

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

2

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

10

11 ¹NORMENT Centre, Institute of Clinical Medicine, University of Oslo and Division of Mental
12 Health and Addiction, Oslo University Hospital, 0407 Oslo, Norway; ²Department of Cognitive
13 Science, University of California, San Diego, La Jolla, CA, USA; United States of America;
14 ³Department of Radiology, University of California, San Diego, La Jolla, CA 92093, United
15 States of America; ⁴Department of Psychiatry, University of California, San Diego, La Jolla,
16 CA, USA; ⁵Department of Neuroscience, University of California San Diego, La Jolla, CA
17 92093; ⁶Center for Multimodal Imaging and Genetics, University of California San Diego, La
18 Jolla, CA 92093, United States of America; ⁷Department of Medical Genetics, Oslo University
19 Hospital, Oslo, Norway; ⁸NORMENT Centre, Department of Clinical Science, University of
20 Bergen, Bergen, Norway

21

22

23

24

25 **Corresponding authors:** Olav B Smeland or Ole A Andreassen

1 Olav B. Smeland M.D. Ph.D.
2 Postdoctoral researcher
3 Division of Mental Health and Addiction
4 University of Oslo and Oslo University Hospital
5 Kirkeveien 166, 0424 Oslo, Norway
6 Email: o.b.smeland@medisin.uio.no
7 Phone: +47 41220844
8

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

9 **Acknowledgements:** National Institutes of Health (NS057198; EB00790); National Institutes
10 of Health NIDA/NCI: U24DA041123; the Research Council of Norway (229129; 213837;
11 248778; 223273; 249711); the South-East Norway Regional Health Authority (2017-112); KG
12 Jebsen Stiftelsen (SKGJ-2011-36).

13

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

16

17 Abstract word count: 233
18 Manuscript word count: 3043

19

20

21

1 **Abstract**

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

22

23

1 **Introduction**

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

18 GWAS typically consist of genome-wide scans of millions of common genetic variants
19 (tag single-nucleotide polymorphisms [SNPs]), estimating the strength of their association with
20 the phenotype of interest in massively-univariate regression analyses. Given the large numbers
21 of SNPs tested, a GWAS must correct for multiple testing and applies a genome-wide
22 significance threshold of $p < 5 \times 10^{-8}$ to avoid false positive findings. The consequence is that only
23 a subset of all involved genetic variants is revealed (i.e., many false negative findings), with a
24 large fraction of the polygenic architecture remaining to be uncovered. This phenomenon was
25 previously labeled “the missing heritability” (Manolio et al. 2009). With increasing GWAS

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

11

12 **Improved discovery of shared loci using conditional false discovery rate**

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

1 et al. 2013), can also be leveraged using condFDR (Lo et al. 2017; Wang et al. 2016b). Since
2 its introduction in 2013 (Andreassen et al. 2013a), the condFDR method has increased genetic
3 discovery in a wide spectrum of complex human traits and disorders, including psychiatric,
4 cardiovascular and neurological disorders, as well as metabolic, psychological and cognitive
5 traits, among others (see Table 1 for a selection of cross-trait condFDR studies) (Andreassen et
6 al. 2013a; Andreassen et al. 2014a; Andreassen et al. 2014c; Andreassen et al. 2013c;
7 Andreassen et al. 2014d; Broce et al. 2018; Broce et al. 2019; Desikan et al. 2015; Drange et
8 al. 2019; Ferrari et al. 2017; Hu et al. 2018; Karch et al. 2018; Le Hellard et al. 2017; LeBlanc
9 et al. 2015; Liu et al. 2013; Lv et al. 2017; McLaughlin et al. 2017; Mufford et al. 2019; Shadrin
10 et al. 2018; Smeland et al. 2019; Smeland et al. 2017a; Smeland et al. 2018; Smeland et al.
11 2017b; van der Meer et al. 2018; Wang et al. 2016a; Winsvold et al. 2017; Witoelar et al. 2017;
12 Yokoyama et al. 2017; Yokoyama et al. 2016; Zuber et al. 2018).

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

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

4 The first step in the condFDR procedure is to construct conditional quantile-quantile
5 (Q-Q) plots, which extends the standard Q-Q plots commonly applied in GWAS. Standard Q-
6 Q plots visualize the enrichment of statistical association relative to that expected under the
7 global null hypothesis by plotting the nominal $-\log_{10}$ p-values of the single SNP association
8 statistics versus their empirical distribution. Conditional Q-Q plots help visualize the cross-trait
9 enrichment between two phenotypes and are constructed by creating subsets of SNPs based of
10 the level of association with the secondary phenotype. Under the global null hypothesis, the
11 nominal p-values will form a straight line plotted as a function of their empirical distribution.
12 Under polygenic association, standard Q-Q plots will be deflected leftwards, while cross-trait
13 enrichment can be seen as successive leftward deflections in conditional Q-Q plots as levels of
14 SNP associations with the secondary phenotype increase. Figure 1a presents a conditional Q-Q
15 plot demonstrating SNP enrichment for the psychiatric disorder bipolar disorder (n=51,710)
16 (Stahl et al. 2019) as a function of the association with intelligence (n=269,867) (Savage et al.
17 2018), adapted from Smeland et al. (2019). A complementary way to assess for cross-trait
18 enrichment is to construct fold-enrichment plots, which provide a more direct visualization of
19 the polygenic enrichment (Figure 1b). The fold enrichment is calculated as the ratio between
20 the $-\log_{10}(p)$ cumulative distribution for a given stratum and the cumulative distribution for all
21 SNPs. Figure 1b shows that for SNPs with p-values below 0.001 in intelligence, there was up
22 to 60-fold enrichment of stronger SNP associations with bipolar disorder in comparison to all
23 SNPs. The enrichment seen in conditional Q-Q plots and fold-enrichment plots reflects
24 increased tail probabilities in the distribution of test statistics and an overabundance of low p-
25 values compared to that expected by chance, which can be directly interpreted in terms of a

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

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

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

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

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

21

22 **Comparison to other cross-trait analytical tools**

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

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

6 The most common approaches for evaluating genetic overlap at the genome-wide level
7 include tools such as polygenic risk scores (Purcell et al. 2009), mixed-model approaches
8 (Cross-Disorder Group of the Psychiatric Genomics et al. 2013; Lee et al. 2012) and LD score
9 regression (Bulik-Sullivan et al. 2015a), which return a single estimate of shared genetic risk
10 between phenotypes. Polygenic risk scores are per-individual risk profiles based on the sum of
11 alleles associated with a phenotype weighted by their effect sizes (Purcell et al. 2009). The
12 polygenic risk score approach uses summary statistics as training data and requires individual
13 genotype data in an independent target sample to test how well the polygenic risk score explains
14 phenotypic variation in the target phenotype. Another traditional measure that estimates the
15 degree of pleiotropy is the genetic correlation, which is defined as the correlation between the
16 genetic influences for a pair of traits, thus indicating the proportion of variance that the two
17 traits share due to genetic causes. Mixed-model approaches (Lee et al. 2012), originally
18 implemented in the Genome-wide Complex Trait Analysis software (GCTA), obtained
19 unbiased estimates of the genetic correlation using individual genotype data, relaxing several
20 limitations of traditional studies based on pedigree data. Estimates of genetic correlation can
21 also be quantified from GWAS summary statistics, using cross-trait LD score regression
22 (Bulik-Sullivan et al. 2015a) and its multivariate extension Genomic SEM (Grotzinger et al.
23 2019). LD score regression aims to distinguish confounding from polygenicity by regressing
24 the association statistics of SNPs on their ‘LD scores’, which is a measure of the amount of
25 genetic variation the SNP represents (Bulik-Sullivan et al. 2015b). Application of LD score

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

17 There is a large class of cross-trait methods aiming to discover specific genomic loci
18 unique or shared between phenotypes inspired by the meta-analysis technique (Willer et al.
19 2010) and its extensions dealing with sample overlap (Han et al. 2016; Lin and Sullivan 2009).
20 For example, the COMBINE approach (Ellinghaus et al. 2012) consists of two separate runs of
21 a same-effect and opposite-effect meta-analysis, both using the inverse variance weighted
22 procedure. In the opposite-effect meta-analysis, the minor and major alleles are flipped in the
23 second dataset to capture bi-allelic SNPs with opposite effect directions in the two phenotypes
24 investigated. This method was later refined and extended to multiple heterogeneous traits using
25 restricted and weighted subset search (ASSET) (Bhattacharjee et al. 2012), which exhaustively

1 explore subsets of studies to achieve the best possible trade-off between specificity and sample
2 size. Its successor, compare-and-contrast meta-analysis (CCMA) (Baurecht et al. 2015), further
3 improved the power to discover associations by combining the subset search approach with
4 trans-ethnic meta-analysis (MANTRA) (Morris 2011). Several alternative approaches explore
5 additional information, including individual-level genotypes (MultiPhen) (O'Reilly et al.
6 2012), phenotypic correlations (TATES) (van der Sluis et al. 2013) or estimated genetic
7 correlations (MTAG) (Turley et al. 2018). A common feature of all techniques based on a meta-
8 analysis framework is that the analysis is performed independently for each SNP, thus requiring
9 a follow-up mechanism to control for multiple testing, such as Bonferroni correction, to avoid
10 false positive findings. The condFDR analysis, on the other hand, directly works with the entire
11 original set of p-values from the two GWAS and intrinsically incorporates multiple testing via
12 the FDR framework (Efron 2010).

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

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

5

6 **Conclusion**

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

15

16

17

18

19

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18

Conflict of Interest Disclosures: O.A.A. has received speaker’s honorarium from Lundbeck and is a consultant for Healthytix. C.C.F. is under employment of Multimodal Imaging Service, dba Healthytix, 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 Healthytix 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.

1 **References**

- 2 Andreassen OA et al. (2013a) Improved detection of common variants associated with
3 schizophrenia by leveraging pleiotropy with cardiovascular-disease risk factors *American*
4 *journal of human genetics* 92:197-209 doi:10.1016/j.ajhg.2013.01.001
- 5 Andreassen OA et al. (2014a) Genetic pleiotropy between multiple sclerosis and
6 schizophrenia but not bipolar disorder: differential involvement of immune-related gene loci
7 *Molecular psychiatry* 20:207-214 doi:10.1038/mp.2013.195
- 8 Andreassen OA et al. (2014b) Identifying common genetic variants in blood pressure due to
9 polygenic pleiotropy with associated phenotypes *Hypertension* 63:819-826
10 doi:10.1161/hypertensionaha.113.02077
- 11 Andreassen OA et al. (2014c) Identifying common genetic variants in blood pressure due to
12 polygenic pleiotropy with associated phenotypes *Hypertension* 63:819-826
13 doi:10.1161/HYPERTENSIONAHA.113.02077
- 14 Andreassen OA, Thompson WK, Dale AM (2013b) Boosting the Power of Schizophrenia
15 Genetics by Leveraging New Statistical Tools *Schizophrenia bulletin*
16 doi:10.1093/schbul/sbt168
- 17 Andreassen OA et al. (2013c) Improved detection of common variants associated with
18 schizophrenia and bipolar disorder using pleiotropy-informed conditional false discovery rate
19 *PLoS genetics* 9:e1003455 doi:10.1371/journal.pgen.1003455
- 20 Andreassen OA et al. (2014d) Shared common variants in prostate cancer and blood lipids
21 *International journal of epidemiology* 43:1205-1214 doi:10.1093/ije/dyu090
- 22 Baurecht H et al. (2015) Genome-wide comparative analysis of atopic dermatitis and psoriasis
23 gives insight into opposing genetic mechanisms *Am J Hum Genet* 96:104-120
24 doi:10.1016/j.ajhg.2014.12.004
- 25 Bhattacharjee S et al. (2012) A Subset-Based Approach Improves Power and Interpretation
26 for the Combined Analysis of Genetic Association Studies of Heterogeneous Traits *The*
27 *American Journal of Human Genetics* 90:821-835
28 doi:<https://doi.org/10.1016/j.ajhg.2012.03.015>
- 29 Brainstorm C et al. (2018) Analysis of shared heritability in common disorders of the brain
30 *Science* 360 doi:10.1126/science.aap8757
- 31 Broce I et al. (2018) Immune-related genetic enrichment in frontotemporal dementia: An
32 analysis of genome-wide association studies *PLoS medicine* 15:e1002487
33 doi:10.1371/journal.pmed.1002487
- 34 Broce IJ et al. (2019) Dissecting the genetic relationship between cardiovascular risk factors
35 and Alzheimer's disease *Acta Neuropathol* 137:209-226 doi:10.1007/s00401-018-1928-6
- 36 Bulik-Sullivan B et al. (2015a) An atlas of genetic correlations across human diseases and
37 traits *Nature genetics* 47:1236-1241 doi:10.1038/ng.3406
- 38 Bulik-Sullivan BK et al. (2015b) LD Score regression distinguishes confounding from
39 polygenicity in genome-wide association studies *Nature genetics* 47:291-295
40 doi:10.1038/ng.3211
- 41 Cross-Disorder Group of the Psychiatric Genomics C et al. (2013) Genetic relationship
42 between five psychiatric disorders estimated from genome-wide SNPs *Nature genetics*
43 45:984-994 doi:10.1038/ng.2711
- 44 Davies G et al. (2018) Study of 300,486 individuals identifies 148 independent genetic loci
45 influencing general cognitive function *Nat Commun* 9:2098 doi:10.1038/s41467-018-04362-x
- 46 Desikan RS et al. (2015) Polygenic Overlap Between C-Reactive Protein, Plasma Lipids, and
47 Alzheimer Disease *Circulation* 131:2061-2069
48 doi:10.1161/CIRCULATIONAHA.115.015489
- 49 Devlin B, Roeder K (1999) Genomic control for association studies *Biometrics* 55:997-1004

1 Drange OK et al. (2019) Genetic Overlap Between Alzheimer's Disease and Bipolar Disorder
2 Implicates the MARK2 and VAC14 Genes *Front Neurosci* 13:220
3 doi:10.3389/fnins.2019.00220
4 Efron B (2007) Size, power and false discovery rates *The Annals of Statistics* 35:1351–1377
5 Efron B (2010) Large-scale inference : empirical Bayes methods for estimation, testing, and
6 prediction. Institute of mathematical statistics monographs, vol 1. Cambridge University
7 Press, Cambridge ; New York
8 Efron B, Tibshirani R (2002) Empirical bayes methods and false discovery rates for
9 microarrays *Genetic epidemiology* 23:70-86 doi:10.1002/gepi.1124
10 Ellinghaus D et al. (2012) Combined analysis of genome-wide association studies for Crohn
11 disease and psoriasis identifies seven shared susceptibility loci *American journal of human*
12 *genetics* 90:636-647 doi:10.1016/j.ajhg.2012.02.020
13 Ferrari R et al. (2017) Genetic architecture of sporadic frontotemporal dementia and overlap
14 with Alzheimer's and Parkinson's diseases *J Neurol Neurosurg Psychiatry* 88:152-164
15 doi:10.1136/jnnp-2016-314411
16 Frei O et al. (2019) Bivariate causal mixture model quantifies polygenic overlap between
17 complex traits beyond genetic correlation *Nat Commun* 10:2417 doi:10.1038/s41467-019-
18 10310-0
19 Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V
20 (2014) Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using
21 Summary Statistics *PLoS genetics* 10:e1004383 doi:10.1371/journal.pgen.1004383
22 Gratten J, Visscher PM (2016) Genetic pleiotropy in complex traits and diseases: implications
23 for genomic medicine *Genome medicine* 8:78 doi:10.1186/s13073-016-0332-x
24 Grotzinger AD et al. (2019) Genomic structural equation modelling provides insights into the
25 multivariate genetic architecture of complex traits *Nature Human Behaviour*
26 doi:10.1038/s41562-019-0566-x
27 Hackinger S, Zeggini E (2017) Statistical methods to detect pleiotropy in human complex
28 traits *Open Biol* 7 doi:10.1098/rsob.170125
29 Han B, Duong D, Sul JH, de Bakker PI, Eskin E, Raychaudhuri S (2016) A general
30 framework for meta-analyzing dependent studies with overlapping subjects in association
31 mapping *Human molecular genetics* 25:1857-1866 doi:10.1093/hmg/ddw049
32 Hernan MA, Robins JM (2006) Instruments for causal inference: an epidemiologist's dream?
33 *Epidemiology (Cambridge, Mass)* 17:360-372 doi:10.1097/01.ede.0000222409.00878.37
34 Hill WD, Davies G, Group CCW, Liewald DC, McIntosh AM, Deary IJ (2016) Age-
35 Dependent Pleiotropy Between General Cognitive Function and Major Psychiatric Disorders
36 *Biological psychiatry* 80:266-273 doi:10.1016/j.biopsych.2015.08.033
37 Holland D et al. (2019) Beyond SNP Heritability: Polygenicity and Discoverability of
38 Phenotypes Estimated with a Univariate Gaussian Mixture Model *bioRxiv*:133132
39 doi:10.1101/133132
40 Hu Y et al. (2018) Identification of Novel Potentially Pleiotropic Variants Associated With
41 Osteoporosis and Obesity Using the cFDR Method *J Clin Endocrinol Metab* 103:125-138
42 doi:10.1210/jc.2017-01531
43 Karch CM et al. (2018) Selective Genetic Overlap Between Amyotrophic Lateral Sclerosis
44 and Diseases of the Frontotemporal Dementia Spectrum *JAMA neurology* 75:860-875
45 doi:10.1001/jamaneurol.2018.0372
46 Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G (2008) Mendelian
47 randomization: using genes as instruments for making causal inferences in epidemiology
48 *Statistics in medicine* 27:1133-1163 doi:10.1002/sim.3034
49 Le Hellard S et al. (2017) Identification of Gene Loci That Overlap Between Schizophrenia
50 and Educational Attainment *Schizophrenia bulletin* 43:654-664 doi:10.1093/schbul/sbw085

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

1 Savage JE et al. (2018) Genome-wide association meta-analysis in 269,867 individuals
2 identifies new genetic and functional links to intelligence *Nature genetics* 50:912-919
3 doi:10.1038/s41588-018-0152-6

4 Schork AJ et al. (2013) All SNPs are not created equal: genome-wide association studies
5 reveal a consistent pattern of enrichment among functionally annotated SNPs *PLoS genetics*
6 9:e1003449 doi:10.1371/journal.pgen.1003449

7 Schork AJ, Wang Y, Thompson WK, Dale AM, Andreassen OA (2016) New statistical
8 approaches exploit the polygenic architecture of schizophrenia--implications for the
9 underlying neurobiology *Curr Opin Neurobiol* 36:89-98 doi:10.1016/j.conb.2015.10.008

10 Schwartzman A, Lin X (2011) The effect of correlation in false discovery rate estimation
11 *Biometrika* 98:199-214 doi:10.1093/biomet/asq075

12 Shadrin AA et al. (2018) Novel Loci Associated With Attention-Deficit/Hyperactivity
13 Disorder Are Revealed by Leveraging Polygenic Overlap With Educational Attainment *J Am*
14 *Acad Child Adolesc Psychiatry* 57:86-95 doi:10.1016/j.jaac.2017.11.013

15 Shi H, Mancuso N, Spendlove S, Pasaniuc B (2017) Local Genetic Correlation Gives Insights
16 into the Shared Genetic Architecture of Complex Traits *American journal of human genetics*
17 101:737-751 doi:10.1016/j.ajhg.2017.09.022

18 Sivakumaran S et al. (2011) Abundant pleiotropy in human complex diseases and traits
19 *American journal of human genetics* 89:607-618 doi:10.1016/j.ajhg.2011.10.004

20 Smeland OB et al. (2019) Genome-wide analysis reveals extensive genetic overlap between
21 schizophrenia, bipolar disorder, and intelligence *Molecular psychiatry* doi:10.1038/s41380-
22 018-0332-x

23 Smeland OB et al. (2017a) Identification of Genetic Loci Jointly Influencing Schizophrenia
24 Risk and the Cognitive Traits of Verbal-Numerical Reasoning, Reaction Time, and General
25 Cognitive Function *JAMA psychiatry* 74:1065-1075 doi:10.1001/jamapsychiatry.2017.1986

26 Smeland OB et al. (2018) Genetic Overlap Between Schizophrenia and Volumes of
27 Hippocampus, Putamen, and Intracranial Volume Indicates Shared Molecular Genetic
28 Mechanisms *Schizophrenia bulletin* 44:854-864 doi:10.1093/schbul/sbx148

29 Smeland OB et al. (2017b) Identification of genetic loci shared between schizophrenia and the
30 Big Five personality traits *Sci Rep* 7:2222 doi:10.1038/s41598-017-02346-3

31 Smith GD, Ebrahim S (2003) 'Mendelian randomization': can genetic epidemiology
32 contribute to understanding environmental determinants of disease? *International journal of*
33 *epidemiology* 32:1-22

34 Smoller JW, Andreassen OA, Edenberg HJ, Faraone SV, Glatt SJ, Kendler KS (2018)
35 Psychiatric genetics and the structure of psychopathology *Molecular psychiatry*
36 doi:10.1038/s41380-017-0010-4

37 Sniekers S et al. (2017) Genome-wide association meta-analysis of 78,308 individuals
38 identifies new loci and genes influencing human intelligence *Nature genetics* 49:1107-1112
39 doi:10.1038/ng.3869

40 Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW (2013) Pleiotropy in complex
41 traits: challenges and strategies *Nature reviews Genetics* 14:483-495 doi:10.1038/nrg3461

42 Stahl EA et al. (2019) Genome-wide association study identifies 30 loci associated with
43 bipolar disorder *Nature genetics* 51:793-803 doi:10.1038/s41588-019-0397-8

44 Sun L, Craiu RV, Paterson AD, Bull SB (2006) Stratified false discovery control for large-
45 scale hypothesis testing with application to genome-wide association studies *Genetic*
46 *epidemiology* 30:519-530 doi:10.1002/gepi.20164

47 Turley P et al. (2018) Multi-trait analysis of genome-wide association summary statistics
48 using MTAG *Nat Genet* 50:229-237 doi:10.1038/s41588-017-0009-4

1 van der Meer D et al. (2018) Brain scans from 21,297 individuals reveal the genetic
2 architecture of hippocampal subfield volumes *Molecular psychiatry* doi:10.1038/s41380-018-
3 0262-7
4 van der Sluis S, Posthuma D, Dolan CV (2013) TATES: Efficient Multivariate Genotype-
5 Phenotype Analysis for Genome-Wide Association Studies *PLOS Genetics* 9:e1003235
6 doi:10.1371/journal.pgen.1003235
7 Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J (2017) 10
8 Years of GWAS Discovery: Biology, Function, and Translation *American journal of human*
9 *genetics* 101:5-22 doi:10.1016/j.ajhg.2017.06.005
10 Wang Y et al. (2016a) Genetic overlap between multiple sclerosis and several cardiovascular
11 disease risk factors *Mult Scler* 22:1783-1793 doi:10.1177/1352458516635873
12 Wang Y et al. (2016b) Leveraging Genomic Annotations and Pleiotropic Enrichment for
13 Improved Replication Rates in Schizophrenia GWAS *PLoS genetics* 12:e1005803
14 doi:10.1371/journal.pgen.1005803
15 Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of
16 genomewide association scans *Bioinformatics* 26:2190-2191
17 doi:10.1093/bioinformatics/btq340
18 Winsvold BS et al. (2017) Shared genetic risk between migraine and coronary artery disease:
19 A genome-wide analysis of common variants *PloS one* 12:e0185663
20 doi:10.1371/journal.pone.0185663
21 Witoelar A et al. (2017) Genome-wide Pleiotropy Between Parkinson Disease and
22 Autoimmune Diseases *JAMA neurology* 74:780-792 doi:10.1001/jamaneurol.2017.0469
23 Yokoyama JS et al. (2017) Shared genetic risk between corticobasal degeneration,
24 progressive supranuclear palsy, and frontotemporal dementia *Acta Neuropathol* 133:825-837
25 doi:10.1007/s00401-017-1693-y
26 Yokoyama JS et al. (2016) Association Between Genetic Traits for Immune-Mediated
27 Diseases and Alzheimer Disease *JAMA neurology* 73:691-697
28 doi:10.1001/jamaneurol.2016.0150
29 Yoo YJ, Pinnaduwa D, Waggott D, Bull SB, Sun L (2009) Genome-wide association
30 analyses of North American Rheumatoid Arthritis Consortium and Framingham Heart Study
31 data utilizing genome-wide linkage results *BMC proceedings* 3 Suppl 7:S103
32 Zhu Z et al. (2018) Causal associations between risk factors and common diseases inferred
33 from GWAS summary data *Nature Communications* 9:224 doi:10.1038/s41467-017-02317-2
34 Zuber V et al. (2018) Identification of shared genetic variants between schizophrenia and lung
35 cancer *Sci Rep* 8:674 doi:10.1038/s41598-017-16481-4
36

37

1 **Box 1: Conditional and conjunctive False Discovery Rate**

2 The ‘enrichment’ seen in the conditional Q-Q plots can be directly interpreted in terms of a
3 Bayesian interpretation of the true discovery rate (TDR = 1 – false discovery rate (FDR)) (Efron
4 2010). More specifically, for a given p-value, under a simple two-group (null and non-null)
5 model, Bayes rule gives the posterior probability of being null as

6
$$\text{FDR}(p) = \pi_0 F_0(p) / F(p), \quad [1]$$

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

12
$$\text{FDR}(p) = \pi_0 p / F(p). \quad [2]$$

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

16
$$\text{Estimated FDR}(p) = \pi_0 p / q, \quad [3]$$

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

19
$$q^* = p/q. \quad [4]$$

20 If π_0 is close to one, the increase in bias going from Eq. [3] to Eq. [4] is minimal. The quantity
21 $1 - p/q$, is therefore biased downward, and hence a conservative estimate of the TDR. Referring
22 to the Q-Q plots, we see that q^* is equivalent to the nominal p-value divided by the empirical
23 quantile, as defined earlier. We can thus read the FDR estimate directly off the Q-Q plot as

24
$$-\log_{10}(q^*) = \log_{10}(q) - \log_{10}(p), \quad [5]$$

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

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

$$4 \quad \text{FDR}(p_1|p_2) = \pi_0(p_2)p_1 / F(p_1|p_2), \quad [6]$$

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

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

$$18 \quad \text{FDR}^{\text{Phenotype1\&Phenotype2}}(p_1, p_2) = \pi_0 F_0(p_1, p_2) / F(p_1, p_2) + \pi_1 F_1(p_1, p_2) / F(p_1, p_2) + \pi_2 F_2(p_1, p_2) \\ 19 \quad / F(p_1, p_2), \quad [7]$$

20 where π_0 is the *a priori* proportion of SNPs null for both phenotypes simultaneously and $F_0(p_1,$
 21 $p_2)$ is the joint null cdf, π_1 is the *a priori* proportion of SNPs non-null for the first phenotype
 22 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*
 23 proportion of SNPs non-null for the second phenotype and null for the first, with joint cdf $F_2(p_1,$
 24 $p_2)$. $F(p_1, p_2)$ is the joint overall mixture cdf for all phenotype 1 and 2 SNPs.

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

$$3 \quad \text{Estimated FDR}_{\text{Phenotype1\&Phenotype2}} =$$

$$4 \quad \max \{ \text{Estimated FDR}_{\text{Phenotype1|Phenotype2}}, \text{Estimated FDR}_{\text{Phenotype2|Phenotype1}} \}, \quad [8]$$

5 where $\text{Estimated FDR}_{\text{Phenotype1|Phenotype2}}$ and $\text{Estimated FDR}_{\text{Phenotype2|Phenotype1}}$ are conservative
 6 (upwardly biased) estimates of Eq. [6]. Thus, Eq (8) is a conservative estimate of $\max \{ p_1/F(p_1|$
 7 $p_2), p_2/F(p_2|p_1) \} = \max \{ p_1 F_2(p_2)/F(p_1, p_2), p_2 F_1(p_1)/F(p_1, p_2) \}$, with $F_1(p_1)$ and $F_2(p_2)$ the
 8 marginal non-null cdfs of SNPs for phenotype 1 and 2, respectively. For enriched samples, p-
 9 values will tend to be smaller than predicted from the uniform distribution, so that

10 $F_1(p_1) \geq p_1$ and $F_2(p_2) \geq p_2$. Then

$$11 \quad \max \{ p_1 F_2(p_2) / F(p_1, p_2), p_2 F_1(p_1) / F(p_1, p_2) \}$$

$$12 \quad \geq [\pi_0 + \pi_1 + \pi_2] \max \{ p_1 F_2(p_2) / F(p_1, p_2), p_2 F_1(p_1) / F(p_1, p_2) \}$$

$$13 \quad \geq [\pi_0 p_1 p_2 + \pi_1 p_2 F_1(p_1) + \pi_2 p_1 F_2(p_2)] / F(p_1, p_2).$$

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

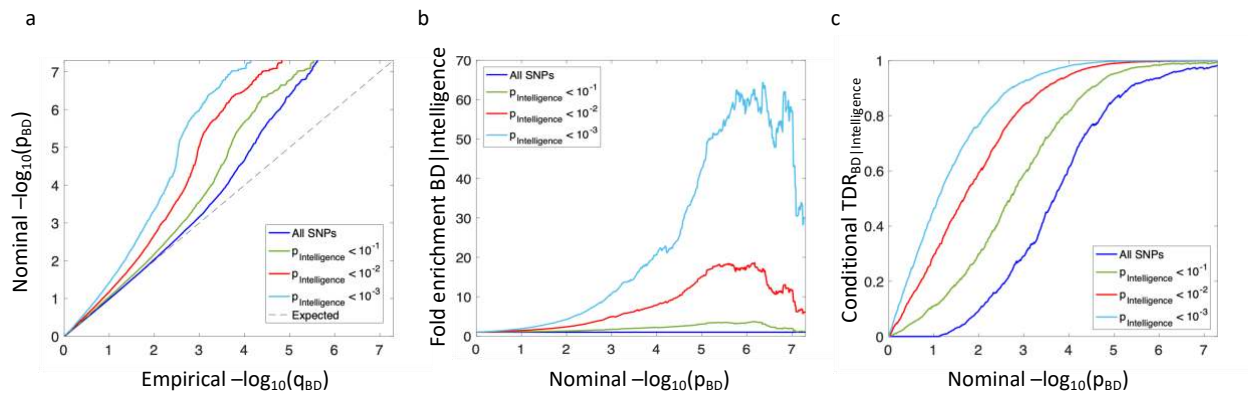
17
 18
 19

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

<i>Primary phenotype</i>	<i>Secondary phenotype</i>	<i>Novel loci for primary phenotype</i>	<i>Citation</i>
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)

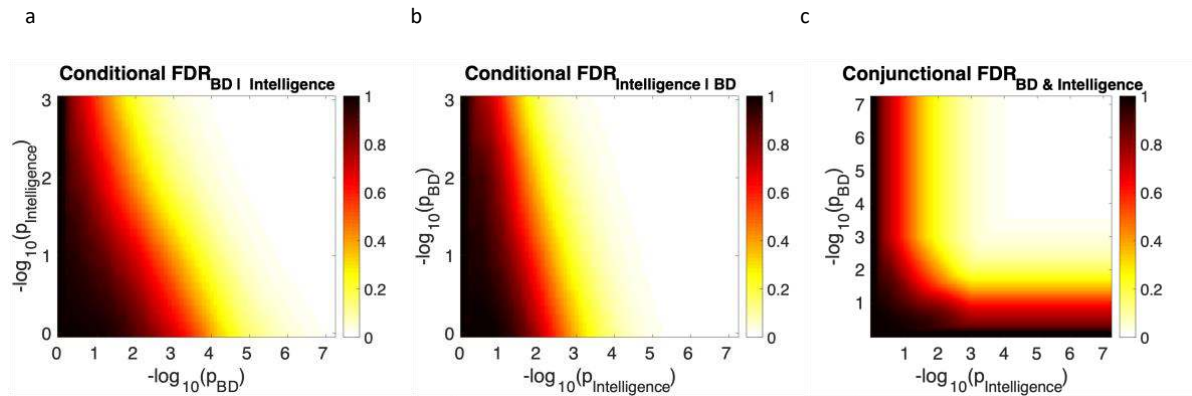
1 Figures

2



3 **Figure 1.** Cross-trait enrichment between bipolar disorder (BD; $n=51,710$) (Stahl et al. 2019)
4 and intelligence ($n=269,867$) (Savage et al. 2018), adapted from Smeland et al. (2019). (a)
5 Conditional Q-Q plot displaying the nominal $-\log_{10}$ p-values of the single SNP association
6 statistics versus their empirical distribution in BD below the standard GWAS threshold of
7 $p < 5 \times 10^{-8}$ as a function of significance of association with intelligence at the level of $p \leq 0.1$,
8 $p \leq 0.01$, $p \leq 0.001$, respectively. The blue line indicates all SNPs. The dashed line indicates
9 the null hypothesis. (b) Fold-enrichment plot of enrichment versus nominal $-\log_{10}$ p-values in
10 BD as a function of association with intelligence. (c) Conditional true discovery rate (TDR)
11 plot illustrating the increase in TDR associated with increased enrichment in BD conditioned
12 on intelligence. The test statistics were corrected for genomic inflation, SNPs were randomly
13 pruned across 500 iterations using a linkage disequilibrium r^2 threshold of 0.1, and the extended
14 major histocompatibility complex region and chromosomal region 8p.23.1 were excluded
15 (Smeland et al. 2019).

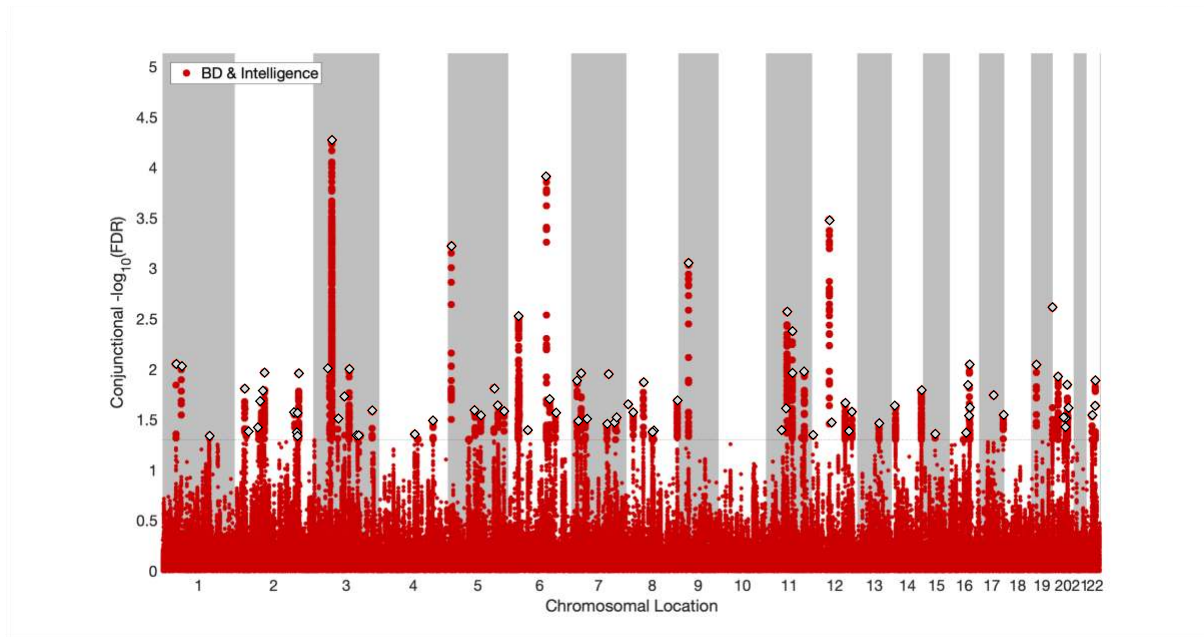
1



2

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

10



1

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

9