Review

Biological Psychiatry

Genetics of Complex Traits in Psychiatry

Joel Gelernter

ABSTRACT

Virtually all psychiatric traits are genetically complex. This article discusses the genetics of complex traits in psychiatry. The complexity is accounted for by numerous factors, including multiple risk alleles, epistasis, and epigenetic effects such as methylation. Risk alleles can individually be common or rare, and can include, for example, single nucleotide polymorphisms and copy number variants that are transmitted or are new mutations, and other kinds of variation. Many different kinds of variation can be important for trait risk, either together in various proportions or as different factors in different subjects. Until more recently, approaches to complex traits were limited, and consequently only a few variants, usually of individually minor effect, were identified. At the present time, a much richer armamentarium exists that includes the routine application of genome-wide association studies and next-generation high-throughput sequencing and the combination of this information with other biologically relevant information, such as expression data. We have also seen the emergence of large meta-analysis and mega-analysis consortia. These developments are extremely important for psychiatric genetics, have advanced the field substantially, and promise formidable gains in the years to come as they are applied more widely.

Keywords: Complex Traits, Genetics, GWAS, G × E, Polymorphisms, Sequencing Studies

http://dx.doi.org/10.1016/j.biopsych.2014.08.005

This article reviews the genetics of complex traits—traits that do not follow the mendelian inheritance patterns of dominant, recessive, or sex-linked, a category encompassing nearly all psychiatric traits. The complexity is accounted for by numerous factors, including multiple risk alleles, epistatic (i.e., gene-gene interaction) effects, and epigenetic effects such as methylation. Risk alleles can individually be common or rare, and can include, for example, single nucleotide polymorphisms (SNPs) and copy number variants (CNV) that are transmitted or are new mutations, and other kinds of variation. Many of the different kinds of variation can be important for a trait, either together or as different factors in different subjects. Until more recently, the approaches to complex traits were limited, and consequently only a few variants, usually of individually minor effect, were identified. A much richer armamentarium exists at the present time that includes the routine application of genome-wide association studies (GWAS) and next-generation high-throughput sequencing (NextGen). We have also seen the emergence of large meta-analysis consortia and studies combining genetic polymorphism data with large data sets (e.g., gene expression in target tissues). These developments are extremely important for psychiatric genetics, have advanced the field substantially, and promise formidable gains in the years to come as they are applied more widely.

We can take schizophrenia as an illustration. This trait has been known to be moderately to highly heritable for almost 50 years. However, "traditional" approaches, including genetic linkage studies, candidate gene studies based on biological hypotheses, and targeted sequencing studies, yielded few replicated risk variants. This situation started to change with the recognition that velocardiofacial syndrome, which is marked by an easily discernible (if complex and variable) cytogenetic finding, shares phenotypic features with schizophrenia (1); the first wave of GWAS; the identification of genome-wide–significant (GWS) evidence for association in meta-analysis of multiple large data sets (2); and the discovery of strong evidence of many risk alleles individually of small effect in mega-analysis studies incorporating the data from many individual GWAS in single large analyses (3,4). A CNV component is well supported at the present time (5,6), and there is evidence of new mutation (7).

GENOME-WIDE STUDIES

Because we do not fully understand the biology of any psychiatric traits, most of the genes that are involved cannot be predicted a priori. Three general methods are used to identify risk genes without prior knowledge of risk mechanisms. They query the entire genome and use statistical methods of inference. Genome-wide linkage studies are the traditional approach to identifying risk loci. These family-based studies require the investigation of polymorphic markers that span the genome, allowing identification of chromosomal risk regions where markers are coinherited with the phenotype of interest. In GWAS, very closely spaced markers are required, typically ≥1 million as implemented at the present time (vs. 400 highly polymorphic markers for linkage) usually studied in unrelated individuals. The intention is to determine the genotype of enough markers such that there is at least one marker within linkage disequilibrium distance of any point in the genome. Current genotyping arrays accomplish this, but there are gaps, especially in genetically older populations that have

SEE COMMENTARY ON PAGE 3

 ISSN: 0006-3223

lower average linkage disequilibrium across the genome (e.g., individuals of African ancestry). Use of tiling arrays to detect CNVs is a related genome-wide approach. A third method, based on deep sequencing of entire exomes or genomes, now becoming popular as sequencing prices decline.

Successful genome-wide linkage studies [e.g., one that identified the X-chromosomal location of a risk gene for Brunner syndrome, a very rare single-gene disorder associated with violent behavior and cognitive deficits (8)] give the chromosomal locations of risk loci but generally do not identify specific genes. Many such investigations resulted only in large regions, rather than genes or risk alleles. In contrast, successful GWAS and sequencing studies can implicate specific genes and risk alleles immediately. The enthusiasm of a prior era for linkage studies of complex traits was borne largely of a lack of other genome-wide methods and was only partially rewarded. There are several examples of genes being identified based on linkage regions [notably, the identification of an *MAOA* mutation as the cause of Brunner syndrome (9) and of a *GABRA2* variation as influencing alcohol dependence risk and related endophenotypes (10)]. However, there are also many examples where there was no such identification; this can be attributed partly to insufficiently powered linkage studies, the inapplicability of the common disease/common variant model that underlies traditional linkage, and genetic differences between familial forms and nonfamilial forms of an illness. The difficulty in identifying a gene out of a linkage peak, together with the expense of recruiting families with multiple affected individuals as well as the now-easy access to genotyping microarrays for GWAS, has led to a dramatic decline in the use of linkage for complex traits.

GWAS

The mythical "ideal" genetic study design might be to obtain DNA from cases and controls, sift through the entire genome, and identify the differences. When all other sources of differences between the samples are accounted for, the distinctions that are left must account for the genetic part of the difference between the particular case and control samples for the phenotype that differentiates them; this is the basic idea of GWAS. The first major GWAS was published in 2005 (11) and identified polymorphic variants associated with age-related macular degeneration. The study included only 146 subjects, and it employed a genotyping microarray that included 106,000 markers that today would be considered unacceptably sparse. In contrast, GWAS performed today more typically employ thousands of samples and millions of markers, imputed (12) as well as the result of direct genotyping. The results have changed our understanding of complex trait genetics. At the start of the GWAS era, many expected that the method would identify the variants responsible for a large part of the genetic risk for most complex traits (13). They should have—if the "common disease/common variant" model was a good approximation of reality.

But risk alleles identified by GWAS for complex traits characteristically account for only a small percentage of the predicted genetic risk. There have been numerous discussions of the explanation of the "missing heritability." Although this question still cannot be answered definitively, an understanding of some of the important factors has emerged. One factor is the nature of the variants studied in GWAS, which are considered "common" variants. The risk for complex traits was previously thought to be most likely composed of the cumulative risk from a set of common variants. The more usual result for GWAS has been the identification of risk alleles with odds ratios of $\leq$1.2. There are exceptions, but these are rare. Initially, investigators concentrated on variants that met Bonferroni-adjusted criteria for GWS, often taken as $p < 5 \times 10^{-8}$, which is a reasonable threshold to identify individual risk alleles that can reproducibly be shown to be associated with a trait. However, there many other true risk variants among the variants that fail to meet this criterion, and it has been shown that when large sets of such variants are taken into account, a much larger portion of trait heritability can be accounted for (14). Another source of "missing heritability" is that part accounted for by rare variants (RVs), which can be defined as alleles having a frequency <1%. The effects of RVs that individually have a large effect on risk (but are relatively unimportant on a population level because they are rare) are important for some traits, as has been revealed by sequencing studies discussed subsequently.

In most situations where GWAS have been applied, they have been successful at identifying risk variants for psychiatric traits. The early years were disappointing for such important traits as schizophrenia and bipolar affective disorder (15), but we now know that well-powered studies (2) can detect risk loci for those traits. However, for adequate power, GWAS may require tens of thousands of subjects, studied in meta-analysis. Interpreting the results presents additional challenges; what to do with a list of genes, each with only a small effect on phenotype, is not obvious.

A brief note about population differences and population stratification is warranted. When candidate gene studies of psychiatric traits were more common, especially in the early days, failures to replicate often seemed to be the rule rather than the exception. There are numerous explanations for these failures, including small sample size, phenotypic heterogeneity, and random chance (16). Another contributor was population stratification—that is, different ancestral populations often have different allele frequencies at marker loci simply because they are different populations (17,18). These allele frequency differences may have nothing to do with the trait under study or with any detectable phenotypic trait. At first, this factor could be controlled only by matching or by using family-controlled designs (e.g., the transmission-disequilibrium test (19)), but the development of statistical methods to control for stratification in samples of unrelated subjects—notably, structured association (20,21) and genomic control (22) methods—revolutionized the field. It has turned out to be critically important to control for stratification in GWAS, and a set of methods has been developed to control for population differences in GWAS as well, most notably principal components methods (23). These methods are responsible for the development of useful GWAS as much as the technical development of dense genotyping microarrays.

GWAS DATA BEYOND SINGLE SNP ANALYSIS: NETWORKS AND RISK SCORES

Numerous approaches have been suggested to aid in interpreting the output of the SNP association content of GWAS.