Breaking the code: Statistical methods and methodological issues in psychiatric genetics

Stringer, S.

Publication date
2015

Document Version
Final published version

Chapter 2

Genome-wide association analysis in schizophrenia

This chapter will be published as: Sven Stringer, Dorien H. Nieman, René S. Kahn, Eske M. Derks (in press). Genome-wide association analysis in schizophrenia. In Appasani, K. (Ed.) *Genome-Wide Association Studies: From Polymorphism to Personalized Medicine (pp. xx-xx)*. Cambridge: Cambridge University Press

Introduction

Schizophrenia is a severe mental disorder with a typical onset in adolescence or young adulthood. Global lifetime prevalence is about 0.3–0.7%[9]. Symptoms can be divided into positive symptoms (e.g., delusions and hallucinations), negative (deficit) symptoms (e.g., anhedonia, blunted affect, and avolition), and disorganization symptoms (e.g. disorganized speech). In addition, the majority of schizophrenia patients show cognitive dysfunctioning. In general, schizophrenia patients have deficits in most cognitive domains (e.g., attention, memory, and executive functioning) approximately one standard deviation below the normative mean[10]. However, there is no specific cognitive profile that distinguishes schizophrenia patients from patients with other Diagnostic and Statistical Manual of Mental Disorders (DSM)[11] diagnoses. Schizophrenia is often preceded by a prodromal period of months to years in which mild psychotic and other symptoms can occur and psycho-social functioning deteriorates.

A short case example:

Michael is a 20 year old philosophy student who is skipping a lot of classes lately. In the class room he hears his name being whispered by fellow students in the front row, although the distance is too far to be able to hear them. During the breaks he hears other students talk and laugh about him. Sometimes he thinks they are conspiring to kill him especially since he also hears them talking about how they are going to get him when he is alone in his room. He is unable to concentrate on what the professor says in the classroom. It is as if he cannot extract the meaning of what is being said. He has suffered from that problem for several years. His grades have decreased during this period and he will probably drop out from university. He also experiences a feeling of

emptiness, which started years ago. Nothing seems to get through to him. Even the birth of his niece left him cold. He experiences a loss of identity. Lately he sometimes has the feeling that someone else is putting thoughts in his head or moves his limbs outside his own will.

Schizophrenia has remained a mental disorder with an unknown etiology, unchanged prevalence and disabling outcomes for the vast majority of the patients. Sustained recovery occurs in less than 14% within the first five years following a psychotic episode and in an additional 16% in a later phase[12, 13]. Throughout Europe, less than 20% of people with schizophrenia are employed[14]. Various treatments, especially pharmacological, have been tried to improve its disabling lifetime course[15]. Antipsychotic medication reduces the positive symptoms but the negative symptoms often remain, sometimes even worsened by antipsychotic medication. One of the causes for lack of progress in scientific understanding and treatment may be that the DSM-IV category of schizophrenia is very broad and therefore psychopathology is too heterogeneous to find a biological substrate[16]. Wessman et al.[17] performed unsupervised clustering of individuals from Finnish schizophrenia families, based on extensive clinical and neuropsychological data, including Structured Clinical Interview for DSM-IV information[18]. The sample consisted of 904 individuals from 288 families with at least one member with schizophrenia. Wessman et al.[17] found several subgroups. One group was characterized by psychotic and mood symptoms and an association with allelic variants of the *DISC1* gene. The other group showed mainly negative and cognitive symptoms and a strong association to several allelic variants in the *DTNBP1* gene. Other examples include the studies of Derks and colleagues[19-21] and Fanous et al.[22], which suggest that genetic associations may be particularly strong for negative and disorganization symptoms.

Summarizing, modeling phenotypic heterogeneity within schizophrenia may provide increased insight into the biological substrate of the diverse symptoms of schizophrenia.

Twin and family studies

The risk of developing schizophrenia is increased in relatives of schizophrenia patients, indicating that familial or genetic factors influence disease risk. In the early nineties, Gottesman[23] showed that disease probability is more strongly increased in first-degree relatives of schizophrenia patients (6-17%) compared to second-degree relatives (2-6%). In agreement with a large genetic component for schizophrenia, the probability of developing schizophrenia is 48% in the monozygotic twin of a proband with schizophrenia. The fact that disease risk is more strongly increased in relatives, who are genetically more alike, indicates that genetic factors are important in causing schizophrenia. Indeed, a meta-analysis of twin studies indicates that the heritability of liability to schizophrenia is 81%[24]. Since the publication of this meta-analysis, Lichtenstein and colleagues investigated the heritability of schizophrenia based on multi-generation register data including information on psychiatric inpatient admissions in Sweden[3]. This sample comprised over nine million unique individuals, including 35,985 probands with schizophrenia. The heritability of schizophrenia was estimated at 64%, which is lower compared to previous estimates. However, the latter estimate is consistent with the study of Wray and Gottesman[25], who reported a heritability estimate of 67% based on Danish population-based cohort data.

Lichtenstein and colleagues did not limit the Swedish registry study to schizophrenia only, but also aimed to investigate whether schizophrenia

and bipolar disorder share a common genetic etiology. In relatives of schizophrenia probands, the risk to develop bipolar disorder was higher than the risk in relatives of healthy probands. Furthermore, in relatives of probands with bipolar disorder, an increased risk of schizophrenia was reported. Based on the relative risks, Lichtenstein and colleagues concluded that the comorbidity between schizophrenia and bipolar disorder was for a large percentage (63%) due to shared genetic effects.

The genetic architecture of schizophrenia is complex. Early linkage and association studies were challenged by the fact that twin and family data are not in agreement with a single major gene effect[26]. Indeed, the identification of causal genetic variants has been less successful than expected. However, genome-wide association (GWA) studies in large samples of patients with schizophrenia and healthy controls have recently resulted in conclusive evidence for the role of several genetic loci (section 3). These studies have also increased our knowledge about the genetic architecture of this severe disorder (section 4).

Genome-wide association studies

In 2009, the first GWA study on schizophrenia was published[27]. In this study, DNA pooling was used to estimate allele frequencies in 574 schizophrenia patients, their parents, and 605 unaffected controls. Although genetic variants in the genes *CCDC60* and *RBP1* were highlighted, no genome-wide significant findings were reported. Kirov et al.[27] predicted that thousands of cases and controls would have to be included to reach statistical significance. Indeed, since then, large consortia have been formed to increase sample size. In 2009, three consortia presented their data in Nature. Stefansson and colleagues,

representing the SGENE-plus consortium, included data of 2,663 cases and 13,498 controls[28]. The International Schizophrenia Consortium (ISC) collected data in 3,322 schizophrenia cases and 3,587 controls from European ancestry[6]. Finally, the Molecular Genetics of Schizophrenia (MGS) sample included a European-ancestry sample (2,681 cases and 2,652 controls) and an African-American sample (1,286 cases and 973 controls)[29].

Even though these studies included thousands of cases and controls, very few genome-wide significant findings were reported for each of the individual samples. To further increase statistical power, the three consortia exchanged GWA study summary results. Meta-analyses provided strong support for the involvement of the Major Histocompatibility Complex (MHC)[6, 28, 29] and for involvement of genetic variants in *NRGN* and *TCF4*[28]. The associations with *NRGN* and *TCF4* point to a causal role of pathways involved in brain development, memory and cognition. The MHC region (6p21.32-p22.1) is a gene-dense region, characterized by high levels of linkage disequilibrium. The MHC region comprises 0.3% of the genome and contains hundreds of genes (i.e., 1.5% of the genes in Online Mendelian Inheritance in Man (OMIM))[30]. It is enriched for genome-wide significant SNP associations while 6.4% of the associations reported in the NHGRI GWAS catalog are located in this region[31, 32].

Due to the high level of linkage equilibrium in the MHC region, multiple SNPs show significant associations with schizophrenia and any of the genes in this region could be functionally related to schizophrenia. Walters et al. studied the association between four SNPs in the MHC region and cognitive deficits, a core component of schizophrenia [33, 34]. The G allele of SNP rs6904071 was found to be associated with delayed

episodic memory and decreased hippocampal volume in schizophrenia patients and controls[33] which supports a causal relation between genetic variants in the MHC region and schizophrenia. However, until the location of the causal genetic variants is more precisely identified, it is impossible to distinguish between several theoretically plausible mechanisms.

Because of the large samples required to detect genetic variants with small effects, in 2011 the samples of the three consortia were combined into a single mega-analysis by the Psychiatric Genomics Consortium (PGC)[35]. This analysis included a stage-I discovery sample of 9,394 cases and 12,462 controls and a stage-II replication sample of 8,442 cases and 21,397 controls. Ten independent SNPs located in seven different loci were significantly associated with case-control status. Two loci were previously implicated in schizophrenia: 6p21.32-p22.1 (MHC) and 18q21.1 (*CCDC68* and *TCF4*). One of the novel loci is located within intron three of *AK094607* which contains the primary transcript for microRNA-137 (*MIR137)*[36]. *MIR137* has been implicated in regulating adult neurogenesis and neuronal maturation[35, 37-39] (see Ripke et al.[35]) and could therefore contribute to brain development abnormalities. This study was followed up in 2013, by including GWAs data from an additional 6,454 subjects from Sweden[32]. Further evidence was provided for eight loci previously implicated in schizophrenia, one locus previously implicated in bipolar disorder, and 13 new risk loci (see Table 1). Two loci (*CACNA1C* and *CACNB2*) implicate a role for calcium signaling in the etiology of schizophrenia. To further test the role of calcium channels, which are involved in learning, memory and synaptic plasticity, Ripke and colleagues performed a gene-set test which indeed showed enrichment of smaller p-values in genes encoding calcium channel subunits[32].

Summarizing, current GWAs data provide convincing evidence for several loci. Recent, yet unpublished, analyses of even larger samples have resulted in an increase in the number of significant hits and we can safely conclude that the GWAs approach has been successful in identification of risk loci that can be followed up in functional studies. At present, GWAs data support the role of *MIR137* and calcium signaling in the etiology of schizophrenia. Genetic variants in the MHC region are strongly associated with schizophrenia, but the biological mechanisms are as yet unknown. Furthermore, large sample sizes were required to detect a relatively small number of common genetic polymorphisms affecting schizophrenia, as effect sizes were small (OR~1.1). This raises the question how many other SNPs affect schizophrenia and with what effect size. In the next section, we therefore discuss what GWA studies have revealed so far about the genetic architecture of schizophrenia.

Table 1. Description of the 22 genome-wide significant loci in the combined analysis.[a]

Chromosomal region	*P* value	Candidate gene in relation to index SNP[b]	Other genes in genomic region defined by LD[c]	Disease associations[d]
Chr. 6: 31,596,138–32,813,768	9.14×10^{-14}	*HLA-DRB9*	MHC class II, many other genes, lincRNA	Many
Chr. 10: 104,487,871–105,245,420	3.68×10^{-13}	*C10orf32-AS3MT*	*CALHM1*, *CALHM2*, *CALHM3*, *CNNM2*, *CYP17A1*, *INA*, *MIR1307*, *NT5C2*, *PCGF6*, *PDCD11*, *SFXN2*, *ST13P13*, *TAF5*, *USMG5*, *WBP1L*	GWAS: blood pressure, coronary artery disease, aneurysm
Chr. 7: 1,827,717–2,346,115	5.93×10^{-13}	*MAD1L1*	*FTSJ2*, *NUDT1*, *SNX8*	
Chr. 1: 98,141,112–98,664,991	1.72×10^{-12}	(*MIR137*, 37 kb)	*DPYD*, lincRNA	*DPYD*: mental retardation
Chr. 12: 2,285,731–2,440,464	5.22×10^{-12}	*CACNA1C*	–	*CACNA1C*: autism, Timothy syndrome, Brugada syndrome 3
Chr. 10: 18,601,928–18,934,390	1.27×10^{-10}	*CACNB2*	*NSUN6*	*CACNB2*: Brugada syndrome 4; GWAS: blood pressure
Chr. 8: 143,297,312–143,410,423	2.19×10^{-10}	*TSNARE1*	–	
Chr. 1: 73,275,828–74,099,273	3.64×10^{-10}	(x10NST00000415686.1, 4 kb)	lincRNA	
Chr. 11: 130,706,918–130,894,976	1.83×10^{-9}	(*SNX19*, 31 kb)	lincRNA	
Chr. 5: 151,888,959–	2.65×10^{-9}	ENST00000503048.1	lincRNA (*GRIA1*)	

152,835,304					
Chr. 5: 152,505,453–152,707,306	4.12×10^{-8}				
Chr. 19: 19,354,937–19,744,079	3.44×10^{-9}	(*MAU2*, 4 kb)	*CILP2*, *GATAD2A*, *GMIP*, *HAPLN4*, *LPAR2*, *MIR640*, *NCAN*, *NDUFA13*, *PBX4*, *SUGP1*, *TM6SF2*, *TSSK6*, *YJEFN3*	GWAS: lipid levels	
Chr. 2: 37,422,072–37,592,628	6.78×10^{-9}	*QPCT*	*C2orf56*, *CEBPZ*, *PRKD3*, *SULT6B1* lincRNA		
Chr. 5: 101,581,848–101,870,822	9.03×10^{-9}	*SLCO6A1*	lincRNA		
Chr. 3: 52,215,002–53,175,017	1.16×10^{-8}	*ITIH3*	*ALAS1*, *ALDOAP1*, *BAP1*, *C3orf78*, *DNAH1*, *GLT8D1*, *GLYCTK*, *GNL3*, *ITIH1*, *ITIH4*, *MIR135A1*, *MIRLET7G*, *MUSTN1*, *NEK4*, *NISCH*, *NT5DC2*, *PBRM1*, *PHF7*, *PPM1M*, *RFT1*, *SEMA3G*, *SFMBT1*, *SPCS1*, *STAB1*, *TLR9*, *TMEM110*, *TNNC1*, *TWF2*, *WDR82*, lincRNA	*GLYCTK*: D-glyceric aciduria, mental retardation; *RTF1*: mental retardation; GWAS: adiponectin, height, waist-hip ratio	
Chr. 2: 145,139,727–145,214,607	1.19×10^{-8}	*ZEB2*	–	*ZEB2*: Mowat-Wilson syndrome, mental retardation	
Chr. 2: 200,628,118–201,293,421	1.21×10^{-8}	*FONG*	*C2orf47*, *C2orf69*, *SPATS2L*, *TYW5*, lincRNA	GWAS: osteoporosis	
Chr. 18: 52,722,378–52,827,668	1.22×10^{-8}	(ENST00000565991.1, 21 kb)	lincRNA (*TCF4*)		
Chr. 2: 233,550,961–233,808,241	1.51×10^{-8}	*C2orf82*	*GIGYF2*, *KCNJ13*, *NGEF*		
Chr. 1: 243,593,066–244,025,999	1.80×10^{-8}	*AKT3*	*CEP170*		
Chr. 1: 243,418,063–243,627,135	2.53×10^{-8}	*SDCCAG8*			
Chr. 12: 123,447,928–123,913,433	2.28×10^{-8}	*C12orf65*	*ABCB9*, *ARL6IP4*, *CDK2AP1*, *MIR4304*, *MPHOSPH9*, *OGFOD2*, *PITPNM2*, *RILPL2*, *SBNO1*, *SETD8*, lincRNA	*C12orf65*: mental retardation; GWAS: high-density lipoprotein, height, head size	
Chr. 8: 89,188,454–89,761,163	3.33×10^{-8}	Intergenic	*MMP16*, lincRNA		
Chr. 5: 60,484,179–60,843,706	3.78×10^{-8}	ENST00000506902.1	*ZSWIM6*, *C5orf43*, lincRNA		

[a]Adapted from Ripke et al.[32].

[b]The gene within which an index SNP is located is given. For intergenic index SNPs, the nearest gene is given in parentheses

[c]Other named genes in the genomic interval.

[d] Data from the NHGRI GWAS catalog[31], OMIM[30], and a compilation of genes related to autism[40] and mental retardation [30, 40, 41]. No data means no Affymetrix U219 probe sets or low expression in peripheral blood. The *CACNB2* association emerged when considering attention deficit/hyperactivity disorder (ADHD), autism, bipolar disorder, major depressive disorder and schizophrenia as affected[42].

Genetic architecture

Although subsequent PGC schizophrenia studies have identified an increasingly large number of genome-wide significant SNPs due to increased sample sizes, Visscher and colleagues estimated the contribution of SNPs robustly associated with schizophrenia to be less than one percent[43]. Despite the fact that additional loci have been found since then, this small percentage is in stark contrast with the 64-81% heritability range estimated by twin studies[3, 24, 25] a discrepancy which has been coined the missing heritability[44]. This section examines several explanations for the missing heritability and discusses their implications for the genetic architecture of schizophrenia.

One explanation for the missing heritability is model misspecification[8, 45]. According to this view the additive models often used to analyze genetic data are oversimplifications of the complex genetic architecture of schizophrenia. Additive models assume that the cumulative genetic effect of SNPs is the sum of the individual genetic effects. This assumption is convenient as it implies that heritability is the sum of genetic effects of individual SNPs. It also implies that no epistasis is present; i.e. no (statistical) interactions between genetic effects are assumed[8, 46-49]. However, if epistasis is present, heritability estimates based on family and twin studies could be overestimated instead of missing[8]. In this view, missing heritability is a problem of overestimated heritability in family and twin studies.

The role of epistasis in missing heritability has been much debated[8, 46-51]. The problem with this debate is that it is currently impossible to empirically quantify the amount of epistasis as most epistatic models are too complex to estimate. The epistasis debate is therefore foremost a

philosophical debate. However, even if we do not believe the additive model to be strictly true, there are pragmatic reasons for using it. The additive model is currently the most parsimonious model to investigate genetic effects and has taught important lessons about the genetic architecture of complex diseases, as we will discuss next.

Most other explanations for missing heritability assume the additive model to be valid and attribute the discrepancy in heritability either to unobserved variants or a lack of statistical power. First, a proportion of the causal variants might not be properly tagged by SNPs on current genotype platforms, which makes it difficult to detect their effects[45, 46]. This is especially true for SNPs with (very) rare variants (minor allele frequency < 1%). Extremely large sample sizes are required to reliably detect effects of rare variants even for relatively large effect sizes. Second, hundreds or thousands of common SNPs could have effect sizes too small to identify individually, but which cumulatively could explain a sizeable portion of the heritability[45, 46]. Although the common and rare variant hypotheses are often contrasted, both might be better viewed as two extremes of a continuum[43]. Ultimately it is the combination of sample size, allele frequency, effect size, and linkage equilibrium structure which determines the power to detect an individual causal variant.

So far, only few rare variants affecting schizophrenia have been identified and none of those were disease-specific[4]. In contrast, several approaches have shown that many common variants with small effect sizes indeed comprise a sizeable proportion of the heritability. For example, risk score analysis, has shown the importance of (currently) non-significant SNPs. In polygenic risk score analysis[6], the SNP effect sizes (i.e., logistic regression weights) from a discovery sample are used

to compute additive genetic risk scores and to predict disease status in an independent target sample. The effect sizes from a GWA analysis of 8,831 schizophrenia cases and 12,067 controls (the PGC-1 schizophrenia sample) showed that a risk score based on all SNPs with a p-value below 0.1 predicts 6% of the variance in a Swedish case-control target sample (based on Nagelkerke pseudo-R^2), whereas a score based on SNPs with a p-value below 0.001 only resulted in a pseudo-R^2 of 3%[32]. Given that the p-value threshold for genome-wide significance is $5x10^{-8}$, these results imply that many non-significant SNPs contribute to schizophrenia. In other words the missing heritability is, at least partly, hidden in thousands of non-significant SNPs with small effects.

Although polygenic risk score analysis provides evidence for the importance of non-significant common SNPs, the pseudo-R^2 does not provide an estimate of the total heritability contributed by these common SNPs. Genome-wide complex trait analysis (GCTA), a second approach, was explicitly designed to estimate the total heritability (on the liability scale) of complex traits based on common SNPs[52-54]. GCTA does so by implementing a linear mixed model (LMM). The LMM is an additive model with the additional assumption that the effect sizes on the liability scale of SNPs come from a normal distribution with mean zero. In other words, most SNPs are assumed to have a near-zero effect size. Under these assumptions, GCTA infers the total heritability contributed by all observed common SNPs, both significant and non-significant. For the PGC-1 schizophrenia sample, GCTA estimated the heritability at 27% (s.e. 2%) assuming a population risk of 0.4%[32]. In the Swedish target sample, the heritability estimate was even 32% (s.e. 3%), possibly due to higher genetic homogeneity in this sample[32].

A third approach which illustrated the importance of common variants is approximate Bayesian polygenic analysis (ABPA)[32, 55, 56]. ABPA is an approximate Bayesian method which, like GCTA, estimates the heritability, but uses polygenic risk score analysis as a model. Note that in Bayesian statistics the credible interval is analogous but not equal to the well-known confidence interval in frequentist statistics. ABPA yielded a heritability estimate of 34% (95% credible interval 31%-37%) assuming a population prevalence of 0.4% for the PGC and Swedish samples combined[32]. In GWA studies it is customary to impute unobserved SNPs using a reference panel in which these SNPs are observed. In this case the Utah residents with ancestry from northern and Western Europe (CEU) reference panel from HapMap 3[57] was used for imputation. However, when the much denser CEU 1000 Genomes reference panel[58] was used and a one percent population prevalence of schizophrenia was assumed the estimated heritability was even 50% (95% credible interval 45%-54%)[32]. Furthermore, unlike GCTA, ABPA can also estimate other parameters of interest. For example, it was estimated that 8,300 (95% credible interval 6,300-10,200) common independent SNPs contributed to the 50% heritability estimate[32]. Moreover, compared to other diseases such as rheumatic arthritis and celiac disease, the estimated effect sizes per SNP are smaller in schizophrenia[32].

In conclusion, all three approaches demonstrate that under the additive model a large part of the missing heritability can be explained. Depending on method, population prevalence, sample (size), and reference panel used for imputation, 52-78% of the missing heritability can be explained[32]. Hence, GWA studies have provided clear evidence that a sizeable proportion of the heritability in schizophrenia can be attributed to common SNPs while most individual effect sizes are too

small to be detected with current sample sizes. It is therefore expected that increasing sample sizes even further will continue to produce additional significant hits. Identifying the many genetic loci that contribute to schizophrenia is only a first step, however. The next challenge will be to investigate to what extent these hundreds or thousands of schizophrenia SNPs point towards common pathways[4].

Genetic overlap with other disorders

GWA studies not only support the contribution of genetic factors to the risk of developing schizophrenia, but also point towards pleiotropy, a shared genetic basis with other diseases. For example, a cross-disorder study by PGC identified four different genome-wide significant loci (see Figure 1) associated with disease in a large cross-disorder meta-analysis in which schizophrenia, bipolar disorder, major depressive disorder, autism spectrum disorder, and attention deficit-hyperactivity disorder PGC samples were combined[42]. The associated SNPs were located on chromosome 3p21 and 10q24, and on two L-type voltage gated calcium-channel subunits (*CACNA1C* and *CACNB2*). Earlier, Huang et al.[59] also performed a cross-disorder GWA analysis of schizophrenia, bipolar disorder, and depression and identified a genome-wide significant SNP (rs6484218) near the adrenomedullin (*ADM*) gene which was most strongly associated with bipolar disorder II. These cross-disorder findings are in line with the evidence from family studies that, for example, schizophrenia and bipolar disorder have a common genetic basis[3].

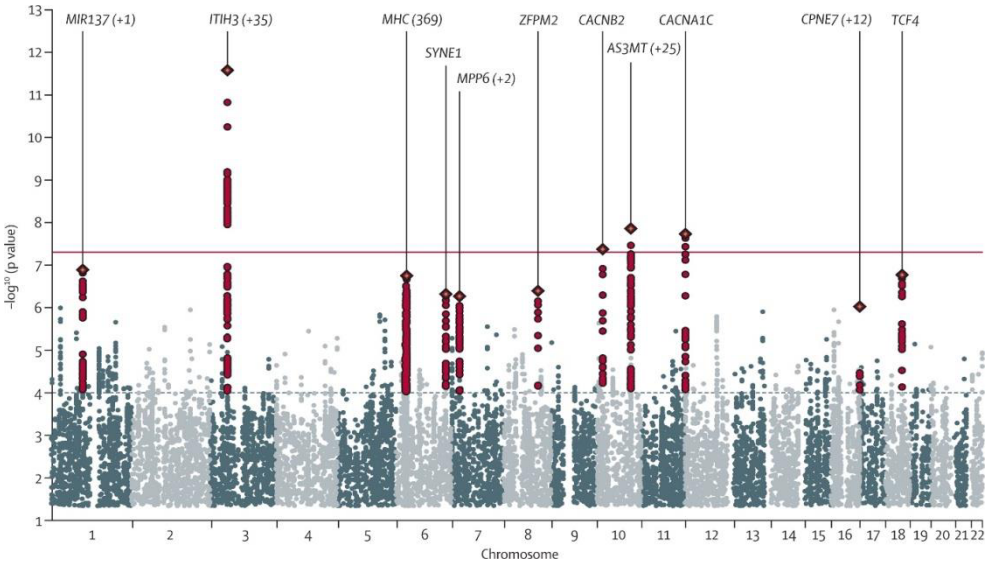

Figure 1. Manhattan plot of cross-disorder meta-analysis. Horizontal line represents genome-wide significance threshold ($p<5\times10^{-8}$). Copied with permission from Smoller et al.[42].

However, due to the relatively small number of significant hits in schizophrenia GWA studies, it is no surprise that a cross-disorder meta-analysis only results in a few genome-wide significant associations. Polygenic risk score analysis can also be applied to estimate the extent of genetic overlap between disorders. For example, the cross-disorder group of the psychiatric genetics consortium performed a risk score analysis using the PGC-I schizophrenia sample as a discovery sample and the PGC samples for bipolar disorder (BD), major depression disorder (MDD), autism spectrum disorder (ASD), and attention deficit-hyperactivity disorder (ADHD) as target samples[42]. If there is no genetic overlap between schizophrenia and bipolar disorder, polygenic risk scores based on a schizophrenia sample should not predict disease status in bipolar disorder. However, if many SNPs affect the liability for both schizophrenia and bipolar disorder, a polygenic risk score based on schizophrenia liability will predict bipolar disorder.

This is indeed what has been found[42]. Schizophrenia risk scores based on SNPs with p-values < 0.3 explained more than 2% of the variance in BPD status (Nagelkerke R^2), which was highly significant ($p < 10^{-16}$). The same analysis for MDD resulted in a predictive value of 0.8%, which was also highly significant ($p < 10^{-16}$). Early onset psychiatric diseases such as ASD and ADHD showed much less genetic overlap with schizophrenia: 0.1% for ASD ($p<0.05$) and 0.0% for ADHD ($p>0.05$). Note that these results refer to overall genome-wide genetic overlap. For example, the low genetic overlap between schizophrenia and ASD does not preclude genetic overlap of small effect size at specific sites such as reported by Voineskos and colleagues[60]. Although a genome-wide analysis of genetic overlap does not specify the location of contributing SNPs, a pathway analysis for all five diseases combined revealed significant enrichment for a set of calcium channel activity genes. More precise mapping of gene pathways affected in schizophrenia is expected as increasingly large sample sizes allow better distinction between relevant and non-relevant SNPs.

Similar to polygenic risk score analysis, GCTA can also be used for cross-disorder genetic studies[2]. Generalizing the linear mixed modeling (LMM), not only the heritability of a single disease can be estimated, but also the coheritability of two diseases. Note that coheritability is a measure of genetic overlap which depends not only on the correlation between genetic effects, but also on the individual heritabilities. In other words, a low coheritability can be due to low correlation in genetic effects as well as low heritabilities in either disease. The coheritabilities of schizophrenia and BPD (15%, s.e. 1%), MDD (9%, s.e. 1%), ASD (3%, s.e. 1%), and ADHD (2%, s.e. 1%) estimated by Lee et al.[2] are shown in Figure 2 and followed the same pattern as the results from polygenic risk score analysis.

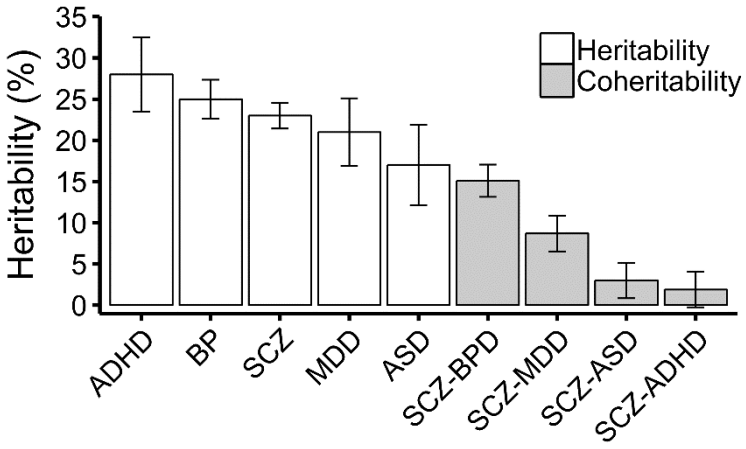

Figure 2. SNP-based heritabilities and coheritabilities with 95% confidence interval for five psychiatric disorders: attention deficit hyperactivity disorder (ADHD), bipolar disorder (BPD), schizophrenia (SCZ), major depression disorder (MDD), and autism spectrum disorder (ASD). Adapted from Lee et al.[2].

The evidence of a genetic overlap between schizophrenia and bipolar disorder that has been provided by family studies[3], has been confirmed in cross-disorder GWA studies. However, analogous to identifying the loci contributing to the heritability of schizophrenia, the evidence for pleiotropy in psychiatric diseases provides a challenge to identify the loci involved and to establish which common biological pathways these diseases have in common[61].

Conclusions

Family, twin, and GWA studies have provided clear evidence that genetic factors play an important role in the onset of schizophrenia. Moreover, genetic studies have contributed hugely to a better appreciation of the complexity of the genetic architecture of schizophrenia. Not only is it now evident that many different loci affect schizophrenia; many of these loci are expected to contribute to the risk of other psychiatric diseases as well. However, many challenges need to be overcome for a thorough understanding of the etiology of schizophrenia. First, many more loci

affecting schizophrenia need to be identified. Second, to understand the biological processes behind schizophrenia it is important to establish the common pathways that contribute to schizophrenia. Third, to biologically differentiate schizophrenia from, for example, bipolar disorder it is necessary to further investigate the similarities and dissimilarities in biological pathways between both diseases.

Nonetheless, the coordinated effort of PGC to continuously increase sample sizes promises to mitigate some of the statistical challenges in genetic studies. Although it is not feasible to identify all loci, Ripke et al.[32] propose a goal for the field to identify the top 2,000 loci for schizophrenia. Furthermore, especially for follow-up studies, next-generation sequencing allows more precise mapping of the location of causal common SNPs[62]. Similarly, family studies based on next-generation sequencing might reveal large rare effect loci which may not contribute much to the heritability, but could provide important clues about the biological processes behind schizophrenia. Finally, other types of genomic information, such as CNVs[63], expression data[64], and epigenetic factors[65], might provide additional biological insight into the etiology of schizophrenia.

Due to the unknown etiology of psychiatric diseases, classification of psychiatric disorders has traditionally been based on clinical observation. By solving the genetic puzzle behind schizophrenia and other psychiatric disease, genetic studies would complement the traditional classification with an alternative classification based on etiological coherence instead of phenotypic coherence. Such a break-through would allow development of drugs that target specific pathways identified for schizophrenia. Therefore genetic studies will continue to play an important and exciting role in schizophrenia research.