Discovery of shared genomic loci using the conditional false discovery rate approach

Olav B Smeland (M.D. Ph.D.) (ORCID: 0000-0002-3761-5215)[1], Oleksandr Frei (Ph.D.) (ORCID: 0000-0002-6427-2625)[1], Alexey Shadrin (Ph.D.)[1], Kevin O'Connell (Ph.D.) (ORCID: 0000-0002-6865-8795)[1], Chun-Chieh Fan (M.D. Ph.D.) (ORCID: 0000-0001-9437-2128)[2,3], Shahram Bahrami (Ph.D.)[1], Dominic Holland (Ph.D.)[3,5,6], Srdjan Djurovic (Ph.D.) (ORCID: 0000-0002-8140-8061)[7,8], Wesley K. Thompson (Ph.D.) (ORCID: 0000-0002-1148-1976)[4], Anders M Dale (Ph.D.) (ORCID: 0000-0002-6126-2966)[2,3,5,6], Ole A Andreassen (M.D. Ph.D.) (ORCID: 0000-0002-4461-3568)[1]

[1]NORMENT Centre, Institute of Clinical Medicine, University of Oslo and Division of Mental Health and Addiction, Oslo University Hospital, 0407 Oslo, Norway; [2]Department of Cognitive Science, University of California, San Diego, La Jolla, CA, USA; United States of America; [3]Department of Radiology, University of California, San Diego, La Jolla, CA 92093, United States of America; [4]Department of Psychiatry, University of California, San Diego, La Jolla, CA, USA; [5]Department of Neuroscience, University of California San Diego, La Jolla, CA 92093; [6]Center for Multimodal Imaging and Genetics, University of California San Diego, La Jolla, CA 92093, United States of America;[7]Department of Medical Genetics, Oslo University Hospital, Oslo, Norway; [8]NORMENT Centre, Department of Clinical Science, University of Bergen, Bergen, Norway

Corresponding authors: Olav B Smeland or Ole A Andreassen

Olav B. Smeland M.D. Ph.D.
Postdoctoral researcher
Division of Mental Health and Addiction
University of Oslo and Oslo University Hospital
Kirkeveien 166, 0424 Oslo, Norway
Email: o.b.smeland@medisin.uio.no
Phone: +47 41220844

Ole A. Andreassen M.D. Ph.D.
Professor of Biological Psychiatry,
Division of Mental Health and Addiction
University of Oslo and Oslo University Hospital
Kirkeveien 166, 0424 Oslo, Norway
Email: o.a.andreassen@medisin.uio.no
Phone: +47 23027350

Acknowledgements: National Institutes of Health (NS057198; EB00790); National Institutes of Health NIDA/NCI: U24DA041123; the Research Council of Norway (229129; 213837; 248778; 223273; 249711); the South-East Norway Regional Health Authority (2017-112); KG Jebsen Stiftelsen (SKGJ-2011-36).

Key Words: conditional false discovery rate, pleiotropy, genetic overlap, polygenic architecture, genetic correlation

Abstract word count: 233
Manuscript word count: 3043

Abstract

In recent years, genome-wide association study (GWAS) sample sizes have become larger, the statistical power has improved and thousands of trait-associated variants have been uncovered, offering new insights into the genetic etiology of complex human traits and disorders. However, a large fraction of the polygenic architecture underlying most complex phenotypes still remain undetected. We here review the conditional false discovery rate (condFDR) method, a model-free strategy for analysis of GWAS summary data, which has improved yield of existing GWAS and provided novel findings of genetic overlap between a wide range of complex human phenotypes, including psychiatric, cardiovascular, and neurological disorders, as well as psychological and cognitive traits. The condFDR method was inspired by Empirical Bayes approaches and leverages auxiliary genetic information to improve statistical power for discovery of single-nucleotide polymorphisms (SNPs). The cross-trait condFDR strategy analyses separate GWAS data, and leverages overlapping SNP associations, i.e. cross-trait enrichment, to increase discovery of trait-associated SNPs. The extension of the condFDR approach to conjunctional FDR (conjFDR) identifies shared genomic loci between two phenotypes. The conjFDR approach allows for detection of shared genomic associations irrespective of the genetic correlation between the phenotypes, often revealing a mixture of antagonistic and agonistic directional effects among the shared loci. This review provides a methodological comparison between condFDR and other relevant cross-trait analytical tools and demonstrates how condFDR analysis may provide novel insights into the genetic relationship between complex phenotypes.

Introduction

Most human traits and disorders have a complex etiology, which is influenced by multiple environmental and genetic factors. While some phenotypes follow simple patterns of Mendelian inheritance, large-scale genome-wide association studies (GWAS) conducted during the last decade have shown that most phenotypes have a complex polygenic architecture, in which genetic risk is accounted for by a large number of genetic variants, each with small effect (Visscher et al. 2017). Accumulating evidence from GWAS demonstrates that many genetic variants influence more than one phenotype, i.e. they exhibit allelic pleiotropy (Sivakumaran et al. 2011; Solovieff et al. 2013). Identification of shared genetic influences between human traits and disorders can be highly valuable to inform disease nosology, epidemiological associations, and diagnostic classification systems, improve treatment strategies, provide biological insights and uncover shared biological underpinnings (Sivakumaran et al. 2011; Solovieff et al. 2013; Visscher et al. 2017). For example, it is now evident that psychiatric disorders share a large proportion of their genetic architecture (Brainstorm et al. 2018; Cross-Disorder Group of the Psychiatric Genomics et al. 2013), suggesting that their etiologies are not fully distinct and hence challenging existing diagnostic guidelines (Smoller et al. 2018).

GWAS typically consist of genome-wide scans of millions of common genetic variants (tag single-nucleotide polymorphisms [SNPs]), estimating the strength of their association with the phenotype of interest in massively-univariate regression analyses. Given the large numbers of SNPs tested, a GWAS must correct for multiple testing and applies a genome-wide significance threshold of $p<5x10^{-8}$ to avoid false positive findings. The consequence is that only a subset of all involved genetic variants is revealed (i.e., many false negative findings), with a large fraction of the polygenic architecture remaining to be uncovered. This phenomenon was previously labeled “the missing heritability” (Manolio et al. 2009). With increasing GWAS

sample sizes, statistical power has improved and more genetic variants have been uncovered (Visscher et al. 2017). However, despite the assembly of very large GWAS samples, often involving hundreds of thousands of participants, most of the polygenic architecture underlying complex human phenotypes remain undetected (Holland et al. 2019). The number of participants needed for a GWAS to fully uncover all genetic variants influencing a given phenotype depends on the unique polygenic architecture underlying that phenotype, which is determined by the number of causal variants involved and the distribution of effect sizes (Holland et al. 2019). For example, it has been estimated that to uncover most of the genetic variants influencing the complex disorders schizophrenia and bipolar disorder, genotypes from more than one million individuals are required (Holland et al. 2019).

Improved discovery of shared loci using conditional false discovery rate

Although the successive incremental increases in GWAS sample sizes have effectively improved the discovery of trait-associated loci, an alternative and more cost-efficient approach is to apply statistical tools that improve the yield of existing GWAS. The conditional false discovery rate (condFDR) is such an approach, which boosts GWAS discovery by leveraging auxiliary genetic information to re-adjust the GWAS test-statistics in a primary phenotype (Andreassen et al. 2013b; Schork et al. 2016). The condFDR method is a model-free strategy for analysis of GWAS summary statistics inspired by the Empirical Bayes statistical framework, which is designed for situations with dense elements, such as the large number of small genetic effects seen in polygenic traits and disorders. Most commonly, the condFDR method has been applied for cross-trait analysis, by leveraging overlapping SNP associations (i.e. cross-trait enrichment) between separate GWAS to re-rank the test-statistics in a primary phenotype conditional on the associations in a secondary phenotype (Andreassen et al. 2013b; Schork et al. 2016). Other auxiliary enrichment sources, such as genomic annotations (Schork

et al. 2013), can also be leveraged using condFDR (Lo et al. 2017; Wang et al. 2016b). Since its introduction in 2013 (Andreassen et al. 2013a), the condFDR method has increased genetic discovery in a wide spectrum of complex human traits and disorders, including psychiatric, cardiovascular and neurological disorders, as well as metabolic, psychological and cognitive traits, among others (see Table 1 for a selection of cross-trait condFDR studies) (Andreassen et al. 2013a; Andreassen et al. 2014a; Andreassen et al. 2014c; Andreassen et al. 2013c; Andreassen et al. 2014d; Broce et al. 2018; Broce et al. 2019; Desikan et al. 2015; Drange et al. 2019; Ferrari et al. 2017; Hu et al. 2018; Karch et al. 2018; Le Hellard et al. 2017; LeBlanc et al. 2015; Liu et al. 2013; Lv et al. 2017; McLaughlin et al. 2017; Mufford et al. 2019; Shadrin et al. 2018; Smeland et al. 2019; Smeland et al. 2017a; Smeland et al. 2018; Smeland et al. 2017b; van der Meer et al. 2018; Wang et al. 2016a; Winsvold et al. 2017; Witoelar et al. 2017; Yokoyama et al. 2017; Yokoyama et al. 2016; Zuber et al. 2018).

The present review focuses on the cross-trait condFDR approach, which returns a condFDR value for each SNP, defined as the probability that a SNP is null in the first phenotype (i.e., that it has no association with the phenotype) given that the p-values in the first and second phenotypes are as small as or smaller than the observed ones. The condFDR estimates are obtained for each nominal SNP p-value in the primary phenotype after computing the stratified empirical cumulative distribution functions (cdfs) of the nominal p-values (Sun et al. 2006; Yoo et al. 2009). The separate strata are determined by the relative enrichment of SNP associations as a function of increased nominal SNP p-values in a secondary phenotype. The standard FDR framework derives from a model that assumes that the distribution of test statistics in a GWAS can be formulated as a mixture of null and non-null effects, with true associations (non-null effects) having more extreme test statistics than false associations (null effects) on average. Given a statistical genetic relationship between two phenotypes, stratification of the test-statistics in a primary phenotype based on the genetic associations with a secondary phenotype

will result in a reduction in the FDR at a given nominal p-value relative to the FDR computed from the unstratified distribution of the primary phenotype p-values alone, and thus re-rank the test statistics.

The first step in the condFDR procedure is to construct conditional quantile-quantile (Q-Q) plots, which extends the standard Q-Q plots commonly applied in GWAS. Standard Q-Q plots visualize the enrichment of statistical association relative to that expected under the global null hypothesis by plotting the nominal $-\log_{10}$ p-values of the single SNP association statistics versus their empirical distribution. Conditional Q-Q plots help visualize the cross-trait enrichment between two phenotypes and are constructed by creating subsets of SNPs based of the level of association with the secondary phenotype. Under the global null hypothesis, the nominal p-values will form a straight line plotted as a function of their empirical distribution. Under polygenic association, standard Q-Q plots will be deflected leftwards, while cross-trait enrichment can be seen as successive leftward deflections in conditional Q-Q plots as levels of SNP associations with the secondary phenotype increase. Figure 1a presents a conditional Q-Q plot demonstrating SNP enrichment for the psychiatric disorder bipolar disorder (n=51,710) (Stahl et al. 2019) as a function of the association with intelligence (n=269,867) (Savage et al. 2018), adapted from Smeland et al. (2019). A complementary way to assess for cross-trait enrichment is to construct fold-enrichment plots, which provide a more direct visualization of the polygenic enrichment (Figure 1b). The fold enrichment is calculated as the ratio between the $-\log_{10}(p)$ cumulative distribution for a given stratum and the cumulative distribution for all SNPs. Figure 1b shows that for SNPs with p-values below 0.001 in intelligence, there was up to 60-fold enrichment of stronger SNP associations with bipolar disorder in comparison to all SNPs. The enrichment seen in conditional Q-Q plots and fold-enrichment plots reflects increased tail probabilities in the distribution of test statistics and an overabundance of low p-values compared to that expected by chance, which can be directly interpreted in terms of a

Bayesian interpretation of the true discovery rate (TDR = 1 - FDR; see Box 1 for mathematical framework) (Efron 2010). This is illustrated in Figure 1c.

To control for spurious (i.e. non-generalizable) enrichment due to population stratification or cryptic relatedness (Devlin and Roeder 1999), all test statistics are corrected using a genomic inflation control procedure leveraging intergenic SNPs, which are relatively depleted for true associations (Schork et al. 2013). Conditional-Q-Q plots and the condFDR computation are conducted after random pruning to approximate independence, by selecting one random SNP per LD block (defined by an $r_2 > 0.1$) averaged over at least 100 iterations (Andreassen et al. 2013b; Schork et al. 2016). Similar to previously described stratified-FDR procedures (Sun et al. 2006; Yoo et al. 2009), the condFDR value is then determined for each SNP by constructing a two-dimensional FDR look-up table where the FDR for SNP associations with the primary phenotype is computed conditionally on the nominal p-values for SNP associations with the secondary phenotype (Box 1). Figure 2a presents the respective condFDR look-up table for bipolar disorder conditional on intelligence, corresponding to the cross-trait enrichment observed in Figure 1.

The conjunctional FDR (conjFDR) is an extension of the condFDR, which allows for discovery of SNPs significantly associated with two phenotypes simultaneously (Andreassen et al. 2013a; Schork et al. 2016). The conjFDR is determined after inverting the roles of the primary and secondary phenotypes and repeating the condFDR procedure. Based on previous conjunction tests for p-value statistics (Nichols et al. 2005), the conjFDR is defined as the maximum of the two condFDR values, providing a conservative estimate of the FDR for a SNP association with both phenotypes jointly (Figure 2c). Thus, in combination the condFDR/conjFDR approaches both improve SNP discovery rates (condFDR) and enable detection of shared genomic loci (conjFDR), respectively. Since the condFDR/conjFDR estimates are based on nominal p-values only, these methods are agnostic to the effect directions

of the individual SNPs, and can detect overlapping SNP associations irrespective of the genome-wide genetic correlation between phenotypes. However, after detecting likely overlapping SNPs, the directional SNP effects in the loci can be determined *post hoc* by comparing the effect-sizes (*z*-scores or odds ratios) between the phenotypes.

The condFDR/conjFDR approaches have some limitations. Although all SNPs are randomly pruned using an LD r^2 threshold of 0.1, complex correlations among the test-statistics may bias the condFDR estimates (Schwartzman and Lin 2011). Hence, given strong SNP associations within long range LD regions, such as the extended major histocompatibility complex (MHC) region, chromosomal region 8p.23.1, the microtubule-associated tau protein (*MAPT*) region or the *APOE* region (Price et al. 2008), these regions should be excluded to avoid artificially inflated genetic enrichment. The condFDR/conjFDR procedures are agnostic about the specific causal variants underlying the overlapping genomic associations, which could arise from both shared or separate causal variants, or "mediated pleiotropy", where one phenotype is causative of the other (Solovieff et al. 2013). Given that the cross-trait enrichment both reflects the extent of polygenic overlap between the phenotypes and the power of the two GWAS analyzed, cross-trait enrichment will be harder to detect if one or both investigated GWAS are inadequately powered. Another important limitation of the condFDR method is that a large fraction of overlapping participants between the investigated GWAS may inflate the cross-trait enrichment, and shared participants should therefore be reduced to a minimum. An extension of condFDR, allowing shared controls, has been proposed (Liley and Wallace 2015).

Comparison to other cross-trait analytical tools

A large number of tools for cross-trait analysis using GWAS data have been developed in recent years, which have been reviewed in detail elsewhere (Gratten and Visscher 2016; Hackinger and Zeggini 2017; Pasaniuc and Price 2017; Schork et al. 2016). In short, the methods

differentiate in terms of the data analyzed (summary statistics versus individual genotype data), the underlying mathematical framework and assumptions, whether they are bivariate or multivariate in nature, and whether they measure overlap at the genome-wide level or across individual SNPs or loci/regions. Here we compare the condFDR/conjFDR approach to a selection of relevant cross-trait analytical tools.

The most common approaches for evaluating genetic overlap at the genome-wide level include tools such as polygenic risk scores (Purcell et al. 2009), mixed-model approaches (Cross-Disorder Group of the Psychiatric Genomics et al. 2013; Lee et al. 2012) and LD score regression (Bulik-Sullivan et al. 2015a), which return a single estimate of shared genetic risk between phenotypes. Polygenic risk scores are per-individual risk profiles based on the sum of alleles associated with a phenotype weighted by their effect sizes (Purcell et al. 2009). The polygenic risk score approach uses summary statistics as training data and requires individual genotype data in an independent target sample to test how well the polygenic risk score explains phenotypic variation in the target phenotype. Another traditional measure that estimates the degree of pleiotropy is the genetic correlation, which is defined as the correlation between the genetic influences for a pair of traits, thus indicating the proportion of variance that the two traits share due to genetic causes. Mixed-model approaches (Lee et al. 2012), originally implemented in the Genome-wide Complex Trait Analysis software (GCTA), obtained unbiased estimates of the genetic correlation using individual genotype data, relaxing several limitations of traditional studies based on pedigree data. Estimates of genetic correlation can also be quantified from GWAS summary statistics, using cross-trait LD score regression (Bulik-Sullivan et al. 2015a) and its multivariate extension Genomic SEM (Grotzinger et al. 2019). LD score regression aims to distinguish confounding from polygenicity by regressing the association statistics of SNPs on their ‘LD scores’, which is a measure of the amount of genetic variation the SNP represents (Bulik-Sullivan et al. 2015b). Application of LD score

regression to the bivariate framework estimates the co-variance in the SNP-heritability between two phenotypes, allowing sample overlap (Bulik-Sullivan et al. 2015a). An alternative approach estimating local genetic correlations based on the fixed-effects model is also available (Shi et al. 2017). The condFDR approach is fundamentally different to these approaches by aiming for discovery of specific genomic loci. However, the condFDR approach similarly focuses on the polygenic fraction that did not reach genome-wide significance to uncover cross-trait enrichment. To fully disentangle the genetic relationship between complex phenotypes it is necessary to complement measures of genetic overlap at the genome-wide level with cross-trait analytical tools allowing detection of individual shared loci regardless of their directional effects. For instance, a recent condFDR study demonstrated substantial cross-trait enrichment between bipolar disorder (Stahl et al. 2019) and intelligence (Savage et al. 2018) (Figure 1) and uncovered a balanced pattern of concordant and discordant directional effects among 79 shared loci identified at conjFDR<0.05 (Figure 3) (Smeland et al. 2019). These findings extend and complies with prior genetic studies reporting no significant genome-wide genetic correlation between the phenotypes (Brainstorm et al. 2018; Davies et al. 2018; Hill et al. 2016; Lencz et al. 2014; Savage et al. 2018; Sniekers et al. 2017; Stahl et al. 2019).

There is a large class of cross-trait methods aiming to discover specific genomic loci unique or shared between phenotypes inspired by the meta-analysis technique (Willer et al. 2010) and its extensions dealing with sample overlap (Han et al. 2016; Lin and Sullivan 2009). For example, the COMBINE approach (Ellinghaus et al. 2012) consists of two separate runs of a same-effect and opposite-effect meta-analysis, both using the inverse variance weighted procedure. In the opposite-effect meta-analysis, the minor and major alleles are flipped in the second dataset to capture bi-allelic SNPs with opposite effect directions in the two phenotypes investigated. This method was later refined and extended to multiple heterogeneous traits using restricted and weighted subset search (ASSET) (Bhattacharjee et al. 2012), which exhaustively

explore subsets of studies to achieve the best possible trade-off between specificity and sample size. Its successor, compare-and-contrast meta-analysis (CCMA) (Baurecht et al. 2015), further improved the power to discover associations by combining the subset search approach with trans-ethnic meta-analysis (MANTRA) (Morris 2011). Several alternative approaches explore additional information, including individual-level genotypes (MultiPhen) (O'Reilly et al. 2012), phenotypic correlations (TATES) (van der Sluis et al. 2013) or estimated genetic correlations (MTAG) (Turley et al. 2018). A common feature of all techniques based on a meta-analysis framework is that the analysis is performed independently for each SNP, thus requiring a follow-up mechanism to control for multiple testing, such as Bonferroni correction, to avoid false positive findings. The condFDR analysis, on the other hand, directly works with the entire original set of p-values from the two GWAS and intrinsically incorporates multiple testing via the FDR framework (Efron 2010).

Another class of methods aim at disentangling LD structure to reveal underlying causal genetic mechanisms. Mendelian Randomization aims to distinguish true pleiotropy from mediated pleiotropy by investigating whether one phenotype is causative to the other (Hernan and Robins 2006; Lawlor et al. 2008; Smith and Ebrahim 2003; Zhu et al. 2018). Mendelian Randomization assigns genetic variants, which are expected to be independent of confounding factors, as instrumental variables to test for causality. Several available Bayesian approaches (Giambartolomei et al. 2014; Pickrell et al. 2016) explore whether two association signals in the same genomic region obtained from two different GWAS share a single causal variant or multiple causal variants. Frei and colleagues performed a similar analysis at the genome-wide level, estimating the proportion of phenotype-specific causal variants and shared variants between complex phenotypes using GWAS summary data, while controlling for shared participants (Frei et al. 2019). The analysis demonstrates how the shared polygenic component may constitute a large fraction of the genetic architecture of one phenotype, while constituting

a smaller fraction of the architecture of a phenotype with larger polygenicity. While the condFDR/conjFDR approach is agnostic about the causal variants underlying the identified associations, it complements these methods by improving the discovery of genomic loci, which can be used to prioritize down-stream analysis.

Conclusion

Accumulating evidence has shown that genetic pleiotropy is pervasive among complex human traits and disorders, providing important insights into etiological relationships. Since its introduction in 2013, application of the condFDR/conjFDR approach has increased yield of existing GWAS and aided the discovery of overlapping genomic loci between polygenic phenotypes. Given that large fractions of the polygenic architecture underlying most complex phenotypes still remain undetected, the condFDR/conjFDR approach represents a cost-effective powerful strategy useful for improving GWAS discovery and help elucidating shared genetic etiologies.

Conflict of Interest Disclosures: O.A.A. has received speaker's honorarium from Lundbeck and is a consultant for Healthlytix. C.C.F. is under employment of Multimodal Imaging Service, dba Healthlytix, in addition to his research appointment at the University of California, San Diego. A.M.D. is a founder of and holds equity interest in CorTechs Labs and serves on its scientific advisory board. He is also a member of the Scientific Advisory Board of Healthlytix and receives research funding from General Electric Healthcare (GEHC). The terms of these arrangements have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policies. Remaining authors have no conflicts of interest to declare.

URLs: The condFDR/conjFDR software is available on https://github.com/precimed/pleiofdr as a MATLAB package, under GPL v3 license.

References

Andreassen OA et al. (2013a) Improved detection of common variants associated with schizophrenia by leveraging pleiotropy with cardiovascular-disease risk factors American journal of human genetics 92:197-209 doi:10.1016/j.ajhg.2013.01.001

Andreassen OA et al. (2014a) Genetic pleiotropy between multiple sclerosis and schizophrenia but not bipolar disorder: differential involvement of immune-related gene loci Molecular psychiatry 20:207-214 doi:10.1038/mp.2013.195

Andreassen OA et al. (2014b) Identifying common genetic variants in blood pressure due to polygenic pleiotropy with associated phenotypes Hypertension 63:819-826 doi:10.1161/hypertensionaha.113.02077

Andreassen OA et al. (2014c) Identifying common genetic variants in blood pressure due to polygenic pleiotropy with associated phenotypes Hypertension 63:819-826 doi:10.1161/HYPERTENSIONAHA.113.02077

Andreassen OA, Thompson WK, Dale AM (2013b) Boosting the Power of Schizophrenia Genetics by Leveraging New Statistical Tools Schizophrenia bulletin doi:10.1093/schbul/sbt168

Andreassen OA et al. (2013c) Improved detection of common variants associated with schizophrenia and bipolar disorder using pleiotropy-informed conditional false discovery rate PLoS genetics 9:e1003455 doi:10.1371/journal.pgen.1003455

Andreassen OA et al. (2014d) Shared common variants in prostate cancer and blood lipids International journal of epidemiology 43:1205-1214 doi:10.1093/ije/dyu090

Baurecht H et al. (2015) Genome-wide comparative analysis of atopic dermatitis and psoriasis gives insight into opposing genetic mechanisms Am J Hum Genet 96:104-120 doi:10.1016/j.ajhg.2014.12.004

Bhattacharjee S et al. (2012) A Subset-Based Approach Improves Power and Interpretation for the Combined Analysis of Genetic Association Studies of Heterogeneous Traits The American Journal of Human Genetics 90:821-835 doi:https://doi.org/10.1016/j.ajhg.2012.03.015

Brainstorm C et al. (2018) Analysis of shared heritability in common disorders of the brain Science 360 doi:10.1126/science.aap8757

Broce I et al. (2018) Immune-related genetic enrichment in frontotemporal dementia: An analysis of genome-wide association studies PLoS medicine 15:e1002487 doi:10.1371/journal.pmed.1002487

Broce IJ et al. (2019) Dissecting the genetic relationship between cardiovascular risk factors and Alzheimer's disease Acta Neuropathol 137:209-226 doi:10.1007/s00401-018-1928-6

Bulik-Sullivan B et al. (2015a) An atlas of genetic correlations across human diseases and traits Nature genetics 47:1236-1241 doi:10.1038/ng.3406

Bulik-Sullivan BK et al. (2015b) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies Nature genetics 47:291-295 doi:10.1038/ng.3211

Cross-Disorder Group of the Psychiatric Genomics C et al. (2013) Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs Nature genetics 45:984-994 doi:10.1038/ng.2711

Davies G et al. (2018) Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function Nat Commun 9:2098 doi:10.1038/s41467-018-04362-x

Desikan RS et al. (2015) Polygenic Overlap Between C-Reactive Protein, Plasma Lipids, and Alzheimer Disease Circulation 131:2061-2069 doi:10.1161/CIRCULATIONAHA.115.015489

Devlin B, Roeder K (1999) Genomic control for association studies Biometrics 55:997-1004

Drange OK et al. (2019) Genetic Overlap Between Alzheimer's Disease and Bipolar Disorder Implicates the MARK2 and VAC14 Genes Front Neurosci 13:220 doi:10.3389/fnins.2019.00220
Efron B (2007) Size, power and false discovery rates The Annals of Statistics 35:1351–1377
Efron B (2010) Large-scale inference : empirical Bayes methods for estimation, testing, and prediction. Institute of mathematical statistics monographs, vol 1. Cambridge University Press, Cambridge ; New York
Efron B, Tibshirani R (2002) Empirical bayes methods and false discovery rates for microarrays Genetic epidemiology 23:70-86 doi:10.1002/gepi.1124
Ellinghaus D et al. (2012) Combined analysis of genome-wide association studies for Crohn disease and psoriasis identifies seven shared susceptibility loci American journal of human genetics 90:636-647 doi:10.1016/j.ajhg.2012.02.020
Ferrari R et al. (2017) Genetic architecture of sporadic frontotemporal dementia and overlap with Alzheimer's and Parkinson's diseases J Neurol Neurosurg Psychiatry 88:152-164 doi:10.1136/jnnp-2016-314411
Frei O et al. (2019) Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation Nat Commun 10:2417 doi:10.1038/s41467-019-10310-0
Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V (2014) Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics PLoS genetics 10:e1004383 doi:10.1371/journal.pgen.1004383
Gratten J, Visscher PM (2016) Genetic pleiotropy in complex traits and diseases: implications for genomic medicine Genome medicine 8:78 doi:10.1186/s13073-016-0332-x
Grotzinger AD et al. (2019) Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits Nature Human Behaviour doi:10.1038/s41562-019-0566-x
Hackinger S, Zeggini E (2017) Statistical methods to detect pleiotropy in human complex traits Open Biol 7 doi:10.1098/rsob.170125
Han B, Duong D, Sul JH, de Bakker PI, Eskin E, Raychaudhuri S (2016) A general framework for meta-analyzing dependent studies with overlapping subjects in association mapping Human molecular genetics 25:1857-1866 doi:10.1093/hmg/ddw049
Hernan MA, Robins JM (2006) Instruments for causal inference: an epidemiologist's dream? Epidemiology (Cambridge, Mass) 17:360-372 doi:10.1097/01.ede.0000222409.00878.37
Hill WD, Davies G, Group CCW, Liewald DC, McIntosh AM, Deary IJ (2016) Age-Dependent Pleiotropy Between General Cognitive Function and Major Psychiatric Disorders Biological psychiatry 80:266-273 doi:10.1016/j.biopsych.2015.08.033
Holland D et al. (2019) Beyond SNP Heritability: Polygenicity and Discoverability of Phenotypes Estimated with a Univariate Gaussian Mixture Model bioRxiv:133132 doi:10.1101/133132
Hu Y et al. (2018) Identification of Novel Potentially Pleiotropic Variants Associated With Osteoporosis and Obesity Using the cFDR Method J Clin Endocrinol Metab 103:125-138 doi:10.1210/jc.2017-01531
Karch CM et al. (2018) Selective Genetic Overlap Between Amyotrophic Lateral Sclerosis and Diseases of the Frontotemporal Dementia Spectrum JAMA neurology 75:860-875 doi:10.1001/jamaneurol.2018.0372
Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G (2008) Mendelian randomization: using genes as instruments for making causal inferences in epidemiology Statistics in medicine 27:1133-1163 doi:10.1002/sim.3034
Le Hellard S et al. (2017) Identification of Gene Loci That Overlap Between Schizophrenia and Educational Attainment Schizophrenia bulletin 43:654-664 doi:10.1093/schbul/sbw085

LeBlanc M et al. (2015) Identifying Novel Gene Variants in Coronary Artery Disease and Shared Genes with Several Cardiovascular Risk Factors Circulation research doi:10.1161/circresaha.115.306629
Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR (2012) Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood Bioinformatics 28:2540-2542 doi:10.1093/bioinformatics/bts474
Lencz T et al. (2014) Molecular genetic evidence for overlap between general cognitive ability and risk for schizophrenia: a report from the Cognitive Genomics consorTium (COGENT) Molecular psychiatry 19:168-174 doi:10.1038/mp.2013.166
Liley J, Wallace C (2015) A pleiotropy-informed Bayesian false discovery rate adapted to a shared control design finds new disease associations from GWAS summary statistics PLoS genetics 11:e1004926 doi:10.1371/journal.pgen.1004926
Lin DY, Sullivan PF (2009) Meta-analysis of genome-wide association studies with overlapping subjects American journal of human genetics 85:862-872 doi:10.1016/j.ajhg.2009.11.001
Liu JZ et al. (2013) Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis Nature genetics 45:670-675 doi:10.1038/ng.2616
Lo MT et al. (2017) Modeling prior information of common genetic variants improves gene discovery for neuroticism Human molecular genetics 26:4530-4539 doi:10.1093/hmg/ddx340
Lv WQ et al. (2017) Novel common variants associated with body mass index and coronary artery disease detected using a pleiotropic cFDR method J Mol Cell Cardiol 112:1-7 doi:10.1016/j.yjmcc.2017.08.011
Manolio TA et al. (2009) Finding the missing heritability of complex diseases Nature 461:747-753 doi:10.1038/nature08494
McLaughlin RL et al. (2017) Genetic correlation between amyotrophic lateral sclerosis and schizophrenia Nat Commun 8:14774 doi:10.1038/ncomms14774
Morris AP (2011) Transethnic meta-analysis of genomewide association studies Genetic epidemiology 35:809-822 doi:10.1002/gepi.20630
Mufford M et al. (2019) Concordance of genetic variation that increases risk for tourette syndrome and that influences its underlying neurocircuitry Transl Psychiatry 9:120 doi:10.1038/s41398-019-0452-3
Nichols T, Brett M, Andersson J, Wager T, Poline JB (2005) Valid conjunction inference with the minimum statistic Neuroimage 25:653-660 doi:10.1016/j.neuroimage.2004.12.005
O'Reilly PF, Hoggart CJ, Pomyen Y, Calboli FCF, Elliott P, Jarvelin M-R, Coin LJM (2012) MultiPhen: Joint Model of Multiple Phenotypes Can Increase Discovery in GWAS PLOS ONE 7:e34861 doi:10.1371/journal.pone.0034861
Pasaniuc B, Price AL (2017) Dissecting the genetics of complex traits using summary association statistics Nature reviews Genetics 18:117-127 doi:10.1038/nrg.2016.142
Pickrell JK, Berisa T, Liu JZ, Ségurel L, Tung JY, Hinds DA (2016) Detection and interpretation of shared genetic influences on 42 human traits Nature genetics 48:709-717 doi:10.1038/ng.3570
Price AL et al. (2008) Long-range LD can confound genome scans in admixed populations American journal of human genetics 83:132-135; author reply 135-139 doi:10.1016/j.ajhg.2008.06.005
Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder Nature 460:748-752 doi:10.1038/nature08185

Savage JE et al. (2018) Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence Nature genetics 50:912-919 doi:10.1038/s41588-018-0152-6
Schork AJ et al. (2013) All SNPs are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs PLoS genetics 9:e1003449 doi:10.1371/journal.pgen.1003449
Schork AJ, Wang Y, Thompson WK, Dale AM, Andreassen OA (2016) New statistical approaches exploit the polygenic architecture of schizophrenia--implications for the underlying neurobiology Curr Opin Neurobiol 36:89-98 doi:10.1016/j.conb.2015.10.008
Schwartzman A, Lin X (2011) The effect of correlation in false discovery rate estimation Biometrika 98:199-214 doi:10.1093/biomet/asq075
Shadrin AA et al. (2018) Novel Loci Associated With Attention-Deficit/Hyperactivity Disorder Are Revealed by Leveraging Polygenic Overlap With Educational Attainment J Am Acad Child Adolesc Psychiatry 57:86-95 doi:10.1016/j.jaac.2017.11.013
Shi H, Mancuso N, Spendlove S, Pasaniuc B (2017) Local Genetic Correlation Gives Insights into the Shared Genetic Architecture of Complex Traits American journal of human genetics 101:737-751 doi:10.1016/j.ajhg.2017.09.022
Sivakumaran S et al. (2011) Abundant pleiotropy in human complex diseases and traits American journal of human genetics 89:607-618 doi:10.1016/j.ajhg.2011.10.004
Smeland OB et al. (2019) Genome-wide analysis reveals extensive genetic overlap between schizophrenia, bipolar disorder, and intelligence Molecular psychiatry doi:10.1038/s41380-018-0332-x
Smeland OB et al. (2017a) Identification of Genetic Loci Jointly Influencing Schizophrenia Risk and the Cognitive Traits of Verbal-Numerical Reasoning, Reaction Time, and General Cognitive Function JAMA psychiatry 74:1065-1075 doi:10.1001/jamapsychiatry.2017.1986
Smeland OB et al. (2018) Genetic Overlap Between Schizophrenia and Volumes of Hippocampus, Putamen, and Intracranial Volume Indicates Shared Molecular Genetic Mechanisms Schizophrenia bulletin 44:854-864 doi:10.1093/schbul/sbx148
Smeland OB et al. (2017b) Identification of genetic loci shared between schizophrenia and the Big Five personality traits Sci Rep 7:2222 doi:10.1038/s41598-017-02346-3
Smith GD, Ebrahim S (2003) 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? International journal of epidemiology 32:1-22
Smoller JW, Andreassen OA, Edenberg HJ, Faraone SV, Glatt SJ, Kendler KS (2018) Psychiatric genetics and the structure of psychopathology Molecular psychiatry doi:10.1038/s41380-017-0010-4
Sniekers S et al. (2017) Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence Nature genetics 49:1107-1112 doi:10.1038/ng.3869
Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW (2013) Pleiotropy in complex traits: challenges and strategies Nature reviews Genetics 14:483-495 doi:10.1038/nrg3461
Stahl EA et al. (2019) Genome-wide association study identifies 30 loci associated with bipolar disorder Nature genetics 51:793-803 doi:10.1038/s41588-019-0397-8
Sun L, Craiu RV, Paterson AD, Bull SB (2006) Stratified false discovery control for large-scale hypothesis testing with application to genome-wide association studies Genetic epidemiology 30:519-530 doi:10.1002/gepi.20164
Turley P et al. (2018) Multi-trait analysis of genome-wide association summary statistics using MTAG Nat Genet 50:229-237 doi:10.1038/s41588-017-0009-4

van der Meer D et al. (2018) Brain scans from 21,297 individuals reveal the genetic architecture of hippocampal subfield volumes Molecular psychiatry doi:10.1038/s41380-018-0262-7
van der Sluis S, Posthuma D, Dolan CV (2013) TATES: Efficient Multivariate Genotype-Phenotype Analysis for Genome-Wide Association Studies PLOS Genetics 9:e1003235 doi:10.1371/journal.pgen.1003235
Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J (2017) 10 Years of GWAS Discovery: Biology, Function, and Translation American journal of human genetics 101:5-22 doi:10.1016/j.ajhg.2017.06.005
Wang Y et al. (2016a) Genetic overlap between multiple sclerosis and several cardiovascular disease risk factors Mult Scler 22:1783-1793 doi:10.1177/1352458516635873
Wang Y et al. (2016b) Leveraging Genomic Annotations and Pleiotropic Enrichment for Improved Replication Rates in Schizophrenia GWAS PLoS genetics 12:e1005803 doi:10.1371/journal.pgen.1005803
Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans Bioinformatics 26:2190-2191 doi:10.1093/bioinformatics/btq340
Winsvold BS et al. (2017) Shared genetic risk between migraine and coronary artery disease: A genome-wide analysis of common variants PloS one 12:e0185663 doi:10.1371/journal.pone.0185663
Witoelar A et al. (2017) Genome-wide Pleiotropy Between Parkinson Disease and Autoimmune Diseases JAMA neurology 74:780-792 doi:10.1001/jamaneurol.2017.0469
Yokoyama JS et al. (2017) Shared genetic risk between corticobasal degeneration, progressive supranuclear palsy, and frontotemporal dementia Acta Neuropathol 133:825-837 doi:10.1007/s00401-017-1693-y
Yokoyama JS et al. (2016) Association Between Genetic Traits for Immune-Mediated Diseases and Alzheimer Disease JAMA neurology 73:691-697 doi:10.1001/jamaneurol.2016.0150
Yoo YJ, Pinnaduwage D, Waggott D, Bull SB, Sun L (2009) Genome-wide association analyses of North American Rheumatoid Arthritis Consortium and Framingham Heart Study data utilizing genome-wide linkage results BMC proceedings 3 Suppl 7:S103
Zhu Z et al. (2018) Causal associations between risk factors and common diseases inferred from GWAS summary data Nature Communications 9:224 doi:10.1038/s41467-017-02317-2
Zuber V et al. (2018) Identification of shared genetic variants between schizophrenia and lung cancer Sci Rep 8:674 doi:10.1038/s41598-017-16481-4

Box 1: Conditional and conjunctional False Discovery Rate

The 'enrichment' seen in the conditional Q-Q plots can be directly interpreted in terms of a Bayesian interpretation of the true discovery rate (TDR = 1 – false discovery rate (FDR)) (Efron 2010). More specifically, for a given p-value, under a simple two-group (null and non-null) model, Bayes rule gives the posterior probability of being null as

$$FDR(p) = \pi_0 F_0(p) / F(p), \qquad [1]$$

where π_0 is the proportion of null SNPs, F_0 is the cumulative distribution function (cdf) of the null SNPs, and F is the cdf of all SNPs, both null and non-null (Efron 2007). Here, we assume the SNP p-values are *a priori* independent and identically distributed. Under the null hypothesis, F_0 is the cdf of the uniform distribution on the unit interval [0,1], so that Eq. [1] reduces to

$$FDR(p) = \pi_0 p / F(p). \quad [2]$$

F can be estimated by the empirical cdf $q = N_p / N$, where N_p is the number of SNPs with p-values less than or equal to p, and N is the total number of SNPs. Replacing F by q in Eq. [2], we get

$$\text{Estimated } FDR(p) = \pi_0 p / q, \quad [3]$$

which is biased upwards as an estimate of the FDR (Efron and Tibshirani 2002). Replacing π_0 in Equation [3] with unity gives an estimated FDR that is further biased upward;

$$q^* = p/q. \qquad [4]$$

If π_0 is close to one, the increase in bias going from Eq. [3] to Eq. [4] is minimal. The quantity 1 – p/q, is therefore biased downward, and hence a conservative estimate of the TDR. Referring to the Q-Q plots, we see that q* is equivalent to the nominal p-value divided by the empirical quantile, as defined earlier. We can thus read the FDR estimate directly off the Q-Q plot as

$$-\log_{10}(q^*) = \log_{10}(q) - \log_{10}(p), \qquad [5]$$

i.e. the horizontal shift of the curves in the Q-Q plots from the expected line x = y, with a larger

shift corresponding to a smaller FDR. To estimate the conditional FDR of a given SNP, we repeat the above procedure for a subset of SNPs with p-values in the secondary GWAS equal to or lower than that observed for the given SNP. Formally, this is given by

$$FDR(p_1|p_2) = \pi_0(p_2)p_1 / F(p_1|p_2), \qquad [6]$$

where p_1 is the p-value for the first phenotype, p_2 is the p-value for the second, and $F(p_1 \mid p_2)$ is the conditional cdf and $\pi_0(p_2)$ the conditional proportion of null SNPs for the first phenotype given that p-values for the second phenotype are p_2 or smaller. The condFDR framework is closely related to the stratified FDR method developed by Sun et al. (2006). Whereas they propose computing FDR separately conditional on membership in pre-defined discrete strata of p-values, here, we condition the estimated FDR on a continuous random variable, the SNP p-values with respect to a second phenotype.

To identify SNPs jointly associated with two phenotypes using conjunctional FDR, the conditional FDR procedure is repeated after inverting the roles of the primary and secondary phenotypes. Similar to previous conjunction tests for p-value statistics (Nichols et al. 2005), the conjunctional FDR estimate is defined as the maximum of both conditional FDR values, which minimizes the effect of a single phenotype driving the common association signal. Formally, the conjunctional FDR is given by

$$FDR_{Phenotype1\&Phenotype2}(p_1, p_2) = \pi_0 F_0(p_1, p_2) / F(p_1, p_2) + \pi_1 F_1(p_1, p_2) / F(p_1, p_2) + \pi_2 F_2(p_1, p_2) / F(p_1, p_2), \qquad [7]$$

where π_0 is the *a priori* proportion of SNPs null for both phenotypes simultaneously and $F_0(p_1, p_2)$ is the joint null cdf, π_1 is the *a priori* proportion of SNPs non-null for the first phenotype and null for the second with $F_1(p_1, p_2)$ the joint cdf of these SNPs, and π_2 is the *a priori* proportion of SNPs non-null for the second phenotype and null for the first, with joint cdf $F_2(p_1, p_2)$. $F(p_1, p_2)$ is the joint overall mixture cdf for all phenotype 1 and 2 SNPs.

Conditional empirical cdfs provide a model-free method to obtain conservative estimates of Eq (7). This can be seen as follows. Estimate the conjunction FDR by

$$\text{Estimated FDR}_{\text{Phenotype1\&Phenotype2}} = \max\{\text{Estimated FDR}_{\text{Phenotype1|Phenotype2}}, \text{Estimated FDR}_{\text{Phenotype2|Phenotype1}}\}, \quad [8]$$

where $\text{Estimated FDR}_{\text{Phenotype1|Phenotype2}}$ and $\text{Estimated FDR}_{\text{Phenotype2|Phenotype1}}$ are conservative (upwardly biased) estimates of Eq. [6]. Thus, Eq (8) is a conservative estimate of $\max\{p_1/F(p_1|p_2), p_2/F(p_2|p_1)\} = \max\{p_1F_2(p_2)/F(p_1, p_2), p_2F_1(p_1)/F(p_1, p_2)\}$, with $F_1(p_1)$ and $F_2(p_2)$ the marginal non-null cdfs of SNPs for phenotype 1 and 2, respectively. For enriched samples, p-values will tend to be smaller than predicted from the uniform distribution, so that $F_1(p_1) \geq p_1$ and $F_2(p_2) \geq p_2$. Then

$$\max\{p_1F_2(p_2) / F(p_1, p_2), p_2F_1(p_1) / F(p_1, p_2)\}$$

$$\geq [\pi_0 + \pi_1 + \pi_2] \max\{p_1F_2(p_2) / F(p_1, p_2), p_2F_1(p_1) / F(p_1, p_2)\}$$

$$\geq [\pi_0p_1p_2 + \pi_1p_2F_1(p_1) + \pi_2p_1F_2(p_2)] / F(p_1, p_2).$$

Under the assumption that SNPs are independent if one or both are null, reasonable for disjoint samples, this last quantity is precisely the conjunctional FDR given in Eq (7). Thus, Eq (8) is a conservative model-free estimate of the conjunctional FDR.

1

Table 1. Selected cross-trait conditional false discovery rate studies

Primary phenotype	*Secondary phenotype*	*Novel loci for primary phenotype*	*Citation*
Schizophrenia	Cardiovascular-disease risk factors	14 at condFDR<0.01	(Andreassen et al. 2013a)
Primary sclerosing cholangitis	Autoimmune diseases	33 at condFDR<0.001	(Liu et al. 2013)
Bipolar disorder	Schizophrenia	2 at condFDR<0.01	(Andreassen et al. 2013c)
Schizophrenia	Multiple sclerosis	5 at condFDR<0.01	(Andreassen et al. 2014a)
Systolic blood pressure	Comorbid traits and diseases	42 at condFDR<0.01	(Andreassen et al. 2014b)
Alzheimer disease	C-reactive protein, plasma lipids	55 at condFDR<0.05	(Desikan et al. 2015)
Coronary artery disease	Cardiovascular-disease risk factors	67 at condFDR<0.01	(LeBlanc et al. 2015)
Alzheimer disease	Autoimmune diseases	Not available	(Yokoyama et al. 2016)
Amyotrophic lateral sclerosis	Schizophrenia	5 at condFDR<0.01	(McLaughlin et al. 2017)
Schizophrenia	Educational attainment	23 at condFDR<0.01	(Le Hellard et al. 2017)
Sporadic frontotemporal dementia	Alzheimer disease, Parkinson disease	13 at condFDR<0.05	(Ferrari et al. 2017)
Schizophrenia	Cognitive traits	13 at conjFDR<0.05	(Smeland et al. 2017a)
Corticobasal degeneration	Progressive supranuclear palsy, frontotemporal dementia	3 at conjFDR<0.05	(Yokoyama et al. 2017)
Amyotrophic lateral sclerosis	Neurodegenerative disorders	22 at condFDR<0.05	(Karch et al. 2018)
Frontotemporal dementia	Autoimmune diseases	5 at conjFDR<0.05	(Broce et al. 2018)
Schizophrenia	Subcortical brain volumes	3 at conjFDR<0.05	(Smeland et al. 2018)
Attention-deficit/ hyperactivity disorder	Educational attainment	4 at condFDR<0.01, 1 at conjFDR<0.05	(Shadrin et al. 2018)
Alzheimer disease	Cardiovascular-disease risk factors	4 at conjFDR<0.05	(Broce et al. 2019)
Schizophrenia, bipolar disorder	Intelligence	20 schizophrenia loci and 4 bipolar disorder loci at conjFDR<0.01	(Smeland et al. 2019)

2

Figures

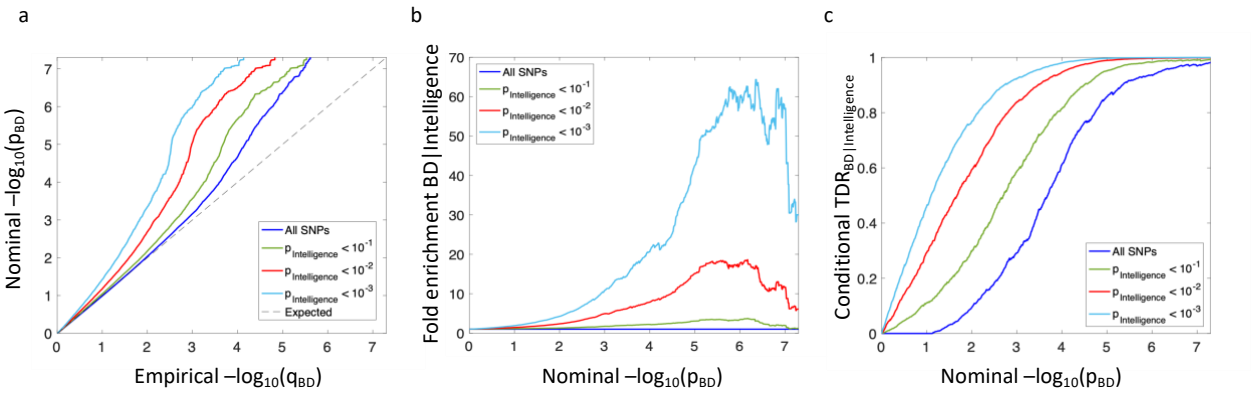

Figure 1. Cross-trait enrichment between bipolar disorder (BD; n=51,710) (Stahl et al. 2019) and intelligence (n=269,867) (Savage et al. 2018), adapted from Smeland et al. (2019). (**a**) Conditional Q-Q plot displaying the nominal -log10 p-values of the single SNP association statistics versus their empirical distribution in BD below the standard GWAS threshold of $p<5\times10^{-8}$ as a function of significance of association with intelligence at the level of $p \leqslant 0.1$, $p \leqslant 0.01$, $p \leqslant 0.001$, respectively. The blue line indicates all SNPs. The dashed line indicates the null hypothesis. **(b**) Fold-enrichment plot of enrichment versus nominal -log10 p-values in BD as a function of association with intelligence. **(c**) Conditional true discovery rate (TDR) plot illustrating the increase in TDR associated with increased enrichment in BD conditioned on intelligence. The test statistics were corrected for genomic inflation, SNPs were randomly pruned across 500 iterations using a linkage disequilibrium r^2 threshold of 0.1, and the extended major histocompatibility complex region and chromosomal region 8p.23.1 were excluded (Smeland et al. 2019).

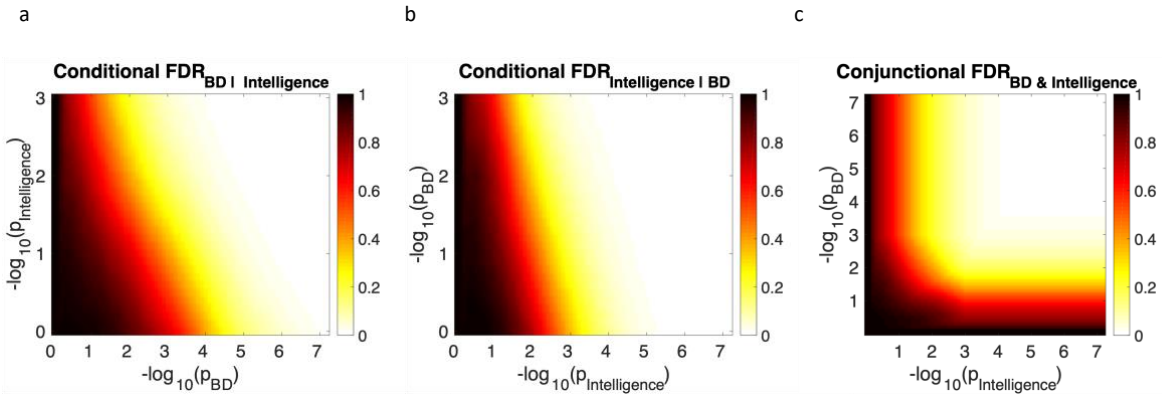

Figure 2. (**a**) Conditional false discovery rate (condFDR) 2D look-up table for SNP associations with bipolar disorder (BD) conditional on SNP associations with intelligence, corresponding to the cross-trait enrichment observed in Figure 1. The FDR in BD SNPs are computed conditionally on the nominal intelligence p-values. **(b**) condFDR 2D look-up table for SNP associations with intelligence conditional on SNP associations with BD. (**c**) Corresponding conjunctional FDR (conjFDR) 2D look-up table for SNP associations shared between BD and intelligence. The color refers to the FDR values.

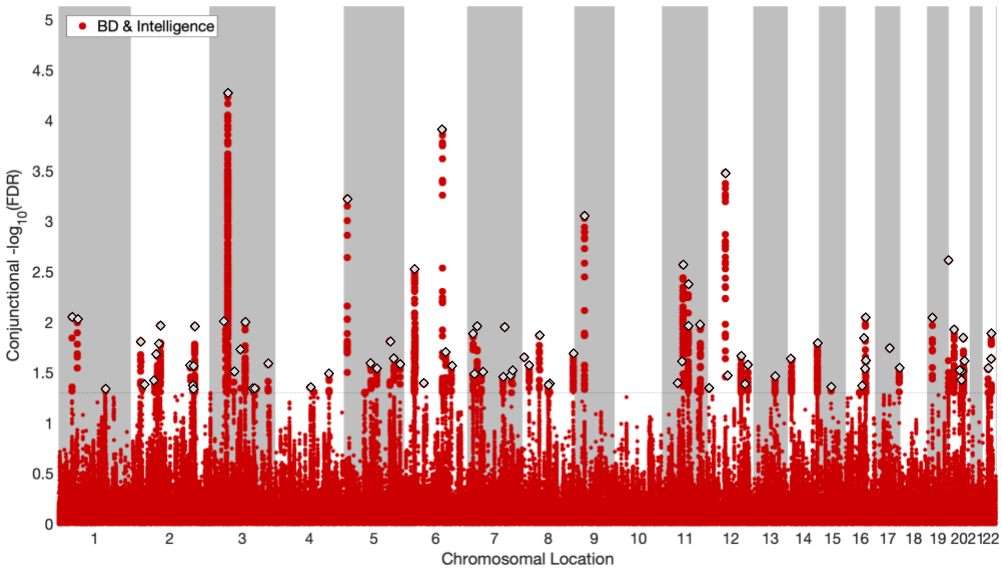

Figure 3. Common genetic variants jointly associated with bipolar disorder (BD; n = 51,710) and intelligence (n = 269,867) at conjunctional false discovery rate (conjFDR) < 0.05, adapted from Smeland et al. (2019). Manhattan plot showing the – $\log_{10}$ transformed conjFDR values for each SNP on the y axis and chromosomal positions along the x axis. The dotted horizontal line represents the threshold for significant shared associations (conjFDR < 0.05, ie, –$\log_{10}$ (conjFDR) > 1.3). Independent lead SNPs are encircled in black. For details, see Supplementary Table 9 in Smeland et al. (2019).