



Published in final edited form as:

*Curr Opin Psychiatry*. 2012 March ; 25(2): 76–82. doi:10.1097/YCO.0b013e32835035dd.

## Genome-wide association studies (GWAS) of schizophrenia: does bigger lead to better results?

Sarah E. Bergen, PhD<sup>1,2,3</sup> and Tracey L. Petryshen, PhD<sup>1,2,3</sup>

<sup>1</sup>Psychiatric and Neurodevelopmental Genetics Unit, Center for Human Genetics Research and Department of Psychiatry, Massachusetts General Hospital, Boston, MA

<sup>2</sup>Department of Psychiatry, Harvard Medical School, Boston, MA

<sup>3</sup>Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA

### Abstract

**Purpose of review**—Numerous genome-wide association studies (GWAS) of schizophrenia have been published in the past six years, with a number of key reports published in the last year. The studies have evolved in scale from small individual samples to large collaborative endeavors. This review aims to critically assess whether the results have improved as the sample size and scale of genetic association studies has grown.

**Recent findings**—Genomic genotyping and increasing sample sizes for schizophrenia association studies has led to parallel increases in the number of risk genes discovered with high statistical confidence. Nearly 20 genes or loci have surpassed the genome-wide significance threshold ( $p = 5 \times 10^{-8}$ ) in a single study, and several have been replicated in more than one GWAS.

**Summary**—Identifying the genetic underpinnings of complex diseases offers insight into the etiological mechanisms leading to manifestation of the disease. New and more effective treatments for schizophrenia are desperately needed, and the ability to target the relevant biological processes grows with our understanding of the genes involved. As the size of GWAS samples has increased, more genes have been identified with high confidence that have begun to provide insight into the etiological and pathophysiological foundations of this disorder.

### Keywords

Schizophrenia; GWAS; association; risk gene; genetic

### Introduction

The genetic basis of schizophrenia has long been known through its tendency to run in families, as well as through adoption studies and twin studies [1]. The heritability (genetic contribution to the phenotypic variance observed) is generally accepted to be in the range of 64–81% [2, 3]. Studies attempting to identify causal genes have evolved concomitant with methodological progress. Initial cytogenetic investigations of large (3 MB or greater) chromosomal deletions, duplications, and rearrangements were followed by linkage studies to assess polymorphisms co-segregating with disease in families, albeit typically with sparse

---

Corresponding author: Dr. Tracey Petryshen, Center for Human Genetic Research, 185 Cambridge Street 6<sup>th</sup> floor, Boston MA 02114, Tel: 617-726-4960, petryshen@chgr.mgh.harvard.edu.

Conflicts of Interest

The authors have no conflicts of interest to declare.

markers that only implicate large chromosomal regions. Candidate gene association studies have implicated myriad putative risk genes; however, these studies are hampered by their single gene focus, reliance on incomplete biological information, and limited gene annotation for much of the genome. Inconsistent results raise questions as to the validity of candidate gene findings, and while inadequate power is one possible reason for intermittent detection of a true genetic association, candidate genes have not been strongly supported by GWAS [4, 5].

It has been only six years since the first schizophrenia GWAS was reported [6]. This short time span has witnessed great strides in understanding the genetic foundations of this disorder, with a number of key findings in the past year. The success of GWAS is often gauged by the number of genes discovered for a disease, but this is only one aspect of what this method has to offer. Associated genes can offer insight into the pathophysiological mechanisms giving rise to the disease phenotype and subsequently yield potential therapeutic targets. Even one GWAS finding leading to an effective treatment is arguably a resounding success. Additionally, while specifically associated single nucleotide polymorphisms (SNPs) generally explain only a small fraction of the genetic variation for a disorder, considering aggregated genetic variation demonstrating association below the typical threshold for genome-wide significance ( $p = 5 \times 10^{-8}$ ) [7] can account for a much larger portion. For schizophrenia, 32–36% of the genetic liability can be explained in this manner [8, 9\*\*]. These sub-threshold associations offer insight into the genetic architecture of complex genetic disorders, and in the case of schizophrenia, hundreds or even thousands of variants with low effect sizes appear to confer risk. Identification of these risk variants necessitates genome-wide genotyping and very large samples.

The ever growing number of subjects included in schizophrenia GWAS and the resultant findings invite a continuous process of evaluation. Several prior reviews specifically address or encompass schizophrenia genetics [10–14]. This review focuses on recent large GWAS and their results, and discusses the implications of robust genetic findings for our understanding of the etiological and pathophysiological changes giving rise to schizophrenia.

### **Genome-wide association studies (GWAS)**

GWAS offer an unbiased assessment of variation across the entire genome, with the capacity to implicate specific variants in disease risk. A genotyped SNP demonstrating disease association may be causal or, more likely, is inherited with and serves as a marker for a causal genetic variant (i.e. in linkage disequilibrium). The genomic region implicated is far more circumscribed for GWAS than for linkage studies, and on a larger scale than candidate gene studies.

In contrast to monogenic disorders in which a single gene with complete or high penetrance causes the disease, the individual genetic effect sizes for nearly all complex genetic disorders are relatively small. The vast majority of genetic disorders are polygenic, involving multiple genes and multiple alleles within these genes. Targeted selection of candidate genes is hindered by incomplete genomic annotation, making the unbiased assessment of genome-wide variation an essential advance in the search for causal disease variants for complex disorders.

Within the GWAS framework, a few distinct study designs exist. Pooled genotyping of many DNA samples offers a cost benefit but diminished ability to detect rare alleles. This has given way to genotyping individual samples following waning genotyping costs and additional analytic possibilities involving haplotypes or epistasis. Individual genotyping also

allows for the potential detection of rare copy number variants (CNVs; e.g., deletions, duplications).

### Sample size impact

Although it is tempting to assume that a high disease heritability offers easier gene-finding capacity, these are actually unrelated. High heritability indicates large contribution to the disease by genes, as opposed to environmental or other non-genetic factors, but does not dictate large effect sizes for individual genes. Diseases with similar heritability can differ dramatically in terms of genetic architecture, and the ability to discover risk variants is dependent upon their effect size, frequency in the population and, of course, the sample power. It is intuitively obvious that increasing sample sizes should enhance power to detect associations [15], and this has been shown for schizophrenia [12] as well as for other diseases and phenotypes. Compared to some traits, such as body mass index assessed in nearly 250,000 people with 42 identified loci [16] or height studied in over 180,000 subjects resulting in 180 significant loci [17], the current numbers of cases in schizophrenia GWAS are low (Table 1). Bolstering sample numbers is the surest way to reveal additional associations.

Recognizing that the number of subjects is the only determinant of sample power under experimental control, many consortia have emerged to combine samples to identify genetic variation contributing to a particular disorder. With respect to schizophrenia, the International Schizophrenia Consortium (ISC), Molecular Genetics of Schizophrenia (MGS), and SGENE are three consortia that have brought together sample collections from multiple sites and additionally collaborated with each other [9\*\*, 23\*\*, 24\*\*]. The Schizophrenia Psychiatric GWAS Consortium (PGC) has been the largest collaboration for this disorder to date, with an initial collection of 21,856 individuals from 17 samples and a replication collection of 29,839 subjects from 19 samples [30\*\*]. Future GWAS of even greater numbers of subjects have been planned.

GWAS on separate samples are often folded into collaborative efforts involving numerous merged samples. Meta-analysis (combining results) and mega-analysis (combining data) yield greater power to detect disease associations with the increased number of subjects. Merging samples, however, does have potential drawbacks. Mixing populations could dilute association signals if recombination has separated a causal variant from a genotyped marker in some of the populations. Similarly, while much of the known genomic variation is observed across many global populations, population-specific variation is substantial and associations could be masked in a mixed sample. Furthermore, the number of directly genotyped SNPs common to all studies may be very low, necessitating imputation of variants not genotyped using a reference panel. Since it is difficult to impute genetic variants located on the sex chromosomes and mitochondrial DNA, they are often removed from association analyses, potentially excluding relevant variation. Imputed SNPs significantly associated with disease should be validated by direct genotyping, adding to research cost and time. In addition to genetic concerns, differences in diagnostic assessments and ascertainment procedures across constituent sites can undercut power gains from multi-sample analyses. While these concerns should be taken into consideration, the benefits of combining samples certainly outweigh the possible detractions.

### Schizophrenia GWAS

Nearly 20 GWAS of schizophrenia were published as of the end of 2011 (Table 1). Although most of the early studies seemed large at the time, they were underpowered to detect genome-wide associations when appropriately correcting for multiple tests, given the effect size of risk genes is generally below 1.2 - much lower than expected. A trio of

publications by the ISC, SGENE, and MGS consortia in 2009 ushered in the initial genome-wide significant findings that implicated the major histocompatibility complex (MHC) region, transcription factor 4 (*TCF4*) and neurogranin (*NRGN*) [9\*\*, 23\*\*, 24\*\*]. *TCF4*, involved in nervous system development, has twice demonstrated subsequent association [29\*, 30\*\*] and the 11q24.2 locus containing *NRGN*, a protein kinase substrate that binds calmodulin, was also significant in the in the PGC primary sample mega-analysis [30\*\*]. Steinberg et al (2011) added *vaccinia-related kinase 2* (*VRK2*), a widely-expressed protein kinase highly expressed in actively dividing cells, in addition to replicating the MHC region and *TCF4* associations [29\*]. The largest schizophrenia GWAS, reported by the PGC, identified five novel loci: 1p23.3 (*MIR137*), 2q32.3 (*PCGEM1*), 8p23.2 (*CSMD1*), 8q21.3 (*MMP16*), and 10q24.32-q24.33 (*CNNM2/NT5C2*) [30\*\*]. Recently, two groups published the first significant hits for schizophrenia GWAS in non-Western populations. Specifically, Yue et al (2011) reported a novel association at 11p11.2 and replicated an association in the MHC region [31\*], and Shi et al (2011) found two new loci at 8p12 and 1q24.2 [32\*], both in Chinese samples. Consistent findings are now beginning to emerge across studies, but since many reports contain overlapping samples, the findings are not completely independent. Several large, rare CNVs have also been associated with schizophrenia, although they are only present in a small proportion of cases [33]. A comprehensive overview of CNVs in schizophrenia is beyond the scope of this review, and has been thoroughly covered elsewhere [14, 34–38].

Although there is a small chance that SNPs exceeding a genome-wide significance p-value of  $5 \times 10^{-8}$  might be false positives (Type I error), the chance of false negatives (Type II error) is almost certain, with many legitimate causal loci likely falling short of this statistical threshold. For example, the initial GWAS report suggesting involvement of *ZNF804A* in schizophrenia yielded a p-value just shy of significance,  $1.61 \times 10^{-7}$  [20], and ultimately surpassed this threshold through meta-analysis of several samples [27\*]. This illustrates one example of how larger samples may reveal significant association for tentatively implicated loci, and many more are sure to follow.

### GWAS Results Provide Insight into Disease Pathophysiology

In addition to identifying specific genetic risk loci for complex genetic disorders, GWAS offer insights into the biological mechanisms giving rise to the disease. Age-related macular degeneration (AMD) is one such success story, in that GWAS identified several genes involved in complement-mediated inflammation [39], a pathway previously unsuspected in this disease and now targeted for therapeutics [40]. Similarly, GWAS of inflammatory bowel disease (IBD) identified SNPs impacting the autophagy pathway and revealed this process as central to the etiology of this disease [41, 42].

Explorations of how GWAS-identified loci mediate schizophrenia risk have begun, but are, for the most part, still in their infancy. The microRNA *MIR137* and four of its gene targets (*TCF4*, *CACNA1C*, *CSMD1*, and *C10orf26*) demonstrated genome-wide significance in either the schizophrenia or combined schizophrenia-bipolar disorder PGC analyses [30\*\*], implicating a shared biological pathway in schizophrenia. *MIR137* functions in regulating adult neurogenesis and neuronal maturation [43–45], suggesting it may contribute to schizophrenia via perturbed developmental processes. Known functions of its four associated targets, such as the expression of *CSMD1* in the nerve growth cone [46], initiation of neuronal differentiation by *TCF4* [47], and the role of voltage-gated calcium channel genes such as *CACNA1C* in interneuron development [48], support this idea. Although the most replicated schizophrenia GWAS association is the MHC locus, its complex composition of many genes functioning in immunity, neurodevelopment, synaptic plasticity, and other processes has thus far hindered investigations relating the association signal in this region to disease pathophysiology.

As risk alleles for schizophrenia are discovered with increasing certainty, the identification of subjects carrying these alleles opens up new avenues of research to elucidate the pathophysiological mechanisms. For example, “deep phenotyping” of subjects carrying risk alleles may connect variation in the clinical phenotype with genetic subtypes. Furthermore, since the associated variants may simply be proxies for near-by causal loci, sequencing the surrounding gene regions in risk allele carriers will yield single nucleotide resolution capable of revealing rare variants, insertions, and deletions, and may pinpoint variants conferring protein-coding or other functional changes. The cellular consequences of these genetic changes can be explored through generation of induced pluripotent stem (iPS) cells [49], and post-mortem brain studies of mRNA expression or protein alterations in subjects with and without the risk alleles. Generating cellular and mouse models with the risk variants or haplotypes and characterizing cellular, behavioral, and other processes also promises to help link the GWAS risk genes to disease mechanisms.

Endophenotypes such as heritable impairments in cognition and prepulse inhibition, and neuroanatomical and functional abnormalities, also offer a starting point to delineate the functional relevance of risk genes, since they are thought to be more directly linked to the genes and less influenced by environmental factors that can influence clinical features. In schizophrenia patients, *ZNF804A* and *TCF4* risk allele carriers are reported to have better cognitive functioning than non-carriers, indicating possible subgroups with relatively spared cognition [50, 51]. The risk variant in *ZNF804A* has also been associated with impaired connectivity of the prefrontal cortex across hemispheres and with the hippocampus in a dosage-dependent manner [52]. As more risk loci are identified, clearer biological patterns will emerge.

### Beyond GWAS - Additional Analyses

In addition to gene discovery, numerous other investigations predicated on GWAS data also justify the collection large samples. Heterogeneity in the clinical presentation of schizophrenia has strongly suggested variability in the etiology of this disease, but most efforts to link genes to symptom profiles have been small in scale. Analyses exploring the genetic relationship to symptom dimensions, age of onset, illness severity, and other features can now be conducted in an unbiased, genome-wide manner in samples well-powered to yield high-confidence results. Also, all models of the genetic architecture of schizophrenia consistent with current knowledge involve the non-additive interaction of genetic variants (i.e. epistasis) [8]. Because the number of tests is effectively multiplied, many more subjects are needed to have adequate power to test epistatic interactions than single variants, and epistasis testing on a genomic scale becomes more tractable with greater numbers of subjects. Furthermore, the ways in which environmental exposures relate to the genetic predisposition to schizophrenia and its manifestation deserve greater attention, and larger samples will facilitate these studies as well.

With increased confidence in risk locus identification by GWAS, there is greater incentive to invest in studies of the mechanisms by which these loci effect biological changes. In addition to primary DNA sequence changes, epigenetic factors such as methylation at CpG islands or acetylation of histones that regulate expression of risk genes can also be investigated. MicroRNAs, such as the recently associated *MIR137*, and other non-coding RNAs, as well as transcription factors that may play a role in pathogenesis “upstream” of the identified risk loci, also merit further scrutiny [30\*\*]. Certainty in the biological pathways identified through pathway analyses will also increase [12], bringing us closer to connecting findings from genetics and neuroscience.



## Conclusion

Schizophrenia association studies have incorporated genome-wide markers and an increasing number of subjects, leading to identification of numerous novel risk loci with high statistical confidence. As risk genes are discovered, research emphasis can shift from gene-finding to exploring the downstream biological mechanisms resulting in schizophrenia, and ultimately facilitate development of new treatment and prevention options. Going beyond case-control analyses to test for epistatic interactions and aggregation of associations in biological pathways will offer additional insight into the distinct etiologic pathways converging on the schizophrenia phenotype. Although some inroads in connecting genetic findings to their neurobiological sequelae have been made, this will be a long-term endeavor. GWAS have brought us closer than ever to understanding the aberrant processes leading to schizophrenia, and larger samples in conjunction with additional analytic and methodological approaches will continue to further our understanding of this disease.

## Acknowledgments

Supported by NIMH grant number P50 MH080272 and the Stanley Medical Research Institute.

## References

1. Kendler KS, Diehl SR. The genetics of schizophrenia: a current, genetic-epidemiologic perspective. *Schizophr Bull.* 1993; 19(2):261–285. [PubMed: 8322035]
2. Lichtenstein P, Yip BH, Bjork C, et al. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet.* 2009 Jan 17; 373(9659):234–239. [PubMed: 19150704]
3. Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry.* 2003 Dec; 60(12):1187–1192. [PubMed: 14662550]
4. Collins AL, Kim Y, Sklar P, et al. Hypothesis-driven candidate genes for schizophrenia compared to genome-wide association results. *Psychol Med.* 2011 Aug; 19:1–10.
5. Need AC, Ge D, Weale ME, et al. A genome-wide investigation of SNPs and CNVs in schizophrenia. *PLoS Genet.* 2009 Feb; 5(2):e1000373. [PubMed: 19197363]
6. Mah S, Nelson MR, Delisi LE, et al. Identification of the semaphorin receptor PLXNA2 as a candidate for susceptibility to schizophrenia. *Mol Psychiatry.* 2006 May; 11(5):471–478. [PubMed: 16402134]
7. 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 May; 32(4):381–385. [PubMed: 18348202]
8. Wray NR, Visscher PM. Narrowing the boundaries of the genetic architecture of schizophrenia. *Schizophr Bull.* 2010 Jan; 36(1):14–23. [PubMed: 19996148]
9. Purcell SM, Wray NR, Stone JL, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature.* 2009 Aug 6; 460(7256):748–752. [PubMed: 19571811]  
This study by the International Schizophrenia Consortium (ISC) of 3,322 cases and 3,587 controls was one of three publications collaborating to produce the first genome-wide significant finding for schizophrenia. Together, the ISC, MGS, and SGENE samples revealed an association signal localized to the MHC region.
10. O'Donovan MC, Craddock NJ, Owen MJ. Genetics of psychosis; insights from views across the genome. *Hum Genet.* 2009 Jul; 126(1):3–12. [PubMed: 19521722]
11. Girard SL, Xiong L, Dion PA, Rouleau GA. Where are the missing pieces of the schizophrenia genetics puzzle? *Curr Opin Genet Dev.* 2011 Jun; 21(3):310–316. [PubMed: 21277191]
12. Kim Y, Zerwas S, Trace SE, Sullivan PF. Schizophrenia genetics: where next? *Schizophr Bull.* 2011 May; 37(3):456–463. [PubMed: 21505112]

13. Cichon S, Craddock N, Daly M, et al. Genomewide association studies: history, rationale, and prospects for psychiatric disorders. *Am J Psychiatry*. 2009 May; 166(5):540–556. [PubMed: 19339359]
14. Lee KW, Woon PS, Teo YY, Sim K. Genome wide association studies (GWAS) and copy number variation (CNV) studies of the major psychoses: What have we learnt? *Neurosci Biobehav Rev*. 2011 Sep 17.
15. Cohen J. A power primer. *Psychol Bull*. 1992 Jul; 112(1):155–159. [PubMed: 19565683]
16. Speliotes EK, Willer CJ, Berndt SI, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet*. 2010 Nov; 42(11):937–948. [PubMed: 20935630]
17. Lango Allen H, Estrada K, Lettre G, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*. 2010 Oct 14; 467(7317):832–838. [PubMed: 20881960]
18. Lencz T, Morgan TV, Athanasiou M, et al. Converging evidence for a pseudoautosomal cytokine receptor gene locus in schizophrenia. *Mol Psychiatry*. 2007 Jun; 12(6):572–580. [PubMed: 17522711]
19. Sullivan PF, Lin D, Tzeng JY, et al. Genomewide association for schizophrenia in the CATIE study: results of stage 1. *Mol Psychiatry*. 2008 Jun; 13(6):570–584. [PubMed: 18347602]
20. O'Donovan MC, Craddock N, Norton N, et al. Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet*. 2008 Sep; 40(9):1053–1055. [PubMed: 18677311]
21. Shifman S, Johannesson M, Bronstein M, et al. Genome-wide association identifies a common variant in the reelin gene that increases the risk of schizophrenia only in women. *PLoS Genet*. 2008 Feb.4(2):e28. [PubMed: 18282107]
22. Kirov G, Zaharieva I, Georgieva L, et al. A genome-wide association study in 574 schizophrenia trios using DNA pooling. *Mol Psychiatry*. 2009 Aug; 14(8):796–803. [PubMed: 18332876]
23. Shi J, Levinson DF, Duan J, et al. Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature*. 2009 Aug 6; 460(7256):753–757. [PubMed: 19571809] This large Molecular Genetics of Schizophrenia (MGS) sample (3,967 cases, 3,626 controls) combined forces with SGENE and ISC to yield an association in the MHC region – the first significant schizophrenia GWAS finding.
24. Stefansson H, Ophoff RA, Steinberg S, et al. Common variants conferring risk of schizophrenia. *Nature*. 2009 Aug 6; 460(7256):744–747. [PubMed: 19571808] As part of the first set of significant GWAS results, the primary SGENE sample (2,663 cases, 13,498 controls) yielded associations with *TCF4* and *NRGN* in addition to the MHC region when meta-analyzed with the concurrently published ISC and MGS samples as well as an additional replication sample.
25. Chen X, Lee G, Maher BS, et al. GWA study data mining and independent replication identify cardiomyopathy-associated 5 (CMYA5) as a risk gene for schizophrenia. *Mol Psychiatry*. 2010 Sep 14.
26. Athanasiu L, Mattingsdal M, Kahler AK, et al. Gene variants associated with schizophrenia in a Norwegian genome-wide study are replicated in a large European cohort. *J Psychiatr Res*. 2010 Sep; 44(12):748–753. [PubMed: 20185149]
27. Williams HJ, Norton N, Dwyer S, et al. Fine mapping of *ZNF804A* and genome-wide significant evidence for its involvement in schizophrenia and bipolar disorder. *Mol Psychiatry*. 2011 Apr; 16(4):429–441. [PubMed: 20368704] This study was predicated on the O'Donovan et al, 2008 finding of sub-threshold association for *ZNF804A*. The focused replication in 21,274 schizophrenia cases and 38,675 controls led to a strong and significant association for this gene.
28. Ikeda M, Aleksic B, Kinoshita Y, et al. Genome-wide association study of schizophrenia in a Japanese population. *Biol Psychiatry*. 2011 Mar 1; 69(5):472–478. [PubMed: 20832056] Notable as the only GWAS of schizophrenia in a Japanese population and the first in non-Western subjects, but no significant results emerged.
29. Steinberg S, de Jong S, Andreassen OA, et al. Common variants at *VRK2* and *TCF4* conferring risk of schizophrenia. *Hum Mol Genet*. 2011 Oct 15; 20(20):4076–4081. [PubMed: 21791550] This report followed up on the results from the prior analyses combining SGENE-plus, ISC, and

MGS data. Meta-analysis of some SNPs in a large replication sample led to replicated associations in the MHC region. *NRGN*, and *TCF4*, in addition to a novel *TCF4* association and *VRK2*.

30. Ripke S, Sanders AR, Kendler KS, et al. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet.* 2011; 43(10):969–976. [PubMed: 21926974] The largest schizophrenia GWAS published so far, incorporating 17 samples in the primary analysis (9,394 cases 12,462 controls) and 19 replication samples (8,442 cases and 21,397 controls), which identified the most genome-wide significant loci. The microRNA *MIR137* and its targets were particularly interesting new associations from this study.
31. Yue WH, Wang HF, Sun LD, et al. Genome-wide association study identifies a susceptibility locus for schizophrenia in Han Chinese at 11p11.2. *Nat Genet.* 2011 Oct 30. A tie for the first schizophrenia GWAS in a Chinese sample, this study replicated the MHC region association and identified a new significant locus at 11p11.2.
32. Shi Y, Li Z, Xu Q, et al. Common variants on 8p12 and 1q24.2 confer risk of schizophrenia. *Nat Genet.* 2011 Oct 30. Published in the same issue as the only other schizophrenia GWAS with Chinese subjects, this study identified two novel significant loci.
33. Sebat J, Levy DL, McCarthy SE. Rare structural variants in schizophrenia: one disorder, multiple mutations, one mutation, multiple disorders. *Trends Genet.* 2009 Dec; 25(12):528–535. [PubMed: 19883952]
34. Bassett AS, Scherer SW, Brzustowicz LM. Copy number variations in schizophrenia: critical review and new perspectives on concepts of genetics and disease. *Am J Psychiatry.* 2010 Aug; 167(8):899–914. [PubMed: 20439386]
35. Kirov G. The role of copy number variation in schizophrenia. *Expert Rev Neurother.* 2010 Jan; 10(1):25–32. [PubMed: 20021318]
36. Tam GW, Redon R, Carter NP, Grant SG. The role of DNA copy number variation in schizophrenia. *Biol Psychiatry.* 2009 Dec 1; 66(11):1005–1012. [PubMed: 19748074]
37. Williams HJ, Owen MJ, O'Donovan MC. Schizophrenia genetics: new insights from new approaches. *Br Med Bull.* 2009; 91:61–74. [PubMed: 19443537]
38. Owen MJ, Williams HJ, O'Donovan MC. Schizophrenia genetics: advancing on two fronts. *Curr Opin Genet Dev.* 2009 Jun; 19(3):266–270. [PubMed: 19345090]
39. Klein RJ, Zeiss C, Chew EY, et al. Complement factor H polymorphism in age-related macular degeneration. *Science.* 2005 Apr 15; 308(5720):385–389. [PubMed: 15761122]
40. Rohrer B, Long Q, Coughlin B, et al. A targeted inhibitor of the alternative complement pathway reduces angiogenesis in a mouse model of age-related macular degeneration. *Invest Ophthalmol Vis Sci.* 2009 Jul; 50(7):3056–3064. [PubMed: 19264882]
41. Rioux JD, Xavier RJ, Taylor KD, et al. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat Genet.* 2007 May; 39(5):596–604. [PubMed: 17435756]
42. Fiocchi. Susceptibility genes and overall pathogenesis of inflammatory bowel disease: where do we stand? *Dig Dis.* 2009; 27(3):226–235. [PubMed: 19786745]
43. Szulwach KE, Li X, Smrt RD, et al. Cross talk between microRNA and epigenetic regulation in adult neurogenesis. *J Cell Biol.* 2010 Apr 5; 189(1):127–141. [PubMed: 20368621]
44. Silber J, Lim DA, Petritsch C, et al. miR-124 and miR-137 inhibit proliferation of glioblastoma multiforme cells and induce differentiation of brain tumor stem cells. *BMC Med.* 2008; 6:14. [PubMed: 18577219]
45. Smrt RD, Szulwach KE, Pfeiffer RL, et al. MicroRNA miR-137 regulates neuronal maturation by targeting ubiquitin ligase mind bomb-1. *Stem Cells.* 2010 Jun; 28(6):1060–1070. [PubMed: 20506192]
46. Kraus DM, Elliott GS, Chute H, et al. CSMD1 is a novel multiple domain complement-regulatory protein highly expressed in the central nervous system and epithelial tissues. *J Immunol.* 2006 Apr 1; 176(7):4419–4430. [PubMed: 16547280]
47. Flora A, Garcia JJ, Thaller C, Zoghbi HY. The E-protein Tcf4 interacts with Math1 to regulate differentiation of a specific subset of neuronal progenitors. *Proc Natl Acad Sci U S A.* 2007 Sep 25; 104(39):15382–15387. [PubMed: 17878293]



48. Jiang M, Swann JW. A role for L-type calcium channels in the maturation of parvalbumin-containing hippocampal interneurons. *Neuroscience*. 2005; 135(3):839–850. [PubMed: 16154277]
49. Brennand KJ, Simone A, Jou J, et al. Modelling schizophrenia using human induced pluripotent stem cells. *Nature*. 2011 May 12; 473(7346):221–225. [PubMed: 21490598]
50. Walters JT, Corvin A, Owen MJ, et al. Psychosis susceptibility gene ZNF804A and cognitive performance in schizophrenia. *Arch Gen Psychiatry*. 2010 Jul; 67(7):692–700. [PubMed: 20603450]
51. Lennertz L, Quednow BB, Benninghoff J, et al. Impact of TCF4 on the genetics of schizophrenia. *Eur Arch Psychiatry Clin Neurosci*. 2011 Nov; 261(Suppl 2):161–165.
52. Esslinger C, Walter H, Kirsch P, et al. Neural mechanisms of a genome-wide supported psychosis variant. *Science*. 2009 May 1.324(5927):605. [PubMed: 19407193]

### Key points

- Schizophrenia GWAS of large subject samples have consistently led to increasing numbers of genetic risk factors with high statistical confidence, and larger samples should allow for the identification of even more risk genes.
- Larger samples will confer greater power for tests of epistatic contributions to disease risk, facilitate discovery of genetic relationships with clinical variation, and enhance certainty in the results of pathway analyses.
- Functional studies in humans and in cellular and animal models of risk genes and specific sequence variants are beginning to connect genetic findings to the neurobiological mechanisms that are abnormal in schizophrenia.
- Next-generation sequencing technologies will eventually overtake SNP arrays as the primary genotyping methodology and advance our understanding of rare variation and population-specific influences on disease risk.

**Table 1**  
Schizophrenia GWAS reports and genes/loci significantly associated with risk at  $p < 5 \times 10^{-8}$

First Author (year)	# Markers	primary sample		replication sample		Locus	Gene/Region	p value	Reference
		# Cases	# Controls	# Cases	# Controls				
Mah (2006) <sup>†</sup>	25,494	320	325	--	--	--	--	--	6
Lenz (2007) <sup>*</sup>	439,511	178	144	--	--	--	--	--	18
Sullivan (2008) <sup>*</sup>	492,900	738	733	--	--	--	--	--	19
O'Donovan (2008) <sup>*</sup>	362,532	479	2,937	6,666	9,897	--	--	--	20
Shifman (2008) <sup>†</sup>	510,552	660	1,100	--	--	--	--	--	21
Kirov (2009) <sup>†</sup>	433,680	574	605	--	--	--	--	--	22
Need (2009) <sup>*</sup>	312,565	871	863	1,460	12,995	--	--	--	5
Shi, J (2009) <sup>**</sup>	696,788-843,798	3,967	3,626	5,327	16,424	6p22.1	MHC region	$9.5 \times 10^{-9}$	23**
Stefansson (2009) <sup>**</sup>	314,868	2,663	13,498	4,999	15,555	6p21.3-6p22.1	MHC region	$1.1 \times 10^{-9}$ to $1.4 \times 10^{-12}$	24**
ISC (2009) <sup>**</sup>	739,995	3,322	3,587	--	--	11q24.2	NRGN	$2.4 \times 10^{-9}$	9**
Chen (2010) <sup>*</sup>	446,225-727,905	1,658	1,655	11,380	14,708	18q21.2	TCF4	$4.1 \times 10^{-9}$	25
Athanasou (2010) <sup>*</sup>	572,888	201	305	2,663	13,780	6p21.3-6p22.1	MHC region	$9.5 \times 10^{-9}$	26
Williams (2011) <sup>**</sup>	176	479	2,937	18,945	38,675	2q32.1	ZNF804A	$2.5 \times 10^{-11}$	27*
Ikeda (2011)	297,645	575	564	1,990	5,389	--	--	--	28*
Steinberg (2011) <sup>**</sup>	39	7,946	19,036	10,260	23,500	2p15.1	VRK2	$1.9 \times 10^{-8}$	29*
PGC (2011) <sup>*</sup>	1,252,901	9,394	12,462	8,442	21,397	6p21.3-6p22.1	MHC region	$1.6 \times 10^{-10}$ to $1.4 \times 10^{-13}$	30**
						11q24.2	NRGN	$2.8 \times 10^{-9}$	
						18q21.2	TCF4	$7.8 \times 10^{-9}$	
						18q21.2	TCF4/CCDC68	$4.2 \times 10^{-9}$	
						1p23.3	MIR137	$1.59 \times 10^{-11}$	
						2q32.3	PCGEM1	$4.65 \times 10^{-8}$	

First Author (year)	# Markers		primary sample		replication sample		Locus	Gene/Region	p value	Reference
	Cases	Controls	Cases	Controls	Cases	Controls				
Yue (2011)	546,561	746	1,599	4,027	5,603	5,603	6p21.3-p22.1	TRIM26/MHC region	$2.18 \times 10^{-12}$	31*
							8p23.2	CSMD1	$4.14 \times 10^{-8}$	
							8q21.3	MMP16	$2.75 \times 10^{-8}$	
							10q24.32	CNNM2	$1.82 \times 10^{-9}$	
							10q24.33	NT5C2	$1.11 \times 10^{-8}$	
							18q21.2	CCDC68	$2.60 \times 10^{-10}$	
Shi, Y (2011)	493,203	3,750	6,468	4,383	4,539	4,539	18q21.2	TCF4	$2.35 \times 10^{-8}$ ‡	32*
							11q24.2	STT3A	$2.91 \times 10^{-8}$ ‡	
							11p11.2	TSPAN18	$4.76 \times 10^{-11}$	
							6p21-p22.1	MHC region	$<5.18 \times 10^{-10}$	
	546,561	746	1,599	4,027	5,603	5,603	8p12	LSM1/WHSC1L1	$1.27 \times 10^{-10}$	
	493,203	3,750	6,468	4,383	4,539	4,539	1q24.2	BRP44	$9.5 \times 10^{-9}$	

‡ Utilized pooled genotyping data

\* Overlapping sample with at least one other study

# Genotype data were meta-analyzed and reported together

^ Follow-up study of SNPs reported in previous GWAS

‡ primary sample p value, not significant after replication analysis