Published in final edited form as:

Am J Psychiatry. 2009 May ; 166(5): 540–556. doi:10.1176/appi.ajp.2008.08091354.

Genomewide association studies: History, rationale and prospects for psychiatric disorders

The Psychiatric GWAS Consortium

Department of Psychiatry and Behavioral Sciences, Stanford University, Palo Alto, Calif.

Abstract

Objective—We review the history and empirical basis of genomewide association studies (GWAS), the rationale for GWAS of psychiatric disorders, results to date, limitations, and plans for GWAS meta-analyses.

Method—Literature review, power analysis, discussion of issues and description of planned studies.

Results—Most of the genomic DNA sequence differences between any two people are common (frequency > 5%) single nucleotide polymorphisms (SNPs). Because of localized patterns of correlation (linkage disequilibrium), 500,000–1,000,000 of these SNPs can test the hypothesis that one or more common variants explain part of the genetic risk for a disease. GWAS technologies can also detect some of the copy number variants (CNVs; deletions and duplications) in the genome. Systematic study of rare variants will require large-scale resequencing studies. GWAS methods have detected a remarkable number of robust genetic associations for dozens of common diseases and traits, leading to new pathophysiological hypotheses, although only small proportions of genetic variance have been explained so far, and therapeutic applications will require substantial further effort. Study design issues, power and limitations are discussed. For psychiatric disorders, there are initial significant findings for common SNPs and rare CNVs. Many other studies are in progress.

Conclusion—GWAS of large samples have detected associations of common SNPs and of rare CNVs to psychiatric disorders. More findings are likely -- larger GWAS samples detect larger numbers of common susceptibility variants (with smaller effects). The Psychiatric GWAS Consortium (of 110 researchers from 54 institutions) is carrying out GWAS meta-analyses for schizophrenia, bipolar disorder, major depressive disorder, autism and attention deficit hyperactivity disorder. Based on results for other diseases, larger samples will be required. The contribution of GWAS will depend on the true genetic architecture of each disorder.

Keywords

Review; genome-wide association study; meta-analysis; attention deficit hyperactivity disorder; autism; bipolar disorder; major depressive disorder; schizophrenia

Introduction

Since 2005 (1), genomewide association studies (GWAS, “j ’ w s”) have produced strongly significant evidence that specific common DNA sequence differences among people influence their genetic susceptibility to over 40 different common diseases. (2). Many of

these findings implicate previously-unsuspected candidate genes and new pathophysiological hypotheses. The method is feasible because millions of human DNA sequence variations have been catalogued, and new technologies developed that can assay over one million variants rapidly and accurately. The first GWAS reports have appeared for psychiatric disorders, and close to 50 GWAS of attention-deficit hyperactivity disorder, autism, bipolar disorder, major depressive disorder and schizophrenia should be completed by the end of 2008, with more to come. The present authors have formed an international consortium of psychiatric GWAS investigators to carry out rapid meta-analyses of these five disorders to maximize power. Here we describe GWAS methods, their rationale and current results for non-psychiatric and psychiatric disorders, and discuss some limitations and uncertainties.

Candidate genes, linkage and linkage disequilibrium

Genetic epidemiology

Before any molecular genetic study is undertaken, the methods of *genetic epidemiology* are used to identify a *phenotype* (observable disease or trait) that is at least partially *heritable*. An introduction to these methods is available online (<http://www.dorak.info/epi/genetepi.html>). Briefly, twin, family and population-based studies are used to estimate heritability, define the most heritable phenotype, and explore interactions between genetic and environmental factors. The current diagnostic definitions of major psychiatric disorders are based in part on twin and family data. Epidemiological data are also critical for defining appropriate control groups for molecular studies. The data for psychiatric disorders suggest that most of the heritable risk is due to *interactions* of combinations of genetic risk variants, each with a relatively small effect on risk.

Candidate genes

When the pathophysiology of a disease is known (e.g., an enzyme deficiency), it may be straightforward to define *candidate genes* and to determine which DNA sequence variants predict who becomes ill. For psychiatric disorders, pathophysiologies are unknown. Most candidate gene hypotheses are based on the effects of psychiatric medications on monoamine neurotransmission, focusing particularly on several *functional polymorphisms* in dopaminergic or serotonergic pathways (i.e., sequence variants that alter relevant receptor proteins or enzymes). (3, 4) None has been shown to be associated with a psychiatric disorder with a level of significance that would lead to general acceptance of a finding.

Positional methods

The alternative strategy is to localize disease-related sequence variation based entirely on its location or *position* in the genome. Before GWAS, available methods included the *genomewide linkage study* (GWLS) and *linkage disequilibrium (LD) mapping* (of which GWAS is a large-scale example). (See Table 1 for definitions, and Table 2 for a timeline of critical developments.)

GWLS became feasible in the 1980s with genomewide “maps” (7) of hundreds of DNA sequence variations (*markers*). *Linkage analysis* (reviewed in (15)), of families with multiple ill members, exploits *within-family correlations* between illness and the alternative sequences (alleles) of the markers that are closest to the disease-related gene(s). Linkage studies led to the discovery of (mostly rare dominant or recessive) mutations for more than 1,600 diseases (Online Mendelian Inheritance in Man, <http://www.ncbi.nlm.nih.gov/Omim/mimstats.html>). They have been less successful for complex (multifactorial/multigenic) disorders. In psychiatric linkage studies (catalogued at <https://slep.unc.edu>), small samples of pedigrees were initially studied in the hope of discovering simpler genetic mechanisms

that would provide clues to pathophysiology. Then, larger studies (hundreds of families) searched for genes with smaller effects. There are diverse opinions about these studies' past success and future prospects. Statistically significant linkages have been reported but have been difficult to replicate, presumably because linkage is much less powerful when risk variants have small effects and there is *heterogeneity* in the underlying genetic factors in different families. Meta-analyses have supported linkage for some disorders. (16-18)

LD mapping relies instead on the *population-wide correlation* between two sequence variants. Most variants are *single nucleotide polymorphisms* (SNPs) (almost always just two alternative nucleic acids at a genomic position). SNP variants that are reasonably common are mutations that occurred thousands of generations ago and then spread, due to chance or natural selection. When a second SNP mutation occurred very close to an earlier one (up to tens of thousands of base pairs [bp] away), then both variant alleles are almost always transmitted to the same children in subsequent generations. *Linkage disequilibrium* is this non-random association of two alleles. Around 20 years ago, it was proposed that LD could be exploited to "map" or identify disease genes, such as in linkage candidate regions (or in recently isolated populations in which LD spans long distances). (19) If one SNP increases the risk of a common disease, then there will be a statistical association in the population between disease and that SNP (*direct association*) and several nearby SNPs (*indirect association*, due to LD).

LD mapping studies have identified plausible *positional candidate genes* in regions of linkage or of cytogenetic abnormalities associated with psychiatric disorders, and these genes have suggested new mechanistic hypotheses. (20) For example, as of April, 2008, there were 1291 published studies of 690 schizophrenia candidate genes (see <http://www.schizophreniaforum.org/res/sczgene/default.asp>). A recent meta-analysis of these studies (3) identified four "strong" psychiatric candidate gene associations based on epidemiological criteria for meta-analysis, but not at what is currently understood to be a genomewide level of statistical significance (see below).

Common SNPs, HapMap and GWAS

Risch and Merikangas (21) noted that small genetic effects could be detected with greater power by association analyses, and proposed that genomewide LD mapping (GWAS) could be applied if technologies were developed to study SNP frequencies in all genes, contrasting in ill cases vs. control subjects, or cases and their parents (associated alleles are transmitted to ill offspring more often than expected by chance). Lander (9) proposed the *common disease common variant (CDCV)* hypothesis. Comparing any two people, most sequence differences are ancient, "common" SNPs (by convention, varying on at least 5% of chromosomes in a population), which Lander argued must confer at least *some* (not all) of the genetic risk for common diseases. He proposed cataloguing them and studying their association to disease in large samples. SNPs become common because they are neutral or favorable with respect to survival (e.g., evolutionary pressures can rapidly increase frequencies of adaptive SNPs in gene-regulating regions). But some have mildly harmful effects, perhaps depending on environmental conditions (e.g., preserving fat during an ice age but leading to obesity in the fast food era). The CDCV-GWAS strategy assumed that many different common SNPs have small effects on each disease, and that some could be found by testing enough SNPs in enough people.

How many SNPs should be tested? Studies of small regions revealed *LD blocks* within which common SNPs are highly correlated (usually less than 10-30,000 bp in Africans, or 30-50,000 in the newer European or Asian populations).(22) This motivated the *HapMap project* (www.hapmap.org) (12), which has validated around 4 million SNPs including 2.8

million of the estimated 10 million common SNPs in major world populations, while creating competition among biotechnology companies to develop high-throughput genotyping technologies. Sequencing and genotyping studies showed that sets of 500,000 (Europeans) to 1,000,000 (Africans) SNPs could “tag” (serve as proxies for) around 80% of common SNPs. (23) Over the last three years, the Affymetrix and Illumina companies have developed “chips” (arrays of assays on glass slides) that assay large SNP sets with high accuracy (0-2% missing data, less than 0.5% errors), at low cost (around US\$500 per subject, around a 2000-fold reduction in cost per genotype in ten years) and rapidly (over 1,000 DNA specimens per week in some labs). The GWAS era has arrived.

Rare SNPs

Common SNPs are unlikely to explain all of the genetic risk for common disorders. An evolutionary model of complex diseases (24) predicts roles for common SNPs and for *multiple rare variants* (such as SNPs) in some genes (*MRV hypothesis*). A rare variant is usually defined by a frequency below 1%, although many are so rare that they are found in only one individual in a sample. (25) Most variants carried by any one person are common SNPs, but if one sequences a chromosomal region in many people, one finds more and more rare SNP sites. The most *deleterious* variants die out or remain rare due to *natural selection*, i.e., they reduce survival. They are found in functional regions, i.e., among the SNPs in exons (protein coding regions) that alter amino acid sequence (*non-synonymous* or nsSNPs), or in promoters (sequences that regulate gene expression). (26, 27) But there are other, poorly-understood functional regions. Many non-coding regions are highly *conserved* across species, suggesting that they have a function. Gene expression can be altered by common, synonymous exonic SNPs (no coding change), and by SNPs in introns (non-coding gene segments). (28) Indeed, most genomic DNA is apparently transcribed into RNA and thus could have unknown regulatory functions. (29) Most rare SNP associations will be missed by current GWAS methods, but it is expected that the 1000 Genomes Project (www.1000genomes.org) will discover most SNPs with 1-5% frequencies, which would permit an extension of GWAS methods into that range. Linkage could detect a locus with rare pathogenic variants in many families.

Rare SNP associations are more likely to be detected by *resequencing* of relevant regions in hundreds or thousands of individuals (by convention, *resequencing*, sometimes now called “medical sequencing,” determines an individual’s DNA sequence, vs. *sequencing* of an organism’s genome). Botstein and Risch (30) encouraged systematic study of nsSNPs in common diseases. Multiple rare pathogenic variants have been discovered by resequencing genes influencing lipid metabolism (31) and hypertension (32), and also genes in which GWAS had already detected common-SNP associations. (33-35) It is anticipated that advances in resequencing technologies will make it feasible to search systematically for rare variant effects in parts of the genome (e.g., linkage regions, all exons, all promoters) and eventually genomewide.

Copy number variants

GWAS technologies can also detect more of the *copy number variants* (CNVs) in the genome than was possible with older cytogenetic methods, by analysis of the relative intensities of the fluorescent labels used in the assays. CNVs are deletions and duplications of DNA segments, of diverse sizes and population frequencies. For example, large deletions on chromosome 22q11 cause the velocardiofacial/DiGeorge syndrome, and 20% of such cases also develop schizophrenia. (36) CNVs tend to arise in regions with repetitive DNA sequences. Some CNVs are common and are transmitted from generation to generation, while others recurrently arise *de novo*. Like rare SNPs, rare CNVs are more likely to be

harmful. (Other structural variants such as inversions and translocations remain difficult to detect.) Large genomewide CNV scans show that CNVs are more common than was previously recognized. (37) Structural variation has not been as comprehensively studied as SNPs, because CNV detection is less accurate, biological confirmation is still costly, and smaller CNVs (less than 100,000 base pairs) are less reliably detected. But technologies are rapidly improving. Significant CNV findings are now being reported for psychiatric disorders as discussed below.

GWAS study design

Study design issues are summarized in Table 3. A GWAS sample, selected based on a well-defined, heritable phenotype, might include case (ill) and control subjects, subjects with a range of values for a continuous phenotypic variable, or probands and both of their parents (trios) or other constellations of relatives. Samples are often limited to a single ancestry (European, Asian, etc.), because some SNPs have markedly different frequencies across populations (and some are not observed in every population), so that some associations can best be detected in homogeneous samples. Each subject is genotyped using a GWAS SNP array. Extensive “quality control” (data cleaning) is required to detect problems that can result in false negative or false positive findings, such as SNPs and DNA specimens that gave poor quality results, or unexpected relatedness among subjects. Case-control differences in ancestry (“population substructure”) can also confound association test results, but this can be corrected statistically based on correlations among SNP genotypes that reflect ancestry. (38) Most studies then test each SNP for association of genotypes to the phenotype, and *impute* the genotypes of other HapMap SNPs, based on the correlations among SNPs in HapMap data. (39-41)

Selection of control groups is critical, beyond the problem of ancestral matching. It is ideal to recruit cases and controls systematically from the same population. This is not always feasible for very large samples of a clinically severe disorder, but controls must be sufficiently comparable to cases to avoid systematic biases. Depending on the phenotype, it might be important to match for such variables as age (e.g., for an Alzheimer’s study) or sex. Information about known gene-environment interactions should be considered, e.g., in studies of substance dependence, controls are usually selected who have used the substance but did not become dependent. When the phenotype is relatively uncommon (e.g., 5% prevalence), little power is lost by studying controls without clinical screening, but for more common disorders, power is increased if ill individuals are excluded from the control group. (40) It is reassuring that in the UK Wellcome Trust Case Control Consortium (WTCCC) GWAS of seven common diseases, robust results were obtained when association was tested using control groups recruited from blood donors or from a population-based birth cohort.

Statistical power of GWAS

A key factor in the recent success of GWAS has been the assembling of large samples with adequate statistical power to detect small effects of common SNPs on disease risks.

Figure 1 illustrates why. The figure legend discusses factors that predict power: sample size, correction for testing many SNPs, population frequency of the risk allele, and its *genotypic relative risk* (GRR). Large GRRs (e.g., 5-10-fold increase in risk to carriers) would have produced large linkage signals. Early GWAS analyses with a few hundred cases were powered to search for risk alleles with GRRs above 2. Only a few such effects were detected. (1) The more typical GWAS has included 1,000-2,000 cases plus a similar number of controls, with power to detect risk alleles that are reasonably common and have GRRs of 1.5-2. The small number of robust findings suggested the need to detect smaller GRRs. (2)

This led to much larger GWAS analyses in collaborative samples, which has proven remarkably successful for many diseases. As discussed in the next section, most of the new, highly significant findings have been for alleles with GRRs of 1.1-1.4, mostly between 1.12-1.20. In this range (Figure 1), good or excellent power requires samples of 8,000-20,000 cases (plus controls), depending on GRR and allele frequency – i.e., larger than any sample collected by a single research group to date.

GWAS findings for non-psychiatric disorders and lessons for psychiatry

Over the past three years, many highly significant GWAS findings have been reported for non-psychiatric disorders. Table 4 summarizes a systematic listing of GWAS findings <http://www.genome.gov/GWASstudies/> (accessed November 15, 2008) provided by the National Institute for Human Genome Research, restricted to findings with p-values less than 5×10^{-8} (42-44). This choice of threshold, and alternatives to it, are discussed in the Table 4 legend. There are 200 distinct findings listed for 59 disorders or traits. Some may be false positives due to chance (every p-value is an estimate of the probability of a false positive result) or to technical problems such as genotyping or analytic errors. But many of these findings have already been replicated in independent samples, and most robust p-values do replicate. These results far exceed all previous robust associations for complex disorders. This confirms that common SNPs explain part of the genetic risk for these disorders, as predicted by the CDCV hypothesis. There are almost certainly also many common SNPs with smaller effects on risk, as well as rare and very rare SNPs and CNVs with diverse effect sizes.

Sample size

Most initial GWAS samples included 500-3,000 cases (plus controls), or as high as 10,657 subjects for a continuous trait. One or more replication samples were usually then studied via collaboration, totaling 2,000-8,000 subjects (cases and controls, or family members). For studies with at least 1,000 cases, most findings involved common alleles (20-80%) with *odds ratios* (ORs, estimates of GRR) between 1.1-1.4, i.e., the range within which there was some power.

Findings for type 2 diabetes (T2D) illustrate the importance of sample size. In late 2007, there were 11 strong candidate genes: 6 discovered by GWAS, 4 based on mechanistic hypotheses, and 1 (TCF7L2) by LD mapping of a linkage region (although TCF7L2 SNPs did not explain the linkage). (47) TCF7L2 has an overall OR of 1.37; it was detected by most (not all) studies. Other T2D loci have allelic ORs between 1.1 and 1.2, requiring from 10,000 to well over 20,000 total subjects for 80% power; each locus was missed by most single studies. For example, in the WTCCC study (2,000 cases, 3,000 controls), these 11 SNPs were ranked from 2 to 26,017 in their strength of association. (47) Zeggini et al. combined over 60,000 subjects to study T2D findings that had not quite reached genomewide significance previously; 6 SNPs (implicating eight different genes) now achieved $p < 5 \times 10^{-8}$, with ORs from 1.09-1.15. (48)

Novel etiologic hypotheses

Most findings have implicated novel genes or regions and suggested new mechanisms. For example, SNPs in FTO (“fat mass and obesity associated” gene) are strongly associated with common obesity. (49, 50) This was surprising, because FTO knockout mice are not obese. Mechanisms are under study, including a role in adipocyte lipolysis. (51) As Todd has noted (52), implicating a gene in disease requires both compelling statistical evidence for association and substantial additional biological evidence.

Insights into phenotypes

FTO also exemplifies the importance of phenotypic variables. T2D is common in obese individuals. FTO SNPs are associated with T2D, but this is due to the association of T2D and body mass index (BMI). (50) The association of FTO with T2D disappears if T2D cases and controls are matched for BMI. (53) Surprising relationships among phenotypes have also been discovered. For example, SNPs on chromosome 8q24.21 are associated with prostate, breast and colorectal cancer, which were not previously thought to be genetically related. (54) The region contains no known genes, so that without a GWAS strategy, it would have been ignored. It is now being intensively studied.

Thus, GWAS has been remarkably successful for many common diseases. Large multicenter samples have usually been required, and larger samples have detected more associations. Only a small part of the genetic risk for any one disease has been explained, but these discoveries have suggested new disease mechanisms and targets for therapy and prevention, although direct therapeutic applications will require substantial additional effort to characterize the biological mechanisms and develop new treatments. Some of the unexplained variance is likely to be due to other common SNPs (those that have smaller effects than can be detected with current sample sizes, or that are not tagged by the arrays, or were missed because of technical or sampling problems). The remaining variance may be due to rare SNPs, CNVs, other unsuspected genomic mechanisms, gene-gene or gene-environment interactions that have not been adequately modeled, and epigenetic effects. The results suggest that the largest possible samples should be studied by GWAS for each of the major psychiatric disorders, to test the hypothesis that common SNPs or detectable CNVs are involved in etiology. Positive findings could lead to important etiologic discoveries.

GWAS of psychiatric disorders

GWAS findings are now emerging for psychiatric disorders (Table 5). The early findings include replicated CNV associations for schizophrenia and for autism, a genome-wide significant association for bipolar disorder that emerged when several datasets were combined, and a significant association in a combined schizophrenia-bipolar dataset.

For schizophrenia, four genome-wide studies of CNVs (55-58) have produced two types of replicated findings:

First, two large studies found two rare deletions that are significantly associated with schizophrenia, on chromosomes 1q21.1 (0.2% of cases) and 15q13.3 (0.3%). (55, 56) The case:control ratios (around 10) suggest major effects on risk, but it is unknown which deleted genes or sequences are responsible, or whether they account for all of the subject's genetic risk. These deletions are also seen (but probably less frequently) in individuals with mental retardation and/or autism, and are typically *de novo* (not inherited from parents). (55) The well-known chromosome 22q11 deletions were also significantly associated with schizophrenia (0.2-0.4% of cases across studies vs. 0% of controls).

Second, the three studies that tested such a hypothesis (56-58) showed that schizophrenia cases have a small but significant increase in their total genome-wide count of rare, long CNVs, suggesting that other pathogenic CNVs exist which are so rare that they are difficult to detect singly.

Three small schizophrenia GWAS (178-738 cases) have tested association to SNPs using individual genotyping, (61-63) and two others (69, 70) used pooled genotyping (not included in Table 3). No genome-wide significant finding has emerged yet for schizophrenia alone, but when the 12 "best" SNPs from a GWAS of 479 cases and 2,937 WTCCC controls

were genotyped in an additional 7,308 schizophrenia cases and 12,834 controls, and the 1,868 WTCCC bipolar disorder cases were added to the analysis, a genomewide significant p-value was seen for a SNP in a gene of unknown function (ZNF804A, zinc finger protein 804A). (63) This will require replication in these disorders both separately and combined. It illustrates the potential importance of cross-diagnosis analyses, although these will increase the problem of multiple testing and thus require very large samples for confirmation.

For autism, three studies have reported association with a rare (1% of cases), large, high-penetrance deletion on chromosome 16p11.2. (65-67) There is also support for the hypothesis of an excess of rare, mostly *de novo* CNVs in around 10% of cases, although their role in autism remains to be proven. (64, 65, 68) Autism GWAS of common SNPs have yet to be reported.

For bipolar disorder, three individual studies (with 1,000-2,000 cases each) failed to detect significant association, but the three datasets combined produced a p-value of 9.1×10^{-9} in ANK3 (ankyrin-G, whose product links membrane proteins such as voltage-dependent sodium channels to the axonal cytoskeleton). (41, 59, 60) A significant association (in DGKH) reported in a smaller study using pooled genotyping was not seen in the larger analysis. (71)

Among the reports that will appear in the near future are the four psychiatric GWAS supported by the Genetic Association Information Network (GAIN, fnih.org) for schizophrenia, bipolar disorder, major depression and ADHD. Details and preliminary results are available online (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap>); we are not permitted to summarize them pending the initial publications by each group of investigators. GAIN is an example of a new emphasis on rapid public sharing of genetic data to accelerate the process of discovery.

The Psychiatric GWAS Consortium

The first set of psychiatric GWAS analyses have demonstrated that this methodology can work for psychiatric disorders. The pattern observed in the bipolar disorder studies is particularly encouraging because it is consistent with what has happened for non-psychiatric diseases: combining several smaller samples produced a significant result, as well as several other findings with modestly significant p-values in each individual study which could prove to be significant as more data become available. (60)

These results support our expectation that multiple definitive association findings will be detected for many psychiatric disorders, often requiring large samples. We therefore organized the Psychiatric GWAS Consortium which includes almost all known GWAS studies to date for SCZ, BD, MDD, ADHD and AUT, contributed by 110 investigators at 54 institutions around the world (Table 6). The PGC has three specific aims:

1. **Within-disorder meta-analyses** of all available GWAS data. These diagnoses are based on definitions that produced maximum heritability estimates in genetic-epidemiological studies. Thus, disorder-specific analyses represent our strongest hypotheses.
2. **Cross-disorder analyses**, including analyses of combinations of disorders and of phenotypes observed in two or more disorders (such as depression or psychosis), based on the recommendations of an expert committee. Because data are insufficient to determine what common, cross-disorder etiologic factors might exist, alternative phenotypes should be explored. GWAS have produced surprising cross-disorder associations such as those for cancers (54) and for inflammatory

bowel diseases (72), which could also exist for psychiatric disorders given the many common symptoms.

3. **Analyses of comorbidities** such as alcohol, nicotine and illicit drug use disorders, which can be studied across multiple case groups.

Additional exploratory analyses will be carried out by analysts from participating research groups, generating new hypotheses that can be tested as more samples become available. All GWAS data used by PGC (unless prohibited by the original consents or IRB decisions) will become available to the scientific community through data repositories.

A central analytic team, in consultation with participating analysts, will carry out uniform QC analyses and imputation of untyped HapMap SNPs (to permit combining of data). The disorder-specific workgroups will design their own primary meta-analyses, with additional workgroups to define other phenotypic and cross-disorder analyses. Analyses will account for ethnic substructure within samples and appropriate pairing of case and control groups.

Depending on the genetic architecture of each disorder, one or more primary analyses could have sufficient power to detect genomewide significant evidence for association. For example, the largest analyses, with approximately 10,000 cases and 10,000 controls, would have 80% power to detect a SNP with a GRR of 1.152 with $p < 5 \times 10^{-8}$, assuming direct association with an allele with a frequency of 0.25, and log-additive inheritance, or 57% power for indirect association with an r^2 of 0.8. Power would be reduced for smaller samples or for less common alleles or recessive effects. Note that if there are many risk alleles in the genome with a sufficient effect size, there would be substantial power to detect at least one of them. We expect to complete interim meta-analyses during 2008 and final analyses within 2009. Updated results will be posted on the PGC website (<http://pgc.unc.edu>).

Discussion

There is a compelling rationale for applying GWAS methods to very large samples for major psychiatric disorders. Given that the pathophysiologies of these disorders are unknown, genomewide studies provide an unbiased way to search the genome for causative factors. Many successful GWAS analyses have combined data from diverse clinical samples and SNP arrays to obtain replicable findings that point to new hypotheses about disease mechanisms and treatment targets. The first significant psychiatric GWAS findings have been reported (Table 5), using large collaborative samples. It is hoped that meta-analyses can produce multiple robust findings for psychiatric disorders.

GWAS SNP arrays “cover” 80% or more of common HapMap SNPs, and regional resequencing data suggest that most unknown common SNPs are also being tested indirectly. Within these limitations, GWAS methods test the CDCV hypothesis. CNVs are also detected, but less systematically or accurately. The PGC meta-analyses will have reasonable power to detect common SNP associations for each disorder within the limitations shown in Figure 1. But it is possible that very few significant associations might be detected for some disorders, or none. How far should we go with GWAS?

Past experience suggests that for some disorders, as many as 20,000-30,000 cases and a similar number of controls (or case-parent trios) could be required to obtain highly robust findings. More datasets will be genotyped in the near future, and NIMH plans to collect additional large schizophrenia and bipolar disorder samples (<http://grants.nih.gov/grants/guide/rfa-files/RFA-MH-08-131.html>). This raises important questions of resource allocation. For example, the next phase of genetic studies will involve a combination of increasingly large GWAS analyses (for common SNP and CNV associations) and

resequencing studies (for rare variants). It is not known how these and other research investments should be optimally balanced.

To the extent that resources are available, we encourage a long-term view, avoiding the well-known pattern of initial exuberance followed by disillusionment. The logic of GWAS has been clear for over ten years. (23) Results have been remarkably consistent with expectations, in the sense that common SNP associations have been discovered for many common disorders, particularly those that have been studied with larger sample sizes. It is true that initial GWAS results have explained only a small part of the etiologic variance for each disease, and it seems certain that studies of CNVs and rare SNPs will also be critical in elucidating disease mechanisms. But it is likely that common SNPs explain a larger portion of the variance than can be determined with existing sample sizes, with many common SNPs, each with small effects, contributing collectively to a major portion of genetic risk (24). As the number of associations increases, the biological pathways underlying risk for each disease become more clear. GWAS methods should be applied systematically to major psychiatric disorders in large samples.

There are many important *caveats*, some of which we note here:

1. Some disorders might not be amenable to GWAS, e.g., if all risk alleles have very low GRRs; or if genetic risks are conferred by multiple rare SNPs or by CNVs too small to be detected reliably. Discoveries for these disorders might only be possible with larger-scale resequencing studies.
2. Current diagnostic categories might be inadequate. *Endophenotypic* variables (neuroimaging, electrophysiological, neuropsychological, biochemical or other markers) might better index the underlying gene effects (73), although none has yet proven more heritable than diagnostic categories. These measures are not usually available in large datasets.
3. Genetic heterogeneity reduces power. Low frequency alleles are examples of heterogeneity (i.e., most cases do not share that risk factor). Power (Figure 1) is best for frequencies above around 20%, and poor at much below 10% unless GRR is high. Heterogeneity might be increased in large multicenter samples; e.g., despite the generally high inter-rater reliability for these disorders, research groups can have diagnostic “biases”, some of which could correlate with specific risk alleles. But power increases with sample size despite some degree of misclassification, which also occurs in many medical disorders for which there are GWAS findings.
4. More needs to be learned about the selection of controls for psychiatric GWAS studies. It remains possible that some findings will be confounded by systematic biases in control groups, such as under-representation of developmental disabilities. In any event, the field will need much larger control groups ascertained by diverse methods and from multiple ethnic populations.
5. For some disorders, there might be no detectable main effects of SNPs, only higher order gene-gene or gene-environment interactions. However, main effects are often detectable even if interactions are erroneously excluded. Explicit tests of interactions (74) or data mining might prove informative.
6. GWAS assays do not interrogate all common variants. For each array type, some assays perform poorly, and some common SNPs are not or cannot be tagged.
7. Improved methods will be needed to provide more systematic information about CNVs and their relationship to disease. Associated CNV regions will require

resequencing studies of large numbers of subjects without CNVs, to determine whether these regions also contain rare, highly penetrant associated variants.

8. There are probably unknown genetic mechanisms. We have only recently recognized the importance of CNVs, micro RNAs, long-range promoters and epigenetic factors (genomic effects other than sequence changes, such as DNA methylation patterns). (75) The discovery that most of the genome is transcribed suggests that many types of functional sequence are undiscovered. (12)

Bearing these risks and caveats in mind, we conclude that GWAS methods have discovered a remarkable set of robust common SNP association findings for a broad range of diseases, now including an initial set of SNP and CNV associations for psychiatric disorders. It is reasonable to predict that studies of sufficiently large samples can produce definitive discoveries of genetic risk factors for psychiatric disorders, and that these discoveries will contribute to the definitive identification of pathophysiological mechanisms for the first time.

Acknowledgments

This article was written by the Psychiatric GWAS Consortium Coordinating Committee, whose members (presented in alphabetical order) take responsibility for its content: Sven Cichon, Ph.D. (University of Bonn, Germany); Nick Craddock, M.D., Ph.D. (Cardiff University); Mark Daly, Ph.D. (Harvard Medical School, Broad Institute); Stephen V. Faraone, Ph.D. (State University of New York Upstate Medical University); Pablo V. Gejman, M.D. (Northshore University HealthSystem and Feinberg School of Medicine of Northwestern University); John Kelsoe, M.D. (University of California, San Diego); Thomas Lehner, Ph.D., M.P.H. (NIMH); Douglas F. Levinson, M.D. (Stanford University); Audra Moran, M.A. (NARSAD, Ex Officio); Pamela Sklar, M.D., Ph.D. (Massachusetts General Hospital, Broad Institute); and Patrick F. Sullivan, M.D. (University of North Carolina at Chapel Hill).

Dr. Faraone receives research support from or has served on the advisory boards of Shire, Eli Lilly, Pfizer, McNeil, and NIH. Dr. Kelsoe is a founder of and holds equity in Psynomics, Inc. Dr. Sullivan has received unrestricted research support from Eli Lilly for genetic research in schizophrenia. Drs. Cichon, Craddock, Daly, Gejman, Lehner, Levinson, Sklar, and Sullivan and Ms. Moran report no competing interests.

Supported by NIMH grant MH-085520. Statistical analyses were conducted using the Genetic Cluster Computer, which is supported by the Netherlands Scientific Organization (NWO 480-05-003, PI Danielle Posthuma), along with a supplement from the Dutch Brain Foundation.

ADHD Working Group: Stephen Faraone, Chair (SUNY-UMU); Richard Anney (Trinity College Dublin); Jan Buitelaar (Radboud University); Josephine Elia (Children's Hospital of Philadelphia); Barbara Franke (Radboud University); Michael Gill (Trinity College Dublin); Hakon Hakonarson (CHOP); Lindsey Kent (St. Andrews University); James McGough (UCLA); Eric Mick (Massachusetts General Hospital/ Harvard University); Laura Nisenbaum (Eli Lilly); Susan Smalley (UCLA); Anita Thapar (Cardiff University); Richard Todd, deceased (Washington University/St. Louis, MO); and Alexandre Todorov (Washington University/St. Louis, MO).

Autism Working Group: Bernie Devlin, Chair (University of Pittsburgh); Mark Daly, Co-Chair (Massachusetts General Hospital/Harvard University); Richard Anney (Trinity College Dublin); Dan Arking (Johns Hopkins University); Joseph D. Buxbaum (Mt. Sinai School of Medicine, New York); Aravinda Chakravarti (Johns Hopkins University); Edwin Cook (University of Illinois); Michael Gill (Trinity College Dublin); Leena Peltonen (University of Helsinki); Joseph Piven (University of North Carolina-Chapel Hill); Guy Rouleau (University of Montreal); Susan Santangelo (Massachusetts General Hospital/Harvard University); Gerard Schellenberg (University of Washington); Steve Scherer (University of Toronto); James Sutcliffe (Vanderbilt University); Peter Szatmari (McMaster University); and Veronica Vieland (Columbus Children's Research Institute).

Bipolar Disorder Working Group: John Kelsoe, Co-Chair (UCSD); Pamela Sklar, Co-Chair, (Massachusetts General Hospital/Harvard University); Ole A. Andreassen (University of Oslo, Norway); Douglas Blackwood (University of Edinburgh, Scotland); Michael Boehnke (University of Michigan); Rene Breuer (CIMH, Mannheim, Germany); Margit Burmeister (University of Michigan); Sven Cichon (University of Bonn, Germany); Aiden Corvin (Trinity College Dublin); Nicholas Craddock (Cardiff University); Manuel Ferreira (Massachusetts General Hospital/Harvard University); Matthew Flickinger (University of Michigan); Tiffany Greenwood (UCSD); Weihua Guan (University of Michigan); Hugh Gurling (University College London); Jun Li (University of Michigan); Eric Mick (Massachusetts General Hospital/Harvard University); Valentina Moskvina (Cardiff University); Pierandrea

Muglia (GlaxoSmithKline); Walter Muir (University of Edinburgh, Scotland); Markus Nothen (University of Bonn, Germany); John Nurnberger (Indiana University); Shaun Purcell (Massachusetts General Hospital/Harvard University); Marcella Rietschel (CIMH, Mannheim); Douglas Ruderfer (Massachusetts General Hospital/Harvard University); Nicholas Schork (UCSD); Thomas Schulze (CIMH, Mannheim); Laura Scott (University of Michigan); Michael Steffens (University of Bonn, Germany); Ruchi Upmanyu (GlaxoSmithKline); and Thomas Wienker (University of Bonn, Germany).

Cross-Disorder Working Group: Jordan Smoller, Co-Chair (Massachusetts General Hospital/Harvard University); Nicholas Craddock, Co-Chair (Cardiff University); Kenneth Kendler, Co-Chair (Virginia Commonwealth University); John Nurnberger (Indiana University); Roy Perlis (Massachusetts General Hospital/Harvard University); Shaun Purcell (Massachusetts General Hospital/Harvard University); Marcella Rietschel (CIMH, Mannheim); Susan Santangelo (Massachusetts General Hospital/Harvard University); and Anita Thapar (Cardiff University).

Major Depressive Disorder Working Group: Patrick Sullivan, Chair (University of North Carolina-Chapel Hill); Douglas Blackwood (University of Edinburgh, Scotland); Dorret Boomsma (Vrije University, Amsterdam); Rene Breuer (CIMH, Mannheim, Germany); Sven Cichon (University of Bonn, Germany); William Coryell (University of Iowa); Eco de Geus (Vrije University, Amsterdam); Steve Hamilton (UCSF); Witte Hoogendijk (Vrije University, Amsterdam); Stafam Kloiber (MPIP Munich); William B. Lawson (Howard University); Douglas Levinson (Stanford University); Cathryn Lewis (IOP, London); Susanne Lucae (MPI-P Munich); Nick Martin (QIMR); Patrick McGrath (Columbia University); Peter McGuffin (IOP, London); Pierandrea Muglia (Glaxo-SmithKline); Walter Muir (University of Edinburgh, Scotland); Markus Nothen (University of Bonn, Germany); James Offord (Pfizer); Brenda Penninx (Vrije University, Amsterdam); James B. Potash (Johns Hopkins University); Marcella Rietschel (CIMH, Mannheim, Germany); William A. Scheftner (Rush University); Thomas Schulze (CIMH, Mannheim); Susan Slager (Mayo Clinic); Federica Tozzi (Glaxo-SmithKline); Myrna M. Weissman (Columbia University); AHM Willemsen (Vrije University, Amsterdam); and Naomi Wray (QIMR).

Schizophrenia Working Group: Pablo Gejman, Chair (Northshore University HealthSystem and Feinberg School of Medicine of Northwestern University); Ole A. Andreassen (University of Oslo, Norway); Douglas Blackwood (University of Edinburgh, Scotland); Sven Cichon (University of Bonn, Germany); Aiden Corvin (Trinity College Dublin); Mark Daly (Massachusetts General Hospital/Harvard University); Ayman Fanous (Washington Veterans Administration Medical Center, Georgetown University, Virginia Commonwealth University); Michael Gill (Trinity College Dublin); Hugh Gurling (UCL); Peter Holmans (Cardiff University); Christina Hultman (Karolinska Institutet); Kenneth Kendler (Virginia Commonwealth University); Sari Kivikko (National Public Health Institute); Claudine Laurent (Pierre and Marie Curie Faculty of Medicine, Paris); Todd Lencz (LIJ); Douglas Levinson (Stanford University); Anil Malhotra (LIJ); Bryan Mowry (Queensland Center for Mental Health Research, University of Queensland); Markus Nothen (University of Bonn, Germany); Mike O'Donovan (Cardiff University); Roel Ophoff (UCLA); Michael Owen (Cardiff University); Leena Peltonen (University of Helsinki); Ann Pulver (Johns Hopkins University); Marcella Rietschel (CIMH, Mannheim); Brien Riley (Virginia Commonwealth University); Alan Sanders (Northshore University HealthSystem and Feinberg School of Medicine of Northwestern University); Thomas Schulze (CIMH, Mannheim); Sibylle Schwab (University of Western Australia); Pamela Sklar (Massachusetts General Hospital/Harvard University); David St. Clair (University of Aberdeen); Patrick Sullivan (University of North Carolina-Chapel Hill); Jaana Suvisaari (University of Helsinki); Edwin van den Oord (Virginia Commonwealth University); Naomi Wray (QIMR); and Dieter Wildenauer (University of Western Australia).

Statistical Analysis and Computational Working Group: Mark Daly, Chair (Massachusetts General Hospital/Harvard University); Phillip Awadalla (University of Montreal); Bernie Devlin (University of Pittsburgh); Frank Dudbridge (MRC-BSU); Arnolfo Frigessi (University of Oslo, Norway); Elizabeth Holliday (QCMHR/University of Queensland); Peter Holmans (Cardiff University); Todd Lencz (LIJ); Douglas Levinson (Stanford University); Cathryn Lewis (IOP, London); Danyu Lin (University of North Carolina-Chapel Hill); Valentina Moskvina (Cardiff University); Bryan Mowry (QCMHR/University of Queensland); Ben Neale (Massachusetts General Hospital/Harvard University); Eve Pickering (Pfizer Pharmaceuticals Group); Danielle Postuma (Vrije University Amsterdam); Shaun Purcell (Massachusetts General Hospital/Harvard University); John Rice (Washington University/St. Louis, MO); Stephan Ripke (MPI-P Munich); Nicholas Schork (UCSD); Jonathan Sebat (CSHL); Michael Steffens (University of Bonn, Germany); Jennifer Stone (Massachusetts General Hospital/Harvard University); Jung-Ying Tzeng (NCSU); Edwin van den Oord (Virginia Commonwealth University); and Veronica Vieland (Columbus Children's Research Institute).

The authors thank their Psychiatric GWAS Consortium colleagues for their contributions. The authors also thank NARSAD for infrastructure support.

References

1. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J. Complement factor H polymorphism in age-related macular degeneration. *Science*. 2005; 308(5720):385–389. [PubMed: 15761122]
2. Manolio TA, Brooks LD, Collins FS. A HapMap harvest of insights into the genetics of common disease. *J Clin Invest*. 2008; 118(5):1590–605. [PubMed: 18451988]
3. Allen NC, Bagade S, McQueen MB, Ioannidis JP, Kavvoura FK, Khoury MJ, Tanzi RE, Bertram L. Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat Genet*. 2008; 40(7):827–34. [PubMed: 18583979]
4. Levinson DF. Meta-analysis in Psychiatric Genetics. *Curr. Psychiatry Rep*. 2005; 7(2):143–151. [PubMed: 15802092]
5. Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet*. 1980; 32(3):314–31. [PubMed: 6247908]
6. Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, Tanzi RE, Watkins PC, Ottina K, Wallace MR, Sakaguchi AY. A polymorphic DNA marker genetically linked to Huntington's disease. *Nature*. 1983; 306:234–238. [PubMed: 6316146]
7. Donis-Keller H, Green P, Helms C, Cartinhour S, Weiffenbach B, Stephens K, Keith TP, Bowden DW, Smith DR, Lander ES, et al. A genetic linkage map of the human genome. *Cell*. 1987; 51(2): 319–37. [PubMed: 3664638]
8. Coon H, Jensen S, Hoff M, Holik J, Plaetke R, Reimherr F, Wender P, Leppert M, Byerley W. A genome-wide search for genes predisposing to manic-depression, assuming autosomal dominant inheritance. *Am J Hum Genet*. 1993; 52:1234–1249. [PubMed: 8503452]
9. Lander ES. The new genomics: global views of biology. *Science*. 1996; 274(5287):536–539. [PubMed: 8928008]
10. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001; 409(6822):860–921. [PubMed: 11237011]
11. Gibbs RA, Belmont JW, Hardenbol P, et al. The International HapMap Project. *Nature*. 2003; 426(6968):789–796. [PubMed: 14685227]
12. Frazer KA, Ballinger DG, Cox DR, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature*. 2007; 449(7164):851–61. [PubMed: 17943122]
13. Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, Tsunoda T, Sato H, Sato H, Hori M, Nakamura Y, Tanaka T. Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nat Genet*. 2002; 32(4):650–4. [PubMed: 12426569]
14. Yamazaki K, McGovern D, Ragoussis J, Paolucci M, Butler H, Jewell D, Cardon L, Takazoe M, Tanaka T, Ichimori T, Saito S, Sekine A, Iida A, Takahashi A, Tsunoda T, Lathrop M, Nakamura Y. Single nucleotide polymorphisms in TNFSF15 confer susceptibility to Crohn's disease. *Hum. Mol. Genet*. 2005; 14(22):3499–3506. [PubMed: 16221758]
15. Teare MD, Barrett JH. Genetic linkage studies. *Lancet*. 2005; 366(9490):1036–44. [PubMed: 16168786]
16. Lewis CM, Levinson DF, Wise LH, DeLisi LE, Straub RE, Hovatta I, Williams NM, Schwab SG, Pulver AE, Faraone SV, Brzustowicz LM, Kaufmann CA, Garver DL, Gurling HM, Lindholm E, Coon H, Moises HW, Byerley W, Shaw SH, Mesen A, Sherrington R, O'Neill FA, Walsh D, Kendler KS, Ekelund J, Paunio T, Lonnqvist J, Peltonen L, O'Donovan MC, Owen MJ, Wildenauer DB, Maier W, Nestadt G, Blouin JL, Antonarakis SE, Mowry BJ, Silverman JM, Crowe RR, Cloninger CR, Tsuang MT, Malaspina D, Harkavy-Friedman JM, Svrakic DM, Bassett AS, Holcomb J, Kalsi G, McQuillin A, Brynjolfson J, Sigmundsson T, Petursson H, Jazin E, Zoega T, Helgason T. Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. *Am. J. Hum. Genet*. 2003; 73(1):34–48. [PubMed: 12802786]
17. McQueen MB, Devlin B, Faraone SV, Nimgaonkar VL, Sklar P, Smoller JW, Abou Jamra R, Albus M, Bacanu SA, Baron M, Barrett TB, Berrettini W, Blacker D, Byerley W, Cichon S, Coryell W, Craddock N, Daly MJ, Depaulo JR, Edenberg HJ, Foroud T, Gill M, Gilliam TC,

Hamshere M, Jones I, Jones L, Juo SH, Kelsoe JR, Lambert D, Lange C, Lerer B, Liu J, Maier W, Mackinnon JD, McInnis MG, McMahon FJ, Murphy DL, Nothen MM, Nurnberger JI, Pato CN, Pato MT, Potash JB, Propping P, Pulver AE, Rice JP, Rietschel M, Scheftner W, Schumacher J, Segurado R, Van Steen K, Xie W, Zandi PP, Laird NM. Combined analysis from eleven linkage studies of bipolar disorder provides strong evidence of susceptibility loci on chromosomes 6q and 8q. *Am J Hum Genet.* 2005; 77(4):582–95. [PubMed: 16175504]

18. Trikalinos TA, Karvouni A, Zintzaras E, Ylisaukko-oja T, Peltonen L, Jarvela I, Ioannidis JP. A heterogeneity-based genome search meta-analysis for autism-spectrum disorders. *Mol Psychiatry.* 2006; 11(1):29–36. [PubMed: 16189507]
19. Xiong M, Guo SW. Fine-scale genetic mapping based on linkage disequilibrium: theory and applications. *Am J Hum Genet.* 1997; 60(6):1513–31. [PubMed: 9199574]
20. Norton N, Williams HJ, Owen MJ. An update on the genetics of schizophrenia. *Curr Opin Psychiatry.* 2006; 19(2):158–64. [PubMed: 16612196]
21. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science.* 1996; 273(5281):1516–1517. [PubMed: 8801636]
22. Pe'er I, Chretien YR, de Bakker PI, Barrett JC, Daly MJ, Altshuler DM. Biases and reconciliation in estimates of linkage disequilibrium in the human genome. *Am J Hum Genet.* 2006; 78(4):588–603. [PubMed: 16532390]
23. Kruglyak L. The road to genome-wide association studies. *Nat Rev Genet.* 2008; 9(4):314–8. [PubMed: 18283274]
24. Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet.* 2001; 69(1):124–37. [PubMed: 11404818]
25. Guthery SL, Salisbury BA, Pungliya MS, Stephens JC, Bamshad M. The structure of common genetic variation in United States populations. *Am J Hum Genet.* 2007; 81(6):1221–31. [PubMed: 17999361]
26. Boyko AR, Williamson SH, Indap AR, Degenhardt JD, Hernandez RD, Lohmueller KE, Adams MD, Schmidt S, Sninsky JJ, Sunyaev SR, White TJ, Nielsen R, Clark AG, Bustamante CD. Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet.* 2008; 4(5):e1000083. [PubMed: 18516229]
27. Stone EA, Sidow A. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. *Genome Res.* 2005; 15(7):978–86. [PubMed: 15965030]
28. Duan J, Wainwright MS, Comeron JM, Saitou N, Sanders AR, Gelernter J, Gejman PV. Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. *Hum. Mol. Genet.* 2003; 12(3):205–216. [PubMed: 12554675]
29. Birney E, Stamatoyannopoulos JA, Dutta A, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature.* 2007; 447(7146):799–816. [PubMed: 17571346]
30. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet.* 2003; 33(Suppl):228–37. [PubMed: 12610532]
31. Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science.* 2004; 305(5685):869–872. [PubMed: 15297675]
32. Ji W, Foo JN, O'Roak BJ, Zhao H, Larson MG, Simon DB, Newton-Cheh C, State MW, Levy D, Lifton RP. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nat Genet.* 2008
33. Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, Abraham C, Regueiro M, Griffiths A, Dassopoulos T, Bitton A, Yang H, Targan S, Datta LW, Kistner EO, Schumm LP, Lee A, Gregersen PK, Barnada MM, Rotter JI, Nicolae DL, Cho JH. A Genome-Wide Association Study Identifies IL23R as an Inflammatory Bowel Disease Gene. *Science.* 2006
34. Li M, Atmaca-Sonmez P, Othman M, Branham KE, Khanna R, Wade MS, Li Y, Liang L, Zarepari S, Swaroop A, Abecasis GR. CFH haplotypes without the Y402H coding variant show

strong association with susceptibility to age-related macular degeneration. *Nat. Genet.* 2006; 38(9):1049–1054. [PubMed: 16936733]

35. Loos RJ, Lindgren CM, Li S, et al. Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat Genet.* 2008; 40(6):768–75. [PubMed: 18454148]
36. Murphy KC, Jones LA, Owen MJ. High rates of schizophrenia in adults with velo-cardio-facial syndrome. *Arch. Gen. Psychiatry.* 1999; 56(10):940–945. [PubMed: 10530637]
37. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, Gonzalez JR, Gratacos M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME. Global variation in copy number in the human genome. *Nature.* 2006; 444(7118):444–54. [PubMed: 17122850]
38. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 2006; 38(8): 904–909. [PubMed: 16862161]
39. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet.* 2007; 39(7):906–13. [PubMed: 17572673]
40. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet.* 2008; 9(5):356–69. [PubMed: 18398418]
41. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature.* 2007; 447(7145):661–78. [PubMed: 17554300]
42. Dudbridge F, Gusnanto A. Estimation of significance thresholds for genomewide association scans. *Genet Epidemiol.* 2008; 32(3):227–34. [PubMed: 18300295]
43. Hoggart CJ, Clark TG, De Iorio M, Whittaker JC, Balding DJ. Genome-wide significance for dense SNP and resequencing data. *Genet Epidemiol.* 2008; 32(2):179–85. [PubMed: 18200594]
44. 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(4): 381–5. [PubMed: 18348202]
45. Sabatti C, Service S, Freimer N. False discovery rate in linkage and association genome screens for complex disorders. *Genetics.* 2003; 164(2):829–833. [PubMed: 12807801]
46. Wakefield J. Bayes factors for genome-wide association studies: comparison with P-values. *Genet Epidemiol.* 2008
47. Frayling TM. Genome-wide association studies provide new insights into type 2 diabetes aetiology. *Nat Rev Genet.* 2007; 8(9):657–62. [PubMed: 17703236]
48. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, de Bakker PI, Abecasis GR, Almgren P, Andersen G, Ardlie K, Bostrom KB, Bergman RN, Bonnycastle LL, Borch-Johnsen K, Burtt NP, Chen H, Chines PS, Daly MJ, Deodhar P, Ding CJ, Doney AS, Duren WL, Elliott KS, Erdos MR, Frayling TM, Freathy RM, Gianniny L, Grallert H, Grarup N, Groves CJ, Guiducci C, Hansen T, Herder C, Hitman GA, Hughes TE, Isomaa B, Jackson AU, Jorgensen T, Kong A, Kubalanza K, Kuruvilla FG, Kuusisto J, Langenberg C, Lango H, Lauritzen T, Li Y, Lindgren CM, Lyssenko V, Marvelle AF, Meisinger C, Midthjell K, Mohlke KL, Morken MA, Morris AD, Narisu N, Nilsson P, Owen KR, Palmer CN, Payne F, Perry JR, Pettersen E, Platou C, Prokopenko I, Qi L, Qin L, Rayner NW, Rees M, Roix JJ, Sandbaek A, Shields B, Sjogren M, Steinthorsdottir V, Stringham HM, Swift AJ, Thorleifsson G, Thorsteinsdottir U, Timpson NJ, Tuomi T, Tuomilehto J, Walker M, Watanabe RM, Weedon MN, Willer CJ, Illig T, Hveem K, Hu FB, Laakso M, Stefansson K, Pedersen O, Wareham NJ, Barroso I, Hattersley AT, Collins FS, Groop L, McCarthy MI, Boehnke M, Altshuler D. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet.* 2008
49. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JR, Elliott KS, Lango H, Rayner NW, Shields B, Harries LW, Barrett JC, Ellard S, Groves CJ, Knight B,

Patch AM, Ness AR, Ebrahim S, Lawlor DA, Ring SM, Ben-Shlomo Y, Jarvelin MR, Sovio U, Bennett AJ, Melzer D, Ferrucci L, Loos RJ, Barroso I, Wareham NJ, Karpe F, Owen KR, Cardon LR, Walker M, Hitman GA, Palmer CN, Doney AS, Morris AD, Smith GD, Hattersley AT, McCarthy MI. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*. 2007; 316(5826):889–94. [PubMed: 17434869]

50. Freathy RM, Timpson NJ, Lawlor DA, Pouta A, Ben-Shlomo Y, Ruokonen A, Ebrahim S, Shields B, Zeggini E, Weedon MN, Lindgren CM, Lango H, Melzer D, Ferrucci L, Paolisso G, Neville MJ, Karpe F, Palmer CN, Morris AD, Elliott P, Jarvelin MR, Smith GD, McCarthy MI, Hattersley AT, Frayling TM. Common variation in the FTO gene alters diabetes-related metabolic traits to the extent expected, given its effect on BMI. *Diabetes*. 2008
51. Shoulders CC. The FTO (fat mass and obesity-associated) gene: big in adipocyte lipolysis? *J Lipid Res*. 2008; 49(3):495–6. [PubMed: 18281390]
52. Todd JA. Statistical false positive or true disease pathway? *Nat. Genet*. 2006; 38(7):731–733. [PubMed: 16804532]
53. Saxena R, Voight BF, Lyssenko V, Burt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop L, Altshuler D, Almgren P, Florez JC, Meyer J, Ardlie K, Bengtsson Bostrom K, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Rastam L, Speliotes EK, Taskinen MR, Tuomi T, Guiducci C, Berglund A, Carlson J, Gianniny L, Hackett R, Hall L, Holmkvist J, Laurila E, Sjogren M, Sterner M, Surti A, Svensson M, Svensson M, Tewhey R, Blumenstiel B, Parkin M, Defelice M, Barry R, Brodeur W, Camarata J, Chia N, Fava M, Gibbons J, Handsaker B, Healy C, Nguyen K, Gates C, Sougnez C, Gage D, Nizzari M, Gabriel SB, Chirn GW, Ma Q, Parikh H, Richardson D, Ricke D, Purcell S. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*. 2007; 316(5829):1331–6. [PubMed: 17463246]
54. Ghoussaini M, Song H, Koessler T, Al Olama AA, Kote-Jarai Z, Driver KE, Pooley KA, Ramus SJ, Kjaer SK, Hogdall E, DiCioccio RA, Whittemore AS, Gayther SA, Giles GG, Guy M, Edwards SM, Morrison J, Donovan JL, Hamdy FC, Dearnaley DP, Arden-Jones AT, Hall AL, O'Brien LT, Gehr-Swain BN, Wilkinson RA, Brown PM, Hopper JL, Neal DE, Pharoah PD, Ponder BA, Eeles RA, Easton DF, Dunning AM. Multiple loci with different cancer specificities within the 8q24 gene desert. *J Natl Cancer Inst*. 2008; 100(13):962–6. [PubMed: 18577746]
55. Stefansson H, Rujescu D, Cichon S, Pietilainen OP, Ingason A, Steinberg S, Fossdal R, Sigurdsson E, Sigmundsson T, Buizer-Voskamp JE, Hansen T, Jakobsen KD, Muglia P, Francks C, Matthews PM, Gylfason A, Halldorsson BV, Gudbjartsson D, Thorgerirsson TE, Sigurdsson A, Jonasdottir A, Jonasdottir A, Bjornsson A, Mattiasdottir S, Blondal T, Haraldsson M, Magnusdottir BB, Giegling I, Moller HJ, Hartmann A, Shianna KV, Ge D, Need AC, Crombie C, Fraser G, Walker N, Lonnqvist J, Suvisaari J, Tuulio-Henriksson A, Paunio T, Touloupoulou T, Bramon E, Di Forti M, Murray R, Ruggeri M, Vassos E, Tosato S, Walshe M, Li T, Vasilescu C, Muhleisen TW, Wang AG, Ullum H, Djurovic S, Melle I, Olesen J, Kiemeny LA, Franke B, Kahn RS, Linszen D, van Os J, Wiersma D, Bruggeman R, Cahn W, Germeys I, de Haan L, Krabbendam L, Sabatti C, Freimer NB, Gulcher JR, Thorsteinsdottir U, Kong A, Andreassen OA, Ophoff RA, Georgi A, Rietschel M, Werge T, Petursson H, Goldstein DB, Nothen MM, Peltonen L, Collier DA, St Clair D, Stefansson K. Large recurrent microdeletions associated with schizophrenia. *Nature*. 2008
56. Stone JL, O'Donovan MC, Gurling H, et al. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*. 2008
57. Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, Nord AS, Kusenda M, Malhotra D, Bhandari A, Stray SM, Rippey CF, Roccanova P, Makarov V, Lakshmi B, Findling RL, Sikich L, Stromberg T, Merriman B, Gogtay N, Butler P, Eckstrand K, Noory L, Gochman P, Long R, Chen Z, Davis S, Baker C, Eichler EE, Meltzer PS, Nelson SF, Singleton AB, Lee MK, Rapoport JL, King MC, Sebat J. Rare Structural Variants Disrupt Multiple Genes in Neurodevelopmental Pathways in Schizophrenia. *Science*. 2008
58. 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(7):880–5. [PubMed: 18511947]
59. Sklar P, Smoller JW, Fan J, Ferreira MA, Perlis RH, Chambert K, Nimgaonkar VL, McQueen MB, Faraone SV, Kirby A, de Bakker PI, Ogdie MN, Thase ME, Sachs GS, Todd-Brown K, Gabriel

SB, Sougnez C, Gates C, Blumenstiel B, Defelice M, Ardlie KG, Franklin J, Muir WJ, McGhee KA, MacIntyre DJ, McLean A, VanBeck M, McQuillin A, Bass NJ, Robinson M, Lawrence J, Anjorin A, Curtis D, Scolnick EM, Daly MJ, Blackwood DH, Gurling HM, Purcell SM. Whole-genome association study of bipolar disorder. *Mol Psychiatry*. 2008; 13(6):558–69. [PubMed: 18317468]

60. Ferreira MAR, O'Donovan MC, Meng YA, Jones IR, Ruderfer DM, Jones L, Fan J, Kirov G, Perlis RH, Green EK, Smoller JW, Grozeva D, Stone J, Nikolov I, Chambert K, Hamshere ML, Nimgaonkar V, Moskvina V, Thase ME, Caesar S, Sachs GS, Franklin J, Gordon-Smith K, Ardlie K, Gabriel SB, Fraser C, Blumenstiel B, Defelice M, Breen G, Gill M, Morris DW, Elkin A, Muir WJ, McGhee KA, Williamson R, MacIntyre D, McLean A, Clair D, VanBeck M, Pereira A, Kandaswamy R, McQuillin A, Collier DA, Bass NJ, Young AH, Lawrence J, Ferrier I, Anjorin A, Farmer A, Curtis D, Scolnick EM, McGuffin P, Daly MJ, Corvin AP, A'Holmans P, Blackwood DH, Consortium WTCC, Gurling HM, Owen MJ, Purcell SM, Sklar P, Craddock N. Collaborative genome-wide association analysis of 10,596 individuals supports a role for Ankyrin-G (ANK3) and the alpha-1C subunit of the L-type voltage-gated calcium channel (CACNA1C) in bipolar disorder. *Nat Genet*. 2008 In press.
61. Lencz T, Morgan TV, Athanasiou M, Dain B, Reed CR, Kane JM, Kucherlapati R, Malhotra AK. Converging evidence for a pseudoautosomal cytokine receptor gene locus in schizophrenia. *Mol Psychiatry*. 2007; 12(6):572–80. [PubMed: 17522711]
62. Sullivan PF, Lin D, Tzeng JY, van den Oord E, Perkins D, Stroup TS, Wagner M, Lee S, Wright FA, Zou F, Liu W, Downing AM, Lieberman J, Close SL. Genomewide association for schizophrenia in the CATIE study: results of stage 1. *Mol Psychiatry*. 2008; 13(6):570–84. [PubMed: 18347602]
63. O'Donovan MC, Craddock N, Norton N, Williams H, Peirce T, Moskvina V, Nikolov I, Hamshere M, Carroll L, Georgieva L, Dwyer S, Holmans P, Marchini JL, Spencer CC, Howie B, Leung HT, Hartmann AM, Moller HJ, Morris DW, Shi Y, Feng G, Hoffmann P, Propping P, Vasilescu C, Maier W, Rietschel M, Zammit S, Schumacher J, Quinn EM, Schulze TG, Williams NM, Giegling I, Iwata N, Ikeda M, Darvasi A, Shifman S, He L, Duan J, Sanders AR, Levinson DF, Gejman PV, Gejman PV, Sanders AR, Duan J, Levinson DF, Buccola NG, Mowry BJ, Freedman R, Amin F, Black DW, Silverman JM, Byerley WF, Cloninger CR, Cichon S, Nothen MM, Gill M, Corvin A, Rujescu D, Kirov G, Owen MJ. Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet*. 2008
64. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, Yamrom B, Yoon S, Krasnitz A, Kendall J, Leotta A, Pai D, Zhang R, Lee YH, Hicks J, Spence SJ, Lee AT, Puura K, Lehtimäki T, Ledbetter D, Gregersen PK, Bregman J, Sutcliffe JS, Jobanputra V, Chung W, Warburton D, King MC, Skuse D, Geschwind DH, Gilliam TC, Ye K, Wigler M. Strong association of de novo copy number mutations with autism. *Science*. 2007; 316(5823):445–9. [PubMed: 17363630]
65. Kumar RA, KaraMohamed S, Sudi J, Conrad DF, Brune C, Badner JA, Gilliam TC, Nowak NJ, Cook EH Jr, Dobyns WB, Christian SL. Recurrent 16p11.2 microdeletions in autism. *Hum Mol Genet*. 2008; 17(4):628–38. [PubMed: 18156158]
66. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, Shago M, Moessner R, Pinto D, Ren Y, Thiruvahindrapadam B, Fiebig A, Schreiber S, Friedman J, Ketelaars CE, Vos YJ, Ficicocioglu C, Kirkpatrick S, Nicolson R, Sloman L, Summers A, Gibbons CA, Teebi A, Chitayat D, Weksberg R, Thompson A, Vardy C, Crosbie V, Luscombe S, Baatjes R, Zwaigenbaum L, Roberts W, Fernandez B, Szatmari P, Scherer SW. Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet*. 2008; 82(2):477–88. [PubMed: 18252227]
67. Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, Saemundsen E, Stefansson H, Ferreira MA, Green T, Platt OS, Ruderfer DM, Walsh CA, Altshuler D, Chakravarti A, Tanzi RE, Stefansson K, Santangelo SL, Gusella JF, Sklar P, Wu BL, Daly MJ. Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med*. 2008; 358(7):667–75. [PubMed: 18184952]
68. Christian SL, Brune CW, Sudi J, Kumar RA, Liu S, Karamohamed S, Badner JA, Matsui S, Conroy J, McQuaid D, Gergel J, Hatchwell E, Gilliam TC, Gershon ES, Nowak NJ, Dobyns WB, Cook EH Jr. Novel submicroscopic chromosomal abnormalities detected in autism spectrum disorder. *Biol Psychiatry*. 2008; 63(12):1111–7. [PubMed: 18374305]

69. Kirov G, Zaharieva I, Georgieva L, Moskvina V, Nikolov I, Cichon S, Hillmer A, Toncheva D, Owen MJ, O'Donovan MC. A genome-wide association study in 574 schizophrenia trios using DNA pooling. *Mol Psychiatry*. 2008
70. Shifman S, Johannesson M, Bronstein M, Chen SX, Collier DA, Craddock NJ, Kendler KS, Li T, O'Donovan M, O'Neill FA, Owen MJ, Walsh D, Weinberger DR, Sun C, Flint J, Darvasi A. Genome-wide association identifies a common variant in the reelin gene that increases the risk of schizophrenia only in women. *PLoS Genet*. 2008; 4(2):e28. [PubMed: 18282107]
71. Baum AE, Akula N, Cabanero M, Cardona I, Corona W, Klemens B, Schulze TG, Cichon S, Rietschel M, Nothen MM, Georgi A, Schumacher J, Schwarz M, Abou Jamra R, Hofels S, Propping P, Satagopan J, Detera-Wadleigh SD, Hardy J, McMahon FJ. A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. *Mol Psychiatry*. 2008; 13(2):197–207. [PubMed: 17486107]
72. Achkar JP, Duerr R. The expanding universe of inflammatory bowel disease genetics. *Curr Opin Gastroenterol*. 2008; 24(4):429–34. [PubMed: 18622155]
73. Gottesman II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry*. 2003; 160(4):636–45. [PubMed: 12668349]
74. Marchini J, Donnelly P, Cardon LR. Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nat. Genet*. 2005; 37(4):413–417. [PubMed: 15793588]
75. Ptak C, Petronis A. Epigenetics and complex disease: from etiology to new therapeutics. *Annu Rev Pharmacol Toxicol*. 2008; 48:257–76. [PubMed: 17883328]
76. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81(3):559–75. [PubMed: 17701901]
77. Costello EJ, Mustillo S, Erkanli A, Keeler G, Angold A. Prevalence and development of psychiatric disorders in childhood and adolescence. *Arch Gen Psychiatry*. 2003; 60(8):837–44. [PubMed: 12912767]
78. Faraone SV, Sergeant J, Gillberg C, Biederman J. The worldwide prevalence of ADHD: is it an American condition? *World Psychiatry*. 2003; 2(2):104–113. [PubMed: 16946911]
79. Faraone SV, Perlis RH, Doyle AE, Smoller JW, Goralnick JJ, Holmgren MA, Sklar P. Molecular genetics of attention-deficit/hyperactivity disorder. *Biol Psychiatry*. 2005; 57(11):1313–23. [PubMed: 15950004]
80. Rutter M. Incidence of autism spectrum disorders: changes over time and their meaning. *Acta Paediatr*. 2005; 94(1):2–15. [PubMed: 15858952]
81. Freitag CM. The genetics of autistic disorders and its clinical relevance: a review of the literature. *Mol Psychiatry*. 2007; 12(1):2–22. [PubMed: 17033636]
82. Weissman MM, Bland RC, Canino GJ, Faravelli C, Greenwald S, Hwu HG, Joyce PR, Karam EG, Lee CK, Lellouch J, Lepine JP, Newman SC, Rubio-Stipec M, Wells JE, Wickramaratne PJ, Wittchen H, Yeh EK. Cross-national epidemiology of major depression and bipolar disorder. *Jama*. 1996; 276(4):293–9. [PubMed: 8656541]
83. McGuffin P, Rijsdijk F, Andrew M, Sham P, Katz R, Cardno A. The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. *Arch Gen Psychiatry*. 2003; 60(5):497–502. [PubMed: 12742871]
84. Hasin DS, Goodwin RD, Stinson FS, Grant BF. Epidemiology of major depressive disorder: results from the National Epidemiologic Survey on Alcoholism and Related Conditions. *Arch Gen Psychiatry*. 2005; 62(10):1097–106. [PubMed: 16203955]
85. Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry*. 2000; 157(10):1552–62. [PubMed: 11007705]
86. Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ. The lifetime history of major depression in women. Reliability of diagnosis and heritability. *Arch Gen Psychiatry*. 1993; 50(11):863–70.
87. Thapar A, McGuffin P. A twin study of depressive symptoms in childhood. *Br J Psychiatry*. 1994; 165(2):259–65. [PubMed: 7953041]
88. Saha S, Chant D, Welham J, McGrath J. A systematic review of the prevalence of schizophrenia. *PLoS Med*. 2005; 2(5):e141. [PubMed: 15916472]

89. Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry*. 2003; 60(12):1187–92. [PubMed: 14662550]
90. McGuffin P, Katz R, Watkins S, Rutherford J. A hospital-based twin register of the heritability of DSM-IV unipolar depression. *Arch Gen Psychiatry*. 1996; 53(2):129–36. [PubMed: 8629888]
91. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics*. 2003; 19(1):149–50. [PubMed: 12499305]

Figure 1. Relationship among power, GRR (multiplicative inheritance) and sample size

The graphs show expected power (91) for a disease with 1% population prevalence ($p = 5 \times 10^{-8}$), depending on minor (less frequent) allele frequency of the tested SNP, sample size (assuming the N of cases shown in the graph legend, and the same N of controls (power is similar for the same N of case-parent trios), and *genotypic relative risk* (GRR), which is the ratio of the risk of disease to carriers of a particular genotype vs. non-carriers (thus, if GRR is 1.2, risk is increased by 20%). The calculations assume indirect association between a tested SNP allele and a risk allele at a correlation (r^2) of 0.8, so that the effective sample sizes are approximately 80% of those shown. A sample of 8,000 cases and 8,000 controls will miss most associated alleles that confer much less than a 20% increase in risk ($GRR \ll 1.2$), whereas 20,000/20,000 would detect most associated alleles with $GRR = 1.12$ and frequency $> 15\text{-}20\%$. Factors that affect power include:

- **GRR.** Power increases with GRR.
- **Allele frequency and LD.** Power increases with the minor allele frequency of the associated SNP and with stronger LD between than SNP and an untested risk allele.
- **Mode of transmission.** Power is greater for dominant and multiplicative (log additive) genetic effects, and less for recessive effects (particularly for rare alleles).
- **Selection of controls.** For diseases with higher prevalence (e.g., $> 5\%$), power increases if controls with the disorder/trait of interest are excluded.(40)
- **Technical artifacts** of all kinds can reduce power.

Table 1

Definitions of terms

Term	Definition
Heritability	Proportion of the variance of a phenotype (disease, trait) that is due to genes, estimated from risks to twins and other relatives
Mendelian disease	Caused by a (usually rare) change (“mutation”) in DNA sequence on one (dominant) or both (recessive) of an individual’s pair of chromosomes
Complex disease	Caused by an interaction of multiple genetic and/or environmental factors
Single nucleotide polymorphism (SNP) [†]	Specific position (among 3.2 billion in the genome) where chromosomes carry different nucleic acids. $\approx$ 11-15 million SNPs (estimated) with frequency $\geq$ 1%. $\approx$ 4 million are catalogued by the HapMap project.
Common SNPs	$\geq$ 5% frequency. $\approx$ 10 million in the genome, $\approx$ 2.8 million on the current HapMap. These SNPs are targeted by GWAS.
Rare variants (rare SNPs) [†]	$<$ 1% frequency, many of them very rare. Rarer SNPs in protein-coding regions tend to be more harmful (frequency constrained by selection).
Copy number variant (CNV)	Chromosomal segment where DNA has been deleted or duplicated. Other structural variants include inversions and translocations.
Common disease-common variant hypothesis (CDCV)	Some of the genetic risk to common diseases is due to common SNPs.
Multiple rare variant hypothesis (MRV)	Some of the genetic risk to common disease is due to many different different rare SNPs, especially in protein coding or gene regulatory regions.
Linkage disequilibrium (LD) between SNPs	Correlation between two SNPs that are close together (an allele of one SNP is usually inherited with a specific allele from the other). LD makes GWAS possible: a subset of common SNPs gives information about most of them.
Genomewide association study (GWAS)	A systematic search common SNPs that influence a disease or trait, using a genomewide SNP array for typing a cohort of individuals. Current arrays also provide information about CNVs.
Genomewide SNP chip (array)	A system for assaying 300,000-1,000,000 SNPs for an individual subject, using an array of bead-based or hybridization assays on a glass slide.

[†]The term “SNP” is sometimes reserved for single-position variants with frequency $\geq$ 1% (i.e., found on at least 1% of chromosomes in a population). For variants with $<$ 1% frequency, the terms “rare variants” and “rare SNPs” are both in use, although “variants” could also refer to other types of sequence changes.

Table 2

Timeline of positional genetic methods from linkage to GWAS

Year	Development	Comment
1980	Proposal to create a genome-wide map of DNA markers for human linkage analysis(5)	Following the discovery of restriction fragment length polymorphism (RFLP) markers, it was proposed that once RFLPs throughout the genome were available, it would be possible to search any genomic region, or the entire human genome, for evidence of genetic linkage.
1983	Linkage mapping and identification of the Huntington's disease gene(6)	The first of the many Mendelian disorders for which genetic linkage was detected followed by identification of specific disease mutations in the linkage region.
1987	First human linkage map(7)	The first genome-wide map of ~ 400 RFLPs ushered in the era of genome-wide linkage studies (GWLS). RFLPs were supplanted by short tandem repeat markers and then SNPs.
1993	First genome-wide linkage study (GWLS) of a psychiatric disorder(8)	Psychiatric GWLS (catalogued at https://slep.unc.edu) produced some convergent linkage evidence, but no definitive evidence for susceptibility genes.
1996	Common disease - common variant (CDCV) hypothesis(9)	The HapMap project grew out of the need to develop a dense set of genetic markers to test this hypothesis.
2001	Draft of the complete human genome sequence(10)	The genome sequence set the stage for all future progress. It stimulated critical advances in genomic sequencing technology and set a new standard of immediate public release of government-supported genomic research data.
2002-2007	International HapMap project(11, 12) (www.hapmap.org)	The project discovered and genotyped (in 270 individuals from three populations) 1.3 million SNPs in Phase I plus 2.1 million in Phase II -- ~ 25-35% of common SNPs in these populations), providing good genome-wide coverage. It spurred advances in SNP assays, making GWAS possible. "HapMap III" provided genotypes in an expanded dataset for the Illumina 1M and Affymetrix 6.0 (900K) SNP sets.
2002	First published GWAS(13)	This study of myocardial infarction used few SNPs (65,761) and cases (94) by current standards.
2005-2007	Availability of high-throughput array-based SNP assays	Affymetrix and Illumina arrays became available, initially with ~ 100,000 SNPs, and currently with up to ~ 1 million SNPs per array plus additional probes for analysis of copy number. These have made it possible to carry out GWAS for many diseases and samples.
2005	First year with multiple GWAS publications	The first small studies using denser SNP sets produced strong associations for macular degeneration(1) and Crohn's disease(14), demonstrating the feasibility and power of GWAS.
2007	Initiation of the 1000 Genomes project (www.1000genomes.org)	This project aims to extend the HapMap to all SNPs with 1% frequency in diverse populations, functional SNPs of lower frequencies, and sequence-level data on structural variants, utilizing multiple high-throughput sequencing technologies.

Table 3

GWAS study design issues and requirements

Issue	Requirement	Comment
Phenotype	Well-defined, adequately heritable disorder (e.g., schizophrenia) or trait (e.g., high-density cholesterol level or neuroticism score).	Power depends on the frequency and effect size for individual variants, not overall heritability
Sample type	Ill cases and controls; or subjects with a range of trait scores (e.g., highest and lowest); or cases and their parents or other relatives.	Cases/controls have more power per subject, but are prone to mis-match biases (e.g., ancestry)
Controls	Match for ancestry, other relevant attributes, e.g., age for an Alzheimer's study, or environmental exposures (e.g., "ever smoked" for a study of nicotine dependence).	For more common disorders, controls with the disorder may be excluded to avoid false negative results (40)
Sample size	Depends on the actual frequency and genetic effect of risk variants in the sample.	Samples up to tens of thousands of subjects have proven useful, but some common risk variants cannot feasibly be detected
SNPs	300,000-1,000,000 common SNPs, depending on ancestry of the sample	Goal is direct or indirect assay of 80% of HapMap II common SNPs with correlation (r^2) ≥ 0.8
Multiple testing	P-value correction for multiple, partially correlated genotyped SNPs, plus imputed data for all HapMap SNPs to permit cross-study comparison and meta-analysis (40, 41)	Genomewide significance threshold $\sim 5 \times 10^{-8}$ (42-44)
Population substructure	World populations differ in frequencies of many SNPs. Case-control ancestry differences can create false positive and negative results.	Match cases/controls for ancestry; apply statistical correction for population differences (38)
Data management	Billions of datapoints to manage.	Requires powerful computers or computer clusters and software (76)
Quality control (QC)	Extensive QC analyses are required to exclude poorly-performing SNPs and DNA specimens, identify duplicate or closely-related specimens, and more subtle assay and sample problems.	Without adequate QC, spurious highly "significant" findings are common.
Detection of CNVs	Computational methods to detect copy number change from intensities of fluorescent labels in assays; additional non-polymorphic assays can be added to improve CNV detection.	CNV detection is less specific, sensitive or accurate than SNP genotype detection. Biological confirmation needed.

Table 4

Significant GWAS findings for non-psychiatric disorders

Type of disease or trait	Unique findings with $p < 10^{-8}$	N of disorders or traits
Autoimmune	12	3
Bone density	10	1
Cancer	37	8
Cardiovascular	5	4
Diabetes - type I	10	1
Diabetes - type II	10	1
Gastrointestinal	25	5
Lipid levels	13	3
Neurological	9	6
Physical traits	28	7
Plasma values	22	10
Other	19	10
Totals	200	59

Data are summarized from <http://www.genome.gov/GWASTudies/> (November 15, 2008). There is no definitive p-value threshold that predicts true positive GWAS findings. Interpretation rests on consistency of replication and/or meta-analysis of cumulative data. A p-value threshold of 5×10^{-8} has been used throughout this review, based on three estimates that assumed that all common SNPs have been tested (42-44), but other thresholds can be defended. Other approaches include False Discovery Rate (45) or Bayes' Factor (41, 46) criteria.

Shown for each category is the number of distinct *findings* (defined here as one or more SNPs in a single chromosomal band, for a specific disease or trait) with $p < 5 \times 10^{-8}$, counting only once those findings reported more than once. Of the 200 findings, 95 had $p < 10^{-12}$, and 58 had $p < 10^{-15}$. Some SNPs or regions have produced findings for different disorders or traits (see text). In many cases, there are additional studies or meta-analyses (not included in this tabulation of GWAS reports) that contain additional findings or updated significance levels.

"Physical traits" includes non-disease traits such as hair and eye color and height. "Plasma values" includes studies of potentially disease-related values (other than lipids) such as C-reactive protein, glucose and IgE. "Other" includes studies related to eye, skin or pulmonary diseases, obesity-related traits, aging, and other traits.

Table 5

Published genomewide association studies of psychiatric disorders

First author, year	Disorder	Initial sample (cases / controls)	Additional information	Genomewide significant findings
Studies of association to SNP genotypes (individual genotyping)				
WTCCC 2007 (41)	BD	1868 / 2938 (UK)		--
Sklar 2008 (59)	BD	1461 / 2008 (US and UK, STEP-UCL)	Replic: 409 US trios, 365 / 351 (Scottish)	--
Ferreira 2008 (60)	BD	4387 / 6209	WTCCC + Sklar (see Ns above) + ED-DUB-STEP2 (1098 / 1267)	P = 9.1 × 10 ⁻⁹ , ANK3 gene (OR = 1.45, AAF = 0.053)
Lenz 2007 (61)	SCZ	178 / 144 (US)		--
Sullivan 2008 (62)	SCZ	738 / 733 (US)	Multiple ancestries	--
O'Donovan 2008 (63)	SCZ	479 / 2937 (UK) + 1865 WTCCC BD cases	Replic: 6829 / 9897 (UK, Eur, US, Aust, Japan, Israel)	With BD included: P = 9.96 × 10 ⁻⁹ , ZNF804A gene (OR = 1.12, AAF = 0.59)
Studies of association to copy number variants				
Walsh 2008 (57)	SCZ	150 / 268	Replic: 83 COS + parents	P = 0.0008, ↑ novel CNVs in cases (15%) vs. controls (5%) (P = 0.03 in COS)
Xu 2008 (58)	SCZ	152 / 159	Sporadic cases	P = 0.00078, ↑ non-inherited CNVs in sporadic cases (9.9%) vs. controls (1.26%)
International Schiz. Consortium 2008 (56)	SCZ	3381 / 3191	--	P = 3 × 10 ⁻⁵ , ↑ CNVs (<1% freq. >100Kb) in cases (1.14/subject) vs. controls (0.99); GWS CNVs: 1q21.1, 22q11.2, 15q13.3
Stefansson 2008 (55)	SCZ	1433 / 33350	Replic: 3285 / 7951	GWS CNVs: 1q21.1, 22q11.2, 15q11.2, 15q13.3
Sebat 2007 (64)	Autism	118 (sporadic) / 196	Some controls from autism families; some AGRE families	↑ <i>de novo</i> CNVs in cases (10%) vs. controls (1%) (note: >1 control per family)
Kumar 2008 (65)	Autism	180 / 372	Replic: 532 / 465	P = 0.044 (uncorrected), ↑ 16p.11.2 deletions in cases (0.6%) vs. controls (0%)
Marshall 2008 (66)	Autism	427 / 500	Replic: 1152 additional controls	GWS ↑ <i>de novo</i> CNVs in cases (7%) vs. controls (1%), ↑ 16p.11.2 deletions in cases (~1%) vs. controls (0%) (P = 0.002).
Weiss 2008 (67)	Autism	751 multiplex AGRE families (1441 cases) + 2814 controls	Replic: 512 / 434, 299 / 18834	↑ 16p.11.2 CNVs in cases (1.1%) vs. controls (0.05%), signif in all 3 samples
Christian 2008 (68)	Autism	397 / 372	Cases are from AGRE families	11.6% of cases had a CNV unique to cases

BD = Bipolar disorder. SCZ = schizophrenia. COS = childhood-onset schizophrenia. Replic = replication sample (for best results). Numbers separated by “/” indicate case / control sample sizes. AGRE = Autism Genetics Resource Exchange (the CNV studies of Sebat, Weiss and Christian all used some families from the AGRE repository, and so are not entirely independent). GWS = genomewide significant evidence for association. OR = odds ratio. AAF = frequency of the associated allele in controls. SNP studies all used SNP arrays with 500,000 SNPs (Affymetrix 500K or 5.0) or 900,000 SNPs (Affymetrix 6.0). Some studies used more than one type of array. CNV studies used array-based comparative genomic hybridization and/or GWAS SNP arrays (Affymetrix or Illumina), with additional confirmation of some or all results with additional method (quantitative PCR, karyotyping and other methods). Studies using pooled genotyping are not included but are cited in the text.

Table 6

Summary of PGC GWAS samples and characteristics of studied disorders

Disorder	Samples	Cases	Controls	Trios or Families	Prevalence	Heritability
Attention deficit hyperactivity disorder	6	1,418	0	2,443	4-12%	70-80%
Autism	6	652	6,000	4,661	0.3-0.6%	90-100%
Bipolar disorder	10	7,075	10,559	0	0.3-1.5%	73-93%
Major depressive disorder	9	12,926	9,618	0	5-18%	31-42% [†]
Schizophrenia	11	9,588	13,500	650	0.2-1.1%	73-90%
Totals	42	31,659	26,945	7,772		

Shown are expected combined sample sizes for meta-analysis of GWAS data by the Psychiatric GWAS Consortium by the end of 2008 (110 investigators from 54 institutions worldwide). Only European-ancestry subjects are shown here; a small number of African-American samples are also available for schizophrenia and bipolar disorder.

Note that the cases are all independent (although independence will be tested using genotypes). However the total N of controls is less than the sum across disorders, because cross-disorder overlap in control groups is excluded from the sum. Within each disorder, only non-overlapping controls are counted. The column for "Trios or families" includes a sample of multiply-affected families for schizophrenia, and trio or sib-pair families (with parents) for ADHD and autism.

References for prevalence and heritability are: ADHD (77, 78) (79), autism (80) (81), bipolar disorder (82) (83), major depressive disorder (84) (85) (86) (87), and schizophrenia (88) (89).

[†]For major depression, higher estimates have been obtained in clinical samples (90) or using repeated interviews (86).