contingencies with which gene products become available to cells—variation that can shape behavioral variation among members of the same species (Young et al., 1999; Insel and Shapiro, 1992). Non-coding, regulatory parts of the genome may be vehicles for innovation on the rapid timescales that shape variation within species in their natural environments. This may be a way in which natural polymorphism is different from the mutations that scientists introduce in the genomes of isogenic model organisms to ascertain their ability to produce strong phenotypes that are outside the range of natural, common variation in phenotypes for members of that species.

An increasing number of genetic results fit a pattern in which rare, protein-disrupting variants cause severe, multiorgan system damage, which in the brain is manifest as significant intellectual disability and often epilepsy. In contrast, common, regulatory variants in the same genes cause milder phenotypes reflecting subsets of the tissues or cell types in which a gene is expressed. For example, voltage-gated calcium channels are essential for the function of the heart and other organs. Rare gain-of-function mutations in the coding sequence of the channel subunit *CACNA1C* cause Timothy Syndrome, a multiorgan disorder whose manifestations include potentially lethal cardiac arrhythmias, immune deficiency, cognitive disability, and autism (Splawski et al., 2004). Common variation in regulatory regions of the *CACNA1C* gene appears to result in localized perturbations of the gene's activity; this variation associates with a quantitative increase in risk of schizophrenia and bipolar disorder (approximately a 15% increase) without apparent association to cardiac or immune phenotypes (Ripke et al., 2013). As another example, disruptive mutations in the *TCF4* coding sequence cause Pitt-Hopkins syndrome, a condition characterized by microcephaly, severe intellectual disability (including, for example, the almost complete absence of language), and altered development of physical structures in many organ systems (Amiel et al., 2007). Specific noncoding variants in introns of *TCF4* associate with increased risk of schizophrenia (Lee et al., 2012) without producing phenotypes in other organ systems. Regulatory sequence may allow tissue- and cell-type-specific perturbations of a gene's activity—modest, quantitative perturbations rather than complete functional knockouts—and likely represent a larger part of natural variation in phenotypes that are observed in the general population. It is likely that common-variant association studies are giving us our first appreciation of how such regulatory, noncoding variation contributes to natural variation in genetically complex disease phenotypes in humans.

Further evidence for the regulatory nature of the variants implicated in common-variant association studies comes from the study of expression QTLs (eQTLs) in human tissues. The common variants that are implicated in genome-wide association studies tend also to associate with quantitative measurements of the expression levels of the same genes, especially when gene expression is measured in the tissue relevant to the disease (Nicolae et al., 2010; Richards et al., 2012).

Epigenetic Expression of Genetic Variation

Progress in the genome-scale analysis of chromatin states now reveals hundreds of thousands of sites across the genome that contain dynamic chromatin marks suggestive of tissue-specific enhancer activity—the ability to regulate the expression of nearby genes in specific tissues (Heintzman et al., 2009; Ernst et al., 2011; Bernstein et al., 2012). Enhancer sites tend to exhibit DNase hypersensitivity, suggesting that they are in open, accessible chromatin; they are also flanked by characteristic histone marks, including monomethylation of H3K4 and acetylation of H3K27 (Heintzman et al., 2009; Ernst et al., 2011; Thurman et al., 2012). Extensive new data from the ENCODE and Epigenomics Roadmap projects now document many ways in which chromatin states and DNA methylation implement the regulatory

instructions that are encoded in genomic sequence, although with a plasticity that makes them also responsive to cell type, cell state, and environment.

Recent studies indicate that associations of disease to common variants in the noncoding regions of genes involve sequence variation in putative enhancers as defined by epigenomic profiling. These relationships follow a tissue-and-disease logic: the common variants that associate to disease phenotypes tend to reside in the tissue-specific enhancers defined experimentally in the tissues thought to be most relevant to each disease (Maurano et al., 2012). Such results reinforce the conclusion that variation in gene regulation at many genomic loci contributes to complex, polygenic disease by acting in a tissue-specific manner.

The epigenomic profiles available in public resources today are derived from homogenized brain tissues that are mixtures of many cell types, including multiple neuronal and glial cell populations. The utilization of genomic sequence elements is ultimately a property of specific cell types, defined by their developmental lineage and functional properties. It will be important to understand how regulatory DNA elements are utilized by each specific cell population, both under baseline and stimulated conditions. An important breakthrough in helping make such studies possible involves the development of techniques, described below, for creating specific populations of human neurons by cellular reprogramming—for example, permitting the production of relatively pure populations of cortical interneurons (Maroof et al., 2013). Such technologies will make it possible to ascertain the specific segments of noncoding DNA that are utilized by each cell population and to connect specific disease-associated variants to perturbations of specific types of neurons and glia.

From Genetic Findings to Biological Insights

The recent success of genetic studies for highly polygenic brain disorders such as schizophrenia creates both a historical scientific opportunity and a formidable challenge for neurobiology. The opportunity inherent in having an initial molecular “parts list” for these disorders is clear. However, the challenges are also substantial. Historically, neurobiologists have investigated gene function by making highly penetrant mutations in individual genes, studying their effects on isogenic backgrounds, often inbred laboratory mouse strains, and focusing on phenotypes that are outside the range of natural phenotypic variation. In this way, a great deal has been learned about some aspects of rare and often severe monogenic diseases, whether of the nervous system or of other organ systems (Shahbazian et al., 2002; Peça et al., 2011). However, as described above, the genetic architecture of common polygenic diseases is quite different from either the severe mutations of rare monogenic disorders or artificial mutations (such as knockouts) made in laboratory mice. The genetic architecture of common polygenic diseases involves natural polymorphisms, including regulatory variants, whose ultimate contribution to phenotype is just one piece of a larger puzzle; such variants segregate on genetic backgrounds that contain many other risk and protective factors.

The resulting challenges have led some to suggest that biology should focus on the component of genetic architecture

that derives from rare, protein-altering mutations that are assumed to have large effects (McClellan and King, 2010). We think that to do so would miss the far larger scientific opportunity emerging from studies of polygenic disorders. Indeed to do so might miss the most important opportunities to address common serious diseases.

We recognize, however, that successful neurobiological analysis of polygenic disorders will require relatively new technologies and experimental approaches at scales that have not been typical for neuroscience. For example, the interrogation of large numbers of disease-associated genes and an even larger number of allelic variants within them, both individually and likely in combination, will require new approaches to living model systems. It would neither be practical nor likely given the modest penetrance of relevant alleles to make thousands of transgenic mice.

Alleles Lead to Genes and Pathways

Before addressing possible approaches to investigating genetic variants of modest penetrance that predispose to common diseases including psychiatric diseases, it is worth arguing that the scientific goals of such research are often misunderstood. It has seemed to some that gene discovery would be valuable, above all, to support new objective approaches to diagnosis, something that is sorely needed for psychiatric disorders (Hyman, 2007). There are at least two central obstacles in the way of this goal. The first is that given the large number of common and, more significantly, rare variants that likely contribute to polygenic disorders such as schizophrenia and autism, a very large catalog of risk alleles would be needed before a genetic test could be used diagnostically with reasonable probability. More important is the problem of pleiotropy, the phenomenon in which one gene can influence multiple phenotypes. For variants ranging from large CNVs to common variants detected by genome-wide association studies (GWASs), there is substantial sharing of genetic risk-conferring alleles across psychiatric disorders including autism, schizophrenia, bipolar disorder, major depressive disorder, and attention-deficit hyperactivity disorder (Sullivan et al., 2012; Smoller et al., 2013). Insofar as genetic tests will come to play a role in diagnosis, they will likely be most useful when combined with phenotypic information such as cognitive testing or imaging.

The far greater benefit of identifying genes is as clues to the biology of disease. While disease-associated alleles can be objects of study in their own right, they are often the most effective tools we have to identify genes relevant to pathogenesis. Beyond that, genes can serve as a tool for discovering pathways or molecular networks involved in the neurobiology of disease or can serve as the basis for molecular target discovery. The high population frequency of a common allele gives geneticists a foothold to rigorously quantify its contribution to a phenotype and to discover the effect in an unbiased genome-wide search. However, the particular allele does not establish an upper bound on the biological significance of the gene. Alleles of small effect routinely point to genes and pathways of large effect. For example, common-variant association studies of human lipid traits identified regulatory variants in an intron of the *HMGCR* gene; the common variants explain only a 3 mg/dl increase in

levels of low-density lipoprotein (LDL) in the blood, representing less than 1% of the heritability of this phenotype (Keebler et al., 2009). But pharmacological manipulations of the same pathway reduce LDL levels by 30%–60% and have done much to reduce deaths from cardiovascular disease. Thus, for example, the identification of risk alleles in GWASs for schizophrenia and bipolar disorder implicated several genes encoding subunits of L-type calcium channels in disease processes (Figure 2). The identification of risk alleles in genetic studies of schizophrenia and autism implicated protein networks in the postsynaptic specializations of excitatory synapses in these disorders (Figure 2). Having found a gene of interest, using rare or common alleles as pointers, neurobiologists are not limited to study the allele by which the gene was found. They can proceed to manipulate the gene and pathways in which its products function in powerful and creative ways.

Living Systems to Investigate Gene Function

The most obvious limitation for use of mouse models to study polygenic disorders, even with remarkably efficient new tools for genome engineering (Wang et al., 2013), is that they are not a high-throughput system. As such, the use of invertebrate animal models such as *Drosophila* or vertebrate models that reproduce more rapidly than mice, such as zebrafish, are likely to prove important—even though high enough throughput will remain a challenge. A second limitation to animal models is posed by evolution. In recent years, there has been increasing awareness, across many disease areas, that drugs that appear efficacious in mouse models often lack efficacy in humans. In nervous system disorders, substantial disillusionment with the ability of animal models to predict treatment efficacy (Nestler and Hyman, 2010; van der Worp et al., 2010) has contributed to many pharmaceutical companies de-emphasizing neurologic and psychiatric disorders. A recent workshop at the Institute of Medicine posed the question, why do many therapeutics show promise in preclinical animal models but then fail to elicit predicted effects when tested in humans (Institute of Medicine, 2013)? A key reason appears to be lack of evolutionary conservation of key molecular networks and circuits. For example, rodents are lissencephalic and lack a well-developed lateral prefrontal cortex, an evolutionarily new region of cortex that supports cognitive control in humans. Moreover, the largest number of disease associations found by GWASs in schizophrenia, for example, are in regulatory regions, the least well-conserved genomic elements between humans and rodents (Church et al., 2009) and indeed across all of evolution. Animal models will remain critical, especially because human brain disorders do not appear to be cell autonomous and, indeed, affect brain circuitry that involves a large number of different cell types. We would argue, however, that keeping both throughput and evolution in mind, it is critical to use the simplest living system possible that does not predispose to the blind alleys posed by phenocopies.

A technology that has recently gained attention for its potential utility in studying the function of human genes and their disease risk alleles is the use of human neurons derived from fibroblasts directly or through a stage of induced pluripotent cells (iPSCs) or from human embryonic stem cells (hESCs) (Son et al., 2011;

Zhang et al., 2013). Such human cell-based systems have the advantage of human transcriptional networks, indeed human genomes. They also have the advantage of permitting high-throughput interrogation of gene function when combined with recent gene engineering technologies (Cong et al., 2013) that are still being improved upon. It is also possible to study the function of one or several genes or gene variants against diverse isogenic background from individuals with no known brain disorder as well as “rescue” patient-derived cells by engineering risk variant out of their cells.

For all this promise, there is a long distance to travel before cell-based systems are optimized as complements to carefully interpreted animal models. The ability to differentiate fibroblasts, iPS cells, or hESC cells into diverse mature neurons and glia remains at an early stage, although significant progress is being made (Son et al., 2011; Rouaux and Arlotta, 2013; Maroof et al., 2013). In addition, technologies needed to compare cultured neurons made by different methods with postmortem human neurons, such as single-cell RNA sequencing or single-cell proteomics, are at different and often early stages of development. For schizophrenia, both cortical pyramidal neurons and certain parvalbumin-expressing interneurons have been implicated in the disease process (Lewis and Sweet, 2009); however, for essentially all neuropsychiatric disorders, including schizophrenia, the relevant cell types to be modeled remain poorly understood (see below). Another challenge, based on the importance of synaptic proteins, described above, in autism and schizophrenia, will be the ability to make replicable and robust small neural circuits in vitro. Such hurdles notwithstanding, the ability to make human neurons in vitro is likely to prove a critically important approach to the study of disease.

Finding the Cell Populations that Matter

Insights into the etiology of an illness often arise when science identifies the specific cell populations in which a disease process begins. Such insights would be of enormous value in schizophrenia, bipolar disorder, and autism, in which the culpable cell populations and circuits are not yet known. Genetics may be particularly valuable in distinguishing phenomena that are *causal* from phenomena that are mere markers for disease progression or biological accommodations to the disease state: since the genetic alleles existed long before the pathology itself, their association to phenotypes must reflect a directional effect on disease processes rather than a response to them. We believe that an important direction in leveraging emerging genetic results in polygenic disease is to use them to find the cell types that matter to each disorder. We propose three ways in which this will be possible. First, some of the implicated genes may have patterns of expression that localize their effects to specific cell populations. Second, regulatory variants may act in geographically restricted ways, altering the expression level in some but not all of the cell populations in which a gene is expressed. Third, some populations of cells may show selective vulnerability to a genetic perturbation, manifesting a difference in cell state. Creative applications of RNA sequencing, for example, analyzing individual cells, have the potential to help uncover such relationships.

High-Throughput Biology and Multidimensional Data Analysis

Another potential strategy is to embrace the highly polygenic nature of risk for many common disorders and to develop biological research strategies that explore the functions of many genes at once, looking for patterns and convergent phenotypes across a large set of biological perturbations. Gene knockdown and overexpression can be pursued in parallel in large numbers of genes in parallel assays. Genome editing, which is being greatly simplified by recent technological innovation (Cong et al., 2013; Fenno et al., 2011; Zhang et al., 2013), may also soon be amenable to high-throughput strategies. The challenge of making sense of the long gene lists emerging from human genetics may be one that is well served by a period of high-throughput biology and multidimensional data analysis, which can be used to develop testable hypotheses about the pathophysiology involved in disease.

Conclusion

The combination of advances in genetic technologies and the creation of large consortia to identify study populations adequate to investigate polygenic disease has led to the first breakthroughs in common, severe psychiatric disorders. As described, schizophrenia is the best studied, with the expectation that more than 100 genome-wide significant associated loci will be known in the near future. Bipolar disorder and autism await the identification of larger populations, but the “playbook” for identifying molecular risk factors in polygenic brain disorders is now clear. As we have described, the exploitation of emerging findings to achieve deep understandings of pathogenesis and to design much needed new treatments is likely to depend on new approaches to neurobiology that are higher throughput than most traditional investigations and the careful development and exploitation of new model living systems, including human neurons in vitro. We are just at the beginning of this journey, which should provide rich new discoveries for basic science, but the pressure to succeed is driven far more by the lack of effective treatments for so many individuals with psychiatric disorders.

ACKNOWLEDGMENTS

This work was made possible by funding from the Stanley Medical Research Institute and by NIH grant R01 HG006855 to S.A.M. We also thank Janet Theurer and Pat Rossi for the artwork. S.E.H. discloses that he serves on the Novartis Science Board and has advised AstraZeneca.

REFERENCES

- Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., and McVean, G.A.; 1000 Genomes Project Consortium. (2010). A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–1073.
- Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A.; 1000 Genomes Project Consortium. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* 491, 56–65.
- Amiel, J., Rio, M., de Pontual, L., Redon, R., Malan, V., Boddaert, N., Plouin, P., Carter, N.P., Lyonnet, S., Munnich, A., and Colleaux, L. (2007). Mutations in TCF4, encoding a class I basic helix-loop-helix transcription factor, are responsible for Pitt-Hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction. *Am. J. Hum. Genet.* 80, 988–993.

- Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U., and Zoghbi, H.Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.* 23, 185–188.
- Anney, R., Klei, L., Pinto, D., Regan, R., Conroy, J., Magalhaes, T.R., Correia, C., Abrahams, B.S., Sykes, N., Pagnamenta, A.T., et al. (2010). A genome-wide scan for common alleles affecting risk for autism. *Hum. Mol. Genet.* 19, 4072–4082.
- Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C., and Snyder, M.; ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74.
- Church, D.M., Goodstadt, L., Hillier, L.W., Zody, M.C., Goldstein, S., She, X., Bult, C.J., Agarwala, R., Cherry, J.L., DiCuccio, M., et al.; Mouse Genome Sequencing Consortium. (2009). Lineage-specific biology revealed by a finished genome assembly of the mouse. *PLoS Biol.* 7, e1000112.
- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823.
- Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P., et al.; Wellcome Trust Case Control Consortium. (2010). Origins and functional impact of copy number variation in the human genome. *Nature* 464, 704–712.
- Darnell, J.C., Van Driesche, S.J., Zhang, C., Hung, K.Y., Mele, A., Fraser, C.E., Stone, E.F., Chen, C., Fak, J.J., Chi, S.W., et al. (2011). FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* 146, 247–261.
- de Ligt, J., Willemsen, M.H., van Bon, B.W., Kleefstra, T., Yntema, H.G., Kroes, T., Vulto-van Silfhout, A.T., Koolen, D.A., de Vries, P., Gilissen, C., et al. (2012). Diagnostic exome sequencing in persons with severe intellectual disability. *N. Engl. J. Med.* 367, 1921–1929.
- Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shores, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473, 43–49.
- Fenko, L., Yizhar, O., and Deisseroth, K. (2011). The development and application of optogenetics. *Annu. Rev. Neurosci.* 34, 389–412.
- Ferreira, M.A., O'Donovan, M.C., Meng, Y.A., Jones, I.R., Ruderfer, D.M., Jones, L., Fan, J., Kirov, G., Perlis, R.H., Green, E.K., et al.; Wellcome Trust Case Control Consortium. (2008). Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat. Genet.* 40, 1056–1058.
- Fu, Y.H., Kuhl, D.P., Pizzuti, A., Pieretti, M., Sutcliffe, J.S., Richards, S., Verkerk, A.J., Holden, J.J., Fenwick, R.G., Jr., Warren, S.T., et al. (1991). Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell* 67, 1047–1058.
- Fu, J., Wolfs, M.G., Deelen, P., Westra, H.J., Fehrmann, R.S., Te Meerman, G.J., Buurman, W.A., Rensen, S.S., Groen, H.J., Weersma, R.K., et al. (2012). Unraveling the regulatory mechanisms underlying tissue-dependent genetic variation of gene expression. *PLoS Genet.* 8, e1002431.
- Girard, S.L., Gauthier, J., Noreau, A., Xiong, L., Zhou, S., Jouan, L., Dionne-Laporte, A., Spiegelman, D., Henrion, E., Diallo, O., et al. (2011). Increased exonic de novo mutation rate in individuals with schizophrenia. *Nat. Genet.* 43, 860–863.
- Gulsuner, S., Walsh, T., Watts, A.C., Lee, M.K., Thornton, A.M., Casadei, S., Rippey, C., Shahin, H., Nimgaonkar, V.L., Go, R.C., et al.; Consortium on the Genetics of Schizophrenia (COGS); PAARTNERS Study Group. (2013). Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* 154, 518–529.
- Handsaker, R.E., Korn, J.M., Nemesh, J., and McCarroll, S.A. (2011). Discovery and genotyping of genome structural polymorphism by sequencing on a population scale. *Nat. Genet.* 43, 269–276.
- Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459, 108–112.
- Hyman, S.E. (2007). Can neuroscience be integrated into the DSM-V? *Nat. Rev. Neurosci.* 8, 725–732.
- Insel, T.R., and Shapiro, L.E. (1992). Oxytocin receptor distribution reflects social organization in monogamous and polygamous voles. *Proc. Natl. Acad. Sci. USA* 89, 5981–5985.
- Institute of Medicine. (2013). Improving the Utility and Translation of Animal Models for Nervous System Disorders: Workshop Summary. (Washington, DC: The National Academies Press).
- International Multiple Sclerosis Genetics Consortium. (2013). Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat. Genet.* Published online September 29, 2013. <http://dx.doi.org/10.1038/ng.2770>.
- Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., Yamrom, B., Lee, Y.H., Narzisi, G., Leotta, A., et al. (2012). De novo gene disruptions in children on the autistic spectrum. *Neuron* 74, 285–299.
- Keebler, M.E., Sanders, C.L., Surti, A., Guiducci, C., Burt, N.P., and Kathiresan, S. (2009). Association of blood lipids with common DNA sequence variants at 19 genetic loci in the multiethnic United States National Health and Nutrition Examination Survey III. *Circ Cardiovasc Genet* 2, 238–243.
- Kirov, G., Pocklington, A.J., Holmans, P., Ivanov, D., Ikeda, M., Ruderfer, D., Moran, J., Chambert, K., Toncheva, D., Georgieva, L., et al. (2012). De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol. Psychiatry* 17, 142–153.
- Lancaster, M.A., Renner, M., Martin, C.A., Wenzel, D., Bicknell, L.S., Hurles, M.E., Homfray, T., Penninger, J.M., Jackson, A.P., and Knoblich, J.A. (2013). Cerebral organoids model human brain development and microcephaly. *Nature* 501, 373–379.
- Lee, S.S., Wan, M., and Francke, U. (2001). Spectrum of MECP2 mutations in Rett syndrome. *Brain Dev.* 23(Suppl 1), S138–S143.
- Lee, S.H., DeCandia, T.R., Ripke, S., Yang, J., Sullivan, P.F., Goddard, M.E., Keller, M.C., Visscher, P.M., and Wray, N.R. Schizophrenia Psychiatric Genome-Wide Association Study Consortium (PGC-SCZ); International Schizophrenia Consortium (ISC); Molecular Genetics of Schizophrenia Collaboration (MGS). (2012). Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat. Genet.* 44, 247–250.
- Lee, S.H., Ripke, S., Neale, B.M., Faraone, S.V., Purcell, S.M., Perlis, R.H., Mowry, B.J., Thapar, A., Goddard, M.E., Witte, J.S., et al.; Cross-Disorder Group of the Psychiatric Genomics Consortium; International Inflammatory Bowel Disease Genetics Consortium (IBDGC). (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* 45, 984–994.
- Lewis, D.A., and Sweet, R.A. (2009). Schizophrenia from a neural circuitry perspective: advancing toward rational pharmacological therapies. *J. Clin. Invest.* 119, 706–716.
- Li, Y., Willer, C., Sanna, S., and Abecasis, G. (2009). Genotype imputation. *Annu. Rev. Genomics Hum. Genet.* 10, 387–406.
- Lim, E.T., Raychaudhuri, S., Sanders, S.J., Stevens, C., Sabo, A., MacArthur, D.G., Neale, B.M., Kirby, A., Ruderfer, D.M., Fromer, M., et al.; NHLBI Exome Sequencing Project. (2013). Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. *Neuron* 77, 235–242.
- MacArthur, D.G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins, L., Habegger, L., Pickrell, J.K., Montgomery, S.B., et al.; 1000 Genomes Project Consortium. (2012). A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 335, 823–828.
- Maroof, A.M., Keros, S., Tyson, J.A., Ying, S.W., Ganat, Y.M., Merkle, F.T., Liu, B., Goulburn, A., Stanley, E.G., Elefany, A.G., et al. (2013). Directed differentiation and functional maturation of cortical interneurons from human embryonic stem cells. *Cell Stem Cell* 12, 559–572.
- Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337, 1190–1195.

- McClellan, J., and King, M.C. (2010). Genetic heterogeneity in human disease. *Cell* 141, 210–217.
- Neale, B.M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K.E., Sabo, A., Lin, C.F., Stevens, C., Wang, L.S., Makarov, V., et al. (2012). Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 485, 242–245.
- Nestler, E.J., and Hyman, S.E. (2010). Animal models of neuropsychiatric disorders. *Nat. Neurosci.* 13, 1161–1169.
- Nica, A.C., Parts, L., Glass, D., Nisbet, J., Barrett, A., Sekowska, M., Travers, M., Potter, S., Grundberg, E., Small, K., et al.; MuTHER Consortium. (2011). The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet.* 7, e1002003.
- Nicolae, D.L., Gamazon, E., Zhang, W., Duan, S., Dolan, M.E., and Cox, N.J. (2010). Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* 6, e1000888.
- O'Roak, B.J., Vives, L., Fu, W., Egerton, J.D., Stanaway, I.B., Phelps, I.G., Carvill, G., Kumar, A., Lee, C., Ankenman, K., et al. (2012a). Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* 338, 1619–1622.
- O'Roak, B.J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B.P., Levy, R., Ko, A., Lee, C., Smith, J.D., et al. (2012b). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 485, 246–250.
- Peça, J., Feliciano, C., Ting, J.T., Wang, W., Wells, M.F., Venkatraman, T.N., Lascola, C.D., Fu, Z., and Feng, G. (2011). Shank3 mutant mice display autistic-like behaviours and striatal dysfunction. *Nature* 472, 437–442.
- 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, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., and Sklar, P.; International Schizophrenia Consortium. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460, 748–752.
- Rauch, A., Wiczorek, D., Graf, E., Wieland, T., Ende, S., Schwarzmayr, T., Albrecht, B., Bartholdi, D., Beygo, J., Di Donato, N., et al. (2012). Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* 380, 1674–1682.
- Richards, A.L., Jones, L., Moskvina, V., Kirov, G., Gejman, P.V., Levinson, D.F., Sanders, A.R., Purcell, S., Visscher, P.M., Craddock, N., et al.; Molecular Genetics of Schizophrenia Collaboration (MGS); International Schizophrenia Consortium (ISC). (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, J.L., Kähler, A.K., Akterin, S., Bergen, S.E., Collins, A.L., Crowley, J.J., Fromer, M., et al.; Multicenter Genetic Studies of Schizophrenia Consortium; Psychosis Endophenotypes International Consortium; Wellcome Trust Case Control Consortium 2; Management Committee; Data and Analysis Group; DNA, Genotyping, Data QC and Informatics Group; Publications Committee. (2013). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat. Genet.* 45, 1150–1159.
- Ronald, A., and Hoekstra, R.A. (2011). Autism spectrum disorders and autistic traits: a decade of new twin studies. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* 156B, 255–274.
- Rouaux, C., and Arlotta, P. (2013). Direct lineage reprogramming of post-mitotic callosal neurons into corticofugal neurons in vivo. *Nat. Cell Biol.* 15, 214–221.
- Ruderfer, D.M., Kirov, G., Chambert, K., Moran, J.L., Owen, M.J., O'Donovan, M.C., Sklar, P., and Purcell, S.M. (2011). A family-based study of common polygenic variation and risk of schizophrenia. *Mol. Psychiatry* 16, 887–888.
- Sanders, S.J., Ercan-Sencicek, A.G., Hus, V., Luo, R., Murtha, M.T., Moreno-De-Luca, D., Chu, S.H., Moreau, M.P., Gupta, A.R., Thomson, S.A., 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.
- Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J., Ercan-Sencicek, A.G., DiLullo, N.M., Parikshak, N.N., Stein, J.L., et al. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 485, 237–241.
- Shahbazi, M., Young, J., Yuva-Paylor, L., Spencer, C., Antalfy, B., Noebels, J., Armstrong, D., Paylor, R., and Zoghbi, H. (2002). Mice with truncated MeCP2 recapitulate many Rett syndrome features and display hyperacetylation of histone H3. *Neuron* 35, 243–254.
- Smoller, J.W., and Finn, C.T. (2003). Family, twin, and adoption studies of bipolar disorder. *Am. J. Med. Genet. C. Semin. Med. Genet.* 123C, 48–58.
- Smoller, J.W., Craddock, N., Kendler, K., Lee, P.H., Neale, B.M., Nurnberger, J.I., Ripke, S., Santangelo, S., and Sullivan, P.F.; Cross-Disorder Group of the Psychiatric Genomics Consortium. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* 381, 1371–1379.
- Son, E.Y., Ichida, J.K., Wainger, B.J., Toma, J.S., Rafuse, V.F., Woolf, C.J., and Eggan, K. (2011). Conversion of mouse and human fibroblasts into functional spinal motor neurons. *Cell Stem Cell* 9, 205–218.
- Splawski, I., Timothy, K.W., Sharpe, L.M., Decher, N., Kumar, P., Bloise, R., Napolitano, C., Schwartz, P.J., Joseph, R.M., Condouris, K., et al. (2004). CaV1.2 calcium channel dysfunction causes a multisystem disorder including arrhythmia and autism. *Cell* 119, 19–31.
- Sullivan, P.F., Kendler, K.S., and Neale, M.C. (2003). Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch. Gen. Psychiatry* 60, 1187–1192.
- Sullivan, P.F., Daly, M.J., and O'Donovan, M. (2012). Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nat. Rev. Genet.* 13, 537–551.
- Suter, B., Treadwell-Deering, D., Zoghbi, H.Y., Glaze, D.G., and Neul, J.L. (2013). Brief report: MECP2 mutations in people without Rett syndrome. *J. Autism Dev. Disord.* Published online August 7, 2013. <http://dx.doi.org/10.1007/s10803-013-1902-z>.
- The Huntington's Disease Collaborative Research Group. (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 72, 971–983.
- Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. *Nature* 489, 75–82.
- van der Worp, H.B., Howells, D.W., Sena, E.S., Porritt, M.J., Rewell, S., O'Collins, V., and Macleod, M.R. (2010). Can animal models of disease reliably inform human studies? *PLoS Med.* 7, e1000245.
- Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., and Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* 153, 910–918.
- Weiss, L.A., Arking, D.E., Daly, M.J., and Chakravarti, A.; Gene Discovery Project of Johns Hopkins & the Autism Consortium. (2009). A genome-wide linkage and association scan reveals novel loci for autism. *Nature* 461, 802–808.
- Xu, B., Ionita-Laza, I., Roos, J.L., Boone, B., Woodrick, S., Sun, Y., Levy, S., Gogos, J.A., and Karayiorgou, M. (2012). De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nat. Genet.* 44, 1365–1369.
- Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42, 565–569.
- Young, L.J., Nilsen, R., Waymire, K.G., MacGregor, G.R., and Insel, T.R. (1999). Increased affiliative response to vasopressin in mice expressing the V1a receptor from a monogamous vole. *Nature* 400, 766–768.
- Zhang, Y., Pak, C., Han, Y., Ahlenius, H., Zhang, Z., Chanda, S., Marro, S., Patzke, C., Acuna, C., Covy, J., et al. (2013). Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* 78, 785–798.