Editorial

Dear Colleagues,

Medical genetics is the study, at the gene level, of the relationship between structures and observed disorders (genetic diagnosis). It seeks, in particular, abnormalities in the genes and chromosomes.

Genomics is the study of the functional expression of the genome, and is not restricted to one particular gene.

The root "gen-" implies specific characteristics related to a person, and thus the identification of specific factors (positive or favorable factors, negative factors leading to fragility or pathology).

Each individual carries specific traits which identify him or her, and which on one hand create a tendency towards the expression of certain pathologies, and on the other hand cause individual differences in therapeutic response, making the individual sensitive to a therapeutic approach, causing a reaction to a particular treatment, making a particular therapeutic approach more desirable, causing drug dependence, etc.

Psychiatric disorders are the most striking example of the abovementioned points.

However, we must also accept the limits of this concept, knowing that in the field of psychiatry, genetic investigations and the resulting knowledge have for the most part provided extremely complex profiles, which are not always able to be implemented on a diagnostic or therapeutic level.

The correlation of genetic data and a genomic approach opens up the way to new paths of investigation, and new possibilities of applying the resulting knowledge, providing a basis for diagnostic conclusions or new therapeutic strategies using this approach.

This issue of *Dialogues in Clinical Neuroscience* gives us an opportunity to renew our acquaintance with this field, and to put the new diagnostic and treatment procedures in context. It provides us with an extremely interesting update on these various methods. We are most grateful to Margret Hoehe, assisted by Deborah Morris-Rosendahl and Nancy Andreasen, for coordinating this issue, and also to the authors who agreed to contribute.

Sincerely yours,

Jean-Paul Macher, MD

Contents

Issue coordinated by: *Margret R. Hoehe with Deborah J. Morris-Rosendahl and Nancy C. Andreasen*

Contributors

Danielle M. Dick, PhD

Author affiliations: Virginia Institute of Psychiatric and Behavioral Genetics; Department of Psychiatry; Department of Human and Molecular Genetics, Virginia Commonwealth University School of Medicine, Richmond, VA; Department of Psychology, Virginia Commonwealth University, Richmond, VA, USA

Arturas Petronis, MD, PhD

Author affiliations: The Krembil Family Epigenetics Laboratory, Centre for Addiction and Mental Health, and University of Toronto, Toronto, Ontario, Canada

Anna C. Need, PhD

Author affiliations: Institute for Genome Sciences and Policy, Center for Human Genome Variation, Duke University, Durham, North Carolina, USA

George M. Church, PhD

Author affiliations: Department of Genetics, Harvard Medical School, Boston, Massachusetts; PersonalGenomes.org, Boston, Massachusetts;

Kári Stefánsson, MD

Author affiliations: deCODE Genetics, Reykjavik; Faculty of Medicine, University of Iceland, Reykjavik, Iceland

David A. Mrazek, MD, FRCPsych

Author affiliations: Chair, Department of Psychiatry and Psychology, Mayo Clinic; Professor of Psychiatry and Pediatrics, Mayo Clinic College of Medicine, Rochester, Minnesota, USA

Joel Gelernter, MD

Author affiliations: Departments of Psychiatry, Genetics, and Neurobiology, Yale University School of Medicine, New Haven, CT, USA; and VA CT Healthcare System

Markus M. Nöthen, MD

Author affiliations: Department of Genomics, Life & Brain Centre and Institute of Human Genetics, University of Bonn, Bonn, Germany

Hans-Hilger Ropers, MD, PhD

Author affiliations: Max Planck Institute for Molecular Genetics, Berlin, Germany

Ted Reichborn-Kjennerud, MD, PhD

Author affiliations: Division of Mental Health, Norwegian Institute of Public Health Institute of Psychiatry, University of Oslo, Norway; Department of Epidemiology, Columbia University, New York, NY, USA

Deborah J. Morris-Rosendahl, PhD

Author affiliations: Institute for Human Genetics, University Clinic Freiburg, Germany

In this issue...

The DNA age is dawning… These past years have seen an unprecedented transformation: from single candidate gene studies to whole genome-based screening approaches to the analysis of whole human genome sequences and their variation in health and disease. Whereas the first human genome took years and US$ 3 billion to materialize, with an effort involving armies of researchers, the US$ 1000 genome is now within reach, and genome sequences within 15 minutes are now envisageable. This will revolutionize our understanding of health and disease. Importantly, personal genomes will become accessible not only to researchers, but in principle to every human individual. Knowledge and insights gained from human DNA, the very code of our existence, in its countless individually different forms, is bound to impact our personal and public lives in significant and yet unpredictable ways: the ways we perceive ourselves and each other, our attitudes and approaches towards life, society, and mankind. Knowledge on personal genomes will prepare the ground to learn about the interplay between nature and nurture, the genome and the environment, in shaping the traits we observe.

The current issue covers much of this ground. We have assembled a group of outstanding international scholars, leaders at the cutting edge in genetics and genomics, who have had, and will have, a decisive impact on these revolutionizing developments.

To prepare the ground, this issue begins with a ***State of the art*** article on "Nature and nurture in neuropsychiatric genetics" (p 7) by Prof Kendler and colleagues. They provide a most comprehensive and advanced review of the fields of genetic epidemiology and molecular genetics, examining the contribution of both genetic and nongenetic risk factors, as well as interactions and correlations between them, to the etiology of psychiatric and behavioral phenotypes. They conclude that, to date, only a few specific genetic variants influencing risk have been unambiguously identified, the results collectively being complex and inconsistent with a single common DNA variant in any gene. Their hopes rely upon contemporary approaches that promise to further elucidate liability genes and variants, as well as their potential inter-relationships with each other and the environment.

Three ***Basic research*** articles follow. Prof Petronis and colleagues (p 25) introduce epigenetic approaches to psychiatric disorders. These, as a relatively new field, address changes in phenotype or gene expression caused by mechanisms other than changes in the actual base composition of DNA sequence. Thus, additional etiological models have become available, compatible with mechanisms by which environmental or other genetic influences translate into lasting imprints on the translation of sequence information. The authors outline how many epidemiological, clinical, and molecular characteristics of psychiatric diseases may be consistent with an epigenetic dysregulation and accordingly review current findings in major psychosis, Alzheimer's disease, and autism spectrum disorders. The next article expands on search strategies to identify genetic risk variants. The initial approach to study single selected candidate genes has been complemented by a wave of genome-wide association studies (GWAS). The establishment of a data base of common human variation, the HapMap, in conjunction with novel high-throughput SNP typing technologies, enables effective screening of the entire genome in order to identify common variants that associate with complex disease. Profs Goldstein and Need (p 37) provide a most comprehensive and critical assessment of the hundreds of GWAS that have been performed in recent years. Whereas some strong effects of common variants could be found in late-onset diseases and in drug response, and the major histocompatibility complex emerged with very strong associations, on the whole, common variation has explained little of the high heritability of neuropsychiatric traits. In contrast, early studies of rare structural variation, copy number variants, have rapidly resulted in a number of genes and loci that strongly associate with neuropsychiatric disorders. The authors conclude that the use of whole-genome sequencing to extend the study of rare variation in neuropsychiatry will greatly advance our understanding of neuropsychiatric genetics. Whole-genome sequencing has now become an increasingly affordable; technology capable of meeting the target cost of US$1000 or less for a diploid human genome sequence is within reach. Motivated by the low return from common single-nucleotide polymorphism analysis, this has caused a significant rise in interest in correlating genome sequences with comprehensive environmental and trait data (GET). Integrating in principle all the various "omes" will catalyze progress in functional genomics and enable systems biology-based insights into the mechanisms of human health and disease. Prof Church and colleagues (p 47) introduce, in a first comprehensive paper, the Personal Genome Project (PGP) conducted at the Department of Genetics at Harvard Medical School. This is currently the most ambitious and visionary human genomic research study, ultimately aiming to recruit as many as 100 000 individuals. The authors examine the PGP's effort to develop a GET database as a public genomics resource broadly accessible to both researchers and research participants, while pursuing the highest standards in research ethics.

In this issue...

Higher accessibility of genome and genotypic data, coupled with increasing public awareness of their potential impact on health, disease prevention and treatment, the interest of individuals in learning about their genomes, themselves, has turned out to be remarkable. In their ***Translational research*** article, Prof Stefánsson and his team (p 61), who have significantly contributed to current association findings in major common, complex diseases, highlight their translational role by offering corresponding genetic tests direct to consumers. The overall predictive power of common sequence variants is expected to be considerable, due to the high incidence of the diseases found associated with them. The authors address the past, present, and future of direct-to-consumer (DTC) genetic tests, as well as current concerns. They conclude that a key translational role of DTC genetic tests may be seen in democratizing privileged knowledge of value for preventive medicine to the public, thereby empowering them.

The first paper on ***Pharmacological aspects*** addresses genetic testing to predict drug response. The question is, to what extent have the numerous, solid findings in the areas of pharmacogenomics in fact been translated into optimization of individual treatment response? Prof Mrazek (p 69) provides an informative overview of the clinical adoption of psychiatric pharmacogenomic testing. He elaborates on the implementation and implications of genotyping highly variable drug-metabolizing enzyme genes from the Cytochrome P450 family, and then goes on to the testing of genes influencing the pharmacodynamic response to medications, including serotonin transporter and receptor genes. He provides moreover a quantification of the clinical utility of pharmacogenomic testing emphasizing its "increasingly clear" cost-effectiveness, as well as ethical considerations established for testing. The following article by Profs Gelernter and Kranzler (p 77) addresses the genetic predisposition to drugs of abuse. With a particular focus on cocaine, opioid, and nicotine dependence, the authors provide a detailed review on the genetics of drug dependence, outlining significant results from linkage, association, and genome-wide association study methodologies. Obviously, future prospects for risk allele identification will also include more extensive sequencing to identify a fuller range of risk variants.

The ***Clinical research*** section, in its first article, extends the appraisal of the genetics of psychiatric phenotypes to a state-of-the-art account of the genetics of the major psychoses, schizophrenia, and bipolar disorder. Prof Nöthen and colleagues (p 85) focus on a detailed summary of systematic genome-wide association and follow-up studies, also outlining the small effects and lack of diagnostic specificity of the common risk variants identified. In addition, they report on the recent—more successful—studies that have detected large, rare structural variants (copy number variants) conferring a greater risk for schizophrenia. Taken together, it is increasingly evident that, for most common diseases, the "Common Disease-Common Variant (CDCV)" hypothesis is insufficient, and the support for GWAS is dwindling. At the same time, remarkable successes have recently been achieved in the field of Mendelian disorders, renewing interest in these diseases. Specifically, the affordability and application of new technologies such as exome or whole-genome sequencing to limited numbers of extreme, clearly defined phenotypes has been shown to lead to new potential causative genes for neuropsychiatric disorders being identified. Prof Ropers (p 95) provides a most modern and instructive review of these developments, while bringing back into focus also the medical importance of these diseases. These must be far more common than generally thought and, moreover, may provide valuable clues to the understanding of common diseases. Efficient strategies for the identification of causative single-gene defects are outlined; in combination with novel genome partitioning and sequencing technologies, these could have far-reaching implications for health care. The third ***Clinical research*** article by Prof Reichborn-Kjennerud (p 103) addresses a somewhat young, but nevertheless fascinating field, the analysis of genetic factors involved in personality disorders (PDs), as supported by genetic epidemiologic studies. This field, though, seems even more difficult and complex than what has been attempted so far, given that it remains a major task to define and classify the phenotypes precisely, a prerequisite for meaningful genetic analyses. Multivariate studies suggest that three genetic and environmental risk factors are common to all PDs. Previous molecular genetic studies, mostly candidate gene association studies, have pointed to genes involved particularly in serotonergic and dopaminergic neurotransmitter pathways.

To conclude with the ***Brief report***, Dr Morris-Rosendahl (p 116) provides a most valuable and detailed glossary of genetic terms relevant to this issue, including useful genetic databases and additional sources of information. This material should help to understand the current contents and prepare the readers for this most exciting field of scientific and human enterprise.

Margret R. Hoehe, MD, PhD

Nature and nurture in neuropsychiatric genetics: where do we stand?

Danielle M. Dick, PhD; Brien Riley, PhD; Kenneth S. Kendler, MD

Both genetic and nongenetic risk factors, as well as interactions and correlations between them, are thought to contribute to the etiology of psychiatric and behavioral phenotypes. Genetic epidemiology consistently supports the involvement of genes in liability. Molecular genetic studies have been less successful in identifying liability genes, but recent progress suggests that a number of specific genes contributing to risk have been identified. Collectively, the results are complex and inconsistent, with a single common DNA variant in any gene influencing risk across human populations. Few specific genetic variants influencing risk have been unambiguously identified. Contemporary approaches, however, hold great promise to further elucidate liability genes and variants, as well as their potential inter-relationships with each other and with the environment. We will review the fields of genetic epidemiology and molecular genetics, providing examples from the literature to illustrate the key concepts emerging from this work.

 Dialogues Clin Neurosci. 2010;12:7-23.

Keywords: *genetics; gene-environment interaction; alcohol dependence; twin study; schizophrenia*

Our knowledge of psychiatric and substance-use genetics comes from two key fields of research, both dynamic areas in rapid change. First, genetic epidemiology asks whether there is risk in excess of the population baseline in the relatives of cases, and, if so, whether the excess risk is attributable to the genetic factors or the environments they share. Beyond simply estimating heritability, genetic epidemiology has evolved to address more sophisticated questions, such as whether liability genes have the same effects across the lifespan, how they may influence multiple disorders, and how they might interact with environmental risks.

Genetic epidemiology of psychiatric and behavioral phenotypes has consistently demonstrated that: i) genetic risk factors are, in aggregate, important etiological components; ii) they cannot completely account for observed risk, meaning these phenotypes are multifactorial traits, with important nongenetic (or environmental) contributing factors; and iii) the risk alleles appear to be of small effect size and to occur in a large number of genes. Psychiatric and behavioral phenotypes are influenced by a large number of risk factors that individually are within the range of normal human variation and produce modest individual increases in risk.

The initial goal of the second major research area, molecular genetics, is to identify genes which influence these

Author affiliations: Virginia Institute of Psychiatric and Behavioral Genetics; Department of Psychiatry; Department of Human and Molecular Genetics, Virginia Commonwealth University School of Medicine, Richmond, VA, USA (Danielle M. Dick*, Brien Riley*, Kenneth S. Kendler); Department of Psychology, Virginia Commonwealth University, Richmond, Virginia, USA (Danielle M. Dick*)
*These authors contributed equally to the review

Address for correspondence: Kenneth S. Kendler MD, Dept Psychiatry MCV, PO Box 980126, Richmond VA 23298, USA
(e-mail: kendler@vcu.edu)

phenotypes and to identify the specific risk variants within them. There are substantial differences in DNA sequences between individuals, and gene identification methods test whether specific alleles at these variable positions are more common in affected than in unaffected individuals, most commonly with linkage studies (in families) and association studies (primarily in case/controls, but also in numerous other designs). We will discuss the underlying causes of these two genetic phenomena, the methods for detecting them, and the limitations of each.

The second goal of molecular genetics is to identify specific risk alleles and to use functional studies to elucidate how a gene functions normally, how the risk allele alters normal function, and how these alterations contribute to disease. The aim of this work is to explain the aggregate genetic risks observed through the effects of risk alleles on gene expression, protein structure and function, and/or biological processes. This area remains largely unsuccessful to date for complex traits generally.

In this review we focus on the basic methods of genetic epidemiology and molecular genetics, and provide examples, across a variety of psychiatric and substance use disorders, of questions currently being addressed. In contrast to this first section on genetic epidemiology, the sections on molecular genetics focus narrowly on schizophrenia, where there is a much longer history of molecular genetic studies, because we judged that emphasizing a single disorder would provide a more coherent example of ongoing research progress and challenges.

Basic genetic epidemiology

The most fundamental question addressed by psychiatric genetic epidemiology is whether a particular trait or disorder shows evidence for genetic influence. Both twin and adoption studies provide methods to address this question and tease apart the degree to which genetic and environmental influences are important on a given outcome. Twin studies accomplish this by comparisons of the similarity of monozygotic twins (MZs; who share 100% of their genetic variation), with dizygotic twins (DZs; who share on average just 50% of their genetic variation). Adoption studies compare similarity among adopted-apart biological relatives, who share genetic variation, but not their environments, and adoptive relatives, who share their environment, but not their genetic makeup. Through these comparisons, we can quantify the degree to which genetic influences contribute to individual differences in risk, a statistic commonly referred to as the heritability of the trait. These study designs have been applied to virtually all psychiatric disorders and to a number of related traits, yielding compelling evidence that genetic influences play a critical role in virtually all psychiatric outcomes. There is considerable variability in the magnitude of genetic influence across different disorders. On the high end are disorders such as schizophrenia, bipolar disorder, and autism, which yield heritability estimates of the order of 80% or higher. Alcohol and other drug dependence shows moderate heritability, in the range of 50% to 60%. On the lower end of the spectrum, though still showing significant evidence of genetic influence, are anxiety and depressive disorders, as well as eating disorders, which yield heritability estimates of ~30% to 40%. So, while there is variability in the magnitude of importance of genetic effects, it is widely accepted that a significant genetic component plays a role in virtually all psychiatric traits. It is a sign of the paradigm shift that has taken place in psychiatry that heritability estimates are no longer considered controversial, since the original studies finding evidence for genetic effects represented strong challenges to predominant views favoring environmental theories on the causation of most psychiatric conditions, ranging from schizophrenia to autism to alcohol dependence—disorders that are all now widely recognized as having genetic components.

While demonstration of heritability played an important role in altering fundamental assumptions about the etiology of psychiatric disorders, if not understood in their proper context, heritability estimates can also have a number of unfortunate side effects. Firstly, the heritability statistic created a dichotomy of genetic versus environmental influence—nature versus nurture. How much is genetic? How much is environmental? This is, as we hope to show, a somewhat arbitrary distinction. Genetic predispositions by necessity are expressed in the context of the organism's environment, and the environment can differentially affect individuals based on their unique genetic makeup. Further, many environments are not simply "imposed" on an individual; rather, individuals play an active role in selecting and shaping their environments. Accordingly, it is generally more informative to elucidate pathways of risk and show how genetic and environmental influences come together in this process, rather than trying to divide influence into that which is genetic and that which is environmental. Secondly,

demonstration of heritability led to the idea that there were genes "for" a given disorder. More complex models that have examined genetic influences across multiple different conditions suggest that the Diagnostic and Statistical Manual of Mental Disorders (DSM) structure of psychiatric diagnoses often does not map onto the underlying genetic architecture of psychiatric traits. Genetic influences appear to be shared across many psychiatric conditions, and likely operate through mediating characteristics that alter risk for a number of different outcomes. Finally, static heritability estimates fail to capture the dynamic nature of genetic and environmental influences on psychiatric outcome. Heritability estimates are specific to the population under study. Lost in heritability estimates are potential differences across environmental conditions, across populations or gender, and across ages. Accordingly, genetic epidemiology has undergone an evolution in the kinds of questions being addressed. No longer is the question simply "Are genetic influences important on Trait X?" or even "How important are genetic influences on Trait X?". Rather, the focus has shifted to addressing the complexities raised here, using the paradigm we have called advanced genetic epidemiology.

Advanced genetic epidemiology

Moving beyond genes versus environment: gene-environment interaction and correlation

Parsing genetic and environmental influences into separate sources represents a necessary oversimplification, as for most traits we know about, genetic and environmental influences are inexorably intertwined. Most measures of the environment show some degree of genetic influence, illustrating the active role that individuals play in selecting and creating their social worlds.[1] To the extent that these choices are impacted upon by an individual's genetically influenced temperaments and behavioral characteristics, an individual's environment is not purely exogenous, but rather, in some sense, is in part an extension and reflection of the individual's genotype. This concept is called gene-environment correlation or, perhaps more descriptively, genetic control of exposure to the environment. It is likely an important process in the risk associated with several psychiatric outcomes. For example, there is considerable evidence for peer deviance being associated with adolescent substance use. However, individuals play an active role in selecting their friends, and multiple genetically informative samples have now demonstrated that a genetic predisposition toward substance use is associated with the selection of other friends who use substances.[2-4] Interestingly, there is evidence that genetic effects on peer-group deviance show a strong and steady increase across development,[5] suggesting that as individuals get older and have increasing opportunities to select and create their own social environment, genetic factors assume increasing importance. Another area where gene-environment correlation is known to play a significant role is in the risk pathways associated with depression. Stressful life events have been consistently associated with the manifestation of depression. However, there is evidence for genetic influence on the occurrence of stressful life events,[6,7] indicating that an individual's predisposition plays a role in the likelihood that they will experience difficulties that are then associated with risk for depressive episodes. For example, research has shown that a genetic liability to major depression increases the risk for a range of stressful life events, particularly those reflecting interpersonal and romantic difficulties.[8] These represent only a couple of areas where individuals are known to play an active role in shaping environmental factors that are associated with subsequent risk for psychiatric problems.

Another way that genetic and environmental influences are linked is via gene-environment interaction or, as we might prefer, genetic control of sensitivity to the environment. In these situations, genetic influences may vary in importance as a function of environmental conditions and/or that the environment differs in importance as a function of an individual's genetic predisposition (these two conceptualizations of gene-environment interaction are indistinguishable statistically). Heritability estimates essentially average across environments; accordingly, if there is reason to believe that the importance of genetic effects might vary as a function of the environment, this information can be incorporated into the twin model to test for significant differences in heritability as a function of the environment. Substance use provides one area where gene-environment interaction effects have been found to be particularly important. Environments that exert more social control and present less opportunity to engage in substance use consistently show reduced evidence for the importance of genetic effects. In this sense, the environment is essentially constraining

the expression of a predisposition toward substance use/problems. This has been demonstrated with respect to enhanced parental monitoring in adolescents,[9] a more religious upbringing,[10] and enhanced community stability,[11] among other factors. One nice example of this can be found in an analysis of the heritability of adolescent smoking across the United States using data from the National Longitudinal Study of Adolescent Health. Genetic influences on daily smoking were lower in states with relatively high taxes on cigarettes and in those with greater controls on vending machines and cigarette advertising, again suggesting the importance of social control mechanisms in moderating the importance of genetic influences on substance use.[12]

Delineating phenotypic boundaries of genetic risk

The rationale of the basic twin design can be expanded to examine the extent to which genetic and environmental factors contribute to the co-occurrence of psychiatric conditions. Comorbidity among psychiatric disorders is common, and multivariate twin studies have helped address the etiological mechanisms that contribute to these observed epidemiological patterns. A fascinating result to emerge from these studies is that psychiatric conditions with distinct clinical presentations (eg, major depression and anxiety) are not necessarily distinct genetically. For example, a study of major depression and generalized anxiety disorder found a genetic correlation of 1.0, suggesting that the same genetic influences impact depression and anxiety, but differences in environmental experiences contribute to the manifestation of different outcomes.[13] An expanded study that examined the genetic and environmental architecture across seven common psychiatric and substance-use disorders found that genetic influences load broadly onto two factors that map onto internalizing disorders (depression, anxiety disorders), and externalizing disorders (alcohol and other drug dependence, childhood conduct problems, and adult antisocial behavior).[14] These findings indicate that while distinguishing these disorders as "separate conditions" in the DSM may be useful for clinical purposes, these categories do not necessarily reflect differences in biological etiology. These findings, along with similar results from phenotypic analyses (eg, refs 15,16) have led some to suggest a reorganization of the "metastructure" of psychiatric disorders in DSM-V.

Another area of investigation examines whether there are differences in the importance of genetic and environmental factors at different stages of the disorder. For example, the development of substance dependence is necessarily preceded by several stages, including the initiation of the substance, the progression to regular use, and the subsequent development of problems, whether they be psychological, social, and/or physiological. Twin studies can investigate the degree to which each of these steps in the pathway of risk is influenced by genetic and/or environmental factors, and the extent to which the same or different genetic/environmental factors impact different stages. For example, data from two population-based, longitudinal Finnish twin studies found that shared environmental factors played a large role in initiation of alcohol use, and a more moderate role on frequency of use, and it was largely the same influences acting across these stages of use. However, there was no significant evidence of shared environmental influences on alcohol problems in early adulthood. Problems were largely influenced by genetic factors that overlapped with genetic influences on frequency of use.[17] In a study from Virginia in male twins, similar results were found for alcohol, cannabis, and nicotine.[18] In the early years of adolescence, shared environmental influences were responsible for nearly all twin resemblance for levels of intake of these psychoactive substances. However, as individuals aged, the impact of shared environment decreased and that of genetic factors increased.

Finally, there is known to be tremendous heterogeneity among individuals with psychiatric conditions. Twin studies can provide insight into whether clinical heterogeneity may reflect differences in etiological risk factors. For example, alcohol dependence with comorbid drug dependence has been found to be a particularly heritable form of the disorder,[19,20] and twin studies have suggested a genetic influence on typical versus atypical forms of major depression.[21]

Changing genetic influence across development

Another active area of research is the clarification of how genetic and environmental influences may change across development. A recent meta-analysis examined published studies with at least two heritability time points across adolescence and young adulthood for eight different behavioral domains. These analyses revealed significant cross-time heritability increases for external-

izing behaviors, anxiety symptoms, depressive symptoms, IQ, and social attitudes, and nonsignificant increases for alcohol consumption and nicotine initiation. The only domain that showed no evidence of heritability changes across time was attention-deficit/hyperactivity disorder.[22] Similarly, in a large study of >11 000 pairs of twins from four countries, the heritability of general cognitive ability was found to increase significantly and linearly from 41% in childhood (9 years) to 55% in adolescence (12 years) and to 66% in young adulthood (17 years).[23] The robust finding of increases in the importance of genetic influences across development likely reflects, in part, active gene-environment correlation, as individuals increasingly select and create their own experiences based on their genetic propensities.

In addition to changes in the relative magnitude of importance of genetic and environmental influences, another dynamic change is that different genes may be acting at different time points. This is nicely illustrated in recent analyses of alcohol use problems, as assessed at five time points from ages 19 to 28 in the Dutch Twin Registry (Kendler et al, in preparation). Kendler and colleagues found strong innovation and attenuation of genetic factors across this age range—indicating that some genetic influences on alcohol problems that were evident at age 19 declined in importance across time, while new genetic influences became important starting at ages 21 and 23. Thus, although the overall heritability of alcohol problems remained fairly stable, it appeared that different genetic factors were important at different timepoints. In analyses in the TCHAD Swedish study which followed twins from ages 9 to 20 across four waves of assessment, large changes were seen in the genetic risk factors for fears and phobias[24] and for symptoms of anxiety and depression,[25] with particularly pronounced evidence for genetic innovation at puberty. These analyses suggest that genetic influences of many psychiatric and substance use disorders are likely to be developmentally dynamic.

Sex differences

Sex differences in the prevalence of psychiatric disorders, and in risk and protective factors associated with psychiatric outcomes, are widespread in epidemiology. Twin studies allow us to investigate the extent to which there are differences in the relative importance of genetic and environmental influences on outcome, and the extent to which *different* genes and/or environments may be important. Large-scale twin studies have suggested, for example, that the genetic risk factors for both depression[26] and alcohol dependence,[27] while correlated, are not entirely the same for males and females. Results from two large twin studies in the US and Sweden agree that the genetic influences of major depression are modestly stronger in women than in men.[26,28]

Do we still need twin studies in the era of gene finding?

As advances in molecular genetics and statistical analysis have made it possible to conduct large-scale projects aimed at identifying the specific genes involved in susceptibility to psychiatric outcome (detailed in the next sections), some have raised questions about the continuing utility of genetic epidemiology. The argument is that heritability has now been established, which provides the foundation and justification for moving beyond twin studies, on to large-scale gene identification projects. However, as detailed in this paper, most twin studies are no longer conducted simply to test for the presence of genetic effects; rather, they focus on the more complex kinds of questions summarized above. These analyses are not only informative about the nature of etiological pathways of risk, but they can also be used to guide gene identification efforts and to further our understanding of the risk associated with specific genes as they are identified.

Currently, gene-finding efforts for psychiatric disorders (and other common, complex medical conditions) have met with limited success. Findings from genetic epidemiology can be used to inform the phenotypes used in gene-finding studies. For example, based on the twin literature (reviewed above) suggesting that much of the predisposition to alcohol dependence is via a broad externalizing factor, externalizing factor scores were created in the Collaborative Study on the Genetics of Alcoholism (COGA) sample, comprised of symptoms of alcohol and other drug dependence, and childhood and adult antisocial behavior, as well as the personality traits of novelty-seeking and sensation-seeking, which also index general behavioral disinhibition. This latent externalizing factor score was then used in both linkage and association analyses, with results compared with analyzing separately the individual symptoms of each of the psychiatric disorders that went into the creation of the

general externalizing score.[29] The results demonstrated that this broader externalizing phenotype was useful in both linkage and association analyses, suggesting that creating phenotypes grounded in the twin literature can aid in identifying susceptibility genes. Twin data has also been used to aid in genetic association studies in the area of internalizing disorders. Using data from the Virginia Adult Twin Study of Psychiatric and Substance Use Disorders, multivariate structural equation modeling was used to identify common genetic risk factors for major depression, generalized anxiety disorder, panic disorder, agoraphobia, social phobia, and neuroticism. Cases and controls were then identified for genetic association studies based on scoring at the extremes of the genetic factor extracted from the twin analysis, with the subsequent association analyses yielding evidence for association with the gene *GAD1*.[30]

Another area where genetic epidemiology intersects with gene identification efforts is in the characterization of risk associated with identified genes. Most major gene identification efforts for psychiatric disorders currently focus on adult psychiatric outcomes. As we identify genes that are reliably associated with these disorders, one of the next interesting research challenges will be to study how risk associated with these genes unfolds across development and in conjunction with the environment. Here, findings from genetic epidemiology can again be useful in developing hypotheses to test the risk associated with specific genes. For example, based on the twin literature suggesting that adult alcohol dependence and childhood externalizing symptoms overlap in large part due to a shared genetic predisposition,[31] genes that were originally identified as associated with adult alcohol dependence (eg, *GABRA2*,[32] *CHRM2*[33]) have been tested for association with externalizing behavior in younger samples of children and adolescents. These studies suggest that children carrying the genetic variants associated with alcohol problems later in life display elevated rates of conduct problems earlier in development, before any association with alcohol dependence has manifested.[34-36] Further, based on the twin literatures suggesting that genetic influences on externalizing behaviors are moderated by parental monitoring[9] and peer deviance,[37,38] further analyses demonstrated that the associations between these genes and externalizing behavior were stronger under conditions of lower parental monitoring and higher peer deviance. Characterizing the risk pathways associated with identified genes will be critical in eventually translating this information into improved prevention and intervention programs.

Gene identification methods

The field of psychiatric genetics has used two different methods to attempt to identify individual risk genes: linkage and association. These are fundamentally different approaches with different study designs applied, until recently, to very different research questions. It is important to understand both in order to understand why association approaches have become the norm in follow-up studies of linkage regions as well as the primary current approach in genome-wide studies.

DNA polymorphisms

Humans are ~99.9% identical at the nucleotide level on average. Molecular genetic studies depend critically on the remaining 0.1% (~3 million nucleotides) where variation occurs between individuals, collectively known as genetic polymorphisms or markers. Linkage studies generally use short tandem repeat polymorphisms (STRs). STR alleles are differing numbers of a repeating unit of nucleotides and have specific sequence lengths and molecular weights as a result, allowing them to be separated and identified. STRs are very common and tend to be extremely polymorphic (ie, to have many alleles—where an allele is one of the possible variants that exist in a population at a particular genetic locus) and therefore to have high heterozygosity (the proportion of individuals who have two different alleles at the marker locus). This high heterozygosity is important for linkage analyses, which require a unique allele at each position on each homologous chromosome to be informative.

In contrast, single nucleotide polymorphisms (SNPs) are changes of a single base or insertion/deletion variation up to a few nucleotides in size. SNPs generally have only two alleles, and have lower heterozygosity and lower information content. Association studies tend to use SNPs as the marker of choice, because alleles of these markers evolve more slowly than those of STRs and preserve more of the evolutionary relationships on which genetic association is based. SNPs can also be used for linkage, but about ten times as many SNPs as STRs are required to capture the linkage information.

Linkage

In marker genotype data from families, new combinations of alleles at a series of markers on individual chromosomes are observed in each generation. This recombination of alleles is observed because there is at least one physical exchange of material (or crossover) between each homologous chromosome pair in every meiosis *(Figure 1)*. Recombination between loci on different chromosomes (because of independent assortment of homologous chromosome pairs) or far apart on the same chromosome (because of crossover at meiosis) is observed 50% of the time. Linkage is observed between loci in close proximity on a chromosome because their alleles are separated by crossover less than 50% of the time.

Mendelian diseases are caused by mutations in a single gene at a single chromosomal location, so disease phenotypes can be treated as marker alleles in linkage analysis. Because these illnesses are rare, for a dominant disorder, the rare risk allele must segregate from one parent (often affected or with family history) into affected offspring, or arise as an even rarer de novo mutation. By following the segregation of marker alleles from the affected lineage into offspring, linkage between markers and phenotypes can be observed when affected offspring inherit a particular set of marker alleles (and thus a specific parental chromosomal segment) compared with their unaffected relatives.

Association

While linkage occurs in families, association is a population-based phenomenon. Genetic association studies test whether specific alleles at variable sites are more common in individuals affected by a disease (cases) than individuals not affected by the disease (controls). This association between allele and phenotype can occur for two reasons. Either the allele being studied directly influences risk for the disorder or, more commonly, the allele is in linkage disequilibrium (LD) with the disease-predisposing allele. Linkage disequilibrium means that specific alleles at two nearby loci tend to occur together in an entire population. Linkage, (the cosegregation of a chromosome region and a disease observed in families), occurs at scales of tens of millions of base pairs because of the limited number of recombinations observed in each generation of a family. Association (and LD) are seen at scales of thousands to tens of thousands of base pairs, because the number of recombinations present in the evolutionary history of a population is large, meaning that the physical distances between loci in LD must be correspondingly small if recombination is to occur rarely (if ever) between them.

LD occurs because a new allele always arises on a specific background chromosome (and its existing haplotype of marker alleles), and will, until separated by recombination, only exist in conjunction with the other alleles present on that background. Over time, the original LD (and thus the genetic association) between more distant loci decays as a result of recombination events, while the rarity of recombination between nearby loci preserves the original LD and association. Association can also be detected spuriously, eg, if observed differences in allele frequency are due to population differences rather than to true association between marker and phenotype. Association approaches are also substantially reduced in power in the presence of allelic heterogeneity (the existence of more than one risk allele at a locus), while this phenomenon has no effect on the detection of linkage.

Challenges associated with gene identification in psychiatric and substance-use disorders

A number of features of psychiatric and behavioral phenotypes contribute to an overall reduction in study power. Association is more powerful, generally for detecting genes of small effect,[39] but the specific features of psychiatric and behavioral phenotypes also reduce the power of association studies.

First, psychiatric phenotypes are almost certainly influenced by multiple common alleles of small effect in many genes. Both linkage and association study designs are more powerful for alleles of large effect size, and are much less powerful when examining highly polygenic phenotypes. Replication studies are hampered by the need for sample sizes larger than the discovery sample (in order to maintain power) and stochastic sampling variation, the expected variation in the extent to which any specific risk factor is present (and association detectable) in any particular sample.

Second, interactions between genes (GxG) or between genes and environmental variables (GxE) seem necessary

to account for observed risks, but we rely heavily on analytic approaches that assess single genes. In a few cases, genes with known molecular interactions with the candidates have also generated replicated association. Environmental risk factors remain largely unknown and are difficult or very expensive to test in many samples.

Third, these phenotypes are common, so the liability alleles seem likely to be common, although increased rates of rare deletions and duplications (structural or copy number variants) in cases have been observed multiple times and suggest that rare variation may also contribute to risk in a proportion of cases. The common risk variants are expected to occur with relatively high frequency in the general population, reducing contrast between affected and unaffected individuals and reducing power. The impact of individual rare structural variants in the subset of cases where they are observed is harder to assess currently, but the observation of an aggregate increase appears robust, further increasing the apparent etiological complexity.

Fourth, the expected frequency of risk alleles and the clinical variability in presentation, course, and outcome suggest that *the etiology of individual cases may be heterogeneous*, derived from different specific genes or alleles between individuals. Allelic heterogeneity substantially reduces the power of association designs.

Fifth, diagnostic boundaries are difficult to draw, and the best phenotype to study is a complex choice. It is critically important to consider this last point and the phenotypes that yield the strongest evidence in some detail.

An example: schizophrenia gene identification

Through 2004, 25 complete or nearly complete genome scans for schizophrenia (in which about 400 individual genetic markers are genotyped at regular intervals over the entire human genome) were published (for review see refs 40,41). None provided evidence for genes of major effect. Some linkage regions were replicated in these studies, and a number of promising genes emerged from sequential linkage and association studies and multiple replication reports. We focus here on those regions with the best replication record and with evidence emerging from other contemporary studies: 22q12-q13, 8p22-p21, 6p24-p22, and 1q32-42. Two additional regions with little support in the primary literature, 2p11.1-q21.1 and 3p25.3-p22.1, were among the most significant in a meta-analysis of schizophrenia genome scans. A number of other regions (including 5q22-q31 and 15q13-q14) have less strong summary evidence but also overlap with evidence from more recent GWAS and structural variation studies.

Chromosome 22q, the VCFS microdeletion, and *COMT*

Chromosome 22q has been widely studied using many different designs. Primary linkage signals were observed in a few samples but have generally been widely replicated. However, the cosegregation of a known microdeletion in the region with a phenotype in which psychosis is a common feature added significantly to interest in this region. Velo-cardio-facial syndrome (VCFS) is caused by two overlapping, recurrent deletions at 22q11. Historically, about 10% of VCFS patients were thought to present with a psychotic phenotype, but more recent studies suggest much higher rates of 25% to 29%.[42,43] Conversely, preliminary results suggest that about 2% of adult onset and 6% of childhood onset schizophrenic patients have microdeletions in this region, in excess of the estimated general population frequency of such deletions of 0.025%.[44] Interest in this region has been further increased recently by studies assessing structural variation (see below). The gene for catechol-O-methyl transferase (*COMT*), involved in the degradation of catecholamines, maps to this region; the enzyme is functionally polymorphic with a variable amino acid, Val158Met, affecting activity. Although widely studied, the results from genetic studies of *COMT* are inconclusive as reviewed recently.[45]

Chromosome 8p22-p21, *NRG1*, and *ERBB4*

Studies of pedigrees from numerous different ethnic backgrounds have detected linkage to schizophrenia on 8p, as did a statistically robust meta-analysis.[46] Although numerous samples support a locus on 8p, comparison between individual studies is consistent with the presence of multiple susceptibility genes, a feature of a number of linkage regions. Almost certainly the most important result on 8p so far is the widely replicated association with the neuregulin 1 (*NRG1*) gene in fami-

lies and case/controls from Iceland.[47] *NRG1* is a large gene with multiple transcripts yielding distinct protein molecules. It is expressed at central nervous system synapses and is involved in the expression and activation of neurotransmitter (including glutamate) receptors. Initial replication studies[48,49] detected association on haplotypes identical or closely related to those identified in the Icelandic cases; 13 additional studies in multiple populations reported association with more variation in associated alleles or haplotypes,[50-62] while nine studies did not.[63-71] A meta-analysis of studies of *NRG1* supported involvement of the gene in schizophrenia liability, but did not provide evidence supporting association of the most prominent marker in the original studies.[72] In a pattern observed for a number of the best supported schizophrenia genes, several studies have also shown association between *NRG1* and bipolar disorder.[62,73,74]

ErbB4, encoded by the *ERBB4* gene, is a receptor for NRG1 and has important roles in neurodevelopment and the modulation of NMDA receptor functioning. Both activation of ErbB4 and suppression of NMDA receptor activation by NRG1 are increased in the prefrontal cortex in individuals with schizophrenia compared with controls.[75] This functional relationship prompted genetic study of *ERBB4*, which demonstrated association in *ERBB4* and evidence of interaction with *NRG1*.[59,76-78] Associated alleles in *ERBB4* alter splice-variant expression[79] and both NRG1 and ErbB4 protein are increased in the brain in schizophrenia. These results may be of particular importance as there is a biologically plausible mechanism for gene x gene interactions, and even if the interaction is not confirmed, both genes impact the glutamatergic system (supporting the widely held view that part of the complexity may be explained by effects at the level of the pathway or system). Important tests of both interaction and system effects unbiased by candidate selection will be undertaken in the current GWAS datasets.

Chromosome 6p24-p22, *DTNBP1*, and the HLA region

Chromosome 6 has a long history in genetic studies of schizophrenia with major shifts in the apparent importance of particular results. Early linkage studies observed evidence of linkage in human leukocyte antigen (HLA) genes in the major histocompatibility complex (MHC) region on chromosome 6p21.3-22.1, but the limited genome coverage (only ~6%) and lack of replication reduced the apparent importance of these findings. The first strong evidence for linkage of schizophrenia to the 6p region came from studies of Irish families with a high density of disease.[80] This study was also important because it addressed the question of diagnostic boundaries in some detail. Evidence for linkage was modest under a narrow diagnostic model, increased substantially as the diagnostic definition broadened to include psychosis spectrum disorders, and fell when the definition was broadened further to include nonspectrum disorders, in keeping with observed risks in relatives for these traits. Multiple independent studies of this region of 6p observed evidence for linkage, as did a multicenter collaborative study[81] and a robust meta-analysis.[46]

The dystrobrevin binding protein 1 or dysbindin (*DTNBP1*) gene was first reported to be associated in the same Irish families.[82,83] Many studies support association in *DTNBP1* in samples from diverse ethnic backgrounds although the markers, alleles and haplotypes associated vary significantly from study to study: 13 studies of 15 independent samples reported significant positive association with schizophrenia (most consistently with common alleles and the highest frequency common allele haplotype),[70,82-93] while 14 studies of 18 independent samples did not.[61,63,85,94-104] A further four studies have also provided positive evidence for association of *DTNBP1* with bipolar disorder.[105-108] Although the function of DTNBP1 in brain is unknown, both RNA[109] and protein[110] expression is reduced in cases.

Chromosome 1q and *DISC1*

Interest in chromosome 1 in schizophrenia began with reports of a balanced 1:11 translocation segregating with serious mental illness in a large pedigree from Scotland.[111] The chromosome 1 breakpoint lies at 1q42.1, and the breakpoint directly disrupts a novel gene, Disrupted in Schizophrenia 1 (*DISC1*).[112] There are now nine positive reports of association of *DISC1* with schizophrenia[74,113-120] and 2 of association with positive symptoms[121,122] suggesting that this gene influences schizophrenia liability in the general population, as well as in the family with the chromosomal anomaly. Other rare variants in this gene besides the breakpoint have also been reported to be associated with schizophrenia[123,124] and association has been reported for additional psychiatric diagnoses, reviewed in ref 125, and for bipolar

disorder.[126] A smaller number of negative reports have also been published.[103,127-130]

Other chromosomal regions and genes

Two additional chromosome regions, 5q22-q31, where association was recently reported in the interleukin-3 (*IL3*) gene[131] and 15q13-q14, where evidence for linkage of an evoked potential abnormality common in patients[132] was supported by five additional studies reporting linkage of schizophrenia to the same narrow region,[133-137] show some overlap with the results of current studies discussed below. Other high-profile candidate genes such as *PRODH2* on 22q[138] and *PPP3CC* on 8p[139] have not replicated well. One exception is *AKT1*,[140] which has similar numbers of positive[141-145] and negative[61,103,146-149] replications.

Genome-wide association studies

By assaying 500 000 to 1 000 000 DNA variants in a single experiment, GWAS provide unbiased genome-wide coverage, avoiding selection of candidate genes. They use an association framework for analysis, avoiding the weaknesses of linkage in complex traits. They impose stringent criteria due to the number of tests performed (typically around $P<5 \times 10^{-8}$ for genome-wide significance). They hold enormous potential to move beyond the identification of single genes (which may show small effects and be difficult to detect individually) toward the simultaneous identification of multiple genes through their interactions or involvement in systems.

Seven GWAS of schizophrenia have been published to date, four of which were small and underpowered. The first (320 cases, 325 controls) was of limited density as it genotyped only 25 000 SNPs in 14 000 known genes, and did not detect any association that reached genome-wide significance[150]; nominal association was reported in the plexin A2 (*PLXNA2*) gene. Only one of four samples tested in three independent studies replicates the association.[151-153] The second (extremely underpowered with 178 cases, 144 controls) identified one genome-wide significant association in the X/Y pseudoautosomal region (a homologous region of the sex chromosomes where recombination can occur), near the interleukin 3 receptor (*IL3R*) gene.[154] Cytokines have been suggested as possible candidates previously and *IL3* (in the 5q linkage region) was associated with schizophrenia in one study.[131] One replication attempt supported association in *IL3R*.[155] The third, using the CATIE[156] sample (738 cases, 733 controls), did not detect any genome-wide significant results in its primary analysis.[157] The fourth, using a multistage design of discovery (479 cases, 2937 controls) and targeted replication (6666 cases, 9897 controls) samples, identified one genome-wide significant SNP in the zinc-finger protein transcription factor *ZNF804A* gene,[158] but only in the meta-analysis including the original sample. One independent replication attempt supported the association of *ZNF804A*, and showed that expression was increased from the associated haplotype.[159]

Three substantially larger GWAS of schizophrenia were published in 2009, in the SGENE+ sample[160] (multiple European sites, 2663 cases/13498 controls), the International Schizophrenia Consortium (ISC) sample[161] (multiple European sites, 3322 cases/3587 controls) and the Molecular Genetics of Schizophrenia (MGS) sample[162] (multiple US sites, European ancestry: 2681 cases/2653 controls; African ancestry: 1286 cases/973 controls), analyzed both separately and together. The one region of the genome with significant overlap in signals from the 3 studies was the MHC region on chromosome 6p21.3-p22.1, site of some of the earliest genetic evidence in schizophrenia discussed above. The SGENE+ sample detected significant association with several markers spanning the MHC region, as well as signals upstream of the neurogranin (*NRGN*) gene on 11q24.2 and in intron four of the transcription factor 4 (*TCF4*) gene on 18q21.2. The ISC sample detected association in ~450 SNPs spanning the MHC region and the myosin XVIIIB (*MYO18B*) gene on 22q and supported *ZNF804A*. The MGS sample did not detect any individual genome-wide significant signals, but detected signals in the range of 10^{-5}-10^{-7} in the *CENTG2* gene (reported deleted in autism cases[163]) on chromosome 2q37.2 and *JARID2* (the gene adjacent to *DTNBP1*) in European-ancestry subjects, and in *ERBB4* and *NRG1* in African-American subjects.

Meta-analysis of data from all European-ancestry MGS, ISC and SGENE samples detected genome-wide significant association signals for 7 SNPs spanning 209 Kb of the MHC region. LD is high between the 7 SNPs and extends over a region of 1.5 Mb on chromosome 6p22.1, making it difficult to determine if the signal is driven by one or many genes. The genic content of this region is not limited to histocompatibility loci, and also includes genes

involved in transcriptional regulation, DNA repair, chromatin structure, G-protein-coupled-receptor signaling and the nuclear pore complex.

Meta-analyses of schizophrenia linkage and association data

The strongest linkage meta-analysis approach ranks 30 cM bins of the genome from most positive to least positive for each study, and then sums the ranks for each bin. Significance levels are calculated by simulation, and this method can identify regions of the genome where modest positive results occur across many studies. Results of this approach supported linkage to chromosomes 6p and 8p among the previously identified regions discussed above.[46] The strongest evidence for a potential locus was on chromosome 2p11.1-q21.1, a region suggested by only a few studies and not widely followed up, and on 3p, the site of an early linkage finding that could never be replicated.

A recent effort has been made to systematize the collection and archiving of association data from studies of schizophrenia, and to provide a framework for continuous updating of both the data and the meta-analytic results[164] in the SzGene database (http://www.szgene.org/). Meta-analyses of the data contained in this resource provided support of varying degrees for 24 SNPs in 16 previously reported genes, including older candidate genes (eg, dopamine receptor 2 (*DRD2*) gene, those resulting from association-based follow-up of linkage data (eg, *DTNBP1*) and one suggested by one of the smaller GWAS (*PLXNA2*). Meta-analyses of schizophrenia GWAS data from at least 15 000 cases and 15 000 controls are scheduled for completion in 2010.

Rare structural variation in schizophrenia

The epidemiological and genetic data above seems most consistent with the common disease/common variant hypothesis of the genetic risks for complex traits and the results of GWAS in other complex traits like type 2 diabetes provided a major validation of this model.[165-168] The alternative common disease/rare variant hypothesis of genetic risks for complex traits has been proposed in schizophrenia,[169] largely based on the reduction in fertility observed in cases. A key focus of research in this area has been the deletions, duplications, and inversions of a few thousand (Kb) to a few million (Mb) base pairs collectively known as structural variants, an area of intense research interest generally since 2004,[170-172] reviewed in ref 173. As a class, these genomic rearrangements are common: ~360 Mb or 12% of the genome is included in structural variation.[174] A few such variants occur at high frequency due to apparent selection in certain contexts,[175,176] but studies of large samples consistently show that the majority of structural variants are rare (~50% detected in only one individual).[174]

The aggregate rate of such rare structural variants is significantly increased in individuals with schizophrenia in all four studies that have examined this question.[177-180] Critically, there is substantial overlap in the regions where excess structural variation is observed, most notably on chromosomes 22q11, 15q13.3 and 1q21.1, with some evidence that neurodevelopmental genes are overrepresented, as in[181] and more recently on 16p11.2.[182] However, even considered in aggregate, structural variants are observed in only 15% of schizophrenia cases, and so cannot account for a substantial fraction of the total population risk. Because they are rare, the true impact of individual structural variants on schizophrenia is difficult to validate and interpret, although the replication of excess structural variation in cases on chromosomes 22q11, 15q13.3, and 1q21.1 is extremely encouraging.

Summary of current gene-finding studies

At both the technical/molecular and statistical/conceptual levels, the science of gene discovery in complex disease genetics is moving rapidly. By the time this paper is published, new developments are sure to have arisen. As is common in science in the state of rapid flux, the direction ahead is far from clear. How will the modest but hard-fought advances obtained in more traditional positional cloning and candidate gene work integrate with the new findings from GWAS? How will the common-variant SNP-based approach inter-relate with the emerging rare-variant copy number variant findings? Will advances in phenotypic assessment or endophenotypes provide critical new insights? How will the burgeoning fields of bioinformatics, expression arrays, and proteomics impact on our gene-finding efforts?

One emerging consensus is that the field needs to move from a "gene-centric" approach toward one that considers "gene networks." For example, many of the candidate genes discussed above are involved in glutamatergic neurotransmission, which may be an important systemic

element in the etiology of schizophrenia. Although a detailed discussion of this theory is outside the scope of this summary, recent reviews of the genetic[183] and neuroscience[184] data and evidence from other studies highlight the positions of the gene products of *NRG1*, *COMT*, and possibly *DTNBP1* among others, in the biochemical and functional pathways influencing the glutamatergic system. Many other possible networks may be involved in the etiology of schizophrenia that, if properly articulated, could aid in our gene-discovery efforts.

Conclusion

We have attempted in this article to review the rapidly evolving field of psychiatric genetics. In the section on genetic epidemiology, we took a conceptual approach focusing on a range of the most interesting questions now being confronted by the field, with the goal of giving the reader a "feel" for the issues. While examining a wide range of disorders, we focused on substance use and externalizing disorders because they clearly illustrated the points we wanted to make. In the section on gene-finding, we decided it would be more useful to "drill down" and illustrate our important themes by focusing on one disorder—schizophrenia.

The major theme that cuts across these two sections is the complexity of the pathways from genetic variation to psychiatric and substance use disorders. Results of the last 20 years have shown that the early prior simple hypothesis of large effect genes that directly causes psychiatric illness was seriously misplaced. We now know that multiple gene variants (as well as—for at least some disorders—genomic rearrangements) are involved at the DNA level. These genetic risk factors then act and interact with each other and with the environment in a complex developmental "dance" to produce individuals at high versus low risk of illness. It is this kind of complexity that the field is now confronting directly.

As one might hope, progress is being made in multiple ways. The field that is moving downward—in a reductionist sense—to more detailed biological mechanisms at the DNA, RNA, and protein levels. These efforts are being driven by rapid technological advances. However, we are straining to develop the conceptual and analytic tools to keep pace with the information generated by these new generation technologies. At the same time, the field is moving out into the environment to clarify the often critical inter-relationship between these two broad classes of risk factors. Equally importantly, it is moving "forward" in emphasizing the importance of time and development.

This can all be confusing and sometimes a bit overwhelming. In a desire to simplify, some, in the "glow" of the new biological tools now available, have devalued the genetic epidemiologic approaches. These approaches, they suggest, focus on "statistics" but not "real genes." However, knowledge gained from genetic epidemiology, in addition to provide a guiding light for molecular approaches, also have their own inherent validity. Studying aggregate genetic risk factors allows us to build etiologic models that can inform prevention efforts, aid policy makers in planning for research programs, and provide critical input into revisions of psychiatric nosology.

We would like to close by emphasizing that knowledge about the role of genetic factors in the etiology of psychiatric illness can be profitably understood from several perspectives. The human mind/brain system—the organ that instantiates psychiatric illness—is surely influenced by processes occurring at the levels of basic molecular biology, neural systems and networks, and psychological, social, and cultural processes.[185] A full understanding of the processes whereby genetic risks lead to the development of psychiatric disorders will surely require considering all these perspectives, each of which contributes a useful viewpoint with methodologies that have important (and different) strengths and limitations.

REFERENCES

1. Kendler KS, Baker JH. Genetic influences on measures of the environment: a systematic review. *Psychol Med*. 2007;37:615-626.
2. Harden KP, Hill JE, Turkheimer E, Emery RE. Gene-environment correlation and interaction in peer effects on adolescent alcohol and tobacco use. *Behav Genet*. 2008;38:339-347.
3. Hill J, Emery RE, Harden KP, Mendle J, Turkheimer E. Alcohol use in adolescent twins and affiliation with substance using peers. *J Abnorm Child Psychol*. 2008;36:81-94.
4. Dick DM, Pagan JL, Holliday C, et al. Gender differences in friends' influences on adolescent drinking: a genetic epidemiological study. *Alcoholism: Clin Exp Res*. 2007;31: 2012-2019.
5. Kendler KS, Jacobson KC, Gardner CO, Gillespie N, Aggen SA, Prescott CA. Creating a social world: a developmental twin study of peer-group deviance. *Arch Gen Psychiatry*. 2007;64:958-965.
6. Kendler KS, Neale M, Kessler R, Heath A, Eaves L. A twin study of recent life events and difficulties. *Arch Gen Psychiatry*. 1993;50:789-796.
7. Plomin R, Lichtenstein P, Pedersen NL, McClearn GE, Nesselroade JR. Genetic influence on life events during the last half of the life span. *Psychol Aging*. 1990;5:25-30.

Lo innato y lo adquirido en la genética neuropsiquiátrica: ¿dónde estamos?

Se piensa que los factores de riesgo tanto genéticos como no genéticos, al igual que las interacciones y correlaciones entre ellos contribuyen a la etiología de los fenotipos psiquiátricos y conductuales. La epidemiología genética confirma consistentemente la participación de genes en estos defectos. Los estudios de genética molecular han resultado menos exitosos en la identificación de genes defectuosos, pero el progreso reciente sugiere que se ha identificado un número de genes específicos que contribuyen al riesgo. En conjunto los resultados son complejos e inconsistentes, al considerar una sola variante común de ADN en algún gen que influya en el riesgo en poblaciones humanas. Son pocas las variantes genéticas específicas que influyen en el riesgo que se han identificado en forma inequívoca. Sin embargo, las aproximaciones actuales son prometedoras respecto a dilucidar más genes y variantes defectuosas, como también sus potenciales interrelaciones entre ellos y con el ambiente. Se revisarán los campos de la genética molecular y de la epidemiología genética, aportando ejemplos de la literatura para ilustrar los conceptos clave que surgen de este trabajo.

L'inné et l'acquis en génétique neuropsychiatrique : où en sommes-nous ?

Des facteurs de risque génétiques et non génétiques, et leurs interactions et leurs corrélations mutuelles, participeraient à l'étiologie des phénotypes psychiatriques et comportementaux. L'implication des gènes de susceptibilité est régulièrement confirmée par l'épidémiologie génétique. Des études de génétique moléculaire ont été moins heureuses dans l'identification des gènes de susceptibilité, mais des progrès récents suggèrent que plusieurs gènes spécifiques participant au risque ont été identifiés. Pris collectivement, les résultats sont complexes et contradictoires avec un variant ADN unique présent dans un gène, influant sur le risque à travers les populations humaines. Les variants génétiques spécifiques influant sur le risque sont peu nombreux à avoir été identifiés sans ambiguïté. Les approches actuelles sont cependant très prometteuses pour l'identification future des gènes de susceptibilité et de leurs variants, de leurs interrelations éventuelles les uns avec les autres et avec l'environnement. Dans cette revue, nous analyserons les domaines de l'épidémiologie génétique et de la génétique moléculaire, des exemples de la littérature illustrant les idées phares de notre travail.

8. Kendler KS, Karkowski-Shuman L. Stressful life events and genetic liability to major depression: genetic control of exposure to the environment? *Psychol Med*. 1997;27:539-547.

9. Dick DM, Viken R, Purcell S, Kaprio J, Pulkkinen L, Rose RJ. Parental monitoring moderates the importance of genetic and environmental influences on adolescent smoking. *J Abnorm Psychol*. 2007;116:213-218.

10. Koopmans JR, Slutske WS, van Baal GC, Boomsma DI. The influence of religion on alcohol use initiation: evidence for genotype X environment interaction. *Behav Genet*. 1999;29:445-453.

11. Dick DM, Rose RJ, Viken RJ, Kaprio J, Koskenvuo M. Exploring gene-environment interactions: Socioregional moderation of alcohol use. *J Abnorm Psychol*. 2001;110:625-632.

12. Boardman JD. State-level moderation of genetic tendencies to smoke. *Am J Public Health*. 2009;99:480-486.

13. Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ. Major depression and generalized anxiety disorder: Same genes, (partly) different environments? *Arch Gen Psychiatry*. 1992;49:716-722.

14. Kendler KS, Prescott CA, Myers J, Neale MC. The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. *Arch Gen Psychiatry*. 2003;60:929-937.

15. Krueger RF. The structure of common mental disorders. *Arch Gen Psychiatry*. 1999;56:921-926.

16. Vollebergh WAM, Iedema J, Bijl RV, de Graaf R, Smit F, Ormel J. The structure and stability of common mental disorders - The NEMESIS Study. *Arch Gen Psychiatry*. 2001;58:597-603.

17. Pagan JL, Rose RJ, Viken RJ, Pulkkinen L, Kaprio J, Dick DM. Genetic and environmental influences on stages of alcohol use across adolescence and into young adulthood. *Behavior Genetics*. 2006;36:483-497.

18. Kendler KS, Schmitt E, Aggen SH, Prescott CA. Genetic and environmental influences on alcohol, caffeine, cannabis, and nicotine use from early adolescence to middle adulthood. *Arch Gen Psychiatry*. 2008;65:674-682.

19. Pickens RW, Svikis DS, McGue M, Lykken DT, Heston LL, Clayton PJ. Heterogeneity in the inheritance of alcoholism. *Arch Gen Psychiatry*. 1991;48:19-28.

20. Johnson EO, van den Bree MB, Pickens RW. Subtypes of alcohol-dependent men: a typology based on relative genetic and environmental loading. *Alcoholism: Clin Exp Res*. 1996;20:1472-1480.

21. Kendler KS, Eaves LJ, Walters EE, Neale MC, Heath AC, Kessler RC. The identification and validation of distinct depressive syndromes in a population-based sample of female twins. *Arch Gen Psychiatry*. 1996;53:391-399.

22. Bergen SE, Gardner CO, Kendler KS. Age-related changes in heritability of behavioral phenotypes over adolescence and young adulthood: a meta-analysis. *Twin Res Hum Genet*. 2007;10:423-433.

23. Haworth CM, Wright MJ, Luciano M, et al. The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Mol Psychiatry*. In press.

24. Kendler KS, Gardner CO, Annas P, Neale MC, Eaves LJ, Lichtenstein P. A longitudinal twin study of fears from middle childhood to early adulthood: evidence for a developmentally dynamic genome. *Arch Gen Psychiatry*. 2008;65:421-429.

25. Kendler KS, Gardner CO, Lichtenstein P. A developmental twin study of symptoms of anxiety and depression: evidence for genetic innovation and attenuation. *Psychol Med*. 2008;38:1567-1575.
26. Kendler KS, Gatz M, Gardner C, Pedersen N. A Swedish national twin study of lifetime major depression. *Am J Psychiatry*. 2006;163:109-114.
27. Prescott CA, Aggen SH, Kendler KS. Sex differences in the sources of genetic liability to alcohol abuse and dependence in a population-based sample of U.S. twins. *Alcohol Clin Exp Res*. 1999;23:1136-1144.
28. Kendler KS. Gender differences in the genetic epidemiology of major depression. *J Gend Specif Med*. 1998;1:28-31.
29. Dick DM, Aliev F, Wang JC, et al. Using dimensional models of externalizing psychopathology to aid in gene identification. *Arch Gen Psychiatry*. 2008;65:310-318.
30. Hettema JM, An SS, Neale MC, et al. Association between glutamic acid decarboxylase genes and anxiety disorders, major depression, and neuroticism. *Mol Psychiatry*. 2006;11:752-762.
31. Slutske WS, Heath AC, Dinwiddle SH, et al. Common genetic risk factors for conduct disorder and alcohol dependence. *J Abn Psychol* 1998;107:363-374.
32. Edenberg HJ, Dick DM, Xuei X, et al. Variations in GABRA2, encoding the a2 subunit of the GABA-A receptor are associated with alcohol dependence and with brain oscillations. *Am J Hum Genet*. 2004;74:705-714.
33. Wang JC, Hinrichs AL, Stock H, et al. Evidence of common and specific genetic effects: Association of the muscarinic acetylcholine receptor M2 (CHRM2) gene with alcohol dependence and major depressive syndrome. *Hum Mol Genet*. 2004;13:1903-1911.
34. Dick DM, Latendresse SJ, Lansford JE, et al. The role of GABRA2 in trajectories of externalizing behavior across development and evidence of moderation by parental monitoring. *Arch Gen Psychiatry*. 2009;66:649-657.
35. Latendresse SJ, Bates J, Goodnight JA, et al. Differential susceptibility to discrete patterns of adolescent externalizing behavior: examining the interplay between CHRM2 and Peer group deviance. *Child Dev*. 2009. In press.
36. Dick DM, Bierut L, Hinrichs AL,et al. The role of GABRA2 in risk for conduct disorder and alcohol and drug dependence across developmental stages. *Behavior Genetics*. 2006;36:577-590.
37. Dick DM, Pagan JL, Viken R, et al. Changing environmental influences on substance use across development. twin research and human genetics 2007;10:315-326.
38. Button TM, Corley RP, Rhee SH, Hewitt JK, Young SE, Stallings MC. Delinquent peer affiliation and conduct problems: a twin study. *J Abnorm Psychol*. 2007;116:554-564.
39. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science*. 1996;273:1516-1517.
40. Riley BP, McGuffin P. Linkage and associated studies of schizophrenia. *Am J Med Genet Sem Med Genet*. 2000;97:23-44.
41. Sullivan PF. The genetics of schizophrenia. *PLoS Med*. 2005;2:e212.
42. Murphy KC, Jones LA, Owen MJ. High rates of schizophrenia in adults with velo-cardio-facial syndrome. *Arch Gen Psychiatry*. 1999;56:940-945.
43. Pulver AE, Nestadt G, Goldberg R, et al. Psychotic illness in patients diagnosed with velo-cardio-facial syndrome and their relatives. *J Nerv Ment Dis*. 1994;182:476-478.
44. Karayiorgou M, Morris MA, Morrow B, et al. Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. *Proc Natl Acad Sci U S A*. 1995;92:7612-7616.
45. Williams HJ, Owen MJ, O'Donovan MC. Is COMT a susceptibility gene for schizophrenia? *Schizophr Bull*. 2007;33:635-641.
46. Lewis CM, Levinson DF, Wise LH, et al. Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. *Am J Hum Genet*. 2003;73:34-48.
47. Stefansson H, Sigurdsson E, Steinthorsdottir V, et al. Neuregulin 1 and susceptibility to schizophrenia. *Am J Hum Genet*. 2002;71:877-892.
48. Stefansson H, Sarginson J, Kong A, et al. Association of neuregulin 1 with schizophrenia confirmed in a Scottish population. *Am J Hum Genet*. 2003;72:83-87.
49. Corvin AP, Morris DW, McGhee K, et al. Confirmation and refinement of an 'at-risk' haplotype for schizophrenia suggests the EST cluster, Hs.97362, as a potential susceptibility gene at the Neuregulin-1 locus. *Mol Psychiatry*. 2004;9:208-213.
50. Williams NM, Preece A, Spurlock G, et al. Support for genetic variation in neuregulin 1 and susceptibility to schizophrenia. *Mol Psychiatry*. 2003;8:485-487.
51. Yang JZ, Si TM, Ruan Y, et al. Association study of neuregulin 1 gene with schizophrenia. *Mol Psychiatry*. 2003;8:706-709.
52. Bakker SC, Hoogendoorn ML, Selten JP, et al. Neuregulin 1: genetic support for schizophrenia subtypes. *Mol Psychiatry*. 2004;9:1061-1063.
53. Li T, Stefansson H, Gudfinnsson E, et al. Identification of a novel neuregulin 1 at-risk haplotype in Han schizophrenia Chinese patients, but no association with the Icelandic/Scottish risk haplotype. *Mol Psychiatry*. 2004;9:698-704.
54. Tang JX, Chen WY, He G, et al. Polymorphisms within 5' end of the Neuregulin 1 gene are genetically associated with schizophrenia in the Chinese population. *Mol Psychiatry*. 2004;9:11-12.
55. Zhao X, Shi Y, Tang J, et al. A case control and family based association study of the neuregulin1 gene and schizophrenia. *J Med Genet*. 2004;41:31-34.
56. Petryshen TL, Middleton FA, Kirby A, et al. Support for involvement of neuregulin 1 in schizophrenia pathophysiology. *Mol Psychiatry*. 2005;10:366-374.
57. Fukui N, Muratake T, Kaneko N, Amagane H, Someya T. Supportive evidence for neuregulin 1 as a susceptibility gene for schizophrenia in a Japanese population. *Neurosci Lett*. 2006;396:117-120.
58. Hall J, Whalley HC, Job DE, et al. A neuregulin 1 variant associated with abnormal cortical function and psychotic symptoms. *Nat Neurosci*. 2006;9:1477-1478.
59. Norton N, Moskvina V, Morris DW, et al. Evidence that interaction between neuregulin 1 and its receptor erbB4 increases susceptibility to schizophrenia. *Am J Med Genet B Neuropsychiatr Genet*. 2006;141:96-101.
60. Thomson PA, Christoforou A, Morris SW, et al. Association of Neuregulin 1 with schizophrenia and bipolar disorder in a second cohort from the Scottish population. *Mol Psychiatry*. 2007;12:94-104.
61. Turunen JA, Peltonen JO, Pietilainen OP, et al. The role of DTNBP1, NRG1, and AKT1 in the genetics of schizophrenia in Finland. *Schizophr Res*. 2007;91:27-36.
62. Georgieva L, Dimitrova A, Ivanov D, et al. Support for Neuregulin 1 as a susceptibility gene for bipolar disorder and schizophrenia. *Biol Psychiatry*. 2008;64:419-427.
63. Hall D, Gogos JA, Karayiorgou M. The contribution of three strong candidate schizophrenia susceptibility genes in demographically distinct populations. *Genes Brain Behav*. 2004;3:240-248.
64. Iwata N, Suzuki T, Ikeda M, et al. No association with the neuregulin 1 haplotype to Japanese schizophrenia. *Mol Psychiatry*. 2004;9:126-127.
65. Thiselton DL, Webb BT, Neale BM, et al. No evidence for linkage or association of neuregulin-1 (NRG1) with disease in the Irish study of high-density schizophrenia families (ISHDSF). *Mol Psychiatry*. 2004;9:777-783.
66. Duan J, Martinez M, Sanders AR, et al. Neuregulin 1 (NRG1) and schizophrenia: analysis of a US family sample and the evidence in the balance. *Psychol Med*. 2005;35:1599-1610.
67. Ingason A, Soeby K, Timm S, et al. No significant association of the 5' end of neuregulin 1 and schizophrenia in a large Danish sample. *Schizophr Res*. 2006;83:1-5.
68. Walss-Bass C, Raventos H, Montero AP, et al. Association analyses of the neuregulin 1 gene with schizophrenia and manic psychosis in a Hispanic population. *Acta Psychiatr Scand*. 2006;113:314-321.
69. Rosa A, Gardner M, Cuesta MJ, et al. Family-based association study of neuregulin-1 gene and psychosis in a Spanish sample. *Am J Med Genet B Neuropsychiatr Genet*. 2007;144B:954-957.
70. Vilella E, Costas J, Sanjuan J, et al. Association of schizophrenia with DTNBP1 but not with DAO, DAOA, NRG1 and RGS4 nor their genetic interaction. *J Psychiatr Res*. 2008;42:278-288.
71. Ikeda M, Takahashi N, Saito S, et al. Failure to replicate the association between NRG1 and schizophrenia using Japanese large sample. *Schizophr Res*. 2008;101:1-8.
72. Munafo MR, Thiselton DL, Clark TG, Flint J. Association of the NRG1 gene and schizophrenia: a meta-analysis. *Mol Psychiatry*. 2006;11:539-546.
73. Green EK, Raybould R, Macgregor S, et al. Operation of the schizophrenia susceptibility gene, neuregulin 1, across traditional diagnostic boundaries to increase risk for bipolar disorder. *Arch Gen Psychiatry*. 2005;62:642-648.

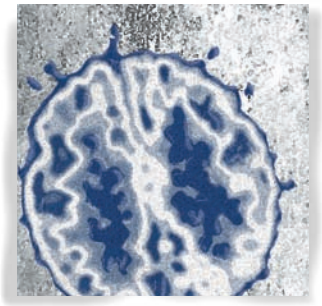
74. Thomson PA, Wray NR, Millar JK, et al. Association between the TRAX/DISC locus and both bipolar disorder and schizophrenia in the Scottish population. *Mol Psychiatry*. 2005;10:657-668.
75. Hahn CG, Wang HY, Cho DS, et al. Altered neuregulin 1-erbB4 signaling contributes to NMDA receptor hypofunction in schizophrenia. *Nat Med*. 2006;12:824-828.
76. Nicodemus KK, Luna A, Vakkalanka R, al. Further evidence for association between ErbB4 and schizophrenia and influence on cognitive intermediate phenotypes in healthy controls. *Mol Psychiatry*. 2006;11:1062-1065.
77. Silberberg G, Darvasi A, Pinkas-Kramarski R, Navon R. The involvement of ErbB4 with schizophrenia: association and expression studies. *Am J Med Genet B Neuropsychiatr Genet*. 2006;141B:142-148.
78. Benzel I, Bansal A, Browning BL, et al. Interactions among genes in the ErbB-Neuregulin signalling network are associated with increased susceptibility to schizophrenia. *Behav Brain Funct*. 2007;3:31.
79. Law AJ, Kleinman JE, Weinberger DR, Weickert CS. Disease-associated intronic variants in the ErbB4 gene are related to altered ErbB4 splice-variant expression in the brain in schizophrenia. *Hum Mol Genet*. 2007;16:129-141.
80. Straub RE, MacLean CJ, O'Neill FA, et al. A potential vulnerability locus for schizophrenia on chromosome 6p24-22: Evidence for genetic heterogeneity. *Nature Genetics*. 1995;11:287-293.
81. Levinson DF, Wildenauer DB, Schwab SG, et al. Additional support for schizophrenia linkage on chromosomes 6 and 8: A multicenter study. *Am J Med Genet Neuropsych Genet*. 1996;67:580-594.
82. Straub RE, Jiang Y, MacLean CJ, et al. Genetic variation in the 6p22.3 gene DTNBP1, the human ortholog of mouse dysbindin, is associated with schizophrenia. *Am J Hum Genet*. 2002;71:337-348.
83. van den Oord E, Sullivan PF, Chen X, Kendler KS, Riley B. Identification of a high risk haplotype for the dystrobrevin binding protein 1 (DTNBP1) gene in the Irish study of high density schizophrenia families. *Mol Psychiatry*. 2003;8:499-510.
84. Schwab SG, Knapp M, Mondabon S, et al. Support for association of schizophrenia with genetic variation in the 6p22.3 gene, dysbindin, in sib-pair families with linkage and in an additional sample of triad families. *Am J Hum Genet*. 2003;72:185-190.
85. Van Den Bogaert A, Schumacher J, Schulze TG, et al. The DTNBP1 (dysbindin) gene contributes to schizophrenia, depending on family history of the disease. *Am J Hum Genet*. 2003;73:1438-1443.
86. Tang JX, Zhou J, Fan JB, Li XW, et al. Family-based association study of DTNBP1 in 6p22.3 and schizophrenia. *Mol Psychiatry*. 2003;8:717-718.
87. Funke B, Finn CT, Plocik AM, et al. Association of the DTNBP1 locus with schizophrenia in a U.S. population. *Am J Hum Genet*. 2004;75:891-898.
88. Kirov G, Ivanov D, Williams NM, et al. Strong evidence for association between the dystrobrevin binding protein 1 gene (DTNBP1) and schizophrenia in 488 parent-offspring trios from Bulgaria. *Biol Psychiatry*. 2004;55:971-975.
89. Numakawa T, Yagasaki Y, Ishimoto T, et al. Evidence of novel neuronal functions of dysbindin, a susceptibility gene for schizophrenia. *Hum Mol Genet*. 2004;13:2699-2708.
90. Williams NM, Preece A, Morris DW, et al. Identification in 2 independent samples of a novel schizophrenia risk haplotype of the dystrobrevin binding protein gene (DTNBP1). *Arch Gen Psychiatry*. 2004;61:336-344.
91. Li T, Zhang F, Liu X, et al. Identifying potential risk haplotypes for schizophrenia at the DTNBP1 locus in Han Chinese and Scottish populations. *Mol Psychiatry*. 2005;10:1037-1044.
92. Tochigi M, Zhang X, Ohashi J, et al. Association study of the dysbindin (DTNBP1) gene in schizophrenia from the Japanese population. *Neurosci Res*. 2006;56:154-158.
93. Tosato S, Ruggeri M, Bonetto C, et al. Association study of dysbindin gene with clinical and outcome measures in a representative cohort of Italian schizophrenic patients. *Am J Med Genet B Neuropsychiatr Genet*. 2007;144B:647-659.
94. DeLuca V, Voineskos D, Shinkai T, Wong G, Kennedy JL. Untranslated region haplotype in dysbindin gene: analysis in schizophrenia. *J Neural Transm*. 2005;112:1263-1267.
95. Holliday EG, Handoko HY, James MR, et al. Association study of the dystrobrevin-binding gene with schizophrenia in Australian and Indian samples. *Twin Res Hum Genet*. 2006;9:531-539.
96. Joo EJ, Lee KY, Jeong SH, Ahn YM, Koo YJ, Kim YS. The dysbindin gene (DTNBP1) and schizophrenia: no support for an association in the Korean population. *Neurosci Lett*. 2006;407:101-106.
97. Liu CM, Liu YL, Fann CS, et al. No association evidence between schizophrenia and dystrobrevin-binding protein 1 (DTNBP1) in Taiwanese families. *Schizophr Res*. 2007;doi:10.1016/j.schres.2007.02.003.
98. Pedrosa E, Ye K, Nolan KA, et al. Positive association of schizophrenia to JARID2 gene. *Am J Med Genet B Neuropsychiatr Genet*. 2007;144B:45-51.
99. Wood LS, Pickering EH, Dechairo BM. Significant support for DAO as a schizophrenia susceptibility locus: examination of five genes putatively associated with schizophrenia. *Biol Psychiatry*. 2007;61:1195-1199.
100. Bakker SC, Hoogendoorn ML, Hendriks J, Verzijlbergen K, Caron S, Verduijn W, et al. The PIP5K2A and RGS4 genes are differentially associated with deficit and non-deficit schizophrenia. *Genes Brain Behav.* 2007;6:113-119.
101. Datta SR, McQuillin A, Puri V, et al. Failure to confirm allelic and haplotypic association between markers at the chromosome 6p22.3 dystrobrevin-binding protein 1 (DTNBP1) locus and schizophrenia. *Behav Brain Funct*. 2007;3:50.
102. Peters K, Wiltshire S, Henders AK, et al. Comprehensive analysis of tagging sequence variants in DTNBP1 shows no association with schizophrenia or with its composite neurocognitive endophenotypes. *Am J Med Genet B Neuropsychiatr Genet*. 2008;147B:1159-1166.
103. Sanders AR, Duan J, Levinson DF, et al. No significant association of 14 candidate genes with schizophrenia in a large European ancestry sample: implications for psychiatric genetics. *Am J Psychiatry*. 2008;165:497-506.
104. Morris DW, McGhee KA, Schwaiger S, et al. No evidence for association of the dysbindin gene [DTNBP1] with schizophrenia in an Irish population-based study. *Schizophr Res*. 2003;60:167-172.
105. Fallin MD, Lasseter VK, Avramopoulos D, et al. Bipolar I disorder and schizophrenia: a 440-single-nucleotide polymorphism screen of 64 candidate genes among Ashkenazi Jewish case-parent trios. *Am J Hum Genet*. 2005;77:918-936.
106. Breen G, Prata D, Osborne S, et al. Association of the dysbindin gene with bipolar affective disorder. *Am J Psychiatry*. 2006;163:1636-1638.
107. Joo EJ, Lee KY, Jeong SH, et al. Dysbindin gene variants are associated with bipolar I disorder in a Korean population. *Neurosci Lett*. 2007;418:272-275.
108. Pae CU, Serretti A, Mandelli L, et al. Effect of 5-haplotype of dysbindin gene (DTNBP1) polymorphisms for the susceptibility to bipolar I disorder. *Am J Med Genet B Neuropsychiatr Genet*. 2007;144B:701-703.
109. Weickert CS, Straub RE, McClintock BW, et al. Human dysbindin (DTNBP1) gene expression in normal brain and in schizophrenic prefrontal cortex and midbrain. *Arch Gen Psychiatry*. 2004;61:544-555.
110. Talbot K, Eidem WL, Tinsley CL, et al. Dysbindin-1 is reduced in intrinsic, glutamatergic terminals of the hippocampal formation in schizophrenia. *J Clin Invest*. 2004;113:1353-1363.
111. St Clair D, Blackwood D, Muir W, et al. Association within a family of a balanced autosomal translocation with major mental illness. *Lancet*. 1990;336:13-16.
112. Millar JK, Wilson-Annan JC, Anderson S, et al. Disruption of two novel genes by a translocation co-segregating with schizophrenia. *Hum Mol Genet*. 2000;9:1415-1423.
113. Hennah W, Varilo T, Kestila M, et al. Haplotype transmission analysis provides evidence of association for DISC1 to schizophrenia and suggests sex-dependent effects. *Hum Mol Genet*. 2003;12:3151-3159.
114. Hodgkinson CA, Goldman D, Jaeger J, et al. Disrupted in schizophrenia 1 (DISC1): association with schizophrenia, schizoaffective disorder, and bipolar disorder. *Am J Hum Genet*. 2004;75:862-872.
115. Callicott JH, Straub RE, Pezawas L, et al. Variation in DISC1 affects hippocampal structure and function and increases risk for schizophrenia. *Proc Natl Acad Sci U S A*. 2005;102:8627-8632.
116. Cannon TD, Hennah W, van Erp TG, et al. Association of DISC1/TRAX haplotypes with schizophrenia, reduced prefrontal gray matter, and impaired short- and long-term memory. *Arch Gen Psychiatry.* 2005;62:1205-1213.
117. Liu YL, Fann CS, Liu CM, et al. A single nucleotide polymorphism fine mapping study of chromosome 1q42.1 reveals the vulnerability genes for schizophrenia, GNPAT and DISC1: Association with impairment of sustained attention. *Biol Psychiatry*. 2006;60:554-562.

118. Zhang F, Sarginson J, Crombie C, Walker N, St CD, Shaw D. Genetic association between schizophrenia and the DISC1 gene in the Scottish population. *Am J Med Genet B Neuropsychiatr Genet*. 2006;141:155-159.
119. Palo OM, Antila M, Silander K, et al. Association of distinct allelic haplotypes of DISC1 with psychotic and bipolar spectrum disorders and with underlying cognitive impairments. *Hum Mol Genet*. 2007;16:2517-2528.
120. Qu M, Tang F, Yue W, et al. Positive association of the Disrupted-in-Schizophrenia-1 gene (DISC1) with schizophrenia in the Chinese Han population. *Am J Med Genet B Neuropsychiatr Genet*. 2007;144:266-270.
121. DeRosse P, Hodgkinson CA, Lencz T, et al. Disrupted in schizophrenia 1 genotype and positive symptoms in schizophrenia. *Biol Psychiatry*. 2007;61:1208-1210.
122. Szeszko PR, Hodgkinson CA, Robinson DG, , et al. DISC1 is associated with prefrontal cortical gray matter and positive symptoms in schizophrenia. *Biol Psychol*. 2007;79:103-110.
123. Sachs NA, Sawa A, Holmes SE, Ross CA, DeLisi LE, Margolis RL. A frameshift mutation in Disrupted in Schizophrenia 1 in an American family with schizophrenia and schizoaffective disorder. *Mol Psychiatry*. 2005;10:758-764.
124. Song W, Li W, Feng J, Heston LL, Scaringe WA, Sommer SS. Identification of high risk DISC1 structural variants with a 2% attributable risk for schizophrenia. *Biochem Biophys Res Commun*. 2008;367:700-706.
125. Hennah W, Thomson P, McQuillin A, et al. DISC1 association, heterogeneity and interplay in schizophrenia and bipolar disorder. *Mol Psychiatry*. 2008;14:865-873.
126. Perlis RH, Purcell S, Fagerness J, et al. Family-based association study of lithium-related and other candidate genes in bipolar disorder. *Arch Gen Psychiatry*. 2008;65:53-61.
127. Kockelkorn TT, Arai M, Matsumoto H, et al. Association study of polymorphisms in the 5' upstream region of human DISC1 gene with schizophrenia. *Neurosci Lett*. 2004;368:41-45.
128. Zhang X, Tochigi M, Ohashi J, et al. Association study of the DISC1/TRAX locus with schizophrenia in a Japanese population. *Schizophr Res*. 2005;79:175-180.
129. Chen QY, Chen Q, Feng GY, et al. Case-control association study of Disrupted-in-Schizophrenia-1 (DISC1) gene and schizophrenia in the Chinese population. *J Psychiatr Res*. 2007;41:428-434.
130. Kim HJ, Park HJ, Jung KH, et al. Association study of polymorphisms between DISC1 and schizophrenia in a Korean population. *Neurosci Lett.* 2008;430:60-63.
131. Chen X, Wang X, Hossain S, et al. Interleukin 3 and schizophrenia: the impact of sex and family history. *Mol Psychiatry*. 2007;12:273-282.
132. Freedman R, Coon H, Myles-Worsley M, et al. Linkage of a neurophysiological deficit in schizophrenia to a chromosome 15 locus. *Proc Natl Acad Sci U S A*. 1997;94:587-592.
133. Riley BP, Makoff A, Mogudi-Carter M, et al. Haplotype transmission disequilibrium and evidence for linkage of the CHRNA7 gene region to schizophrenia in southern African Bantu families. *Am J Med Genet Neuropsychiatric Genet*. 2000;96:196-201.
134. Liu CM, Hwu HG, Lin MW, et al. Suggestive evidence for linkage of schizophrenia to markers at chromosome 15q13-14 in Taiwanese families. *Am J Med Genet*. 2001;105:658-661.
135. Tsuang DW, Skol AD, Faraone SV, et al. Examination of genetic linkage of chromosome 15 to schizophrenia in a large Veterans Affairs Cooperative Study sample. *Am J Med Genet*. 2001;105:662-668.
136. Xu J, Pato MT, Torre CD, et al. Evidence for linkage disequilibrium between the alpha 7-nicotinic receptor gene (CHRNA7) locus and schizophrenia in Azorean families. *Am J Med Genet*. 2001;105:669-674.
137. Gejman PV, Sanders AR, Badner JA, Cao Q, Zhang J. Linkage analysis of schizophrenia to chromosome 15. *Am J Med Genet*. 2001;105:789-793.
138. Liu H, Heath SC, Sobin C, et al. Genetic variation at the 22q11 PRODH2/DGCR6 locus presents an unusual pattern and increases susceptibility to schizophrenia. *Proc Natl Acad Sci U S A.* 2002;99:3717-3722.
139. Gerber DJ, Hall D, Miyakawa T, , et al. Evidence for association of schizophrenia with genetic variation in the 8p21.3 gene, PPP3CC, encoding the calcineurin gamma subunit. *Proc Natl Acad Sci U S A.* 2003;100:8993-8998.
140. Emamian ES, Hall D, Birnbaum MJ, Karayiorgou M, Gogos JA. Convergent evidence for impaired AKT1-GSK3beta signaling in schizophrenia. *Nat Genet*. 2004;36:131-137.
141. Ikeda M, Iwata N, Suzuki T, et al. Association of AKT1 with schizophrenia confirmed in a Japanese population. *Biol Psychiatry*. 2004;56:698-700.
142. Schwab SG, Hoefgen B, Hanses C, et al. Further evidence for association of variants in the AKT1 gene with schizophrenia in a sample of European sib-pair families. *Biol Psychiatry*. 2005;58:446-450.
143. Bajestan SN, Sabouri AH, Nakamura M, et al. Association of AKT1 haplotype with the risk of schizophrenia in Iranian population. *Am J Med Genet B Neuropsychiatr Genet*. 2006;141:383-386.
144. Xu MQ, Xing QH, Zheng YL, et al. Association of AKT1 gene polymorphisms with risk of schizophrenia and with response to antipsychotics in the Chinese population. *J Clin Psychiatry*. 2007;68:1358-1367.
145. Thiselton DL, Vladimirov VI, Kuo PH, et al. AKT1 is associated with schizophrenia across multiple symptom dimensions in the Irish study of high density schizophrenia families. *Biol Psychiatry*. 2008;63:449-457.
146. Ohtsuki T, Inada T, Arinami T. Failure to confirm association between AKT1 haplotype and schizophrenia in a Japanese case-control population. *Mol Psychiatry*. 2004;9:981-983.
147. Ide M, Ohnishi T, Murayama M, et al. Failure to support a genetic contribution of AKT1 polymorphisms and altered AKT signaling in schizophrenia. *J Neurochem*. 2006;99:277-287.
148. Liu YL, Fann CS, Liu CM, et al. Absence of significant associations between four AKT1 SNP markers and schizophrenia in the Taiwanese population. *Psychiatr Genet*. 2006;16:39-41.
149. Norton N, Williams HJ, Dwyer S, et al. Association analysis of AKT1 and schizophrenia in a UK case control sample. *Schizophr Res*. 2007;97:271-276.
150. Mah S, Nelson MR, DeLisi LE, et al. Identification of the semaphorin receptor PLXNA2 as a candidate for susceptibility to schizophrenia. *Mol Psychiatry*. 2006;11:471-478.
151. Fujii T, Iijima Y, Kondo H, et al. Failure to confirm an association between the PLXNA2 gene and schizophrenia in a Japanese population. *Prog Neuropsychopharmacol Biol Psychiatry*. 2007;31:873-877.
152. Takeshita M, Yamada K, Hattori E, et al. Genetic examination of the PLXNA2 gene in Japanese and Chinese people with schizophrenia. *Schizophr Res*. 2008;99:359-364.
153. Budel S, Shim SO, Feng Z, Zhao H, Hisama F, Strittmatter SM. No association between schizophrenia and polymorphisms of the PlexinA2 gene in Chinese Han Trios. *Schizophr Res*. 2008;99:365-366.
154. Lencz T, Morgan TV, Athanasiou M, et al. Converging evidence for a pseudoautosomal cytokine receptor gene locus in schizophrenia. *Mol Psychiatry*. 2007;12:572-580.
155. Sun S, Wang F, Wei J, et al. Association between interleukin-3 receptor alpha polymorphism and schizophrenia in the Chinese population. *Neurosci Lett*. 2008;440:35-37.
156. Lieberman JA, Stroup TS, McEvoy JP, et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N Engl J Med*. 2005;353:1209-1223.
157. Sullivan PF, Lin D, Tzeng JY, et al. Genomewide association for schizophrenia in the CATIE study: results of stage 1. *Mol Psychiatry*. 2008;13:570-584.
158. O'Donovan MC, Craddock N, Norton N, et al. Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet*. 2008;40:1053-1055.
159. Riley B, Thiselton D, Maher BS, et al. Replication of association between schizophrenia and ZNF804A in the Irish Case–Control Study of Schizophrenia sample. *Mol Psychiatry*. 2010;15:29-37.
160. Stefansson H, Ophoff RA, Steinberg S, et al. Common variants conferring risk of schizophrenia. *Nature*. 2009;460:744-747.
161. Purcell SM, Wray NR, Stone JL, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460:748-752.
162. Shi J, Levinson DF, Duan J, et al. Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature*. 2009;460:753-757.
163. Wassink TH, Piven J, Vieland VJ, et al. Evaluation of the chromosome 2q37.3 gene CENTG2 as an autism susceptibility gene. *Am J Med Genet B Neuropsychiatr Genet.* 2005;136B:36-44.
164. Allen NC, Bagade S, McQueen MB, et al. Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat Genet*. 2008;40:827-834.

165. Saxena R, Voight BF, Lyssenko V, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*. 2007;316:1331-1336.
166. Scott LJ, Mohlke KL, Bonnycastle LL, et al. A genome-wide association study of type 2 diabetes in finns detects multiple susceptibility variants. *Science*. 2007;316:1341-1345.
167. Sladek R, Rocheleau G, Rung J, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature*. 2007;445:881-885.
168. Steinthorsdottir V, Thorleifsson G, Reynisdottir I, et al. A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat Genet*. 2007;39:770-775.
169. McClellan JM, Susser E, King MC. Schizophrenia: a common disease caused by multiple rare alleles. *Br J Psychiatry*. 2007;190:194-199.
170. Sebat J, Lakshmi B, Troge J, et al. Large-scale copy number polymorphism in the human genome. *Science*. 2004;305:525-528.
171. Iafrate AJ, Feuk L, Rivera MN, et al. Detection of large-scale variation in the human genome. *Nat Genet*. 2004;36:949-951.
172. Tuzun E, Sharp AJ, Bailey JA, et al. Fine-scale structural variation of the human genome. *Nat Genet*. 2005;37:727-732.
173. Bailey JA, Eichler EE. Primate segmental duplications: crucibles of evolution, diversity and disease. *Nat Rev Genet*. 2006;7:552-564.
174. Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. *Nature*. 2006;444:444-454.
175. Gonzalez E, Kulkarni H, Bolivar H, et al. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science*. 2005;307:1434-1440.
176. Perry GH, Dominy NJ, Claw KG, et al. Diet and the evolution of human amylase gene copy number variation. *Nat Genet*. 2007;39:1256-1260.
177. Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M. Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet*. 2008;40:880-885.
178. Walsh T, McClellan JM, McCarthy SE, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science*. 2008;320:539-543.
179. Stefansson H, Rujescu D, Cichon S, et al. Large recurrent microdeletions associated with schizophrenia. *Nature*. 2008;455:232-236.
180. International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*. 2008;455:237-241.
181. Rujescu D, Ingason A, Cichon S, et al. Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum Mol Genet*. 2009;18:988-996.
182. McCarthy SE, Makarov V, Kirov G, et al. Microduplications of 16p11.2 are associated with schizophrenia. *Nat Genet*. 2009;41:1223-1227.
183. Harrison PJ, Owen MJ. Genes for schizophrenia? Recent findings and their pathophysiological implications. *Lancet*. 2003;361:417-419.
184. Moghaddam B. Bringing order to the glutamate chaos in schizophrenia. *Neuron*. 2003;40:881-884.
185. Kendler KS, Parnas J. *Philosophical Issues in Psychiatry: Explanation, Phenomenology and Nosology*. First ed. Baltimore, MD: Johns Hopkins University Press; 2008.

Epigenetic approaches to psychiatric disorders

Carolyn Ptak, BScH; Arturas Petronis, MD, PhD



Psychiatric diseases place a tremendous burden on affected individuals, their caregivers, and the health care system. Although evidence exists for a strong inherited component to many of these conditions, dedicated efforts to identify DNA sequence-based causes have not been exceptionally productive, and very few pharmacologic treatment options are clinically available. Many features of psychiatric diseases are consistent with an epigenetic dysregulation, such as discordance of monozygotic twins, late age of onset, parent-of-origin and sex effects, and fluctuating disease course. In recent years, experimental technologies have significantly advanced, permitting in-depth studies of the epigenome and its role in maintenance of normal genomic functions, as well as disease etiopathogenesis. Here, we present an epigenetic explanation for many characteristics of psychiatric disease, review the current literature on the epigenetic mechanisms involved in major psychosis, Alzheimer's disease, and autism spectrum disorders, and describe some future directions in the field of psychiatric epigenomics.

Dialogues Clin Neurosci. 2010;12:25-35.

Epigenetics, complex disease, and the brain

In general, epigenetics refers to the regulation of DNA sequences that does not involve alteration of their actual base composition. Transcription and numerous other genomic functions are epigenetically controlled via heritable, but potentially reversible, changes in modification of DNA and histones (acetylation, methylation, phosphorylation, etc),[1] and epigenomics is the application of these processes across the genome.

The normal functioning of genomes is tightly connected to their epigenetic regulation, and epimutations can be harmful in the presence of impeccable DNA sequences. The epigenetic theory of complex non-Mendelian disease is based on three key postulates. Firstly, an organism's epigenetic status is far more dynamic than its DNA sequence, and may be altered by a number of factors, such as environment, developmental programs,[2] or even as a result of stochasticity.[3] Secondly, certain epigenetic signals may be inherited transgenerationally with DNA sequence[4] and may account for heritability of some traits

Keywords: *epigenomics; psychiatry; DNA methylation; histone; psychosis; schizophrenia; bipolar disorder: Alzheimer's disease; autism*

Author affiliations: The Krembil Family Epigenetics Laboratory, Centre for Addiction and Mental Health, and University of Toronto, Toronto, Ontario, Canada

Address for correspondence: Dr Arturas Petronis, The Krembil Family Epigenetics Laboratory, Centre for Addiction and Mental Health, 250 College St, room R28, Toronto, Ontario M5T 1R8, Canada
(e-mail: arturas_petronis@camh.net)

Selected abbreviations and acronyms

AD	*Alzheimer's disease*
ASD	*autism spectrum disorders*
BD	*bipolar disorder*
DNMT	*DNA methyltransferase*
GABA	*γ-aminobutyric acid*
GAD	*glutamate decarboxylase*
HDAC	*histone deacetylase*
LOAD	*late-onset Alzheimer's disease*
RTT	*Rett syndrome*
SZ	*schizophrenia*

and diseases.[5] Thirdly, epigenetic regulation is required in the maintenance of proper genomic function, for example, regulation of gene activity, inactivation of parasitic DNA elements, and chromosomal segregation.[6] Epigenetic factors greatly affect phenotype—even genes that are free of mutations may become harmful if they are not expressed at the appropriate time and at the required level. Combined, these points provide a solid, mechanistic basis for a cohesive interpretation of various epidemiological, clinical, and molecular features of complex diseases.

The molecular epigenetic mechanisms are complex and highly intertwined. At the most basic level, methyl groups may be bound to cytosines at the C_5 carbon, usually within cytosine/guanine dinucleotides (CpG), which are established and maintained by the DNA methyltransferase (DNMT) family of enzymes. This is believed to be the most stable epigenetic mark, due to the covalent nature of the modification.[7] Additionally, another DNA modification, hydroxymethylcytosine, has very recently been discovered in Purkinje neurons and other cells of the brain, and it may also play a role in epigenetic regulation of neural function.[8]

DNA is wrapped around octamers of basic histone proteins, each consisting of a core and N-terminus, to form nucleosomes. Numerous modifications of these proteins influence the condensation of chromatin, which can be open (transcriptionally active) or closed (inactive). Histone acetyltransferases (HATs) acetylate lysine residues on the N-terminal tail of histone proteins, neutralizing the positive charge of the protein and decreasing its affinity for DNA. As a result, the chromatin relaxes and the transcription machinery gains access to previously restricted sites.[9] Acetyl groups can be removed by histone deacetylases (HDACs), resulting in chromatin condensation and transcriptional inactivation.[10] The presence of an N-terminal methyl-CpG-binding domain (MBD) allows proteins, such as methyl-CpG-binding protein 2 (MeCP2), to bind methylated sites on DNA and complex with HDACs and the corepressor SIN3A. The complex facilitates histone deacetylation and downstream gene silencing from the methylated CpG site. Histone methylation can result in either gene activation or repression, depending on the specific lysine or arginine that is modified.[11] Another family of enzymes, the histone demethylases, such as lysine-specific demethylase 1 (LSD1), are capable of removing this methyl group from the lysine residues of histone and nonhistone proteins.[12]

A hallmark of non-Mendelian disease, discordance of monozygotic (MZ) twins, has traditionally been attributed to differential environmental factors activating a disease state in one of the genetically predisposed co-twins[13]; however, very few of these factors have been identified. Alternately, MZ twin discordance may be due to the partial stability of epigenetic factors, as disease-relevant epigenetic dissimilarity can accumulate quite readily between cotwins.[5,14,15] Another non-Mendelian peculiarity, sexual dimorphism, is the differential susceptibility to a disease between males and females. It is observed in many psychiatric conditions, such as Alzheimer's disease, schizophrenia, alcoholism, and mood and anxiety disorders.[16] Although the exact mechanism by which they predispose or protect from a disease is currently unknown, there is a great deal of evidence that sex hormones exert control of gene expression via epigenetic modifications; thus it is hypothesized that sexual dimorphism in many disease states may be the result of sex hormone-induced differences in the epigenetic status of key genes.[17,18] Furthermore, the degree of risk for acquiring certain complex diseases may depend on the sex of the affected parent, as in schizophrenia,[19] Alzheimer's disease (AD),[20] autism,[21] and bipolar disorder (BD).[22] Genomic imprinting, an epigenetic mechanism in which differential epigenetic modification of genes occurs depending on their parental origin,[23] is thought to be the source of such parent-of-origin effects. Diseases affecting cell growth, development, and behavior may result from disruption of the normal imprinting pattern.[24]

In the epigenetic model of complex disease, it is assumed that a primary epigenetic disruption takes place during the maturation of the germline, and this pre-epimutation increases an organism's risk of acquiring a disease. The

pre-epimutation may be tolerated and it may not be sufficient to cause the disease itself, but with time, perhaps even decades, small misregulations add up until a threshold is crossed and the individual experiences phenotypic changes that meet diagnostic criteria for a clinical disorder. The age of disease onset may depend on the effects of tissue differentiation, stochastic factors, hormones, and likely some external environmental factors (nutrition, infections, medications, addictions, etc).[6,25,26] Severity of epimutations may fluctuate over time, due to their reversible nature, known to clinicians as "remission" and "relapse." It is also possible that epimutations may regress back to the norm with aging, which presents partial recovery, eg, reduction of psychopathology in elderly psychiatric patients.

Although there are very few studies investigating the role of epigenetic factors in psychiatric diseases, there is an increasing body of experimental evidence that epigenetic signals play a critical role in neuronal development, differentiation, and communication, as well as synaptic plasticity in general[27]; these processes are fundamental for normal brain activity, such as learning and memory.[28,29] The known epigenetic modifiers, Polycomb (PcG), and Trithorax (TrxG) proteins, have been shown to influence synaptic plasticity,[30,31] and cascade activation during memory formation in the mitogen activated protein kinase (MAPK) pathway appears to trigger H3K14 acetylation.[32] Additionally, pharmacologic inhibitors of epigenetic processes have had documented effects on long-term potentiation (LTP), an increase in efficiency of synaptic transmission, in the mammalian brain. DNMT inhibitors, such as zebularine, impair induction of LTP in mouse hippocampus,[33] while HDAC inhibitors (HDACi), such as sodium butyrate and trichostatin A (TSA), have been shown to enhance LTP in rat hippocampus[32] and amygdala.[34] Taken together, this theoretical and experimental evidence suggest that epigenetic regulation is essential for neural and brain functioning, and putative epimutations may play a role in etiopathogenesis of complex psychiatric disease.

Psychiatric epigenetics and epigenomics

Major psychosis

Major psychosis is a classification that encompasses both schizophrenia (SZ) and BD—two conditions that seem to be related etiologically.[35] SZ is a multifactorial disease characterized by disordered thinking and concentration that results in psychotic thoughts (delusions and hallucinations), inappropriate emotional responses, erratic behavior, as well as social and occupational deterioration,[36] while BD represents a category of mood disorders, in which affected individuals experience episodes of mania or hypomania interspersed with periods of depression, and may also suffer from delusions and hallucinations. Thus far, traditional gene- and environment-based approaches have not been very successful in deciphering the clinical, molecular, and epidemiological aspects of psychosis, such as MZ discordance (41% to 65% for SZ,[37] ~60% BD[38]), sexual dimorphism, parent-of-origin effects, fluctuating disease course with periods of remission and relapse, and peaks of susceptibility to the disease that correspond to periods of major hormonal changes in the organism.[25] Classically, psychosis research was aimed at defining genetic and environmental risk factors, but despite significant evidence of a heritable component derived from twin and adoption studies,[39,40] many molecular genetic findings have not been replicated, and significant heterogeneity and small effect sizes are thought to plague genetic association studies.[41]

Recently, the first epigenomic study of major psychosis utilizing CpG-island microarrays was released by Mill et al,[42] providing a large-scale overview of DNA methylation differences in the brain associated with SZ and BD. DNA extracted from the frontal cortex was subjected to enrichment of the unmethylated fraction using the methylation-sensitive restriction enzymes, and adaptor ligation coupled with PCR amplification. The amplicons (multiple copies of the unmethylated genomic DNA) were interrogated on 12 192 feature CpG-island microarrays. The data was normalized, assigned raw *P* values based on a *t* statistic, and then converted to false discovery rates (FDR). Indeed, in cortex they discovered differences at loci involved in glutamatergic and γ-aminobutyric acid (GABA)-ergic neurotransmission, brain development, mitochondrial function, stress response, and other disease-related functions, many of which correspond to psychosis-related changes in steady-state mRNA. In relation to the glutamatergic hypothesis, a lower degree of DNA methylation was observed in SZ and combined male psychosis (SZ and BD) samples at two glutamate receptor genes, *NR3B* and the α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor (AMPA) receptor-subunit gene *GRIA2*; the dysregulation of AMPA and N-

methyl-D-aspartic acid (NMDA) receptors is an etiological component of major psychosis, and it has been shown that *GRIA2* expression is altered in the prefrontal cortex and striatum of SZ patients.[43] Hypomethylation was also detected at the vesicular glutamate transporter (*VGLUT2*) in SZ females, and at secretogranin II (*SCG2*), which encodes a neuronal vesicle protein that stimulates glutamate release. A higher degree of methylation was observed in SZ females at *VGLUT1*, a transporter protein that is downregulated in SZ brains,[44] and the glutaminase enzyme, *GLS2*, in SZ males, which has previously been shown to exhibit altered expression in cases of SZ.[45] In synergy with glutamatergic pathways, GABAergic pathways also show dysregulation in cases of major psychosis. Detected disruptions in such pathways included hypermethylation at the RNA-binding regulator of GABA(B) receptors, *MARLIN-1*, in SZ, BD, and psychosis females, the G protein-coupled inwardly rectifying potassium channel linked to GABA neurotransmission, *KCNJ6*, in SZ and psychosis males, as well as the *HELT* locus in SZ and BD females, which is known to determine GABAergic over glutamatergic neuronal fate in the mesencephalon. Several other intriguing loci were highlighted, such as the hypermethylation at *WNT1*, a gene critical for neurodevelopment that is differentially expressed in SZ brains,[46] in females affected with major psychosis, and at *AUTS2* in SZ males, which spans a translocation breakpoint associated with autism and mental retardation. A highly significant hypermethylation was detected in both male and female samples at two loci: *RPP21*, which encodes a component of ribonuclease P, a complex that forms t-RNA molecules via 5'-end cleavage, and *KEL*, which encodes the Kell blood-group glycoprotein and causes McLeod Syndrome when incorrectly expressed; SZ symptoms are manifested as part of McLeod Syndrome. Network and gene ontology (GO) analyses were performed in order to determine relationships between the functionally linked pathways from the microarray dataset. The network analysis revealed a lower degree of modularity of DNA methylation "nodes" in the major psychosis samples, indicating that there is some degree of systemic epigenetic dysregulation involved in the disorder. From the GO analysis, several categories were highlighted, including those involved in epigenetic processes, transcription, and development, as well as brain development in female BD and SZ samples, and in those related to stress response in male BD samples.[46] To date, this is the largest and most comprehensive epigenomic study of major psychosis—the data presented supports epigenetic mechanisms underlying broader hypotheses of major psychosis and uncovers some new avenues for future exploration.

Both SZ and BD have also been examined using the candidate gene approach, as epigenetic downregulation of genes is emerging as a possible underlying mechanism of the GABAergic neuronal dysfunction in SZ. One of the more intensively investigated SZ-related genes is *RELN*, which is involved in neuronal development and cell signaling, and has been found to be hypermethylated in cases of SZ.[47] However, no differences were observed at this locus in a replication attempt,[46,48] and the focus seems to be shifting to other candidate genes, namely the 67 kDa glutamate decarboxylase (*GAD67*, aka *GAD1*) and *DNMT1*. GAD67 catalyzes the conversion of glutamic acid to GABA. In cases of SZ, the levels of this enzyme and several others involved in GABAergic neurotransmission, such as *GAD65* and GABA plasma membrane transporter-1 (*GAT-1*), display decreased mRNA levels, as determined by real-time quantitative polymerase chain reaction (qPCR) and in situ hybridization.[49-52] In addition to aberrant methylation at this locus, an analysis of the microarray collection of the National Brain Databank (USA) has shown that decreased GAD67 mRNA levels strongly correlated with upregulated *HDAC1* in the prefrontal cortices of SZ subjects.[53] Oddly enough, at the *GAD67* promoter, SZ patients have been shown to display an ~8-fold deficit in repressive chromatin-associated DNA methylation.[54] In the prefrontal cortex of 41 SZ patients, another histone modification, H3-(methyl)arginine 17 (H3meR17) was found to exceed control levels by 30%, and this was associated with downregulated metabolic gene expression.[55] So, while it is apparent that histone modifications are involved in the development of SZ, their exact mechanism is not entirely clear. Hypermethylation of *GAD67* is believed to occur via DNMT1,[56,57] a maintenance methyltransferase enzyme that is upregulated in the GABAergic neurons and peripheral blood lymphocytes of SZ patients, along with the *de novo* methyltransferase, DNMT3a.[56,58] Interestingly, nicotine has been shown to decrease DNMT1 mRNA expression in cortical and hippocampal GABAergic neurons in mice—this decrease results in *GAD67* promoter demethylation, and is inversely related to an upregulation of cortical GAD67 protein.[59] This information is highly relevant, as SZ patients tend to smoke tobacco at a rate that is 2- to 4-

fold higher than in the general population,[60] and are possibly drawn to the nicotine content for its effects on the aforementioned pathway.

Less information is available on BD; genomic imprinting has been suggested by statistical genetics, but molecular approaches have not yielded the imprinted disease genes.[61] A recent study applied methylation-sensitive representational difference analysis (MS-RDA) to lymphoblastoid cells derived from twins discordant for BD.[62] One detected gene, named peptidylprolyl isomerase E-like (*PPIEL*), was hypomethylated in BD-affected twins, while a region of the spermine synthase (*SMS*) gene was hypermethylated versus unaffected twins; it has yet to be determined if either of these regions are biologically and functionally significant. In combined studies of epigenetics and DNA sequence, some interesting developments have been observed. It has recently been shown that rare G variants of a G/A polymorphism in the potassium chloride co-transporter 3 gene (*SLC12A6*) may represent risk factors for BD.[63] Eventually, it was discovered that variants containing the G allele were methylated at the adjacent cytosine, and this accompanied a decrease in gene expression in human lymphocytes.[64] This hints at a functional link between epigenetics and genetic variation, and the association with BD is believable, as *SLC12A6* mutations underlie another psychiatric disorder, Andermann syndrome, which is an autosomal recessive motor-sensory neuropathy associated with developmental and neurodegenerative defects.[65] It is interesting to note that BD provides a unique opportunity to investigate epigenetic variation between two extreme forms of the same disease—depression and mania. A study design of this variety would unfortunately be limited to the use of peripheral blood, buccal epithelial cells, and fibroblasts as experimental tissues, but nonetheless, it would be incredibly interesting to determine the state of the epigenome during manic and depressive states, in the same individual when the same genetic and environmental impacts are present.

Alzheimer's disease

AD is a neurodegenerative disorder and the most common form of dementia in the elderly; it is characterized by the accumulation of intracellular neurofibrillary tangles (NFT) and extracellular amyloid plaques in the brain.[66] AD often presents with psychiatric symptoms such as memory loss, mood swings, and irritability that increase in severity as the disease progresses. While the phenotype of this condition is well documented, the molecular mechanisms largely remain unknown. The majority of AD research focuses on dysregulation of fibers and proteins, such as epsilon4 allele of apolipoprotein E (APOE), but little ground has been gained in regards to determining the actual origins of their dysfunction.[67] In the rare early-onset form of AD (EOAD), genetic factors play a more defined role, with mutations in amyloid-beta precursor protein (*APP*) and the presenilin genes (*PSEN1*, *PSEN2*) showing a clear connection to the disease.[68] However, since EOAD does not represent the majority of all cases, accounting for only ~5% of the total,[69] this genetic model is not normally applicable.

Similar to other complex diseases, late-onset AD (LOAD), the more common form of the illness that affects individuals over 65 years of age, demonstrates a considerable number of non-Mendelian features. Some of these anomalies include dominance of sporadic over familial cases,[70] discordance of MZ twins,[71] differential susceptibility and course of illness in males and females,[16,18] parent-of-origin effects[72] and, clearly, the late age of onset that is not easily explained by genetic causes alone. Consistent with the epigenetic hypothesis, abnormal levels of folate and homocysteine, signs of dysregulated methylation maintenance, have been detected in the brain of AD subjects. LOAD is a particularly interesting target from the epigenetics of aging perspective, as the epigenome may become deregulated in old age.[73] Using a MethyLight approach, it was shown that a large number of genes increase in methylation with age in control subjects, including several implicated in AD and SZ (*GAD1*, *PSEN1*, *BDNF*, *DRD2*, *GABRA2*, *HOXA1*, *NTF3*, *LDLR*, and *S100A2*), whereas *Alu* and other repetitive elements showed a significant decrease in DNA methylation that was limited to the first decade of life.[74] Of the fifty loci investigated, two displayed significant changes in methylation status with age in AD subjects: *SORBS3* gained methylation over time and is more likely to be methylated in AD patients, while *S100A2* displays a complex chronology, but results in a slow, stochastic methylation decrease later in life (*ibid*). *SORBS3* encodes a neuronal/glial cell adhesion molecule and *S100A2* encodes a calcium binding protein from the S100 family. As part of normal brain aging, S100A2 protein accumulates in corpora amylacea, or polyglucosan bodies; subjects with neurodegenerative disorders experience a much greater accumulation of corpora

amylacea,[75] and this is consistent with the eventual decrease in *S100A2* methylation.[74]

In a study dedicated to DNA methylation analysis of AD candidate genes, it was found that the twelve analyzed loci were epigenetically different in the brains of LOAD cases versus controls, particularly at the locus for transcription factor A (*TFAM*), a key activator of mitochondrial transcription in mammals. Other candidates, such as *PSEN1*, *APOE*, *DNMT1*, and *MTHFR*, displayed an enhanced "distance" in LOAD subjects.[76] This concept of distance is part of the theory of "epigenetic drift," in which an affected individual has an epigenetic status at some gene(s) that is distanced from the norm, and this distance increases with age.[76] Of the CpG-rich regions analyzed, the majority were unmethylated, and it appears possible that very small alterations in methylation level could accumulate over time, ultimately affecting gene regulatory functions and causing disease. Age-related alteration of methylation status is a global phenomenon, not necessarily limited to particular disease susceptibility genes. Another study examined the methylation changes in 807 arbitrarily selected genes from two cohorts from Utah and Iceland, taking DNA samples at two timepoints from each subject, spaced either 11 or 16 years apart. In these two populations, they observed time-dependent changes in global DNA methylation within the same individual, with 8% to 10% of individuals in showing changes that were greater than 20 percent; both gains and losses of methylation were detected.[77] Similarly, the Boston Normative Aging Study measured DNA methylation in the blood of 718 elderly subjects (55 to 92 years of age) over a span of 8 years. A progressive loss of DNA methylation in repetitive elements was found, particularly in Alu repeats, and this linear decline highly correlated with time since the first measurement.[78] A seemingly innocuous early-life epigenetic change in some critical gene involved in AD etiology, for example, the amyloid precursor protein (APP) locus, could potentially become pathologic when subjected to epigenetic drift as the subject ages. Although the molecular mechanisms leading to early-life methylation disturbances have not yet been identified, the possibility of early epimutation and epigenetic drift should not be ignored as an etiological candidate for LOAD.

Autism spectrum disorders

Autism and related developmental disorders, such as Asperger's and Rett syndromes, fall under the broader class of autism spectrum disorders (ASD), where "spectrum" reflects the observed continuum of severity or impairment experienced. These disorders become apparent in young children and persist into adulthood, with deficits in social cognition regarded as the most characteristic feature of ASD, leading to restrictions in social communication.[79] While autism itself is believed to have a particularly strong inherited basis relative to other developmental psychiatric syndromes,[80] DNA sequence factors in the etiology of ASD are still largely unknown.[81] Evidence supports a contribution of imprinted genes in ASD, as well as paternal transmission (reviewed in ref 82), and perhaps the combination of this information and the lack of identified genetic markers will stimulate future epigenetic and epigenomic studies of ASD.

Rett syndrome (RTT), a division of ASD, has been extensively studied and arises from loss of function mutations at the locus for methyl-CpG-binding protein-2 (*MeCP2*), a transcriptional repressor that silences methylated genes[83] and may participate in RNA splicing.[84] Mouse models have been very useful in delineating the relationship between disturbances to *MeCP2* and the disease.[85] In mice, deletion of MeCP2 mimics RTT syndrome, leading to locomotor impairments and reductions in brain size.[86,87] Mice with a truncated MeCP2 protein, similar to that of RTT patients, developed many features of RTT, such as tremors, motor impairments, hypoactivity, increased anxiety-related behavior, seizures, kyphosis, and stereotypic forelimb motions; these mice also presented hyperacetylation on histone H3,[88] illustrating that chromatin abnormalities exist in this disorder. In astrocytes cultured from a mouse model of RTT, MeCP2 deficiency causes significant abnormalities in *BDNF* regulation, cytokine production, and neuronal dendritic induction. Whereas previous experiments have only focused on neurons, this evidence suggests that astrocytes may also represent therapeutic targets for RTT.[89]

The classic form of autism also appears to be connected to *MeCP2* expression. Coding mutations affecting the protein are rarely detected in autism, but significantly increased *MeCP2* promoter methylation has been found in autistic male frontal cortex compared with controls, and this inversely correlated with protein expression[90]; aberrant promoter methylation at *MeCP2* has also been detected in female brain DNA.[91] Similarly, loss of methyl-CpG binding protein 1 (MBD1), leads to autism-like behavioral deficits in mice, namely reduced social

interaction, learning deficits, anxiety, defective sensory motor gating, depression, and abnormal brain serotonin activity.[92] Also, a novel mutation has been discovered in the Jumonji AT-rich interactive domain 1C (JARID1C) gene of a child with autism. While very preliminary, this discovery is interesting, as *JARID1C* is believed to be a histone demethylase specific for di- and trimethylated histone 3 lysine 4 (H3K4), as well as a transcriptional repressor for the ASD-associated genes *SCN2A*, *CACNA1H*, *BDNF*, and *SLC18A1*.[93] Finally, another interesting hypothesis relating epigenetics to ASD concerns the observation that autistic children exhibit improved behavior communication during febrile episodes.[94] It may be the case that fever restores the modulatory functions of the intact, but dysregulated locus coeruleus-noradrenergic (LC-NA) system that is present in ASD. The fact that the state of the LC-NA system can be switched back and forth, combined with evidence that imprinted genes within the LC-NA are tightly epigenetically regulated and susceptible to environmental interference,[95] suggests that dynamic epigenetic remodeling processes may regulate the malfunctioning pathways in ASD.[96]

Epigenetic treatment opportunities

Epigenetic drug strategies are currently employed to treat a collection of cancer subtypes, and these medications are now being considered in the treatment of psychiatric disease, as well. The DNMT inhibitor, doxorubicin, has been used to increase reelin and GAD67 expression in neuronal precursor cells, and it was shown that reelin gene expression correlated with the dissociation of DNMT1 and MeCP2 from its promoter, as well as an increased level of histone H3 acetylation.[97] Other studies have shown that HDAC inhibition enhances learning and memory following neurodegeneration induced by traumatic brain injury,[98] and also shows some therapeutic efficacy in rodent models of neurodegenerative conditions, such as Huntington's disease,[99] multiple sclerosis,[100] and Parkinson's disease.[101] One of the downstream effects of HDAC inhibition is upregulation of p21,[102] a cyclin-dependent kinase inhibitor that appears to play an important protective role against oxidative stress and DNA damage.[103] Valproate, a compound utilized for its anticonvulsant and mood-stabilizing properties, also exhibits HDAC activity and has been successfully implemented as a treatment for epilepsy,[104] BD,[105] and, less commonly, SZ.[106] Like valproate, it has been discovered that several drugs have previously unknown epigenetic modifying properties, and the list continues to grow. While such medications are promising, their pleiotropy, transient effects, and nonspecific alterations to the entire epigenome limit them for the time being.

A substantial challenge to the field of epigenomics of psychiatric and other diseases involves the identification and verification of inhibitors for specific histone-modifying enzymes. Once developed, these compounds should provide higher therapeutic efficiency versus the nonspecific therapeutics that are currently in use, such as suberoylanilide hydroxamic acid (SAHA). The development of small, targeted molecules to specific disease-causing epimutations may resolve some of these issues but, of course, the molecules themselves must first be identified. Alternately, discovery of the downstream effects of epimutations in vivo may nominate particular proteins, to which drug interventions can be applied in a more traditional style, using molecules to exert agonistic and antagonistic effects on the protein products of epigenetically misregulated genes. Knowledge of the three dimensional structures of DNA- and histone-modifying enzymes is mounting and, through the use of fragment-based drug design and ligand motif-based libraries,[107] virtual screening technologies may soon become a feasible option. In the search for target-specific ligands, high-throughput screening of small organic molecule libraries is a useful tool.[108] A recent study utilized a 125 000 small molecule library to screen for specific inhibitors against histone lysine methyltransferases (HMTases). The compound discovered was BIX-01294 (diazepinquinazolinamine derivative), an incredibly specific inhibitor of the target enzyme, euchromatic G9a HMTase, that was able to significantly lower promoter-proximal H3K9me2 marks in mouse embryonic stem cells.[109]

In addition to small molecules, RNA and proteins may also be utilized in the design of effective epigenetic drugs. One strategy focuses on RNA interference (RNAi), in which endogenously produced small interfering RNAs (siRNAs) are incorporated into an RNA-induced silencing complex (RISC) that targets and destroys homologous mRNA, thus preventing protein production.[110] A siRNA with the ability to knock down beta-secretase (*BACE1*) in Huntington's and AD has been developed, as has one against the *SCA1* gene in spinocerebellar ataxia.[111] However, before these RNAs can become effective treatment options, the issues of

nonspecific silencing of partially homologous genes, safe delivery, and inhibition of microRNA (miRNA) must first be resolved. Although the exact mechanisms by which RNAi affects local chromatin structure, gene silencing, and heterochromatin assembly is unknown,[112] it still holds much promise as a therapeutic technique. Another promising technology utilizes zinc-finger proteins (ZFPs), which can recognize specific DNA sequences and bind to short stretches of DNA (~9–18 basepairs), depending on their particular domains.[113] This feature could theoretically allow targeted ZFPs, attached to a DNA- or histone-modifying enzyme,[114] to bind an epimutated site and permit the enzyme to correct the misregulation at that location alone. The damaging global epigenetic effects observed with current drugs would not occur, in this case.

The ability to target etiological disease epimutations and identify epigenetic biomarkers for psychiatric diseases would be another incredibly beneficial development. Biotechnologies are advancing at an amazing rate, and already allow for genome-wide detection of the patterns of DNA methylation and histone modifications. Fully mapped epigenomes in different tissues and cells will facilitate the discovery of disease epimutations and the mechanisms of their pathological action, thus providing the basis for etiological treatment.

Concluding remarks

The role of epigenetic mechanisms in psychiatric diseases is only beginning to solidify, but it is already evident in major psychosis, AD, ASD, and several other conditions not described in this review, such as Rubinstein-Taybi syndrome,[115] addiction,[116,117] Huntington's disease,[118] and Fragile X syndrome.[119] Maintenance of DNA methylation and histone modifications is crucial for normal neurodevelopment and functioning of the brain—dysregulation of these components is highly deleterious to the subject and can predispose to any of the aforementioned disease phenotypes. Previous studies of psychiatric conditions have concentrated on the contributions of genetic and environmental factors but, while DNA sequence and external influences may play an important role in disease etiology, the impact of gene regulation via epigenetic mechanisms on neural function also cannot be ignored. Rather, the interplay between epigenetics, DNA sequences, and environment should become the focus of future work, adopting such concepts as differentially methylated "epi-alleles"[120] and environmental effects on DNA methylation and chromatin modifications.[2] As technologies advance, next-generation sequencing and comprehensive microarrays will become much more affordable, allowing researchers to perform larger, more in-depth epigenomic studies. Perhaps, in the near future, identification of epigenetic biomarkers and operationalization of new, effective diagnostics and treatments will become feasible for psychiatric and various other complex diseases. ❑

Acknowledgements: This project was supported by the National Institute of Mental Health (R01 MH074127; R01 MH088413), and the Canadian Institutes for Health and Research (CIHR). CP is a CIHR Graduate Fellow, and AP is Senior Fellow, of the Ontario Mental Health Foundation.

REFERENCES

1. Henikoff S, Matzke MA. Exploring and explaining epigenetic effects. *Trends Genet*. 1997;13:293-295.
2. Weaver IC, Cervoni N, Champagne FA, et al. Epigenetic programming by maternal behavior. *Nat Neurosci*. 2004;7:847-854.
3. Riggs AD, Xiong Z, Wang L, LeBon JM. Methylation dynamics, epigenetic fidelity and X chromosome structure. *Novartis Found Symp*. 1998;214:214-25; discussion 25-32.
4. Richards EJ. Inherited epigenetic variation--revisiting soft inheritance. *Nat Rev Genet*. 2006;7:395-401.
5. Kaminsky ZA, Tang T, Wang SC, et al. DNA methylation profiles in monozygotic and dizygotic twins. *Nat Genet*. 2009;41:240-245.
6. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet*. 2003;33(suppl):245-254.
7. Davey CA, Sargent DF, Luger K, Maeder AW, Richmond TJ. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. *J Mol Biol*. 2002;319:1097-1113.
8. Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*. 2009;324:929-930.
9. Margueron R, Trojer P, Reinberg D. The key to development: interpreting the histone code? *Curr Opin Genet Dev*. 2005;15:163-176.
10. Thiagalingam S, Cheng KH, Lee HJ, et al. Histone deacetylases: unique players in shaping the epigenetic histone code. *Ann N Y Acad Sci*. 2003;983:84-100.
11. Klose RJ, Zhang Y. Regulation of histone methylation by demethylimination and demethylation. *Nat Rev Mol Cell Biol*. 2007;8:307-318.
12. Nicholson TB, Chen T. LSD1 demethylates histone and non-histone proteins. *Epigenetics*. 2009;4:129-132.
13. Reiss D, Plomin R, Hetherington EM. Genetics and psychiatry: an unheralded window on the environment. *Am J Psychiatry*. 1991;148:283-291.
14. Petronis A, Gottesman II, Kan P, et al. Monozygotic twins exhibit numerous epigenetic differences: clues to twin discordance? *Schizophr Bull*. 2003;29:169-178.
15. Fraga MF, Ballestar E, Paz MF, et al. Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci U S A*. 2005;102:10604-10609.
16. Seeman MV. Psychopathology in women and men: focus on female hormones. *Am J Psychiatry*. 1997;154:1641-1647.
17. Auger AP, Jessen HM. Corepressors, nuclear receptors, and epigenetic factors on DNA: a tail of repression. *Psychoneuroendocrinology*. 2009;34 (suppl 1):S39-S47.

Aproximaciones epigenéticas a los trastornos psiquiátricos

Las enfermedades psiquiátricas determinan un enorme costo para los individuos afectados, sus cuidadores y el sistema de atención de salud. Aunque existe evidencia de un fuerte componente hereditario para muchas de estas condiciones, los esfuerzos dedicados a identificar las causas en base a las secuencias de ADN no han resultado especialmente productivos y son muy pocas las opciones de tratamientos farmacológicos que están clínicamente disponibles. Muchas características de las enfermedades psiquiátricas son concordantes con una falta de regulación epigenética, como la discordancia de los gemelos monocigóticos, la edad tardía de aparición, los efectos del sexo y de los padres biológicos y el curso fluctuante de la enfermedad. Recientemente las tecnologías experimentales han avanzado de manera significativa, permitiendo estudios a fondo del epigenoma y de su papel en el mantenimiento de las funciones genómicas normales, como también en la etiopatogenia de las enfermedades. En este artículo se presenta una explicación epigenética para muchas características de la enfermedad psiquiátrica, se revisa la literatura actual acerca de los mecanismos epigenéticos involucrados en las principales psicosis, la Enfermedad de Alzheimer, y los trastornos del espectro autístico, y se describen algunas líneas a futuro en el campo de la epigenómica psiquiátrica.

Approches épigénétiques des troubles psychiatriques

Les maladies psychiatriques pèsent considérablement sur les individus atteints, leurs soignants et sur le système de santé. Même s'il existe des arguments pour une forte héritabilité de beaucoup de ces troubles, les efforts portés sur l'identification des causes liées à une séquence ADN n'ont pas été très productifs et très peu de traitements pharmacologiques sont disponibles. De nombreuses caractéristiques des maladies psychiatriques concordent avec une dysrégulation épigénétique, comme une discordance entre jumeaux monozygotes, un début tardif, des effets liés au sexe et aux origines parentales et une évolution fluctuante de la maladie. Ces dernières années, des avancées significatives des technologies expérimentales ont permis d'étudier en profondeur l'épigénome et son rôle dans le maintien des fonctions génomiques normales comme dans l'étiopathogenèse de la maladie. Nous présentons dans cet article une explication épigénétique pour de nombreuses caractéristiques des maladies psychiatriques, nous analysons la littérature actuelle sur les mécanismes épigénétiques impliqués dans les psychoses majeures, la maladie d'Alzheimer et les troubles autistiques et nous donnons quelques perspectives dans le domaine de l'épigénomique psychiatrique.

18. Kaminsky Z, Wang SC, Petronis A. Complex disease, gender and epigenetics. *Ann Med*. 2006;38:530-544.

19. Ohara K, Xu HD, Mori N, et al. Anticipation and imprinting in schizophrenia. *Biol Psychiatry*. 1997;42:760-766.

20. Bassett SS, Avramopoulos D, Fallin D. Evidence for parent of origin effect in late-onset Alzheimer disease. *Am J Med Genet*. 2002;114:679-686.

21. Lamb JA, Barnby G, Bonora E, et al. Analysis of IMGSAC autism susceptibility loci: evidence for sex limited and parent of origin specific effects. *J Med Genet*. 2005;42:132-137.

22. Schulze TG, Chen YS, Badner JA, et al. Additional, physically ordered markers increase linkage signal for bipolar disorder on chromosome 18q22. *Biol Psychiatry*. 2003;53:239-243.

23. Barlow DP. Gametic imprinting in mammals. *Science*. 1995;270:1610-1613.

24. Delaval K, Wagschal A, Feil R. Epigenetic deregulation of imprinting in congenital diseases of aberrant growth. *Bioessays*. 2006;28:453-459.

25. Petronis A. The origin of schizophrenia: genetic thesis, epigenetic antithesis, and resolving synthesis. *Biol Psychiatry*. 2004;55:965-970.

26. Sutherland JE, Costa M. Epigenetics and the environment. *Ann N Y Acad Sci*. 2003;983:151-160.

27. Levenson JM, Sweatt JD. Epigenetic mechanisms: a common theme in vertebrate and invertebrate memory formation. *Cell Mol Life Sci*. 2006;63:1009-1016.

28. Lubin FD, Roth TL, Sweatt JD. Epigenetic regulation of BDNF gene transcription in the consolidation of fear memory. *J Neurosci*. 2008;28:10576-10586.

29. Liu L, van Groen T, Kadish I, Tollefsbol TO. DNA methylation impacts on learning and memory in aging. *Neurobiol Aging*. 2009;30:549-560.

30. Orlando V. Polycomb, epigenomes, and control of cell identity. *Cell*. 2003;112:599-606.

31. Kim SY, Levenson JM, Korsmeyer S, Sweatt JD, Schumacher A. Developmental regulation of Eed complex composition governs a switch in global histone modification in brain. *J Biol Chem*. 2007;282:9962-9972.

32. Levenson JM, O'Riordan KJ, Brown KD, et al. Regulation of histone acetylation during memory formation in the hippocampus. *J Biol Chem.* 2004;279:40545-40559.

33. Levenson JM, Roth TL, Lubin FD, et al. Evidence that DNA (cytosine-5) methyltransferase regulates synaptic plasticity in the hippocampus. *J Biol Chem*. 2006;281:15763-15773.

34. Yeh SH, Lin CH, Gean PW. Acetylation of nuclear factor-kappaB in rat amygdala improves long-term but not short-term retention of fear memory. *Mol Pharmacol*. 2004;65:1286-1292.

35. Ivleva E, Thaker G, Tamminga CA. Comparing genes and phenomenology in the major psychoses: schizophrenia and bipolar 1 disorder. *Schizophr Bull*. 2008;34:734-742.

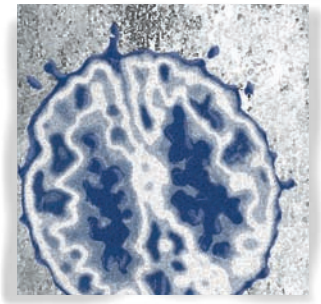
36. Andreasen NC. Symptoms, signs, and diagnosis of schizophrenia. *Lancet*. 1995;346:477-481.

37. Craddock N, O'Donovan MC, Owen MJ. The genetics of schizophrenia and bipolar disorder: dissecting psychosis. *J Med Genet*. 2005;42:193-204.
38. Craddock N, Jones I. Genetics of bipolar disorder. *J Med Genet*. 1999;36:585-594.
39. Bertelsen A, Gottesman, II. Schizoaffective psychoses: genetical clues to classification. *Am J Med Genet*. 1995;60:7-11.
40. Cardno AG, Gottesman, II. Twin studies of schizophrenia: from bow-and-arrow concordances to star wars Mx and functional genomics. *Am J Med Genet*. 2000;97:12-17.
41. O'Donovan MC, Craddock NJ, Owen MJ. Genetics of psychosis; insights from views across the genome. *Hum Genet*. 2009;126:3-12.
42. Mill J, Tang T, Kaminsky Z, et al. Epigenomic profiling reveals DNA-methylation changes associated with major psychosis. *Am J Hum Genet*. 2008;82:696-711.
43. Gupta DS, McCullumsmith RE, Beneyto M, et al. Metabotropic glutamate receptor protein expression in the prefrontal cortex and striatum in schizophrenia. *Synapse*. 2005;57:123-131.
44. Eastwood SL, Harrison PJ. Decreased expression of vesicular glutamate transporter 1 and complexin II mRNAs in schizophrenia: further evidence for a synaptic pathology affecting glutamate neurons. *Schizophr Res*. 2005;73:159-172.
45. Bruneau EG, McCullumsmith RE, Haroutunian V, Davis KL, Meador-Woodruff JH. Increased expression of glutaminase and glutamine synthetase mRNA in the thalamus in schizophrenia. *Schizophr Res*. 2005;75:27-34.
46. Miyaoka T, Seno H, Ishino H. Increased expression of Wnt-1 in schizophrenic brains. *Schizophr Res*. 1999;38:1-6.
47. Grayson DR, Jia X, Chen Y, et al. Reelin promoter hypermethylation in schizophrenia. *Proc Natl Acad Sci U S A*. 2005;102:9341-9346.
48. Tochigi M, Iwamoto K, Bundo M, et al. Methylation status of the reelin promoter region in the brain of schizophrenic patients. *Biol Psychiatry*. 2008;63:530-533.
49. Bullock WM, Cardon K, Bustillo J, Roberts RC, Perrone-Bizzozero NI. Altered expression of genes involved in GABAergic transmission and neuromodulation of granule cell activity in the cerebellum of schizophrenia patients. *Am J Psychiatry*. 2008;165:1594-1603.
50. Volk DW, Lewis DA. Impaired prefrontal inhibition in schizophrenia: relevance for cognitive dysfunction. *Physiol Behav*. 2002;77:501-505.
51. Hashimoto T, Volk DW, Eggan SM, et al. Gene expression deficits in a subclass of GABA neurons in the prefrontal cortex of subjects with schizophrenia. *J Neurosci*. 2003;23:6315-63126.
52. Huang HS, Matevossian A, Whittle C, et al. Prefrontal dysfunction in schizophrenia involves mixed-lineage leukemia 1-regulated histone methylation at GABAergic gene promoters. *J Neurosci*. 2007;27:11254-11262.
53. Sharma RP, Grayson DR, Gavin DP. Histone deactylase 1 expression is increased in the prefrontal cortex of schizophrenia subjects: analysis of the National Brain Databank microarray collection. *Schizophr Res*. 2008;98:111-117.
54. Huang HS, Akbarian S. GAD1 mRNA expression and DNA methylation in prefrontal cortex of subjects with schizophrenia. *PLoS One*. 2007;2:e809.
55. Akbarian S, Ruehl MG, Bliven E, et al. Chromatin alterations associated with down-regulated metabolic gene expression in the prefrontal cortex of subjects with schizophrenia. *Arch Gen Psychiatry*. 2005;62:829-840.
56. Veldic M, Caruncho HJ, Liu WS, et al. DNA-methyltransferase 1 mRNA is selectively overexpressed in telencephalic GABAergic interneurons of schizophrenia brains. *Proc Natl Acad Sci U S A*. 2004;101:348-353.
57. Veldic M, Kadriu B, Maloku E, et al. Epigenetic mechanisms expressed in basal ganglia GABAergic neurons differentiate schizophrenia from bipolar disorder. *Schizophr Res*. 2007;91:51-61.
58. Zhubi A, Veldic M, Puri NV, et al. An upregulation of DNA-methyltransferase 1 and 3a expressed in telencephalic GABAergic neurons of schizophrenia patients is also detected in peripheral blood lymphocytes. *Schizophr Res*. 2009;111:115-122.
59. Satta R, Maloku E, Zhubi A, et al. Nicotine decreases DNA methyltransferase 1 expression and glutamic acid decarboxylase 67 promoter methylation in GABAergic interneurons. *Proc Natl Acad Sci U S A.* 2008;105:16356-16361.
60. Lising-Enriquez K, George TP. Treatment of comorbid tobacco use in people with serious mental illness. *J Psychiatry Neurosci*. 2009;34:E1-E2.
61. McGowan PO, Kato T. Epigenetics in mood disorders. *Environ Health Prev Med*. 2008;13:16-24.
62. Kuratomi G, Iwamoto K, Bundo M, et al. Aberrant DNA methylation associated with bipolar disorder identified from discordant monozygotic twins. *Mol Psychiatry*. 2008;13:429-441.
63. Meyer J, Johannssen K, Freitag CM, et al. Rare variants of the gene encoding the potassium chloride co-transporter 3 are associated with bipolar disorder. *Int J Neuropsychopharmacol*. 2005;8:495-504.
64. Moser D, Ekawardhani S, Kumsta R, et al. Functional analysis of a potassium-chloride co-transporter 3 (SLC12A6) promoter polymorphism leading to an additional DNA methylation site. *Neuropsychopharmacology*. 2009;34:458-467.
65. Uyanik G, Elcioglu N, Penzien J, et al. Novel truncating and missense mutations of the KCC3 gene associated with Andermann syndrome. *Neurology*. 2006;66:1044-1048.
66. Sennvik K, Fastbom J, Blomberg M, et al. Levels of alpha- and beta-secretase cleaved amyloid precursor protein in the cerebrospinal fluid of Alzheimer's disease patients. *Neurosci Lett*. 2000;278:169-172.
67. Kim J, Basak JM, Holtzman DM. The role of apolipoprotein E in Alzheimer's disease. *Neuron*. 2009;63:287-303.
68. Liddell MB, Lovestone S, Owen MJ. Genetic risk of Alzheimer's disease: advising relatives. *Br J Psychiatry*. 2001;178:7-11.
69. Manzano S, Gonzalez J, Marcos A, Matias-Guiu J. [Genetics and Alzheimer's disease]. *Neurologia*. 2009;24:83-89.
70. Grunblatt E. Commonalities in the genetics of Alzheimer's disease and Parkinson's disease. *Expert Rev Neurother*. 2008;8:1865-1877.
71. Gatz M, Reynolds CA, Fratiglioni L, et al. Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry*. 2006;63:168-174.
72. Bassett SS, Avramopoulos D, Perry RT, et al. Further evidence of a maternal parent-of-origin effect on chromosome 10 in late-onset Alzheimer's disease. *Am J Med Genet B Neuropsychiatr Genet*. 2006;141B:537-540.
73. Dolinoy DC, Das R, Weidman JR, Jirtle RL. Metastable epialleles, imprinting, and the fetal origins of adult diseases. *Pediatr Res*. 2007;61:30R-37R.
74. Siegmund KD, Connor CM, Campan M, et al. DNA methylation in the human cerebral cortex is dynamically regulated throughout the life span and involves differentiated neurons. *PLoS One*. 2007;2:e895.
75. Hoyaux D, Decaestecker C, Heizmann CW, et al. S100 proteins in Corpora amylacea from normal human brain. *Brain Res*. 2000;867:280-288.
76. Wang SC, Oelze B, Schumacher A. Age-specific epigenetic drift in late-onset Alzheimer's disease. *PLoS One*. 2008;3:e2698.
77. Bjornsson HT, Sigurdsson MI, Fallin MD, et al. Intra-individual change over time in DNA methylation with familial clustering. *JAMA*. 2008;299:2877-83.
78. Bollati V, Schwartz J, Wright R, et al. Decline in genomic DNA methylation through aging in a cohort of elderly subjects. *Mech Ageing Dev*. 2009;130:234-239.
79. Kumbier E, Domes G, Herpertz-Dahlmann B, Herpertz SC. [Autism and autistic disorders: Historical and current aspects]. *Nervenarzt*. 2010;81:55-65.
80. Hallmayer J, Glasson EJ, Bower C, et al. On the twin risk in autism. *Am J Hum Genet.* 2002;71:941-946.
81. Piggot J, Shirinyan D, Shemmassian S, Vazirian S, Alarcon M. Neural systems approaches to the neurogenetics of autism spectrum disorders. *Neuroscience*. 2009;164:247-256.
82. Crespi B. Genomic imprinting in the development and evolution of psychotic spectrum conditions. *Biol Rev Camb Philos Soc*. 2008;83:441-493.
83. Yusufzai TM, Wolffe AP. Functional consequences of Rett syndrome mutations on human MeCP2. *Nucleic Acids Res*. 2000;28:4172-4179.
84. Young JI, Hong EP, Castle JC, et al. Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. *Proc Natl Acad Sci U S A.* 2005;102:17551-17558.
85. Kriaucionis S, Bird A. DNA methylation and Rett syndrome. *Hum Mol Genet*. 2003;12 Spec No 2:R221-R227.
86. Guy J, Hendrich B, Holmes M, Martin JE, Bird A. A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. *Nat Genet*. 2001;27:322-326.

87. Collins AL, Levenson JM, Vilaythong AP, et al. Mild overexpression of MeCP2 causes a progressive neurological disorder in mice. *Hum Mol Genet*. 2004;13:2679-2689.
88. Shahbazian M, Young J, Yuva-Paylor L, et al. Mice with truncated MeCP2 recapitulate many Rett syndrome features and display hyperacetylation of histone H3. *Neuron*. 2002;35:243-254.
89. Maezawa I, Swanberg S, Harvey D, LaSalle JM, Jin LW. Rett syndrome astrocytes are abnormal and spread MeCP2 deficiency through gap junctions. *J Neurosci*. 2009;29:5051-5061.
90. Nagarajan RP, Hogart AR, Gwye Y, Martin MR, LaSalle JM. Reduced MeCP2 expression is frequent in autism frontal cortex and correlates with aberrant MECP2 promoter methylation. *Epigenetics*. 2006;1:e1-e11.
91. Nagarajan RP, Patzel KA, Martin M, et al. MECP2 promoter methylation and X chromosome inactivation in autism. *Autism Res*. 2008;1:169-178.
92. Allan AM, Liang X, Luo Y, et al. The loss of methyl-CpG binding protein 1 leads to autism-like behavioral deficits. *Hum Mol Genet*. 2008;17:2047-2057.
93. Adegbola A, Gao H, Sommer S, Browning M. A novel mutation in JARID1C/SMCX in a patient with autism spectrum disorder (ASD). *Am J Med Genet A.* 2008;146A:505-511.
94. Curran LK, Newschaffer CJ, Lee LC, et al. Behaviors associated with fever in children with autism spectrum disorders. *Pediatrics*. 2007;120:e1386-e1392.
95. Plagge A, Isles AR, Gordon E, et al. Imprinted Nesp55 influences behavioral reactivity to novel environments. *Mol Cell Biol*. 2005;25:3019-3026.
96. Mehler MF, Purpura DP. Autism, fever, epigenetics and the locus coeruleus. *Brain Res Rev*. 2009;59:388-392.
97. Kundakovic M, Chen Y, Costa E, Grayson DR. DNA methyltransferase inhibitors coordinately induce expression of the human reelin and glutamic acid decarboxylase 67 genes. *Mol Pharmacol*. 2007;71:644-653.
98. Dash PK, Orsi SA, Moore AN. Histone deactylase inhibition combined with behavioral therapy enhances learning and memory following traumatic brain injury. *Neuroscience*. 2009;213:47-59.
99. Hockly E, Richon VM, Woodman B, et al. Suberoylanilide hydroxamic acid, a histone deacetylase inhibitor, ameliorates motor deficits in a mouse model of Huntington's disease. *Proc Natl Acad Sci U S A*. 2003;100:2041-2046.
100. Camelo S, Iglesias AH, Hwang D, et al. Transcriptional therapy with the histone deacetylase inhibitor trichostatin A ameliorates experimental autoimmune encephalomyelitis. *J Neuroimmunol*. 2005;164:10-21.
101. Chen PS, Peng GS, Li G, et al. Valproate protects dopaminergic neurons in midbrain neuron/glia cultures by stimulating the release of neurotrophic factors from astrocytes. *Mol Psychiatry*. 2006;11:1116-1125.
102. Dagtas AS, Edens ER, Gilbert KM. Histone deacetylase inhibitor uses p21(Cip1) to maintain anergy in CD4(+) T cells. *Int Immunopharmacol*. 2009;9:1289-1297.
103. Langley B, D'Annibale MA, Suh K, et al. Pulse inhibition of histone deacetylases induces complete resistance to oxidative death in cortical neurons without toxicity and reveals a role for cytoplasmic p21(waf1/cip1) in cell cycle-independent neuroprotection. *J Neurosci*. 2008;28:163-176.
104. Gilad R, Izkovitz N, Dabby R, et al. Treatment of status epilepticus and acute repetitive seizures with i.v. valproic acid vs phenytoin. *Acta Neurol Scand*. 2008;118:296-300.
105. Bowden CL. Spectrum of effectiveness of valproate in neuropsychiatry. *Exp Rev Neurother*. 2007;7:9-16.
106. Sajatovic M, Coconcea N, Ignacio RV, et al. Adjunct extended-release valproate semisodium in late life schizophrenia. *Int J Geriatr Psychiatry*. 2008;23:142-147.
107. Hajduk PJ, Greer J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat Rev Drug Discov*. 2007;6211-219.
108. Strausberg RL, Schreiber SL. From knowing to controlling: a path from genomics to drugs using small molecule probes. *Science*. 2003;300:294-295.
109. Kubicek S, O'Sullivan RJ, August EM, et al. Reversal of H3K9me2 by a small-molecule inhibitor for the G9a histone methyltransferase. *Mol Cell*. 2007;25:473-481.
110. Waggoner D. Mechanisms of disease: epigenesis. *Semin Pediatr Neurol*. 2007;14:7-14.
111. Pushparaj PN, Melendez AJ. Short interfering RNA (siRNA) as a novel therapeutic. *Clin Exp Pharmacol Physiol*. 2006;33:504-510.
112. Bernstein E, Allis CD. RNA meets chromatin. *Genes Dev*. 2005;19:1635-1655.
113. Jamieson AC, Miller JC, Pabo CO. Drug discovery with engineered zinc-finger proteins. *Nat Rev Drug Discov*. 2003;2:361-368.
114. Xu GL, Bestor TH. Cytosine methylation targetted to pre-determined sequences. *Nat Genet*. 1997;17:376-378.
115. Roelfsema JH, White SJ, Ariyurek Y, et al. Genetic heterogeneity in Rubinstein-Taybi syndrome: mutations in both the CBP and EP300 genes cause disease. *Am J Hum Genet.* 2005;76:572-580.
116. Philibert RA, Gunter TD, Beach SR, Brody GH, Madan A. MAOA methylation is associated with nicotine and alcohol dependence in women. *Am J Med Genet B Neuropsychiatr Genet*. 2008;147B:565-570.
117. Bonsch D, Lenz B, Fiszer R, et al. Lowered DNA methyltransferase (DNMT-3b) mRNA expression is associated with genomic DNA hypermethylation in patients with chronic alcoholism. *J Neural Transm*. 2006;113:1299-1304.
118. Stack EC, Del Signore SJ, Luthi-Carter R, et al. Modulation of nucleosome dynamics in Huntington's disease. *Hum Mol Genet*. 2007;16:1164-1175.
119. Warren ST. The Epigenetics of Fragile X Syndrome. *Cell Stem Cell*. 2007;1:488-489.
120. Schilling E, El Chartouni C, Rehli M. Allele-specific DNA methylation in mouse strains is mainly determined by cis-acting sequences. *Genome Res*. 2009;19:2028-2035.

Whole genome association studies in complex diseases: where do we stand?

Anna C. Need, PhD; David B. Goldstein, PhD



Hundreds of genome-wide association studies have been performed in recent years in order to try to identify common variants that associate with complex disease. These have met with varying success. Some of the strongest effects of common variants have been found in late-onset diseases and in drug response. The major histocompatibility complex has also shown very strong association with a variety of disorders. Although there have been some notable success stories in neuropsychiatric genetics, on the whole, common variation has explained little of the high heritability of these traits. In contrast, early studies of rare copy number variants have led rapidly to a number of genes and loci that strongly associate with neuropsychiatric disorders. It is likely that the use of whole-genome sequencing to extend the study of rare variation in neuropsychiatry will greatly advance our understanding of neuropsychiatric genetics.

 Dialogues Clin Neurosci. 2010;12:37-46.

Keywords: *genome-wide association study; rare variant; neuropsychiatric; schizophrenia; sequencing*

Author affiliations: Institute for Genome Sciences and Policy, Center for Human Genome Variation, Duke University, Durham, North Carolina, USA

In recent years, hundreds of genetic association studies have sought to explore the relationship between common genetic variation and disease, biological characteristics, or drug response. The basic premise of these studies is that the diseases (or traits) are not caused by single gene variants of strong effect, such as, for instance, sickle-cell anemia or cystic fibrosis, but rather that some "manageable" number of common variants have an important influence on the trait under question. Part of the motivation for this perspective is the "common disease, common variant" (CDCV) theory.[1,2] Once a genetic variant has been found to be associated, there are a number of possible uses for the information. If the effect of the genetic variant is strong enough, perhaps in combination with lifestyle or other environmental factors, it might be used to predict risk of the disease. Alternatively, the associated variant(s) may be used to try to predict response to a particular medication. Finally, if the effect size of the genetic variant is very small and thus not useful for either of these purposes, it may still be of use in identifying a disease-associated gene or genetic pathway that could illuminate disease pathophysiology or implicate new therapeutic targets. Here we review the current status of genome-wide association studies, with a particular focus on neuropsychiatric disorders.

Genome-wide association studies

Genome-wide association studies (GWAS), are a way of performing genetic association studies without prior

Address for correspondence: David B. Goldstein, Institute for Genome Sciences and Policy, Center for Human Genome Variation, Duke University, 450 Research Drive, Durham, NC 27708, USA
(e-mail: d.goldstein@duke.edu)

hypotheses about which genes are likely to be involved. To do this, arrays of single-nucleotide polymorphisms (SNPs) that cover the whole genome are used. Although there are thought to be approximately 10 million common SNPs in the genome,[3] it is not necessary to genotype each one of these individually to get information about most of them. This is because, due to the way that human populations have migrated and genetic variants have arisen, many of the variants are associated with each other or "linked." Thus, in European and Asian populations, if you genotype one variant, you are gaining information about 10 to 20 other variants simultaneously. This is called "tagging" (the genotyped variants "tag" the ungenotyped, linked variants), and was brought to the genome-wide scale by the HapMap project, which has genotyped millions of common SNPs in four populations to create a detailed map of how common genetic variants relate to one another.[3-5] A significant motivation for the HapMap project was the idea that common variants make up an important part of the genetic contribution to common diseases (the CDCV hypothesis). While some theoretical arguments were marshaled in support of this hypothesis—and indeed, even before the HapMap project a handful of examples were known—there was no way to know a priori how general the CDCV hypothesis might turn out to be. For this reason, a systematic investigation of common variation was judged by much of the community (including these authors) but not all[6] to be a sensible beginning to the study of human disease genetics. The result is that a true genome-wide study can be performed by actually genotyping as few as 300 000 to 1 million SNPs.[7,8] However, because so many tests are being performed, it is necessary to obtain a very strongly significant *P* value to be sure that the result is really significant. This is known as "genome-wide significance" and the consensus is that this should be about 10^{-8} or less.[9] Because the effects sizes of common variants are generally small, it is usually necessary to include a large number of subjects in the study in order to have the power to detect a genome-wide significant *P* value *(Figure 1)*.

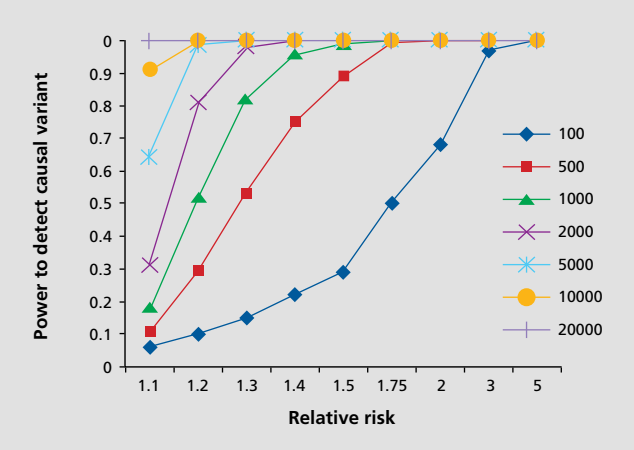


Figure 1. The power to detect a causal variant that is perfectly tagged by a genotyped marker (assuming dominant model, minor allele frequency=0.2, frequency of disease is 1% and equal numbers of cases and controls). To have a good chance of detecting a variant with a relative risk of 1.2, about 2000 cases and controls are needed.

Major discoveries with GWAS

The success of GWAS has been very variable for different disease areas. Some diseases have found common variants with very strong effects, and managed to track these down to the causal variant. An inspiring example is an intronic variant in *BCL11A* that was found in two GWAS studies to associate with fetal hemoglobin (HbF) levels in healthy adults,[10,11] and also to modify the presentation of β-thalassemia, and associate with HbF levels in patients with sickle-cell disease.[11] This finding was soon followed up with a functional study that showed that the variant associated with high HbF[12] reduced the expression of *BCL11A*,[13] and that reduction of *BCL11A* expression caused increase in levels of gamma-globin in adult human red blood progenitor cells, which led to increased levels of HbF.[13] These findings clearly suggest that *BCL11A* serves as an inhibitor of HbF production and that directed repression of *BCL11A* could be developed as a clinical tool to ameliorate the presentation of thalassemias and sickle-cell disease. These findings in turn have led to further understanding of developmental and species-specific changes in globin regulation.[14]

On the less inspirational side, however, other diseases, like hypertension, have been thoroughly and carefully investigated using huge numbers of patients and controls with very little progress.[15] Here we outline some of the highest impact findings of GWAS and where (if anywhere) they have led us.

As might be expected by the laws of natural selection, there are not many common genetic variants that confer a strong predisposition to common diseases. Such variants would be expected to have been selected against, and thus maintained at low population frequen-

cies. However, there are some phenotypes that might be expected to have dodged the purifying effects of selection. These include common diseases that do not onset until old age, and response to drugs that the body has not historically had to interact with. Accordingly, some of the strongest effects of common variants on disease have been found in association with ailments with an onset during the postreproductive years, and with drug response.

Genetic variants that affect late-onset diseases

One of the most well-known genetic risk factors is the E4 variant of the apolipoprotein E gene, *ApoE*, which greatly increases the risk of Alzheimer's disease (AD) and reduces the age of onset in a dose-dependent manner.[16-18] The effect of this variant is so strong that it was, in fact, discovered before the GWAS era, but it has since been confirmed as the most important predictor of late-onset AD in a number of genome-wide analyses,[19-22] one with fewer than 500 cases and controls reporting a *P* value of 1×10^{-40}.[21] However, despite the definitive effects of this genetic variant on AD and the length of time that we have known about it, it is still not clear how the variant mediates its effects,[23] and it has not yet led to improved treatment.

One of the very earliest novel discoveries of GWAS was the association of an amino acid substitution in the complement factor H gene, *CFH*, with age-related macular degeneration, a very common form of blindness that affects the elderly. This genetic association was found with a tiny sample size: 96 cases and 50 controls, and carrying two copies of the risk variant increases the risk of illness up to 7 times.[24] The associated variant does itself seem to be functional, changing the binding properties of the protein, although it is not yet exactly understood how the variant contributes to disease,[25] nor how this can be utilized in novel treatments.

A third very strong disease-associated common genetic variant is in the *LOXL1* gene in exfoliation glaucoma, another very common form of age-related blindness. The associated variant was discovered in a set of only 75 cases, and individuals homozygous for the risk haplotypes are thought to be at 700-fold increased risk of exfoliation glaucoma when compared with homozygotes of the low-risk haplotype. However, because the risk haplotype is so common, this translates to just a 2.5-fold increase risk from the population average.[26] The two variants contributing to the risk haplotype are both protein-coding changes, and the same variants have now been associated with disease in multiple populations,[27-40] suggesting that these are the causal variants, although the degree of penetrance, and the risk haplotype, have been reported to differ in Australia and Japan.[28,29,35,37,38,41,42] Unfortunately, the very high frequency of the risk haplotype in the general population currently precludes these markers from being used to predict disease, but it is hoped that a better understanding of the role of *LOXL1* in optical pathophysiology may lead to advances in treatment.[40]

Genetic variants that affect drug response

Genetic variants affecting drug response can have very strong effects, and often occur in the genes that would be most expected to be involved.[43] Thus, pharmacogenetics was one of the more successful areas of genomics before the GWAS area, and a number of strong genetic influencers of drug response have been known for some time.[44] GWAS have added at least three pharmacogenetic associations of considerable strength and importance.

Flucloxacillin-induced liver injury

Idiosyncratic drug reactions are the most common cause of liver failure in the US.[45] Flucloxacillin is an antibiotic drug commonly used to treat *Staphylococcus aureus* infections, but it has a relatively high incidence of causing liver injury (6.1 per 100 000 users) in comparison with other antibiotics such as penicillin.[46] This has previously led to restrictions on its use.[46] A GWAS was performed on 51 patients with flucloxacillin-induced liver injury and 487 controls, in which a huge signal was seen for a missense polymorphism in the *HCP5* gene ($P = 8.7\ 10^{-33}$).[47] Through linkage disequilibrium, the association was traced to the *HLA-B*5701* allele, the presence of which increased the likelihood of flucloxacillin-induced liver injury by 80 times.[47] Since the general frequency of the associated allele in the European population is only about 5%, and it was present in 84% of cases, this variant could potentially be used to screen out people at high risk of liver injury before flucloxacillin is prescribed. However, due to the rarity of the hepatoxicity, this would result in a high false-positive rate. A proposed alternative is to use the genotyping of this variant as a diagnos-

tic marker in suspected cases of hepatoxicity so that the patient can be rapidly switched to alternative antibiotics.[47]

Statin-induced myopathy

Taking statin therapy to reduce the levels of low-density lipoprotein cholesterol has been shown to reduce the likelihood of cardiovascular events, such as heart attack and stroke.[48] Occasionally, however, statins, particularly at high doses, can cause serious myopathy, which may lead to hospitalization or death.[49] In August 2008 a GWAS that included only 85 cases and 90 controls revealed a SNP in the *SLCO1B1* gene, which accounted for more than 60% of cases of myopathy.[50] Carrying one C at this locus increases the risk of statin-induced myopathy by 4.5 times, and CC homozygotes have a 17-fold greater risk than TT homozygotes. This has been suggested as a genetic test to identify vulnerable individuals before offering high-dose simvastatin therapy.[51]

Hepatitis-C treatment response

One of the most recent, and perhaps the most clinically significant, of any GWAS to date is the association of a SNP close to the *IL28B* gene with response to treatment for hepatitis C.[52] In this study, Ge et al focused on who is cured by treatment, and found that the good response genotype is associated with a greater than 80% chance of clearance in European-Americans, while the poor response genotype is associated with only about a 30% chance. A follow-up study found that the polymorphism also influences natural clearance of hepatitis C and shows very sharp geographic differentiation.[53] This suggests that the variant may be common in the population because the "good response" allele conferred protection against one or more viruses and hence was positively selected. This variant is a very good candidate to use as a pharmacogenetic predictor of treatment response before beginning hepatitis C treatment, since the procedure is long and often associated with adverse effects.[54]

The major histocompatibility complex

Setting aside the old-age or pharmacogenetic associations, many of the strongest reported GWAS associations of common variants with common disease involve markers in the major histocompatibility complex (MHC). These associations are too extensive to discuss in detail in this review, but include autoimmune diseases, infectious diseases, neuropsychiatric disorders, and variability in normal traits such as height.[55] A number of hypotheses have been put forward to explain why variants conferring disease risk at this locus have been maintained at high frequency in the population. One suggestion is that the disease-associated variants have been selected for because they confer resistance to particular infectious agents, either now or historically. An alternative hypothesis is that each locus that confers risk for one common disease is maintained at high frequency because it confers protection against one or more other common diseases. For example, the HLA gene *DQB1*0602*, which encodes the β chain for the HLA class II molecule DQ6, is protective against diabetes,[56] but a strong risk factor for narcolepsy[57] and multiple sclerosis.[58]

GWAS in neuropsychiatry

Neuropsychiatric traits have been among the most disappointing GWAS results. Despite many GWAS, most associated variants have either not withstood significance correction for multiple testing, or else have failed to replicate. In general, where replicable effects have been found, they have required very large sample sizes and the effects have been small.

There have been some notable success stories, however. Two GWAS have revealed strong and replicable genetic influences on restless legs syndrome (RLS), a condition characterized by an unpleasant and irresistible urge to move the legs, particularly while resting and during the evening and night. Both studies, one on Icelandic individuals and one on a more mixed European cohort, implicated *BTBD9*.[59,60] The European study also found an association with two other loci: MEIS1 and a locus encompassing *MAP2K5/LBXCOR1*.[60] The associations with *MEIS1* and *BTBD9* were quickly replicated in two subsequent studies,[61,62] but the *MAP2K5/LBXCOR1* appears to be weaker, showing a borderline significance in one study only.[62] Although the risk associated with *MEIS1* and *BTBD9* (ranging from 1.5 to 3.7[59,60,62,63]) is substantially lower than those described above, they do appear to be real and highly significant risk factors for RLS. Nevertheless, the biology underlying the associations remains unclear. The associated variants do not

appear to have any obvious function, and a thorough search for putative functional variants in all coding exons and across intron-exon boundaries revealed no obviously causal variant.[64]

Another positive GWAS finding in neuropsychiatry is with narcolepsy, a disorder that causes disrupted sleep patterns, with the patient often feeling excessively tired during the day, and suffering sudden sleep attacks. Pre-GWAS studies had connected the disorder to an MHC class II antigen called *HLA-DQB1*0602*, and about 85% of narcoleptics carry this antigen.[65] However, there remained unexplained heritability. Very recently, a GWAS study was done on 807 cases and 1074 controls, all positive for *HLA-DQB1*0602*. A significant association of three SNPs in the T cell receptor alpha locus was found, which was then replicated in the same study in 1057 further cases and 1104 controls.[66] Further analysis showed a single SNP was responsible for the association, although it is not clear whether this variant is itself causal or how it may contribute to disease. This association is of particular interest because it adds considerable weight to the view that narcolepsy is an autoimmune disease, and as such, it would be the first autoimmune disease to be associated with a T-cell receptor locus. This finding also opens up the possibility of immunotherapy as a future treatment for narcolepsy.

Other neuropsychiatric diseases for which definite, replicated effects of common SNPs have been found include schizophrenia, associated with MHC markers, *NRGN* and *TCF4* (12 945 cases and 34 591 controls, ORs=1.24, 1.15,1.23),[67,68] bipolar disorder, associated with *ANK3* and *CACNA1C* (4 387 cases and 6 209 controls, ORs=1.45 and 1.18)[69], and autism, associated with SNPs at 5p14.1 (3 101 family members, 204 cases and 6 941 controls, OR=1.19).[70,71] However, all of these were discovered with very large sample sizes and account for very little of the very high heritability of these conditions.

Rare variants

Although studies of common variation in neuropsychiatric disease may be underwhelming, the opposite is true for rare variation. Although the SNP chips used for GWAS comprise only polymorphisms that are reasonably common (~≥5%), their data can be used to find other types of non-SNP variants—specifically copy number variants (CNVs)—with much lower frequency. CNVs are duplications or deletions of large stretches of DNA—ranging in size from just a few hundred base pairs to many megabases. To detect such variants, the intensity data from the SNP chips is examined to determine whether particular stretches of SNPs are less intense than expected (or absent), which would indicate a deletion, or more intense than expected, which suggests a duplication.[72] Because the CNVs are identified on an individual-by-individual basis, very rare CNVs, even those present in a single individual, can be found. This has allowed us for the first time to examine the role of rare variation in common disease (albeit just a tiny fraction of the total amount of rare variant in a cohort). The majority of investigations of copy number variation to date have been in neuropsychiatric disease and, happily, they have led immediately to real, replicable and very strong associations. A summary of CNVs recently strongly associated with neuropsychiatric disease is shown in *Table I*.

These variants confer considerable risk, but they are not completely penetrant. Although the specific variants are very rare in the general population, they are occasionally seen in controls *(Table I)*, and where families have been examined, the variants are often inherited from unaffected or only mildly affected parents.[73-77] Additionally, as can be seen in *Table I*, many of the variants have been associated with more than one neuropsychiatric condition. This is consistent with the characteristics of neuropsychiatrically-associated rare variants that were found before the GWAS era, such as *DISC1* in schizophrenia, which associated with a range of phenotypes from psychiatrically normal to suicide, recurrent major depression, and schizophrenia.[78] It seems that these variants, rather than predisposing to a specific neuropsychiatric condition, may strongly confer some sort of "neural vulnerability," the ultimate manifestation of which depends on other interacting genetic and environmental factors. Because, to date, the only rare variants that we have been able to associate with neuropsychiatric illness are very large deletions and duplications, it is not clear whether this lack of specificity will be a general rule, or is somehow related to the size of the lesion. However, there is some evidence from the associations with common SNPs that this is a characteristic of the disease rather than the size of the associated variant. For instance, bipolar-associated common variants in *CACNA1C* may also confer risk of depression and schizophrenia.[79]

The future for neuropsychiatric genetics

There are two, not incompatible, possible directions for neuropsychiatric genetics research. One approach is to continue searching for common variants of small effect size using much larger cohorts in the tens or hundreds of thousands. This has been suggested as a future direction for schizophrenia genetics.[80] Although this will require a considerable effort, there are already established worldwide collaborations for schizophrenia,[68,80] so very large collections should be achievable in the relatively near future. The disadvantages of this approach are that if such huge sample sizes are needed to discover them, the effect sizes of the associated variants must be very small *(Figure 1)*, and they will be present at a similar frequency in unaffected controls. This makes further study of the effects of the variants very difficult or impossible. However, proponents of this approach correctly suggest that although the associated variant may have a very small effect, the gene it is in may have a big impact on disease when targeted by novel pharmaceuticals.

A second argument in favor of proceeding with GWAS in very large samples is that neuropsychiatric researchers have long expressed concern that clinical diagnostic criteria do not reflect the biological underpinnings of the disease, and that diseases such as schizophrenia may in fact represent multiple different disorders with different genetic contributors. Thus, only with very large sample sizes would one expect to obtain sufficient numbers of any one genetically homogenous subgroup to obtain a genome-wide significant association. However, as discussed above, all genetic variants that have been associated with neuropsychiatric disease so far seem to be very nonspecific. Where they are found in multiple patients

CNV	Copy # linked to disease	Seen in schizophrenia patients?	Seen in autism patients?	Seen in epilepsy patients?	Seen in patients with mental retardation?	Other disorders	Reported in controls*?	Lead candidate genes
1q21.1	Deletion and duplication	Deletion (0.23-0.29%)[82-84]	Yes[75,94]	Yes[77,83,95,] unpublished data	Yes[95,96]	Congenital heart disease[77,97,98], micro- and macrocephaly,[75,77] neuroblastoma,[99] other[75]	Deletion, 0.02% (8/41,199),[84] frequency for dup unclear	*HYDIN* paralog[77]; *GJA8*[84]
15q11	Deletion	Yes (0.61%)[84,100]	Yes[101]	No	Yes[102]	Deletions of this region cause Angelman and Prader Willi syndrome	Yes, (0.19%) (79/41,194) (0.19%)	*CYFIP1*[84]
15q13.1	Duplication and deletion	Yes[81,83,84, 103]	NR	NR	NR	NR	no data	APBA2
15q13.3	Deletion	Yes (0.17%-0.27%)[83,84]	Yes (0.31%)[84,86,104]	Yes (1-1.3%)[73,74,110]	Yes[73,106]	Various including mild developmental delay, heart defects[73]	Yes, 0.02% (8/39,800)[74,84]	*CHRNA7*[83,84]
16p11.2	Deletion and duplication	Yes[94]	Yes (0.6%, del only,[107] 1% dup + del[94])	Yes[110,111]	Yes[76,108]	Various neuropsychiatric and developmental[76,108]	Yes, 0.01% del, 0,03% dup[94]	*SEZ6L2*[110,111]
16p13.11	Deletion	Yes[81]	Yes[112]	Yes (0.6%) unpublished data	Yes (0.5%)[87,111]	NR	NR (0/3313),[87,96] unpublished data but inherited from unaffected parents[87]	*NDEL*[81]
CNTN4	Deletion and duplication	NR	Yes[85,114,115]	NR	NR	NR	no data	
NRXN1	Deletion	Yes (0.19%)[81-83,104,120-122]	Yes[85,116-118]	NR	Yes[122,123]	NR	Yes, 0.04% (17/42054)[115]	

Table I. Copy number variants (CNV) strongly associated with neuropsychiatric disorders.. Frequencies are given only when the CNV was found in a large case-control study design. *Controls may not have been carefully screened for neuropsychiatric illness. NR, not reported; Dup, duplication; Del, Deletion

with a single diagnosis (eg, schizophrenia), they do not segregate patients into any clear diagnostic categories either by disease presentation or drug response. Additionally, they tend to associate with multiple neuropsychiatric conditions *(Table I)*.

The alternative approach is to further investigate the role of rare variants in neuropsychiatric disease. To date, the only type of rare variation that has been identifiable on a genome-wide scale has been large CNVs, and already we have found many strong associations.[81-87] It is likely that when we can identify the totality of rare variation in an individual using whole-genome sequencing, many more rare variants will be found to be definitely associated with neuropsychiatric illness. Fortunately, this is rapidly becoming a reality, and the first sequencing studies in neuropsychiatric illness are already underway. For confirmation and follow-up, this approach will definitely benefit from very large cohorts collected for GWAS, but the ideal discovery samples will be rather different. With this approach, we hope to find variants with very large effect sizes and high penetrance. This means that it will be much more straightforward to understand how the variants exert their effects and what genetic and environmental factors influence them. To do this, the priority will be patients and relatives that can be reapproached for further study after potentially causal variants have been identified. Additionally, since initial sequencing attempts will be expensive, it is worth, at first at least, selecting patients who are most likely to carry highly penetrant genetic variants. These include severely ill, treatment-resistant patients[88] and patients with a strong family history of mental illness. Thus, this approach benefits from close collaboration between geneticists and psychiatrists and a thorough understanding of each sequenced patient and his or her relatives.

Although it is hoped that whole-genome sequencing will lead swiftly to a clearer understanding of neuropsychiatric disease, there are many challenges ahead. Not least is a very well-characterized psychiatrically normal control cohort. And, as with any new technology, there are considerable technical challenges, such as the use of whole-genome data to identify copy number variation. However, software is constantly developing and it is doubtful that these will be limiting factors for long.[89-92] There are also "genomic" challenges: there are many regions of the genome on which we tend not to focus, such as remote enhancer regions, upstream open reading frames, and chromatin binding sites, which are likely to be functional and affected by rare variation. However, using Mendelian diseases as a model, it is reasonable to expect that many of the most important variants will be in or very close to exons.[93] Thus, neuropsychiatric geneticists should be able to gorge themselves on the low-hanging fruit for some time to come.

In summary, there have been many GWAS success stories in which common variants have been found to associate definitely with complex diseases. In most cases, however, the mechanism underlying the association is not well understood, and they have not yet led to strong predictive tests or to novel treatments. Neuropsychiatric disease, in particular, has so far benefited little from large-scale analysis of common variants. Use of GWAS data to examine rare copy number variants, however, rapidly led to multiple strong and highly penetrant associations with neuropsychiatric illness. However, the associated variants are not completely penetrant and tend to be associated with multiple neuropsychiatric conditions. Detailed studies of patients and their relatives will be necessary to understand what factors affect the manifestation of the phenotype. Despite this recent success, we can still only account for a very small amount of the heritability of neuropsychiatric conditions. Further investigation of rare variation using whole-genome sequencing is likely to significantly advance the field.

REFERENCES

1. Chakravarti A. Population genetics--making sense out of sequence. *Nat Genet*. 1999;21(1 suppl):56-60.
2. Reich DE, Lander ES. On the allelic spectrum of human disease. *Trends Genet*. 2001;17:502-510.
3. Frazer KA, Ballinger DG, Cox DR, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature*. 2007;449:851-861.
4. A haplotype map of the human genome. *Nature*. 2005;437:1299-1320.
5. The International HapMap Project. *Nature*. 2003;426:789-796.
6. Terwilliger JD, Hiekkalinna T. An utter refutation of the "fundamental theorem of the HapMap". *Eur J Hum Genet*. 2006;14:426-437.
7. Need AC, Goldstein DB. Genome-wide tagging for everyone. *Nat Genet*. 2006;38:1227-1228.
8. Johnson GC, Esposito L, Barratt BJ, et al. Haplotype tagging for the identification of common disease genes. *Nat Genet*. 2001;29:233.
9. McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet*. 2008;9:356-369.
10. Menzel S, Garner C, Gut I, et al. A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. *Nat Genet*. 2007;39:1197-1199.
11. Uda M, Galanello R, Sanna S, et al. Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. *Proc Natl Acad Sci U S A*. 2008;105:1620-1625.

Estudios de asociación del genoma completo en enfermedades complejas: ¿dónde estamos?

En los últimos años se han realizado cientos de estudios de asociación del genoma completo tratando de identificar variantes comunes que se asocien con enfermedades complejas, los que han tenido logros variables. En enfermedades de aparición tardía y en la respuesta a fármacos se han encontrado algunos de los efectos más potentes de variantes comunes. El complejo mayor de histocompatibilidad también ha mostrado una asociación muy fuerte con una variedad de trastornos. Aunque han existido algunos casos destacados de éxito en la genética neuropsiquiátrica, en conjunto, la variación común ha explicado sólo parte de la alta herencia de estos rasgos. Por otra parte, los estudios iniciales de variantes raras del número de la copia han conducido rápidamente a asociaciones potentes entre un número de genes y loci con trastornos neuropsiquiátricos. Es posible que el empleo de la secuenciación de todo el genoma se extienda al estudio de variaciones raras en neuropsiquiatría y se progrese enormemente en la comprensión de la genética neuropsiquiátrica.

Les études d'association sur le génome entier dans les maladies complexes : où en sommes-nous ?

Ces dernières années, des centaines d'études d'association sur le génome entier ont tenté d'identifier des variants communs associés aux maladies complexes, ceci avec un succès mitigé. Certains des effets les plus marqués des variants communs ont été retrouvés dans les maladies à début tardif et dans la réponse au médicament. Le complexe majeur d'histocompatibilité a montré également une très forte association avec différents troubles. Malgré quelques succès notables en génétique neuropsychiatrique, dans l'ensemble, la très haute héritabilité de ces caractères a été peu expliquée par les variants communs. Au contraire, les premières études de variations rares du nombre de copies ont permis rapidement d'affirmer une forte association de nombreux gènes et loci à des maladies neuropsychiatriques. Il est probable que l'utilisation du séquençage du génome entier pour améliorer l'étude des variations rares en neuropsychiatrie va permettre de faire avancer de manière significative notre compréhension de la génétique neuropsychiatrique.

12. Lettre G, Sankaran VG, Bezerra MA, et al. DNA polymorphisms at the BCL11A, HBS1L-MYB, and beta-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. *Proc Natl Acad Sci U S A*. 2008;105:11869-11874.

13. Sankaran VG, Menne TF, Xu J, et al. Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. *Science*. 2008;322:1839-1842.

14. Sankaran VG, Xu J, Ragoczy T, et al. Developmental and species-divergent globin switching are driven by BCL11A. *Nature*. 2009;460:1093-1097.

15. Charchar F, Zimmerli L, Tomaszewski M. The pressure of finding human hypertension genes: new tools, old dilemmas. *J Hum Hypertens*. 2008;22:821-828.

16. Strittmatter WJ, Saunders AM, Schmechel D, et al. Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proc Natl Acad Sci U S A*. 1993;90:1977-1981.

17. Corder EH, Saunders AM, Strittmatter WJ, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science*. 1993;261:921-923.

18. Tsai MS, Tangalos EG, Petersen RC, et al. Apolipoprotein E: risk factor for Alzheimer disease. *Am J Hum Genet*. 1994;54:643-649.

19. Coon KD, Myers AJ, Craig DW, et al. A high-density whole-genome association study reveals that APOE is the major susceptibility gene for sporadic late-onset Alzheimer's disease. *J Clin Psychiatry*. 2007;68:613-618.

20. Bertram L, Lange C, Mullin K, et al. Genome-wide association analysis reveals putative Alzheimer's disease susceptibility loci in addition to APOE. *Am J Hum Genet*. 2008;83:623-632.

21. Feulner TM, Laws SM, Friedrich P, et al. Examination of the current top candidate genes for AD in a genome-wide association study. *Mol Psychiatry*. 2009 Jan 6. Epub ahead of print.

22. Abraham R, Moskvina V, Sims R, et al. A genome-wide association study for late-onset Alzheimer's disease using DNA pooling. *BMC Med Genomics*. 2008;1:44.

23. Kim J, Basak JM, Holtzman DM. The role of apolipoprotein E in Alzheimer's disease. *Neuron*. 2009;63:287-303.

24. Klein RJ, Zeiss C, Chew EY, et al. Complement factor H polymorphism in age-related macular degeneration. *Science*. 2005;308:385-389.

25. Boon CJ, van de Kar NC, Klevering BJ, et al. The spectrum of phenotypes caused by variants in the CFH gene. *Mol Immunol*. 2009;46:1573-1594.

26. Thorleifsson G, Magnusson KP, Sulem P, et al. Common sequence variants in the LOXL1 gene confer susceptibility to exfoliation glaucoma. *Science*. 2007;317:1397-1400.

27. Fingert JH, Alward WL, Kwon YH, et al. LOXL1 mutations are associated with exfoliation syndrome in patients from the midwestern United States. *Am J Ophthalmol*. 2007;144:974-975.

28. Hewitt AW, Sharma S, Burdon KP, et al. Ancestral LOXL1 variants are associated with pseudoexfoliation in Caucasian Australians but with markedly lower penetrance than in Nordic people. *Hum Mol Genet*. 2008;17:710-716.

29. Hayashi H, Gotoh N, Ueda Y, Nakanishi H, Yoshimura N. Lysyl oxidase-like 1 polymorphisms and exfoliation syndrome in the Japanese population. *Am J Ophthalmol*. 2008;145:582-585.

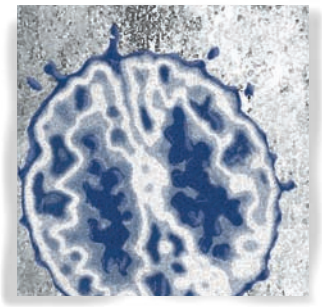
30. Fan BJ, Pasquale L, Grosskreutz CL, et al. DNA sequence variants in the LOXL1 gene are associated with pseudoexfoliation glaucoma in a U.S. clinic-based population with broad ethnic diversity. *BMC Med Genet*. 2008;9:5.

31. Yang X, Zabriskie NA, Hau VS, et al. Genetic association of LOXL1 gene variants and exfoliation glaucoma in a Utah cohort. *Cell Cycle (Georgetown, Tex)*. 2008;7:521-524.
32. Ramprasad VL, George R, Soumittra N, Sharmila F, Vijaya L, Kumaramanickavel G. Association of non-synonymous single nucleotide polymorphisms in the LOXL1 gene with pseudoexfoliation syndrome in India. *Mol Vis*. 2008;14:318-322.
33. Pasutto F, Krumbiegel M, Mardin CY, et al. Association of LOXL1 common sequence variants in German and Italian patients with pseudoexfoliation syndrome and pseudoexfoliation glaucoma. *Invest Ophthalmol Vis Sci*. 2008;49:1459-1463.
34. Aragon-Martin JA, Ritch R, Liebmann J, et al. Evaluation of LOXL1 gene polymorphisms in exfoliation syndrome and exfoliation glaucoma. *Mol Vis*. 2008;14:533-541.
35. Ozaki M, Lee KY, Vithana EN, et al. Association of LOXL1 gene polymorphisms with pseudoexfoliation in the Japanese. *Invest Ophthalmol Vis Sci*. 2008;49:3976-3980.
36. Mossbock G, Renner W, Faschinger C, Schmut O, Wedrich A, Weger M. Lysyl oxidase-like protein 1 (LOXL1) gene polymorphisms and exfoliation glaucoma in a Central European population. *Mol Vis*. 2008;14:857-861.
37. Mori K, Imai K, Matsuda A, et al. LOXL1 genetic polymorphisms are associated with exfoliation glaucoma in the Japanese population. *Mol Vis*. 2008;14:1037-1040.
38. Mabuchi F, Sakurada Y, Kashiwagi K, Yamagata Z, Iijima H, Tsukahara S. Lysyl oxidase-like 1 gene polymorphisms in Japanese patients with primary open angle glaucoma and exfoliation syndrome. *Mol Vis*. 2008;14:1303-1308.
39. Lemmela S, Forsman E, Onkamo P, et al. Association of LOXL1 gene with Finnish exfoliation syndrome patients. *J Hum Genet*. 2009;54:289-297.
40. Wolf C, Gramer E, Muller-Myhsok B, et al. Lysyl oxidase-like 1 gene polymorphisms in german patients with normal tension glaucoma, pigmentary glaucoma and exfoliation glaucoma. *J Glaucoma*. 2009. In press.
41. Fuse N, Miyazawa A, Nakazawa T, Mengkegale M, Otomo T, Nishida K. Evaluation of LOXL1 polymorphisms in eyes with exfoliation glaucoma in Japanese. *Mol Vis*. 2008;14:1338-1343.
42. Tanito M, Minami M, Akahori M, et al. LOXL1 variants in elderly Japanese patients with exfoliation syndrome/glaucoma, primary open-angle glaucoma, normal tension glaucoma, and cataract. *Mol Vis*. 2008;14:1898-1905.
43. Goldstein DB, Tate SK, Sisodiya SM. Pharmacogenetics goes genomic. *Nat Rev Genet*. 2003;4:937.
44. Dervieux T, Meshkin B, Neri B. Pharmacogenetic testing: proofs of principle and pharmacoeconomic implications. *Mutat Res*. 2005;573:180-1894.
45. Ostapowicz G, Fontana RJ, Schiodt FV, et al. Results of a prospective study of acute liver failure at 17 tertiary care centers in the United States. *Ann Intern Med*. 2002;137:947-954.
46. Li L, Jick H, Jick SS. Updated study on risk of cholestatic liver disease and flucloxacillin. *Br J Clin Pharmacol*. 2009;68:269-270.
47. Daly AK, Donaldson PT, Bhatnagar P, et al. HLA-B*5701 genotype is a major determinant of drug-induced liver injury due to flucloxacillin. *Nat Genet.* 2009;41:816-819.
48. Bowman L, Armitage J, Bulbulia R, Parish S, Collins R. Study of the effectiveness of additional reductions in cholesterol and homocysteine (SEARCH): characteristics of a randomized trial among 12064 myocardial infarction survivors. *Am Heart J.* 2007;154:815-23.
49. Thompson PD, Clarkson P, Karas RH. Statin-associated myopathy. *JAMA*. 2003;289:1681-1690.
50. Link E, Parish S, Armitage J, et al. SLCO1B1 variants and statin-induced myopathy--a genomewide study. *N Engl J Med*. 2008;359:789-799.
51. Mitka M. Researchers worry about myopathy risk for patients taking high-dose simvastatin. *JAMA*. 2009;301:261-262.
52. Ge D, Fellay J, Thompson AJ, et al. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature*. 2009;461:798-801.
53. Thomas DL, Thio CL, Martin MP, et al. Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. *Nature*. 2009;461:798-801.
54. McHutchison JG, Lawitz EJ, Shiffman ML, et al. Peginterferon alfa-2b or alfa-2a with ribavirin for treatment of hepatitis C infection. *N Engl J Med*. 2009;361:580-593.
55. Hindorff LA, Junkins HA, Mehta JP, Manolio TA. A Catalog of Published Genome-Wide Association Studies. Cited 2009. Available at: www.genome.gov/gwastudies. Accessed 8/10/2009.
56. Erlich HA, Griffith RL, Bugawan TL, Ziegler R, Alper C, Eisenbarth G. Implication of specific DQB1 alleles in genetic susceptibility and resistance by identification of IDDM siblings with novel HLA-DQB1 allele and unusual DR2 and DR1 haplotypes. *Diabetes*. 1991;40:478-481.
57. Matsuki K, Grumet FC, Lin X, et al. DQ (rather than DR) gene marks susceptibility to narcolepsy. *Lancet*. 1992;339:1052.
58. Etzensperger R, McMahon RM, Jones EY, Fugger L. Dissection of the multiple sclerosis associated DR2 haplotype. *J Autoimmun*. 2008;31:201-207.
59. Stefansson H, Rye DB, Hicks A, et al. A genetic risk factor for periodic limb movements in sleep. *N Engl J Med*. 2007;357:639-647.
60. Winkelmann J, Schormair B, Lichtner P, et al. Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions. *Nat Genet*. 2007;39:1000-1006.
61. Vilarino-Guell C, Farrer MJ, Lin SC. A genetic risk factor for periodic limb movements in sleep. *N Engl J Med*. 2008;358:425-427.
62. Kemlink D, Polo O, Frauscher B, et al. Replication of restless legs syndrome loci in three European populations. *J Med Genet*. 2009;46:315-318.
63. Desautels A, Turecki G, Montplaisir J, et al. Evidence for a genetic association between monoamine oxidase A and restless legs syndrome. *Neurology*. 2002;59:215.
64. Vilarino-Guell C, Chai H, Keeling BH, et al. MEIS1 p.R272H in familial restless legs syndrome. *Neurology*. 2009;73:243-245.
65. Mignot E. Genetic and familial aspects of narcolepsy. *Neurology*. 1998;50(2 suppl 1):S16-S22.
66. Hallmayer J, Faraco J, Lin L, et al. Narcolepsy is strongly associated with the T-cell receptor alpha locus. *Nat Genet*. 2009;41:708-711.
67. Shi J, Levinson DF, Duan J, et al. Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature*. 2009;460:753-757.
68. Stefansson H, Ophoff RA, Steinberg S, et al. Common variants conferring risk of schizophrenia. *Nature*. 2009;460:744-747.
69. Ferreira MA, O'Donovan MC, Meng YA, et al. Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat Genet*. 2008;40:1056-1058.
70. Ma D, Salyakina D, Jaworski JM, et al. A genome-wide association study of autism reveals a common novel risk locus at 5p14.1. *Ann Hum Genet*. 2009;73(Pt 3):263-273.
71. Wang K, Zhang H, Ma D, et al. Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature*. 2009;459:528-533.
72. Wang K, Li M, Hadley D, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res*. 2007;17:1665-1674.
73. van Bon BW, Mefford HC, Menten B, et al. Further delineation of the 15q13 microdeletion and duplication syndromes: a clinical spectrum varying from non-pathogenic to a severe outcome. *J Med Genet*. 2009;46:511-523.
74. Dibbens LM, Mullen S, Helbig I, et al. Familial and sporadic 15q13.3 microdeletions in idiopathic generalized epilepsy: precedent for disorders with complex inheritance. *Hum Mol Genet*. 2009 ;18:3626-3631.
75. Brunetti-Pierri N, Berg JS, Scaglia F, et al. Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nat Genet*. 2008;40:1466-1471.
76. Bijlsma EK, Gijsbers AC, Schuurs-Hoeijmakers JH, et al. Extending the phenotype of recurrent rearrangements of 16p11.2: deletions in mentally retarded patients without autism and in normal individuals. *Eur J Med Genet*. 2009;52:77-87.
77. Mefford HC, Sharp AJ, Baker C, et al. Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N Engl J Med*. 2008;359:1685-1699.
78. St Clair D, Blackwood D, Muir W, et al. Association within a family of a balanced autosomal translocation with major mental illness. *Lancet*. 1990;336:13-16.
79. Green EK, Grozeva D, Jones I, et al. The bipolar disorder risk allele at CACNA1C also confers risk of recurrent major depression and of schizophrenia. *Mol Psychiatry*. 2009. In press.

80. Purcell SM, Wray NR, Stone JL, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460:748-752.
81. Need AC, Ge D, Weale ME, et al. A genome-wide investigation of SNPs and CNVs in schizophrenia. *PLoS genetics*. 2009;5:e1000373.
82. Walsh T, McClellan JM, McCarthy SE, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science*. 2008;320:539-543.
83. Stone JL, O'Donovan MC, Gurling H, et al. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*. 2008;455:237-241.
84. Stefansson H, Rujescu D, Cichon S, et al. Large recurrent microdeletions associated with schizophrenia. *Nature*. 2008;455:232-236.
85. Glessner JT, Wang K, Cai G, et al. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature*. 2009;459:569-573.
86. Pagnamenta AT, Wing K, Akha ES, et al. A 15q13.3 microdeletion segregating with autism. *Eur J Hum Genet*. 2009;17:687-692.
87. Hannes FD, Sharp AJ, Mefford HC, et al. Recurrent reciprocal deletions and duplications of 16p13.11: The deletion is a risk factor for MR/MCA while the duplication may be a rare benign variant. *J Med Genet*. 2008;46:223-242.
88. Gottesman I. *Schizophrenia Genesis, the Origins of Madness*. New York, NY: W.H. Freeman and Company; 1991.
89. Alkan C, Kidd JM, Marques-Bonet T, et al. Personalized copy number and segmental duplication maps using next-generation sequencing. *Nat Genet*. 2009;41:1061-1067.
90. Yoon S, Xuan Z, Makarov V, Ye K, Sebat J. Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome Res*. 2009;19:1586-1592.
91. Xie C, Tammi MT. CNV-seq, a new method to detect copy number variation using high-throughput sequencing. *BMC Bioinformatics*. 2009;10:80.
92. Korn JM, Kuruvilla FG, McCarroll SA, et al. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet*. 2008;40:1253-1260.
93. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet*. 2003;33(suppl):228.
94. Weiss LA, Shen Y, Korn JM, et al. Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med*. 2008;358:667-675.
95. Sharp AJ, Hansen S, Selzer RR, et al. Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nat Genet*. 2006;38:1038-1042.
96. de Vries BB, Pfundt R, Leisink M, et al. Diagnostic genome profiling in mental retardation. *Am J Hum Genet*. 2005;77:606-616.
97. Christiansen J, Dyck JD, Elyas BG, et al. Chromosome 1q21.1 contiguous gene deletion is associated with congenital heart disease. *Circ Res*. 2004;94:1429-1435.
98. Greenway SC, Pereira AC, Lin JC, et al. De novo copy number variants identify new genes and loci in isolated sporadic tetralogy of Fallot. *Nat Genet*. 2009;41:931-935.
99. Diskin SJ, Hou C, Glessner JT, et al. Copy number variation at 1q21.1 associated with neuroblastoma. *Nature*. 2009;459:987-991.
100. Kirov G, Grozeva D, Norton N, et al. Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. *Hum Mol Genet*. 2009;18:1497-1503.
101. Doornbos M, Sikkema-Raddatz B, Ruijvenkamp CA, et al. Nine patients with a microdeletion 15q11.2 between breakpoints 1 and 2 of the Prader-Willi critical region, possibly associated with behavioural disturbances. *Eur J Med Genet*. 2009;52:108-115.
102. Murthy SK, Nygren AO, El Shakankiry HM, et al. Detection of a novel familial deletion of four genes between BP1 and BP2 of the Prader-Willi/Angelman syndrome critical region by oligo-array CGH in a child with neurological disorder and speech impairment. *Cytogenetic Genome Res*. 2007;116:135-140.
103. Kirov G, Gumus D, Chen W, et al. Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. Hum M et al. Microdeletion/duplication at 15q13.2q13.3 among individuals with features of autism and other neuropsychiatric disorders. *J Med Genet*. 2009;46:242-248.
105. Helbig I, Mefford HC, Sharp AJ, et al. 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. *Nat Genet*. 2009;41:160-162.
106. Sharp AJ, Mefford HC, Li K, et al. A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. *Nat Genet*. 2008;40:322-328.
107. Kumar RA, KaraMohamed S, Sudi J, et al. Recurrent 16p11.2 microdeletions in autism. *Hum Mol Genet*. 2008;17:628-638.
108. Ghebranious N, Giampietro PF, Wesbrook FP, Rezkalla SH. A novel microdeletion at 16p11.2 harbors candidate genes for aortic valve development, seizure disorder, and mild mental retardation. *Am J Med Genet A*. 2007;143A:1462-1471.
109. Weber YG, Jacob M, Weber G, Lerche H. A BFIS-like syndrome with late onset and febrile seizures: suggestive linkage to chromosome 16p11.2-16q12.1. *Epilepsia*. 2008;49:1959-1964.
110. Kumar RA, Marshall CR, Badner JA, et al. Association and mutation analyses of 16p11.2 autism candidate genes. *PloS one*. 2009;4:e4582.
111. McCarthy SE, Makarov V, Kirov G, et al. Microduplications of 16p11.2 are associated with schizophrenia. *Nat Genet*. 2009;41:1223-1227.
112. Ullmann R, Turner G, Kirchhoff M, et al. Array CGH identifies reciprocal 16p13.1 duplications and deletions that predispose to autism and/or mental retardation. *Hum Mutat*. 2007;28:674-682.
113. Fernandez T, Morgan T, Davis N, et al. Disruption of Contactin 4 (CNTN4) results in developmental delay and other features of 3p deletion syndrome. *Am J Hum Genet*. 2008;82:1385.
114. Roohi J, Montagna C, Tegay DH, et al. Disruption of contactin 4 in three subjects with autism spectrum disorder. *J Med Genet*. 2009;46:176-182.
115. Kirov G, Rujescu D, Ingason A, Collier DA, O'Donovan MC, Owen MJ. Neurexin 1 (NRXN1) deletions in Schizophrenia. *Schizophr Bull*. 2009;35:851-854.
116. Rujescu D, Ingason A, Cichon S, et al. Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum Mol Genet*. 2009;18:988-996.
117. Vrijenhoek T, Buizer-Voskamp JE, van der Stelt I, et al. Recurrent CNVs disrupt three candidate genes in schizophrenia patients. *Am J Hum Genet*. 2008;83:504-510.
118. Bucan M, Abrahams BS, Wang K, et al. Genome-wide analyses of exonic copy number variants in a family-based study point to novel autism susceptibility genes. *PLoS genetics*. 2009;5:e1000536.
119. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, et al. Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet*. 2008;82:477-488.
120. Kim HG, Kishikawa S, Higgins AW, et al. Disruption of neurexin 1 associated with autism spectrum disorder. *Am J Hum Genet*. 2008;82:199-207.
121. Zahir FR, Baross A, Delaney AD, et al. A patient with vertebral, cognitive and behavioural abnormalities and a de novo deletion of NRXN1alpha. *J Med Genet*. 2008;45:239-243.
122. Friedman JM, Baross A, Delaney AD, et al. Oligonucleotide microarray analysis of genomic imbalance in children with mental retardation. *Am J Hum Genet*. 2006;79:500-513.

Personal genomes in progress: from the Human Genome Project to the Personal Genome Project

Jeantine E. Lunshof, PhD; Jason Bobe, MS; John Aach, PhD; Misha Angrist, PhD; Joseph V. Thakuria, MD; Daniel B. Vorhaus, JD, MA; Margret R. Hoehe, MD, PhD; George M. Church, PhD



The dawning of a new decade is an appropriate time to reflect on the tremendous progress that has been made in human genomic research. In 2010, with whole-genome sequencing becoming increasingly affordable, the promise of large-scale human genomic research studies involving hundreds, thousands, and even hundreds of thousands of individuals is rapidly becoming a reality. The next generation of human genomic research will occur on a scale that would have been nearly unfathomable at the start of the last decade, when the publication of the Human Genome Project's first draft results was still pending.

The cost of a diploid human genome sequence has dropped from about $70M to $2000 since 2007—even as the standards for redundancy have increased from 7x to 40x in order to improve call rates. Coupled with the low return on investment for common single-nucleotide polylmorphisms, this has caused a significant rise in interest in correlating genome sequences with comprehensive environmental and trait data (GET). The cost of electronic health records, imaging, and microbial, immunological, and behavioral data are also dropping quickly. Sharing such integrated GET datasets and their interpretations with a diversity of researchers and research subjects highlights the need for informed-consent models capable of addressing novel privacy and other issues, as well as for flexible data-sharing resources that make materials and data available with minimum restrictions on use. This article examines the Personal Genome Project's effort to develop a GET database as a public genomics resource broadly accessible to both researchers and research participants, while pursuing the highest standards in research ethics.

Dialogues Clin Neurosci. 2010;12:47-60.

Keywords: *Personal Genome Project; personal genomics; DNA sequencing technology; whole-genome sequencing; phenome; envirome; microbiome; GET data set; open consent; public genome; ELSI*

Author affiliations: Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA (Jason Bobe*, John Aach, George M. Church**); PersonalGenomes.org, Boston, Massachusetts, USA (Jason Bobe, Daniel B. Vorhaus, Joseph V. Thakuria, George M. Church**); European Centre for Public Health Genomics, FHML, Maastricht University, Maastricht, The Netherlands; Department of Molecular Cell Physiology, VU University Amsterdam, The Netherlands (Jeantine E. Lunshof*); Institute for Genome Sciences & Policy, Duke University, Durham, North Carolina, USA (Misha Angrist); Harvard Medical School, Massachusetts General Hospital, Boston, Massachusetts, USA (Joseph V. Thakuria); Robinson, Bradshaw & Hinson, P.A., Charlotte, North Carolina, USA (Daniel B. Vorhaus); Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Berlin, Germany (Margret R. Hoehe**) * Co-first authors ** Co-last authors

Address for correspondence: Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115, USA, Tel: +1 617/432-7562; PersonalGenomes.org, 77 Avenue Louis Pasteur, Boston, MA 02115 USA (e-mail: jason@personalgenomes.org)

When the Human Genome Project published its draft results on June 26, 2000, it published a compound human genome sequence containing genetic information from several volunteers. Seventy percent of the final sequence was obtained from one anonymous individual, while the remaining 30% came from a number of different individuals. From the first amalgamated human genome sequence—which was refined in 2003 and continues to be updated and refined to this day—private and public research efforts have gone on to sequence numerous individual human genomes with increasing speed and detail and decreasing time and cost. The acceleration of whole-genome sequencing in the research context necessitates new perspectives and models that enable scientists and society to learn as much as possible from this rapidly expanding dataset while still respecting important ethical, legal, and social norms.

The Personal Genome Project (PGP),[1] an ambitious research study directed by faculty members in the Department of Genetics at Harvard Medical School, aims to recruit as many as 100 000 informed participants to contribute genomic sequence data, tissues, and extensive environmental, trait, and other information to a publicly accessible and identifiable research database.

In this review we describe the Personal Genome Project itself, focusing on its unique structural features and the rationale behind the project's design. We also elucidate the changing scientific and social landscape that makes the PGP's model of open consent and public data access increasingly important to the furtherance of human genomic research.

The PGP's mission

In contrast to research studies that focus on small subsets of traits within narrowly defined human populations exhibiting single diseases, the PGP was conceived with an expansive mission. From the outset, the mission of the project *(Table I)* has been to develop a broad-based, longitudinal, and participatory research study that will facilitate a comprehensive understanding of the project's participants at the genomic level and beyond.

The PGP is constructed with the recognition that our desire to truly understand the genesis of most complex human traits—from dread diseases to the talents and quirks that make us each uniquely human—could only be satisfied by examining genomic information in context and by surrounding it with the richest possible data from the widest possible array of supplemental sources. By supplementing genomic sequence data with the collection and analysis of tissues and extensive environmental and trait data, and by making these data publicly accessible to researchers worldwide, the PGP aims to improve understanding of the ways in which genomes plus environments ultimately equal traits *(Figure 1)*.

The PGP is more than just a research repository. In addition to its publicly accessible research database, the PGP, which is supported by the nonprofit PersonalGenomes.org, also works to disseminate genomic technology and knowledge at a global level, thereby producing tangible and widely available improvements in the understanding and management of human health and disease. The PGP also

The Personal Genome Project's Mission Statement

The mission of the Personal Genome Project is to encourage the development of personal genomics technology and practices that:

- are effective, informative, and responsible
- yield identifiable and improvable benefits at manageable levels of risk
- are broadly available for the good of the general public

To achieve this mission we will build a framework for prototyping and evaluating personal genomics technology and practices at increasing scales. In support of this goal, we will:

- develop a broad vision for how personal genomes may be used to improve the understanding and management of human health and disease
- provide educational and informational resources for improving general understanding of personal genomics and its potential
- recruit individuals interested in obtaining and openly sharing their genome sequences, related health and physical information, and reporting their experiences as a participant of the project on an ongoing basis
- develop technologies to improve the accessibility of personal genome sequencing
- foster dialog with research communities, industries, and public and governmental bodies with interests in personal genomics, and related ethical, legal, and social issues (ELSI)
- develop tools for interpreting genomic information and correlating it with personal medical and biological information

Table I. PGP's Mission Statement, available at: htttp://www.personalgenomes.org/mission.html.[1]

finds itself at the forefront of discourse surrounding the ethical, legal, and social issues (ELSI) associated with large-scale whole-genome sequencing, particularly in the areas of privacy, informed consent, and data accessibility. The PGP is, and is intended to be, a research project that is constantly in progress, exploring the boundaries of human genomic research in a way that produces maximal advances in scientific understanding and public understanding and well-being, while striving to reach beyond what is minimally required to satisfy its ethical, legal, and social obligations to its participants. In the sections that follow we report on unique aspects of the PGP relating to technology development, integrative genomics, and human subject research protocols, as well as describe the development and current state of the PGP.

Key developments in human genome sequencing

The PGP derives its impetus and importance from historic breakthroughs in understanding and analysis of DNA. DNA comprises only a very small fraction of a cell (~3% dry weight *E. coli*), and its role as the molecule primarily responsible for transmission of genetic traits was not recognized until a series of discoveries beginning in the 1940s. The emergence in 1953 of a clear concept of DNA as a double-helical structure comprising a pair of complementary strings of four elementary bases (the nucleotides A, C, G, and T) crystallized interest in determining the DNA sequences of genes and the sequence differences responsible for disease, and set the stage for over four decades of development of ever more efficient and comprehensive sequencing methods. *Table II* describes this history by a set of milestones that take one from the early beginnings of DNA sequencing up through delivery of draft human genome sequences in 2001 to 2003. In the 38 years between 1965, when Robert Holley and colleagues at Cornell and the US Department of Agriculture sequenced a 77 nt RNA gene after 4 years of effort, and 2003, when the public Human Genome Project (HGP) declared that it had met its goals regarding delivery of a ~3Gbp human genome sequence, the size of DNA sequence that could be accommodated by sequencing technology improved ~30 million-fold.

Post-HGP sequencing—towards whole diploid genomes

Notably, the HGP had delivered only a single human genome sequence that was a composite built from a small number of deidentified individuals, while the competing nonpublic human genome project merged in data from an identified individual (Craig Venter); both were haploid estimates. As recognized from the beginning of the HGP, many additional resources would be needed to understand the functions of the genes laid out in these "reference" human genomes, and to identify the sequence differences between individuals that contribute to individual traits, health, and disease. Indeed, as the HGP ended, projects were already under way to identify large numbers of genetic differences from the HGP-derived reference genome in different human populations that could sub-

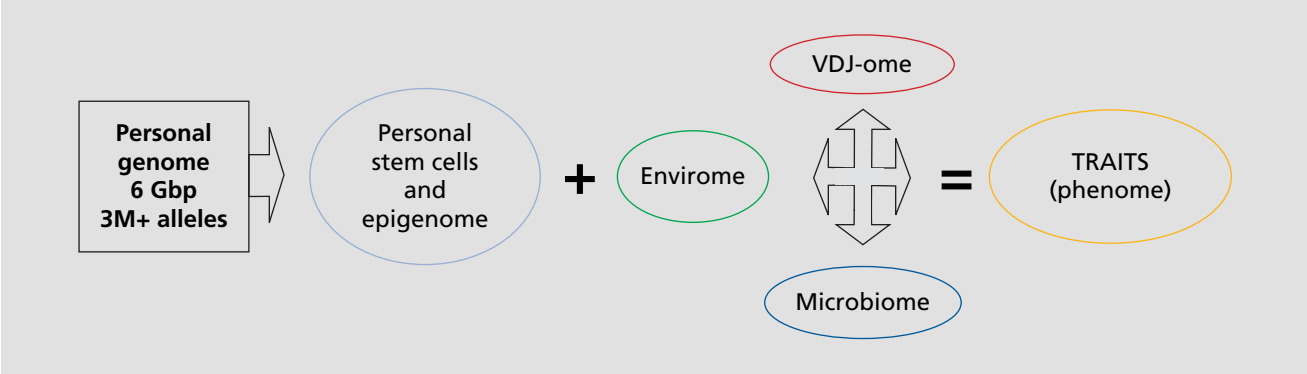


Figure 1. Genome + Environment = Traits (GET) equation. Envirome: the totality of environmental influences; VDJ-ome: the DNA sequences of the entire repertoire of an individual's immunoglobulin and T-cell receptors, which reflect a lifetime of antigenic exposures; Microbiome: the billions of commensal, symbiotic, and pathogenic micro-organisms that share our body space; Epigenome: the totality of programmed biochemical and structural modifications to genomic DNA that regulate organism or phenotype development. (see overview in *Table III*).

sequently be analyzed using low-cost array methods in large numbers of individuals, a strategy that has since given rise to more than 480 published genome-wide association studies.[16,17] At the same time, however, interest was rising in the second approach: to significantly improve DNA sequencing technology to a point where an individual's entire genome could be sequenced at very low cost. A combination of two kinds of arguments were advanced supporting this approach, focusing on functional utility and economics, respectively.

The gist of the functional arguments was that sequencing of individuals is intrinsically more informative and flexible than array-based interrogation of known sites of variation and that, variation aside, any improvements in sequencing cost and capability could be quickly applied to numerous general aspects of biology that are critical to understanding gene function, traits, and health and disease.[18,19] The relative advantages of sequencing have long been recognized. Unlike array analyses, sequencing: (i) does not require variations to be preidentified; (ii) can more readily accommodate more complex variations than single nucleotide changes and very short inserts or deletions; and (iii) need not focus on variations that are common in large populations vs rare or unique variations. In consequence, as sequencing technology has improved, it has increasingly been integrated into association studies of variation.[20-23]

However, these advantages of sequencing were counterbalanced by their high cost, a situation well illustrated by the $3 billion US cost of the HGP itself. It is here that economic arguments were advanced suggesting that dramatic improvements in sequencing were feasible that might ultimately enable an individual's genome to be sequenced for 1000 to 10 000 USD.[18] On an empirical level, sequencing technology has appeared to exhibit a historical trend of exponentially decreasing costs with time as measured by sequenced base pairs per dollar at a given error rate, a situation frequently compared with "Moore's Law" in computing,[24] which noted that computing power measured by the integrated circuit transistor density doubled roughly every 2 years at constant cost *(Figure 2)*.[18,25] To get genome sequencing costs down to $1000 would require cost and throughput improvements of an additional 4 to 5 orders of magnitude, so the question of economic feasibility ultimately turned on whether new methods could enable this very large improvement.

Here, the HGP again gave grounds for optimism, for even though the HGP itself only achieved 100-fold improvements, it achieved this largely by refining, miniaturizing, and robotically scaling up, but not fundamentally changing, a Sanger sequencing method initially developed over 20 years earlier *(Table II)*. If such methods were capable of 100-fold improvement, considerably greater improvements might be expected from more radically changing sequencing chemistry, signal generation and detection, and instrumentation in ways that could integrate some of the vast advances in chemistry and

Date	Event	Size of sequence (bp)	Reference
1957	First sequence mutation identified responsible for disease	1 amino acid (sickle cell vs normal hemoglobin)	(Ingram 1957[2])
1965	First sequence of a single complete gene	77 bases	(Holley, Apgar et al 1965[3])
1976-1977	Sequencing of first viral genomes	3562 bases (MS2 RNA phage) 5375 bases (φ X174 DNA phage)	(Fiers, Contreras et al 1976[4]; Sanger, Air et al 1977[5])
1975-1977	Maxam/Gilbert and Sanger DNA sequencing methods		(Sanger and Coulson 1975[6]; Maxam and Gilbert 1977[7]; Sanger, Nicklen et al 1977[8])
1994	First commercial bacterial genome sequence	1.7Mbp *(Helicobacter pylori)*	(Nature Genetics, May 1996[9])
1995	First published bacterial genome sequence	1.83Mbp *(Haemophilus influenzae)*	(Fleischmann, Adams et al 1995[10])
1998-2000	Genome sequences of first animals	100Mbp *(Caenorhabditis elegans)* 120Mbp *(Drosophila melanogaster)*	(C. elegans Sequencing Consortium 1998,[11] Adams, Celniker et al 2000[12])
2001	Two draft sequences of human genome	~3Gbp	(Lander, Linton et al 2001,[13] Venter, Adams et al 2001[14])
2003	Completion of public Human Genome Project		(Collins, Morgan et al 2003[15])

Table II. Development of DNA sequencing.

enzymology, optics and electronics, materials science, microfabrication, and process control that had accrued over the preceding 20 years and been put to good use in many other fields. The HGP also directly provided an important resource for realizing this strategy: the reference human genome sequence itself, as this could serve as a template against which reads obtained by new technologies could be located, allowing new human genomes to be assembled at least initially by "resequencing" vs de novo assembly. This reduces the burden on new sequencing methods by allowing them to generate useful data with shorter reads and higher base call error rates than would generally be needed for de novo assembly, although de novo assembly of genomes using new sequencing technology remains an important goal.

Next-generation sequencing

Researchers were quick to work out sequencing approaches along the lines indicated in these arguments, and commercial products emerged soon, giving rise to *next-generation* sequencing (NGS). Soon granting agencies promised funding for support, and a ~10M USD competition was announced for rapid, accurate genomic sequencing, generating increased coalescence around target goals for dramatic improvements to sequencing technology.[26,27,28] Detailed reviews and comparisons of NGS approaches have been published.[18,29,30]

Among the earliest NGS methods were polony sequencing (the Polonator) and 454 Life Sciences.[31,32,33] Both methods amplify DNA templates onto microbeads that are packed onto two-dimensional arrays for sequencing, thereby achieving enormous economies of scale compared with Sanger sequencing, and each achieved ~25-fold better cost per bp compared with HGP *(Figure 2)*. However, each uses different sequencing chemistry and arraying technology, giving rise to many technical trade-offs. Together they proved the general point that great improvements in sequencing efficiency were indeed within reach, but also that the precise character and degree of improvement would depend closely on the novel technologies employed and the ingenuity with which they could be integrated. A second wave of development introduced methods by Illumina and ABI that, by very different means, have improved the utility and costs, *(Figure 2)*[34,35] and hence use of these systems is becoming widespread for both large scale and "deep" sequencing applications, and both are under continuous development. Two complete cancer genomes were recently sequenced, one with each platform.[36,37] Further rounds of innovation have yielded a diverse set of newer NGS methods. For instance, a number of "single-molecule" sequencing methods are now available or in development. These methods avoid the need to make thousands to millions of copies of DNA template molecules on microbeads or surfaces to assure that sequencing operations generate sufficient signal to read individual bases accurately, and instead use highly sensitive optics to detect bases at the single molecule level; this allows even denser packing of DNA templates and further efficiencies in sequencing chemistry. While Helicos Biosciences has commercialized a single-molecule system that simply arrays single template molecules on a surface and uses sequencing cycle similar to the methods above, Pacific Biosciences is developing a system in which enzymes and templates are tethered to the bottom of nanofabricated wells and which monitors the signals generated by sequencing chemistry in real-time vs artificial cycles.[38,39] Here, the nanofabricated wells enable substantially increased accuracy of single molecule base incorporation events. Finally, on another track, the company Complete Genomics, Inc has developed a method whereby very compact self-assembling amplicons of template DNAs called "nanoballs" are flowed onto a nanofabricated grid of ~300nm spots at 700 to 1300 nm center-to-center distances. Three complete human genomes were sequenced with this method (as of January 2010) with an average consumable cost of $4400 and as low as $1500 for 40X coverage.[40]

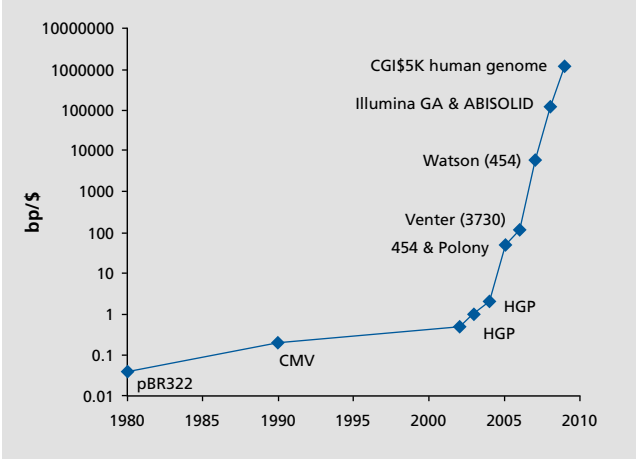


Figure 2. Exponential trend of sequencing costs in base pairs per USD (bp/$), a trend often compared with Moore's Law (see text). See ref 25 for details.

Towards affordable personal genomes

These developments suggest that technology capable of meeting the cost target of $1000 or less for a diploid human genome sequence is within reach. Indeed, the in-depth resequencing of individual human genomes has now been demonstrated several times by NGS developers to demonstrate that their methods have come of age. There are now published full genome sequences for at least seven individuals,[40] with some having been sequenced by more than one method. There are also tens—and perhaps hundreds—of additional unpublished or partly published genomes (see, eg, refs 36,37),while the lower-coverage 1000 Genomes Project[20,21] continues. Clearly, the age of personal genomics is now close at hand.

The PGP

As described in the first section, one of the PGP's central aims is to develop a publicly available, fully consented database containing comprehensive human genome and phenome data for its research participants. Such integrated datasets are fundamental drivers of progress in functional genomics and enable systems biology-based insights into the mechanisms of human health and disease.[41] PGP studies will look beyond inherited genomes to include somatic and epigenetic variation data, as well as relevant microbiome, transcriptome, immunity-reflecting "VDJ-ome" and phenome data to develop comprehensive profiles. By developing high-resolution data profiles for each participant, and multiplying that by a large (up to 100 000) participant population, the PGP will also generate valuable data describing the kinds and distributions of variation that exist in populations. Although an improved understanding of human health and disease is a central aim of the PGP, its focus is considerably broader and will enable research into the social and behavioral sciences using personal genomic data. Finally, the PGP's flexible study protocol and public and distributed approach to research enables it to keep pace with sequencing and other technological advances while simultaneously driving these developments.

Integrated personal genomes: inherited, somatic, environmental genomics

If the PGP is to fulfill its mission to address the multidimensional complexity of human biology, it must encompass multiple interacting "-omes." For example, a person's diet will have a profound influence upon her or his somatic gene expression as well as the genomic and proteomic activity of the person's microbiome. It will also affect the metabolome. Similarly, an individual's environmental exposures to pollutants will have a direct bearing on her or his immunological response and therefore, on the VDJ-ome. Germline alleles will affect how one metabolizes drugs, which will have myriad effects on an individual's physiological and behavioral phenotypes.

Genomes (vs exomes)

In its early phase, given the then-current cost of genomic sequencing, the PGP planned to focus on exomes rather than whole genomes as a way to affordably expand the project to large numbers of participants. Despite representing only 1% to 2% of the 6 billion base pairs in a human genome, the exome contains all protein-coding exons and therefore provides access to the majority of known functional variants.[48,49,50] However, continued improvements in genomic sequencing have produced price declines that have rendered whole-genome sequencing significantly cheaper per base pair than exome sequencing. The PGP, as a result, has determined that whole-genome sequencing is cost-justified given the relatively high price of exomes and the additional information supplied by whole-genome sequences of PGP participants.[51] See also *Table III* for the various "omes."

Phenomes

Detailed phenotype data is required to categorize and, ultimately, understand the phenotypes that the PGP seeks to explore. However, the vastness of the human phenome, defined as the physical totality of human traits at all levels, from the molecular to the behavioral, will require new strategies that permit high-throughput trait collection while yielding accurate and standardized phenotypic data. With regard to the cellular and molecular phenotypes, the PGP collects participant tissue samples and develops cell lines that are then deposited and publicly accessible through established biobanks.[52,53]
As the PGP expands it is exploring Web-based, high-throughput behavioral phenotype data-collection models pioneered by leading public and private researchers. While the reliability and validity of self-reported traits is a concern, particularly for phenome research con-

ducted online,[54,55] Web-based assessments provide distinct opportunities for "dynamic phenotyping" based on a particular individual's prior genotype-phenotype associations.[56] The multimodal capabilities of Web-based trait collection instruments, combined with their low cost of implementation at large scales, seem likely to accelerate the ability of studies like the PGP to effectively explore new corners of the human phenome.

The PGP is also taking advantage of recent advancements in health information technologies to assist participants and researchers alike in structuring and accessing the massive amounts of personalized data generated by the project. The emergence of online Personally Controlled Health Record (PCHR) platforms and other novel tools enables individuals to collect and manage their own health data—including health history, medication, allergy, immunization, biometric and other data types[57,58,59]—and can be developed for integrated data entry, access and dissemination by both the individual and third-party researchers or data providers, including health care providers.

Enviromes

The picture of genome and phenome is incomplete without the envirome. The envirome can be described as the totality of equivalent environmental influences contributing to all disorders and organisms.[60] The mode of response of an organism to the environment that is reflected in its phenotype is constrained by its unique set of genetic variations and the environmental influences on gene expression. Therefore, a comprehensive approach is required to describe the envirome systematically in conjunction with genome and phenome information. The relevant envirome data is too large and complex to be reported, managed, or analyzed manually. The creation of phenome-genome and genome-envirome networks has been suggested in order to relate phenome and envirome information to potential disease-associated genes.[61]

Microbiomes

Even though microbial cells are estimated to outnumber human cells in a single individual by a factor of ten, we know very little about the microbes that live in and on us, including what mixture of bacteria, viruses, and other micro-organisms constitute a "normal" human microbiome and how those organisms impact different biological states.[62] Major efforts such as the Human Microbiome Project are under way to characterize the microbiota at different body sites in humans and to assess how variation in microbial communities is associated with states of health and disease.[63] The PGP takes advantage of the unique availability of comprehensive participant profiles and uses them to explore interactions between host genetic and phenotypic variability alongside the genomic variation in the microbes that colonize them.[64]

The VDJ-ome

The Church Lab at Harvard Medical School is developing techniques for characterizing the repertoires of B- and T-cell receptors in individual humans from blood samples and correlated across time with personal exposure histories, with an ultimate goal of characterizing individuals repertoires of *linked* VDJ and VJ sequences.

Personal genome: Entire diploid human genome of a single individual representing 6 billion base pairs.
Exome: All exons, representing 1% to 2% of the entire human genome.
Phenome: Set of all traits in an organism, at all levels, or one of its subsystems, including morphology, physiology, and behavior.[42,43]
Envirome: The totality of equivalent environmental influences contributing to all disorders and organisms.[44]
Microbiome (human): The ecological community of commensal, symbiotic, and pathogenic microorganisms that share our body space.[45]
VDJ-ome: The repertoire of rearranged V, D, and J genome segments present in an individuals's B and T immune cells at any given time (see *Table IV*).
Transcriptome: The set of all RNA molecules, including mRNA, rRNA, tRNA, and noncoding RNA produced in one or a population of cells.[46]
Epigenome: The totality of programmed biochemical and structural modifications to genomic DNA that regulate organism or phenotype development.
Metabolome: Total set of metabolites generated by an organism, or subsystem.
Proteome: The entire set of proteins expressed by a genome, cell, tissue or organism at a given time under defined conditions. There are more proteins than genes.[47]

Table III. The "omes."

These techniques will be directly applicable to PGP participants and their self-reported data, and will yield a database of unprecedented depth describing the diversity and time development of human immune responses of large numbers of individuals in their life contexts.

The adaptive immune system

The adaptive immune system enables individuals to respond to their unique exposure histories to pathogens and environmental antigens, and possibly to cancerous mutations in their own cells, by generating and modulating expression of $>10^{12}$ unique antibodies from B cells and T cell receptors.[65] Antibody diversity derives from programmed stochastic rearrangements in maturing B cells of ~40 V, 23 D, and ~5 J functional genomic segments into VDJ heavy chains, and ~35 V and ~5 J segments into VJ light chains (κ or λ) in B cells, that are further randomized by somatic hypermutation; a similar process occurs in T cells.[66] NGS methods are now allowing researchers to identify and analyze expressed VDJ sequences in depth.[67]

Table IV. The adaptive immune system and the VDJ-ome.

Tissue reprogramming

The PGP also applies advances in tissue reprogramming techniques to tissue samples collected from PGP participants. Cells from collected somatic tissues are reprogrammed into induced pluripotent stem (iPS) cells[68] and made to differentiate into the cell types that are targeted for functional analysis. These methods enable experimental access to diverse tissue types that would otherwise be unobtainable from human subjects but are routinely analyzed in model organisms, and thus, PGP participants can effectively serve as *human model organisms*. By examining multiple cell types from a single individual, differences in physiological states within and between tissues can be compared within a single PGP participant and/or across the entire PGP cohort. This approach also permits researchers to elucidate connections between genetic variation and variation in other molecular traits, such as gene expression or epigenetic modifications.[69] Stored fibroblast cell lines provide researchers with access to renewable supplies of different tissue types from PGP participants.

The PGP: from personal to public genomes

The potential benefits arising from large-scale and integrated human genomic datasets are immense.[70] The utility of such research, however, depends upon the responsible development and widespread availability of such comprehensive datasets, which in turn depends on describing and addressing the various ethical, legal and social challenges. Those challenges include a standard set that are inherent to any research involving human subjects, as well as certain challenges that are unique to "public genomics"[71] research involving publicly available, identifiable whole-genome sequence data, such as the model pioneered by the PGP. We use the term "public genomics" to denote research studies that possess the following three critical attributes.

Integrated data

The various data types, including genomic and phenomic or trait data, are accessible in a linked format, such as a PCHR or other integrated data structure. Through this explicit linkage of data it is possible to ascertain the complete list of available traits and genetic variants for any given participant. Integration also facilitates participant-researcher interactions, longitudinal study and recontact and, crucially, simultaneous investigation of the full range of complex trait associations. Although participants need not be explicitly identified, integrated data sets that include both genomic and phenomic data will be identifiable in most cases. For this reason, participants must be made explicitly aware of the probability that they will be identified with their publicly available data, rendering promises of perfect privacy, anonymity, or confidentiality impermissible within the public genomics model. However, the promise of privacy need not give way to a promise of publicity.

Open access

Data sets and tissues are made publicly available with minimal or no access restrictions (including researcher qualifications and cost), and are generally transferable outside the original research study to be utilized by and combined with data from third parties. Well-developed data structures and intellectual property licenses are important components of this characteristic. Developing datasets that are not only publicly available but also easily portable fosters the development of a genomic commons, allows data validation by third parties, and enables the use and application of data in novel contexts that may not be foreseeable at the time of collection, thereby

facilitating hypothesis generation, encouraging serendipity and broadening the genomic research community.

Voluntary and informed participation

Satisfaction of the first two criteria publication of an integrated dataset in an open-access format necessitates that a premium be placed on receiving truly voluntary and informed consent from participants in public genomics research projects. Given the yet-unknown outcomes and the potential personal, familial, and social risks associated with such research, enrollment is only acceptable under an informed consent protocol that is specially designed to meet the highest standards of human research subjects protection in view of these conditions.

The study protocol

The PGP aims to produce public genomics research—and to develop and evaluate associated technologies and research—on a large and expanding scale. In October of 2008, the PGP published the first integrated set of DNA sequences, traits, and tissues collected from ten participants (the "PGP-10") enrolled in a pilot study initiated in 2005. Today, the PGP is incrementally expanding its cohort toward 100 000 participants. More than 12 000 individuals had registered to participate in the PGP as of February 2010. In the following section we highlight significant features of the PGP study protocol as it is implemented for the enrollment of the first 100 participants ("PGP-100") and summarized in *Table V*.

Public genomes: adding to ELSI

The practice of public genomics poses its own challenges, especially for the organization and governance of human subjects' research, forcing us to critically reassess current frameworks and practices. In order to pursue innovative research in a responsible manner, the PGP has devel-

Eligibility screening	• Review and sign "mini-consent" form. • Eligibility questionnaire about family circumstances and privacy preferences. • Entrance exam to ensure informed consent; includes potential risks of participating, project protocols, and basic genetics. • Review of full PGP consent form. • Submit information or delete account.
Pre-enrollment	• Consent to participate. • Collection of baseline trait data via questionnaire and a personal health record. Includes allergies, immunizations, medical history, medications, physical traits and measurements, diet, ethnicity/ancestry, lifestyle, and environmental exposures. • Participants asked to make a financial pledge (does not impact enrollment decisions). • Identity verification and provision of mailing address. • Submission of application for enrollment. Individuals selected to continue the enrollment process will receive an enrolment kit by mail, including saliva collection materials.
Enrollment	• Participants may be interviewed by one or more PGP staff to verify identity and consent, confirm familiarity with study protocols, and/or review trait questionnaire responses. Blood samples, saliva sample, and/or skin cells may be collected. • Tissue samples prepared for DNA sequencing and other biological analyses. • Participants opt-in to have their profiles made available on a publicly accessible Web site, or withdraw from the study. • Establishment, distribution and analysis of cell lines for research.
Ongoing participation	• Information collected for 25 years. Participants can leave the study at any time. • Data Safety Monitoring Board monitors the impacts of the PGP on enrolled participants. Quarterly emails inquire about adverse events. • Additional trait data and tissue samples may be requested periodically.

Table V. Overview of PGP study protocol.

Adapted from ref 52: Angrist M. Eyes wide open: the personal genome project, citizen science and veracity in informed consent. *Pers Med*. 2009;6:691-699. Copyright © Future Medicine 2009

oped a number of project-specific tools and resources relevant to ELSI.

Open consent

The "open consent" model developed by the PGP is designed to address the set of challenges associated with the creation of datasets where it may be possible to identify individual participants with their genomic and other data. The open consent model assumes that, in such a context, conventional assurances of anonymity, privacy and confidentially are impossible and should not serve as any part of the foundation for the informed consent protocol.[72,73] Due to the structure of public genomics projects such as the PGP, and their associated datasets, while privacy and confidentiality can be *protected* they cannot and should not be *guaranteed* to participants. This practice ensures veracity, which we regard as a necessary—though not sufficient—prerequisite for the exertion of substantive autonomy. It is only through veracity that the criteria underlying truly *informed* consent can be satisfied.

Open consent is therefore based on complete openness and transparency with regard to all aspects of participation, including the potential for reidentification and the reality that there may be other risks that are unidentifiable at the time of consent. Predicting *all potential* risks is by definition impossible and even a list of known possible risks is unlikely ever to be comprehensive.

Data sharing—and the risks of public genomes

The PGP's informed consent process begins with an extensive pre-enrollment educational examination designed to ensure a potential participant's ability to understand the specific nature of the data collected and the risks presented by public genomics research. For individuals who demonstrate the needed proficiency, the specific informed consent agreement that follows includes a lengthy but "noncomprehensive list of hypothetical scenarios that could pose risks" for participants and their families *(Table VI)*. Participants are warned that "the complete set and magnitude of the risks that the public availability of [your genomic data] poses to you and your relatives is not known at this time." It is crucial that participants understand that once identifying genetic and trait data and tissues are released into the public domain for the express intent of broad dissemination and use by third parties it will be, in all likelihood, impossible to effect a meaningful retraction at a later date.

The PGP's informed consent agreements and broader study protocol are developed in continuous close interaction with the Harvard Medical School Committee on Human Studies. The project is also overseen by an independent Data Safety Monitoring Board. Removing potentially disingenuous promises of anonymity, privacy, and confidentiality, while seeking to comprehensively and openly describe both known and unknown risks of participation, helps to ensure that research participants are as

Potential risks of participation in the PGP as described in the consent form (Abbreviated)
• The risks of public disclosure of your genetic and trait information could affect your employment, insurance and financial well-being and social interactions for you and your family.
• Anyone with sufficient knowledge and resources could take your DNA sequence data and/or posted trait information and use that data, with or without modification, to: (i) infer paternity or other features of your genealogy; (ii) claim statistical evidence that could affect your employment, insurance or ability to obtain financial services; (iii) claim relatedness to criminals or incriminate relatives; (iv) make synthetic DNA and plant it at a crime scene, or otherwise use it to falsely identify you; or (v) reveal the possibility of a disease or unknown propensity for a disease.
• Whether or not it is lawful to do so, you could be subject to actual or attempted employment, insurance, financial, or other forms of discrimination or negative treatment on the basis of the public disclosure of your genetic and trait information by the PGP or by a third party.
• The distribution of your cell lines could result in the creation and further distribution by a third party of additional cell lines, organs, or tissues containing your DNA for research, commercial, clinical, or other uses, including certain forms of assisted reproduction, some of which you may find objectionable or upsetting.
• If you have previously made available or intend to make available genetic information in a confidential setting, for example in another research study or in a clinical trial, the data that you provide as part of the PGP may be used, on its own or in combination with your previously shared data, to identify you as a participant in otherwise confidential genetic research or trials.

Table VI. Potential risks of participation.

informed as possible about the nature of public genomics research and, simultaneously, safeguards the trustworthiness of scientists and of scientific research in general.

Return of research data to participants

Research volunteers have been traditionally treated as "objects" of study who have no intrinsic rights to the data generated by their participation.[74] Today, we see that study participants are increasingly asking for access to their data[75] and that available information and communication technologies have turned the return of research results into a feasible option. While some researchers adhere to the traditional viewpoint that research subjects should not or cannot receive identifiable research data, some have suggested legal and ethical grounds for finding that researchers possess the obligation to inform their participants of certain results, particularly when they are clinically actionable.[76] However, defining the scenarios in which research results should be reported—and how to report such results—remains a challenging issue. The medical, financial, and psychosocial risks of disclosing variants of known and unknown clinical significance require that a careful distinction be made between those variants in which convincing clinical observational data exists and those in which disease association is less robust; a distinction that can influence both when and how to return results. Other concerns that have been voiced include the uncertainty surrounding regulations governing the return of genomics research results directly to participants, the impact of false-positive and/or false-negative results, as well as the "incidentalome,"[77] and in the context of commercial direct-to-consumer testing, the concern that obtaining results could lead to a "raiding of the medical commons."[78]

As new models of genomic research and commerce emerge, new mechanisms for communicating results to participants are also being explored. Many of these new models embrace a high level of involvement from their participants and, in return, may rely on some combination of education, informed consent, and intermediation to return data in a responsible fashion.[79]

The public genomics model adopted by the PGP utilizes the first two approaches while foregoing the third, opting to return data directly to research participants without the *required* intervention of an intermediary. The advantages of direct data return and participant communication are blunted by the partial shifting of the interpretative burden from the clinician to the researcher. The PGP has approached this issue by focusing on data disclosure via the Preliminary Research Report (PRR), which contains a noncomprehensive list of genetic variants present in the participant's DNA sequence data currently thought to have a likelihood of clinical relevance among individuals possessing such variants.

This preliminary identification of potentially significant variants is not intended to substitute in any way for professional medical advice, diagnosis or treatment. It leverages current knowledge by combining an evolving set of filtering algorithms and the use of existing variant databases— neither of which can be expected to have 100% accuracy in identifying truly pathogenic variants given the gaps in current scientific understanding. Participants are specifically instructed to confirm any potentially significant findings in consultation with their health care provider. It is possible that the increased rate of data return from public genomics research—as well as from commercial providers of personal genomic data—will help speed the creation of universal standards for clinical genomic interpretation that will help shift some of the interpretative burden back away from public genomics researchers.

Outlook: the PGP from 10 to 100 000

After publishing initial data from its first 10 participants in 2008, the PGP has continued to broaden the scope of the information it is collecting and publishing while simultaneously commencing the next stages of participant enrollment. From exome to whole-genome sequence data, the development and release of the GET-EvidenceBase tool[80] for generation of Preliminary Research Reports, and the publication of substantial scholarship based on the PGP data generated to date, the project's progress has been substantial. The PGP is now supported by PersonalGenomes.org, a 501(c)(3) non-profit charity that coordinates the international efforts of the PGP with other collaborative public genomics research projects around the world. Both the PGP and PersonalGenomes.org continue to strive to develop and disseminate genomic technologies, phenotyping strategies, and knowledge on a global scale and to produce tangible and widely available improvements in the understanding and management of human health in a responsible fashion.

Avances en el genoma personal: desde el Proyecto Genoma Humano al Proyecto Genoma Personal

El costo de una secuencia del genoma humano diploide se ha reducido desde cerca de 70 millones de dólares a 2000 dólares desde 2007, aunque los estándares de la redundancia han aumentado de 7 a 40 veces para mejorar los índices de demanda de genotipo. Junto con el bajo retorno de inversión para los polimorfismos de nucleótidos únicos comunes, esta situación ha causado un aumento significativo del interés en correlacionar las secuencias genómicas con una completa información ambiental y de rasgos (GAR). El costo de las fichas médicas electrónicas, de las imágenes y de la información microbiológica, inmunológica y conductual también está reduciéndose rápidamente. El compartir tal conjunto de información y sus interpretaciones con una diversidad de investigadores y sujetos de investigación pone de relieve la necesidad de contar con modelos de consentimiento informado capaces de estar orientados hacia nuevos temas de privacidad y otros, además de flexibilizar los recursos de datos compartidos que permitan disponer de materiales e información con mínimas restricciones de uso. Este artículo examina el esfuerzo del Proyecto de Genoma Personal para desarrollar una base de datos de GAR como un recurso de genómica pública ampliamente accesible tanto a investigadores como a participantes de las investigaciones, respetando los estándares más elevados de la ética de la investigación.

Les progrès du génome personnel : de l'étude du génome humain à l'étude du génome personnel

Le coût de séquençage d'un génome diploïde humain a chuté de 70 millions de dollars à 2 000 dollars depuis 2007, bien que les standards de redondance aient augmenté de 7 à 40 fois afin d'améliorer le taux d'identification des bases. Associé au faible retour sur investissement des polymorphismes de simples nucléotides (SNP), cette situation explique l'intérêt accru pour la corrélation des séquences des génomes avec des données complètes environnementales et de traits (GET). Les coûts des enregistrements numériques médicaux, de l'imagerie et des données microbiennes, immunologiques et comportementales chutent aussi rapidement. Le partage de telles bases de données GET intégrées et de leurs interprétations avec un grand nombre de chercheurs et de sujets de recherche souligne la nécessité de modèles de consentement éclairé nécessaires à cette nouvelle protection des données personnelles et autres problématiques, en plus des besoins de flexibilité des ressources requises pour le partage des données, permettant en plus une utilisation peu restrictive de ces matériels et données. Cet article analyse les efforts du Projet du Génome Personnel afin de développer une base de données GET en tant que ressource génomique publique, largement accessible à la fois aux chercheurs et aux participants à la recherche, tout en respectant les standards les plus élevés de l'éthique de la recherche.

REFERENCES

1. Personal Genome Project. http://www.personalgenomes.org. Accessed January 30, 2010.

2. Ingram VM. Gene mutations in human haemoglobin: the chemical difference between normal and sickle cell haemoglobin. *Nature*. 1957;180:326-328.

3. Holley RW, Apgar J, Everett GA, et al. Structure of a ribonucleic acid. *Science*. 1965;147:1462-1465.

4. Fiers W, Contreras R, Duerinck F, et al. Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature*. 1976;260:500-507.

5. Sanger F, Air GM, Barrell BG, et al. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature*. 1977;265:687-695.

6. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol*. 1975;94:441-8.

7. Maxam AM, Gilbert W. A new method for sequencing DNA. *Proc Natl Acad Sci U S A.* 1977;74:560-564,

8. Sanger F, Nicklen S, Cowlson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*. 1977;74:5463-5467,

9. Anonymous. Editorial. Capitalizing on the genome. *Nat Genet*. 1996;13:1.

10. Fleischmann RD, Adams MD, White Oet al. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science*. 1995;269:496-512.

11. C. elegans Sequencing Consortium. Genome sequence of the nematode C. elegans: a platform for investigating biology. *Science*. 1998;282:2012-2018.

12. Adams MD, Celniker SE, Holt RA, et al. The genome sequence of Drosophila melanogaster. *Science*. 2000;287:2185-2195.

13. Lander ES, Linton LM, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409:860-921.

14. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. *Science*. 2001;291:1304-1351.

15. Collins FS, Morgan M, Patrinos A. The Human Genome Project: lessons from large-scale biology. *Science*. 2003;300:286-290.

16. International HapMap Consortium. A haplotype map of the human genome. *Nature*. 2005;437:1299-320.

17. Hindorff LA, Junkins HA, Mehta JP, Manolio TA, et al. A Catalog of Published Genome-Wide Association Studies. Available at: http://www.genome.gov/ gwastudies. Accessed 30 January, 2010.

18. Shendure J, Mitra RD, Varma C, Church GM, et al. Advanced sequencing technologies: methods and goals. *Nat Rev Genet*. 2004;5:335-344.

19. Kahvejian A, Quackenbush J, Thompson JF. What would you do if you could sequence everything? *Nat Biotechnol*. 2008;26:1125-1133.

20. 1000 Genomes Project. 2007 Meeting Report: A Workshop to Plan a Deep Catalog of Human Genetic Variation. Available at: http://www.1000genomes.org/docs/1000Genomes-MeetingReport.pdf. Accessed January 30, 2010.

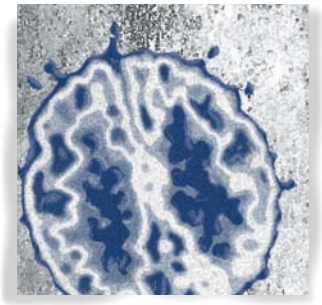
21. Altshuler D, Daly MJ, Lander ES, et al. Genetic mapping in human disease. *Science*. 2008;322:881-888.

22. Siva N. 1000 Genomes project. *Nat Biotechnol*. 2008;26:256.
23. Nejentsev S, Walker N, Riches D, Egholm M, Todd JA. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science*. 2009;324:387-389.
24. Moore GE. Cramming more components onto integrated circuits. *Electronics*. 1965;38:114-117.
25. Carr PA, Church GM. Genome engineering. *Nat Biotechnol*. 2009;27:1151-1162.
26. National Human Genome Research Institute. 2004. Near-Term Technology Development for Genome Sequencing (RFA-HG-04-002. Available at: http://grants.nih.gov/grants/guide/rfa-files/RFA-HG-04-002.html. Accessed 30 January, 2010.
27. National Human Genome Research Institute. 2004. Revolutionary Genome Sequencing Technologies -- the $1000 Genome (RFA-HG-04-003. Available at: http://grants.nih.gov/grants/guide/rfa-files/RFA-HG-04-003.html. Accessed January 30, 2010.
28. X PRIZE Foundation. 2006.Archon X PRIZE for Genomics: X PRIZE Foundation Announces Largest Medical Prize in History. Available at: http://genomics.xprize.org/press-release/x-prize-foundation-announces-largest-medical-prize-in-history. Accessed January 30, 2010.
29. Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet*. 2010;11:31-46.
30. Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol*. 2008;26:1135-1145.
31. Ultra-Low Cost Sequencing Technology. Available at: http://arep.med.harvard.edu/Polonator/. Accessed February 5, 2010.
32. Shendure J, Porreca GJ, Reppas NB, et al. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*. 2005;309:1728-1732.
33. Margulies M, Egholm M. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 2005;437:376-380.
34. Applied Biosystems Inc. 2010. ABI product literature. Available at https://docs.appliedbiosystems.com/pebiodocs/00113233.pdf. Accessed January 22, 2010.
35. Applied Biosystems Inc 2010. The ABI SoLID 3 System: Enabling the Next Generation of Science. Available at: http://www3.appliedbiosystems.com/cms/groups/mcb_marketing/documents/generaldocuments/cms_061241.pdf. Accessed January 30, 2010.
36. Bentley DR, Balasubramanian S, Swerdlow HP, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008;456:53-59.
37. Pleasance ED, Cheetham RK, Stephens PJ, et al. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature*. 2010;463:191-196.
38. Pleasance ED, Stephens PJ, O'Meara S, et al. A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature*. 2010;463:184-190.
39. Harris TD, Buzby PR, Babcock H. Single-molecule DNA sequencing of a viral genome. *Science*. 2008;320:106-109.
40. Eid J, Fehr A, Gray J, et al. Real-time DNA sequencing from single polymerase molecules. *Science*. 2009;323:133-138.
41. Drmanac R, Sparks AB, Callow MJ, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science*. 2010;327:78-81.
42. Church GM. The Personal Genome Project. *Mol Systems Biol*. 2005;1:2005.0030.
43. Freimer N, Sabatti C. The human phenome project. *Nat Genet*. 2003;34:15–21.
44. Mahner M, Kary M. What exactly are genomes, genotypes and phenotypes? And what about phenomes? *J Theor Biol*. 1997;186:55–63.
45. Anthony JC, Eaton WW, Henderson AS. Looking to the future in psychiatric epidemiology. *Epidemiol Rev*. 1995;17:240-242.
46. Lederberg J, McCray AT. 'Ome Sweet 'Omics—a genealogical treasury of words. *Scientist*. 2001;15:8.
47. Transcriptome. Available at: http://en.wikipedia.org/wiki/Transcriptome. Accessed January 31, 2010.
48. Proteome. Available at: http://en.wikipedia.org/wiki/Proteome. Accessed January 31, 2010.
49. Ng PC, Levy S, Huang J, et al. Genetic variation in an individual human exome. *PLoS Genet*. 2008;4:e1000160.
50. Kryukov GV, Shpunt A, Stamatoyannopoulos JA, Sunyaev SR. Power of deep, all-exon resequencing for discovery of human trait genes. *Proc Natl Acad Sci U S A.* 2009;106:3871-3876.
51. Ng SB, Turner EH, Robertson PD, et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*. 2009;461:272-276.
52. Angrist M. Eyes wide open: the personal genome project, citizen science and veracity in informed consent. *Pers Med*. 2009;6:691-699.
53. Coriell Cell Repositories at the Coriell Institute for Medical Research. Available at: http://www.coriell.org/. Accessed January 31, 2010.
54. dbGaP: the database of Genotypes and Phenotypes. Available at: http://www.ncbi.nlm.nih.gov/gap. Accessed 31 January, 2010.
55. Merrill RM, Richardson JS. Validity of self-reported height, weight, and body mass index: findings from the National Health and Nutrition Examination Survey, 2001-2006. *Prev Chronic Dis*. 2009;6:A121.
56. Porter SC, Manzi SF, Volpe D, Stack AM. Getting the data right: information accuracy in pediatric emergency medicine. *Qual Saf Health Care*. 2006;15:296-301.
57. Bilder RM, Sabb FW, Cannon TD, et al. Phenomics: the systematic study of phenotypes on a genome-wide scale. *Neuroscience*. 2009;164:30-42.
58. GoogleHealth. Available at: https://www.google.com/health/. Accessed January 31, 2010.
59. MS Healthvault. Available at: http://www.healthvault.com/. Accessed January 31, 2010.
60. Indivo™ The Personally Controlled Health Record. Available at: http://indivohealth.org/. Accessed January 31, 2010.
61. Anthony JC, Eaton WW, Henderson AS. Looking to the future in psychiatric epidemiology. *Epidemiol Rev*. 1995:17;240-242.
62. Butte AJ, Kohane IS. Creation and implications of a phenome-genome network. *Nat Biotechnol*. 2006;24:55-62.
63. Turnbaugh PJ, Gordon JI. A core gut microbiome in obese and lean twins. *Nature*. 2009;457:480–484.
64. Peterson J. The NIH Human Microbiome Project. *Genome Res*. December 2009;19:2317-2323.
65. Sommer MO, Dantas G, Church GM. Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science*. 2009;325:1128-1131.
66. Lefranc MP, Giudicelli V, Ginestoux C, et al. IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Res*. 2009;37:D1006-1012.
67. Number of functional IG and TR genes per haploid genome. Available at: http://www.imgt.org/textes/IMGTrepertoire/LocusGenes/tabgenes/human/geneNumber.html#functional. Accessed 5 February, 2010.
68. Weinstein JA, Jiang N, White RA 3rd, Fisher DS, Quake SR. High-throughput sequencing of the zebrafish antibody repertoire. *Science*. 2009;324:807-810.
69. Park IH, Arora N, Huo H, et al. Disease-specific induced pluripotent stem cells. *Cell*. 2008;134:877-886.
70. Lee JH, Park IH, Gao Y, et al. A robust approach to identifying tissue-specific gene expression regulatory variants using personalized human induced pluripotent stem cells. *PLoS Genet*. 2009;5:e1000718.
71. Collins F S.The case for a US prospective cohort study of genes and environment. *Nature*. 2004;429:475–477.
72. Conley JM, Doerr AK, Vorhaus DB. Enabling responsible public genomics. *Health-Matrix: Journal of Law-Medicine*. In press.
73. Lunshof JE, Chadwick R, Vorhaus DB, Church GM. From genetic privacy to open consent. *Nat Rev Genet*,. 2008; 9:406-411.
74. Overview of PGP consent forms. Available at: http://www.personalgenomes.org/consent/ . Accessed January 31, 2010.
75. Renegar G, Webster CJ, Stuerzebecher S, et al. Returning genetic research results to individuals: points-to-consider. *Bioethics*. 2006;20:24-36.
76. Murphy J, Scott J, Kaufman D, Geller G, LeRoy L, Hudson K. Public expectationsfor return of results from large-cohort genetic research. *Am J Bioeth*. 2008;8:36-43.
77. Wolf SM, Lawrenz FP, Kahn JP, et al. Managing 0*J Law Med Ethics*. 2008;Summer 2008:2-31.

78. Kohane IS, Masys DR, Altman RB. The incidentalome: a threat to genomic medicine. *JAMA*. 2006 12;296:212-215
79. McGuire AL, Burke W. An unwelcome side effect of direct-to-consumer personal genome testing: raiding the medical commons. *JAMA*. 2008;300:2669-2671.
80. Kohane IS, Mandl KD, Taylor PL, Holm IA, Nigrin DJ, Kunkel LM. Reestablishing the researcher-patient compact. *Science*. 2007;316:836-837.
81. Trait-o-matic. Available at: http://snp.med.harvard.edu. Accessed February 4, 2010.

The past, present, and future of direct-to-consumer genetic tests

Agnar Helgason, DPhil; Kári Stefánsson, MD



Technological advances in the field of human genetics have resulted in a wave of discoveries of common DNA sequence variants that are associated with a risk of common complex diseases, such as heart attack, that account for a substantial proportion of morbidity, mortality, and health care costs in most contemporary populations. The overall predictive power of these sequence variants can be considerable, due to the high incidence of these diseases and the sheer number of associations that have been discovered. Health care providers have been slow to utilize this knowledge for preventative medicine. However, several companies have taken on a translational role by offering genetic tests based on these discoveries direct to consumers. In this paper, we review the current state and future prospects of such genetic tests, as scientists involved both in the discovery of disease associations and the development of genetic tests.

 Dialogues Clin Neurosci. 2010;12:61-68.

Keywords: *genetic test; consumer; genome-wide association study; disease risk; ancestry*

Author affiliations: deCODE Genetics, Reykjavik, Iceland (Agnar Helgason, Kári Stefánsson); Department of Anthropology, University of Iceland, Reykjavik, Iceland (Agnar Helgason); Faculty of Medicine, University of Iceland, Reykjavik, Iceland (Kári Stefánsson)

A surge of discoveries in the genetics of disease

The past 4 years have yielded an unparalleled number of discoveries in the field of the genetics of human disease. In particular, huge strides have been made in the discovery of DNA sequence variants that are associated with risk of common complex diseases, such as type 2 diabetes, myocardial infarction, Crohn's disease, breast cancer, and prostate cancer. Prior to 2006, in spite of huge efforts by dedicated researchers, only a few sequence variants had been found to be associated with these diseases that had been adequately verified by replication in well-powered studies. By October of 2009, according to the Catalog of Genome-Wide Association Studies,[1,2] the number of replicated associations of sequence variants with these diseases was 44, 10, 59, 14, and 28, respectively (obtained by counting unique variants reported as statistically significant in more than one study, or in at least two population samples from the same study). *Figure 1* shows the total count of replicated disease associations by year of publication, based on the same criteria, as reported by this database. These numbers will almost certainly continue to rise at a rapid rate. This sudden progress is due to the advent of genome-wide association studies (GWAS), made possible by a combination of at least four key developments. The first is the wealth of knowledge that has been produced about sequence variation in the human genome, in large

Address for correspondence: deCODE Genetics, Sturlugata 8, 101 Reykjavik, Iceland
(e-mail: agnar@decode.is/kstefans@decode.is)

part due to the Human Genome Project[6] and the ensuing HapMap project.[7] The second is the development of high-throughput microarray genotyping technologies, that now enable researchers to simultaneously and relatively cheaply assess genotypes at hundreds of thousands of single-nucleotide polymorphisms (SNPs) across the genome. The third is the availability of collections of DNA samples from large numbers of individuals who suffer from the diseases of interest and controls from the same populations. Another factor that has contributed to the success of GWAS is the close and fruitful cooperation between research groups and journals in defining conservative and robust standards for the verification of disease association signals obtained using this approach. The recent discoveries of sequence variants associated with the risk of complex diseases represent an important step in the task of understanding their biology, of which we are still remarkably ignorant. While some of the newly discovered associations were found in genes already suspected of playing a role in etiology, most are in, or close to, genes with no prior connection to the disease in question. These latter discoveries, in particular, represent important new points of departure for more focused research into the biology and etiology of these diseases.

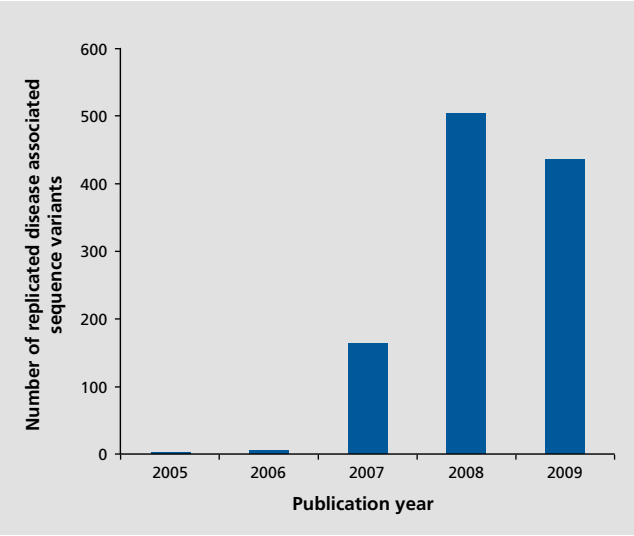


Figure 1. The number of replicated sequence variants associated with diseases and medically relevant traits by publication year of first report in genome-wide association studies according to the Catalog of Genome-Wide Association Studies on October 20th 2009. Counts were obtained by counting unique variants reported as statistically significant (at the level of $P < 1.0 \times 10^{-5}$) in more than one study or in at least two population samples from the same study. They represent a total of 1108 replicated associations to sequence variants based on 132 diseases and medically relevant traits. Note that in some cases the Catalog of Genome-Wide Association Studies reports multiple correlated single-nucleotide polymorphisms (SNPs) from the same genomic region, and thus the numbers shown are likely to be overestimates. However, the 1108 associated SNPs belong to 871 different genomic regions, and this latter number is likely to be an underestimate of the overall number of associations discovered. This is because some regions are known to contain multiple independent associations to the same trait—for example, region 8q24 and prostate cancer.[3] Also, not all bona fide disease-associated sequence variants are included in the catalog, for example several that have been reported for age-related macular degeneration.[4,5]

While the discovery rate of new disease-associated variants shows no signs of decline, there is good reason to believe that much of the lowest-hanging fruit has already been picked. These are the common sequence variants that have an easily detected impact on disease risk, given the existing sample sizes of cases and controls (ie, with an odds ratio of more than 1.1) and that are covered by the existing microarray genotyping platforms. Some researchers argue for continuation of the GWAS approach, with larger sample sizes to detect more common variants with small effect.[8] Others argue for a change of strategy, pointing out that the combined effects of variants that are likely to be found with more GWAS only account for a part of the overall heritability of the diseases concerned.[9] Proposals have been made to pay greater attention to rare variants, copy number variants, epigenetic factors, or epistatic effects between unlinked sequence variants. At least some of these aims will be achieved in the near future, as further technological developments make full genome sequencing and more comprehensive microarray genotyping platforms realistic options for large-scale disease studies.

Translation of disease association findings for public use

Clearly, there is more to be found, and it seems obvious to us that all of the aforementioned lines of research should be pursued. However, at the same time as geneticists continue their hunt for new disease-associated sequence variants and attempt to determine the functional relevance of the variants they have already discovered, they must address an equally pressing issue of practical concern in relation to existing knowledge. To date, more than 1000 sequence variants have been discovered with robustly verified disease associations to

tens of major complex diseases.[1] These diseases account for a substantial proportion of morbidity, mortality, and health care costs in most contemporary populations. Even though the functional impact of most of these variants is not yet understood, their association with disease imbues them with intrinsic value for both the general public and health care practitioners. This is because the strength of their association in a population reflects the degree to which they can be used to improve predictions of disease risk for individuals.

Disease risk stratification (by age, sex, weight, and other biological markers) forms the cornerstone of effective screening programs. It also serves an important role in increasing the health awareness of individuals, thereby promoting the adoption of preventative measures and leading to earlier diagnosis. The overall result is not only a healthier population, but a reduction in the burgeoning cost of health care, much of which stems from late treatment of preventable diseases. It follows that considerable additional health benefit and cost-effectiveness could be achieved through the addition of recently discovered genetic factors. Although such genetic tests are available, they have yet to be routinely adopted by major health care providers. This may be partly due to a lack of familiarity with such tests, or with their scientific basis, or due to an inherent resistance that stems from financial or organizational concerns. However, genetic tests are also available directly to the public through the Internet, where they have been positively received through a combination of health concerns, curiosity, and recreational interest in genetics and ancestry.

The authors of this article are members of a research team at deCODE Genetics that has contributed to the ongoing wave of discoveries in the genetics of complex disease and intends to remain at the forefront of this field in the coming years. However, from 2007 deCODE Genetics has also been a leader in using this knowledge to develop tests to evaluate genetic risk, both in the form of diagnostic products aimed at health care providers, and a personal genome scan, under the name of deCODEme, which is sold directly to consumers via the Internet. In what follows, we will outline the nature of the genetic tests provided by deCODE Genetics and others, and the value we believe they can bring by informing individuals about their health prospects and motivating them to take preventative measures where possible or to seek early diagnosis. We will also discuss some of the concerns raised by such genetic tests.

The nature of the genome and possibilities of genetic testing

Before moving on to a description of the kinds of genetic tests that are currently available to health care providers and the general public, it may be helpful to consider the nature of the genome. This can help us to understand what kind of information the genome can offer and thereby what kind of genetic tests are possible. The genome constitutes an extremely complex set of instructions for the assembly and maintenance of an organism, within some normal range of environmental conditions. For humans these instructions are almost exclusively encoded in the sequence of the roughly 3 billion nucleotides that make up the genome.

We may consider a human being as a vast collection of phenotypic traits, ranging from, for example, height and skin pigmentation to less perceptible features such as blood insulin levels or the build-up of amyloid plaques in brain tissue. All such traits are the outcome of an interaction between instructions from one or more parts of the genome and some set of environmental factors. Most phenotypic traits exhibit some variation among individuals that reflects underlying differences in DNA sequence and differences in exposure to environmental conditions. In some cases, differences between individuals exposed to normal environmental conditions are solely due to DNA sequence variants from a single gene. An example of a trait that is fully determined by sequence variants, and is inherited in accordance with simple Mendelian rules of inheritance, is the capacity to metabolize the amino acid phenylalanine, that when lacking, results in the disease phenylketonuria. More often, however, trait variation among individuals can be traced to many DNA sequence variants and environmental factors.

The power to correctly predict traits such as the development of disease for individuals using a genetic test (that is, the clinical validity of the test) depends on the nature of the relationship between genotype and phenotype. Many of the key human diseases, the so-called common complex diseases, are substantially affected by environmental factors. This means that the predictive power of genetic tests for these diseases will be less than for "simple" traits such as phenylketonuria (although the validity for such tests could be boosted by including known environmental risk factors). Nonetheless, the potential health and economic rewards gained from

improving risk predictions for diseases that affect large numbers of individuals in a population are substantial. No matter how many sequence variants and environmental factors contribute to a given phenotypic trait, all other things being equal, the accuracy of prediction is always increased by the inclusion of just one truly associated sequence variant.

Diseases may be defined as the fraction of variation in physiological function that lies outside the normal range, such that either the quality of life is impaired, or the probability of untimely death is raised to an unacceptable level. It is no coincidence that diseases are the focus of most existing genetic tests, because they have been the primary focus to date of research into genotype-phenotype associations in humans. However, once reliable information has been gathered about an association between any phenotypic trait and a set of sequence variants, it becomes possible to develop a genetic test to estimate genetic propensity for that trait. In fact, there are already several genetic tests on the market that include nondisease traits such as eye color, male-pattern baldness, bitter taste perception, and drug metabolism. Such traits are often ignored by commentators on genetic testing, but they are likely to play a larger role in this field as our understanding of the functional impact of sequence variants on normal human phenotypic variation advances.

In addition to enabling the assessment of genetic predispositions to particular phenotypic traits, the genome holds a record of ancestry and genealogical relationships between all people. This information is inscribed into the genome as it is replicated and transmitted from parents to offspring with a small number of changes in the form of mutations and recombination events. In the world of consumer genetic tests, ancestry has been a leading area of interest, with considerable sales of tests based on the uniparentally inherited genetic material from mitochondria and Y chromosomes.[10]

The use of genetic markers to verify the existence of close family relationships between individuals is a relatively trivial task that is routinely performed in forensic laboratories and in tests that are already available to the public. The introduction of microarray genotyping platforms with hundreds of thousands of SNPs is likely to facilitate the development of more powerful algorithms to explicitly test for more distant genealogical relationships between individuals. Such data are already used in genetic tests in conjunction with recently developed statistical methods from population genetics to provide detailed assessments of ancestry and admixture. The results from these analyses are in effect summary analyses of the genealogical relationship of an individual to different populations from around the world. As the magnitude of comparative data from such populations grows and the number of sequence variants assessed increases (for example, through full genome sequencing), there will be considerable improvements in the detail and accuracy of ancestry assessments and genealogical testing.

Genetic tests of ancestry are typically defined as recreational by commentators, and somehow qualitatively different from tests that evaluate disease risk.[11,12] However, it is important to bear in mind that in the genome, information about function and ancestry is inexorably intertwined. Sequence variants that are used in an ancestry test today, on the basis of having no known function, may well be found to be associated with a disease or medically relevant traits tomorrow. Moreover, to the extent that disease risks vary between populations due to differing frequencies of the underlying associated sequence variants, it follows that tests of ancestry are in effect tests, albeit low-powered tests, of genetic risk of disease.

The past and present of genetic testing

At present, a wide range of genetic tests are available either through health care providers or by companies direct to the consumer (DTC). Indeed, there are now so many that it is impossible to provide a comprehensive overview of them all. To begin with, most tests were performed by university or hospital laboratories within the scope of health care provision. They were based on rare and highly penetrant sequence variants that are strongly associated with a particular disease. Included in this category are tests used for prenatal and newborn screening, diagnostic testing for chromosome abnormalities, carrier testing, and predictive testing for particular conditions such as Huntington's disease and hemochromatosis. In time, companies offering such tests emerged, in some cases established by people from the aforementioned universities or hospitals. An early example is the company Myriad Genetics, that patented and marketed predictive tests for breast cancer based on variants in the BRCA genes (that confer a roughly fivefold risk of developing this disease).

One thing the tests available through health care providers have in common is that the variants tested are rare and highly penetrant (ie, their clinical validity is high). Consequently, very few individuals from the general population would be expected to receive positive results from such tests—as is the case, for example, in population screening for phenylketonuria mutations. In many cases, however, such tests are provided on the basis of clinical diagnosis or familial risk, which increases the fraction of positive results from the tests. For individuals who receive positive results, the implications tend to be a very high probability of disease. Thus, test-takers often meet a genetic counselor prior to tests and after in light of a positive result (with treatment if applicable). The use of tests for rare and highly penetrant sequence variants is widespread among health care providers in most countries. Many have also adopted predictive tests for breast cancer (using variants in the BRCA gene), and Alzheimer's disease, based on more common variants in the APOE gene that confer a fourfold risk. However, they have been slow or reluctant to take advantage of the recent wave of robustly replicated GWAS discoveries of variants associated with increased disease risk ranging from 1.05-fold to 7-fold.

In spite of the reluctance of health care providers to adopt genetic tests for common diseases, a growing number of companies have been harnessing findings from GWAS and other genetic studies to design tests that are sold DTC, mainly through the Internet. Most currently available DTC tests are based on a handful of sequence variants and focus on a specific application, such as ancestry, family relationships, or the testing of genetic risk for particular diseases. Often a particular company will offer several such small tests covering one or more of these areas. On top of this, a few companies are now offering DTC personal genome scans. These products are mostly based on large-scale and cost-effective microarray genotyping platforms that test 500 000 to 1 million SNPs from across the entire genome, but in a few cases on more expensive full-genome sequencing. The four best-known providers of DTC personal genome scans are, in order of appearance on the market, deCODEme (the product the authors are involved with), 23 and me, Navigenics, and Knome. Such products typically provide consumers with estimates of risk or predispositions to many diseases or traits, in addition to a range of tests of ancestry and family relationships through a secure personal Web site account. Aside from the breadth of tests that such a large number of SNPs can offer, one advantage of the personal genome scans for consumers is that they can (at least in principle) obtain updated estimates for all tests as new discoveries are made. Thus, the initial purchase of the test may be viewed as an investment that yields interest in the form of accumulating knowledge from new discoveries. To date, the companies providing DTC personal genome scans have been fairly active in updating the tests offered on their Web sites. In addition, customers can download their genotypes onto their own computers and analyze the data for themselves (for example, by uploading genotypes on Web sites such as SNPedia or seeking advice on layman Web sites such as DNA-forums.org).

Addressing the concerns of some scientists about DTC genetic tests

Many of the recent commentaries about DTC genetic testing in the scientific literature have focused on the personal genome scans[11-17] and in particular on concerns about the disease risk estimates they provide to customers. These concerns may be divided into two main themes.

First, there are questions about the validity of the tests. The most common criticism here is that because the risk conferred by each variant used in these tests tends to be low (odds ratio <2), then the accuracy in predicting risk of disease for individuals (ie, their clinical validity) will also be low.[11-13,15,18] Such comments either explicitly or implicitly compare the new tests with older tests that are already in use by health care providers in a way that is highly misleading. If we take a multifactorial disease such as heart attack, then it is self-evident that one cannot design a genetic test with predictive power as great as that for Huntington's disease, which is rarer, solely caused by mutations in one gene, and is negligibly affected by environmental factors. However, it is wrong to think that the predictive power of genetic tests based on GWAS findings is insignificant and not clinically relevant. Take, for example, the test for heart attack in the deCODEme personal genome scan, which includes eight independent SNPs with strong evidence for association to this disease[19-23] at the time of writing. This test alone can allow for the identification of 10% of people of European decent who have at least 1.4 times greater risk of developing a heart attack than the average person in the population. The average relative risk in this group is

1.6. Since the lifetime risk of suffering a heart attack is 42% for men over the age of 40,[24] it follows that this test can identify men who have, on average, a lifetime risk of 67% of developing the most lethal disease of man.[25] There is a dramatic difference between 42% and 67% lifetime risk. For example, individuals who are in the top quintile of the concentration of LDL cholesterol have only 1.3 times the population average risk of having a heart attack. Hence, a genetic test based on recently discovered sequence variants can identify people with added risk of heart attack that is twice that of those who are at the top of the cholesterol curve. In this context, it is important to keep in mind that the ability to assess risk of the heart attack by measuring serum cholesterol has transformed cardiology into the most important field of preventive medicine. Given the rather high lifetime risk of heart attack in most industrialized populations, then it follows that even a small increase in predictive power for each individual can be valuable.

Comparable estimates of average relative risk in the top 10% of people for other diseases included in the deCODEme genome scan are, for example, 5.2 for age-related macular degeneration, 1.8 for type 2 diabetes and 3.0 for Crohn's disease. There is good reason to believe that the predictive power of these genetic tests will increase in the near future. On the one hand, we expect additional associated sequence variants to be discovered. On the other hand, because risk estimates in current tests are typically only stratified by sex and ethnicity, advances can be made through the inclusion of other relevant background variables (for example, the waist-to-hip ratio and smoking history in the case of heart attack). In both cases, further epidemiological research is needed. However, contrary to the views of some commentators,[11,13] these are not grounds for delay. The value of the discoveries so far, as reflected in the aforementioned example of heart attack, unequivocally warrants their use in genetic tests.

The second main theme of concern raised by commentators relates to the capacity of consumers to understand and cope with disease risk estimates from tests.[8,9,12-14] Underlying these concerns is a somewhat paternalistic and patronizing view that information about disease risk is dangerous to the general public unless mediated in person by medical experts. Among the alleged dangers to the public are anxieties from overinterpreting risk estimates, which could lead to increased demands on health care providers and unnecessary medical procedures. We are not aware of any evidence that has been reported in support of this view, but there is at least some indirect evidence against it.[26]

However, even if such information were to provoke anxiety, the right of regulators or medical experts to prevent access to it is questionable. Consider the following analogy of a hypothetical company that provides individuals with reliable estimates of their risk of being mugged, murdered, run over, or burgled based on their age, sex, and address. There is no question that someone living in a rough inner-city area with limited economic means would have considerably greater risk on all counts than people living in more affluent areas. Would this be a legitimate reason for preventing individuals from seeking such potentially anxiety-provoking information?

Another issue raised by commentators is that of clinical utility,[10,13] that is, the extent to which knowledge of increased risk can reduce the burden of a disease through prevention or treatment. Although frequently raised in discussions of DTC genetic tests, this issue is really only relevant within the scope of health care provision (for example in the case of Huntington's disease). Thus, for DTC genetic tests, clinical utility is a secondary issue when balanced against peoples' right to seek information about themselves at their own cost. Given that such tests are in accordance with the accepted scientific literature and adhere to consumer laws (ie, that they deliver what their providers promise), then it is hard to see how regulators could prevent the public from buying them.

The challenge for providers of DTC genetic tests

In our opinion, the key to the success of DTC genetic tests for consumers, the companies that provide them, and regulators, is clarity and transparency. Whether tests report disease risk estimates, ancestry analyses, or evaluation of genealogical relationships, the information used to motivate consumers to buy tests and then explain the results should be as clear and accurate as possible. In particular, the probabilistic rather than deterministic nature of disease risk estimation must be unambiguous and comprehensible to the layman (and to medical experts). A key task is also to use the scientific literature in an accurate and responsible manner, for example by including only sequence variants with associations to disease that have been robustly replicated.

One way to uphold such standards is through transparency, ie, by providing information about all the sequence variants used and the parameter values for risk models and their sources in the scientific literature. Most of the current providers of DTC personal genome scans have followed this approach, to a greater or lesser extent.

If such basic ground rules are adhered to, we believe that DTC genetic tests can provide considerable value to the general public, in particular while tests based on disease-associated variants discovered through GWAS are not available through health care providers. From a public health perspective, there is real preventative value to be gained from making people aware of their health and the risks posed to it. It is true that many of the lifestyle changes recommended for prevention through the application of DTC genetic tests could benefit many individuals, regardless of whether they take such tests. However, as most health care providers know, people are generally reluctant to change their lifestyle, even in the face of stern advice from medical experts. We would argue that when genetic risk factors are added on to conventional lifestyle risk factors in motivating people to take preventative measures, the outcome provides a greater impetus to act. Of course, from the perspective of personal autonomy, even if people choose to disregard advice about disease prevention, their right to seek information about genetic risk should prevail.

It is also important to highlight the educational nature of the Web sites of many companies that offer DTC genetic tests. They usually contain detailed information on hundreds of Web pages about diseases, ancestry, and genetic discoveries and methods that are used to provide results. This information is typically available to anyone through various front-end Web pages, where potential buyers can explore the kind of information they would receive as customers. Anyone can therefore learn a great deal about diseases, ancestry, and genetics without paying for a test. Whether the decision to buy a test is motivated by health concerns, recreational curiosity, or vanity, the consumer is almost certain to gain not only an increased understanding of genetics in general, but also what the recent wave of discoveries in the human genetics of disease and ancestry mean for them personally.

Conclusion

We believe that DTC genetic tests play a key translational role for the science of genetics, democratizing and disseminating privileged knowledge to the public. No matter how clichéd it sounds, knowledge is power. While some medical experts may complain about patients armed with results from DTC genetic tests or information about disease symptoms from the internet,[13] we believe that a knowledgeable public is an empowered public.

REFERENCES

1. Hindorff LA, Junkins HA, Mehta JP, Manolio TA. A Catalog of Published Genome-Wide Association Studies. Available at: www.genome.gov/gwastudies. Accessed November 2009.
2. Hindorff LA, Sethupathy P, Junkins HA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*. 2009;106:9362-9367.
3. Gudmundsson J, Sulem P, Manolescu A, et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet*. 2007;39:631-637.
4. Gold B, Merriam JE, Zernant J, et al. Variation in factor B (BF) and complement component 2 (C2) genes is associated with age-related macular degeneration. *Nat Genet*. 2006;38:458-462.
5. Maller JB, Fagerness JA, Reynolds RC, Neale BM, Daly MJ, Seddon JM. Variation in complement factor 3 is associated with risk of age-related macular degeneration. *Nat Genet*. 2007;39:1200-1201.
6. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409:860-921.
7. Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P. A haplotype map of the human genome. *Nature*. 2005;437:1299-1320.
8. Hirschhorn JN. Genomewide association studies--illuminating biologic pathways. *N Engl J Med*. 2009;360:1699-1701.
9. Goldstein DB. Common genetic variation and human traits. *N Engl J Med*. 2009;360:1696-1698.
10. Bolnick DA, Fullwiley D, Duster T, et al. Genetics. The science and business of genetic ancestry testing. *Science*. 2007;318:399-400.
11. Evans JP, Green RC. Direct to consumer genetic testing: avoiding a culture war. *Genet Med*. 2009;11:568-569.
12. Farkas DH, Holland CA. Direct-to-consumer genetic testing: two sides of the coin. *J Mol Diagn*. 2009;11:263-265.
13. Hunter DJ, Khoury MJ, Drazen JM. Letting the genome out of the bottle--will we get our wish? *N Engl J Med.* 2008;358:105-107.
14. Kaye J. The regulation of direct-to-consumer genetic tests. *Hum Mol Genet*. 2008;17(R2):R180-R183.
15. Lenzer J, Brownlee S. Knowing me, knowing you. *BMJ*. 2008;336:858-860.
16. Rogowski WH, Grosse SD, Khoury MJ. Challenges of translating genetic tests into clinical and public health practice. *Nat Rev Genet*. 2009;10:489-495.
17. Henrikson NB, Bowen D, Burke W. Does genomic risk information motivate people to change their behavior? *Genome Med*. 2009;1:37.
18. Kraft P, Hunter DJ. Genetic risk prediction--are we there yet? *N Engl J Med*. 2009;360:1701-1703.
19. Helgadottir A, Thorleifsson G, Manolescu A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science*. 2007;316:1491-1493.
20. Kathiresan S, Voight BF, Purcell S, et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet*. 2009;41:334-341.

Pasado, presente y futuro de las pruebas genéticas dirigidas a los usuarios

Los avances tecnológicos en el campo de la genética humana han producido una ola de descubrimientos de variantes de la secuencia común del ADN que están asociadas con un riesgo de enfermedades comunes y complejas, como el ataque cardíaco, que dan cuenta de un porcentaje significativo de morbilidad, mortalidad y costos de salud en la mayoría de las poblaciones actuales. El poder predictor en conjunto de estas variantes de la secuencia puede ser considerable, debido a la alta incidencia de estas enfermedades y al reducido número de asociaciones que se han descubierto. Los proveedores de atención de salud han utilizado lentamente este conocimiento para la medicina preventiva. Sin embargo, algunas empresas han asumido un papel translacional al ofrecer pruebas genéticas basadas en estos descubrimientos dirigidas a los usuarios. En este artículo se revisa tanto el estado actual y las perspectivas futuras de tales pruebas genéticas, como a los científicos involucrados en el descubrimiento de las asociaciones de enfermedades y en el desarrollo de las pruebas genéticas.

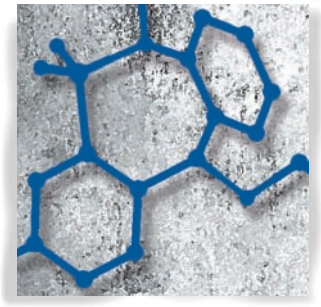
Les tests génétiques en vente libre : passé, présent, futur

Les avancées technologiques en génétique humaine ont permis une série de découvertes de variants fréquents de séquence d'ADN associés à un risque de maladies complexes courantes, comme la crise cardiaque, responsables d'une morbi-mortalité et de coûts liés à la santé très importants chez la plupart de nos contemporains. La puissance prédictive globale de ces variants génétiques peut être considérable en raison de la fréquence élevée de ces maladies et du seul nombre d'associations découvertes. Les acteurs du système de santé ont mis du temps à utiliser ces données en médecine préventive. Plusieurs laboratoires ont néanmoins accepté de servir d'intermédiaire en proposant directement au grand public des tests génétiques basés sur ces découvertes. En tant que scientifiques impliqués à la fois dans la découverte d'associations de maladies et dans le développement de tests génétiques, nous analysons dans cet article l'état actuel et le devenir de tels tests.

21. Erdmann J, Grosshennig A, Braund PS, et al. New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nat Genet*. 2009;41:280-282.
22. Samani NJ, Erdmann J, Hall AS, et al. Genomewide association analysis of coronary artery disease. *N Engl J Med*. 2007;357:443-453.
23. Gudbjartsson DF, Bjornsdottir US, Halapi E, et al. Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. *Nat Genet*. 2009;41:342-347.
24. Lloyd-Jones DM, Larson MG, Beiser A, Levy D. Lifetime risk of developing coronary heart disease. *Lancet*. 1999;353:89-92.
25. *The Global Burden of Disease - 2004 Update*. Geneva, Switzerland: World Health Organization; 2008.
26. Cassidy MR, Roberts JS, Bird TD, et al. Comparing test-specific distress of susceptibility versus deterministic genetic testing for Alzheimer's disease. *Alzheimers Dement*. 2008;4:406-413.

Psychiatric pharmacogenomic testing in clinical practice

David A. Mrazek, MD, FRCPsych



The clinical adoption of psychiatric pharmacogenomic testing has taken place rapidly over the past 7 years. Initially, drug-metabolizing enzyme genes, such as the cytochrome P450 2D6 gene (CYP2D6), *were identified. Genotyping the highly variable cytochrome P450 2D6 gene now provides clinicians with the opportunity to identify both poor metabolizers and ultrarapid metabolizers of 2D6 substrate medications. Subsequently, genes influencing the pharmacodynamic response of medications have been made available for clinical practice. Among the earliest "target genes" was the serotonin transporter gene* (SLC6A4) *which has variants that have been shown to influence the clinical response of patients of European ancestry when they are treated with selective serotonin reuptake inhibitors. Genotyping of some of the serotonin receptor genes is also available to guide clinical practice. The quantification of the clinical utility of pharmacogenomic testing is evolving, and ethical considerations for testing have been established. Given the increasingly clear cost-effectiveness of genotyping, it has recently been predicted that pharmacogenomic testing will routinely be ordered to guide the selection and dosing of psychotropic medications.*

 Dialogues Clin Neurosci. 2010;12:69-76.

Clinical pharmacogenomics consists of the application of research that links measurable genetic variants with the prediction of drug response.[1] Every medical specialty can utilize the results of pharmacogenomic probe studies to inform the adoption of individualized pharmacotherapy. However, psychiatric pharmacotherapy is particularly likely to benefit from the introduction of pharmacogenomic testing, because there are many psychotropic agents available for selection that target specific symptoms.

The terms pharmacogenetics and pharmacogenomics are currently used interchangeably. However, with the growing understanding that multiple intragenic variations should be considered in making predictions related to medication response, the use of the term pharmacogenomics has become more frequently chosen to designate the process of using documented genetic variation to guide medication selection and dosing.

Historically, psychiatrists have used empirical strategies to select medications. In the best practices, the choice of medications has evolved based on a rational trial-and-error process that has used clinical indicators to select medications and then relied on documenting treatment responses to titrate the optimal dose for a particular patient. Psychiatrists learn to "start low and go slow" in order to minimize side effects. They also know that it is

Keywords: *pharmacogenomic testing; cytochrome P450 gene; serotonin transporter gene* (SLC6A4)*; serotonin receptor 2A gene* (HTR2A)*; serotonin receptor 2C gene* (HTR2C)*; poor metabolizer phenotype; ultrarapid metabolizer phenotype*

Author affiliations: Chair, Department of Psychiatry and Psychology, Mayo Clinic; Professor of Psychiatry and Pediatrics, Mayo Clinic College of Medicine, Rochester, Minnesota, USA

Address for correspondence: David A. Mrazek, Mayo Clinic College of Medicine, 200 First Street SW, Rochester, MN 55905, USA
(e-mail: mrazek.david@mayo.edu)

necessary to provide their patients with an "adequate" trial of each medication. Unfortunately, these strategies can result in a 3- to 4-week interval during which the patient continues to experience symptoms. In recent years, the potential iatrogenic harm associated with psychotropic medications has become increasingly obvious, with "black-box warnings" being attached to antidepressants, antipsychotic medications, stimulants, and mood stabilizers.

Despite a growing awareness of this potential harm, there are powerful pressures to try to accelerate the achievement of therapeutic benefit. At the most basic level, patients are impatient. They do not want to wait a month to achieve symptom relief. Additionally, with an increasing focus on the relief of specific symptoms, strategies using multiple psychotropic medications have become a standard of practice. Research supports the common practice of augmenting an initial medication with a second psychotropic drug.[2] However, there is no scientifically available evidence to support the practice of using four or five psychotropic medications simultaneously. Nevertheless, patients routinely receive multiple psychotropic medications in an attempt to identify the "right combination." While some patients do achieve a good therapeutic response using this trial-and-error approach to individualized medicine, it is also true that others become overmedicated or suffer from iatrogenic side effects.

Pharmacogenomic testing provides an innovative strategy to improve the likelihood of selecting an effective psychotropic medication. The earliest medical texts recognize that individual patients experience quite dramatically different responses to the same drug. There is also a longstanding observation that unusual drug responses can occur in members of the same family. The identification of specific gene variants associated with idiosyncratic responses is about 50 years old,[3] and the recognition that some psychiatric patients metabolize antidepressants at dramatically different rates has been documented for several decades.[4] However, with the use of newer antidepressant medications that rarely have life-threatening complications, the relatively expensive practice of monitoring the serum levels of newer antidepressant medications has become uncommon in the United States. This change has occurred despite the fact that serum levels of these newer agents also have dramatic variations based on the metabolic capacity of each patient.

A decade ago, the cost of genotyping began to become more affordable, and individual laboratories initiated pharmacogenomic testing that would provide genotyping of individual cytochrome P450 genes. However, there was no standard or well-validated methodology for the genotyping of these informative genes. There was also considerable variability in the interpretation of the results. In 2004, the US Food and Drug Administration (FDA) approved the use of a new product, the AmpliChip.[5] The introduction of the AmpliChip provided reference laboratories with a standard method for identifying variations in two of the cytochrome P450 genes: cytochrome P450 2D6 *(CYP2D6)* and cytochrome P450 2C19 *(CYP2C19)*. The approval of the AmpliChip was an important landmark in the history of psychiatric pharmacogenomic testing, and within 3 years, *CYP2D6* and *CYP2C19* were being genotyped by every reference laboratory in the country. However, this advance also highlighted some of the challenges associated with the introduction of clinical testing. One of the most obvious challenges that must be addressed is how to begin to assess new variants of these two genes in updated versions of the assay. Ideally, the methodology for establishing drug-metabolizing phenotypes should be updated regularly based on new molecular genetic findings showing how new genotypic variants influence gene function. Also, the clarification of the predictive capacity of previously identified gene variants influencing gene function is similarly evolving, and newly identified associations between gene structure and function should ideally be incorporated into algorithms that define the metabolic capacity of psychiatric patients.

The evolution of pharmacogenomic research should inform modifications in pharmacogenomic testing. However, an implication of the rapid increase in our knowledge base is that these new studies demonstrate limitations in the accuracy of older genotyping methodologies that were designed prior to the discovery of more recent variants. What is often not well appreciated is that even older pharmacogenomic methods provided important information for many patients, as these early innovations were a major advance over psychopharmacological practice without pharmacogenomic insights. However, as newer methodologies have further improved the accuracy of the prediction of medication response, the clinical utility of pharmacogenomic testing continues to increase.

Pharmacogenomic testing in psychiatric practice initially focused on identifying pharmacokinetic variability that

would influence the responses of patients who had atypical genotypes. Pharmacokinetic variation influences the concentration of a drug at its sites of action. Pharmacogenomic testing of drug-metabolizing enzyme genes provides a prediction of how an individual patient will metabolize a specific psychotropic medication. More recently, the focus of pharmacogenomic testing has expanded to include determining variability in the pharmacodynamic response of a patient to a specific medication. This variability reflects the capacity of the individual patient to respond to adequate exposure to the drug. Prediction of response is estimated based on the documentation of variations in "target genes" that code for receptors and transporters that influence the response of the patient to a particular medication.

This review will first identify the most widely genotyped drug-metabolizing enzyme genes that influence the pharmacokinetic metabolic capacity of a patient. Then, it will focus on genes that influence the pharmacodynamic responses of individual patients, before concluding with a brief discussion of the clinical utility of pharmacogenomic testing and some of the ethical considerations related to its routine use.

Pharmacogenomic testing to establish the metabolic capacity of psychiatric patients

Many genes code for enzymes that influence drug response. However, only the clinical implications of genotyping four of the most commonly tested cytochrome P450 genes will be reviewed. The focus of this discussion will be the clinical benefit for the patient of identifying individualized molecular variations, and the implications for those patients who have a quite significant decrement in their capacity to metabolize specific psychotropic medications. Identifying these individual patients provides clinicians with a clear method of minimizing side effects. This determination of decreased metabolic capacity is the most obvious benefit of pharmacogenomic testing, but implications of the pharmacogenomic testing for patients with increased metabolic capacity will also be discussed, as these patients are less likely to respond to specific psychotropic medications.

The cytochrome P450 2D6 gene *(CYP2D6)*

CYP2D6 was the first drug-metabolizing enzyme gene that was genotyped to identify psychiatric patients with increased or decreased metabolic capacity. It is located on chromosome 22 and consists of 4382 nucleotides. *CYP2D6* codes for an enzyme that is composed of 497 amino acids.

The CYP2D6 enzyme plays a primary role in the metabolism of more than 70 substrate medications, including twelve psychotropic medications. *CYP2D6* is one of the most highly variable drug-metabolizing enzyme genes. However, many of the other 29 P450 drug-metabolizing enzyme genes are also highly variable. The specific genetic variations that define variable phenotypes can be located on a Web site maintained by the Karolinska Institute (http://www.cypalleles.ki.se/). Each newly identified variant is included on the Web site after confirmation that it is unique.

There are currently 75 distinct *CYP2D6* alleles posted on this site, as well as an additional 55 *CYP2D6* variants that closely resemble one of the primary variants. Traditionally, these variants have been classified as being normal, deficient, or inactive drug-metabolizing alleles. Additionally, some alleles have more recently been demonstrated to code for an increased amount of enzyme which enhances the metabolic activity of the patient. Furthermore, patients can have a variable number of copies of *CYP2D6*. The most common number of copies of *CYP2D6* that patients carry is two. However, some patients have only one copy and, rarely, none at all. It is also possible to have more than two copies, and one patient has been reported to have 13 copies.[4] The development of several different classification systems to categorize 2D6 substrate metabolic capacity of patients into four phenotypic categories has been problematic. The use of alternative methodologies by different research teams has made it more difficult to study the implications of this variability.

The most important *CYP2D6* phenotype to identify is the poor 2D6 substrate metabolizer phenotype. Patients who are poor metabolizers are at increased risk for adverse events when they are prescribed 2D6 substrate medications, because of their low metabolic capacity. Patients are now classified as poor metabolizers if they have two inactive alleles, or one inactive allele and one deficient allele.

The second most clinically important *CYP2D6* phenotype is the ultrarapid metabolizer phenotype. Patients are ultrarapid metabolizers if they have either three or more active copies of *CYP2D6* or two or more enhanced copies of *CYP2D6*. They are unlikely to respond to 2D6

substrate medications at standard doses because their ability to rapidly metabolize these medications makes it difficult to sustain therapeutic serum levels.

The third clinically important *CYP2D6* phenotype is the intermediate metabolizer phenotype. These patients have one normal copy of *CYP2D6*, and one copy that is either deficient or inactive. While these patients can normally benefit from 2D6 substrate medications at low-to-moderate doses, they are at increased risk for the development of side effects at higher doses because of their decreased metabolic capacity, and they are more at risk for enzyme inhibition as a consequence of drug-drug interactions. When intermediate metabolizers are exposed to powerful 2D6 inhibitors such as paroxetine or fluoxetine, their metabolic capacity can be further decreased to the level of a poor metabolizer.[6]

There are many psychotropic medications metabolized by the 2D6 enzyme. Specifically, this enzyme:

- primarily metabolizes five antidepressants: fluoxetine, paroxetine, venlafaxine, desipramine, and nortriptyline
- substantially metabolizes amitriptyline, imipramine, doxepin, duloxetine, trazodone, and mirtazapine
- primarily metabolizes risperidone and four of the typical antipsychotic medications: chlorpromazine, thioridazine, perphenazine, and haloperidol
- has substantial involvement in the metabolism of aripiprazole and olanzapine
- primarily metabolizes atomoxetine and dextroamphetamine.

Beyond the prescription of psychotropic medications, psychiatric patients are given many other 2D6 substrate medications. Specifically, dextromethorphan is a cough suppressant that is metabolized by the 2D6 enzyme. Patients who are poor metabolizers of 2D6 substrate medications are at increased risk for cognitive side effects if taking standard doses of preparations that contain dextromethorphan. Another example is codeine, which is a prodrug. A prodrug must be converted to an active metabolite in order to have a therapeutic effect. Patients who are poor 2D6 metabolizers do not receive analgesic benefit from codeine because they do not metabolize codeine to morphine. Tamoxifen is also a prodrug that is the most frequently prescribed treatment for breast cancer. Poor metabolizers have little or no benefit from tamoxifen because they are not able to metabolize tamoxifen to endoxifen.[7,8] Additionally, paroxetine, fluoxetine, or bupropion should not be given to patients who are receiving tamoxifen because they inhibit the 2D6 enzyme. Giving these inhibitors to intermediate metabolizers can convert them to functional poor metabolizers. Consequently, they become unable to produce endoxifen.[9]

The cytochrome P450 2C19 gene *(CYP2C19)*

CYP2C19 was the second drug-metabolizing enzyme gene that was widely genotyped to identify patients with increased or decreased metabolic capacity. It is a large gene located on chromosome 10. It consists of 90 209 nucleotides, but codes for an enzyme that contains only 490 amino acids.

The identification of patients with low 2C19 metabolic capacity is clinically important because it allows clinicians to decrease the risk of iatrogenic side effects.

The 2C19 enzyme:

- primarily metabolizes citalopram, escitalopram, clomipramine, and amitriptyline
- has substantial involvement in the metabolism of sertraline, imipramine, nortriptyline, and doxepin
- plays an important role in the metabolism of clozapine and a minimal role in the metabolism of thioridazine
- is the primary enzyme involved in the metabolism of diazepam.

Recently, a new variant of *CYP2C19* has been identified which has enhanced function.[10] Patients who are homozygous for this new allele are less likely to respond to 2C19 substrate medications at standard doses. The identification of ultrarapid 2C19 metabolizers can be helpful in evaluating patients who do not respond to standard doses of any of these psychotropic medications.

The cytochrome P450 2C9 gene *(CYP2C9)*

CYP2C9 is located on chromosome 10 in relative close proximity to *CYP2C19*. However, it is only about half the size of *CYP2C9* as it consists of 50 708 nucleotides. Like *CYP2C19*, *CYP2C9* codes for an enzyme that contains 490 amino acids.

CYP2C9 is a drug-metabolizing enzyme gene that is less routinely genotyped to identify the increased or decreased metabolic capacity of psychiatric patients for 2C9 substrate medications. It does not play a primary role in the metabolism of any currently prescribed psychotropic medications. However, the 2C9 enzyme provides the only secondary pathway for the metabolism of fluoxetine, so patients who are poor metabolizers of

both 2D6 substrates and 2C9 substrates are at very high risk for adverse effects if treated with standard dose of fluoxetine.

The cytochrome P450 1A2 gene *(CYP1A2)*

CYP1A2 is a less well-studied drug-metabolizing enzyme gene, but it codes for an enzyme that plays an important role in the metabolism of fluvoxamine. It is also involved in the metabolism of duloxetine and olanzapine. *CYP1A2* is located on chromosome 15 and consists of 7758 nucleotides. *CYP1A2* codes for an enzyme that is composed of 516 amino acids.
A somewhat atypical aspect of the CYP1A2 gene is that there are alleles of this gene that are inducible by smoking tobacco or consuming cruciferous vegetables, such as cabbage or Brussels sprouts. Consequently, patients who smoke tobacco and have two alleles of *CYP1A2* that are inducible by their smoking can be difficult to maintain on 1A2 substrate medications. A relatively common problem occurs when these patients are treated with olanzapine or clozapine on an inpatient psychiatric unit that does not allow them to smoke. When they begin to smoke after they are discharged, their serum level drops and their psychotic symptoms often reoccur. In some populations of European ancestry, as many as 25% of the population can have an inducible ultrarapid *CYP1A2* phenotype.

Pharmacogenomic testing to identify variability in pharmacodynamic responses

A goal of individualized molecular psychopharmacology is to identify medications for an individual patient that will not only be safe, but will be effective. Progress in making predictions of medication response has occurred, and while the goal of being able to predict this response with certainty has not been achieved, we can make increasingly accurate probabilistic predictions of the likelihood of response. Psychiatrists are familiar with this limitation. While hundreds of randomized clinical trials of psychotropic medications have been conducted to identify effective psychotropic drugs, the results of these trials only provide assurance that for a sample of patients there is reasonable likelihood that the medication will be of more benefit than a placebo. While selective serotonin reuptake blockers are among the most widely prescribed medications in the world, many patients do not respond. Specifically, the largest clinical effectiveness study of citalopram reported that less than 30% of the entire sample of patients experienced a complete remission of their symptoms.[11] While the ultimate goal of pharmacodynamically designed pharmacogenomic testing is to identify a drug for a specific patient that will definitely be effective, at the current stage of our understanding, it is only possible to identify a medication that is more likely to be effective.

The serotonin transporter gene *(SLC6A4)*

SLC6A4 is located on chromosome 17 and consists of 37 800 nucleotides. It codes for an enzyme that is composed of 630 amino acids.
SLC6A4 is the most widely genotyped pharmacogenomic "target" gene. A meta-analysis of studies of the relationship between the more active long form of the indel promoter variant of this gene and responses to selective serotonin reuptake inhibitors[12] confirmed the early finding that the long form is associated with a more rapid and better response.[13] However, this has not consistently been demonstrated in patients of Asian ancestry.[14,15] The importance of ancestral heritage has been further demonstrated by multiple analyses of the large STAR*D effective treatment study.
Analyses that did not consider ancestral background did not demonstrate a significant association,[16] while those that focused on patients who identified themselves as "white" but not "Hispanic" did confirm the relationship that patients who were homozygous for the more active long form of the indel promoter polymorphism were more likely to respond to citalopram. Other variants, such as rs25531[17] and the second intronic VNTR[18] are likely to influence the activity level of the gene and, consequently, its response to medications that block its ability to reuptake serotonin in the synapses of the central nervous system.

The serotonin receptor 2A gene *(HTR2A)*

HTR2A is located on chromosome 13 and consists of 62 663 nucleotides. Despite its large size, it codes for an enzyme that is composed of only 471 amino acids.
There have been a series of studies examining the association between variants of *HTR2A* and antidepressant response. A large study examining the response of depressed patients of European ancestry to citalopram

found that a positive response to citalopram was associated with having a copy of the adenine allele of rs7997012.[19]

Another study examining a different *HTR2A* variant, rs6313, reported that patients who were homozygous for the cytosine allele were less likely to tolerate taking paroxetine than those who had one or more copies of the thymine allele.[20]

A series of studies have reported a better response to clozapine in patients who had the thymine allele of rs6313. The thymine allele of rs6313 has also been associated with a lower risk for the development of extrapyramidal side effects when taking antipsychotic medications.[21-23]

The serotonin receptor 2C gene *(HTR2C)*

HTR2C is a very large gene that is located on the X chromosome and consists of 326 074 nucleotides. However, it codes for a protein product that is composed of only 458 amino acids.

Variations in the *HTR2C* gene have been associated with a better clinical response to clozapine. Specifically, patients with schizophrenia who have a copy of the cytosine allele of rs6318 have achieved better control of their psychotic symptoms than patients with the guanine allele.[24,25] However, this same variant has been associated with a higher risk for the development of extrapyramidal side effects in patients who are taking typical antipsychotic medications.[26]

An increased risk for the development of weight gain has been linked to a different *HTR2C* variant. Specifically, the cytosine allele of rs518147 is associated with increased weight gain, while the thymine allele is conceptualized as providing protection against weight gain.[27-29]

The clinical utility of pharmacogenomic testing in psychiatric practice

Assessing the clinical utility of pharmacogenomic testing is an ongoing process, given that the accuracy of genotyping is continually improving, and new research is identifying additional genetic variants that influence medication responses. Reports of adverse responses to 2D6 substrate medications in patients with decreased 2D6 metabolic capacity support the use of testing at this most basic level. Specifically, poor 2D6 metabolizers have had quite dramatic side effects to 2D6 substrate medications[3] and some toxic reactions have been lethal.[30,31] However, there have been no large randomized clinical trials to demonstrate the clinical utility of pharmacogenomic testing. Such trials would reinforce the use of testing. However, it is unlikely that these trials will ever be conducted because, by definition, they are not designed to concentrate on those patients who are the most likely to benefit from pharmacogenomic testing. Trials that screen vulnerable populations and identify patients at risk for suboptimal responses to medications are a more efficient method to address the clinical usefulness of testing patients with decreased metabolic capacity. These screened patients could then be enrolled in protocols designed to provide optimal response for their specific genotypes and predicted pharmacogenomic phenotypes.

Ethical considerations for pharmacogenomic testing in psychiatric practice

The provision of pharmacogenomic testing involves relatively few risks, but ethical safeguards are still important to consider. These are essentially the same considerations that are important to think through when ordering any laboratory test that has the potential to direct a treatment decision.

First, clinical pharmacogenomic testing requires obtaining appropriate consent. This has become a guiding principle for all diagnostic and therapeutic procedures. Clinicians should provide the basic rationale for proceeding with pharmacogenomic testing so that their patients have the opportunity to provide explicit informed consent.

Secondly, as a component of obtaining clinical consent, it must be clear that clinical testing is a voluntary procedure. This is true for virtually all clinical laboratory testing with the relatively rare exceptions of mandatory testing that can identify a condition with a potential negative influence on the public health of the community. A common example of compulsory testing is the monitoring of infections in order to prevent contagion.

A third principle is that clinicians must insure the confidentiality of sensitive medical information that becomes a part of the medical record of the patient. This is true whether the information is derived from a pathological specimen that reveals a malignant carcinoma or from magnetic resonance imaging that demonstrates atrophy

of the hippocampus. The security of the medical record is the responsibility of the clinician.

Finally, any diagnostic medical procedure must have an acceptable level of reliability. The degree of accuracy of any clinical laboratory testing is dependent on a number of key variables. Two of these variables are the seriousness of the prognosis for the patient if the test is positive and the efficacy of available treatments. In designing the treatment plan for a potentially lethal condition that is likely to respond well to a relatively benign intervention if it is administered early in the course of the illness, a laboratory test with high sensitivity is desirable. The most important objective in this situation is to identify as quickly as possible those patients who will benefit from treatment.

Future developments that will influence pharmacogenomic testing in psychiatric practice

In the 2009 presidential lecture of the American Psychiatric Association, it was predicted that pharmacogenomic testing would become a part of everyday psychiatric practice.[32] Ironically, in many academic health centers, pharmacogenomic testing has been utilized since 2004—the time of the introduction of the AmpliChip. Over the intervening years, early adopters have integrated pharmacogenomic testing into their inpatient protocols and ultimately into their outpatient practices. However, this testing has not yet been included in many clinical guidelines.

Pharmacogenomic testing is an innovation, and it takes time for innovations to become integrated into standard practice. While it is difficult to predict with accuracy just how quickly pharmacogenomic testing will become an essential component of clinical psychopharmacological practice, there is no question that this will happen. Ironically, given advances in our ability to sequence genes both rapidly and inexpensively, there will come a time in the near future when most patients will know their 2D6 phenotype in the same way as today they know their blood type. However, well before we reach a state of universal awareness of our informative genotypes, our patients will no longer accept avoidable side effects, and will demand basic pharmacogenomic testing prior to taking antidepressant or antipsychotic medications.

REFERENCES

1. Mrazek DA. *Psychiatric Pharmacogenomics*. New York, NY: Oxford University Press; 2010.
2. Thase ME, Friedman ES, Biggs MM, et al. Cognitive therapy versus medication in augmentation and switch strategies as second-step treatments: a STAR*D report. *Am J Psychiatry*. 2007;164:739-752.
3. Weinshilboum R. Inheritance and drug response. *N Engl J Med*. 2003;348:529-537.
4. Dalen P, Dahl ML, Ruiz ML, Nordin J, Bertilsson L. 10-Hydroxylation of nortriptyline in white persons with 0, 1, 2, 3, and 13 functional CYP2D6 genes. *Clin Pharmacol Ther*. 1998;63:444-452.
5. Mrazek D. Out of the pipeline: pharmacogenomic DNA chip. *Curr Psychiatr*. 2005;4:67-73.
6. Preskorn S. Reproducibility of the in vivo effect of the selective serotonin reuptake inhibitors on the in vivo function of cytochrome P450 2D6: an update (part II). *J Psychiatr Pract*. 2003;9:150-158.
7. Goetz MP, Knox SK, Suman VJ, et al. The impact of cytochrome P450 2D6 metabolism in women receiving adjuvant tamoxifen. *Breast Cancer Res Treat*. 2007;101:113-21.
8. Schroth W, Goetz MP, Hamann U, et al. Associaiton between CYP2D6 polymorphisms and outcomes among women with early stage breast cancer treated with tamoxifen. *JAMA*. 2009;302:1429-1436.
9. Henry NL, Stearns V, Flockhart DA, Hayes DF, Riba M. Drug interactions and pharmacogenomics in the treatment of breast cancer and depression. *Am J Psychiatry*. 2008;165:1251-1255.
10. Sim SC, Risinger C, Dahl ML, et al. A common novel CYP2C19 gene variant causes ultrarapid drug metabolism relevant for the drug response to proton pump inhibitors and antidepressants. *Clin Pharmacol Ther*. 2006;79:103-113.
11. Trivedi M, Rush A, Wisniewski S, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry*. 2006;163:28.
12. Serretti A, Kato M, De Ronchi D, Kinoshita T. Meta-analysis of serotonin transporter gene promoter polymorphism (5-HTTLPR) association with selective serotonin reuptake inhibitor efficacy in depressed patients. *Mol Psychiatry.* 2007;12:247-257.
13. Serretti A, Zanardi R, Rossini D, Cusin C, Lilli R, Smeraldi E. Influence of tryptophan hydroxylase and serotonin transporter genes on fluvoxamine antidepressant activity. *Mol Psychiatry*. 2001;6:586-592.
14. Yoshida K, Itoa K, Satoa K, et al. Influence of the serotonin transporter gene-linked polymorphic region on the antidepressant response to fluvoxamine in Japanese depressed patients. *Prog Neuropsychopharmacol Biol Psychiatry*. 2002;26:383-386.
15. Kim DK, Lim SW, Lee S, et al. Serotonin transporter gene polymorphism and antidepressant response. *Neuroreport*. 2000;11:215-219.
16. Peters E, Slager S, McGrath P, Knowles J, Hamilton S. Investigation of serotonin-related genes in antidepressant response. *Mol Psychiatry*. 2004;9:879-889.
17. Hu XZ, Rush AJ, Charney D, et al. Association between a functional serotonin transporter promoter polymorphism and citalopram treatment in adult outpatients with major depression. *Arch Gen Psychiatry*. 2007;64:783-792.
18. Mrazek DA, Rush AJ, Biernacka JM, et al. SLC6A4 variation and citalopram response. *Am J Med Genet Part B.* 2009;150:341-351.
19. McMahon FJ, Buervenich S, Charney D, et al. Variation in the gene encoding the serotonin 2A receptor is associated with outcome of antidepressant treatment. *Am J Hum Genet*. 2006;78:804-814.
20. Murphy GM, Kremer C, Rodrigues HE, Schatzberg AF. Pharmacogenetics of antidepressant medication intolerance. *Am J Psychiatry*. 2003;160:1830-1835.

Pruebas farmacogenómicas en la práctica clínica psiquiátrica

La incorporación en psiquiatría clínica de las pruebas farmacogenómicas ha ocurrido rápidamente en los últimos siete años. Inicialmente se identificaron genes de enzimas metabolizadoras de fármacos, como el gen del citocromo P450 2D6. La tipificación del gen del citocromo P450 2D6 que es altamente variable da la oportunidad actualmente a los clínicos de identificar a los metabolizadores pobres y los ultrarrápidos para las sustancias que son sustrato del 2D6. Con posterioridad se ha podido disponer en la práctica clínica de genes que influyen en la respuesta farmacodinámica de los medicamentos. Entre los primeros "genes blanco" estuvo el gen del transportador de serotonina (SLC6A4), el cual tiene variantes que han demostrado que influyen en la respuesta clínica de los pacientes con ancestros europeos cuando son tratados con inhibidores selectivos de la recaptura de serotonina. La tipificación de algunos de los genes del receptor de serotonina también está disponible para guiar la práctica clínica. La cuantificación de la utilidad clínica de las pruebas farmacogenómicas está en desarrollo y se han establecido las consideraciones éticas para su realización. Considerando la cada vez más clara costo-eficacia de la tipificación génica, recientemente se ha pronosticado que las pruebas farmacogenómicas se solicitarán de rutina para orientar la selección y dosificación de los fármacos psicotrópicos.

Évaluation pharmacogénomique psychiatrique en pratique clinique

L'évaluation pharmacogénomique psychiatrique s'est rapidement imposée en pratique clinique au cours de ces 7 dernières années. Les gènes d'enzymes métabolisant les médicaments, comme le cytochrome P450 2D6 (CYP2D6), ont d'abord été identifiés. Le génotypage de ce gène très variable permet maintenant aux cliniciens d'identifier des métaboliseurs lents et des métaboliseurs ultrarapides des substrats du 2D6. Des gènes influant sur la réponse pharmacodynamique des médicaments sont ensuite devenus disponibles en pratique clinique. Parmi les premiers « gènes cibles », le gène du transporteur de la sérotonine (SLC6A4) possède des variants qui influent sur la réponse clinique des patients d'ascendance européenne lorsqu'ils sont traités avec des inhibiteurs sélectifs de la recapture de la sérotonine. Le génotypage de certains gènes du récepteur de la sérotonine est également disponible pour guider la réponse clinique. La quantification de l'utilité clinique de l'évaluation pharmacogénomique évolue et fait l'objet de considérations éthiques. Il a été récemment prédit qu'en raison de l'évidente rentabilité croissante du génotypage, l'évaluation pharmacogénomique devrait faire partie des examens de routine pour sélectionner et ajuster la posologie des médicaments psychotropes.

21. Segman RH, Heresco-Levy U, Finkel B, et al. Association between the serotonin 2A receptor gene and tardive dyskinesia in chronic schizophrenia. *Mol Psychiatry.* 2001;6:225-229.

22. Tan EC, Chong SA, Mahendran R, Dong F, Tan CH. Susceptibility to neuroleptic-induced tardive dyskinesia and the T102C polymorphism in the serotonin type 2A receptor. *Biol Psychiatry*. 2001;50:144-147.

23. Lattuada E, Cavallaro R, Serretti A, Lorenzi C, Smeraldi E. Tardive dyskinesia and DRD2, DRD3, DRD4, 5-HT2A variants in schizophrenia: an association study with repeated assessment. *Int J Neuropsychopharmcol* 2004;7:489-493.

24. Sodhi MS, Arranz MJ, Curtis D, et al. Association between clozapine response and allelic variation in the 5-HT2C receptor gene. *Neuroreport*. 1995;7:169-172.

25. Arranz MJ, Munro J, Birkett J, et al. Pharmacogenetic prediction of clozapine response. *Lancet*. 2000;355:1615-1616.

26. Gunes A, Scordo MG, Jaanson P, Dahl ML. Serotonin and dopamine receptor gene polymorphisms and the risk of extrapyramidal side effects in perphenazine-treated schizophrenic patients. *Psychopharmacology* 2007;190:479-484.

27. Reynolds GP, Zhang ZJ, Zhang XB. Association of antipsychotic drug-induced weight gain with a 5-HT2C receptor gene polymorphism. *Lancet.* 2002;359:2086-2087.

28. Templeman LA, Reynolds GP, Arranz B, San L. Polymorphisms of the 5-HT2C receptor and leptin genes are associated with antipsychotic drug-induced weight gain in Caucasian subjects with a first-episode psychosis. *Pharmacogenet Genomics.* 2005;15:195-200.

29. De Luca V, Muller DJ, Hwang R, et al. HTR2C haplotypes and antipsychotics-induced weight gain: X-linked multimarker analysis. *Hum Psychopharmacol*. 2007;22:463-467.

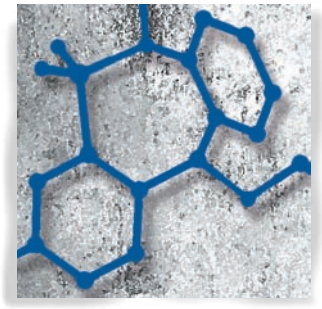
30. Sallee FR, DeVane CL, Ferrell RE. Fluoxetine-related death in a child with cytochrome P-450 2D6 genetic deficiency. *J Child Adolesc Psychopharmacol*. 2000;10:27-34.

31. Koski A, Ojanpera I, Vuori E, Sajantila A. A fatal doxepin poisoning associated with a defective CYP2D6 genotype. *Am J Forensic Med Pathol*. 2007;28:259-261.

32. Stotland NL. Presidential address. *Am J Psychiatry*. 2009;166:1100-1104.

Genetics of drug dependence

Joel Gelernter, MD; Henry R. Kranzler, MD



Drug-dependence disorders (we focus here on cocaine, opioid, and nicotine dependence) are genetically influenced. Risk genes have been located based primarily on genetic linkage studies, and identified primarily based on genetic association studies. In this article we review salient results from linkage, association, and genome-wide association study methodologies, and discuss future prospects for risk allele identification based on these, and on newer, methodologies. Although considerable progress has been made, it is likely that the application of more extensive sequencing than has previously been practical will be required to identify a fuller range of risk variants.

 Dialogues Clin Neurosci. 2010;12:77-84.

Keywords: *drug dependence; genetics; genome-wide; candidate gene; complex trait*

Author affiliations: Departments of Psychiatry, Genetics, and Neurobiology, Yale University School of Medicine, New Haven, CT, USA; and VA CT Healthcare System (Joel Gelertner); Departments of Psychiatry and Genetics and Developmental Biology, University of Connecticut School of Medicine, Farmington, CT, USA (Henry R. Kranzler)

Address for correspondence: Joel Gelernter, Yale University School of Medicine, VA CT Healthcare Center 116A2, 950 Campbell Avenue,West Haven CT 06516, USA (e-mail: joel.gelernter@yale.edu)

It is well established that risk for many substance-dependence traits is genetically influenced; this is the case for each specific substance that has been studied. This has been determined using the methods of genetic epidemiology, the most relevant of which, for this purpose, are twin and adoption studies. We discuss relevant findings from genetic epidemiologic studies of drug use and use disorders below.

In considering drug dependence, we include the most commonly used illegal substances (primarily cocaine, opioids, marijuana, and methamphetamine) and also nicotine, a legal drug that is the dependence-causing substance in tobacco. Alcohol dependence (AD) shares many risk genes with the drug-dependence disorders, but is beyond the scope of the present article. We have recently reviewed AD genetics elsewhere.[1]

As is usual for complex traits, risk for drug dependence is influenced by both genetic and environmental factors. Compared with most other kinds of traits though, environmental factors, most obviously exposure to the substance, are crucial—you cannot become heroin-dependent, for example, if you live in an environment with no access to heroin. Because the availability of illegal substances of abuse varies over the world (to a much greater extent than the availability of either alcohol or tobacco), and also varies with time as a function of secular trends in substance use that are determined by fads, trends in law enforcement, and other factors, patterns of substance dependence are very different across the globe. Genetic epidemiologic studies have helped to clarify the important implications of this environmental variation for genetic studies.

Family studies have shown substantially higher rates of drug abuse among siblings (particularly those whose par-

Selected abbreviations and acronyms

AA	*African-American*
AD	*alcohol dependence*
CD	*cocaine dependence*
EA	*European-American*
GWAS	*genome-wide association study*
ND	*nicotine dependence*
OD	*opioid dependence*
SD	*substance dependence*
SNP	*single-nucleotide polymorphism*

ents were positive for substance abuse) than among individuals in the community.[2-5] Such studies have also provided evidence for both general familial aggregation for substance-use disorders and substance-specific aggregation across a wide range of drugs, including nicotine, opioids, cocaine, and cannabis.[6,7] However, while these designs demonstrate that drug-dependence disorders are familial, they cannot distinguish between genetic and environmental contributions to this familiality. A demonstration of genetic contributions to these and other disorders requires other designs, prominently twin and adoption studies.

Two adoption studies conducted by Cadoret et al[8,9] showed that the only biological factor that was significantly associated with drug abuse in the proband was an alcohol problem in first-degree relatives. However, Tsuang et al,[10] studying a sample of more than 3000 twin pairs, found a significantly greater pairwise concordance rate for monozygotic (MZ) than dizygotic (DZ) twins for abuse of marijuana, stimulants, cocaine, and for all drugs combined. Using twin pairs ascertained through the Virginia Twin Registry, Kendler et al examined concordance rates for drug use and dependence among more than 800 female-female pairs.[11-13] Model fitting showed that twin resemblance for liability to the use of cocaine, cannabis, hallucinogens, opioids, and sedatives was due to both genetic and family environmental factors. Liability to abuse or dependence on cocaine and cannabis was due only to genetic factors. In contrast, however, in another study by Kendler et al[14] of the use and misuse of six classes of illicit drugs by nearly 1200 male-male twin pairs, model fitting revealed that one common genetic factor exerted a potent influence on risk for both substance use and misuse for all six substances. There was a modest effect on risk of substance-specific genetic factors seen for substance use, but in contrast to other studies cited above, not for abuse or dependence. A single common shared environmental factor was also found to exert an effect on risk of substance use, and to a lesser extent, on risk of abuse dependence.

Despite some contradictory findings, overall, the data from adoption, twin, and family studies support a substantial genetic contribution to drug dependence, including the existence of genetic factors specific to each of these disorders, and factors common to these disorders and other forms of substance dependence.

It is only common genetic factors (that is, those that influence more than one substance) that are likely to be important worldwide (genetic factors specific to substances will vary because the specific substances vary). Whether genes relevant to drugs of abuse that have some similarities in their mechanisms of action, such as cocaine (important, eg, in the US) and methamphetamine (predominant, eg, in Thailand, and important in certain regions in the US) will prove to overlap, is still an open question. Further, different risk factors may be important in different populations (discussed in ref 1). In the small number of instances where similar SD traits have been studied in different populations, the genetic factors uncovered have not been identical.

Thus, gene mapping for substance-dependence (SD) traits is complicated. Some risk alleles identified may be important only for specific substances of abuse and others, only for certain populations. So why try to map genes for SD traits? First, SD is a huge cause of morbidity and mortality worldwide; that is, it is a very important problem that deserves to be studied despite its complexity. Second, despite all of the *a priori* reasons to believe that it would be exceedingly difficult to identify genes and validate the findings, the track record for SD genetics as a field is really very good. Below, we will review some recent results that support this claim.

Linkage studies

Genome-wide linkage studies, the traditional approach to identifying risk loci, provide chromosomal locations for risk-influencing loci based on the observation of coinheritance of marker alleles and the disease trait in families. To be comprehensive, linkage studies employ markers that map throughout the entire genome. This approach has been used for cocaine, opioid, and nicotine dependence, and for related traits.

We are aware of only one linkage study of cocaine dependence (CD); we studied a sample of small families

each with at least one subject affected with CD, which included 528 full and 155 half sibpairs and was 45.5% European-American (EA) and 54.5% African-American (AA).[15] We completed an autosomal genome-wide linkage scan for the CD diagnosis, cocaine-induced paranoia, and cocaine-related subphenotypes derived using cluster analytic methods. The subtyping procedure was used to identify more genetically homogeneous subgroups of subjects in which the effects of individual risk loci might be more prominent. For CD, we found "suggestive" linkage signals on chromosome 10, in the full sample, and on chromosome 3, in the EA part of the sample. Much stronger results were obtained for the cluster-derived subtypes, including genome-wide-significant lod scores for membership in the "Heavy Use, Cocaine Predominant" cluster on chromosome 12 and for membership in the "Moderate Cocaine and Opioid Abuse" cluster on chromosome 18. In AA families only, we observed a genome-wide-significant lod score on chromosome 9 for the trait of cocaine-induced paranoia. Genome-wide significance was defined on the basis of Lander and Kruglyak's 1995 criteria.[16]

There have been three independent genome-wide linkage studies of opioid dependence (OD). We studied 393 small families each with at least one individual affected with OD.[17] We completed a genome-wide linkage scan for DSM-IV OD, and, as for the CD study, for cluster-defined phenotypes, a heavy-opioid-use cluster, and a non-opioid-using cluster. The strongest results were, again, seen with the cluster-defined traits: for the "heavy opioid users" cluster there was a genome-wide-significant linkage for EA and AA subjects combined, on chromosome 17. For the "nonopioid users" cluster, there was a genome-wide-significant linkage elsewhere on chromosome 17, for EA subjects only. Lachman et al[18] studied a mixed US sample of 305 OD-affected sibling pairs, and identified evidence for linkage on a region of chromosome 14 overlying the neurexin 3 gene (*NRXN3*). They also identified a male-specific linkage peak on chromosome 10q. Finally, Glatt et al[19] studied a sample of nearly 400 independent affected sibling pairs ascertained in China near the Golden Triangle, one of Asia's largest illicit opium-producing areas, but did not identify any strongly-supported linkage signals, despite the presumed genetic homogeneity of the sample. The strongest signal they observed was on chromosome 4q.

There have been numerous genome-wide linkage scans for smoking and related phenotypes, reviewed in ref 20. Han et al[21] completed genome scan meta-analysis (GSMA) of genome-wide linkage scans for nicotine dependence (ND) and related traits, pooling all available independent genome scan results on smoking behavior. To minimize locus heterogeneity, subgroup analyses of the smoking behavior assessed by the Fagerstrom Test for Nicotine dependence (FTND) and maximum number of cigarettes smoked in a 24-hour period (MaxCigs24) were also carried out. Fifteen genome scan results were available for analysis, including 10 253 subjects in 3 404 families. The primary GSMA across all smoking behavior identified a genome-wide "suggestive" linkage in chromosome 17q24.3-q25.3. But the strongest result derived from the subgroup analysis of MaxCigs24 (including 966 families with 3 273 subjects), which identified a genome-wide significant linkage in 20q13.12-q13.32. *CHRNA4*, a strongly supported ND candidate gene, is located in this interval; Li et al[22] previously reported on association of *CHRNA4* variants to ND.

A high level of statistical support for a genetic linkage is very valuable, but the ultimate proof that a disease-influencing locus underlies a statistical linkage peak is the identification of a risk gene in the peak that accounts for the linkage signal. The next step is typically genetic association analysis, ie, evaluation of a set of markers that map under the linkage peak for association with the trait. Genetic association provides another degree of statistical evidence, but eventually, proof of a disease-gene relationship must rely on demonstration of a functional effect of a variant or variants at the risk locus. ND is the furthest of all drug-dependence (DD) traits along this pathway, with numerous loci supported on the basis of statistical genetic association evidence, and some of these loci have received the higher level of support of functional data.

Association studies

Strategy for single-nucleotide polymorphism (SNP) selection plays a key role in association study outcome. In general, variants predicted to have functional consequences, eg, because they alter predicted amino acid sequence, have been favored for study; alternatively, researchers often try to capture most of the genetic variation at a locus via selection of haplotype tagging SNPs followed by haplotype reconstruction. It is important to recognize some of the limitations of these strategies at the outset. Although most common putatively functional

SNPs are known, rarer SNPs may have large phenotypic effects, and there are many such variants yet to be discovered. Not all functional SNPs are easily recognized as such. SNPs vary by population, and populations differ in the extent to which common genetic variation has been identified. The same population variation is reflected in differences in haplotype structure. Finally, haplotype reconstruction is almost always accomplished via computer algorithms, and the results are estimated. With these limitations in mind, we discuss several examples of genetic associations with DD phenotypes, focusing on interesting physiological candidates and on replicated findings.

Association of variants that map at or near the D2 dopamine receptor (*DRD2* locus) with drug or alcohol dependence was proposed many years ago and has been widely debated. We identified a "suggestive" linkage peak for ND at the region of chromosome 11 that includes the *NCAM1-TTC12-ANKK1-DRD2* gene cluster.[23] The inconsistent results with *DRD2* may be attributable to an indirect effect—observed association could actually be mediated through variation at a nearby locus in linkage disequilibrium with *DRD2*. To test this hypothesis, we genotyped 43 SNP markers in a region including *DRD2* and the three adjacent genes, in an SD linkage sample of >1600 subjects. We found very strong evidence of association of multiple SNPs at *TTC12* and *ANKK1* in two different populations, EAs and AAs (minimal P=0.0007 in AAs and minimal P=0.00009 in EAs), and highly significant association of a single haplotype (set of markers) spanning *TTC12* and *ANKK1* to ND in the pooled sample (P=0.0000001). Thus, a risk locus for ND maps to a region that spans *TTC12* and *ANKK1*. The exact localization of the risk haplotype depends on the disease definition, and whether and which co-occurring diagnoses are present in the study sample.[24]

These results support the hypothesis that the *DRD2* findings could be attributable to variants in nearby loci. Such variants could reflect either functional variation that affect those loci (and not *DRD2*), or relatively distant regulatory regions important for *DRD2* function. The *ANKK1* finding in ND has been replicated.[25]

Another set of risk loci that are of interest in relation to the risk of drug dependence are those encoding proteins that regulate or mediate opioidergic function. All of the opioid receptor genes have been reported to be associated with substance dependence liability. A functional polymorphism in *OPRM1* (Asn40Asp), which encodes the mu-opioid receptor, has been the most extensively studied in this regard, though the association is controversial. Although multiple studies have shown a significant allelic association with DD, they are nearly evenly divided between those showing a significant excess of the Asp40 allele among cases[26,27] and those showing a significant excess of the Asp40 allele among controls.[28-30] Consequently, meta-analyses of that literature failed to show a reliable association of the SNP with either OD[31] or any SD disorder.[32] However, Zhang et al[33] examined 13 SNPs spanning the coding region of *OPRM1* in a sample of EAs with AD and/or DD and 338 EA healthy controls. The SNPs formed two haplotype blocks. There were significant differences between cases and controls in allele and/or genotype frequencies for SNPs in Block I and in Block II, after correction for multiple testing. Haplotypes constructed from five tag SNPs differed significantly in frequency between both AD and DD subjects and controls. Logistic regression analyses in which the sex and age of subjects and alleles, genotypes, haplotypes, or diplotypes of the five tag SNPs were considered confirmed the association between *OPRM1* variants and SD.

Zhang et al[34] also examined the genes encoding the other two opioidergic receptors: *OPRD1* (which encodes the delta receptor) and *OPRK1* (which encodes the kappa receptor). Eleven SNPs spanning *OPRD1* were examined in EAs with AD, CD, and/or OD, and control subjects. Although nominally significant associations were observed for five SNPs with SD, only the association of the nonsynonymous variant G80T with OD remained significant after correction for multiple testing. Haplotype analyses with six tag SNPs indicated that a specific haplotype was significantly associated with AD and OD (P<0.001). In logistic regression analyses, controlling for sex and age, this haplotype had a risk effect on AD and, to a much greater extent, on OD. In addition, seven SNPs covering *OPRK1* were examined in the majority of subjects and although there were no significant differences in allele, genotype, or haplotype frequency distributions between cases and controls, a specific *OPRK1* haplotype was significantly associated with AD, but not DD. In summary, these findings demonstrated a robust positive association between *OPRD1* variants and SD, particularly OD.

Finally, Zhang et al[35] studied *POMC*, the gene that encodes pro-opioimelanocortin, from which functionally different peptides are derived via tissue-specific post-

translational processing; of particular relevance here are two principal elements of the hypothalamic-pituitary-adrenal axis: adrenocorticotropin (ACTH) and β-endorphin. Five SNPs spanning *POMC* were examined in independent family and case-control samples of EAs and AAs. The families were ascertained based on a pair of siblings affected with cocaine and/or opioid dependence. Case-control studies included cases affected with AD, CD and/or OD and controls. Family-based analyses revealed an association of one SNP (rs6719226) with OD in AA families, and a different SNP (rs6713532) with CD in EA families. Case-control analyses demonstrated an association of rs6713532 with AD or CD. Moreover, the minor allele of a third SNP was a risk factor for CD or OD in AAs, and for AD, CD, or OD in EAs. Logistic regression analyses in which sex and age were considered and population stratification analyses confirmed these findings. Additionally, specific haplotypes increased risk for CD in AAs and OD in EAs.

In summary, as might be expected given that the brain's opioidergic system plays a central role in reinforcement, which has important implications for addiction,[36] variation in a number of functional candidate genes encoding opioidergic proteins have been implicated in dependence on alcohol, cocaine, and opioids. Assuming independent replication of these findings, a key question to be addressed is the nature of gene-gene and gene by environment interactions to which risk of SD is attributable.

Other studies have demonstrated associations with the cannabinoid receptor gene (*CNR1*),[37-39] neurexin 1 (*NRXN1*),[40] and a set of alcohol-metabolizing enzymes.[41] A clear pattern emerges from the examination of this sampling of candidate gene associations with SD: insofar as genes with known function are concerned, there are no big surprises with respect to physiology. (This can not be said about genes without clearly delineated functional roles, such as *ANKK1*, which was identified, not incidentally, based on its position, rather than its function.) This highlights the limitations of the candidate gene approach, which is often inherently biased by prior knowledge about physiology. Unbiased studies have greater potential to reveal new mechanisms of addiction, and that is a key attraction of the genome-wide association study (GWAS) methodology discussed below.

GWASs are an alternative to linkage for locating genes anywhere in the genome without prior hypotheses. GWAS designs are of interest due to their potential to identify risk loci of relatively small effect, much smaller than through linkage strategies. (In fact, one controversy engendered by the widespread adoption of GWAS designs is that often risk alleles are identified that have such a small effect—typically with odds ratios less than 1.2—that it is hard to know what to do with them once they have been identified.) A second advantage of GWASs is that they may be based on case-control samples, which are easier to recruit than family sampling schemes, which must be deployed to prepare for linkage. Family samples are more difficult to recruit (markedly so for many kinds of SD because of the tendency of these disorders to fragment families) and can introduce certain kinds of bias. The first GWAS for a specific SD trait, excluding studies that used a pooling methodology exclusively (see ref 42), examined ND.[43] This study employed a two-stage design; first pooled DNA was used to screen 2.4 million SNPs; second, >30 000 SNPs selected from the first stage were screened individually in ~1000 each cases and controls. Numerous genes were identified as possibly associated to ND, including both novel genes and genes that were previously considered candidates based on known physiology (eg, cholinergic receptor, nicotinic, beta 3, *CHRNB3*). The latter finding has been confirmed in larger studies: subsequent GWASs have demonstrated highly significant associations between variation in the nicotinic receptor gene cluster *CHRNA5-CHRNA3-CHRNB4* and ND and related traits[44,45] and with lung cancer.[46,47]

In a hypothesis-generating study, we studied a set of 5633 SNP markers in 1699 subjects from 339 AA families and 334 EA families ascertained through a sib pair meeting DSM-IV criteria for either CD or OD. This is considered a sparse marker set for the purposes of GWAS. It is expected to interrogate <10% of the genome, thus, cannot be considered to be a study of truly genome-wide depth. Associations between these markers and five substance dependence traits (CD, OD, AD, ND, and cocaine-induced paranoia) were assessed by family-based association tests (FBAT). The top-ranked result was an association of a specific SNP in the *MANEA* gene with cocaine-induced paranoia. This study provided an initial SD trait-specific blueprint of associated regions for future candidate gene studies. There are, at the time of this writing, no published GWAS studies for several of these traits. The *MANEA* finding was replicated and extended in a larger sample.[48]

Discussion

We identify two main ways to account for the relatively consistent results seen in this field. First, diagnosis can be made with high reliability. Second, the phenotypes are relatively straightforward because they are, in their essence, pharmacogenetic. That is, SD phenotypes reflect genetic moderation of the subjective response to drugs of abuse.

While results in this research field have been relatively consistent, most of the genetic risk for DD has yet to be attributed to specific alleles. Initially, it was thought that the GWAS was the answer to the problem. But application in other complex traits (eg, schizophrenia, bipolar affective disorder, autism) has revealed a more complex picture, such that even clinical samples that should have been adequately powered have fallen short of providing definitive and significant results. The explanation for this situation may reside in the fundamental genetic architecture of some complex traits. GWAS is based on a common-disease-influenced-by-common-allele model. However, we are now learning that many phenotypes are influenced instead by sets of variants, in sets of loci, each of which is rare on a population level. Such variants are likely to be uncovered only by extensive sequencing of affected and unaffected individuals. Copy number variation (CNV) is another mechanism that is proving to be important in modulating disease risk. Such variation is important for at least some behavioral traits; for example, Sebat et al[49] have reported on the relationship of CNV to autism, and several groups have reported association of rare structural variants with schizophrenia.[50-53]

We have seen several successful examples of genetic association identified following a linkage finding, a sequence that demonstrates the main utility of genetic linkage. But there have also been surprisingly many instances when strong genetic association has not been identified readily. There are many ways to account for such a circumstance—genetic heterogeneity, random variation, and population variation, to name a few. Another intriguing possibility has become more prominent of late. The linkage-to-association-to-gene model is premised basically on the common disease-common variant model discussed above. This model may not be as applicable as was thought; there is increasing evidence that heritability may be accounted for by many rare variants in either a single locus, or a set of related loci. Since linkage depends on the identification of coinheritance of trait and marker within families, it stands to reason that a set of different rare variants could be detected by linkage (even if the responsible variants differed greatly between families in the discovery set). Such variants would be very resistant to discovery by ordinary tagging haplotype association strategies. Similarly, such variants would be expected to be refractory to discovery by GWAS methodology. Deep sequencing studies have successfully accounted for the "missing" genetic variance in some cases. For example Nejentsev et al[54] found a set of individually rare variants at the *IFIH1* locus that affect risk for type 1 diabetes, following up on a GWAS study. Ji et al[55] started with a set of genes known to have large effects on blood pressure in a small number of severely affected families, and sequenced them in a large number of unrelated individuals. Rare variants with smaller effects on blood pressure were identified. These findings are likely to be relevant for SD genetics research as well, inasmuch as deep sequencing of candidate loci in many unrelated individuals may be necessary to account for a greater proportion of the genetic risk than is presently known.

Whole-genome sequencing is becoming progressively less expensive, and will surely ultimately be feasible for locating genetic variants that increase risk for complex genetic traits, albeit at the risk of daunting statistical problems. Sequencing of expressed sequences only ('whole exome") may be a valuable interim step. Ng et al[56] have demonstrated the feasibility of this approach. In summary, new developments in a variety of genetic methods and in the accumulating molecular evidence of the genetic risk for SD promise to yield greater insights into the etiology of these disorders, bringing into relief the environmental contributions and creating opportunities for prevention and new therapeutic options.

Acknowledgements: This work was supported in part by NIDA grants R01 DA12849 and R01 DA12690, and funds from the U.S. Department of Veterans Affairs (the VA Medical Research Program, and the VA Connecticut-Massachusetts Mental Illness Research, Education and Clinical Center (MIRECC).

Genética de la dependencia de drogas

Los trastornos de dependencia de drogas (aquí se revisa cocaína, opioides y nicotina) tienen una influencia genética. Primariamente la localización de los genes de riesgo se ha basado en estudios de ligamiento y la identificación en estudios de asociación genética. En este artículo se revisan resultados destacados sobre metodologías de estudio de ligamiento, asociación y asociación del genoma completo y se discuten las posibilidades futuras para la identificación del riesgo de alelos en base a estas y otras metodologías más recientes. Aunque se ha realizado un progreso considerable, es probable que la aplicación de secuenciaciones más amplias que las que se han practicado previamente serán requeridas para identificar un mayor rango de variantes de riesgo.

Génétique de la toxicomanie

Il existe une prédisposition génétique à la toxicomanie (dans cet article, nous nous intéresserons à la dépendance à la cocaïne, aux opioïdes et à la nicotine). Des gènes de ce risque ont été localisés initialement à partir d'études de liaisons génétiques et identifiés à partir d'études d'association génétique. Nous présentons dans cet article les principaux résultats issus d'études d'association sur le génome entier, de liaisons et d'associations ainsi que les perspectives d'identification d'un allèle du risque fondées sur ces méthodologies et de plus récentes. Malgré des progrès considérables, un séquençage plus étendu que celui effectué auparavant sera probablement nécessaire pour identifier une gamme plus complète de variants du risque.

REFERENCES

1. Gelernter J, Kranzler HR. Genetics of alcohol dependence. *Hum Genet*. 2009;126:91-99. Review.
2. Kosten TR, Rounsaville BJ, Kosten TA, Merikangas K. Gender differences in the specificity of alcoholism transmission among the relatives of opioid addicts. *J Nerv Ment Dis*. 1991;179:392-400.
3. Rounsaville BJ, Kosten TR, Weissman MM, et al. Psychiatric disorders in relatives of probands with opiate addiction. *Arch Gen Psychiatry*. 1991;48:33-42.
4. Luthar SS, Anton SF, Merikangas, KR, Rounsaville BJ Vulnerability to drug abuse among opioid addicts' siblings: Individual, familial, and peer influences. *Compr Psychiatry*. 1992;33:190-196.
5. Luthar SS, Rounsaville BJ. Substance misuse and comorbid psychopathology in a high-risk group: A study of siblings of cocaine misusers. *Int J Addict*. 1993;28:415-434.
6. Bierut LJ, Dinwiddie SH, Begleiter H, et al. Familial transmission of substance dependence: alcohol, marijuana, and habitual smoking. *Arch Gen Psychiatry*. 1998;55:982-988.
7. Merikangas KR, Stolar M, Stevens DE, et al. Familial transmission of substance use disorders. *Arch Gen Psychiatry*. 1998;55:973-979.
8. Cadoret RJ, Troughton E, O'Gorman TW, Heywood E. An adoption study of genetic and environmental factors in drug abuse. *Arch Gen Psychiatry*. 1986;43:1131-1136.
9. Cadoret RJ, Yates WR, Troughton E, Woodworth G, Stewart MA. 1 Adoption study demonstrating two genetic pathways to drug abuse. *Arch Gen Psychiatry*. 1995;52:42-52.
10. Tsuang MT, Lyons MJ, Eisen SA, et al. Genetic influences on DSM-III-R drug abuse and dependence: a study of 3,372 twin pairs. *Am J Med Genet B Neuropsychiatr Genet*. 1996;67:473-477.
11. Kendler KS and Prescott CA. Cocaine use, abuse and dependence in a population-based sample of female twins. *Br J Psychiatry*. 1998;173:345-350.
12. Kendler KS, Prescott CA. Cannabis use, abuse, and dependence in a population-based sample of female twins. *Am J Psychiatry*. 1998;155:1016-1022.
13. Kendler KS, Karkowski LM, Neale MC, Prescott CA. Illicit psychoactive substance use, heavy use, abuse and dependence in a US population-based sample of male twins. *Arch Gen Psychiatry*. 2000;57:261-269.
14. Kendler KS, Jacobson KC, Prescott CA, Neale MC. Specificity of genetic and environmental risk factors for use and abuse/dependence of cannabis, cocaine, hallucinogens, sedatives, stimulants, and opiates in male twins. *Arch Gen Psychiatry*. 2003;160:687-695.
15. Gelernter J, Panhuysen C, Weiss R, et al. Genomewide linkage scan for cocaine dependence and related traits: Linkages for a cocaine-related trait and cocaine-induced paranoia. *Am J Med Genet Part B Neuropsychiatr Genet*. 2005;136B:45-52.
16. LanderE, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet*. 1995;11:241–247.
17. Gelernter J, Panhuysen C, Wilcox M, et al. Genomewide linkage scan for opioid dependence and related traits. *Am J Hum Genet*. 2006;78:759-769.
18. Lachman HM, Fann CS, Bartzis M. Genomewide suggestive linkage of opioid dependence to chromosome 14q. *Hum Mol Genet*. 2007;1;16:1327-1334.
19. Glatt SJ, Lasky-Su JA, Zhu SC, et al. Genome-wide linkage analysis of heroin dependence in Han Chinese: results from Wave Two of a multi-stage study. *Drug Alcohol Depend*. 2008;98:30-34.
20. Li MD, Ma JZ, Beuten J. Progress in searching for susceptibility loci and genes for smoking-related behaviour. *Clin Genet*. 2004;66:382-392.
21. Han S, Gelernter J, Luo X, and Yang BZ. Meta-analysis of 15 genomewide linkage scans of smoking behavior. *Biol Psychiatry*. 2010;67:12-19.
22. Li MD, Beuten J, Ma JZ, et al. Ethnic- and gender-specific association of the nicotinic acetylcholine receptor alpha4 subunit gene (CHRNA4) with nicotine dependence. *Hum Mol Genet*. 2005;14:1211-1219.
23. Gelernter J, Panhuysen C, Weiss R, et al. Genomewide linkage scan for nicotine dependence: Identification of a chromosome 5 risk locus. *Biol Psychiatry*. 2007;61:119-126.
24. Yang B-Z, Kranzler HR, Zhao H, Gruen JR, Luo X, Gelernter J. Haplotypic variants in DRD2, ANKK1, TTC12 and NCAM1 are associated with comorbid alcohol and drug dependence. *Alcohol Clin Exp Res*. 2008;32:2117-2127.
25. Huang W, Payne TJ, Ma JZ, et al. Significant association of ANKK1 and detection of a functional polymorphism with nicotine dependence in an African-American sample. *Neuropsychopharmacology*. 2009;34:319-330.
26. Szeto CY, Tang NL, Lee DT, Stadlin A. Association between mu opioid receptor gene polymorphisms and Chinese heroin addicts. *Neuroreport*. 2001;12:1103-1106.

27. Bart G, Heilig M, LaForge K, et al. Substantial attributable risk related to a functional mu-opioid receptor gene polymorphism in association with heroin addiction in central Sweden. *Mol Psychiatry*. 2004;9:547-549.
28. Bond C, LaForge KS, Tian M, et al. Single-nucleotide polymorphism in the human mu opioid receptor gene alters beta-endorphin binding and activity: possible implications for opiate addiction. *Proc Natl Acad Sci U S A*. 1998;95:9608-9613.
29. Schinka JA, Town T, Abdullah L, et al. A functional polymorphism within the mu-opioid receptor gene and risk for abuse of alcohol and other substances. *Mol Psychiatry*. 2002;7:224-228.
30. Tan, EC, Tan CH, Karupathivan U, Yap EP. Mu opioid receptor gene polymorphisms and heroin dependence in Asian populations. *Neuroreport*. 2003;14:569-572.
31. Glatt SJ, Bousman C, Wang RS, et al. Evaluation of OPRM1 variants in heroin dependence by family-based association testing and meta-analysis. *Drug Alcohol Depend*. 2007;90:159-165.
32. Arias A, Feinn R, Kranzler HR. Association of an Asn40Asp (A118G) polymorphism in the mu-opioid receptor gene with substance dependence: a meta-analysis. *Drug Alcohol Depend*. 2006; 83:262-268.
33. Zhang H, Luo X, Kranzler HR, et al. Association between two mu-opioid receptor gene (OPRM1) haplotype blocks and drug or alcohol dependence. *Hum Mol Genet*. 2006;15:807-819.
34. Zhang H, Kranzler HR, Yang B-Z, Luo X, Gelernter J. The OPRD1 and OPRK1 loci in alcohol or drug dependence: OPRD1 variation modulates substance dependence risk. *Mol Psychiatry*. 2008;13:531-543.
35. Zhang H, Kranzler HR, Weiss RD, et al. Pro-opiomelanocortin gene variation related to alcohol or drug dependence: evidence and replications across family- and population-based studies. *Biol Psychiatry*. 2009;66:128-136.
36. Koob GF. Neurobiology of addiction. In: *Textbook of Substance Abuse Treatment*. 4th ed. Washington, DC: American Psychiatric Publishing, Inc; 2008:3-16.
37. Zhang PW, Ishiguro H, Ohtsuki T, et al. Human cannabinoid receptor 1: 5' exons, candidate regulatory regions, polymorphisms, haplotypes and association with polysubstance abuse. *Mol Psychiatry*. 2004;9:916-931.
38. Zuo L, Kranzler HR, Luo X, Covault J, Gelernter J. CNR1 variation modulates risk for drug and alcohol dependence. *Biol Psychiatry*. 2007;62:616-626.
39. Zuo L, Kranzler HR, Luo X, et al. Interaction between two independent CNR1 variants increases risk for cocaine dependence in European Americans: a replication study in family-based sample and population-based sample. *Neuropsychopharmacology*. 2009;34:1504-1513.
40. Nussbaum J, Xu Q, Payne TJ, et al. Significant association of the neurexin 1 gene (NRXN1) with nicotine dependence in European and African American smokers. *Hum Mol Genet*. 2008;17:1569-1577.
41. Luo X, Kranzler HR, Zuo L, Wang S, Schork NJ, Gelernter J. Multiple ADH genes modulate risk for drug dependence in both African- and European-Americans. *Hum Mol Genet*. 2007;16:380-390.
42. Uhl GR, Drgon T, Liu QR, et al. Genome-wide association for methamphetamine dependence: convergent results from 2 samples. *Arch Gen Psychiatry*. 2008;65:345-355.
43. Bierut LJ, Madden AF, Breslau N, et al. Novel genes identified in a high-density genome wide association study for nicotine dependence. *Hum Mol Genet*. 2007;16:1:24-35.
44. Berrettini W, Yuan X, Tozzi F, et al. Alpha-5/alpha-3 nicotinic receptor subunit alleles increase risk for heavy smoking. *Mol Psychiatry*. 2008;13:368-373.
45. Thorgeirsson TE, Geller F, Sulem P, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature*. 2008;452:638-642.
46. 46.Amos CI, Wu X, Broderick P, et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet*. 2008;40:616-622.
47. Hung RJ, McKay JD, Gaborieau V, et al. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature*. 2008;452:633-637.
48. Farrer LA, Kranzler HR, Yu Y, et al. Association of variants in the α-endomannosidase (MANEA) gene with cocaine-related behaviors. *Arch Gen Psychiatry*. 2009;66:267-274.
49. Sebat J, Lakshmi B, Malhotra D, et al. Strong association of de novo copy number mutations with autism. *Science*. 2007;316:445-449.
50. Walsh T, McClellan JM, McCarthy SE, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science*. 2008;320:539-543.
51. Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M. Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet*. 2008;40:880-885.
52. Stefansson H, Rujescu D, Cichon S, et al. Genetic Risk and Outcome in Psychosis (GROUP). Large recurrent microdeletions associated with schizophrenia. *Nature*. 2008;455:232-236.
53. International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*. 2008;455:237-241.
54. Nejentsev S, Walker N, Riches D, Egholm M, Todd JA. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science*. 2009;324:387-389.
55. Ji W, Foo JN, O'Roak BJ, et al. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nat Genet*. 2008;40:495-496.
56. Ng SB, Turner EH, Robertson PD, et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*. 2009;461:272-276.

New findings in the genetics of major psychoses

Markus M. Nöthen, MD; Vanessa Nieratschker, PhD; Sven Cichon, PhD; Marcella Rietschel, MD

Schizophrenia and bipolar disorder have a largely unknown pathophysiology and etiology, but they are highly heritable. Although linkage and association studies have identified a series of chromosomal regions likely to contain susceptibility genes, progress in identifying causative genes has been largely disappointing. However, rapid technological advances are beginning to lead to new insights. Systematic genome-wide association and follow-up studies have reported genome-wide significant association findings of common variants for schizophrenia and bipolar disorder. The risk conferred by individual variants is small, and some variants confer a risk for both disorders. In addition, recent studies have identified rare, large structural variants (copy number variants) that confer a greater risk for schizophrenia. This review summarizes recent developments in genetic research into schizophrenia and bipolar disorder, and discusses possible future directions in this field.

Dialogues Clin Neurosci. 2010;12:85-93.

Keywords: *schizophrenia; bipolar disorder; heritability; genetic variant; linkage study; candidate gene study; copy number variation, genome-wide association study*

Schizophrenia and bipolar affective disorder (bipolar disorder, manic depression) are major psychiatric disorders. They profoundly affect thought, perception, emotion, and behavior, and their symptoms cause significant social and/or occupational dysfunction. The World Health Organization ranks both disorders among the top 10 leading causes of the global burden of disease for the 15-to-44 age group.

Schizophrenia and bipolar disorder are illnesses with a largely unknown pathophysiology and etiology. However, genetic epidemiology has demonstrated that modern psychiatric diagnostic criteria define disorders that are highly heritable. Estimates of heritability range between 70% and 90% for schizophrenia[1] and 60% and 80% for bipolar disorder.[2] It is generally accepted that the inheritance of psychiatric disorders is complex. Multiple genetic and environmental factors contribute to the development of a disorder[3-9] and it is possible that gene-gene interactions also occur.[10,11]

Extensive efforts have been made over the past 20 years to identify the susceptibility genes for psychiatric disorders on a molecular genetic level, although this has proven to be a far more difficult undertaking than was first anticipated. Until recently, the linkage approach and microscopic cytogenetic studies were the only available

Author affiliations: Department of Genomics, Life & Brain Centre and Institute of Human Genetics, University of Bonn, Bonn, Germany (Markus M. Nöthen, Sven Cichon); Division of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Mannheim, Germany (Vanessa Nieratschker, Marcella Rietschel)

Address for correspondence: Prof Dr med Markus M. Nöthen, Director, Institute of Human Genetics University of Bonn, Sigmund-Freud-Strasse 25, D-53127 Bonn, Germany
(e-mail: markus.noethen@uni-bonn.de)

methods of systematically searching the genome. A disadvantage of these two methods is their low level of resolution. Linkage studies have identified a series of chromosomal regions that are likely to contain susceptibility genes, and highly promising association findings have been obtained for several genes in these regions (eg, *neuregulin 1* [*NRG1*], *G72/G30* locus, *dystrobrevin-binding protein 1* [*DTNBP1*]).[12-14] However, it has not yet been possible to identify any genetic variant that confers a direct functional effect and which is consistently associated with disease across populations. Cytogenetic studies have also generated some highly promising candidate genes such as the *disrupted-in-schizophrenia-1* gene (*DISC1*).[15] Subsequent studies have reported highly interesting findings regarding the function of these genes and their associated pathways.[16]

Recently, however, important advances have been made as a result of rapid developments in technologies that are able to decipher the variability of the human genome at high resolution, and which allow systematic investigation of the impact of such variability in large samples. This article summarizes these developments in genetic research into schizophrenia and bipolar disorder, and discusses possible future directions in this field.

Genome-wide association studies

The introduction of the genome-wide association study (GWAS) is the result of enormous technological advances. GWASs involve the use of arrays that simultaneously genotype several hundred thousand single nucleotide polymorphisms (SNPs) per individual. This enables a hypothesis–free search of every gene and most intergenic regions of the genome in samples of unrelated patients and controls. In this respect GWASs resemble genome-wide linkage studies (genome scans), but they have several major advantages: (i) they are not dependent on the recruitment of families; (ii) they have better resolution since (in contrast to linkage) they detect linkage disequilibrium with susceptibility variants, which usually extends over smaller genomic regions (in the range of a few ten thousand base pairs); and (iii) they have greater power to detect small genetic effects. In contrast to linkage studies, however, they are restricted to the investigation of common variants, since SNPs with low minor allele frequencies are poorly represented on currently available arrays. A serious difficulty in evaluating the results of GWASs is the issue of multiple testing. A large number of SNPs may be tested within the same study for their association with a disease, and this generates many nominally significant findings that are actually false positives. It is therefore necessary to correct for multiple testing to achieve the level of genome-wide significance. This level is dependent upon the number of SNPs analyzed, and the threshold for currently available GWA chips is approximately 5×10^{-8} (660 000 to 1 000 000 SNPs).[17-19] This correction method is very conservative since the association findings of each SNP are considered to be independent, and the haplotype structure of the genome is not taken into account. Conservative correction for multiple testing reduces the risk of false-positive findings, but hampers the detection of true association signals that represent small effects on disease risk.

Following the publication of the first GWAS in age-related macular degeneration,[20] successful GWASs have been conducted for a variety of common, complex diseases including type 2 diabetes, myocardial infarction, breast cancer, and Crohn's disease (for details of all published studies see http://www.genome.gov/gwastudies/).

Schizophrenia

The first GWASs for schizophrenia have recently been published.[21-30] Three of these studies used pooled DNA samples.[21,26,27] The best supported variants in these three studies failed to achieve genome-wide significance [21,26,27] *(Table I)*. This is a cost-effective method of performing GWASs and has proved to be effective in identifying disease genes (eg, refs 31,32). However, due to errors in DNA quantification, this method is less sensitive than individual genotyping and has less power. Furthermore, the evaluation of data is limited to the study of (estimated) allele frequencies at the level of individual SNPs. This method does not detect the effect of haplotypes, interactions between SNPs, or the effects of genotypes that do not show differences in allele frequencies. The first individual-genotyping-based GWAS of schizophrenia involved a very small sample of 178 cases and 144 controls.[29] The best hit was for a variant near the *colony-stimulating factor-2 receptor alpha* (*CSF2RA*) gene, but this did not achieve genome-wide significance.[29] The second GWAS of this type included 738 patients and 733 controls. Although a few signals coincided with genomic regions that had been implicated in previous linkage studies of schizophrenia, this study found no genome-

Study	N° SNPs analyzed	Supported gene	Supported variant	Genomic region	*P* value discovery	N° samples discovery	*P* value combined	N° samples, replication/ meta-analysis
Mah et al (2006)	~ 25 000	*plexin A2* (*PLXNA2*)	rs752016	1q32.2	0.006	320 cases 325 controls	0.035	200 cases (EA) 230 controls (EA)
Lencz et al (2007)	~ 500 000	*colony stimulating factor receptor 2 alpha* (*CSF2RA*)	rs4129148	Xp22.33 Yp22.32	3.7×10^{-7}	178 cases 144 controls	ND	ND
Sullivan et al (2008)	~ 500 000	nearest gene: *angiotensin II receptor-associated protein* (*AGTRAP*)	rs4846033	1p36.22	4.4×10^{-6}	738 cases 733 controls	ND	ND
O´Donovan et al (2008)	~ 500 000	*zinc finger protein 804A* (*ZNF804A*)	rs1344706	2q32.1	1.8×10^{-6}	479 cases 2937 controls	1.6×10^{-7}	7308 cases 12834 controls
Shifman et al (2008)	~ 500 000	*reelin* (*RELN*)	rs7341475	7q22.1	2.9×10^{-5} (in females)	745 cases 2644 controls	8.8×10^{-7} (in females)	2274 cases 4401 controls
Kirov et al (2009)	~ 550 000	*coiled coiled domain containing 60* (*CCDC60*)	rs11064768	12q24.23	1.2×10^{-6}	574 trios	ND	ND
Need et al (2009)	~ 550 000	*ADAMTS like 3* (*ADAMTSL3*)	rs2135551	15q25.2	1.3×10^{-6}	871 cases 863 controls	NR	1460 cases 12995 controls
Shi et al (2009)	~ 600 000	*ArfGAP with GTPase domain, ankyrin repeat and PH domain 1* (*AGAP1*)	rs13025591	2q37.2	4.6×10^{-7} (in EA)	2681 cases 2653 controls (EA)		ND
		v-erb-a erythroblastic leukemia viral oncogene homolog 4 (*avian*) (*ERBB4*)	rs1851196	2q34	2.1×10^{-6} (in AA)	1286 cases 973 controls (AA)		ND
		major histocompatibility complex (*MHC*)	rs9272219 rs9272535 **rs13194053**	6p21.32 6p21.32 6p22.1	ND ND 1.4×10^{-2} (in EA)		6.9×10^{-8} 8.9×10^{-8} 9.5×10^{-9}	8008 cases (EA) 19077 controls (EA)
		cluster of histone protein genes						
The International Schizophrenia Consortium (2009)	~ 1 000 000	*myosin XVIIIB* (*MYO18B*)	rs5761163	22q12.1	3.4×10^{-7}	3322 cases 3587 controls	ND 9.5×10^{-9}	8008 cases 19077 controls
		major histocompatibility complex (*MHC*)	rs13194053	6p22.1	ND			
Stefansson et al (2009)	~ 300 000	*major histocompatibility complex* (*MHC*)	**5 variants**	6p21.3 – 6p22.1	0.0027-0.00023	2663 cases 13498 controls	**1.1×10^{-9}-1.4×10^{-12}**	12945 cases 34591 controls
		neurogranin (*NRGN*)	**rs12807809**	11q24.2	0.00045		**2.4×10^{-9}**	
		transcription factor 4 (*TCF4*)	**rs9960767**	18q21.1	0.0011		**4.1×10^{-9}**	

Table I. Published genome-wide association studies (GWASs) for schizophrenia.[21-30,32] The number of variants investigated, the best associated single-nucleotide polymorphism(s)–SNP(s)—found and the gene(s) containing the SNP(s), the corresponding *P* value(s), and the number of cases and controls in the discovery and the replication/meta-analysis sample are all given. Genome-wide significant findings are highlighted in bold. EA, European Ancestry Individuals; AA, African-American Individuals; ND, no data available; NR, no replication

wide significant association.[30] O'Donovan et al initially performed a GWAS using a moderately sized patient sample (n=479). They then performed a follow-up study of 12 markers with a *P* value $\leq 10^{-5}$ in a much larger sample to enhance the statistical power.[25] Strong evidence for replication was obtained for 3 of these 12 markers ($P \leq 5 \times 10^{-4}$), although the best supported variant still failed to achieve genome-wide significance *(Table I)*. The highest-ranking SNP identified in this study is located in an intron of the *zinc finger protein 804A* gene (*ZNF804A*), a putative transcription factor which had never been implicated previously in the risk for schizophrenia. The case sample was then extended to include bipolar patients. The *P* value for the total sample surpassed the level of genome-wide significance ($P=9 \times 10^{-9}$). The association between *ZNF804A* and schizophrenia has recently been replicated by the International Schizophrenia Consortium,[24] and *ZNF804A* is therefore a promising susceptibility gene for schizophrenia. A recent imaging genetics study of *ZNF804A* risk genotypes has provided evidence in support of these genetic findings. This study demonstrated that healthy carriers of *ZNF804A* risk genotypes display pronounced gene-dosage-dependent alterations in functional coupling between the hippocampus and the dorsolateral prefrontal cortex (DLPFC) across the two hemispheres, which mirrors findings in patients.[33]

Three recent multicenter studies have provided important insights. The initial findings of these three studies failed to surpass the level of genome-wide significance. However, a meta-analysis was then performed using the best hits from the European data of these studies and data from a replication study by Stefansson et al.[22] This revealed a cluster of genome-wide significant SNPs in the major histocompatibility (MHC) region of chromosome 6p22.1 that were in substantial linkage disequilibrium.[22-24] These results provide evidence that the immunological system may play a role in the pathogenesis of schizophrenia. Furthermore, a variant upstream of *neurogranin* (*NRGN*; $P=2.4 \times 10^{-9}$) and a SNP in *transcription factor 4* (*TCF4*; $P= 4.1 \times 10^{-9}$) achieved genome-wide significance in Stefansson et al's study.[22] These studies demonstrate that GWASs of large samples can overcome limitations in power and detect common risk variants for complex psychiatric disorders.

In the study by the International Schizophrenia Consortium, it was demonstrated that possible risk variants may have been among the nominally significant SNPs that failed to reach genome-wide significance. Nominally significant SNPs were grouped into a "set of score alleles" and analyzed in an independent case-control sample, and it was shown that they distinguished cases from controls.[24] This study also demonstrated that this set of genes distinguished bipolar cases from controls, thus providing further evidence for a genetic overlap between schizophrenia and bipolar disorder. Although these SNPs explained only approximately 3% of the variance in schizophrenia risk, this may be regarded as a step towards molecular genetic evidence for the polygenic inheritance of schizophrenia.

Bipolar disorder

Six GWASs have been published to date for bipolar disorder[34-39] *(Table II)* including the landmark study by the Wellcome Trust Case Control Consortium (WTCCC) which investigated seven common disorders.[36] These studies were all based upon individual genotyping, with the exception of the study by Baum et al[39] which involved DNA pooling. Although there has been some inconsistency across studies in terms of their most associated genomic regions,[35-39] meta-analyses of some of these studies have revealed common association signals. A meta-analysis of the Baum et al[39] and the WTCCC[36] datasets found a consistent association between bipolar disorder and variants in the genes *junction adhesion molecule 3* (*JAM3*) (rs10791345, $P=1 \times 10^{-6}$), and *solute carrier family 39 (zinc transporter), member 3* (*SLC39A3*) (rs4806874, $P=5 \times 10^{-6}$).[40] A combined analysis of the Sklar et al[35] and WTCCC[36] studies, which included a total of 4387 patients and 6209 controls, identified the first genome-wide significant association signal for bipolar disorder for *ankyrin 3, node of Ranvier* (*ANK3*) (rs10994336, $P=9.1 \times 10^{-9}$).[34] The second most strongly associated region was marked rs1006737 in *calcium channel, voltage-dependent, L type, alpha 1C subunit CACNA1C* ($P=7 \times 10^{-8}$). Further independent support for *ANK3* rs10994336 has recently been obtained by Schulze et al[41] in samples from Germany and the United States (US); this study also found evidence for allelic heterogeneity at the *ANK3* locus.

Although GWASs of bipolar disorder have identified a number of potentially relevant genetic variants, the widely acknowledged formal threshold for genome-wide significance of $P=5 \times 10^{-8}$ has only been surpassed so far for variation in *ANK3*.

Study	N° SNPs analyzed	Supported gene	Supported variant	Genomic region	*P* value discovery	N° samples, discovery	*P* value combined	N° samples, replication/ meta-analysis
Baum et al (2007)	~ 550 000	*diacylglycerol kinase eta (DGKH)*	rs1012053	13q14.11	0.0002	461 cases 563 controls	1.5×10^{-8}	772 cases 876 controls
Welcome Trust Case Control Consortium (WTCCC; 2007)	~ 500 000	*partner and localizer of BRCA2 (PALB2)*	rs42059	7q21.3	6.3×10^{-8}	1868 cases 2938 controls	ND	ND
Sklar et al (2008)	~ 400 000	*tetraspanin-8 (TSPAN8)* *myosin5B (MYO5B)* *voltage-dependent calcium channel, L-type, alpha 1C subunit (CACNA1C)*	rs1705236 rs4939921 rs1006737	12q21.1 18q21.1 12p13.33	6.1×10^{-7} 1.7×10^{-7} 8.8×10^{-4}	1461 cases 2008 controls	NR NR 3.1×10^{-6}	3329 cases 4946 controls
Ferreira et al (2008)	~ 1 800 000 (imputed) ~ 300 000 (genotyped)	ankyrin G (ANK3) *voltage-dependent calcium channel, L-type, alpha 1C subunit (CACNA1C)*	**rs10994336** **rs1938526** rs1006737	10q21.2 10q21.2 12p13.33	0.0002 0.0002 0.0108	1098 cases 1267 controls	**9.1×10^{-9}** **1.3×10^{-8}** 7.0×10^{-8}	4387 cases 6209 controls
Scott et al (2009)	~ 550 000	*inter-alpha (globulin) inhibitor H1 (ITIH1)* *multiple C2 domains, transmembrane 1 (MCTP1)* *nuclear factor 1 A-type (NF1A)*	rs1042779 rs17418283 rs472913	3p21.1 5q15 1p32.1	ND	2076 cases 1676 controls	1.8×10^{-7} 1.3×10^{-7} 2.0×10^{-7}	3683 cases 14507 controls
Smith et al (2009)	~ 700 000	*nck-associated protein 5 (NAP5)* *dpy-19-like 3 (DPY19L3)*	rs10193871 rs2111504	2q21.2 19q13.11	9.8×10^{-6} 1.5×10^{-6}	1001 cases 1033 controls (EA) 345 cases 670 controls (AA)	ND	ND

Table II. Published genome-wide association studies (GWASs) for bipolar disorder.[34-39] The number of variants investigated, the best associated single-nucleotide polymorphism(s)–SNP(s)—found and the gene(s) containing that SNP(s), the corresponding *P* value(s), and the number of cases and controls in the discovery and the replication/meta-analysis sample are all given. Genome-wide significant findings are highlighted in bold. EA, European Ancestry Individuals; AA, African-American Individuals; ND, no data available; NR, no replication

Future studies involving larger samples, the pooling of datasets, and higher statistical power are expected to identify additional specific risk factors for bipolar disorder and schizophrenia.

Copy number variations

Small chromosomal aberrations (microdeletions and microduplications, collectively known as copy number variations, CNV) may confer a risk for schizophrenia, as illustrated by the 22q11.2 deletion syndrome (22q11.2DS). This is a common microdeletion syndrome with congenital and late-onset features. Patients have a high risk for neuropsychiatric diseases including psychotic disorders and major depression.[42-44] It has not been possible to correlate the extent of the deletion with the occurrence of schizophrenia in these patients, and there is experimental evidence that increased susceptibility may require the altered expression of several genes within the 22q11.2 region.[45,46] This may explain why no replicable results have

been obtained from attempts to implicate individual genes within the deletion region as susceptibility genes for schizophrenia.[47]

Schizophrenia

The application of new technologies such as comparative genomic hybridization (CGH) and SNP arrays in GWASs has enabled the identification of small chromosomal aberrations on a genome-wide scale. Initial studies reported an increased rate of aberrations in schizophrenia[48,49] and subsequent studies have implicated specific chromosomal regions.[28,50-54] Implicated aberrations include microdeletions in chromosomal regions 1q21.1, 2p16.3, 15q11.2, and 15q13.3, as well as microduplications in chromosomal regions 15q13.1 and 16p11.2. Although all of these variants are observed more frequently in patients than in controls (with odds ratios of >10 for some variants), the frequency of each individual variant in schizophrenia patients is low (<1%). Further studies are required to determine the penetrance and mutation rate of these aberrations, as well as their phenotypic spectrum. Research has shown that some variants also occur more frequently in patients with other central nervous system phenotypes such as autism, mental disability, and epilepsy.[55-58] The mechanisms that underlie the phenotypic outcome however, remain unknown. The fact that de novo mutations are found in a proportion of patients with CNVs supports the hypothesis that the negative effect on reproductive fitness observed in schizophrenia patients may be at least partly offset by the occurrence of new mutations.

Bipolar disorder

There have been few CNV studies of bipolar disorder.[59-61] Lachman et al investigated a mixed cohort of Caucasian patients (n=227) and controls (n=276) from the Czech Republic and the United States, and found that CNVs involving the gene *glycogen synthase kinase 3 beta* (*GSK3beta*) were significantly increased in patients compared with controls.[59] Using a European American sample of 1001 BD patients and 1034 controls, Zhang et al investigated singleton microdeletions (ie, those occurring only once in the total dataset of patients and controls) of more than 100 kb and found that they were overrepresented in patients.[60] The effect was strongest in a subgroup of patients with an early onset of mania (<18 years of age). A recent study of a three-generation Older Amish pedigree with segregating affective disorder[61] identified a set of 4 CNVs on chromosomes 6q27, 9q21, 12p13, and 15q11 that were enriched in affected family members and which altered the expression of neuronal genes.

No CNV with a genetic effect comparable to those identified for neuropsychiatric disorders such as schizophrenia or autism has yet been identified for bipolar disorder. In view of the limited number of studies performed, it is not possible to evaluate the influence of CNVs on disease development.

Outlook

The first GWASs of schizophrenia and bipolar disorder have recently been published, and many more are in progress. Large international collaborations have been initiated to combine GWAS data sets in order to increase statistical power, the largest being the Psychiatric GWAS Consortium, which is expected to publish its first results in 2010 (The Psychiatric GWAS Consortium Steering Committee 2009). Currently available research findings suggest that the variants identified through GWASs confer only small individual risks. The major limitation of GWASs is that they are only able to investigate common variants. If a large fraction of the genetic contribution is conferred by rare variants, other approaches will be necessary to identify them. A successful first step in this direction has been the identification of associations between rare CNVs and psychiatric diseases, in particular schizophrenia. However, due to methodological constraints, this approach remains restricted to the investigation of aberrations of at least several thousand base pairs. Continuing technological developments will provide future studies with increasing resolution, and the availability of low-cost whole genome sequencing technology will ultimately make it possible to obtain the complete genomic sequences of large patient samples for comparison with controls. In principle, this will allow the systematic identification of rare variants that are associated with disease risk, although the existence of a myriad of rare variants in the human genome will render this a complex task. It is hoped that some rare variants confer a larger disease risk, as this will facilitate the detection of association in large case-control samples. Rare variants with small disease risk may be extremely difficult to detect, since prohibitively large sample sizes may be required to demonstrate any significant association.

It is likely, however, that even after the identification of all common and rare risk variants a substantial fraction of the familial clustering will remain unexplained. This "missing heritability" in complex diseases is the subject of intense debate and several potential explanations have been proposed, including epistasis and epigenetic mechanisms.[62-64] It will be necessary to apply specific research strategies to further investigate this issue, although these may require prohibitively large sample sizes or tissue samples that are difficult to access in human subjects.

It is not yet clear whether any of the association findings identified by GWASs represent causal variants. Systematic resequencing of the associated genomic regions will provide a comprehensive overview of such variants. In cases where association findings are due to linkage disequilibrium, it is possible that the causal variants have a stronger genetic effect than has been previously suspected. It is also theoretically possible that a given association finding is not attributable to a common causal variant. A simulation study has shown that the "synthetic" effect of multiple rare variants may be responsible for signals detected for common variants. It has also been shown that the location of these variants may be relatively far (up to 2 megabases) from the site identified in GWASs.[65] If this were the case for an associated locus, resequencing over large genomic distances in large samples would be required to identify the true causative variants. Ultimately, it is necessary to identify a direct functional effect for each potential causal variant, such as an effect on the function or expression of a gene.

GWASs performed to date have indicated that certain genes contribute to a susceptibility to both schizophrenia and bipolar disorder. It is clear that some of these genes convey a rather nonspecific susceptibility that overlaps diagnostic boundaries, and it is highly probable that this also overlaps with other psychiatric disorders. Other genes, however, convey specific effects. Future studies of the phenotypic dimensions that are most strongly associated with a specific gene will include analysis of clinical symptoms and endophenotypes. The latter may be particularly suited to guiding researchers in the selection of the most promising phenotypes for animal studies.[66]

The identification of disease-associated genes is likely to increase our knowledge of the underlying pathophysiology of psychiatric disorders in an as-yet unforeseen manner. The identification of biological pathways has the potential to revolutionize diagnostics and treatment. ❑

REFERENCES

1. Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry*. 2003;60:1187-1192.
2. Craddock N, Forty L. Genetics of affective (mood) disorders. *Eur J Hum Genet*. 2006;14:660-668.
3. Gottesman, II, Hanson DR. Human development: biological and genetic processes. *Annu Rev Psychol*. 2005;56:263-286.
4. Kato T, Kuratomi G, Kato N. Genetics of bipolar disorder. *Drugs Today (Barc).* 2005;41:335-344.
5. Hattori E, Liu C, Zhu H, Gershon ES. Genetic tests of biologic systems in affective disorders. *Mol Psychiatry*. 2005;10:719-740.
6. Levinson DF. The genetics of depression: a review. *Biol Psychiatry*. 2006;60:84-92.
7. McGuffin P. Nature and nurture interplay: schizophrenia. *Psychiatr Prax*. 2004;31 (suppl 2):S189-193.
8. Abdolmaleky HM, Thiagalingam S, Wilcox M. Genetics and epigenetics in major psychiatric disorders: dilemmas, achievements, applications, and future scope. *Am J Pharmacogenomics*. 2005;5:149-160.
9. Burmeister M, McInnis MG, Zollner S. Psychiatric genetics: progress amid controversy. *Nat Rev Genet*. 2008;9:527-540.
10. Thapar A, Harold G, Rice F, Langley K, O'Donovan M. The contribution of gene-environment interaction to psychopathology. *Dev Psychopathol*. 2007;19:989-1004.
11. van Os J, Rutten BP, Poulton R. Gene-environment interactions in schizophrenia: review of epidemiological findings and future directions. *Schizophr Bull*. 2008;34:1066-1082.
12. Chumakov I, Blumenfeld M, Guerassimenko O, et al. Genetic and physiological data implicating the new human gene G72 and the gene for D-amino acid oxidase in schizophrenia. *Proc Natl Acad Sci U S A*. 2002;99:13675-13680.
13. Stefansson H, Sigurdsson E, Steinthorsdottir V, et al. Neuregulin 1 and susceptibility to schizophrenia. *Am J Hum Genet*. 2002;71:877-892.
14. Straub RE, Jiang Y, MacLean CJ, et al. Genetic variation in the 6p22.3 gene DTNBP1, the human ortholog of the mouse dysbindin gene, is associated with schizophrenia. *Am J Hum Genet*. 2002;71:337-348.
15. Millar JK, Wilson-Annan JC, Anderson S, et al. Disruption of two novel genes by a translocation co-segregating with schizophrenia. *Hum Mol Genet*. 2000;9:1415-1423.
16. Jaaro-Peled H, Hayashi-Takagi A, Seshadri S, Kamiya A, Brandon NJ, Sawa A. Neurodevelopmental mechanisms of schizophrenia: understanding disturbed postnatal brain maturation through neuregulin-1-ErbB4 and DISC1. *Trends Neurosci*. 2009;32:485-495.
17. Dudbridge F, Gusnanto A. Estimation of significance thresholds for genomewide association scans. *Genet Epidemiol*. 2008;32:227-234.
18. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol*. 2008;32:381-385.
19. Hoggart CJ, Clark TG, De Iorio M, Whittaker JC, Balding DJ. Genome-wide significance for dense SNP and resequencing data. *Genet Epidemiol*. 2008;32:179-185.
20. Klein RJ, Zeiss C, Chew EY, et al. Complement factor H polymorphism in age-related macular degeneration. *Science*. 2005;308:385-389.
21. Kirov G, Zaharieva I, Georgieva L, et al. A genome-wide association study in 574 schizophrenia trios using DNA pooling. *Mol Psychiatry*. 2009;14:796-803.

Nuevos hallazgos en la genética de las principales psicosis

La esquizofrenia y el trastorno bipolar tienen una etiología y una fisiopatología en gran medida desconocidas, pero son altamente heredables. Aunque los estudios de ligamiento y de asociación han identificado una serie de regiones cromosómicas que contienen probablemente genes susceptibles, el progreso en la identificación de genes causales ha sido muy decepcionante. Sin embargo, los rápidos avances tecnológicos están dando origen a nuevos conocimientos. Los estudios sistemáticos de asociación del genoma completo y de seguimiento han informado acerca de hallazgos de asociación significativa del genoma completo y variantes comunes para la esquizofrenia y el trastorno bipolar. El riesgo que determinan las variantes individuales es pequeño y algunas de ellas confieren un riesgo para ambos trastornos. Además, estudios recientes han identificado variantes estructurales largas y raras (variantes de número de copias) que otorgan un mayor riego para la esquizofrenia. Esta revisión resume los desarrollos recientes en la investigación genética de la esquizofrenia y del trastorno bipolar y discute las posibles direcciones futuras en este campo.

Nouvelles découvertes en génétique des principales psychoses

La physiopathologie et l'étiologie de la schizophrénie et des troubles bipolaires restent largement méconnues mais fortement héréditaires. Des études de liaison et d'association ont identifié des séries de régions chromosomiques contenant probablement des gènes de susceptibilité, mais l'identification des gènes de causalité est extrêmement décevante. Des avancées technologiques rapides commencent cependant à voir le jour. Des études systématiques d'association sur le génome entier et de suivi ont découvert une association significative au niveau du génome entier de variants communs pour la schizophrénie et les troubles bipolaires. Le risque inhérent aux variants individuels est faible, et certains variants comportent un risque pour les deux pathologies. De plus, des études récentes ont identifié des variants structurels importants et rares (CNV = copy number variants*, variants du nombre de copies) liés à un risque plus élevé de schizophrénie. Cet article résume les avancées récentes de la recherche génétique concernant la schizophrénie et les troubles bipolaires et analyse les perspectives possibles dans ce domaine.*

22. Stefansson H, Ophoff RA, Steinberg S, et al. Common variants conferring risk of schizophrenia. *Nature*. 2009;460:744-747.
23. Shi J, Levinson DF, Duan J, et al. Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature*. 2009;460:753-757.
24. Purcell SM, Wray NR, Stone JL, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460:748-752.
25. O'Donovan MC, Craddock N, Norton N, et al. Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet*. 2008;40:1053-1055.
26. Shifman S, Johannesson M, Bronstein M, et al. Genome-wide association identifies a common variant in the reelin gene that increases the risk of schizophrenia only in women. *PLoS Genet*. 2008;4:e28.
27. Mah S, Nelson MR, Delisi LE, et al. Identification of the semaphorin receptor PLXNA2 as a candidate for susceptibility to schizophrenia. *Mol Psychiatry*. 2006;11:471-478.
28. Need AC, Ge D, Weale ME, et al. A genome-wide investigation of SNPs and CNVs in schizophrenia. *PLoS Genet*. 2009;5:e1000373.
29. Lencz T, Morgan TV, Athanasiou M, et al. Converging evidence for a pseudoautosomal cytokine receptor gene locus in schizophrenia. *Mol Psychiatry*. 2007;12:572-580.
30. Sullivan PF, Lin D, Tzeng JY, et al. Genomewide association for schizophrenia in the CATIE study: results of stage 1. *Mol Psychiatry*. 2008;13:570-584.
31. Johnson C, Drgon T, Liu QR, et al. Pooled association genome scanning for alcohol dependence using 104,268 SNPs: validation and use to identify alcoholism vulnerability loci in unrelated individuals from the collaborative study on the genetics of alcoholism. *Am J Med Genet B Neuropsychiatr Genet*. 2006;141B:844-853.
32. Liu QR, Drgon T, Walther D, et al. Pooled association genome scanning: validation and use to identify addiction vulnerability loci in two samples. *Proc Natl Acad Sci U S A*. 2005;102:11864-11869.
33. Esslinger C, Walter H, Kirsch P, et al. Neural mechanisms of a genome-wide supported psychosis variant. *Science*. 2009;324:605.
34. Ferreira MA, O'Donovan MC, Meng YA, et al. Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat Genet*. 2008;40:1056-1058.
35. Sklar P, Smoller JW, Fan J, et al. Whole-genome association study of bipolar disorder. *Mol Psychiatry*. 2008;13:558-569.
36. WTCCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,00 shared controls. *Nature*. 2007;447:661-678.
37. Scott LJ, Muglia P, Kong XQ, et al. Genome-wide association and meta-analysis of bipolar disorder in individuals of European ancestry. *Proc Natl Acad Sci U S A*. 2009;106:7501-7506.
38. Smith EN, Bloss CS, Badner JA, et al. Genome-wide association study of bipolar disorder in European American and African American individuals. *Mol Psychiatry*. 2009;14:755-763.

39. Baum AE, Akula N, Cabanero M, et al. A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. *Mol Psychiatry*. 2008;13:197-207.
40. Baum AE, Hamshere M, Green E, et al. Meta-analysis of two genome-wide association studies of bipolar disorder reveals important points of agreement. *Mol Psychiatry*. 2008;13:466-467.
41. Schulze TG, Detera-Wadleigh SD, Akula N, et al. Two variants in Ankyrin 3 (ANK3) are independent genetic risk factors for bipolar disorder. *Mol Psychiatry*. 2009;14:487-491.
42. Green T, Gothelf D, Glaser B, et al. Psychiatric disorders and intellectual functioning throughout development in velocardiofacial (22q11.2 deletion) syndrome. *J Am Acad Child Adolesc Psychiatry*. 2009;48:1060-1068.
43. Karayiorgou M, Morris MA, Morrow B, et al. Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. *Proc Natl Acad Sci U S A.* 1995;92:7612-7616.
44. Bassett AS, Chow EW, Husted J, et al. Clinical features of 78 adults with 22q11 Deletion Syndrome. *Am J Med Genet A*. 2005;138:307-313.
45. Sivagnanasundaram S, Fletcher D, Hubank M, Illingworth E, Skuse D, Scambler P. Differential gene expression in the hippocampus of the Df1/+ mice: a model for 22q11.2 deletion syndrome and schizophrenia. *Brain Res*. 2007;1139:48-59.
46. Meechan DW, Maynard TM, Gopalakrishna D, Wu Y, LaMantia AS. When half is not enough: gene expression and dosage in the 22q11 deletion syndrome. *Gene Expr*. 2007;13:299-310.
47. Glaser B, Moskvina V, Kirov G, et al. Analysis of ProDH, COMT and ZDHHC8 risk variants does not support individual or interactive effects on schizophrenia susceptibility. *Schizophr Res*. 2006;87:21-27.
48. Walsh T, McClellan JM, McCarthy SE, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science*. 2008;320:539-543.
49. Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M. Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet*. 2008;40:880-885.
50. Rujescu D, Ingason A, Cichon S, et al. Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum Mol Genet*. 2009;18:988-996.
51. Stefansson H, Rujescu D, Cichon S, et al. Large recurrent microdeletions associated with schizophrenia. *Nature*. 2008;455:232-236.
52. Kirov G, Gumus D, Chen W, et al. Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. *Hum Mol Genet*. 2008;17:458-465.
53. International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*. 2008;455:237-241.
54. McCarthy SE, Makarov V, Kirov G, et al. Microduplications of 16p11.2 are associated with schizophrenia. *Nat Genet*. 2009;41:1223-1227.
55. Mefford HC, Sharp AJ, Baker C, et al. Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N Engl J Med*. 2008;359:1685-1699.
56. Helbig I, Mefford HC, Sharp AJ, et al. 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. *Nat Genet*. 2009;41:160-162.
57. Ben-Shachar S, Lanpher B, German JR, et al. Microdeletion 15q13.3: a locus with incomplete penetrance for autism, mental retardation, and psychiatric disorders. *J Med Genet*. 2009;46:382-388.
58. Miller DT, Shen Y, Weiss LA, et al. Microdeletion/duplication at 15q13.2q13.3 among individuals with features of autism and other neuropsychiatric disorders. *J Med Genet*. 2009;46:242-248.
59. Lachman HM, Pedrosa E, Petruolo OA, et al. Increase in GSK3beta gene copy number variation in bipolar disorder. *Am J Med Genet B Neuropsychiatr Genet*. 2007;144B:259-265.
60. Zhang D, Cheng L, Qian Y, et al. Singleton deletions throughout the genome increase risk of bipolar disorder. *Mol Psychiatry*. 2009;14:376-380.
61. Yang S, Wang K, Gregory B, et al. Genomic landscape of a three-generation pedigree segregating affective disorder. *PLoS ONE*. 2009;4:e4474.
62. Maher B. Personal genomes: the case of the missing heritability. *Nature*. 2008;456:18-21.
63. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461:747-753.
64. Slatkin M. Epigenetic inheritance and the missing heritability problem. *Genetics*. 2009;182:845-850.
65. Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. Rare variants create synthetic genome-wide associations. *PLoS Biol*. 8:e1000294.
66. Gould TD, Gottesman, II. Psychiatric endophenotypes and the development of valid animal models. *Genes Brain Behav*. 2006;5:113-119.

Single gene disorders come into focus—again

Hans-Hilger Ropers, MD, PhD

For more than 10 years, genome research has focused on finding genetic risk factors for common disorders, based on the "common disease—common variant (CDCV)" hypothesis—the intuitive but unproven assumption that for most of the common disorders like dementia, diabetes, coronary heart disease, autism, and hypertension, there are common genetic risk factors. Since 2007, after many years of growing frustration with the disappointing results of genome-wide association studies (GWAS), associated markers were identified for a wide variety of complex disorders; this was hailed as a decisive breakthrough in this field. However, these associations were only found after massively increasing cohort sizes and marker densities, meaning that the vast majority of the associated risk factors have small effects and that they are of no diagnostic and prognostic relevance. Moreover, many markers were found to be located in noncoding sequences, and thus, very few pro-

In the early 1990s, when the second 5-year plan for the Human Genome Project—which requested more money than any previous research project in biology—was written, common disorders were presented as the future target of genome research. This was a clever move to ensure continued public support for this endeavor, which had been justified previously by the prospect that it would lead to the diagnosis, prevention, and therapy of severe, but mostly rare, Mendelian disorders. Today, more than 15 years later, after billions of dollars have been spent on genome-wide association studies (GWAS), very few major genetic risk factors for common diseases have been identified, and the enthusiasm for large GWAS is dwindling. At the same time, there is renewed interest for studying single gene disorders, which are now considered by some as a better clue to the understanding of common diseases. While this is probably true, Mendelian disorders are also important in their own right, since they must be far more common than generally thought. As discussed here, various efficient strategies exist for the elucidation of single gene defects, and their systematic application in combination with novel genome partitioning and massive parallel sequencing techniques, will have far-reaching implications for health care.

Dialogues Clin Neurosci. 2010;12:95-102.

Keywords: *single gene disorder; genome research; mental retardation; next-generation sequencing; health care*

Author affiliations: Max Planck Institute for Molecular Genetics, Berlin, Germany

Address for correspondence: Max Planck Institute for Molecular Genetics, Ihnestrasse 73, D-14195 Berlin, Germany
(e-mail: ropers@molgen.mpg.de)

vided novel insights into the underlying pathogenetic mechanisms. Ironically, therefore, very shortly after this "breakthrough," there is growing support for the notion that for most common disorders, the CDCV hypothesis must be wrong.[1,2]

This is certainly true for mental retardation (MR)—the biggest unsolved problem of clinical genetics and the largest socioeconomic burden of health care—where most severe forms are due to defined chromosomal abnormalities or single gene defects, instead of resulting from multifactorial inheritance, ie, the interaction of many different gene variants and environmental factors. However, there is increasing evidence that single gene defects also play a significant, previously underestimated, role in other complex disorders. This has led to growing uneasiness about the validity of the idea that GWAS is the preferred approach for identifying sequence variants in the human genome that predispose to, or cause, disease. Moreover, it has raised serious doubts about the strategy, first proposed in the early 1990s and uncritically adopted by leading genome centers worldwide, to focus exclusively on complex disorders.

After the introduction of massive parallel next-generation sequencing techniques, there are now indications for a paradigm shift in this field, with a renewed focus on single gene disorders. At a recent meeting,[3] two groups reported on their efforts to unravel the molecular basis of Mendelian disorders by sequencing all exons in the genomes of patients and their unaffected parents. Moreover, leading genome researchers expressed their belief that instead of GWAS, whole genome sequencing-based, large-scale elucidation of single gene disorders will be the strategy of choice for shedding more light on the molecular architecture of common disorders.

In the late 1980s, before common disorders were proclaimed as the central target of genome research, along with overly optimistic assumptions about the medical implications of this research, the revolutionary and costly project to elucidate the structure of the human genome had been justified by the prospect that it would lead to unambiguous diagnosis, prevention, and, eventually, therapies for severe Mendelian disorders. Now, almost 20 years after the official commencement of the Human Genome Project, and 6 years after its completion, it appears that genome research is coming around full circle by once again focusing on single gene defects.

Single gene defects are important for health care

Single gene defects have significance in their own right. In contrast to many complex disorders such as type 2 diabetes and obesity, which are lifestyle-related, become manifest only later in life, or are relatively mild, single gene disorders are mostly severe, early-onset conditions, necessitating lifelong care and support. Moreover, single gene disorders are far more numerous than generally assumed, and as a group, they are certainly not rare.

According to OMIM, the comprehensive catalogue of human traits that are inherited in a Mendelian fashion (http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim), only slightly more than 2500 human genes have been linked to disease, and there are approximately 3500 Mendelian diseases for which the molecular cause is not yet known. It is likely, however, that this is a wide underestimate, and that the number of genes which are indispensable for normal embryonic and postnatal development, homeostasis, and aging is much higher.

In mice with induced defects of single genes (ie, "knockout mice"), conspicuous (disease) phenotypes or embryonic lethality are the rule rather than the exception, as discussed elsewhere.[2] In humans, the proportion of gene defects that are associated with recognizable disorders must be even higher, because relatively subtle (eg, behavioral) abnormalities are readily detectable in man, even without specific clinical examination. Milder mutations in the same genes known to cause embryonic lethality when affected by loss-of-function mutations may be compatible with life but also cause disease.

Functional considerations and empirical data from model organisms suggest that most disease-associated gene defects are inherited as recessive traits. At least in Western societies, this means that most patients will be isolated cases, due to small family sizes and the fact that in these populations, parental consanguinity is rare. In sporadic cases without specific, previously described combinations of clinical symptoms, single gene defects are unlikely to be considered as the underlying cause. In particular, this holds for patients with complex disorders and presumed multifactorial inheritance. Thus, as discussed for MR, it is likely that many Mendelian disorders have not been identified yet because in the well-studied Western populations, they do not segregate in families. Irrespective of family sizes and parental con-

sanguinity, this also holds for all severe autosomal dominant disorders conferring a significant reproductive disadvantage (eg, severe mental handicaps). Most of these patients will carry new mutations and therefore will be isolated cases as well.

For most common diseases, the possibility that there is a sizable "contamination" by monogenic forms has not been excluded, and the proportion of cases that are due to single gene defects is hitherto unknown. As indicated above, this does not hold for MR, however. Prompted by the early observation that males are more often affected than females,[4] and by the description of several large families where MR segregated in an X-linked fashion (see ref 5, for example), the hypothesis that single gene defects on the X chromosome play a major role in MR was put forward in the early 1970s.[6,7]

Since the 1990s, genetic research into the molecular causes of MR has focused on X-chromosomal genes,[8] and at the time of writing (September 2009), mutations in 90 X-chromosomal genes have been implicated in Mendelian forms of MR, demonstrating that this condition is extremely heterogeneous. Surprisingly, screening of several hundred families with X-linked MR (XLMR) has revealed that these 90 genes account for at most 50% of all mutations[9]; see also ref 10. This means that there must be many more genes on the X chromosome which are indispensable for the normal function of the human brain. The X chromosome carries about 4% of all human genes, and even though there is evidence suggesting that on the X, the density of MR genes is higher than on autosomes,[11] extrapolation of these data suggests that defects in several thousand human genes may give rise to cognitive dysfunction. However, the systematic search for these autosomal MR genes has only just begun, as discussed below.

There is increasing evidence that single gene defects also have important roles in the etiology of other complex disorders. For example, several homozygous deletions were recently described in autistic offspring of healthy consanguineous parents,[12] strongly suggesting that autosomal recessive gene defects are important causes of autism, too. In view of the growing molecular evidence that MR, autism, and schizophrenia are etiologically related,[2,13] it is likely that many cases of schizophrenia are also due to a variety of single gene defects. There is reason to believe that the same holds true for many other complex diseases that are generally considered multifactorial.[14]

Systematic elucidation of single gene disorders

There are various efficient strategies for elucidating the molecular defects underlying Mendelian disorders, as discussed in detail elsewhere.[2] Most of them consist of two steps, the chromosomal and regional mapping of the relevant defect and the search for mutations in positional and functional candidate genes.

Disease-associated balanced chromosome rearrangements

Systematic breakpoint mapping and cloning in patients with disease-associated balanced chromosome rearrangements (DBCRs) has been employed by several groups to identify genes that are truncated or inactivated by the rearrangement *(Figure 1a)*. Most de novo balanced chromosome rearrangements can be identified by conventional karyotyping, and, with an incidence of 1 in 2000, they are not rare. About 6% of these are associated with MR or other clinical abnormalities, which means that in the European Union, with its 495 million inhabitants, there must be almost 15 000 patients with de novo DBCRs, and even more familial cases. So far, only a small percentage of these patients have been identified, which argues for systematic karyotyping in all patients where a genetic cause of the disorder cannot be ruled out. Unfortunately, however, the ongoing substitution of conventional karyotype analyses with array CGH techniques (see below) means that balanced chromosome rearrangements will no longer be detected upon routine cytogenetic examination.

Mapping of chromosomal breakpoints has been facilitated by the availability of an ordered set of large overlapping genomic clones that serve as probes for fluorescent in situ hybridization (FISH). Still, determining the precise sequence of the breakpoint region remained quite time-consuming. Recently, Chen et al[15] have overcome this problem by preparative sorting of derivative chromosomes followed by next-generation sequencing in three mentally retarded patients with DBCRs, which enabled the identification of three novel candidate genes for MR. In follow-up studies, they showed that it is even possible (by paired-end sequencing) to identify breakpoint-spanning DNA fragments in total genomic DNA, ie, without prior sorting of chromosomes.[16]

Screening for microdeletions and duplications

Small deletions, barely detectable by high-resolution karyotyping, illuminated the way to pinpointing the Duchenne muscular dystrophy gene[17]; later on, microdeletions were instrumental in the identification of many other disease genes. Through the recent introduction of array-based comparative genomic hybridization (array CGH), screening of the entire human genome for submicroscopic copy number variants (CNVs) has become possible, thereby providing a very powerful new strategy for finding the molecular defects underlying Mendelian disorders *(Figure 1b)*. Employing tiling path BAC arrays or, more recently, high-density oligonucleotide arrays, apparently causative de novo microdeletions or duplications can be found in more than 10% of mentally retarded patients,[18] which means that these small variations are about as common as chromosome rearrangements that can be seen under the microscope. Recurrent CNVs that are flanked by low-copy repeats account for about half of the cases (B. de Vries, Nijmegen, personal communication, 2009), and for many of these new "genomic disorders," [19] both deletions and duplications have been observed.

Apart from CNVs causing disease, eg, by disturbing the stoichiometry of protein complexes or by unmasking recessive gene defects,[20] the vast majority of CNVs occur in healthy individuals, and most of them are functionally neutral polymorphisms. Using tiling oligonucleotide microarrays to detect CNVs greater than 450 basepairs, Conrad et al[21] have identified, on average, more than 1000 validated CNVs when comparing genomes of two unrelated individuals.

However, not all CNVs can be assigned unambiguously to one of these two groups. There is a third category of CNVs which are neither functionally neutral nor strictly pathogenic; they are significantly more common in patients with specific disorders than in healthy individuals. One of the first CNVs of this kind observed, a recurrent, sometimes familial 1 to 2 Mb deletion/duplication on chromosome 16p13, was detected in a cohort of 300 patients with autism spectrum disorder and/or MR.[22] Follow-up studies[23] have shown that this CNV, and another on chromosome 15q11.2, are among the most common and important risk factors for MR and autism known to date, both raising the risk for these diseases about 5-fold. Moreover, according to a recent report, the dup16p13.1 is also a significant risk factor for schizophrenia.[13] This CNV encompasses the *NDE1* gene, which interacts with *DISC1*, a known schizophrenia susceptibility gene, and has also been implicated in Asperger syndrome, as discussed elsewhere.[2] Thus, there is no sharp demarcation line separating functionally neutral polymorphisms and clinically relevant CNVs, and distinguishing them is not a trivial task (see below).

Linkage mapping

X-linked disorders are easily recognizable because of their characteristic pattern of inheritance. This is why they are over-represented in OMIM, and why the underlying molecular defect has been elucidated in many

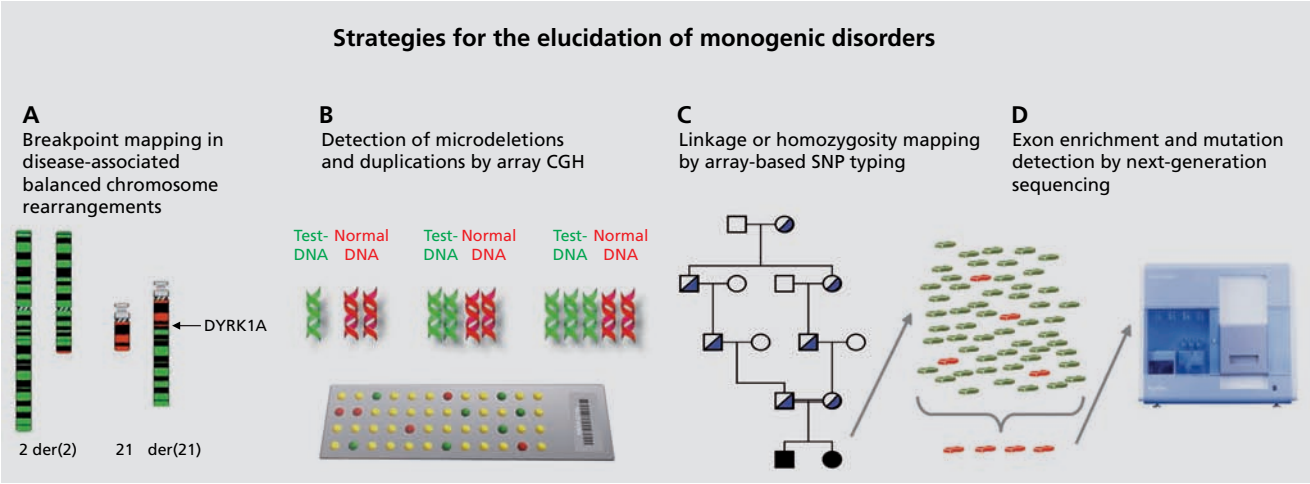


Figure 1. Strategies for the elucidation of monogenic disorders. CGH, comparative genome hybridization; SNP, single-nucleotide polymorphism

instances, as already discussed for X-linked MR. Autosomal dominant disorders also run in families, if they are not lethal in early life, or are so severe that affected individuals do not reproduce. For this reason, they are also easily identifiable, which explains why so many of them are known. In contrast, autosomal recessive disorders are likely to be under-represented, because in Western populations, most patients are isolated cases; the monogenic nature of these disorders is thus not recognized, as discussed above.

Homozygosity mapping in large, consanguineous families is the strategy of choice for mapping recessive disorders *(Figure 1c)*. Such families are common in predominantly Islamic countries of the "consanguinity belt"[24] that extends from Morocco into India. Significantly elevated miscarriage rates and a two-to-threefold higher prevalence of MR and congenital malformations in these countries are generally ascribed to malnutrition and poor standards of hygiene. However, there is evidence that these disorders are also more common in Muslim families living abroad, such as Turkish families in Germany and families from Pakistan in the UK, which suggests that recessive gene defects are another important cause.

Specific forms of autosomal recessive MR (ARMR) that are due to primary microcephaly have been investigated by homozygosity mapping in consanguineous families from Pakistan and India, which led to the identification of 7 loci and 5 microcephaly genes.[25-27] Similarly, large-scale homozygosity mapping in consanguineous Iranian families has revealed numerous novel loci and several new genes for nonsyndromic ARMR, which is thought to be more common than syndromic forms.[28-31] These studies showed that nonsyndromic ARMR is extremely heterogeneous, thereby refuting earlier speculations that, analogous to the fragile X syndrome in X-linked MR and to connexin 26 mutations in nonsyndromic deafness (eg, see ref 32), there might be frequent forms of this disorder. There is recent evidence, however, that ARMR is not quite as heterogeneous as previously suggested. Systematic homozygosity mapping and mutation screening in 250 Iranian families has identified numerous new loci for ARMR and several allelic mutations in the relevant genes (Kuss, Kahrizi, Tzschach, Najmabadi, Ropers et al, unpublished). Analogous studies have also greatly expanded our knowledge of recessive defects in other diseases such as deafness, and there is now evidence that recessive forms also exist in autism and other frequent disorders that are considered to be multifactorial.

Identification of functional candidate genes

Many of the clinically relevant deletions detected by array CGH are larger than 1 to 2 Mb, and most linkage intervals are even larger, often comprising several hundred genes. This renders mutation screening of all genes in these intervals very time-consuming and costly. Numerous software packages have been developed, including PosMed, Endeavour, and Polyphen (see ref 2) that can be employed to identify and prioritize functional candidate genes corresponding to the relevant disease phenotype. The utility of these programs depends on the specificity of the phenotype; not unexpectedly, their performance is still relatively poor for nonsyndromic MR, but much better for easily recognizable syndromes. Undoubtedly, it will improve once more is known about regulatory pathways and the interaction partners of genes and proteins.

As mutation detection techniques are rapidly evolving, sometimes either functional or positional information may suffice for finding specific gene defects. For example, fine-tuning of synaptic transmission is essential for proper brain function, and there are about 1200 proteins that are expressed predominantly in the synapse. Even with conventional Sanger sequencing techniques, screening of all synapse proteins to isolate gene defects responsible for brain dysfunction is no longer an impossible task,[33] and novel technologies are around the corner, which will further facilitate large-scale mutation screening (see below).

Why not search for the mutation directly?

In a recent attempt to identify nearly all genes involved in X-linked MR in one sweep, an international consortium has employed Sanger sequencing to screen 208 families with X-linked MR for mutations in more than 700 fully annotated X-chromosomal genes.[10] This heroic effort has revealed recurrent truncating mutations in 9 novel XLMR genes, and, notably, also almost 1000 missense changes. Some of these are allelic and probably functionally relevant, eg, there are several such mutations in the IQSEC2 gene, which codes for a guanine nucleotide exchange factor.[34] Recent follow-up studies revealed apparently pathogenic CNVs in >10% of the

families,[35] but for more than half of the families studied, the causative molecular defect is still unknown.

This pioneering study has highlighted the possibilities, but also some of the problems, that researchers will face when trying to identify a single pathogenic mutation in an entire genome full of mostly neutral sequence variants. As shown by two independent studies,[36,37] the coding portion of individual genomes contains approximately 10 000 nonsynonymous nucleotide changes, even after excluding those that are known as single-nucleotide polymorphisms (SNPs). These figures should dampen the enthusiasm of those proposing to elucidate unknown monogenic disorders by whole-genome sequencing of single patients and their healthy parents, using exon enrichment and next-generation sequencing techniques *(Figure 1d)*,[3] even though, admittedly, some of the underlying defects may be detectable in this way, depending on the nature of the relevant mutation.

There are now various efficient methods for the enrichment of exons or defined genomic intervals, including custom-made oligonucleotide arrays, commercial enrichment kits based on hybridization in solution, or advanced PCR-based techniques (for details, see the recent review by Tucker et al[38]). Preparative chromosome sorting and next-generation sequencing[39] is another attractive alternative for facilitating mutation detection when the chromosomal location of the defect is known. An advantage of this approach is that it will allow us to detect mutations everywhere on the relevant chromosome, including introns and intergenic sequences. Moreover, sequencing of sorted chromosomes yields a more even coverage than other enrichment strategies that involve PCR amplification (Chen, Wrogemann, Hu, Haas, Ropers et al, unpublished).

Each of these methods has its limitations, however, and the same holds for next-generation sequencing techniques with their usually small read length, which is a problem for (re)sequencing of repeat-rich genome segments. Still, in combination, genome partitioning methods and next-generation sequencing techniques are a great asset for the detection of mutations in defined genomic intervals, which has been one of the stumbling blocks for the large-scale elucidation of single gene disorders.

Conclusions and outlook

With the implementation of these novel methods, the stage is set for the systematic identification of single gene defects, which is overdue and will have far-reaching implications for health care. Recessive disorders likely represent the bulk of the disorders that are hitherto unknown, but they are easily overlooked in industrialized countries because most of the patients will be isolated cases, particularly those without clearly distinguishable phenotypes. Their identification and recruitment is much easier in countries where large families and parental consanguinity are common, but due to more urgent problems, like the scarcity of clean drinking water, malnutrition, or high perinatal and infant mortality, the diagnosis, prevention, and therapy of single gene defects is not high on their agenda, even though these disorders are even more common in these countries than they are in outbred Western populations. This argues for collaborations between emerging and industrialized countries, as exemplified by the long-standing collaboration between our group and an effective Iranian partner, which was instrumental in the elucidating the gene defects responsible for several recessive forms of MR, thereby paving the way for the diagnosis, prevention and—eventually—therapy of these disorders.

So far, recessive disorders are considered too rare to justify carrier screening, but this is likely to change as soon as there is a reliable and inexpensive test for all recessive disorders. According to leading manufacturers, "third-generation" sequencing technologies that enable sequencing of the entire human genome for less than $5000 US will be on the market by the end of 2010 or early in 2011, which indicates that carrier tests for all known recessive disorders will be available sooner rather than later. Indeed, the (US) National Center for Genome Resources has recently teamed up with the Beyond Batten Disease Foundation to develop such a test for approximately 448 single gene defects using available next-generation sequencing technology. With such a carrier test at hand, premarital screening can be offered to rule out the possibility that both spouses are heterozygous for defects in the same gene, and prevention programs can be set up, similar to the successful prevention of Tay-Sachs disease in Ashkenazim, which was initiated in the 1970s.[40]

Whole genome sequencing (WGS) is not only the method of choice for the large-scale elucidation of Mendelian disorders, but it is also a superior alternative for risk factor screening in complex diseases, because it is not fraught with the inherent limitations of GWAS.[2,41] There is no doubt that there exist genetic factors which

predispose individuals to disease without sufficing for disease manifestation, as discussed for CNVs that are risk factors for MR, autism, and schizophrenia. Another telling example is a deletion on chromosome 1q that seems to be a necessary but not sufficient prerequisite for thrombocytopenia/absent radius syndrome.[42] CNVs predisposing for disease can only be identified efficiently by large case-control studies; attempts to find them by investigating the normal variation, ie, by excluding all CNVs present in healthy individuals, are bound to fail because risk factors for common disorders will be found in the healthy controls, too. From the health care point of view it is unfortunate, therefore, that large sums were invested to generate inventories of normal CNVs, instead of focusing on disease-relevant CNVs right from the start—and the same criticism applies to the even more costly "1000 genome project," which uses GWS to study the normal genome variation in 1000 healthy individuals.

It is a commonly held view that mild forms of MR are multifactorial, while severe forms are largely due to catastrophic genetic defects, including chromosomal aberrations and mutations of single genes. Lehrke[6,7] assumed that MR genes and genes determining the IQ were identical, and others speculated that risk factors for mild MR might be allelic variants of these genes,[43,44] exerting a moderate effect on the IQ. As the number of MR genes is increasing, and in view of the novel methods for high-throughput mutation detection, everything seems to be in place for putting these ideas to the test. ❑

Acknowledgements: The author would like to thank Sarah Shoichet, Vera Kalscheuer, Andreas Tzschach, and Reinhard Ullmann for critically reading the manuscript, and Gabriele Eder for secretarial assistance.

REFERENCES

1. Maher B. Personal genomes: the case of the missing heritability. *Nature*. 2008;456:18-21.
2. Ropers HH. New perspectives for the elucidation of genetic disorders. *Am J Hum Genet*. 2007;81:199-207.
3. Check Hayden E. Genomics shifts focus to rare diseases. *Nature*. 2009;461:458.
4. Penrose LS, ed. *A Clinical and Genetic Study of 1280 Cases of Mental Defect*. London, UK: HM Stationery Office; 1938.
5. Renpenning H, Gerrard JW, Zaleski WA, Tabata T. Familial sex-linked mental retardation. *Can Med Assoc J.* 1962;87:954-956.
6. Lehrke R. Theory of X-linkage of major intellectual traits. *Am J Ment Defic*. 1972;76:611-619.
7. Lehrke RG. X-linked mental retardation and verbal disability. *Birth Defects Orig Artic Ser*. 1974;10:1-100.
8. Ropers HH, Hamel BC. X-linked mental retardation. *Nat Rev Genet*. 2005;6:46-57.
9. de Brouwer AP, Yntema HG, Kleefstra T, et al. Mutation frequencies of X-linked mental retardation genes in families from the EuroMRX consortium. *Hum Mutat*. 2007;28:207-208.
10. Tarpey PS, Smith R, Pleasance E, et al. A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. *Nat Genet*. 2009;41:535-543.
11. Zechner U, Wilda M, Kehrer-Sawatzki H, Vogel W, Fundele R, Hameister H. A high density of X-linked genes for general cognitive ability: a run-away process shaping human evolution? *Trends Genet*. 2001;17:697-701.
12. Morrow EM, Yoo SY, Flavell SW, et al. Identifying autism loci and genes by tracing recent shared ancestry. *Science*. 2008;321:218-223.
13. Ingason A, Rujescu D, Cichon S, et al. Copy number variations of chromosome 16p13.1 region associated with schizophrenia. *Mol Psychiatry*. 2009 Sep 29 [Epub ahead of print].
14. Passarge E. Wither polygenic inheritance: mapping Hirschsprung disease. *Nat Genet*. 1993;4:325-326.
15. Chen W, Kalscheuer V, Tzschach A, et al. Mapping translocation breakpoints by next-generation sequencing. *Genome Res*. 2008;18:1143-1149.
16. Chen W, Ullmann R, Langnick C, et al. Breakpoint analysis of balanced chromosome rearrangements by next-generation paired-end sequencing. *Eur J Hum Genet*. In press.
17. Monaco AP, Bertelson CJ, Middlesworth W, et al. Detection of deletions spanning the Duchenne muscular dystrophy locus using a tightly linked DNA segment. *Nature*. 1985;316:842-845.
18. Koolen DA, Pfundt R, de Leeuw N, et al. Genomic microarrays in mental retardation: a practical workflow for diagnostic applications. *Hum Mutat*. 2009;30:283-292.
19. Lupski JR. Genomic disorders ten years on. *Genome Med*. 2009;1:42.
20. Vissers LE, van Ravenswaaij CM, Admiraal R, et al. Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nat Genet*. 2004;36:955-957.
21. Conrad DF, Pinto D, Redon R, et al. Origins and functional impact of copy number variation in the human genome. *Nature*. 2009 Oct 7 [Epub ahead of print].
22. Ullmann R, Turner G, Kirchhoff M, et al. Array CGH identifies reciprocal 16p13.1 duplications and deletions that predispose to autism and/or mental retardation. *Hum Mutat*. 2007;28:674-682.
23. Mefford HC, Cooper GM, Zerr T, et al. A method for rapid, targeted CNV genotyping identifies rare variants associated with neurocognitive disease. *Genome Res*. 2009;19:1579-1585.
24. Bittles AH. A community genetics perspective on consanguineous marriage. *Community Genet*. 2008;11:324-330.
25. Bond J, Roberts E, Springell K, et al. A centrosomal mechanism involving CDK5RAP2 and CENPJ controls brain size. *Nat Genet*. 2005;37:353-355.
26. Cox J, Jackson AP, Bond J, Woods CG. What primary microcephaly can tell us about brain growth. *Trends Mol Med*. 2006;12:358-366.
27. Kumar A, Girimaji SC, Duvvari MR, Blanton SH. Mutations in STIL, encoding a pericentriolar and centrosomal protein, cause primary microcephaly. *Am J Hum Genet*. 2009;84:286-290.
28. Garshasbi M, Hadavi V, Habibi H, et al. A defect in the TUSC3 gene is associated with autosomal recessive mental retardation. *Am J Hum Genet*. 2008;82:1158-1164.
29. Garshasbi M, Motazacker MM, Kahrizi K, et al. SNP array-based homozygosity mapping reveals MCPH1 deletion in family with autosomal recessive mental retardation and mild microcephaly. *Hum Genet*. 2006;118:708-715.
30. Moheb LA, Tzschach A, Garshasbi M, et al. Identification of a nonsense mutation in the very low-density lipoprotein receptor gene (VLDLR) in an Iranian family with dysequilibrium syndrome. *Eur J Hum Genet*. 2008;16:270-273.
31. Motazacker MM, Rost BR, Hucho T, et al. A defect in the ionotropic glutamate receptor 6 gene (GRIK2) is associated with autosomal recessive mental retardation. *Am J Hum Genet*. 2007;81:792-798.
32. Rabionet R, Zelante L, Lopez-Bigas N, et al. Molecular basis of childhood deafness resulting from mutations in the GJB2 (connexin 26) gene. *Hum Genet*. 2000;106:40-44.

Los trastornos por un gen único vuelven a estar en el foco de la investigación

A comienzos de los años 1990, cuando se escribió la segunda fase de cinco años del Proyecto del Genoma Humano –el cual ha requerido más dinero que cualquier otro proyecto de investigación en biología- los trastornos comunes fueron presentados como los blancos futuros de la investigación del genoma. Esta fue una medida ingeniosa para asegurar un financiamiento público continuo para este esfuerzo, el cual se había justificado previamente por las perspectivas que conducirían al diagnóstico, la prevención y el tratamiento de los trastornos Mendelianos, que si bien son graves ocurren con escasa frecuencia. Hoy día, después de más de quince años y luego de haber gastado billones de dólares en los estudios de asociación del genoma completo (EAGC), se han identificado muy pocos factores de riesgo genético importantes para las enfermedades comunes, y el entusiasmo por grandes EAGC está disminuyendo. Al mismo tiempo, hay un renovado interés en el estudio de trastornos por un gen único, los cuales son considerados ahora por algunos investigadores como una mejor pista para la comprensión de las enfermedades comunes. Aunque esto es probablemente cierto, los trastornos Mendelianos también son importantes por derecho propio, ya que ellos deben ser mucho más comunes de lo que generalmente se piensa. Como se discute aquí, existen varias estrategias eficientes para aclarar los defectos de un gen único y su aplicación sistemática en combinación con nuevas técnicas de división del genoma y de secuenciación paralela masiva, tendrán efectos de gran alcance para los cuidados en salud.

Les maladies monogéniques attirent à nouveau l'attention

*Au début des années 90, lorsque le second plan quinquennal du projet du génome humain (*Human Genome Project*) (plus dispendieux que tout autre projet de recherche précédent en biologie) a été écrit, les maladies courantes furent présentées comme la future cible de la recherche sur le génome. C'était une manœuvre intelligente destinée à s'assurer d'un soutien publique prolongé pour cette tentative, préalablement justifié par la perspective du diagnostic, de la prévention et du traitement des maladies mendéliennes sévères mais rares pour la plupart. Aujourd'hui, après plus de 15 ans et des milliards de dollars dépensés pour des études d'association sur le génome entier (GWAS pour* Genome-Wide Association Studies*), très peu de facteurs majeurs de risque génétique pour les maladies courantes ont été identifiés et l'enthousiasme pour les grandes études d'association faiblit. Au même moment, il existe un regain d'intérêt pour l'étude des maladies monogéniques, considérées maintenant par certains comme une meilleure piste pour la compréhension des maladies courantes. C'est probablement le cas, d'autant que les maladies mendéliennes sont elles-mêmes importantes puisque beaucoup plus fréquentes qu'on ne le pense généralement. Dans cet article, nous examinons les différentes stratégies efficaces pour comprendre les anomalies monogéniques et leur application systématique en association aux nouvelles techniques de partition du génome et de séquençage parallèle massif. Ces stratégies auront des implications considérables pour la Santé.*

33. Laumonnier F, Cuthbert PC, Grant SG. The role of neuronal complexes in human X-linked brain diseases. *Am J Hum Genet*. 2007;80:205-220.
34. Shoubridge C, Tarpey P, Abidi F, et al. Mutations in IQSEC2, a guanine nucleotide exchange factor for ARF6, cause non-syndromic mental retardation. Paper presented at: 14th International Workshop on Fragile X and X-linked Mental Retardation. Praia do Forte, Brasil, September 15-19, 2009.
35. Raymond FL, Tarpey P, Whibley A, et al, on behalf of the iGOLD study. A systematic, large-scale resequencing and copy number variant screen of the X Chromosome coding exons in mental retardation. Paper presented at: 14th International Workshop on Fragile X and X-linked Mental Retardation. Praia do Forte, Brasil, September 15-19, 2009.
36. Ng PC, Levy S, Huang J, et al. Genetic variation in an individual human exome. *PLoS Genet*. 2008;4:e1000160.
37. Wheeler DA, Srinivasan M, Egholm M, et al. The complete genome of an individual by massively parallel DNA sequencing. *Nature*. 2008;452:872-876.
38. Tucker T, Marra M, Friedman JM. Massively parallel sequencing: the next big thing in genetic medicine. *Am J Hum Genet*. 2009;85:142-154.
39. Bentley DR, Balasubramanian S, Swerdlow HP, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008;456:53-59.
40. Kaback M, Lim-Steele J, Dabholkar D, Brown D, Levy N, Zeiger K. Tay-Sachs disease--carrier screening, prenatal diagnosis, and the molecular era. An international perspective, 1970 to 1993. The International TSD Data Collection Network. *JAMA*. 1993;270:2307-2315.
41. Terwilliger JD, Hiekkalinna T. An utter refutation of the "fundamental theorem of the HapMap". *Eur J Hum Genet*. 2006;14:426-437.
42. Klopocki E, Schulze H, Strauss G, et al. Complex inheritance pattern resembling autosomal recessive inheritance involving a microdeletion in thrombocytopenia-absent radius syndrome. *Am J Hum Genet*. 2007;80:232-240.
43. Mandel JL, Chelly J. Monogenic X-linked mental retardation: is it as frequent as currently estimated? The paradox of the ARX (Aristaless X) mutations. *Eur J Hum Genet*. 2004;12:689-693.
44. Ropers HH. X-linked mental retardation: many genes for a complex disorder. *Curr Opin Genet Dev*. 2006;16:260-269.

The genetic epidemiology of personality disorders

Ted Reichborn-Kjennerud, MD, PhD

Genetic epidemiologic studies indicate that all ten personality disorders (PDs) classified on the DSM-IV axis II are modestly to moderately heritable. Shared environmental and nonadditive genetic factors are of minor or no importance. No sex differences have been identified. Multivariate studies suggest that the extensive comorbidity between the PDs can be explained by three common genetic and environmental risk factors. The genetic factors do not reflect the DSM-IV cluster structure, but rather: i) broad vulnerability to PD pathology or negative emotionality; ii) high impulsivity/low agreeableness; and iii) introversion. Common genetic and environmental liability factors contribute to comorbidity between pairs or clusters of axis I and axis II disorders. Molecular genetic studies of PDs, mostly candidate gene association studies, indicate that genes linked to neurotransmitter pathways, especially in the serotonergic and dopaminergic systems, are involved. Future studies, using newer methods like genome-wide association, might take advantage of the use of endophenotypes.

 Dialogues Clin Neurosci. 2010;12:103-114.

Keywords: *personality disorder; axis I disorder; genetics; twin study; molecular genetic study; candidate gene*

The introduction of personality disorders (PDs) as diagnostic categories on a separate axis (Axis II) in the third edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III) in 1980[1] had a dramatic effect on the level of interest in these disorders among researchers, and the number of published articles increased substantially. However, the number of genetic epidemiologic studies of the DSM PDs has remained limited compared with studies on both clinical disorders like schizophrenia, depression, and anxiety disorders (which are classified on Axis I in DSM), and on normal personality traits.[2-4]

The understanding of the role of genetic factors in the etiology of disorders and traits is inseparably linked to classification, since a precise definition of the phenotype is a prerequisite for all successful genetic studies. In this review we will focus on PDs as they are classified in the DSM; a system that serves many purposes, and is not specifically designed for genetic studies. This is a problem not only for the genetics of PDs, and the search for better phenotypes for genetic studies of mental disorders is especially well illustrated in the literature on schizophrenia (eg, refs 5, 6).

The goal of psychiatric genetic epidemiology is to understand the role of genetic and environmental factors in the etiology of mental disorders.[7] In this paper we will focus mainly on the genetic factors. After a brief outline of the current DSM axis II PD classification, we will

Author affiliations: Division of Mental Health, Norwegian Institute of Public Health Institute of Psychiatry, University of Oslo, Norway; Department of Epidemiology, Columbia University, New York, NY, USA

Address for correspondence: Ted Reichborn-Kjennerud, Director, Department of Adult Mental Health, Norwegian Institute of Public Health, Box 4404 Nydalen, N-0403 Oslo, Norway
(e-mail: ted.reichborn-kjennerud@fhi.no)

evaluate the evidence for genetic influences on PDs and examine quantitative genetic studies that explore the specificity of the genetic effects, ie, to what extent genetic risk factors are shared between PDs, or between PDs and axis I disorders. Molecular genetic studies that aim to identify gene variants associated with PDs will then be reviewed. It is likely that PDs, like most other psychiatric disorders, are etiologically complex, ie, that they are influenced by a number of genetic and environmental risk factors. Studies examining the interplay between genes and the environment will be addressed both in relation to quantitative and molecular methods. Finally, future directions will be discussed.

The classification of personality disorders

A PD is defined by DSM-IV as an enduring pattern of inner experience and behavior that deviates markedly from the expectations of the individual's culture, is pervasive and inflexible, has an onset in adolescence or early adulthood, is stable over time, and leads to distress or impairment.[8] The DSM-IV classification includes 10 categorical PD diagnoses grouped into three clusters: A or the "odd-eccentric," B or the "dramatic-emotional," and C or the "anxious-fearful."[8] Cluster A includes paranoid, schizoid, and schizotypal PD, and Cluster B antisocial, borderline, histrionic, and narcissistic PD, while cluster C includes avoidant, dependent, and obsessive-compulsive PD. Appendix B includes two additional disorders: depressive and passive-aggressive PDs.

Although the classification of PDs in DSM-IV is more empirically based than in former versions, there are several controversial issues that are unresolved. Substantial co-occurrence between the DSM PDs has consistently been found in both clinical[9] and community samples.[10,11] The majority of individuals with a PD receive more than one PD diagnosis, and this high degree of overlap seriously challenges the descriptive validity of the PD classification. Comorbidity with Axis I disorders is also extensive, and results from both clinical and population-based studies indicate that the key features in the DSM-IV definition (stability over time and early age of onset) do not distinguish PDs from axis I disorders.[12] The underlying validity of the DSM axis I - axis II division has therefore been questioned (eg, refs 12-14). The higher order clustering system has serious limitations, and has not been consistently validated,[8] and factor analytic studies often do not find support for this three-factor structure.[15] One of the most controversial and long-standing issues in the field of PD classification is, however, whether PDs should be conceptualized dimensionally or as discrete categories. There seems to be a general agreement that PDs are best classified dimensionally,[16-18] and several alternative systems are discussed for DSM-V (see ref 19).

Basic quantitative studies

In quantitative genetics, which include family, twin, and adoption studies, the degree to which individual liability to a disorder results from familial effects (in family studies) or genetic and environmental factors (in twin and adoption studies) is estimated. Twin studies have been most commonly used to examine the effects of genetic risk factors on mental disorders, including PDs, and sophisticated analytical models and statistical tools have been developed.[20,21] The proportion of phenotypic differences between individuals (or proportion of variance) in a particular population that can be attributed to genetic differences is called *heritability*. In the classical twin model the total variance in a phenotype is partitioned into three variance components, each accounted for by three latent variables: additive genetic, shared environment, and individual-specific environment. This implies that the genetic and environmental effects are not directly measured, ie, we do not know which specific genes or environmental factors influencing the phenotype. Genetic effects are usually additive, meaning that the independent effects of different alleles or loci act in an additive way to increase risk for the disorder or trait, but they can also be nonadditive, which means that different alleles or loci interact with other alleles or loci (epistasis) or different alleles in the same locus (dominance). Shared environment includes all environmental exposures that contribute to making twins similar, and individual-specific or unique environment includes all environmental exposures that make them different, plus measurement error.

Modern twin studies are based on the liability-threshold model,[22] which assumes that a large number of genetic and environmental risk factors with small individual effects are involved, resulting in a distribution of liability or risk in the population that approximates normality. A dichotomous disorder will appear when a certain threshold is exceeded. Twin studies can be used regardless of whether PDs are defined categorically or dimen-

sionally, but the statistical power is higher if the phenotype is ordinal or continuous.[23]

Normal and abnormal personality traits

Normal personality traits have repeatedly been shown to be influenced by genetic factors with heritability estimates ranging from approximately 30% to 60%.[24,25] The genetic effects are mainly additive, but nonadditive contributions of a smaller magnitude have been identified in studies with sufficient statistical power.[24] Shared environmental factors are usually found to be of minor on no importance.[24] Similar heritability estimates have been found for a dimensional classification of personality disorders based on self-report.[26] Numerous studies have shown relatively high correlations between DSM PDs and normal personality traits of the five-factor model, which includes five broad bipolar domains of extraversion (vs introversion), agreeableness (vs antagonism) conscientiousness (vs impulsivity), neuroticism (vs emotional stability), and openness (vs closedness to experience),[27] but the extent to which this is due to genetic factors is not known.

DSM personality disorders

Cluster A

Prior studies have suggested that familial/genetic factors contribute to the etiology of the three PDs making up the DSM Cluster A.[28] A series of twin studies that examine various measures of schizoid, schizotypal, and paranoid-like traits using self-report questionnaires have nearly uniformly found significant heritability for these traits and failed to find shared environmental effects (eg, refs 29-33). Heritabilities are typically in the range of 35% to 60%. In a twin study using structured interview data, but based on a clinical sample, Torgersen et al[34] found lower heritability estimates for paranoid PD (28%) and schizoid PD (29%), but much higher heritability for schizotypal PD (61%). The method of ascertainment and the relatively low number of participants make the estimates from this study uncertain. In a more recent population-based study of dimensional representations of the DSM-IV cluster A PDs based on structured interviews, Kendler et al[35] estimated heritability to be 21% for paranoid, 28% for schizotypal, and 26% for schizoid PD. No shared environmental effects or sex differences were found.

In twin studies unreliability of measurement will decrease the heritability estimates. Although the inter-rater reliability in Kendler et al's abovementioned study was excellent, the test-retest reliability or stability of measurement for PDs has been shown to be imperfect.[36] It is also likely that genetic and environmental risk factors assessed by self-report questionnaires vs interviews are different. A second study from the same sample was therefore undertaken.[37] Data from a previous self-report questionnaire study were used in addition to the abovementioned interview data to account for unreliability of measurement by using two measures differing in both time and mode of assessment. The estimated heritabilities were substantially higher than in the first study: 66% for paranoid, 55% to 59% for schizoid, and 72% for schizotypal PD.

Cluster B

Antisocial PD-like measures have been extensively studied using genetic epidemiological methods. In a meta-analysis of 51 twin and adoption studies on antisocial behavior based largely on records, self-report, and family report, Rhee & Waldman[38] found that the variance could most parsimoniously be explained by additive genetic factors (32%), nonadditive genetic factors (9%), shared environmental factors (16%) and individual-specific environmental factors (43%). There were no significant differences in the magnitude of genetic and environmental influences for males and females.

In a review of family studies on borderline PD, White et al[39] found the disorder to aggregate in families. However, significant methodological problems made the results uncertain. Distel et al estimated that additive genetic factors explained 42% of the variance in borderline PD features assessed by self-report questionnaire, using data from three countries.[40] Non-shared environment accounted for the rest. In a subsequent extended twin-family study by the same group the heritability of borderline PD features was found to be 45%, but the genetic effects were both additive (21%) and dominant (24%).[41] Nonadditive effects are difficult to detect using the classical twin model due to lack of statistical power.[23] However, such effects have been found for normal personality traits in twin-sibling studies with large samples.[42]

Results from a twin study based on structured interviews in a clinical sample suggest that heritability estimates for borderline, histrionic, and narcissistic PD were high, 69%, 63%, and 77% respectively.[34] More recently, however, Torgersen et al[43] conducted a population-based twin study

of dimensional representations of the DSM-IV cluster B PDs. Heritability was estimated to be 38% for antisocial PD, 31% for histrionic PD, 24% for narcissistic PD and 35% for borderline PD. No shared environmental influences or sex or effects were found.

Cluster C

A family study of the anxious-fearful cluster indicated significant familiality for DSM-III avoidant and dependent PD,[44] and in a clinically based twin study, heritability estimates for avoidant, dependent, and obsessive-compulsive PD were found to be 28%, 57%, and 77%, respectively.[34] Results from a population-based study of dimensional representations of DSM-IV Cluster C PDs,[45] however, indicated that heritability estimates were similar for avoidant PD (35%), but lower for dependent (31%) and for obsessive-compulsive PD (27%), again illustrating the importance of method of ascertainment. This discrepancy is probably in part due to difference in methods of ascertainment. No shared environmental effects or sex differences have been found for cluster C PDs.

Disorders in Appendix B

In a population-based twin study of depressive PD, Ørstavik et al[46] found that liability could best be explained by additive genetic and unique environmental factors alone, with heritability estimates of 49% in females and 25% in males. Unlike the results for the other DSM-IV PDs, both quantitative and qualitative sex-differences were found corresponding to findings from studies on major depression.[47] Significant familial aggregation has also been found for DSM-IV passive aggressive PD.[48]

Multivariate studies

If heritability has been established, several more complex models can be employed to explore the nature and mode of action of the genetic risk factors.[7] Multivariate analyses, which comprise models where several phenotypes are included and different structures of the latent factors can be specified,[20] can be used to estimate to what extent genetic and environmental risk factors are

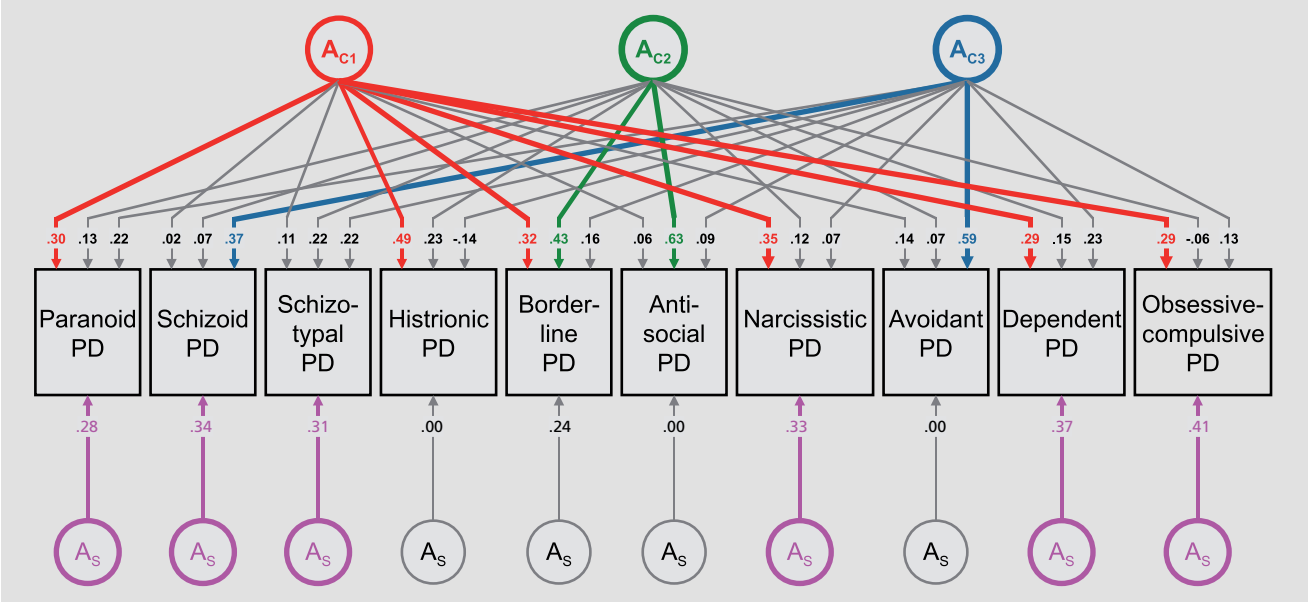


Figure 1. Genetic parameter estimates from best fitting model for ten DSM-IV personality disorders. Path estimates are standardized regression coefficients, so they must be squared to equal the proportion of variance accounted for in the dependent variable. A stands for additive genetic effects. The subscripts C and S stand, respectively, for common factor and disorder-specific effects. The first, second and third genetic common factors are indicated by the subscripts $_{C1}$, $_{C2}$ and $_{C3}$. Paths with values ≥ +0.28 (which account for ≥ 8% of phenotypic variance) are colored with the first, second, and third common factor indicated by, respectively, red, green, and blue and the disorder-specific factors by magenta. Paths not exceeding the + 0.28 cutoff are depicted in gray.

From ref 52: Kendler KS, Aggen SH, Czajkowski N, et al. The structure of genetic and environmental risk factors for DSM-IV personality disorders a multivariate twin study. *Arch Gen Psychiatry*. 2008;65:1438-1446. Copyright © American Medical Association 2008

specific to a given PD or shared in common with other PDs or axis I disorders, and thus to investigate sources of comorbidity.[49,50] By including measures of the same phenotypes on different points in time, they can also be used to determine if genetic effects differ over time in a developmental perspective.

DSM-IV personality disorders

Cluster A PDs have been found to aggregate in families of probands with schizophrenia (see below). Familial coaggregation has also been found for borderline PD and antisocial PD[39] and for borderline PD and all the other cluster B PDs,[51] as well as for the DSM-III cluster C PDs.[44] A population-based twin study including all PDs within cluster B indicated that borderline PD and antisocial PD appeared to share genetic risk factors above and beyond those shared in common with the other cluster B disorders,[43] and a twin study of cluster C PDs suggested that genetic factors influencing obsessive-compulsive PD appeared to be relative specific to this disorder.[45] Kendler et al, in the only population-based multivariate twin study including all 10 DSM-IV PDs that has been published,[52] found that the best-fitting model included three genetic and three environmental factors in addition to disorder-specific factors. The structure of the genetic factors is shown in *Figure 1*. The first genetic factor (A_{C1}) had high loadings on PDs from all 3 clusters including paranoid, histrionic, borderline, narcissistic, dependent, and obsessive-compulsive PD. This factor probably reflects a broad vulnerability to PD pathology and/or negative emotionality, and is related to genetic liability to the normal personality trait neuroticism. The second genetic factor (A_{C2})was quite specific with substantial loadings only on borderline and antisocial PD. This is consistent with the results from the above-mentioned family studies,[39] and suggests genetic liability to a broad phenotype for impulsive/aggressive behavior. The third factor identified (A_{C3})had high loadings only on schizoid and avoidant PD. This can be interpreted in several ways. It might in part reflect genetic risk for schizophrenia spectrum pathology (see below). From the perspective of the five-factor model of normal personality it reflects genetic liability for introversion.[53] Finally, it is noteworthy that obsessive-compulsive PD had the highest disorder-specific genetic loading, which parallels prior findings that this PD shares little genetic and environmental liability with the other cluster C PDs.

The results are also to a large extent consistent with a prior multivariate twin study of the dimensional classification system of personality disorder trait mentioned above[26] in which Livesley et al identified four genetic factors loading on four phenotypic dimensions called "emotional dysregulation," "dissocial behavior," "inhibition," and "compulsivity."

Taken together these results indicate that genetic risk factors for DSM-IV PDs do not reflect the cluster A, B, and C typology. However, this is well reflected in the structure of the environmental risk factors, suggesting that the comorbidity of PDs within clusters is due to environmental experiences.

Personality disorders and Axis I disorders

Several lines of evidence indicate specific axis I/axis II relationships,[54,55] suggesting that common genetic or environmental liability factors might predispose to several disorders within clusters that transcend the axis I/axis II division.[13,49,56]

Schizophrenia

A number of family and adoption studies have examined the risk for paranoid, schizoid, and schizotypal PDs in relatives of schizophrenic and control probands. While a few studies can be found where all three cluster A PDs are at increased risk in relatives of schizophrenic probands,[57,58] more common are studies that find that only schizotypal PD[59-63] or schizotypal PD and paranoid PD[64] have a significant familial relationship with schizophrenia. Taken together, these results suggest that schizotypal PD has the closest familial relationship to schizophrenia, followed by paranoid and schizoid PD, and are consistent with the hypothesis that a common genetic risk factor for cluster A PDs reflects—in the general population—the liability to schizophrenia.[35] The extended phenotype believed to reflect this genetic liability to schizophrenia is often described by the term schizophrenia spectrum. Schizotypal PD has been suggested to be the prototypical disorder in this spectrum.[65] In a recent family study, Fogelson et al[66] showed that avoidant PD, currently classified in DSM cluster C, also occurred more frequently in relatives of probands with schizophrenia even after controlling for schizotypal and paranoid PD. This replicates findings from earlier studies,[58,67] and suggest that avoidant PD should also be

included in this spectrum. It is also in part in accordance with the results from the multivariate study by Kendler et al described above,[52] where avoidant and schizoid PD share genetic liability.

Internalizing disorders

Mood and anxiety disorders (often called internalizing disorders) share genetic and environmental liability factors with each other,[68] and with the normal personality trait neuroticism,[69] which correlates strongly with several PDs, especially in cluster B and C.[53]
Family studies indicate that borderline PD and major depression share familial risk factors.[51,70] In a population-based multivariate twin study of major depression and DSM-IV PDs, Reichborn-Kjennerud et al[71] found that dimensional representations of borderline PD from cluster B, avoidant PD from cluster C, and paranoid PD from cluster A were all independently and significantly associated with increased risk for major depression. Multivariate twin modeling indicated that one latent factor accounted for the genetic covariance between major depression and the three PDs. The genetic correlations between major depression and borderline, avoidant, and paranoid PD were respectively +0.56, +0.22, and +0.40. No sex differences or shared environmental effects were found. These results indicate that vulnerability to general PD pathology and major depression are closely related. In a bivariate twin study, Ørstavik et al[72] found that a substantial part of the covariation between major depressive disorder and depressive PD was accounted for by genetic factors with a genetic correlation of 0.56. Results from another population-based twin study, investigating the sources of co-occurrence between social phobia and of avoidant PD in females, indicated that social phobia and avoidant PD were influenced by identical genetic factors, whereas the environmental factors influencing the two disorders were uncorrelated.[73] This suggests that an individual with high genetic liability will develop avoidant PD versus social phobia entirely as a result of environmental risk factors unique to each disorder, which is in accordance with the hypothesis of underlying psychobiological dimensions cutting across the axis I/ axis II classification system.

Substance-use disorders

Numerous family, adoption and twin studies have demonstrated that antisocial PD, conduct disorder, and substance-use disorders (often called externalizing disorders) share a common genetic liability (eg, refs 68,74). In a family-twin study, Hicks et al[75] found that a highly heritable (80%) general vulnerability to all the externalizing disorders accounted for most of the familial resemblance. Disorder-specific vulnerabilities were detected for conduct disorder, alcohol dependence, and drug dependence, but not for antisocial PD. The same group also reported an association between externalizing disorders and reduced amplitude of the P3 component of the brain event-related potential, suggesting that this could be a common biological marker for the biological vulnerability to these disorders.[76]

Longitudinal studies

Most of the genetic studies that have investigated changes in genetic influences on PDs over time have used measures related to antisocial PD. The following examples illustrate the potential of longitudinal quantitative genetic methods. In a twin study, Lyons et al[77] demonstrated that the genetic influence on symptoms of DSM-III-R antisocial PD was much more prominent in adulthood than in adolescence. Silberg et al[78] studying twins between 10 and 17 years of age found a single genetic factor that influenced antisocial behavior beginning at age 10 through young adulthood, a shared environmental effect beginning in adolescence, a transient genetic effect at puberty and genetic influences specific to adult antisocial behavior. In another recent twin study of externalizing disorders, biometric analyses revealed increasing genetic variation and heritability for men but a trend toward decreasing genetic variation and increasing environmental effects for women.[79]

Gene-environment interplay

In the traditional models of disease etiology in psychiatric epidemiology the causal pathway is conceptualized as moving from the environment to the organism. However, since genes influence behavior, genetic factors can indirectly influence or control exposure to the environment,[20] called *gene-environment correlation*.[20,80,81] Genetic factors can also control an individual's sensitivity to the environment, ie, genetic factors influence or alter an organism's response to environmental stressors.[20,80,81] This is usually called *gene-environment interaction*. In quantitative studies of gene-environment interplay, genetic factors are

either inferred (eg, disorder in biological parent in adoption studies) or modeled as a latent variable.[80,82]

Twin and adoption studies have provided much of the evidence for gene-environment correlations by demonstrating genetic influences for a number of measures of the environment.[80] Overall, the evidence from twin and adoption studies suggests that gene-environment correlations are mediated by heritable personality traits and possibly PDs.[81,83,84]

The initial indications that gene-environment interaction was likely to be operating came from adoption and twin studies.[85] Gene-environment interaction was demonstrated in an adoption study as early as in 1974, when Crowe[86] found that early institutional care was a risk factor for later antisocial behavior only when a genetic risk factor was present. In another adoption study, Cadoret et al[87] found significant gene-environment interaction by showing that there was a negligible risk for antisocial behavior from a genetic risk alone (antisocial behavior in the biological parent), no effect of an adverse adoptive family environment alone, but a substantial effect when both were present. The finding was replicated in a later study with a larger number of adoptees,[88] Jaffe et al,[89] using a twin design, found significant gene-environment interaction with respect to childhood maltreatment and the development of antisocial behavior, and in a twin study Tuvblad et al[90] demonstrated a significant gene-environment interaction by showing that the heritability for adolescent antisocial behavior is higher in socioeconomic advantaged environments. Using an advanced family design, Feinberg et al[91] recently found an interaction of genotype and both parental negativity and low warmth predicting antisocial behavior. Significant gene-environment interaction has also been demonstrated in schizophrenia spectrum disorders. In an adoption study Tienari et al[92] showed that there was a significant association between disordered rearing and the diagnosis of schizophrenia spectrum disorder in the offspring of mothers with but not in offspring of mothers without the diagnoses. In a community based twin study, Hicks et al demonstrated a significant gene-environment interaction with a number of environmental risk factors showing that greater environmental adversity was associated with increased genetic risk for antisocial PD and substance use disorders.[93] Significant gene-environment interaction has also been demonstrated in quantitative studies of anxiety and mood disorders.[81]

Molecular genetic studies

Traditionally, linkage and association studies have been most commonly used for mapping disease loci.[94] Most of the molecular genetic studies of PDs has been done using hypothesis-driven candidate gene association studies[95] focusing on particular genes related to the neurotransmitter pathways, especially in the serotonergic and dopaminergic systems. Although the number of genetic association studies are increasing exponentially, only a very small fraction of positive results are replicated.[96,97] Until further replications are published the results reviewed below must therefore be considered tentative.

Cluster A

Consistent with the hypothesis that schizophrenia and related PDs are linked to dopaminergic dysfunction, Rosmond et al[98] found that Cluster A PDs were associated with a polymorphism in the gene coding for the dopamine 2 receptor (DRD2). Building on results from quantitative genetic studies indicating that common genetic risk factors exist for schizotypal PD and schizophrenia, Stefanis et al[99] examined the potential impact of SNPs within the four most prominent candidate genes for schizophrenia. Dysbindin (DTNBP1) and D-amino-acid oxidase (DAAO) both showed associations with symptoms of schizotypy. Similarly, Fanous et al[100] using a linkage approach, found that a subset of schizophrenia susceptibility genes also affect schizotypy in nonpsychotic relatives. Significant associations with schizotypal personality traits have also been found in several studies with polymorphisms in the gene coding for catechol-O-methyltransferase (COMT)[100,102,103] an enzyme involved in the degradation of catecholamines, and linked to the etiology of schizophrenia.[104]

Cluster B

Multiple lines of evidence suggest that dysfunction in the serotonin (5-HT) system is associated with impulsivity, aggression, affective lability, and suicide. Genes linked to the function of this neurotransmitter can therefore be considered possible candidate genes for borderline and antisocial PD. Kennedy and coworkers found that borderline PD was associated with polymorphisms in the serotonin transporter gene (*5-HTTLPR*),[105] and polymorphisms in the gene coding for the catabolic enzyme

monoamine oxidase A (MAOA), involved in the regulation of biogenic amines like serotonin, norepinephrine, and dopamine,[106] but not polymorphisms in the gene coding for the serotonin 5-HT2A receptor.[107] Recently the group has conducted a gene-gene interaction study with a number of polymorphisms in seven serotonin genes (including the three mentioned above), concluding that "serotonin genes and their interaction may play a role in the susceptibility to borderline PD."[108] Other groups have reported similar findings. A main effect of the *5-HTTLPR* polymorphism on borderline PD has been found in bulimic women,[109] and Lyons-Ruth et al found a significant relationship between the short *5HTTLPR* allele and both borderline and antisocial PD,[110] but other studies have failed to find an association between this polymorphism and cluster B PDs.[111] Polymorphisms in the MAOA gene have been found to be associated with cluster B PDs,[112] and antisocial traits.[113] Tryptophan hydroxylase is the rate-limiting enzyme in the serotonin metabolic pathway. Two genes related to this enzyme, the tryptophan hydroxylase 1 and 2 genes (TPH1and TPH2), have been associated with borderline PD[114] and personality traits related to emotional instability, as well as to cluster B and cluster C PDs.[115] Taken together, these findings suggest that borderline and antisocial PD and possibly also the other cluster B PDs, are influenced by genes regulating the serotonergic system. They are also consistent with the finding of shared genetic influence on borderline PD and antisocial PD, and on borderline PD and the other cluster B PDs found in multivariate twin studies.[43,52]

Cluster C

It has previously been suggested that the 5-HTTLPR polymorphism was associated with anxiety-related traits,[116] but later studies have yielded conflicting results (see ref 117). Patients diagnosed with cluster C PDs, have not been found to be significantly higher in the frequency of the short form allele of the 5-HTTLPR.[111]Recent results, on the other hand, indicate that variations in the COMT gene contribute to genetic risk shared across a range of anxiety-related phenotypes.[118,119] Joyce[120] found an association between avoidant and obsessive-compulsive PD symptoms and the dopamine D3 receptor (DRD3) polymorphism. In a later study and a meta-analysis, the finding for obsessive-compulsive symptoms were replicated, leading the authors to conclude that DRD3 may contribute to the development of obsessive-compulsive PD.[121]

Gene-environment interplay

Few studies of gene-environment correlation using measured genes and measured environments have been published. Dick et al[121] found that individuals who had a polymorphism in a gene (GABRA2) associated with alcohol dependence were less likely to be married, in part because they were at higher risk for antisocial PD and were less likely to be motivated by a desire to please others. Other results confirm the existence of gene-environment correlation with measured genes in both the dopaminergic and serotonergic system, and provide preliminary support for the finding that correlations are mediated by behavioral and personality characteristics.[84]

Gene-environment interaction studies using identified susceptibility genes rather than unmeasured latent genetic factors can provide more secure estimates.[84] Based on results from quantitative genetic studies showing gene-environment interaction for antisocial behavior, Caspi et al[123] studied the association between childhood maltreatment, and a functional polymorphism in the promoter region of the MAOA gene on antisocial behavior assessed through a range of categorical and dimensional measures using questionnaire and interview data plus official records. The results showed no main effect of the gene, a main effect for maltreatment and a substantial and significant interaction between the gene and adversity. The maltreated children whose genotype conferred low levels of MAOA expression more often developed conduct disorder and antisocial personality than children with a high activity MAOA genotype. Foley et al[124] replicated this finding and extended the initial analysis by showing that the gene-environment interaction could not be accounted for by gene-environment correlation. Other studies have failed to replicate the gene-environment interaction effect (eg, ref 125). In a recent meta-analysis, however, the original finding was replicated. In addition the findings was extended to include childhood (closer in time to the maltreatment), and the possibility of a spurious finding was ruled out by accounting for gene-environment correlation.[126] The interaction between MAOA and childhood maltreatment in the etiology of antisocial PD appear to be one of the few replicated findings in the molecular genetics of PDs.

Future directions

Information from genetic epidemiologic studies can contribute to improvement in the validity of diagnoses of mental disorders, and thereby a more empirically based classification system.[49,56,127] Several lines of evidence, including multivariate twin studies, have shown that common axis I disorders can be divided into two main groups (internalizing and externalizing) based on shared etiological factors.[49,68] Currently an alternative classification system are being considered for DSM-V based on the hypothesis that, in addition to phenotypic similarity, spectra or clusters of disorder can be identified based on shared liability or risk factors.[56] Such clusters transcend the axis I-axis II division. Multivariate twin studies, including a comprehensive number of axis I and axis II disorders, could provide new important insights relevant to this proposal and further clarify the etiology of mental disorders by identifying genetic and environmental risk factors shared in common between groups of disorders.

Methods like genome-wide association studies,[128] analyses of copy-number variation,[129] studies of rare genetic variants,[130] epigenetic methods,[131] and deep sequencing of genomic regions[132] have not yet been applied to PDs, and will hopefully contribute to our understanding of the genetic etiology of these disorders in the future. One problem is, however, that the current phenotypes might be inadequate.[128] It is highly unlikely that the new DSM-V classification of PDs will provide a solution. A strategy that has been proposed to increase the rate of success for molecular genetics in psychiatry is the use of endophenotypes, defined as a heritable characteristic that is along the pathway between a disorder and genotype.[5] Although the strategy has not yet proven to be successful,[133] it has been suggested that this approach should be applied to the study of PDs by using clinical dimensions like for example affective instability, impulsivity, and aggression instead of diagnoses.[134]

REFERENCES

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 3rd ed. Washington, DC: American Psychiatric Association; 1980.
2. McGuffin P, Moffitt T, Thapar A. Personality disorders. In: McGuffin P, Owen MJ, Gottesman II, eds. *Psychiatric Genetics and Genomics*. Oxford, UK: Oxford University Press; 2002:183-210.
3. Livesley WJ, Jang KL. The Behavioral genetics of personality disorder. *Ann Rev Clin Psychol*. 2008;4247-274.
4. Reichborn-Kjennerud T. Genetics of personality disorders. *Psychiatr Clin N Am*. 2008;31:421-440.
5. Gottesman II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry*. 2003;160:636-645.
6. Braff DL, Freedman R, Schork NJ, Gottesman II. Deconstructing schizophrenia: an overview of the use of endophenotypes in order to understand a complex disorder. *Schizophr Bull*. 2007;33:21-32.
7. Kendler KS. Psychiatric genetics: a methodologic critique. *Am J Psychiatry*. 2005;162:3-11.
8. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington, DC: American Psychiatric Association; 1994.
9. Oldham JM, Skodol AE, Kellman HD, Hyler SE, Rosnick L, Davies M. Diagnosis of DSM-III-R personality disorders by two structured interviews: patterns of comorbidity. *Am J Psychiatry*. 1992;149:213-220.
10. Coid J, Yang M, Tyrer P, Roberts A, Ullrich S. Prevalence and correlates of personality disorder in Great Britain. *Br J Psychiatry*. 2006;188:423-431.
11. Lenzenweger MF, Lane MC, Loranger AW, Kessler RC. DSM-IV personality disorders in the national comorbidity survey replication. *Biol Psychiatry*. 2007;62:553-564.
12. Krueger RF. Continuity of axes I and II: toward a unified model of personality, personality disorders, and clinical disorders. *J Personality Disord*. 2005;19:233-261.
13. Siever LJ, Davis KL. A psychobiological perspective on the personality disorders. *Am J Psychiatry*. 1991;148:1647-1658.
14. Widiger TA. Personality disorder and Axis I psychopathology: the problematic boundary of Axis I and axis II. *J Personality Disord*. 2003;17:90-108.
15. Sheets E, Craighead WE. Toward an empirically based classification of personality pathology. *Clinl Psychol Sci Pract*. 2007;14:77-93.
16. Oldham JM, Skodol AE. Charting the future of axis II. *J Personal Disord*. 2000;14:17-29.
17. Widiger TA, Samuel DB. Diagnostic categories or dimensions? A question for the Diagnostic and Statistical Manual of Mental Disorders - Fifth Edition. *J Abnorm Psychol*. 2005;114:494-504.
18. Widiger TA, Trull TJ. Plate tectonics in the classification of personality disorder - shifting to a dimensional model. *Am Psychologist*. 2007;62:71-83.
19. Krueger RF, Skodol AE, Livesley WJ, Shrout PE, Huang YQ. Synthesizing dimensional and categorical approaches to personality disorders: refining the research agenda for DSM-V Axis II. *Int J Methods Psych Res*. 2007;16:S65-S73.
20. Kendler KS. Twin studies of psychiatric illness: an update. *Arch Gen Psychiatry*. 2001;58:1005-1014.
21. Neale MC, Boker SM, Xie G, Maes HH. Mx Statistical Modeling. 5th. 1999. Box 710 MCV, Richmond, VA 23298: Medical College of VA of VA Commonwealth Univ. Ref Type: Computer Program.
22. Falconer DS. The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Ann Hum Genet*. 1965;29:51-76.
23. Neale MC, Eaves LJ, Kendler KS. The power of the classical twin study to resolve variation in threshold traits. *Behav Genet*. 1994;24:239-258.
24. Bouchard TJ, Loehlin JC. Genes, evolution, and personality. *Behav Genet*. 2001;31:243-273.
25. Ando J, Suzuki A, Yamagata S, et al. Genetic and environmental structure of Cloninger's temperament and character dimensions. *J Personal Disord*. 2004;18:379-393.
26. Livesley WJ, Jang KL, Vernon PA. Phenotypic and genetic structure of traits delineating personality disorder. *Arch Gen Psychiatry*. 1998;55:941-948.
27. Saulsman LM, Page AC. The five-factor model and personality disorder empirical literature: a meta-analytic review. *Clin Psychol Rev*. 2004;23:1055-1085.
28. Parnas J, Licht D, Bovet P. Cluster A personality disorders: a review. In: Maj M, Akiskal H, Mezzich JE, Okasha A, eds. *Personality Disorders*. New York, NY: John Wiley & Sons Ltd; 2005:1-124.
29. Kendler KS, Heath A, Martin NG. A genetic epidemiologic study of self-report suspiciousness. *Comp Psychiatry*. 1987;28:187-196.

La epidemiología genética de los trastornos de personalidad

Los estudios de epidemiología genética señalan que los diez trastornos de personalidad (TP) clasificados en el eje II del DSM-IV tienen una herencia leve a moderada. Los factores ambientales compartidos y genéticos no aditivos son de importancia menor o carecen de ésta. No se han identificado diferencias por sexo. Los estudios multivariados sugieren que la amplia comorbilidad entre los TP se puede explicar por tres factores de riesgo ambientales y genéticos comunes. Los factores genéticos no reflejan la estructura de grupos del DSM-IV, pero sí: 1) la alta vulnerabilidad para la patología de los TP o para la emocionalidad negativa, 2) la alta impulsividad/baja afabilidad y 3) la introversión. Los factores de riesgo genéticos y ambientales comunes contribuyen a la comorbilidad entre parejas o grupos de trastornos de los ejes I y II. Los estudios de genética molecular de los TP, principalmente los estudios de asociación de genes candidatos, señalan que están involucrados los genes vinculados a los sistemas de neurotransmisión, principalmente serotoninérgicos y dopaminérgicos. Estudios a futuro, que utilicen métodos más nuevos como la asociación del genoma completo, pueden aprovechar el empleo de endofenotipos.

Epidémiologie génétique des troubles de la personnalité

Des études d'épidémiologie génétique montrent que les 10 troubles de la personnalité (TP) classés sur l'axe II du DSM-IV sont légèrement à modérément transmissibles. Les facteurs génétiques non additifs et les facteurs environnementaux partagés sont de peu ou sans importance et il n'y a pas de différences selon le sexe. Des études multivariées suggèrent que trois facteurs de risque génétiques et environnementaux courants peuvent expliquer la comorbidité importante entre les TP. Les facteurs génétiques ne reflètent pas la structure en cluster *du DSM-IV mais plutôt : 1) une grande vulnérabilité aux TP ou à une émotivité négative ; 2) une impulsivité importante/peu d'amabilité ; 3) une introversion. Des facteurs de susceptibilité génétiques et environnementaux communs participent à la comorbidité entre les paires ou les groupes des troubles de l'axe I et de l'axe II. Des études de génétique moléculaire des TP, pour la plupart des études d'association de gène candidat, montrent que sont impliqués les gènes liés aux voies des neurotransmetteurs, surtout dans les systèmes sérotoninergiques et dopaminergiques. Des études futures, utilisant la méthodologie de recherche d'associations sur génome entiers pourraient bénéficier de l'utilisation d' endophénotypes.*

30. Claridge G, Hewitt JK. A Biometrical study of schizotypy in a normal population. *Person Individ Diff.1987;8:303-312*.

31. Kendler KS, Hewitt JK. The structure of self-report schizotypy in twins. *J Personal Disord*. 1992;61-17.

32. Linney YM, Murray RM, Peters ER, Macdonald AM, Rijsdijk F, Sham PC. A quantitative genetic analysis of schizotypal personality traits. *Psychol Med*. 2003;33:803-816.

33. Jang KL, Woodward TS, Lang D, Honer WG, Livesley WJ. The genetic and environmental basis of the relationship between schizotypy and personality - a twin study. *J Nerv Ment Dis*. 2005;193:153-159.

34. Torgersen S, Lygren S, Oien PA, et al. A twin study of personality disorders. *Comp Psychiatry*. 2000;41:416-425.

35. Kendler KS, Czajkowski N, Tambs K, et al. Dimensional representations of DSM-IV Cluster A personality disorders in a population-based sample of Norwegian twins: a multivariate study. *Psychol Med*. 2006;36:1583-1591.

36. McGlashan TH, Grilo CM, Sanislow CA, et al. Two-year prevalence and stability of individual DSM-IV criteria for schizotypal, borderline, avoidant, and obsessive-compulsive personality disorders: toward a hybrid model of axis II disorders. *Am J Psychiatry*. 2005;162:883-889.

37. Kendler KS, Myers J, Torgersen S, Neale MC, Reichborn-Kjennerud T. The heritability of cluster A personality disorders assessed by both personal interview and questionnaire. *Psychol Med*. 2007;37:655-665.

38. Rhee SH, Waldman ID. Genetic and environmental influences on antisocial behavior: a meta-analysis of twin and adoption studies. *Psychol Bull*. 2002;128:490-529.

39. White CN, Gunderson JG, Zanarini MC, Hudson JI. Family studies of borderline personality disorder: A review. *Harv Rev Psychiatry*. 2003;11:8-19.

40. Distel MA, Trull TJ, Derom CA, et al. Heritability of borderline personality disorder features is similar across three countries. *Psychol Med*. 2008;38:1219-1229.

41. Distel MA, Rebollo-Mesa I, Willemsen G, et al. Familial resemblance of borderline personality disorder features: genetic or cultural transmission? *Plos One*. 2009;4:e5334.

42. Keller MC, Coventry WL, Heath AC, Martin NG. Widespread evidence for non-additive genetic variation in Cloninger's and Eysenck's personality dimensions using a twin plus sibling design. *Behav Genet*. 2005;35:707-721.

43. Torgersen S, Czajkowski N, Jacobson K, et al. Dimensional representations of DSM-IV cluster B personality disorders in a population-based sample of Norwegian twins: a multivariate study. *Psychol Med*. 2008;38:1617-1625.

44. Reich JH. Familiality of DSM-III dramatic and anxious personality clusters. *J Nerv Ment Dis*. 1989;177:96-100.

45. Reichborn-Kjennerud T, Czajkowski N, Neale MC, et al. Genetic and environmental influences on dimensional representations of DSM-IV cluster C personality disorders: a population-based multivariate twin study. *Psychol Med*. 2007;37:645-653.

46. Orstavik RE, Kendler KS, Czajkowski N, Tambs K, Reichborn-Kjennerud T. Genetic and environmental contributions to depressive personality disorder in a population-based sample of Norwegian Twins. *J Affect Dis*. 2007;99:181-189.
47. Kendler KS, Gatz M, Gardner CO, Pedersen NL. A Swedish national twin study of lifetime major depression. *Am J Psychiatry*. 2006;163:109-114.
48. Czajkowski N, Kendler KS, Jacobson KC, Tambs K, Roysamb E, Reichborn-Kjennerud T. Passive-aggressive (negativistic) personality disorder: a population-based twin study. *J Personal Disord*. 2008;22:109-122.
49. Krueger RF, Markon KE. Reinterpreting comorbidity: a model-based approach to understanding and classifying psychopathology. *Ann Rev Clin Psychol*. 2006;2111-133.
50. Neale MC, Kendler KS. Models of comorbidity for multifactorial disorders. *Am J Hum Genet*. 1995;57:935-953.
51. Zanarini MC, Barison LK, Frankenburg FR, Reich B, Hudson JI. Family history study of the familial coaggregation of borderline personality disorder with axis I and nonborderline dramatic cluster axis II disorders. *J Personal Disord.* 2009;23:357-369.
52. Kendler KS, Aggen SH, Czajkowski N, et al. The structure of genetic and environmental risk factors for DSM-IV personality disorders: a multivariate twin study. *Arch Gen Psychiatry*. 2008;65:1438-1446.
53. Saulsman LM, Page AC. The five-factor model and personality disorder empirical literature: A meta-analytic review. *Clin Psychol Rev*. 2004;23:1055-1085.
54. Tyrer P, Gunderson J, Lyons M, Tohen M. Extent of comorbidity between mental state and personality disorders. *J Personal Disord*. 1997;11:242-259.
55. Dolan-Sewell RT, Krueger RF, Shea MT. Co-occurrence with syndrome disorders. In: Livesley WJ ed. *Handbook of Personality Disorders: Theory, Research and Treatment*. New York, NY: The Guilford Press; 2001:84-104.
56. Parnas J, Cannon TD, Jacobsen B, Schulsinger H, Schulsinger F, Mednick SA. Lifetime DSM-III-R diagnostic outcomes in the offspring of schizophrenic mothers. Results from the Copenhagen High-Risk Study. *Arch Gen Psychiatry*. 1993;50:707-714.
57. Andrews G, Goldberg DP, Krueger R, et al. Exploring the feasibility of a meta-structure for DSM-V and ICD-11: could it improve utility and validity? *Psychol Med*. 2009;391993-2000.
58. Kendler KS, McGuire M, Gruenberg AM, O'Hare A, Spellman M, Walsh D. The Roscommon Family Study. III. Schizophrenia-related personality disorders in relatives. *Arch Gen Psychiatry*. 1993;50:781-788.
59. Kety SS, Wender PH, Jacobsen B, et al. Mental illness in the biological and adoptive relatives of schizophrenic adoptees. Replication of the Copenhagen Study in the rest of Denmark. *Arch Gen Psychiatry*. 1994;51:442-455.
60. Onstad S, Skre I, Edvardsen J, Torgersen S, Kringlen E. Mental disorders in first-degree relatives of schizophrenics. *Acta Psychiatr Scand*. 1991;83:463-467.
61. Asarnow RF, Nuechterlein KH, Fogelson D, et al. Schizophrenia and schizophrenia-spectrum personality disorders in the first-degree relatives of children with schizophrenia: the UCLA family study. *Arch Gen Psychiatry*. 2001;58:581-588.
62. Tienari P, Wynne LC, Laksy K, et al. Genetic boundaries of the schizophrenia spectrum: Evidence from the Finnish adoptive family study of schizophrenia. *Am J Psychiatry*. 2003;160:1587-1594.
63. Torgersen S, Onstad S, Skre I, Edvardsen J, Kringlen E. "True" schizotypal personality disorder: a study of co-twins and relatives of schizophrenic probands. *Am J Psychiatry*. 1993;150:1661-1667.
64. Baron M, Gruen R, Rainer JD, Kane J, Asnis L, Lord S. A family study of schizophrenic and normal control probands: implications for the spectrum concept of schizophrenia. *Am J Psychiatry*. 1985;142:447-455.
65. Siever L, Davis KL. The pathopysiology of schizophrenia disorders: perspectives from the spectrum. *Am J Psychiatry*. 2004;161:398-413.
66. Fogelson DL, Nuechterlein KH, Asarnow RA, et al. Avoidant personality disorder is a separable schizophrenia-spectrum personality disorder even when controlling for the presence of paranoid and schizotypal personality disorders - the UCLA family study. *Schizophr Res*. 2007;91:192-199.
67. Asarnow RF, Nuechterlein KH, Fogelson D, et al. Schizophrenia and schizophrenia-spectrum personality disorders in the first-degree relatives of children with schizophrenia: the UCLA family study. *Arch Gen Psychiatry*. 2001;58:581-588.
68. Kendler KS, Prescott CA, Myers J, Neale MC. The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. *Arch Gen Psychiatry*. 2003;60:929-937.
69. Hettema JM, Neale MC, Myers JM, Prescott CA, Kendler KS. A population-based twin study of the relationship between neuroticism and internalizing disorders. *Am J Psychiatry*. 2006;163:857-864.
70. Riso LP, Klein DN, Anderson RL, Ouimette PC. A family study of outpatients with borderline personality disorder and no history of mood disorder. *J Personal Disord*. 2000;14:208-217.
71. Reichborn-Kjennerud T, Czajkowski N, Røysamb E, Orstavik RE, et al. Major depression and dimensional representations of DSM-IV personality disorders: a population-based twin study. *Psychol Med*. 2009 Nov 17;Epub ahead of print.
72. Ørstavik RE, Kendler KS, Czajkowski N, Tambs K, Reichborn-Kjennerud T. The relationship between depressive personality disorder and major depressive disorder: a population-based twin study. *Am J Psychiatry*. 2007;164:1866-1872.
73. Reichborn-Kjennerud T, Czajkowski N, Torgersen S, et al. The relationship between avoidant personality disorder and social phobia: a population-based twin study. *Am J Psychiatry*. 2007;164:1722-1728.
74. Krueger RF, Hicks BM, Patrick CJ, Carlson SR, Iacono WG, McGue M. Etiologic connections among substance dependence, antisocial behavior, and personality: modeling the externalizing spectrum. *J Abnorm Psychol*. 2002;111:411-424.
75. Hicks BM, Krueger RF, Iacono WG, McGue M, Patrick CJ. Family transmission and heritability of externalizing disorders - a twin-family study. *Arch Gen Psychiatry*. 2004;61:922-928.
76. Hicks BM, Bernat E, Malone SM, et al. Genes mediate the association between P3 amplitude and externalizing disorders. *Psychophysiology*. 2007;44:98-105.
77. Lyons MJ, True WR, Eisen SA, et al. Differential heritability of adult and juvenile antisocial traits. *Arch Gen Psychiatry*. 1995;52:906-915.
78. Silberg JL, Rutter M, Tracy K, Maes HH, Eaves L. Etiological heterogeneity in the development of antisocial behavior: the Virginia twin study of adolescent behavioral development and the young adult follow-up. *Psychol Med*. 2007;37:1193-1202.
79. Hicks BM, Blonigen DM, Kramer MD, et al. Gender differences and developmental change in externalizing disorders from late adolescence to early adulthood: a longitudinal twin study. *J Abnorm Psychol*. 2007;116:433-447.
80. Kendler KS, Prescott CA. *Genes, Environment, and Psychopathology*. New York, NY: The Guilford Press; 2006.
81. Rutter M, Moffitt TE, Caspi A. Gene-environment interplay and psychopathology: multiple varieties but real effects. *J Child Psychol Psychiatry*. 2006;47:226-261.
82. Purcell S, Sham P. Variance components models for gene-environment interaction in quantitative trait locus linkage analysis. *Twin Res*. 2002;5:572-576.
83. Kendler KS, Baker JH. Genetic influences on measures of the environment: a systematic review. *Psychol Med*. 2007;37:615-626.
84. Jaffee SR, Price TS. Gene-environment correlations: a review of the evidence and implications for prevention of mental illness. *Mol Psychiatry*. 2007;12:432-442.
85. Tsuang MT, Bar JL, Stone WS, Faraone SV. Gene-environment interactions in mental disorders. *World Psychiatry*. 2004;3:73-83.
86. Crowe RR. An adoption study of antisocial personality. *Arch Gen Psychiatry*. 1974;31:785-791.
87. Cadoret RJ, Cain CA, Crowe RR. Evidence for gene-environment interaction in the development of adolescent antisocial behavior. *Behav Genet*. 1983;13:301-310.
88. Cadoret RJ, Yates WR, Troughton E, Woodworth G, Stewart MA. Genetic-environmental interaction in the genesis of aggressivity and conduct disorders. *Arch Gen Psychiatry*. 1995;52:916-924.
89. Jaffee SR, Caspi A, Moffitt TE, et al. Nature X nurture: genetic vulnerabilities interact with physical maltreatment to promote conduct problems. *Dev Psychopathol*. 2005;17:67-84.
90. Tuvblad C, Grann M, Lichtenstein P. Heritability for adolescent antisocial behavior differs with socioeconomic status: gene-environment interaction. *J Child Psychol Psychiatry*. 2006;47:734-743.

91. Feinberg ME, Button TMM, Neiderhiser JM, Reiss D, Hetherington EM. Parenting and adolescent antisocial behavior and depression - evidence of genotype x parenting environment interaction. *Arch Gen Psychiatry*. 2007;64:457-465.
92. Tienari P, Wynne LC, Sorri A, et al. Genotype-environment interaction in schizophrenia-spectrum disorder - long-term follow-up study of Finnish adoptees. *Br J Psychiatry*. 2004;184:216-222.
93. Hicks BM, South SC, Dirago AC, Iacono WG, McGue M. Environmental adversity and increasing genetic risk for externalizing disorders. *Arch Gen Psychiatry*. 2009;66:640-648.
94. Sham P, McGuffin P. Linkage and association. In: McGuffin P, Owen MJ, Gottesman II eds. *Psychiatric Genetics & Genomics*. Oxford, UK: Oxford University Press; 2002:55-73.
95. Zondervan KT, Cardon LR. Designing candidate gene and genome-wide case-control association studies. *Nat Protocols*. 2007;2:2492-2501.
96. Ioannidis JPA, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. *Nature Genetics.* 2001;29:306-309.
97. NCI-NHGRI Working Group on Replication in Association Studies. Replicating genotype-phenotype associations. *Nature*. 2007;447:655-660.
98. Rosmond R, Rankinen T, Chagnon M, et al. Polymorphism in exon 6 of the dopamine D-2 receptor gene (DRD2) is associated with elevated blood pressure and personality disorders in men. *J Hum Hypertens*. 2001;15:553-558.
99. Stefanis NC, Trikalinos TA, Avramopoulos D, et al. Impact of schizophrenia candidate genes on schizotypy and cognitive endophenotypes at the population level. *Biol Psychiatry*. 2007;62:784-792.
100. Fanous AH, Neale MC, Gardner CO, et al. Significant correlation in linkage signals from genome-wide scans of schizophrenia and schizotypy. *Mol Psychiatry*. 2007;12:958-965.
101. Avramopoulos D, Stefanis NC, Hantoumi I, Smyrnis N, Evdokimidis I, Stefanis CN. Higher scores of self reported schizotypy in healthy young males carrying the COMT high activity allele. *Mol Psychiatry*. 2002;7:706-711.
102. Stefanis NC, van Os J, Avramopoulos D, et al. Variation in catechol-O-methyltransferase val met genotype associated with schizotypy but not cognition: A population study in 543 young men. *Biol Psychiatry*. 2004;56:510-515.
103. Schurhoff F, Szoke A, Chevalier F, et al. Schizotypal dimensions: An intermediate phenotype associated with the COMT high activity allele. *Am J Med Genet B-Neuropsych Genet*. 2007;144B:64-68.
104. Lewandowski KE. Relationship of catechol-o-methyltransferase to schizophrenia and its correlates: Evidence for associations and complex interactions. *Harv Rev Psychiatry*. 2007;15:233-244.
105. Ni XQ, Chan K, Bulgin N, et al. Association between serotonin transporter gene and borderline personality disorder. *J Psych Res*. 2006;40:448-453.
106. Ni XQ, Sicard T, Bulgin N, et al. Monoamine oxidase A gene is associated with borderline personality disorder. *Psychiatr Genet*. 2007;17:153-157.
107. Ni XQ, Bismil R, Chan K, et al. Serotonin 2A receptor gene is associated with personality traits, but not to disorder, in patients with borderline personality disorder. *Neurosci Lett*. 2006;408:214-219.
108. Ni X, Chan D, Chan K, McMain S, Kennedy JL. Serotonin genes and gene-gene interactions in borderline personality disorder in a mached case-control study. *Prog Neuropharmacol Biol Psychiatry*. 2009;33:128-133.
109. Steiger H, Richardson J, Joober R, et al. The 5HTTLPR polymorphism, prior maltreatment and dramatic-erratic personality manifestations in women with bulimic syndromes. *J Psychiatr Neurosci*. 2007;32:354-362.
110. Lyons-Ruth K, Holmes BM, Sasvari-Szekely M, Ronai Z, Nemoda Z, Pauls D. Serotonin transporter polymorphism and borderline or antisocial traits among low-income young adults. *Psychiatr Genet*. 2007;17:339-343.
111. Jacob CP, Strobel A, Hohenberger K, et al. Association between allelic variation of serotonin transporter function and neuroticism in anxious cluster C personality disorders. *Am J Psychiatry*. 2004;161:569-572.
112. Jacob CP, Muller J, Schmidt M, et al. Cluster B personality disorders are associated with allelic variation of monoamine oxidase a activity. *Neuropsychopharmacology*. 2005;30:1711-1718.
113. Williams LM, Gatt JM, Kuan SA, et al. A polymorphism of the MAOA gene is associated with emotional brain markers and personality traits on an antisocial index. *Neuropsychopharmacology*. 2009;34:1797-1809.
114. Wilson ST, Stanley B, Brent DA, Oquendo MA, Huang Y, Mann JJ. The tryptophan hydroxylase-1 A218C polymorphism is associated with diagnosis, but not suicidal behavior, in borderline personality disorder. *Am J Med Genet Part B.* 2009;150B;202-208.
115. Gutknecht L, Jacob C, Strobel A, et al. Tryptophan hydroxylase-2 gene variation influences personality traits and disorders related to emotional dysregulation. *Int J Neuropsychopharmacol*. 2007;10:309-320.
116. Lesch KP, Bengel D, Heils A, et al. Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science*. 1996;274:1527-1531.
117. Munafo MR, Clark T, Flint J. Does measurement instrument moderate the association between the serotonin transporter gene and anxiety-related personality traits? A meta-analysis. *Mol Psychiatry*. 2005;10:415-419.
118. Stein MB, Fallin MD, Schork NJ, Gelernter J. COMT polymorphisms and anxiety-related personality traits. *Neuropsychopharmacology*. 2005;30:2092-2102.
119. Hettema JM, An SS, Bukszar J, et al. Catechol-O-methyltransferase contributes to genetic susceptibility shared among anxiety spectrum phenotypes. *Biol Psychiatry*. 2008;64:302-310.
120. Joyce PR, Rogers GR, Miller AL, Mulder RT, Luty SE, Kennedy MA. Polymorphisms of DRD4 and DRD3 and risk of avoidant and obsessive personality traits and disorders. *Psychiatr Res*. 2003;119:1-10.
121. Light KJ, Joyce PR, Luty SE, et al. Preliminary evidence for an association between a dopamine D3 receptor gene variant and obsessive-compulsive personality disorder in patients with major depression. *Am J Med Genet B-Neuropsych Genet*. 2006;141B:409-413.
122. Dick DM, Agrawal A, Shuckit MA, et al. Marital status, alcohol dependence, and GRBRA2: evidence for gene-environment correlation and interaction. *J Stud Alcohol*. 2006;67:185-194.
123. Caspi A, McClay J, Moffitt TE, et al. Role of genotype in the cycle of violence in maltreated children. *Science*. 2002;297:851-854.
124. Foley DL, Eaves LJ, Wormley B, et al. Childhood adversity, monoamine A geneotype, and risk for conduct disorder. *Arch Gen Psychiatry*. 2004;61:738-744.
125. Huizinga D, Haberstick BC, Smolen A, et al. Childhood maltreatment, subsequent antisocial behavior, and the role of monoamine oxidase A genotype. *Biol Psychiatry*. 2006;60:677-683.
126. Kim-Cohen J, Caspi A, Taylor A, et al. MAOA, maltreatment, and gene-environment interaction predicting children's mental health: new evidence and a meta-analysis. *Mol Psychiatry*. 2006;11:903-913.
127. Kendler KS. Reflections on the relationship between psychiatric genetics and psychiatric nosology. *Am J Psychiatry*. 2006;163:1138-1146.
128. Cichon S, Craddock N, Daly M, et al. Genomewide association studies: history, rationale, and prospects for psychiatric disorders. *Am J Psychiatry*. 2009;166:540-556.
129. Cook EH, Scherer SW. Copy-number variations associated with neuropsychiatric conditions. *Nature*. 2008;455:919-923.
130. Goldstein DB. Common genetic variation and human traits. *N Engl J Med*. 2009;360:1696-1698.
131. Stuffrein-Roberts S, Joyce PR, Kennedy MA. Role of epigenetics in mental disorders. *Aust N Z J Psychiatry*. 2008;42:97-107.
132. lint J, Munafo MR. The endophenotype concept in psychiatric genetics. *Psychol Med*. 2007;37:163-180.
133. Flint J, Munafo MR. The endophenotype concept in psychiatric genetics. *Psychol Med*. 2007;37:163-180.
134. Siever LJ. Endophenotypes in the personality disorders. *Dialogues Clin Neurosci*. 2005;7:139-151.

Brief report

A glossary of relevant genetic terms

Deborah J. Morris-Rosendahl, PhD

Allele
One of two or more alternate forms of a gene or marker at a particular locus on a chromosome.

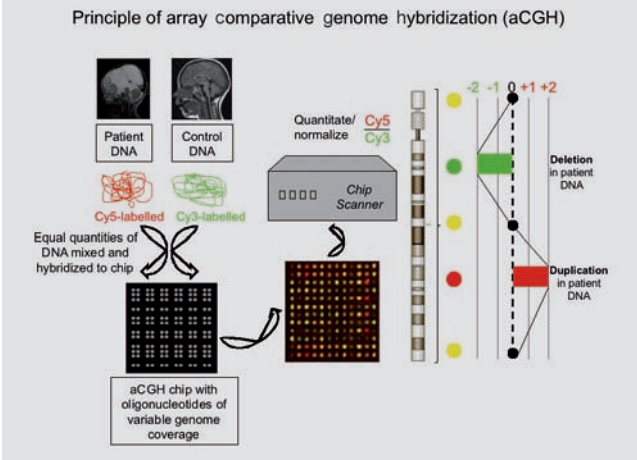
Anticipation (genetic)
Apparent earlier age of onset and increased severity of a disease in successive generations, eg, Huntington's disease.

cDNA
Single-stranded complementary DNA, ie, a DNA molecule synthesized from a RNA template by reverse transcription of RNA.

Common disorder common variant (CDCV) hypothesis
A theory that many common diseases are caused by common alleles that individually have little effect, but in concert confer a high risk.

Complex disease
A disorder in which the cause is considered to be a combination of genetic effects and environmental influences.

Comparative genome hybridization (CGH)
CGH is a molecular-cytogenetic method for the analysis of copy number changes (gains or losses) in the DNA content of a given individual's DNA.



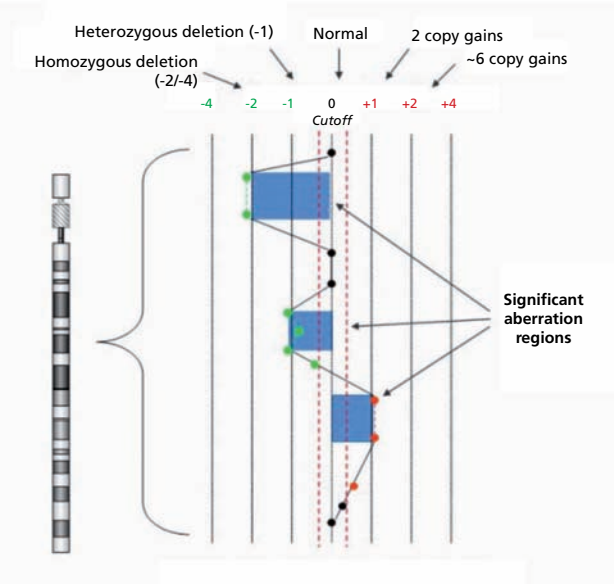
Compound heterozygosity
Heterozygosity for two different mutant alleles of a gene, often the case for autosomal recessive disorders.

Copy number variation (CNV)
A segment of DNA in which copy number differences have been found by comparison of two or more genomes. The segment may range from one kilobase to several megabases in size. The variation is usually due to deletion or duplication.

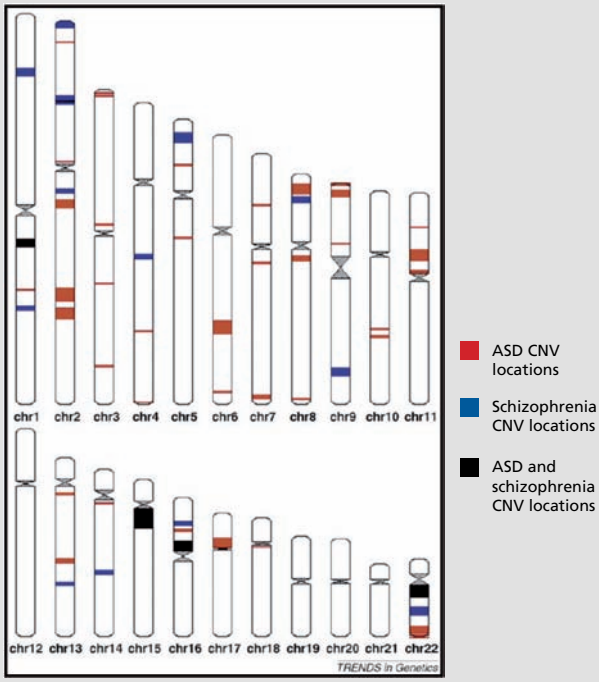
Author affiliations: Institute for Human Genetics, University Clinic Freiburg, Germany

Address for correspondence: Priv.-Doz. Dr. Deborah J. Morris-Rosendahl, Universitaetsklinikum Freiburg, Institut fuer Humangenetik Breisacherstr. 33, D-79106 Freiburg, Germany
(e-mail: deborah.morris-rosendahl@uniklinik-freiburg.de)

Copy number variation (CNV)



CNV association findings in schizophrenia and autism spectrum disorder (ASD)



Adapted from: Merikangas AK, Corvin AP, Gallagher L. Copy number variants in neurodevelopmental disorders: promises and challenges. *Trends Genet*. 2009;25:536-544. Copyright © Elsevier 2009

Deep resequencing

A technique for sequencing a gene in several thousand subjects, typically using one of the new high-throughput sequencing technologies.

Epigenetics

Heritable changes to DNA structure that do not alter the underlying DNA sequence, eg, DNA methylation.

Epigenomics

The application of epigenetics to the whole genome.

Exome

The approximately 1% of the human genome that comprises all exons and therefore the entire protein-coding region of the genome.

Genetic association

The nonrandom occurrence of a genetic marker (usually a particular allele of a polymorphism) with a trait, which suggests an association between the genetic marker (or marker close to it) and disease pathogenesis.

Genome

In eukaryotes, the basic (monoploid) chromosome set, consisting of a species-specific number of linkage groups and the genes contained therein. For example, in humans, the genome consists of the 24 different chromosomes (22 autosomes, X and Y chromosomes). The mitochondrial DNA is usually considered to be a separate "mitochondrial" genome.

Genome-wide association study (GWAS)

A test for the association between genetic polymorphisms spread evenly over the entire genome, and a disease. Usually at least 300 000 markers are required to adequately cover the genome.

Genotype

The genetic constitution with respect to the alleles at one or more pairs of genetic loci under observation. The genotype of an individual is the sum total of the genetic information contained on the chromosomes, as distinguished from the individual's phenotype (idiotype).

Brief report

Haploid
A single genome or set of chromosomes (eg, in human gametes, n=23), compared to the normal diploid (double) set of chromosomes (n=46).

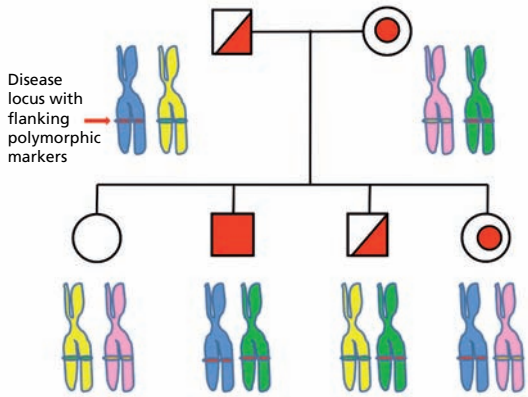
Haplotype
A combination of alleles at closely linked gene loci that are inherited together.

Hemizygous
When one or more genes is present in only one, instead of two copies, eg, men are hemizygous for most genes on the X and Y chromosomes.

Heterozygous
Having different alleles for one or more genes in homologous chromosome segments, as opposed to being homozygous with identical alleles at these loci.

Linkage
Genetic linkage refers to the observation that two or more genes located on the same chromosome are inherited together. The ratio of being transmitted together versus being separated allows an estimate of their distance from each other (recombination fraction).

Linkage analysis

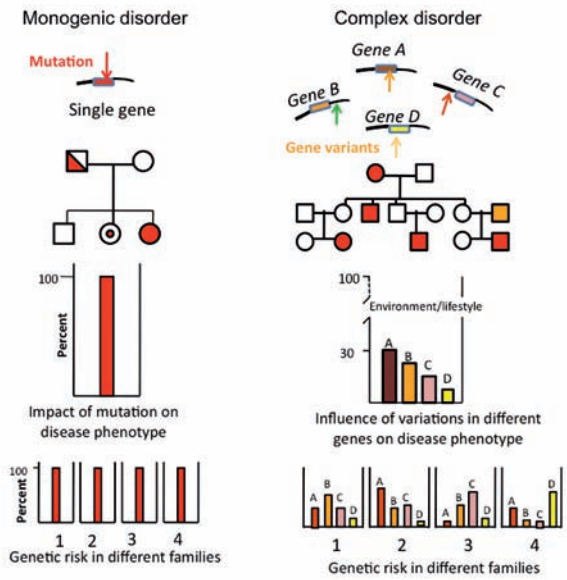


Linkage disequilibrium (LD)
Alleles at different loci that are inherited together more frequently or less frequently than expected by their individual frequencies are said to show linkage disequilibrium.

Methylation (of DNA)
The attachment of a methyl group to DNA. In vertebrates, this typically occurs at CpG sites (cytosine-phosphate-guanine sites) in the DNA sequence, resulting in the conversion of cytosine to 5-methylcytosine.

Monogenic disorder
Disorder caused by one or more mutations in a single gene, eg, cystic fibrosis (mutations in the *CFTR* gene). Such disorders are also sometimes referred to Mendelian diseases.

Monogenic vs complex disease



Penetrance
The frequency (in percent) with which a dominant or homozygous recessive gene or gene combination manifests itself in the phenotype of the carriers.

Pharmacogenetics
A branch of genetics which deals with the genetic variability in individual responses to drugs and drug metabolism.

Phenocopy
A nonhereditary, phenotypic modification (caused by special environmental conditions) that mimics a similar phenotype caused by a gene mutation.

Phenotype
The observable properties (structural and functional) of an organism, produced by the interaction between the

organism's genotype and the environment in which it finds itself.

Pleiotropy
Genes or mutations that result in the production of multiple, apparently unrelated, effects at the phenotypic level. For example, patients with phenylketonuria, caused by mutations in the *PAH* (phenylalanine hydroxylase) gene, have reduced hair and skin pigmentation in addition to mental retardation, resulting from toxic levels of phenylalanine.

Polymorphism (genetic)
A chromosome or DNA variant that is observed in natural populations. A gene locus is defined as polymorphic if a rare allele has a frequency of 0.01 (1%) or more.

Positional cloning
Finding disease genes based on knowledge of their chromosomal location (usually found via linkage analysis in families with the disease) as opposed to knowledge of the function of the gene or protein encoded by the gene.

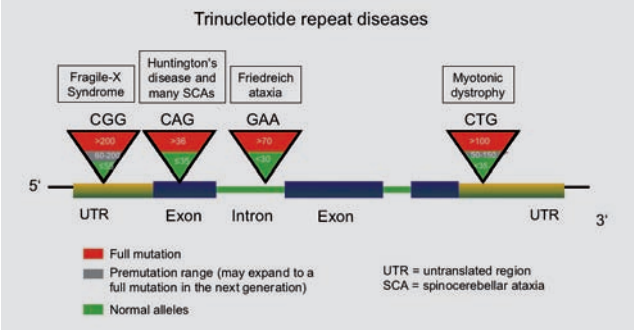
Second- or next-generation sequencing (also referred to as high-throughput sequencing)
New techniques that have increased the speed and decreased the cost of DNA sequencing by two orders of magnitude, enabling the sequencing of the entire genomes of many individuals.

Single nucleotide polymorphism (SNP)
Heritable polymorphism resulting from a single base pair change. SNPs generally have only two alleles.

Structural variant
Structural genomic variation includes any genetic variant that alters chromosomal structure, including inversions, translocations, duplications and deletions. Duplications and deletions, collectively known as CNVs (see *copy number variation*) are the most common form of structural variation in the human genome.

Synonymous nucleotide change/non-synonymous nucleotide change
A change in the DNA sequence which does not result in the change in the amino acid sequence, eg, GTT>GTC both code for Valine (Val or V). A nonsynonymous change results in the coding of a different amino acid (eg, GTT>GAT results in Val>Asp).

Trinucleotide repeat expansion
An increased number of contiguous trinucleotide repeats (eg, CAG, CGG) in the DNA sequence from one generation to the next. When the expansion extends into the pathological range, this type of mutation causes diseases such as Huntington's disease, fragile X syndrome, myotonic dystrophy, and many forms of spinocerebellar ataxia.



X-inactivation
The random, early embryological, inactivation of one of the X chromosomes in females, so that the expression of X-chromosomal genes is the same as that in males.

Useful genetic databases

National Center for Biotechnology information (NCBI)
http://www.ncbi.nlm.nih.gov/
Provides links to many other databases, including many of the databases below.

Online Mendelian Inheritance in Man (OMIM)
http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim
A comprehensive, authoritative, and timely compendium of human genes and genetic phenotypes.

Genome Database (GDB)
http://gdbwww.gdb.org/
Gene and protein sequence database.

UCSC Genome Browser/Bioinformatics site
http://genome.ucsc.edu/index.html
Provides the reference sequence and working draft assemblies for a large number of genomes. The browser

has many useful tools, eg, for searching for sequences within a genome, and comparing sequences within and between genomes.

ENSEMBL Database
http://www.ensembl.org
A genome database for vertebrates and other species, providing gene sequence data as well as chromosomal localization overviews and some information regarding transcripts and proteins.

db-SNP Polymorphism Repository
http://www.ncbi.nlm.nih.gov/SNP/
Database for single nucleotide polymorphisms and other classes of minor genetic variation.

The SNP Consortium Ltd. (TSC)
http://snp.cshl.org/
A non-profit foundation organized to develop up to 300 000 single nucleotide polymorphisms (SNPs) distributed evenly throughout the human genome and to make the information related to these SNPs available to the public without intellectual property restrictions.

European Bioinformatics Institute (EBI)
http://www2.ebi.ac.uk/
A centre for research and services in bioinformatics, which is part of the European Molecular Biology Laboratory (EMBL).

Gene Cards Database
http://bioinformatics.weizmann.ac.il/cards/
Summarizes most available information on a particular gene, with links to many other databases, eg, protein databases

International HapMap Project
http://hapmap.ncbi.nlm.nih.gov/
A partnership of scientists and funding agencies from Canada, China, Japan, Nigeria, the United Kingdom, and the USA to develop a public resource that will help researchers find genes associated with human disease and response to pharmaceuticals.

The Human Genome Variation Database
http://hgvbase.cgb.ki.se/
A reference for the nomenclature of genetic variation; also provides links to various mutation databases

The Human Gene Mutation Database
http://archive.uwcm.ac.uk/uwcm/mg/hgmd0.html
Database of published mutations for different disease genes.

The Pharmacogenomics and Pharmacogenetics Knowledgebase
http://pharmgkb.org/do/serve?id=home.welcome

The Human Variome Project
http://www.humanvariomeproject.org/
Web site of the global initiative to collect and curate all human genetic variation affecting human health.

Useful glossary references

Rieger R, Michaelis A, Green MM. *Glossary of Genetics, Classical and Molecular*. Berlin, Germany: Springer-Verlag; 1991.

Passarge E. *Colour Atlas of Genetics*. Stuttgart, Germany: Thieme Verlag; 2007.