Published in final edited form as:
*Biol Psychiatry*. 2015 January 1; 77(1): 36–42. doi:10.1016/j.biopsych.2014.08.005.

# Genetics of complex traits in psychiatry

**Joel Gelernter**
Yale Univ School of Medicine, Departments of Psychiatry, Genetics, and Neurobiology and VA CT Healthcare Center

## 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 (SNPs) and copy number variants (CNV) 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 recently, our approaches to complex traits were limited, and consequently only a small number of variants, usually of individually minor effect, were identified. Currently, we have a much richer armamentarium that includes the routine application of genomewide association studies (GWAS) and next-generation high throughput sequencing (NextGen); 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 moved the field forward substantially, and promise formidable gains in the years to come as they are applied more widely.

### Keywords

genetics; complex traits; GWAS; polymorphisms; GxE; sequencing studies

## Introduction

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



**Financial Disclosures**
The author reports no biomedical financial interests or potential conflicts of interest.

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 recently, our approaches to complex traits were limited, and consequently only a small number of variants, usually of individually minor effect, were identified. Currently, we have a much richer armamentarium that includes the routine application of genomewide 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 datasets regarding, for example, gene expression in target tissues. These developments are extremely important for psychiatric genetics, have moved the field forward 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. Yet “traditional” approaches – genetic linkage studies, candidate gene studies based on biological hypotheses, targeted sequencing studies – yielded few replicated risk variants. This started to change with the recognition that velocardiofacial syndrome (VCFS), which is marked by an easily discernible (if complex and variable) cytogenetic finding, shares phenotypic features with schizophrenia (1); with the first wave of GWAS; with the identification of genomewide-significant (GWS) evidence for association in meta-analysis of multiple large datasets (2), and then the discovery of strong evidence of many risk alleles individually of small effect in mega-analysis studies incorporating the data from many individual GWAS studies in single large analyses (3), (4). Now, a CNV component is well-supported (5, 6) and there is evidence of new mutation (7).

## Genomewide 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*. There are now three general methods used to identify risk genes without prior knowledge of risk mechanisms. They query the entire genome and use statistical methods of inference. Genomewide 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 co-inherited with the phenotype of interest. Genomewide association studies (GWAS) require very closely spaced markers, typically a million or more as implemented presently (vs. as few as 400 highly polymorphic markers for linkage) usually studied in unrelated individuals. The intention is to genotype 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 lower average linkage disequilibrium across the genome, e.g. African-ancestry individuals. Use of tiling arrays to detect copy number variants is a related genomewide approach. A third method, based on deep sequencing of entire exomes or genomes, is now taking hold as sequencing prices decline.

Successful genomewide linkage studies (e.g., the one that identified the X-chromosomal location of a risk gene for Brunner’s 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. 'Accordingly, many were left only with 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 genomewide 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's syndrome (9), and of *GABRA2* variation as influencing alcohol dependence risk and related endophenotypes (10)); but also many examples where there was no such identification. This can be attributed in part to insufficiently powered linkage studies, the inapplicability of the common disease/common variant model that underlies traditional linkage, and genetic differences between familial forms of an illness and their nonfamilial forms. The difficulty in identifying a gene out of a linkage peak, together with the expense of recruiting families with multiple affected individuals and the now-easy access to genotyping microarrays for GWAS, has led to a dramatic decline in the use of linkage for complex traits.

## Genomewide association studies (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 the GWAS. The first major GWAS was published in 2005 (11) and identified polymorphic variants associated with age-related macular degeneration. The study included a total of only 146 subjects, and it employed a genotyping microarray that included 106,000 markers that would today be considered unacceptably sparse. In contrast, the GWAS of today more typically employs thousands of samples and millions of markers, imputed (12) as well as directly genotyped. 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." While 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 once thought to be most likely composed of the cumulative risk from a set of common variants. In fact, the more usual result for GWAS has been the identification of risk alleles with odds ratios of 1.2 or less. There are exceptions, but these are fairly rare. Initially, investigators concentrated on variants that met Bonferroni-adjusted criteria for genomewide significance, often taken as $p<5\times10^{-8}$, which is a reasonable threshold to identify individual risk alleles that can reproducibly be shown to be associated to a trait. However, there many other true risk variants among those 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 we can define 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 below.

GWAS has been successful at identifying risk variants for psychiatric traits in most situations where the method has been applied. 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; unfortunately, for adequate power, they 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 genes 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 this, including small sample size, phenotypic heterogeneity, and random chance (16). Another contributor, though, was population stratification – i.e. different ancestral populations often have different allele frequencies at marker loci simply because they are different populations (17, 18). This may have nothing to do with the trait under study, or indeed with any detectable phenotypic trait. At first, this could only be controlled by matching or by using family-controlled designs (such as the transmission-disequilibrium test, or TDT (19)) but the development of statistical methods to control for stratification in samples of unrelated subjects, notably, the 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 principle components methods (23). As much as the technical development of dense genotyping microarrays, these methods are responsible for the development of useful GWAS.

## 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 studies. The GWS results are, so to speak, the tip of the iceberg; but beyond that, since these studies often are at the very limit of adequate power, any given study will identify only a small subset of alleles that truly affect the phenotype under study. If the risk contribution of all identified variants is summed, the result is much less than the predicted heritability of the disorder – the "missing heritability" problem discussed above. Then where is the rest of the genetic risk? Many variants that are true risk variants do not meet GWS. This has been demonstrated in several ways. In a 2010 paper (24), it was demonstrated that the cumulative contribution to risk of a large set of SNPs for a complex trait (height) could greatly increase the amount of heritability explained. This approach has also been applied to psychiatric traits, estimating, e.g., the amount of risk variance for schizophrenia accounted for by common SNPs (25). The amount of risk for schizophrenia accounted for by each chromosome was proportional to its length – evidence

for the large number of common variant risk alleles. The genetic risk score (GRS) approach (26) compiles data from a set of polymorphic markers with a statistical cutoff much less stringent than what is required to determine individual statistical significance. The GRS is valuable in understanding the composition of the genetic risk for a trait, and in demonstrating overlapping genetic contributions to different traits (schizophrenia and bipolar affective disorder, in the above-referenced article).

So far we can account for some of the measured heritability with common SNPs identified as significant via GWAS; and by more through incorporation of the many SNPs that cannot be individually identified as significant (14). This essentially exhausts the possible contribution of common SNP variants acting individually. We can also consider RVs, common variants that are not SNPs, and the interactive effects of different variants (e.g., epistasis) (27). Studies of multiple SNP interactions are limited by the need to control for multiple comparisons -- a statistical issue, and also a computational challenge. We can also look to epigenetic variation (modifications to DNA – such as methylation - that do not affect basepair sequence, but which affect gene expression) (28). Tangible results have come from deep sequencing studies and studies of copy number variation, described below. Epigenetic studies relevant for behavioral traits face the obstacle that for these kinds of studies it would be preferable to study DNA derived from the organ of interest, i.e., brain.

Additional information can be obtained via pathway (29) or network (30) analysis. In pathway analysis, the contributions of possible risk variants at sets of genes that are biologically related are considered to identify associations on a higher-order, presumably functional, level, compared to single SNP associations. This has been applied to several psychiatric traits, for example, schizophrenia (31) and opioid dependence (32). GWAS data can be combined with expression data, to focus on variants that have specific detectable functional effects (33). Using additional biological data together with GWAS or sequence data can create valuable leverage. For example, one study using this approach demonstrated that GWAS-identified schizophrenia risk loci tend to be brain-expressed (34). Gene co-expression network analysis was applied to comparison of brains from subjects with autism and controls (35), first with expression array data and identification of modules of genes that tend to be co-expressed in autism; then combined with autism GWAS data, which associated an expression module (expression [RNA] data) with a set of risk variants (GWAS DNA data). Study of autism co-expression networks at different brain regions and time points considered together with risk genes identified previously (“spatiotemporal convergence”) led to implication of a set of glutamatergic projection neurons in a specific brain region (36). Studies using these kinds of designs have shown great utility and are increasing rapidly, as are statistical methods to incorporate different kinds of annotation data flexibly (37). The use of brain expression data for the study of psychiatric traits is, however, substantially limited by the availability of suitable postmortem material.

## Pharmacogenomics

Pharmacogenomics relates gene variants to differences in drug response; really strong examples relevant to treatment in psychiatry have been lacking until recently. *GADL1* variation was demonstrated to have a strong effect on lithium response in a Chinese sample

(38) – a finding that appears likely to be population-specific, on the basis of population-based allele frequency differences. Kranzler et al (39) showed that a *GRIK1* (the kainate GluK1 receptor subunit gene) SNP predicts clinical response to topiramate in European-American heavy drinkers. Drug response phenotypes are likely to be much less complex that psychiatric diagnosis phenotypes – in some cases a great deal of clinical variation hinges on a single genetic variant, which could be the point of interaction between the drug and a particular biological target. These findings have clear clinical implications for the relevant populations.

Some of the strongest GWAS results relevant to psychiatry come from substance dependence traits, consistent with their pharmacogenetic nature. For example, a region of chromosome 15 containing a cluster of genes encoding nicotinic receptors, has been associated to nicotine dependence and related traits (such as number of cigarettes smoked per day and risk of lung cancer) in numerous studies (e.g., ref. (40–42)). An alcohol dependence GWAS showed association of a functional *ADH1B* variant with trait in European-Americans with $p=1.17\times10^{-31}$ (43). Even in these cases, the identified specific risk alleles account for only a small part of the genetic risk.

## The Psychiatric Genomics Consortium

The Psychiatric Genomics Consortium is a large collaboration conducting meta- or mega-analyses of large GWAS datasets of psychiatric traits – some of their work has been discussed above. They have been highly successful at moving the field forward, especially for schizophrenia and bipolar affective disorder, two initial foci. Use of very large sample sizes – tens of thousands of individuals – has made it possible to map numerous GWS loci down to very low individual odds ratios. But in as much as they combine samples from many research groups incorporating different methods for ascertainment and diagnosis, they have little flexibility with diagnosis definition, needing to settle on a common consensus that can be used for all contributed studies. Also, because they include as many available samples as possible, opportunities for external replication are greatly limited. In 2013, the PGC reported 22 GWS risk loci for schizophrenia (44), which they updated to 108 in 2014(45) as new samples were added to the analysis. Availability of such large samples has led to numerous other findings, including, for example, study of the genetic overlap between disorders for which substantial genetic and phenotypic information is available (46).

## Studies based on DNA sequencing

GWAS microarrays can identify only the single nucleotide variants included on the arrays by design, and certain CNVs. Some genotyping arrays include RVs, but other RVs that are present in very few individuals, or even single individuals because they are new mutations, require sequencing for discovery. Both of these kinds of variation are known to be important for some psychiatric complex traits, but their overall importance cannot yet be determined. RVs, and in some cases CNVs, can be identified via sequencing.

There are two basic sequencing strategies: hypothesis-based sequencing of targeted regions and the broader strategy of covering the whole genome or exome, analogous to GWAS. With targeted sequencing, the study is tethered to prior knowledge, but when a gene is

strongly implicated – e.g. by prior linkage or common-variant associations – it can be productive, as for a post-genetic linkage study that identified RVs at the *CHRNA4* (α4 nicotinic acetylcholine receptor) locus that was protective with respect to nicotine dependence (47). Similar results were reported for RVs at the *CHRNB4* locus (48). Taken individually, RVs rarely provide enough information to demonstrate genetic association. Therefore, various gene-based and binning approaches, where, for example, sets of variants predicted to have functional effects are considered together, are generally employed (49, 50).

Exome sequencing can also be used to discover RVs. Whole exome sequencing (WES) was first used in 2009 for clinical diagnosis (51); it is now in common clinical use. Many proven disease-causing variants reside in exonic sequences -- generally missense or nonsense variants, insertions and deletions, and variants that affect splicing – so WES is a reasonable step to take after GWAS. This approach trades knowledge of non-exomic regions for the increased power of a larger sample. New mutations may be detected readily by sequencing an affected subject and both parents; variants not present in either parent are new mutations. WES of subjects with autism (e.g., refs. (52–55) showed an increased burden of gene variants expected to disrupt protein function in affected individuals. On the other hand, WES has shown that, while moderately-rare variants are unlikely to be very important in schizophrenia (56), new mutations (with multiple *de novo* events detected in affected individuals at several specific loci) may be (7). WES has also been useful in the study of cancer, where somatic mutations specific to tumors can be characterized, e.g., for meningiomas (57). These are valuable for subtyping and selecting treatments.

Sequencing complete genomes is much more data-intensive than exome sequencing. This is especially evident since the ENCODE project (58) demonstrated the importance of intergenic regulatory regions, so that it is apparent that WES alone will not be sufficient to identify all of the important genetic variants. The cost of this method is dropping, and it has seen application for some complex genetic traits. While this method has value in identifying new disease-influencing variants (59), it is still more widely used in studies of somatic DNA.

## Gene-environment interaction; Higher-order genetic variation

Gene-by-environment (GxE) interaction refers to environmental effects on a phenotype that differ depending on the subject's genotype, and is an important factor in determining risk for some psychiatric phenotypes. Whether such effects can be discovered reliably is a subject of debate (60, 61); however some GxE effects have been observed quite consistently. Interaction between variation in *SLC6A4* (the serotonin transporter protein gene) and stress-related phenotypes is a case in point. In a prospective, longitudinal study of a representative birth cohort, Caspi et al. (62)) found that subjects with one or two copies of the "short" allele of a common *SLC6A4* functional polymorphism (which has acquired the trivial name 5-HTTLPR), reported more symptoms related to depression following stressful life events than individuals homozygous for the "long" allele. The "short" allele has been shown to be less functionally active than the "long" under several models (starting with Lesch et al (63)). There have since been numerous studies of GxE effects for this variant in different

populations; Kaufman et al. (21) reported similar results, but in an adolescent population. PTSD is in some ways a model for the study of GxE effects in psychiatric illness, as a specific environmental stressor is required for the emergence of the trait. Further, 5-HTTLPR is a promising candidate for PTSD risk (e.g., refs. (64, 65). More generally, GxE studies of different candidate loci have proliferated, and have apparently started to explain the biological mechanism by which these effects occur (66).

Higher-order genetic variation – copy number variation of comparatively long stretches of DNA – is a major component of human genetic variation (67), and is important for the genetic risk of some psychiatric traits. Both bipolar affective disorder and schizophrenia are characterized by increased *de novo* CNVs (5). Initially CNV studies were done primarily with costly tiling arrays. This method has been supplanted for the time being by interpreting intensity differences in SNP genotyping microarrays. While this is a less reliable method, the lower cost has made it possible to use larger samples. Studies using many key methods are summarized in Table 1.

## Conclusion

Complex trait genetics research has benefited greatly from the introduction of a new set of laboratory and analytic methods over the past several years, most notably GWAS and high-throughput NextGen sequencing. GWAS illustrates a progression in itself – increasing densities of SNPs in available arrays, imputation methods that increase effective analyzable SNP density, and the use of larger samples– both in individual studies, and in meta- and mega-analysis. The result has been a fundamental advance in our understanding of the genetics of complex traits, including behavioral traits, as longstanding theories have become testable and have proven insufficient – for sample, the “common disease-common variant” hypothesis. There is no common genetic architecture for the set of psychiatric traits. The complexity of major psychotic disorders such as schizophrenia and bipolar affective disorder is high, despite their high heritability. Some substance dependence traits, which are basically pharmacogenetic traits, are more straightforward. Thus, relatively important individual risk loci have been identified for alcohol and nicotine dependence. New mutation has been revealed as an important disease mechanism for autism spectrum disorders (52).

GWAS, although less in fashion presently than in the years following its introduction, has identified risk-influencing loci in nearly every case where the method has been applied, albeit sometimes requiring very large samples. This method is an important step in our understanding of the genetic etiology of a complex trait. Many key psychiatric traits still have not yet been subjected to GWAS -- e.g., methamphetamine dependence – and until the GWAS is done, little can be said about the contribution of common variants or indeed about the overall genetic architecture of the trait.

As GWAS of most major complex traits are completed and expanded, some but not all of the heritability of each trait is typically accounted for. What is the next step? Genome or exome sequencing is required to identify many RVs and all new mutations that may individually have a large effect on phenotype (in an individual if not on the population level). Sequencing can also identify other kinds of variants, e.g., CNVs. Whether the

cumulative effect of RVs is more or less important than the cumulative effect of common variants for any individual psychiatric trait is a topic for future research. While the laboratory cost of sequencing is dropping rapidly, data management and analysis costs are falling more slowly, and as such constitute an increasing proportion of the cost of using whole genome sequence data. Combining genetic data from GWAS or sequencing with other biologically important data has been very fruitful in identifying relevant networks and prioritizing possible risk variants.

Beyond GWAS and whole genome sequencing we must consider epigenetics – modifications to DNA other than changes in the nucleotide sequence, which can affect function and be either acquired or transmitted. This is a very important factor in regulating gene expression, and is a key facet in understanding cancer biology. The application of epigenetics to psychiatric traits is limited because optimally it involves study of the tissue of greatest importance for a given trait, and there is a lack of suitable human brain tissue for study. Interesting findings have emerged from peripheral tissues – for example, childhood maltreatment has been shown to impact methylation throughout the genome (68), which could have effects on gene regulation and risk for a variety of illnesses later in life. Use of stem cells may provide a solution to this problem. All in all, there is every reason to expect that in the coming years, complex traits research will be as productive as the highly successful recent past.

## Acknowledgments

Supported in part by National Institutes of Health grants R01 DA12690, R01 DA12849, R01 AA11330, R01 AA017535, and the VA Connecticut MIRECC. Dr. Henry R. Kranzler made helpful comments on the manuscript.

## References

1. Murphy KC, Owen MJ. Velo-cardio-facial syndrome: a model for understanding the genetics and pathogenesis of schizophrenia. The British journal of psychiatry: the journal of mental science. 2001; 179:397–402. [PubMed: 11689394]
2. Shi J, Levinson DF, Duan J, Sanders AR, Zheng Y, Pe'er I, et al. Common variants on chromosome 6p22.1 are associated with schizophrenia. Nature. 2009; 460:753–757. [PubMed: 19571809]
3. Schizophrenia Psychiatric Genome-Wide Association Study C . Genome-wide association study identifies five new schizophrenia loci. Nature genetics. 2011; 43:969–976. [PubMed: 21926974]
4. Ripke S, O'Dushlaine C, Chambert K, Moran JL, Kahler AK, Akterin S, et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nat Genet. 2013; 45:1150–1159. [PubMed: 23974872]
5. Malhotra D, McCarthy S, Michaelson JJ, Vacic V, Burdick KE, Yoon S, et al. High frequencies of de novo CNVs in bipolar disorder and schizophrenia. Neuron. 2011; 72:951–963. [PubMed: 22196331]
6. Malhotra D, Sebat J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. Cell. 2012; 148:1223–1241. [PubMed: 22424231]
7. Xu B, Ionita-Laza I, Roos JL, Boone B, Woodrick S, Sun Y, et al. De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. Nature genetics. 2012; 44:1365–1369. [PubMed: 23042115]
8. Brunner HG, Nelen MR, van Zandvoort P, Abeling NG, van Gennip AH, Wolters EC, et al. X-linked borderline mental retardation with prominent behavioral disturbance: phenotype, genetic localization, and evidence for disturbed monoamine metabolism. American journal of human genetics. 1993; 52:1032–1039. [PubMed: 8503438]

9. Brunner HG, Nelen M, Breakefield XO, Ropers HH, van Oost BA. Abnormal behavior associated with a point mutation in the structural gene for monoamine oxidase A. Science (New York, NY). 1993; 262:578–580.
10. Edenberg HJ, Dick DM, Xuei X, Tian H, Almasy L, Bauer LO, et al. Variations in GABRA2, encoding the alpha 2 subunit of the GABA(A) receptor, are associated with alcohol dependence and with brain oscillations. American journal of human genetics. 2004; 74:705–714. [PubMed: 15024690]
11. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, et al. Complement factor H polymorphism in age-related macular degeneration. Science (New York, NY). 2005; 308:385–389.
12. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS genetics. 2009; 5:e1000529. [PubMed: 19543373]
13. Pritchard JK, Cox NJ. The allelic architecture of human disease genes: common disease-common variant...or not? Human molecular genetics. 2002; 11:2417–2423. [PubMed: 12351577]
14. Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM, et al. Genome partitioning of genetic variation for complex traits using common SNPs. Nature genetics. 2011; 43:519–525. [PubMed: 21552263]
15. Consortium WTCC . Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007; 447:661–678. [PubMed: 17554300]
16. Gelernter J, Goldman D, Risch N. The A1 allele at the D2 dopamine receptor gene and alcoholism. A reappraisal. JAMA: the journal of the American Medical Association. 1993; 269:1673–1677.
17. Gelernter J, Kranzler H, Cubells JF. Serotonin transporter protein (SLC6A4) allele and haplotype frequencies and linkage disequilibria in African- and European-American and Japanese populations and in alcohol-dependent subjects. Human genetics. 1997; 101:243–246. [PubMed: 9402979]
18. Gelernter J, Kranzler H, Cubells JF, Ichinose H, Nagatsu T. DRD2 allele frequencies and linkage disequilibria, including the -141CIns/Del promoter polymorphism, in European-American, African-American, and Japanese subjects. Genomics. 1998; 51:21–26. [PubMed: 9693029]
19. Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). American journal of human genetics. 1993; 52:506–516. [PubMed: 8447318]
20. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. Association mapping in structured populations. American journal of human genetics. 2000; 67:170–181. [PubMed: 10827107]
21. Kaufman J, Yang BZ, Douglas-Palumberi H, Houshyar S, Lipschitz D, Krystal JH, et al. Social supports and serotonin transporter gene moderate depression in maltreated children. Proceedings of the National Academy of Sciences of the United States of America. 2004; 101:17316–17321. [PubMed: 15563601]
22. Bacanu SA, Devlin B, Roeder K. The power of genomic control. American journal of human genetics. 2000; 66:1933–1944. [PubMed: 10801388]
23. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nature genetics. 2006; 38:904–909. [PubMed: 16862161]
24. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nature genetics. 2010; 42:565–569. [PubMed: 20562875]
25. Lee SH, DeCandia TR, Ripke S, Yang J, Sullivan PF, Goddard ME, et al. Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. Nature genetics. 2012; 44:247–250. [PubMed: 22344220]
26. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, et al. International Schizophrenia C. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009; 460:748–752. [PubMed: 19571811]
27. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: Genetic interactions create phantom heritability. Proceedings of the National Academy of Sciences of the United States of America. 2012; 109:1193–1198. [PubMed: 22223662]

28. Weder N, Zhang H, Jensen K, Yang BZ, Simen A, Jackowski A, et al. Child abuse, depression, and methylation in genes involved with stress, neural plasticity, and brain circuitry. Journal of the American Academy of Child and Adolescent Psychiatry. 2014; 53:417–424. e415. [PubMed: 24655651]
29. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity. 2005; 95:221–227. [PubMed: 16077740]
30. Han S, Yang BZ, Kranzler HR, Liu X, Zhao H, Farrer LA, et al. Integrating GWASs and Human Protein Interaction Networks Identifies a Gene Subnetwork Underlying Alcohol Dependence. American journal of human genetics. 2013; 93:1027–1034. [PubMed: 24268660]
31. O'Dushlaine C, Kenny E, Heron E, Donohoe G, Gill M, Morris D, et al. Molecular pathways involved in neuronal cell adhesion and membrane scaffolding contribute to schizophrenia and bipolar disorder susceptibility. Molecular psychiatry. 2011; 16:286–292. [PubMed: 20157312]
32. Gelernter J, Kranzler HR, Sherva R, Koesterer R, Almasy L, Zhao H, et al. Genome-wide association study of opioid dependence: multiple associations mapped to calcium and potassium pathways. Biological psychiatry. 2014; 76:66–74. [PubMed: 24143882]
33. de Jong S, van Eijk KR, Zeegers DW, Strengman E, Janson E, Veldink JH, et al. Expression QTL analysis of top loci from GWAS meta-analysis highlights additional schizophrenia candidate genes. European journal of human genetics: EJHG. 2012; 20:1004–1008. [PubMed: 22433715]
34. Richards AL, Jones L, Moskvina V, Kirov G, Gejman PV, Levinson DF, et al. Schizophrenia susceptibility alleles are enriched for alleles that affect gene expression in adult human brain. Molecular psychiatry. 2012; 17:193–201. [PubMed: 21339752]
35. Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, et al. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature. 2011; 474:380–384. [PubMed: 21614001]
36. Willsey AJ, Sanders SJ, Li M, Dong S, Tebbenkamp AT, Muhle RA, et al. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell. 2013; 155:997–1007. [PubMed: 24267886]
37. Chung DYC, Li C, Gelernter J, Zhao H. GPA: A statistical approach to prioritizing GWAS results by integrating pleiotropy information and annotation data. 2014 submitted ms.
38. Chen CH, Lee CS, Lee MT, Ouyang WC, Chen CC, Chong MY, et al. Variant GADL1 and response to lithium therapy in bipolar I disorder. The New England journal of medicine. 2014; 370:119–128. [PubMed: 24369049]
39. Kranzler HR, Covault J, Feinn R, Armeli S, Tennen H, Arias AJ, et al. Topiramate treatment for heavy drinkers: moderation by a GRIK1 polymorphism. The American journal of psychiatry. 2014; 171:445–452. [PubMed: 24525690]
40. Thorgeirsson TE, Gudbjartsson DF, Surakka I, Vink JM, Amin N, Geller F, et al. Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. Nature genetics. 2010; 42:448–453. [PubMed: 20418888]
41. Tobacco Genetics C . Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nature genetics. 2010; 42:441–447. [PubMed: 20418890]
42. Liu JZ, Tozzi F, Waterworth DM, Pillai SG, Muglia P, Middleton L, et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. Nature genetics. 2010; 42:436–440. [PubMed: 20418889]
43. Gelernter J, Kranzler HR, Sherva R, Almasy L, Koesterer R, Smith AH, et al. Genome-wide association study of alcohol dependence: significant findings in African- and European-Americans including novel risk loci. Molecular psychiatry. 2014; 19:41–49. [PubMed: 24166409]
44. Ripke S, O'Dushlaine C, Chambert K, Moran JL, Kahler AK, Akterin S, et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nature genetics. 2013; 45:1150–1159. [PubMed: 23974872]
45. Schizophrenia Working Group of the Psychiatric Genomics C . Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014; 511:421–427. [PubMed: 25056061]
46. Smoller JW, Craddock N, Kendler K, Lee PH, Neale BM, et al. Cross-Disorder Group of the Psychiatric Genomics C. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet. 2013; 381:1371–1379. [PubMed: 23453885]

47. Xie P, Kranzler HR, Krauthammer M, Cosgrove KP, Oslin D, Anton RF, et al. Rare nonsynonymous variants in alpha-4 nicotinic acetylcholine receptor gene protect against nicotine dependence. Biological psychiatry. 2011; 70:528–536. [PubMed: 21683344]
48. Haller G, Druley T, Vallania FL, Mitra RD, Li P, Akk G, et al. Rare missense variants in CHRNB4 are associated with reduced risk of nicotine dependence. Human molecular genetics. 2012; 21:647–655. [PubMed: 22042774]
49. Sham PC, Purcell SM. Statistical power and significance testing in large-scale genetic studies. Nature reviews Genetics. 2014; 15:335–346.
50. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-Variant Association Analysis: Study Designs and Statistical Tests. American journal of human genetics. 2014; 95:5–23. [PubMed: 24995866]
51. Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106:19096–19101. [PubMed: 19861545]
52. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature. 2012; 485:237–241. [PubMed: 22495306]
53. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature. 2012; 485:242–245. [PubMed: 22495311]
54. O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nature genetics. 2011; 43:585–589. [PubMed: 21572417]
55. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature. 2012; 485:246–250. [PubMed: 22495309]
56. Need AC, McEvoy JP, Gennarelli M, Heinzen EL, Ge D, Maia JM, et al. Exome sequencing followed by large-scale genotyping suggests a limited role for moderately rare risk factors of strong effect in schizophrenia. American journal of human genetics. 2012; 91:303–312. [PubMed: 22863191]
57. Clark VE, Erson-Omay EZ, Serin A, Yin J, Cotney J, Ozduman K, et al. Genomic analysis of non-NF2 meningiomas reveals mutations in TRAF7, KLF4, AKT1, and SMO. Science (New York, NY). 2013; 339:1077–1080.
58. Consortium EP, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489:57–74. [PubMed: 22955616]
59. Gudmundsson J, Sulem P, Gudbjartsson DF, Masson G, Agnarsson BA, Benediktsdottir KR, et al. A study based on whole-genome sequencing yields a rare variant at 8q24 associated with prostate cancer. Nature genetics. 2012; 44:1326–1329. [PubMed: 23104005]
60. Kendler KS. The stress of internship and interactions with stress. Archives of general psychiatry. 2010; 67:566–567. discussion 568–569. [PubMed: 20530005]
61. Keller MC. Gene x environment interaction studies have not properly controlled for potential confounders: the problem and the (simple) solution. Biological psychiatry. 2014; 75:18–24. [PubMed: 24135711]
62. Caspi A, Sugden K, Moffitt TE, Taylor A, Craig IW, Harrington H, et al. Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. Science (New York, NY). 2003; 301:386–389.
63. Lesch KP, Bengel D, Heils A, Sabol SZ, Greenberg BD, Petri S, et al. Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. Science (New York, NY). 1996; 274:1527–1531.
64. Xie P, Kranzler HR, Farrer L, Gelernter J. Serotonin transporter 5-HTTLPR genotype moderates the effects of childhood adversity on posttraumatic stress disorder risk: a replication study. American journal of medical genetics Part B, Neuropsychiatric genetics: the official publication of the International Society of Psychiatric Genetics. 2012; 159B:644–652.
65. Xie P, Kranzler HR, Poling J, Stein MB, Anton RF, Brady K, et al. Interactive effect of stressful life events and the serotonin transporter 5-HTTLPR genotype on posttraumatic stress disorder

diagnosis in 2 independent populations. Archives of general psychiatry. 2009; 66:1201–1209. [PubMed: 19884608]

66. Klengel T, Mehta D, Anacker C, Rex-Haffner M, Pruessner JC, Pariante CM, et al. Allele-specific FKBP5 DNA demethylation mediates gene-childhood trauma interactions. Nat Neurosci. 2013; 16:33–41. [PubMed: 23201972]
67. Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, et al. Paired-end mapping reveals extensive structural variation in the human genome. Science (New York, NY). 2007; 318:420–426.
68. Yang BZ, Zhang H, Ge W, Weder N, Douglas-Palumberi H, Perepletchikova F, et al. Child abuse and epigenetic mechanisms of disease risk. American journal of preventive medicine. 2013; 44:101–107. [PubMed: 23332324]

## Table 1

Studies illustrating key methods, mostly pertaining to psychiatric traits

| Study | Year | Phenotype | Design | Key findings |
|---|---|---|---|---|
| Brunner et al. (8) | 1993 | Violence, cognitive deficits | Genetic linkage | Risk gene mapped to region of *MAOA* and *MAOB* genes on the X chromosome |
| Pritchard et al.(20) | 2000 | n/a | Structured association | SNP marker sets can be used to ascertain and correct for population structure |
| Edenberg et al.(10) | 2004 | Alcohol dependence, EEG | Gene identification from linkage peak | *GABRA2* alleles associated to alcohol dependence |
| Caspi et al.(62) | 2003 | Depression (symptoms) | GxE | *SLC6A4* allele predisposed to depression in the presence of environmental stressors |
| Gelernter et al.(43) | 2014 | Alcohol dependence (symptom count) | GWAS | Association with *ADH1B* and other alcohol-metabolizing enzyme loci |
| Yang et al.(24) | 2010 | Height | Use of large sets of SNPs to predict risk (“Visscher analysis”) | 45% of height variant explained by large set of GWAS genotypes |
| Shi et al.(2) | 2009 | Schizophrenia | Large consortium GWAS | Chromosome 6p22.1 markers associated with schizophrenia |
| O’Roak et al.(55) | 2012 | Autism | Exome sequencing of family trios | De novo risk mutations, many or which map to β-catenin/chromatin remodeling protein network, associated with autism risk |