Psychiatric Genetics and the Structure of Psychopathology

Jordan W. Smoller, MD ScD1,2,3, Ole A. Andreassen, MD PhD5,6, Howard J. Edenberg, PhD6, Stephen V. Faraone, PhD7, Stephen J. Glatt, PhD,7 Kenneth S. Kendler, MD8,9

Second Revision 10/23/17

1Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA
2Department of Psychiatry, Massachusetts General Hospital, Boston, MA
3Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA
4 NORMENT - KG Jebsen Centre, University of Oslo, Oslo, Norway
5Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway
6Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN
7Departments of Psychiatry and of Neuroscience and Physiology, SUNY Upstate Medical University, Syracuse, NY;
8 Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA.
9Department of Psychiatry, Virginia Commonwealth University, Richmond VA, USA.

Correspondence to:
Jordan W. Smoller, MD, ScD
Simches Research Building
185 Cambridge St.
Boston, MA 02114
Phone: 617-724-0835
Fax: 617-643-3080
Email: jsmoller@mgh.harvard.edu

Keywords: psychiatric genetics; nosology; heritability; cross-disorder; GWAS

This is the author's manuscript of the article published in final edited form as:

Abstract

For over a century, psychiatric disorders have been defined by expert opinion and clinical observation. The modern DSM has relied on a consensus of experts to define categorical syndromes based on clusters of symptoms and signs, and, to some extent, external validators such as longitudinal course and response to treatment. In the absence of an established etiology, psychiatry has struggled to validate these descriptive syndromes, and to define the boundaries between disorders and between normal and pathologic variation. Recent advances in genomic research, coupled with large-scale collaborative efforts like the Psychiatric Genomics Consortium, have identified hundreds of common and rare genetic variations that contribute to a range of neuropsychiatric disorders. At the same time, they have begun to address deeper questions about the structure and classification of mental disorders: To what extent do genetic findings support or challenge our clinical nosology? Are there genetic boundaries between psychiatric and neurologic illness? Do the data support a boundary between disorder and normal variation? Is it possible to envision a nosology based on genetically informed disease mechanisms? This review provides an overview of conceptual issues and genetic findings that bear on the relationships among and boundaries between psychiatric disorders and other conditions. We highlight implications for the evolving classification of psychopathology and the challenges for clinical translation.
The fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM5\(^1\)) is the latest update of a prevailing diagnostic approach to psychiatric illness. The DSM consists of symptoms and signs of illness, often complemented by requirements for a particular duration and associated distress or disability. Diagnoses are arrived at by checklists, with diseases defined by presence of some minimal number of criteria, often leaving substantial clinical heterogeneity within disorders. While useful clinically, the validity of the DSM boundaries is uncertain. The underlying genetics do not appear to precisely support such definitions. The genetic overlap among clinically-defined psychiatric syndromes was first demonstrated by family and twin studies. More recently, the work of large-scale collaborations, most notably the Psychiatric Genomics Consortium (PGC), has enabled DNA-level studies of genetic relationships among psychiatric disorders and related traits. Here, we review the state of the genetic structure of psychiatric disorders and its implications for psychiatric nosology.

**What does Genetic Epidemiology tell us about the Structure of Psychopathology?**

While the observation that psychiatric disorders ran in families can be traced back to antiquity, psychiatric genetics as a scientific discipline arose in the 20\(^{th}\) century with the first major family, twin, and adoption studies being published, respectively, in 1916\(^2\), 1928\(^3\) and 1966\(^4\). Family studies using modern methodology show substantial familial aggregation for all major psychiatric disorders including depression, schizophrenia, bipolar disorder, and alcohol dependence, as well as many other syndromes such as panic disorder, ADHD, drug abuse, autism, and obsessive-compulsive disorders\(^5\). Early twin studies, based on hospital or national twin registries, showed evidence of strong heritable effects for schizophrenia and bipolar disorder. Later, population-based twin studies showed
substantial genetic influences on conditions such as major depression, eating and anxiety disorders, alcoholism, ADHD and personality disorders\textsuperscript{6,7}. Adoption studies are more difficult to perform, but show heritable risks for schizophrenia (e.g. \textsuperscript{8}), ADHD (e.g. \textsuperscript{9}) and alcohol dependence (e.g. \textsuperscript{10}).

An important question first addressed by family studies was whether there was familial overlap among different disorders. When examined with sufficient power, the answer was nearly always that a range of psychiatric disorders cluster together in families. Examples include bipolar disorder, ADHD and depression; depression and anxiety; mood disorders, ADHD and substance abuse; and schizophrenia and schizotypal personality disorder. When this question was examined in twin and adoption studies, similar overlaps were found, and could be attributed to heritable factors that at least partially overlap (Table 1). One population-based twin study showed that common axis-I and axis-II (personality) psychiatric disorders had a coherent underlying genetic structure that reflected just two major dimensions, illustrating the extensive sharing of heritable risk factors across disorders\textsuperscript{11}. In sum, family and twin studies have long suggested that heritable influences on psychopathology transcend diagnostic boundaries; our genes don’t seem to have read the DSM.

**How Does Molecular Genetics Confirm and Extend Genetic Epidemiologic Findings?**

Studies in the early years of psychiatric molecular genetics tested the parsimonious hypothesis that single variants of large effect would underlie the etiology of psychiatric disorders, but these were not realized except for rare cases. The method of genome-wide association studies (GWAS) gave us a tool for identifying common DNA risk variants (single
nucleotide polymorphisms, SNPs) along with rare copy-number variants (CNVs). This method began to produce results with schizophrenia in 2008 and 2009. Since then, an avalanche of molecular genetic data has generated trustworthy findings for schizophrenia, bipolar disorder, major depressive disorder, ADHD and autism. A broad picture of the genetic architecture of these disorders is emerging. A large portion of the genetic risk for each disorder appears to result from many common SNPs of individually small effect size, with additional contributions from relatively rare (but somewhat more penetrant) CNVs. With the advent of full genome and exome sequencing, rare single-nucleotide variants (SNVs) are also being discovered. One lesson from this work is that very large studies and mega-analyses are required to find these variants. Such analyses have become possible because of a sea change in the culture of psychiatric genomic research. Investigators across the world have come together to share, and often jointly analyze, genetic data through large-scale collaborations and consortia, most notably the PGC and iPSYCH. The identification by GWAS of common variants reliably associated with psychiatric disorders has followed a trajectory similar to that of other common complex biomedical disorders. Early efforts from individual research teams studying relatively small samples yielded little fruit, power analyses suggesting the need for much larger samples. Consortia were formed to aggregate samples sufficient to reliably detect small effects of common variants. The value of such efforts has been made clear by the growing catalogue of robustly associated common variants, including more than 100 associated with schizophrenia in a PGC analysis of nearly 37,000 cases and more than 113,000 controls. Individual SNPs associated with these diseases explain only a tiny fraction of the heritable
variance for psychiatric disorders, with most individual risk-predisposing alleles associated with odds ratios of 1.1 or less. However, the aggregation of these effects into polygenic risk scores (PRSs)\(^{19}\) that capture the additive effects of thousands of SNPs accounts for substantially larger fractions of the heritable variance. A range of statistical methods allow the estimation of heritability based on genomewide SNPs, often referred to as *SNP-chip heritability* \(h^2_{SNP}\)\(^{19}\). Applied to psychiatric disorders, these methods show that the genetic architecture includes a substantial contribution of common variation, though the estimates vary from less than 10% (for anxiety disorders and PTSD)\(^{20,21}\) to more than 20% (for ADHD, bipolar disorder and schizophrenia)\(^{22,23}\) (see Table 1). These estimates are lower than those derived from twin studies, in part because \(h^2_{SNP}\) only includes effects due to common variants.

Genomic studies also show that rare, *de novo* variations (arising in the gametes or embryo) contribute to psychiatric disorders. Thus far, this has most convincingly been demonstrated for autism spectrum disorder and schizophrenia. For example, the largest analysis of CNV data, by the PGC’s CNV and Schizophrenia Workgroups, comprising more than 40,000 subjects, identified eight CNV loci associated with schizophrenia risk\(^{24}\). Although these loci were rare (most with a frequency of <0.5% among cases), their effects were much larger (ORs ranging from 3.8 to more than 67) than those seen with individual common variants. They include a deletion on chromosome 22 which is the cause of 22qdel syndrome (velocardiofacial syndrome) and has long been recognized as carrying a high risk for psychosis. The contribution of specific rare SNVs has been harder to establish, though large-scale sequencing of the exome has shown that the aggregate burden of rare protein-altering variants is elevated in schizophrenia\(^{25}\), autism, \(^{26}\) and Tourette syndrome.
In addition, mutations in several genes have been associated with autism spectrum disorders (e.g. CHD8, SCN2A, SHANK3, GRIN2B)\textsuperscript{28,29} and, to a lesser extent, schizophrenia (SETD1A)\textsuperscript{30}.

Genome-wide studies have documented that psychiatric disorders are highly polygenic, reflecting a combination of thousands of common variants of individually small effect and rarer variants of larger effect. With the disorders best characterized, especially schizophrenia, it is increasingly clear that a substantial fraction of genetic risk is the result of common SNP variation. As sample sizes grow, additional variants are expected to be found. With these discoveries, however, has arisen the challenge of translating these molecular findings into insights about etiologic pathways.

**Cross-disorder studies**

Similar to family and twin studies, GWAS first focused on single disorders and then examined if risk-conferring variants for one disorder affect risks for others; \textit{i.e.}, pleiotropy. (Figure 1 depicts common approaches for evaluating cross-disorder genetic effects). Given the large number of human phenotypes and the limited number of genes, it is not surprising that pleiotropy is a common phenomenon for many traits and disorders, by no means specific to psychiatry\textsuperscript{31}. Among the earliest evidence of shared molecular genetic influences across neuropsychiatric disorders were findings that rare CNVs are associated with multiple disorders including autism, ADHD, and schizophrenia as well as epilepsy and intellectual disability\textsuperscript{32}. GWAS data have also been used to examine the cross-disorder sharing of common variants. For example, the International Schizophrenia Consortium reported that a PRS derived from GWAS of schizophrenia was strongly associated with risk
for bipolar disorder. Subsequently, PRS have been used to demonstrate cross-disorder genetic overlap of a wide range of psychiatric phenotypes. For example, schizophrenia PRS has been associated with psychotic experience and schizoaffective disorder as well as related phenotypes such as cognitive ability, sensory motor gating, working memory brain activation (fMRI signal), childhood neurodevelopmental impairments, major depressive disorder, ADHD and conduct/oppositional defiant disorder, adolescent anxiety disorder, and PTSD.

SNP-chip heritability methods have been extended to estimate the genetic correlation between pairs of disorders averaged over all common SNPs. For example, the Cross-Disorder Group of the PGC reported significant genetic correlations of schizophrenia with bipolar disorder (0.68), major depressive disorder (0.43) and autism (0.16), and between major depressive disorder and bipolar disorder (0.47) and ADHD (0.32) (see Table 1). A recent analysis of of the Big Five personality traits further demonstrates specific patterns of genetic correlation with psychiatric disorders. For a database presenting the many statistically significant molecular genetic correlations among psychiatric disorders and other phenotypes see LD Hub.

Individual common variants have also shown pleiotropic effects across clinically distinct disorders, providing initial clues to a shared biology. Network and pathway analyses have pointed to biological pathways that show genetic association across diagnostic groups. For example, genes related to calcium channel signaling, histone methylation, synaptic function, and immune function have been implicated in both mood disorders and schizophrenia.
These results suggest that susceptibility to each psychiatric disorder, as currently defined by DSM, is influenced by many genetic risk factors and that any given psychiatric disorder will share some genetic risk factors with others. This risk-sharing extends beyond the genome to the environment, with early developmental insults and trauma implicated in risk for several disorders as well. This knowledge has large implications for understanding the pathology of psychiatric disorders. However, advances in our statistical methods and a range of in vitro and in vivo experimental studies are needed to translate the shared genetic and environmental association signals into molecular genetic mechanisms.

**What does it all mean?**

*Psychiatric disorders are highly polygenic*

Family and twin studies, now complemented by molecular genetics, show that genetic variation accounts for a substantial portion of the risk for psychiatric disorders, and that some of the heritability is shared among disorders. However, many variants remain to be discovered, and this will require much larger sample sizes, and additional cross-disorder studies. The history of schizophrenia genetics provides a compelling example that simply expanding sample size is a winning strategy for driving discovery. But there are obstacles. To make more progress, funders will have to accept the drudgery of “normal science” rather than uncritically requiring innovative methods of gene discovery.

*Discovering biological mechanisms from common variants will be challenging.*

Going from a GWAS-associated locus to identifying causal variation underlying the association is a challenge, but one that can be overcome with significant effort. A good example is Sekar et al.’s study of common variation associated with schizophrenia in the
major histocompatibility complex, which identified functional alleles of the complement component 4 (C4) genes and implicated their role in microglia-mediated synaptic pruning. Identifying the causal alleles required complex fine-mapping followed by functional studies of gene expression in mouse and human brain, and high-resolution immunohistochemistry.

Another issue is that most common risk variants individually contribute only a very small portion of overall susceptibility to common complex psychiatric disorders. Mechanistic studies must eventually integrate the effects of many risk loci. Fortunately, powerful new technologies for studying systems genomics and neuroscience (including the use of induced pluripotent stem cells, gene-editing methods, and optogenetic interrogation of brain circuits) provide new opportunities for dissecting the functional effects of risk variants and pathways discovered from them. As the growing catalog of established genetic variants converge on biological pathways, new opportunities arise for identifying treatment targets. For decades, psychiatric drug development has focused on pathways modulated by drugs that had been serendipitously discovered. By moving the focus of drug discovery to genetically validated functional pathways, GWAS can point to new “druggable” targets, enabling the development of novel therapeutics for specific or cross-diagnostic clinical characteristics.

Discovering biological mechanisms from rare variants of larger-effect may be easier. Unlike common SNPs, rare disease-associated mutations often have direct functional consequences, making them especially suitable for mechanistic studies. Some have argued the mechanisms discovered via rare variants will not be relevant to the large majority of patients, but empirical examples show that a biological pathway implicated by a rare variant is relevant to mechanisms implicated by common variation. For example, rare
variants of \textit{PCSK9} cause a rare autosomal dominant familial hypercholesterolemia. Drugs that inhibit PCSK9 protein lower cholesterol and are a viable treatment for common forms of hypercholesterolemia and atherosclerosis.\textsuperscript{49}

\textit{Genetic studies challenge the DSM paradigm.}

The pervasive cross-disorder heritability of psychiatric disorders challenges the DSM paradigm which, from its inception, emphasized hierarchical diagnoses disallowing diagnoses of some disorders in the presence of others. Twin data point to genetic hierarchies with a general psychopathology factor, internalizing and externalizing factors, along with unique sources of heritability for each disorder\textsuperscript{50}. Twin studies have also reported genetic correlations between disorders and the normal range of personality variation (e.g. \textsuperscript{51}) and support the view that some disorders are the extreme of a continuous trait in the population \textsuperscript{52, 53}. Genomic studies are also questioning the idea that psychiatric conditions are discrete entities by demonstrating substantial molecular genetic correlations across diagnostic categories and between disorders and normal ranges of phenotypic variation in the population\textsuperscript{54-57}. In some cases, these findings challenge fundamental assumptions of our clinical nosology. For example, the separation of schizophrenia and bipolar disorder has been a foundational distinction for psychiatric classification, dating to Kraepelin’s work more than a century ago. The modern DSM defines these disorders as mutually exclusive, belonging to different classes of mental illness. However, genomic studies have shown that, at a genetic level, these conditions are highly overlapping, with a genetic correlation of nearly 0.70.
Taken together, genetic findings suggest that the structure of much psychopathology is defined by dimensional variation in the population, as is the case for hypertension and hypercholesterolemia, where the same genes often contain variants that cause diagnosable disorders and other variants that impact on variation in the “normal” range. By contrast, schizophrenia may be more complex as the disorder may represent a concatenation of several dimensions of risk including liability to psychosis, cognitive difficulties and social dysfunction. Although diagnostic categories will continue to be needed from a practical standpoint, future iterations of psychiatric nosology may be usefully informed and refined by incorporating our emerging understanding of the etiologic overlap among clinical syndromes. This is, indeed, a premise of NIMH’s nascent Research Domain Criteria (RDoC) initiative. A former NIMH Director created controversy when he described the DSM as lacking validity and that NIMH would be “re-orienting its research away from DSM categories”. Others (e.g.) have called for a shift from what Kendler called the ‘soft’ symptom-based, etiologically blind diagnoses of the DSM to ‘hard’ diagnoses based on etiologically based biological features. Will genetic data lead to ‘hard’ empirically derived diagnoses? In isolation, probably not, but they, and mechanisms learned from them, may aid in the revision or formulation of novel diagnostic criteria in the future. They may also suggest novel hypotheses about the biological basis of psychiatric disorders. For example, a recent large genomic analysis found substantial genetic correlations between anorexia nervosa and a range of metabolic traits (including measures of cholesterol and lipids, fasting insulin, fasting glucose, insulin resistance, and leptin levels), suggesting that the disorder might be both a psychiatric and metabolic syndrome.
The idea that genetic data should inform diagnostic nosologies is not new. Robins and Guze\textsuperscript{63} included family history as one criterion of a validation method that stimulated the design of structured diagnostic criteria. Kendler\textsuperscript{64} cautioned that purely data-driven scientific nosologies could not address fundamental issues facing nosologists; e.g., how to integrate information from different types of “validators”. Tsuang et al.\textsuperscript{65} suggested that psychiatric genetics could play a limited role by creating a nosology for genetic research aimed at better defining distinct genetic entities. Although such approaches are useful for researchers, the idea that psychiatric genetic findings will revolutionize the clinical psychiatric nosology has been questioned\textsuperscript{66}. This view suggests that, rather than leading to breakthroughs in genetic nosology, genetic data will, as it has in the past, incrementally help the DSM evolve via a process that Kendler\textsuperscript{67} described as ‘epistemic iteration,’ whereby the evidence base sequentially iterates with the acquisition of new data to provide a better approximation of the latent, but unknown, structure of psychopathology. Alternatively, a more data-driven DSM process might consider a revolutionary recasting of diagnostic categories as dimensional entities with well-defined thresholds demarcating wellness from subthreshold and clinically significant disorders. This would provide clinicians the categories they need within a framework that better corresponds to the latent structure of psychopathology. However, the study of risk genes, which impact only on the liability to illness, will not, of itself, permit us to empirically define the boundaries between illness and health or between closely related disorders. Doing so would require including other empirical evidence regarding both diagnostic validity and clinical utility.

*Genetic complexity is a challenge for identifying clinically relevant biomarkers.*
In 1980, when DSM-III was released, 21 papers were published on biomarkers in psychiatry. Thirty-five years later, that number had grown to 1,555. This search for objective measures for defining disorders and their underlying pathophysiological processes reflects the field’s growing discomfort with the subjectively assessed signs and symptoms that define DSM disorders. Although a Google patent search yields over 8,000 relevant patents, with the exception of mutations causing rare, syndromal forms of psychiatric disorders and some useful genetic predictors of pharmacokinetics, this intensive search for biomarkers has not been impacted clinical practice.

Genomic studies indicate that there are many causal genes and pathways, which implies that there will be many peripheral markers of disease, as has been shown for ADHD. This suggests that multifactorial biomarkers will be needed. Initial attempts to do this with genomic studies (e.g. ) have been intriguing, but their interpretation is clouded by methodological concerns. Thus, the current use of biomarkers to aid in diagnosis or genetic counseling in psychiatry has been limited to rare syndromic forms of disorders. In the future, PRS may become useful biomarkers if their precision and predictive value can be improved. By aggregating genetic effects across many genes (and, indirectly, biological pathways) improved PRS may provide an informative summary marker despite the underlying genetic complexity.

**Where Do We Go From Here?**

The first and most obvious agenda for future research is to pursue ever-larger genomewide common and rare variant studies of psychiatric disorders. Empirical evidence and simulations show that for GWAS there is a sample-size-dependent inflection point
beyond which the number of genomewide significant loci increases linearly.\textsuperscript{71} For SCZ, the inflection point was seen at about 15,000 cases, but for other, perhaps more complex or heterogenous disorders such as MDD and SUDs, the inflection point may not be reached until as many as 75,000 to 100,000 cases are examined.\textsuperscript{71} The numbers needed depend on a phenotype’s heritability and polygenicity and the effect sizes of the contributing SNPs.\textsuperscript{72} In the case of rare variants, effect sizes may be larger, but their rarity again requires large sample sizes (more than 25,000 cases) to allow detection of a sufficient number of risk variant carriers.\textsuperscript{73,74} The third wave of the PGC (PGC3), now underway, aims to enable this next generation of larger-scale GWAS and pathway analyses.\textsuperscript{75} Currently, the PGC encompasses ten disorders and aims for 100,000 cases each. Though some have questioned the value of continuing to pursue larger GWAS, the evidence to date for psychiatric disorders and other complex diseases suggests that such efforts will continue to be important for identifying additional genes and through them the biological pathways that underlie disease. This will provide a foundation for novel diagnostic and therapeutic approaches, and for realizing the hope for personalized treatment.\textsuperscript{17,76}

Expanding the size and scope of genomewide common and rare variant studies will also help elucidate the genetic architecture of disorder-specific and cross-disorder genetic effects. Ongoing analyses by the PGC Cross-Disorder Workgroup are focusing on characterizing the association and functional significance of specific variants and pathways related to the nine disorders enumerated above, as well as looking for pleiotropic effects of variants detected. In addition, the PGC3’s Brainstorm initiative is linking the PGC with other GWAS consortia to examine the genetic relationships among psychiatric disorders, neurologic disorders, and dimensional measures of personality, cognition, and brain
structure and function. Initial Brainstorm analyses across 23 brain disorders (with a total N = 842,820) indicate that genetic overlap is stronger among traditionally defined psychiatric disorders than among neurologic disorders (e.g. multiple sclerosis, stroke, migraine) or between psychiatric and neurologic disorders, supporting the clinical demarcation between neurologic and psychiatric disorders. The average genetic correlation based on genomewide SNP data was +0.21 among eight psychiatric disorders (autism spectrum disorder, schizophrenia, bipolar disorder, major depressive disorder, ADHD, anorexia nervosa, obsessive-compulsive disorder, and Tourette syndrome), substantially higher than that observed among the ten neurologic disorders (0.06). Schizophrenia showed broad sharing with the other psychiatric disorders while autism and Tourette syndrome appeared to be the most genetically distinct. Strikingly, none of the neurologic disorders were significantly genetically correlated with the psychiatric disorders with the exception of migraine. However, the low genetic correlation between the psychiatric and neurologic disorders could in part result from diagnostic differences as the psychiatric disorders were assessed using symptoms and signs and the neurological disorders often utilized imaging and/or neurophysiologic measures.

We can also now capitalize on growing resources in electronic health records (EHR) and population-based registries linked to genomic data to examine the pleiotropic effects of psychiatric genetic risk variants across the phenome. For example, the eMERGE network, a consortium of biobanks linked to EHR data, and large-scale cohort studies like the UK Biobank, which recently released genomewide data for 500,000 participants, and the recently launched All of Us Program of the NIH Precision Medicine Initiative, offer opportunities to conduct well-powered phenomewide association studies (“PheWas”) of
common and rare variants. These analyses may reveal unexpected etiologic relationships between psychiatric and other diseases or traits. The *All of Us* Research Program seeks to collect broad and deep phenotypic data (using EHRs, mobile technologies, surveys, and biospecimens) along with genomic data in a longitudinal cohort of 1 million or more Americans, and will offer an even larger resource for testing hypotheses about the spectrum of genotype-phenotype relationships for medical and psychiatric illness. Longitudinal population-based cohorts with genomic data can identify trajectories of disorder risk and gene-environment interplay. Therefore, the ability to recall participants in biobanks or large-scale cohorts for further deep phenotyping based on genotype is critical.

To date, most genomic studies have been restricted to cross-sectional, case-control analyses. The focus has been on phenotypic characterization, with little emphasis on the collection of high quality information on critical environmental risk factors. A more complete understanding of the etiology and structure of psychopathology will require an understanding of how genetic and environmental risk factors act and interact across development. To answer these questions will require study designs that attend to environmental risks and/or take an explicitly developmental perspective. For example, when in human development do risk alleles and pathways exert their effects? Might some pathways confer risk to a broad vulnerability to psychopathology and others to more differentiated disorders? Are there sensitive periods of development when environmental risk factors act or interact more potently to confer risk? What are the molecular or cellular mechanisms by which genetic variation and environmental exposures confer risk over the lifespan? These and other questions may be addressed with data from epidemiological
birth cohorts (e.g., the Avon Longitudinal Study of Parents and Children, ALSPAC\textsuperscript{85}), population-based registries (e.g., those available in Scandinavian countries \textsuperscript{86}) as well as large cohorts consented for follow-up (e.g., the UK Biobank\textsuperscript{79}, \textit{All of Us}\textsuperscript{80}).

Polygenic methods (e.g. PRS) may capture more complex combinations of disorder-specific and cross-disorder variants as indices of genetic vulnerability, and be very useful in longitudinal and developmental studies. The PGC3 analytic plan includes PRS analyses of nine disorders in a longitudinal sample of nearly 14,000 twins followed from age 9 to age 24. PGC analyses will also examine how environmental exposures modify genetic risk (PRS x environment interaction studies) to influence risk of a range of disorders. Although allelic-additive models are the most impervious to model misspecification, it should ultimately be possible to account for additional missing heritability in psychiatry by appropriately modeling dominant and recessive alleles, as well as gene-gene and gene-environment interactions.

Incorporating data beyond the categorical diagnostic variables that have been the predominant focus of psychiatric genetic studies to date will allow better characterization of the structure of psychopathology. To this end, investigators are undertaking studies of dimensional traits (e.g., RDoC measures) and incorporating neuroimaging phenotypes that may identify genetic underpinnings of neural, cognitive, affective and social phenotypes that transcend diagnostic boundaries. For example, recent analyses have demonstrated genetic overlaps between psychiatric disorders and measures of brain structure, including shared genetic influences for schizophrenia and thickness of the left superior frontal gyrus, a region where thinning and volume loss have previously been associated with the disorder \textsuperscript{87}. The international ENIGMA consortium has brought together genomic and brain imaging
data for more than 30,000 individuals spanning a broad range of psychiatric disorders, enabling a growing catalogue of discoveries about the genetic basis of brain structure and function and their relationship to psychopathology. Other studies have examined the genetic relationship between disorder and normal variation in quantitative traits. For example, common and rare genetic variants associated with autism spectrum disorder influence dimensions of social cognition, cognition/intelligence, and communication abilities in the general population. Similarly, genetic risk scores for ADHD predict attention problems in population-based samples of children.

Efforts to dissect the fundamental intermediate phenotypes underlying risk of psychiatric disorder face important challenges. Most importantly, we still do not know which are the most relevant levels of analyses and which of the large number of possible intermediate traits are causally related to mental illness. The domains enumerated in the RDoC framework represent one of several possible approaches. Genetic data may help resolve the causal status of putative intermediate phenotypes. For example, a candidate intermediate phenotype for which robust genetic associations are known can be analyzed using Mendelian randomization (which uses the associated variants as instrumental variables) to determine whether it is causally linked to a disorder. This approach has been used successfully in other areas, providing evidence, for example, that central adiposity is causally related to coronary heart disease while HDL is not. In the realm of psychiatry, recent analyses have suggested that cannabis use maybe a causal risk factor for schizophrenia. Because this approach is biased when there is pleiotropy (which is widespread in psychiatric genetics), other methods that can evaluate causality in presence of genetic correlation may be needed.
Finally, having a more complete understanding of the genetic basis of psychiatric disorders will allow us to determine whether genetic variation, in concert with other more traditional variables, can improve prediction of clinically relevant outcomes. To date, the variance explained by common and rare variants, individually or as aggregate genetic risk scores, has been insufficient to be clinically useful for diagnosis or for prediction of clinical course. An exception is autism spectrum disorder, where genetic evaluation, including testing for structural and rare mutations, is recommended as part of the diagnostic process by the American College of Medical Genetics and Genomics. In general, genetic biomarkers will only be clinically useful if they add value or efficiency beyond established non-genetic diagnostic procedures or risk factor profiles. As the power and precision of polygenic risk profiles improve and as more powerful rare variant studies allow us to fractionate disorder heterogeneity, this may be achievable for some disorders. For example, recent analyses support the existence of autism spectrum disorder subtypes that differ by their genetic architecture. Specific, highly penetrant structural and rare exonic mutations may represent genetic subtypes of heterogeneous disorders. In addition, a more complete characterization of the pleiotropic effects of psychiatric risk variants may have implications for genetic counseling. Numerous rare structural and single nucleotide variants have already been shown to influence a range of psychiatric disorders. This catalogue of pleiotropic variants is expected to increase as the PGC3 conducts whole-genome sequencing analyses of pedigrees densely affected by multiple psychiatric disorders to identify rare variants of strong effect. Given the substantial co-morbidity and overlap in genetic contributions to psychiatric disorders, studies should invest in phenotyping across disorders. Some, such as substance use disorders and depression,
likely contribute to many common medical diseases, so studies of those diseases (e.g., liver
disease, heart disease, cancers) should also gather information on lifetime patterns of
substance use, abuse and dependence and symptoms of depression. We will also need to
convince funders that GWAS will lead to a better mechanistic understanding and,
ultimately, better prevention and treatment.

Conclusions

A substantial and growing body of genetic research has begun to elucidate the
underlying structure of psychopathology. Before the modern era of genomic research,
family and twin studies demonstrated that all major psychiatric disorders aggregate in
families and are heritable. Over the past decade, the success of large-scale genomic studies
has confirmed several key principles: 1) psychiatric disorders are highly polygenic,
reflecting the contribution of hundreds to thousands of common variants of small effect
and rare (often de novo) SNVs and CNVs; 2) the delineation of disease mechanisms from
variant discovery is a difficult problem that requires integrating genetic information across
multiple levels of functional data from “omics” to neurobiology to clinical phenotypes; and
3) genetic influences on psychopathology commonly transcend the diagnostic boundaries
of our clinical DSM nosology. At the level of genetic etiology, there are no sharp boundaries
between diagnostic categories or between disorder and normal variation. In the coming
years, ever-larger studies incorporating DNA sequencing, environmental exposures, and
phenome-wide analyses will facilitate a more granular understanding of the genetic
etiology and phenotypic spectrum of mental illness.
The highly interconnected nature of cellular networks in the brain suggests that many, perhaps most, variants that regulate or participate in these pathways have some non-zero effect across psychiatric disorders (a concept called “network pleiotropy”\textsuperscript{76}) Thus, as naïve single gene and oligogenic models for psychiatric disorders have been replaced by the reality of hundreds if not thousands of risk variants, the field is facing a problem of daunting complexity. However, it is better to confront directly this challenging reality than comfort ourselves with unrealistically simplistic models. Improvements in our ability to link genetic variation to gene function, sort out signal from noise with evolving statistical methods, and build in vitro and in vivo biological models to explore the “curse of polygenicity” should help with the translation from molecular genetic discoveries into etiologic insights. Progress might be slow, but molecular genetic methods remain one of our few tools that can provide us direct insight into etiological pathways that underlie the structure of psychopathology.
Acknowledgments

Supported in part by NIH awards K24MH094614 (JWS), the Research Council of Norway (223273) and KG Jebsen Stiftelsen (OAA). Dr. Smoller is a Tepper Family MGH Research Scholar. “All of Us” is a service mark of the U.S. Department of Health and Human Services. The authors thank Nicholas Merriam for assistance with producing the figure.

Disclosures:

Dr. Smoller is an unpaid member of the Scientific Advisory Board of PsyBrain Inc. and the Bipolar/Depression Research Community Advisory Panel of 23andMe. Dr. Andreassen has received speaker’s honorarium from Lundbeck. The remaining authors have no disclosures.
Figure Legend

**Figure 1. Methods commonly used to evaluate genetic overlap between phenotypes.**

A. At the DNA variant level, individual loci (e.g. SNPs, rare mutations, or CNVs) may show evidence of pleiotropic association with two (e.g. P₁, P₂) or more phenotypes. At the level of biological pathways, gene sets assigned to a pathway may be enriched in association signals beyond chance expectation across multiple phenotype (e.g. P₁, P₂, P₃). B. Genetic risk scores (or polygenic risk scores, PRS) are developed in a “discovery” GWAS sample and computed for each individual in an independent “target” sample. The genetic risk score for each individual (i) in the target sample is computed as the product of the number of risk alleles (X) at each SNP (j) multiplied by that SNP’s association effect size (βⱼ) and summing over all SNPs. The left hand plot shows the distribution of genetic risk scores in cases and controls in an independent target set for the same phenotype as that of the discovery sample. To examine cross-phenotype overlap, the discovery risk score is applied to target samples of other phenotypes. The proportion of variance explained by the discovery GWAS (R²) for each target phenotype (e.g. P₂, P₃, P₄) is shown in the plot on the right hand side. C. Genetic correlation between phenotypes (ranging from -1.0 to +1.0) can be estimated using multiple methods that compare genetic and phenotypic similarity among unrelated individuals. The figure shows a hypothetical genetic correlation matrix between multiple pairs of phenotypes (P₁ – P₈).
References


52. Plomin R, Haworth CM, Davis OS. Common disorders are quantitative traits. *Nat Rev Genet* 2009; 10(12): 872-878.


106. Mathews CA, Grados MA. Familiality of Tourette syndrome, obsessive-compulsive disorder, and attention-deficit/hyperactivity disorder: heritability analysis in a large


Table 1. Heritability and Genetic Correlation Estimates for Selected Psychiatric Disorders

<table>
<thead>
<tr>
<th>Disorder</th>
<th>ASD</th>
<th>ADHD</th>
<th>SCZ</th>
<th>BD</th>
<th>MDD</th>
<th>AN</th>
<th>OCD</th>
<th>PTSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASD</td>
<td>0.17 (0.12-0.22)†</td>
<td>-0.13 (0.38-0.12)§</td>
<td>0.18 (0.08-0.28)‡</td>
<td>0.06 (0.09-0.22)§</td>
<td>0.14 (-0.06-0.34)§</td>
<td>0.009 (-0.11-0.13)§</td>
<td>0.07 (-0.13-0.27)§</td>
<td>N/A</td>
</tr>
<tr>
<td>ADHD</td>
<td>0.74 (0.70-0.87)§</td>
<td>0.28 (0.23-0.33)‡</td>
<td>0.18 (0.03-0.39)§</td>
<td>0.53 (0.28-0.78)§</td>
<td>0.43 (0.06-0.80)§</td>
<td>0.17 (-0.05-0.39)§</td>
<td>-0.07 (-0.24-0.10)§</td>
<td>N/A</td>
</tr>
<tr>
<td>SCZ</td>
<td>N/A</td>
<td>N/A</td>
<td>0.23 (0.21-0.25)‡</td>
<td>0.83 (0.76-0.90)§</td>
<td>0.46 (0.34-0.58)§</td>
<td>0.19 (0.10-0.28)§</td>
<td>0.34 (0.22-0.46)§</td>
<td>0.33 (0.17-0.49)§</td>
</tr>
<tr>
<td>BD</td>
<td>0.24 (0.24-0.29)§</td>
<td>0.33 (0.32-0.39)§</td>
<td>0.68 (0.67-0.73)§</td>
<td>0.25 (0.23-0.27)‡</td>
<td>0.64 (0.48-0.80)§</td>
<td>0.11 (-0.05-0.27)§</td>
<td>0.19 (0.18-0.44)§</td>
<td>0.16 (-0.06-0.38)§</td>
</tr>
<tr>
<td>MDD</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>0.65 (0.58-0.75)§</td>
<td>0.21 (0.17-0.25)‡</td>
<td>0.14 (-0.11-0.39)§</td>
<td>0.25 (0.07-0.43)§</td>
<td>0.34 (0.10-0.58)§</td>
</tr>
<tr>
<td>AN</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>0.56 (0.50-0.62)‡</td>
<td>0.53 (0.19-0.86)§</td>
<td>N/A</td>
</tr>
<tr>
<td>OCD</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>0.52 (0.26-0.81)§</td>
<td>0.37 (0.23-0.51)‡</td>
<td>N/A</td>
</tr>
<tr>
<td>PTSD</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>1.0§</td>
<td>0.45 (0.30-0.60)§</td>
<td>0.18 (0.06-0.30)§</td>
</tr>
</tbody>
</table>

On the diagonal: Heritability estimates (and 95% CI) from pedigree/twin studies (lower, shaded orange) and genomewide SNP analyses (upper, shaded light blue). The SNP-based estimates reflect only the common variant component of heritability. Also shown are pairwise genetic correlations from pedigree/twin studies (below diagonal) and genomewide SNP analyses (above diagonal). Bolded entries: 95% CI does not include 0.0. ASD: autism spectrum disorder; ADHD: attention deficit-hyperactivity disorder; SCZ: schizophrenia; BD: bipolar disorder; MDD: major depressive disorder; AN: anorexia nervosa; OCD: obsessive-compulsive disorder; PTSD: posttraumatic stress disorder; N/A: not available. Sources: †22, ‡43, §77, ¶21, ¶99, ¶100, ¶21, ¶101, ¶102, ¶103, ¶104, ¶105, ¶106, ¶107, ¶108, ¶109.
A. Variant Level

B. Genetic Risk Score = \sum x_{ij} \beta_j

C. Cross-Phenotype Association

Blue – Positive Genetic Correlation; Red – Negative Genetic Correlation