Analysis

Genetic cross-disorder analysis in psychiatry: from methodology to clinical utility

Dick Schijven, Jan H. Veldink and Jurjen J. Luykx

Summary

Genome-wide association studies have uncovered hundreds of loci associated with psychiatric disorders. Cross-disorder studies are among the prime ramifications of such research. Here, we discuss the methodology of the most widespread methods and their clinical utility with regard to diagnosis, prediction, disease aetiology and treatment in psychiatry.

Copyright and usage

In the field of psychiatric genetics, thousands of affected and healthy individuals have been included in genome-wide association studies (GWASs) to test disease associations at common genetic variants, of which single-nucleotide polymorphisms (SNPs) are the most abundant. These occur throughout the genome, on average once every couple of hundreds of base pairs, and have minor allele frequencies >1% in the general population. Currently, arrays are available to genotype over two-million SNPs simultaneously. Genotypes of SNPs that lie close together in the genome are highly likely to be co-inherited and therefore show high correlation, resulting in blocks of SNPs that are in linkage disequilibrium. Linkage disequilibrium can be exploited to impute SNPs that were not directly genotyped on an array, which at present increases the number of SNPs in GWAS data-sets to up to 10–15 million variants. Moreover, an association signal usually covers a genomic region of variable size, called a locus, wherein the associated variants can be in linkage disequilibrium with one or more causal variants. Most SNPs are known to have small effect sizes in mental illness (odds ratios between 1.03 and 1.10)[1,2] but they collectively account for about 20–30% of the heritability per disorder (SNP-based heritability),[3] of which only a fraction is currently explained by genome-wide significant SNPs discovered in GWASs (association P-value $<5 \times 10^{-8}$), implying that a plethora of SNPs which have not yet been discovered by GWASs due to insufficient power underlies these 'complex polygenic' traits.[1,4] Interestingly, many loci are not exclusively associated with a single disorder but with two or even more, suggesting aetiological similarities between diseases. This opens up opportunities to evaluate the shared genetic basis of psychiatric disorders. Thus SNP data are the starting point for many cross-disorder analyses and are usually available as either individual-level data-sets or post-GWAS summary-level data. Individual-level data include SNP genotypes, phenotypes and other relevant information for each participant separately. Summary-level data (or summary statistics) comprise the final results of a GWAS, including per-SNP information on their association with the trait studied (usually odds ratios and P-values). Summary-level data are easier to obtain through public sharing (e.g. www.ebi.ac.uk/gwas/downloads/summary-statistics) and computationally easier to handle, but contain less detailed information than individual-level data. Cross-disorder studies aim to capture both pleiotropy and genetic correlation, which are related but essentially different concepts. Pleiotropy defines a single genetic variant affecting more than one phenotype, whereas genetic correlation describes the sharing of polygenic disease architectures at a genome-wide level. Cross-disorder studies can be used in several clinical domains, namely: (a) diagnostics (disorder classification and patient stratification), (b) prognosis (prediction of clinical course and outcome of a disorder) and (c) treatment (discovery of drug targets and treatment tailoring through new insights into disease aetiology). Here, we outline the current and potential future contributions of widely applied approaches in genetic cross-disorder studies for the field of psychiatry, ordered by the clinical domain they target (also see Table 1).

Diagnostics

Genetic correlations are the most widely used estimate to describe the extent to which variance in liability for two disorders is attributable to additive genetic effects and thus expresses the proportion of shared SNP-based heritability between two traits. Initially, methods to estimate genetic correlation from GWAS data relied on individual-level genotype data (e.g. bivariate restricted maximum likelihood, REML), using a genetic relationship matrix that captures genetic similarity between distantly related individuals which can be correlated to phenotypic similarities for two diseases.[17] This was followed by methods which use summary-level data, where linkage disequilibrium score regression (LDSC) is widely applied and genetic correlations are based on similarities in effect sizes and effect directions of SNPs that are shared between GWAS summary statistics of two phenotypes.[5] Bivariate REML and especially LDSC are most accurate for traits with a polygenic architecture, rendering both well applicable to psychiatric disorders. The genetic correlation estimate falls within the range of −1 and 1, where 1 represents the unlikely scenario that shared liability is caused by exactly the same risk SNPs and −1 by the similarly implausible situation where exactly identical SNPs increase risk for one disease while decreasing risk for another. A null correlation indicates absence of shared disease liability due to overlapping genetic architecture. Genetic correlation estimates are based on all SNPs present in two GWAS data-sets or summary-level results regardless of their strength of association with a disease (as sub-threshold variants capture a substantial amount of the SNP-based heritability). The ability to test these correlations at large scale has provided insights into the landscape of disease classifications both within psychiatric disorders as well as for psychiatric disorders relative to other disease groups. Significant genetic correlations are often observed between psychiatric disorders, with schizophrenia and bipolar disorder peaking at approximately 0.8.[3,5,6] Surprising

Table 1	Clinical utility of genetic cross-disorder analysis in psychiatry		
Clinical domain	Main cross-disorder genetics approaches	Utility	Example
Diagnostics	Genetic correlation, PRS analysis	Describe polygenic overlap between psychiatric disorders	Genetic correlation between schizophrenia and bipolar disorder estimated at 0.80.[3,5,6]
		Describe genetic heterogeneity within one psychiatric disorder	Schizoaffective bipolar disorder is most strongly associated to schizophrenia polygenic risk, followed by bipolar disorder-I and bipolar disorder-II (relative risk based on schizophrenia PRSs of 1.37, 1.30 and 1.04, respectively).[7]
		Describe polygenic overlap with phenotypes outside psychiatry	Schizophrenia genetically correlates with amyotrophic lateral sclerosis (0.14).[8]
			MDD (0.11) and ADHD (0.21) have a significant genetic correlation with body mass index.[6]
	PRS analysis	Identification of clinical subgroups in a heterogeneous disorder (patient stratification)	Schizophrenia PRSs define a gradient in types of psychosis in bipolar disorder.[7]
Prognosis	PRS analysis	Prediction of clinical outcome	People with first-episode psychosis who later develop schizophrenia have higher schizophrenia PRSs than those who develop other psychotic disorders.[9]
		Prediction of predisposition for a disorder	OCD, schizophrenia, schizophrenia–bipolar disorder and MDD PRSs have predictive value for obsessive–compulsive symptoms in a population sample.[10]
			Future possibility to assess risk for post-trauma psychopathology in groups of healthy military personnel using PRSs.
Treatment	Combined analysis (mega- or meta-analysis)	Detection of shared disease variants that might converge in shared disease pathways	Four pleiotropic loci identified between five psychiatric disorders (ASD, ADHD, bipolar disorder, MDD and schizophrenia).[11]
			A total of 114 shared loci identified between schizophrenia and bipolar disorder, and 4 loci with divergent effects.[12]
			A total of 96 pleiotropy-informed loci identified for depressive symptoms, neuroticism and subjective well-being.[13]
		Detection of shared disease variants that are targetable by drugs, or that can induce treatment side effects	Gene targets of two drugs show enrichment for variants associated with multiple cognitive phenotypes and intellectual disability.[14]
			Variants associated with clozapine-induced agranulocytosis have been associated with genes involved in adverse response to statins.[15]
	PRS analysis	Prediction of treatment response	Schizophrenia PRSs are inversely correlated with lithium response in bipolar disorder.[16]

PRS, polygenic risk score; MDD, major depressive disorder; ADHD, attention-deficit hyperactivity disorder; OCD, obsessive–compulsive disorder; ASD, autism spectrum disorder.

genetic correlations between psychiatric traits and other phenotypes presumed to be aetiologically unrelated have been described, suggesting that biological mechanisms involved in psychiatric disorders overlap with those of phenotypes that are clinically distinct. Examples include significant positive genetic correlations between body mass index on the one hand and major depressive disorder (MDD) (0.11) and attention-deficit hyperactivity disorder (ADHD) (0.21) on the other.[6] The positive genetic correlation between schizophrenia and amyotrophic lateral sclerosis (ALS) of 0.14 is the first described genome-wide correlation between a psychiatric and a neurological disorder.[8]

Polygenic risk score (PRS) analysis is another widely applied method to assess polygenic overlap between diseases. This technique uses estimates of SNPs derived from summary-level data of a discovery GWAS to calculate per-individual scores based on the number of effect alleles carried and weighted for the effect sizes at all overlapping SNPs in an individual-level target GWAS data-set. It is possible to calculate these scores on all SNPs tested for association in the discovery GWAS ($P \leq 1$), but a selection is often made based on strength of association (e.g. $P < 5 \times 10^{-8}$). These scores thus capture the combined effects of many SNPs and, in cross-disorder studies, PRSs reflecting the polygenic risk for a discovery disease are tested for association with the phenotypic measure of a different target disease.[18] For example, schizophrenia PRSs explain up to 2.4% of the phenotypic variance in bipolar disorder.[11,19] Strikingly, schizophrenia PRSs were more strongly associated with schizoaffective bipolar disorder, followed by bipolar disorder type I and bipolar disorder type II (risk ratios of 1.37, 1.30 and 1.04, respectively), showing that aetiological differences

exist within psychiatric disorders.[7] It is furthermore particularly interesting to compare individuals at the extreme ends of the range of PRSs in a study population. For instance, when PRSs calculated using summary-level schizophrenia GWAS data were applied to an independent schizophrenia case-control target GWAS data-set, individuals in the highest PRS decile had up to 20-fold increased odds of schizophrenia compared with individuals in the lowest PRS decile,[20] again showing the substantial combined effect of many SNPs. In addition to this within-trait example, the observation of increased odds of disease is to a lesser extent applicable to correlated diseases in between-disorder studies (e.g. highest decile schizophrenia PRSs increasing odds of ALS up to 1.3 compared with lowest decile PRSs in a case-control cohort).[8] Explained variances based on PRSs are always much lower than genetic correlation estimates, which has been demonstrated for schizophrenia and bipolar disorder (2.4 *v.* 80%). These methods are conceptually different: whereas genetic correlations describe the proportion of shared genetic background based on theoretical full SNP-based heritabilities, PRSs are based on actual SNP effect estimates and applied to explain phenotypic variance that is only partly attributable to genetic factors. PRSs can thus never explain more variance than the heritability of the target disease or, when applied in cross-disorder analysis, the proportion of the heritability shared with the discovery disease.[21]

As with genetic correlations, PRSs can be useful to establish genetic links between disorders and define the genetic landscape of psychiatric disorders. In a population of unselected samples and when calculated based on a large number of SNPs, PRSs show a normal distribution. Therefore, the true diagnostic utility

of PRSs as a quantitative phenotype lies in disease classification and the more fine-grained stratification of patients. For example, bipolar disorder with mood-incongruent psychosis shows a stronger correlation to schizophrenia PRSs than bipolar disorder with mood-congruent psychosis and bipolar disorder without psychosis.[7] On a similar note, higher schizophrenia PRSs associate with psychotic features and earlier age at onset in a sample of people with bipolar disorder, whereas bipolar disorder PRSs show a positive correlation with manic symptoms in people with schizophrenia.[12] Future increases in GWAS sample sizes will provide more accurate effect estimates for SNPs, which in turn can further empower PRSs and their specific relation to psychiatric symptoms across the boundaries of DSM-5 (2013) disorder classifications.[4]

Prognosis

The ability to predict the course of a disorder in psychiatric patients could prove highly valuable as we do not currently have accurate course prediction models for these disorders. For psychotic disorders, the potential of PRSs to predict the conversion from at-risk states to clinical diagnoses has been shown by schizophrenia PRSs being significantly higher in individuals with first-episode psychosis later diagnosed with schizophrenia than in first-episode patients diagnosed with other psychotic disorders, although both groups had higher PRSs than healthy controls.[9] A cross-disorder example includes a population-based sample where PRSs for obsessive–compulsive disorder (OCD), schizophrenia, MDD and combined schizophrenia–bipolar disorder predicted subclinical obsessive–compulsive symptoms, thereby showing the potential of these scores to identify individuals who have increased risk of developing OCD and other psychiatric disorders.[10] With regard to psychiatric conditions where environmental factors play a pivotal role, such as post-trauma psychopathology, PRSs for several disorders (such as post-traumatic stress disorder, depression and anxiety) could ideally be used to screen healthy at-risk groups, e.g. military personnel, and take preventative measures for those with high risk scores. However, their currently low accuracy and sensitivity in clinical phenotypes preclude application of such PRSs to risk prediction at the level of the individual. Therefore, integrating PRSs into a diagnostic workup that includes other data, such as psychiatric signs and symptoms, may in the near future improve prognostic accuracy.[22]

Treatment

The previously discussed polygenic methods describe the extent to which two disorders correlate, but do not pinpoint specific loci with cross-disorder effects. Information on specific shared disease loci can be obtained by merging individual-level or summary-level data from different disorders in a single association analysis, combining all cases into a single phenotype. Such mega- or meta-analyses increase power to detect pleiotropic loci that were below the genome-wide significance thresholds in single-disorder GWASs. One of the most illustrative cross-disorder studies performed between five psychiatric disorders with the highest heritability estimates (autism spectrum disorder [ASD], ADHD, bipolar disorder, MDD and schizophrenia) identified four shared genome-wide significant loci with pleiotropic effects which were not genome-wide significant when analyses were performed on disorders separately,[11] clearly indicating the increased power in sets of disorders that share genetic loci. Moreover, a combined analysis of schizophrenia and bipolar disorder individual-level GWAS data-sets has identified 114 loci implicating synaptic and neuronal pathways shared between the two, although these loci

exert different effect sizes in both disorders. Interestingly, this study also identified four loci involved with divergent effects between these disorders.[12] Recently, techniques have been developed to perform powerful combined analyses of polygenic traits with summary-level data, e.g. resulting in the discovery of 96 pleiotropy-informed loci in a combined study of depressive symptoms, neuroticism and subjective well-being.[13]

Despite the low effect sizes of such pleiotropic variants, they may ultimately prove highly valuable in clinical settings as they inform possible cross-disorder therapeutic targets, empowering drug repurposing. This has for example been shown in a GWAS meta-analysis of phenotypes related to general cognitive ability, where discovered loci were enriched in genes associated with intellectual disability and in gene targets of two pharmacological compounds, prioritising these substances as possible cognitive enhancers.[14] In addition, PRSs can also be used to predict treatment response, as illustrated by the inverse correlation between schizophrenia PRSs and lithium response in people with bipolar disorder.[16] Finally, side effects can sometimes be explained by pleiotropic variants, as illustrated by the association of clozapine-induced agranulocytosis with genes previously linked to adverse reactions to statins.[15]

Caveats in genetic cross-disorder studies

Cross-disorder studies should be interpreted with some caution as observations can be driven by other factors than true overlap at causal variants. First, as opposed to a scenario of true biological genetic overlap, a significant genetic correlation could also be observed if a causal variant directly affects disorder A, whereas disorder B is merely caused by disorder A (mediated pleiotropy). In this scenario, there is no direct effect of the causal variant on disorder B and observed pleiotropy is artificial. Given the differential ages at onset, for psychiatric disorders this would imply that early onset disorders occur first (ASD, ADHD, intellectual disability) and then cause disorders with later onset (MDD, schizophrenia, bipolar disorder), a notion currently lacking support from the scientific literature. In addition, disorder-specific causal variants in different genes might be tagged by the same SNP due to linkage disequilibrium. This SNP could then show an association to both phenotypes although the causal variants are actually different, resulting in spurious pleiotropy.[23] Second, observed pleiotropy can result from external confounding factors, such as population differences or assortative mating. The latter plays a role in psychiatric disorders, where non-random mating between and within disorders is observed.[24] Third, the above-mentioned cross-disorder analysis approaches only use information on SNPs captured in GWAS data, thus omitting other sources of genetic variation, such as rare variants. Fourth, because of overlap in clinical symptoms, the risk of diagnostic misclassification is particularly pressing in psychiatry and can lead to false positive findings in cross-disorder studies when patient cohorts are not homogenous. Methodology has been developed to identify heterogeneous subgroups of patients in GWAS data-sets (e.g. BUHMBOX) to detect such confounding in cross-disorder analyses.[25] Finally, genotype data of the same healthy controls are often used in multiple GWASs of different disorders. When this overlap is not taken into account in cross-disorder analyses combining these data-sets, a false positive correlation – not reflecting shared genetic risk between diseases but merely a correlation induced by genetically identical people – may arise. It is therefore imperative to identify and exclude duplicate participants in cross-disorder studies using individual-level genotype data. Alternatively, various novel cross-disorder methods have been developed that include correction for inflated test results due to

sample overlap and allow for the use of only summary-level data.[5,13] Overall, a considerable amount of cross-disorder analyses can nowadays be performed using summary-level data without the requirement of access to individual-level genotype data. As new GWAS results are continuously being published, summary-level data are almost always made publicly available in line with many journals' manuscript acceptance conditions, whereas individual-level data are often available upon request. Sharing of full GWAS results remains essential for cross-disorder analyses by groups and consortia interested in matching genetic data of other studies with their own data-sets.

Conclusion

A range of techniques have been employed to unravel important cross-disorder genetic findings in psychiatry, resulting in intriguing new clinical perspectives at the levels of diagnostics, prognosis and treatment development and prediction.

Dick Schijven, MSc, Researcher, Department of Psychiatry, Department of Neurology and Neurosurgery and Department of Translational Neuroscience, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht University, the Netherlands; **Jan H. Veldink**, MD, PhD, Professor, Department of Neurology and Neurosurgery, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht University, the Netherlands; **Jurjen J. Luykx**, MD, PhD (iD), Assistant Professor, Department of Psychiatry and Department of Translational Neuroscience, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht University; and GGNet Mental Health, Apeldoorn, the Netherlands

Correspondence: Jurjen J. Luykx, Brain Center Rudolf Magnus, Universiteitsweg 100, 3584CG Utrecht, the Netherlands. Email: J.Luykx@umcutrecht.nl

References

1 Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature* 2009; **461**: 747–53.

2 Nishino J, Kochi Y, Shigemizu D, Kato M, Ikari K, Ochi H, et al. Empirical bayes estimation of semi-parametric hierarchical mixture models for unbiased characterization of polygenic disease architectures. *Front Genet* 2018; **9**: 115.

3 Cross-Disorder Group of the Psychiatric Genomics Consortium. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* 2013; **45**: 984–94.

4 Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 years of GWAS discovery: biology, function, and translation. *Am J Hum Genet* 2017; **101**: 5–22.

5 Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet* 2015; **47**: 1236–41.

6 Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, et al. Analysis of shared heritability in common disorders of the brain. *Science* 2018; **360**: eaap8757.

7 Allardyce J, Leonenko G, Hamshere M, Pardiñas AF, Forty L, Knott S, et al. Association between schizophrenia-related polygenic liability and the occurrence and level of mood-incongruent psychotic symptoms in bipolar disorder. *JAMA Psychiatry* 2018; **75**: 28–35.

8 McLaughlin RL, Schijven D, van Rheenen W, van Eijk KR, O'Brien M, Kahn RS, et al. Genetic correlation between amyotrophic lateral sclerosis and schizophrenia. *Nat Commun* 2017; **8**: 14774.

9 Vassos E, Di Forti M, Coleman J, Iyegbe C, Prata D, Euesden J, et al. An examination of polygenic score risk prediction in individuals with first-episode psychosis. *Biol Psychiatry* 2017; **81**: 470–7.

10 Zilhão NR, Abdellaoui A, Smit DJA, Cath DC, Hottenga JJ, Boomsma DI. Polygenic prediction of obsessive compulsive symptoms. *Mol Psychiatry* 2018; **23**: 168–9.

11 Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* 2013; **381**: 1371–9.

12 Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium. Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. *Cell* 2018; **173**: 1705–16.

13 Turley P, Walters RK, Maghzian O, Okbay A, Lee JJ, Fontana MA, et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat Genet* 2018; **9**: 283.

14 Lam M, Trampush JW, Yu J, Knowles E, Davies G, Liewald DC, et al. Large-scale cognitive GWAS meta-analysis reveals tissue-specific neural expression and potential nootropic drug targets. *Cell Rep* 2017; **21**: 2597–613.

15 Legge SE, Hamshere ML, Ripke S, Pardinas AF, Goldstein JI, Rees E, et al. Genome-wide common and rare variant analysis provides novel insights into clozapine-associated neutropenia. *Mol Psychiatry* 2017; **22**: 1502–8.

16 Amare AT, Schubert KO, Hou L, Clark SR, Papiol S, Heilbronner U, et al. Association of polygenic score for schizophrenia and HLA antigen and inflammation genes with response to lithium in bipolar affective disorder: a genome-wide association study. *JAMA Psychiatry* 2018; **75**: 65–74.

17 Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics* 2012; **28**: 2540–2.

18 Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res* 2007; **17**: 1520–8.

19 The International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009; **460**: 748–52.

20 Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511**: 421–7.

21 Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM. Pitfalls of predicting complex traits from SNPs. *Nat Rev Genet* 2013; **14**: 507–15.

22 Peyrot WJ, Milaneschi Y, Abdellaoui A, Sullivan PF, Hottenga JJ, Boomsma DI, et al. Effect of polygenic risk scores on depression in childhood trauma. *Br J Psychiatry* 2014; **205**: 113–9.

23 O'Donovan MC, Owen MJ. The implications of the shared genetics of psychiatric disorders. *Nat Med* 2016; **22**: 1214–9.

24 Nordsletten AE, Larsson H, Crowley JJ, Almqvist C, Lichtenstein P, Mataix-Cols D. Patterns of nonrandom mating within and across 11 major psychiatric disorders. *JAMA Psychiatry* 2016; **73**: 354–61.

25 Han B, Pouget JG, Slowikowski K, Stahl E, Lee CH, Diogo D, et al. A method to decipher pleiotropy by detecting underlying heterogeneity driven by hidden subgroups applied to autoimmune and neuropsychiatric diseases. *Nat Genet* 2016; **48**: 803–10.