Examination of a Novel Expression-Based Gene-SNP Annotation Strategy to Identify Tissue-Specific Contributions to Heritability in Multiple Traits

Travis Mize (trmi1868@colorado.edu)
University of Colorado Boulder

Luke Evans
University of Colorado Boulder  https://orcid.org/0000-0002-7458-1720

Article

Keywords: genetics, gene expression, tissue-specific, heritability, LDSC, annotation

Posted Date: July 6th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1769771/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Version of Record: A version of this preprint was published at European Journal of Human Genetics on November 29th, 2022. See the published version at https://doi.org/10.1038/s41431-022-01244-1.
Abstract

Complex traits show clear patterns of tissue-specific expression influenced by single-nucleotide polymorphisms (SNPs), yet current strategies aggregate SNP effects to genes by employing simple physical proximity-based windows. Here, we examined whether incorporating only those SNPs with effects on tissue-specific cis-expression would improve our ability to detect trait-relevant tissues across 31 complex traits using stratified linkage disequilibrium score regression (S-LDSC). We found that a physical proximity annotation produced more significant tissue enrichments and larger S-LDSC regression coefficients, as compared to an expression-based annotation. Furthermore, we showed that our expression-based annotation did not outperform an annotation strategy in which an equal number of randomly chosen SNPs were annotated to genes within the same genomic window, suggesting extensive redundancy among SNP effect estimates due to linkage disequilibrium. That said, current sample sizes limit estimation of cis-genetic SNP effects; therefore, we recommend reexamination of the expression-based annotation when larger tissue-specific expression datasets become available. Finally, we report new and updated tissue enrichment estimates across 31 complex traits, such as significant heritability enrichment of the frontal cortex for cognitive performance, educational attainment, and intelligence, providing further evidence of this structure's importance in higher cognitive function.

Introduction

Regulation of gene expression is one mechanism of heritable variation of complex traits\textsuperscript{1–5}, yet integration of expression-related processes in identifying genetic associations and influences on traits remains incomplete. Single-nucleotide polymorphisms (SNPs) located in or near a gene (i.e., in cis) may influence that gene's function through alterations of transcription regulatory elements\textsuperscript{6}, mRNA splicing\textsuperscript{7}, translation\textsuperscript{8,9}, and many other factors\textsuperscript{10,11}. However, the functional effects of most SNPs on complex traits, particularly those without clear protein-coding changes, remain unknown\textsuperscript{12}. Despite this, there is clear evidence that genetic influences on tissue-and cell-type-specific expression affect complex traits\textsuperscript{13–16}. This has led to discussions of pleiotropic effects across tissues\textsuperscript{17} and how additional transcriptomic complexity may influence our ability to estimate expression-mediated heritability\textsuperscript{14}.

Recently, Finucane et al. (2018)\textsuperscript{2} used Genotype Tissue Expression (GTEx)\textsuperscript{18} RNA-seq data to identify trait-relevant tissues and cell types across 48 traits and diseases through the application of stratified linkage disequilibrium score regression (S-LDSC)\textsuperscript{19}. They identified sets of genes specifically expressed in tissues or brain regions, within which SNPs contribute significantly to heritable variation in complex traits. As is standard practice, they utilized a 100kb physical window within which all SNPs were annotated to genes. Previous studies suggest that while cis-acting factors influence tissue-specific gene expression\textsuperscript{18,20}, causal genes are frequently further than 100kb away from associated SNPs\textsuperscript{21}, and the addition of functional information (e.g., expression quantitative trait loci) can improve causal inference in genomic annotation\textsuperscript{12}. Together, these studies imply that incorporating variants estimated to influence tissue-specific expression, in addition to specifying a larger genomic window, may capture additional
transcriptomic information resulting in heightened model specificity and improved estimates of tissue specific expression contribution to heritability of complex traits.

There are several approaches to estimate SNP effects on tissue-specific gene expression\cite{4,14,15,18} and it is therefore possible to annotate variants to genes based on their predicted functional impact on expression, foregoing the assumption of the physically nearest genes being causal. Here, we sought to examine whether the annotation of SNPs to genes based on their estimated effects on expression\cite{1} would identify novel expression-trait estimates of tissue-specific enrichment while leveraging the largest publicly available genome-wide association study (GWAS) summary statistics to date. We hypothesized the expression-based annotation would improve the ability to identify trait-relevant tissues by increasing gene-SNP annotation specificity while simultaneously incorporating a larger genomic window. We also updated estimates of partitioned cis-expression SNP heritability ($h^2_{SNP}$) of specifically expressed genes using the most recent available GWAS across 31 complex traits.

**Methods**

**Overview of Cell-Type-Specific Stratified LDSC and Specifically Expressed Genes**

Cell-type-specific S-LDSC\cite{2,19} models the combined heritable contribution of a SNP and those with which it is in LD, such that $E [\chi^2_i] = 1 + Na + N \sum_k \tau(l(i,k))$, where $E [\chi^2_i]$ is the expectation of the association test statistic for SNP $i$, $N$ is the sample size of the GWAS, $a$ is a measure of confounding bias (e.g., population stratification), and $l(i,k)$ is SNP $i$'s LD score for a functional category $k$. The regression coefficient ($\tau$) for the $k^{th}$ functional annotation estimates the contribution of that category's specific expression to $h^2_{SNP}$ enrichment conditional on all other annotations. More specifically, when $\tau = 0$ there is no enrichment, when $\tau < 0$ there is a decrease in the per-SNP heritability while accounting for other annotations, and when $\tau > 0$ there is an increase in the per-SNP heritability while accounting for other annotations. The 53 baseline functional annotations incorporate genetic information not specific to cell type from gene structure (e.g., promoter, super enhancer, intronic, exonic) and methylation patterns (e.g., histone marks, chromatin structure) to increase the accuracy of estimated enrichment\cite{19}. Refer to Finucane et al. (2015)\cite{19} for a complete description of modeling and baseline functional annotation procedures.

Sets of genes specifically expressed in individual tissues and brain regions using GTEx v6p expression data were identified by Finucane et al.\cite{2}, which we downloaded from https://alkesgroup.broadinstitute.org/LDSCORE/ and used in all subsequent analyses. The baseline annotation strategy is based on physical proximity of SNPs to genes of interest, annotating all SNPs ± 100kb surrounding each gene of interest (Fig. 1).
We applied, following Finucane et al.\textsuperscript{2}, two series of gene sets: 1) genes with unique expression within a category of similar tissues, as compared to the expression of other tissues (e.g., cortex vs. all non-brain tissues), and 2) genes with unique expression within a given brain region, as compared to the expression of all other available brain regions. For simplicity, we refer to gene sets from 1) as multi-tissue and 2) as within-brain. Refer to Finucane et al.\textsuperscript{2} for a complete description of the methods employed to identify these sets of genes.

We accessed available GWAS summary statistics for 31 phenotypes\textsuperscript{22–41} (Supp. Table 1) and performed S-LDSC using both the physical and expression-based annotation approaches (described below). Due to the limited samples of publicly released GWAS summary statistics, only data from individuals of European descent were included in this study.

**Expression-Based Annotation**

We identified SNPs with evidence of expression effects in GTEx v7 data in each of 48 tissues based on pre-computed gene expression weights estimated in Functional Summary-based Imputation (FUSION)\textsuperscript{1} and available at http://gusevlab.org/projects/fusion/. FUSION applies four statistical models (best linear unbiased prediction (BLUP), top1, lasso regression, and elastic net regression), each with distinct assumptions of polygenicity, to estimate tissue-specific gene expression weights ± 500kb from the transcription start site of all genes for which there was an available expression weight. For each gene, FUSION then selects the highest performing model based on the model’s $R^2$ and its corresponding p-value. Briefly, BLUP includes all non-zero effect SNPs, top1 incorporates the single largest effect SNP, lasso regression generates a large effect SNP sparse model, and elastic net regression tests across a spectrum of SNP inclusion ranging from BLUP to top1. Thus, for each gene, a set of SNPs was identified that had non-zero cis-expression effects within 500kb upstream and downstream of the transcription start site, and within specific tissues or brain regions.

We annotated these putatively expression-influencing SNPs to the specifically expressed sets of genes identified by Finucane et al.\textsuperscript{2} (Fig. 1) and implemented gene-set enrichment analyses in S-LDSC to test for trait-specific tissue relevance (see below). As LDSC performs poorly when annotations have too few SNPs\textsuperscript{16}, all gene expression weights for a given tissue, regardless of their estimated cis-genetic expression $h^2_{SNP}$, were included to ensure a sufficient number of SNPs were incorporated to control for type 1 error using S-LDSC, as well as to assure the highest degree of overlap with the Finucane et al.\textsuperscript{2} sets of genes (Supp. Table 2). We then directly compared our expression-based annotation approach to the standard physical proximity gene-SNP annotation using estimates published by Finucane et al.\textsuperscript{2} (see below).

For each tissue, specifically expressed sets of genes were incorporated as a 54th functional annotation category alongside the 53 baseline functional categories\textsuperscript{19} plus an additional annotation category that consisted of either all genes examined in the GTEx v6p gene expression dataset from which the specifically expressed gene-sets were derived, as specified in Finucane et al.\textsuperscript{2} (when using physical...
proximity annotation), or all genes for which there were available expression weights for the expression-based annotation. To correct for multiple tests, a false discovery rate (FDR) < 5% was applied across multi-tissue S-LDSC analyses, as in Finucane et al.\textsuperscript{2}.

For traits with multiple implicated brain regions based on the cross-tissue analyses, a within-brain analysis was conducted to control for overlap between brain expression gene sets implemented in the multi-tissue analyses, as described above and similar to Finucane et al.\textsuperscript{2}. If no brain regions were identified as significant for a given trait in the multi-tissue analyses, that trait was excluded from consideration in the within-brain analyses. All background LD model functional annotations remained the same between the multi-tissue and within-brain analyses. Once again, to correct for multiple tests, an FDR < 5% was applied across within-brain S-LDSC analyses.

**Testing specificity of LDSC coefficients between expression- and proximity-based SNP annotations**

To assess whether the expression-based annotation procedure led to increased specificity of heritable contribution relative to physical proximity-based annotation we performed permutation tests. We used data from schizophrenia\textsuperscript{40} (Fig. 2) as a baseline phenotype to compare estimated $h_{SNP}^2$ enrichment when applying expression-based annotation against a set of randomly chosen SNPs annotated within an equivalently sized genomic window (1mb). For each gene, and each permutation, the same number of random SNPs were annotated equivalent to the number of SNPs with non-zero expression weights from the best performing expression prediction model. For each permutation, linkage disequilibrium scores were re-calculated and $h_{SNP}^2$ LDSC regression coefficients estimated. We performed 1000 permutations to generate a normal distribution of S-LDSC coefficients for a single tissue, representing a null distribution to test whether the specific choice of SNPs used, namely, those with evidence of expression effects, provides more information than a random set of SNPs of equal number. We compared our expression-based LDSC regression coefficients to this null distribution using a one-tailed significance test ($\rho = 0.05$). We performed this permutation procedure for three separate scenarios, applied to the 2014 schizophrenia GWAS summary statistics: 1) nonsignificant $h_{SNP}^2$ enrichment when using the expression-based annotation for the within-brain frontal cortex gene set, 2) significant $h_{SNP}^2$ enrichment when using the expression-based annotation for the within-brain cerebellum gene set, and 3) significant $h_{SNP}^2$ enrichment when using both the expression-based and physical proximity annotation for the multi-tissue cerebellum gene set. We chose these scenarios as they represent a range of results and possible outcomes.

**Results**

Ensuring consistency between our proximity-based annotation results and prior work, we first replicated the Finucane et al. multi-tissue schizophrenia S-LDSC analyses (correlation of the total number of SNPs included for each tissue, $r = 1$, Supp. Table 3, and the correlation of the proportion of SNP heritability
explained across all tissues, \( r > 0.99 \), Supp. Table 3). This small difference is most likely due to minor variability in reported GWAS summary statistics over time.

Next, we sought to examine differences in S-LDSC coefficients across prior and newly reported GWAS for the same phenotypes. While some traits were highly consistent, such as height (LDSC regression coefficient estimates based on the new vs. prior GWAS summary statistics \( r = 0.9968 \), Supp. Figure 1), comparisons of new vs. prior GWAS for other traits, such as Alzheimer’s disease (\( r = 0.3265 \), Supp. Figure 2), strongly differed (Supp. Table 4).

For tissue enrichment analyses, at least 100,000 SNPs were mapped to 46 of the 48 tissues when using an expression-based annotation (Supp. Table 2), with pancreas and whole blood falling below this threshold (83,239 and 87,720 SNPs, respectively), suggesting that for these 46 tissues, S-LDSC regression coefficients are likely well controlled for type 1 error. We found the expression-based annotation resulted in fewer identified tissues or brain regions that contribute significantly to \( h^2_{SNP} \) in complex traits when compared to the physical proximity-based annotation. Across the multi-tissue analyses, of the 31 phenotypes examined, 18 had at least one significant tissue when employing a physical proximity annotation, whereas only seven phenotypes had at least one significant tissue when using an expression-based annotation (FDR < 5%, Supp. Figures 3–20, Supp. Table 5). All tissue and trait combinations with significant expression-based annotation enrichments were also identified using the physical proximity annotation with the single exception of ovary tissue in Tourette syndrome (Supp. Figure 19). Of the phenotypes examined, schizophrenia (both sets of published GWAS summary statistics), educational attainment, and intelligence identified all 13 brain regions as significant, representing the maximum individual tissues identified for any trait.

Within-brain analyses were conducted for the 16 phenotypes with significant \( h^2_{SNP} \) contribution of at least one brain region identified in the multi-tissue analyses. We identified significant contributions of specific brain regions in nine and four phenotypes when using a physical proximity and expression-based annotation, respectively (FDR < 5%, Supp. Figures 21–30, Supp. Table 6). All significant expression-based annotations overlapped with a significant physical proximity annotation, with three exceptions: cortex in major depressive disorder (Supp. Figure 27) and cerebellum in the two schizophrenia datasets\(^{40,42}\) (Supp. Figures 28 and 29).

Permutation tests, to assess whether the expression-based annotation procedure led to increased specificity of heritable contribution relative to physical proximity-based annotation, suggested that SNP-gene annotations based on SNP expression effects do not differ from randomly chosen SNPs within the same regions. In all three instances examined, the strength of the regression coefficient using expression-based annotation was no different than when annotating SNPs to genes at random (all \( p > 0.32 \), Supp. Figures 31–33, Supp. Table 7).

**Discussion**
We tested whether expression-based annotation of SNPs to genes with tissue-specific gene expression influences the specificity of tissue enrichment estimates utilizing S-LDSC and updated estimates of partitioned $h_{SNP}^2$ across 31 different complex traits. We found little evidence that annotating SNPs to genes based on evidence of expression positively impacts $h_{SNP}^2$ estimates of S-LDSC. In both the multi-tissue and within-brain analyses, the physical proximity annotation resulted in more instances of significant tissue enrichment and larger S-LDSC regression coefficients, as compared to the expression-based annotation. There were only four occurrences in which the expression-based annotation identified significant tissue enrichments not found when employing the physical proximity annotation: 1) ovary tissue for Tourette syndrome in multi-tissue analyses, 2) brain cortex for major depressive disorder in within-brain analyses, 3) cerebellum for schizophrenia (Daner) in within-brain analyses, and 4) cerebellum for schizophrenia (Clozuk) in within-brain analyses.

Permutations suggest that the expression-based annotation did not outperform an annotation strategy in which an equal number of randomly chosen SNPs were annotated to genes within the same genomic window (Supp. Table 7). This may have resulted for several reasons. The choice of SNPs within the windows surrounding the specifically expressed genes makes little difference, perhaps because of extensive LD among SNPs surrounding these genes. Alternatively, while genetic effects on gene regulation clearly impact complex traits, non-regulatory effects do as well, and their effects are ignored when only including expression-regulating SNPs. Finally, we included expression-based annotations for all genes with available FUSION weights in order to include as many SNPs for each annotation as possible to properly control for known type 1 error inflation in S-LDSC when annotations have too few SNPs, as well as to maximize the number of genes overlapping the specifically expressed genes of each tissue or brain region by Finucane et al. Unfortunately, roughly half of the genes implicated by Finucane et al. did not have significant cis-genetic SNP heritability of expression (as estimated in FUSION) and may have provided additional noise in our $h_{SNP}^2$ enrichment analyses. As tissue-specific expression sample sizes increase, we expect the number of genes with significant cis-$h_{SNP}^2$ to also increase, as well as the accuracy of their expression prediction models. As such, we predict that as these resources continue to grow, expression-based annotation may result in higher specificity of identified relevant tissues and better estimates of partitioned $h_{SNP}^2$. Therefore, we suggest reexamination of the expression-based annotation in estimating tissue-specific enrichment when larger expression reference panel sample sizes become available.

In addition to our comparison of annotation strategies, here we also report new and updated S-LDSC tissue enrichments across a wide variety of traits (Supp. Tables 5 and 6). For example, we report significant heritability enrichment of the frontal cortex for cognitive performance, educational attainment, and intelligence while controlling for expression in other brain regions, corroborating evidence of this structure's importance in higher cognitive function. Additionally, we found stronger tissue-specific heritability enrichment estimates across several traits compared to those originally reported by Finucane et al. (2018), such as those for schizophrenia, reiterating the importance of GWAS sample size in S-LDSC tissue enrichment analyses.
Identification of heritable contributions to complex traits remains pertinent as we transition from macro-level estimates of heritability to a finer-scale of tissue- and cell-type relevancy. Here, we have attempted to build upon current gene-SNP annotation strategies through the incorporation of estimated effects on gene expression, while simultaneously providing updated and new tissue-specific enrichment estimates across 31 complex traits. While our expression-based annotation did not improve our ability to identify trait-relevant tissues, we suggest further examination of this approach as tissue- and cell type-specific transcriptome reference panels continue to grow.

**Declarations**

This work was supported by the National Institutes of Health (grant numbers T32 DA017637 to John K. Hewitt, R01 AG046938-06 to Chandra A. Reynolds, R01 DA044283-01 to Scott I. Vrieze, and R01 MH100141-06 to Matthew C. Keller) and the University of Colorado Boulder Institute for Behavioral Genetics.

**Supplementary Material**

Online supplementary material includes seven Tables and 33 Figures.

**Funding**

This work was supported by the National Institutes of Health (grant numbers T32 DA017637 to John K. Hewitt, R01 AG046938-06 to Chandra A. Reynolds, R01 DA044283-01 to Scott I. Vrieze, and R01 MH100141-06 to Matthew C. Keller) and the University of Colorado Boulder Institute for Behavioral Genetics.

**Declaration of Interests**

The authors declare no conflict of interest.

**Acknowledgements**

We thank the participants and researchers of the original studies whose summary statistics and expression weights were utilized in this work. We also thank the authors of FUSION for generating the expression weights, as well as the Finucane lab for the specifically expressed gene sets. This work utilized the Summit supercomputer, which is supported by the National Science Foundation (awards ACI-1532235 and ACI-1532236), the University of Colorado Boulder, and Colorado State University. The Summit supercomputer is a joint effort of the University of Colorado Boulder and Colorado State University.

**Data Availability**

GWAS summary statistics (Supp. Table 2), FUSION expression weights (http://gusevlab.org/projects/fusion/), and S-LDSC specifically expressed genes
(https://alkesgroup.broadinstitute.org/LDSCORE/) are available online. All data generated herein is reported in the supplementary tables.

**Code Availability**

All code is available upon request.

**Author Contribution Statement**

TJM and LME contributed to study conception, design, manuscript preparation, and approval of the final version of the manuscript. TJM conducted analyses and curated datasets.

**Ethical Approval**

This research was approved by the CU IRB under protocol 18-0091.

**Competing Interests**

The authors declare no competing interests.

**References**

1. Gusev A, Ko A, Shi H, et al. Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet. 2016;48(3):245–252.
2. Finucane HK, Reshef YA, Anttila V, et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. Nat Genet. 2018;50(4):621–629. doi:10.1038/s41588-018-0081-4
3. Ongen H, Brown AA, Delaneau O, Panousis NI, Nica AC, Dermitzakis ET. Estimating the causal tissues for complex traits and diseases. Nat Genet. 2017;49(12):1676–1683. doi:10.1038/ng.3981
4. Gusev A, Mancuso N, Won H, et al. Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. Nat Genet. 2018;50(4):538–548. doi:10.1038/s41588-018-0092-1
5. Heinzen EL, Ge D, Cronin KD, et al. Tissue-Specific Genetic Control of Splicing: Implications for the Study of Complex Traits. PLOS Biol. 2008;6(12):e1000001. doi:10.1371/journal.pbio.1000001
6. Guo Y, Jamison DC. The distribution of SNPs in human gene regulatory regions. BMC Genomics. 2005;6(1):140. doi:10.1186/1471-2164-6-140
7. Shen LX, Basilion JP, Stanton VP. Single-nucleotide polymorphisms can cause different structural folds of mRNA. Proc Natl Acad Sci U S A. 1999;96(14):7871–7876.
8. Beaudoin JD, Perreault JP. 5′-UTR G-quadruplex structures acting as translational repressors. Nucleic Acids Res. 2010;38(20):7022–7036. doi:10.1093/nar/gkq557
9. Xu H, Wang P, You J, et al. Screening of Kozak-motif-located SNPs and analysis of their association with human diseases. Biochem Biophys Res Commun. 2010;392(1):89–94. doi:10.1016/j.bbrc.2010.01.002

10. Robert F, Pelletier J. Exploring the Impact of Single-Nucleotide Polymorphisms on Translation. Front Genet. 2018;9:507. doi:10.3389/fgene.2018.00507

11. Calvo SE, Pagliarini DJ, Mootha VK. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proc Natl Acad Sci U S A. 2009;106(18):7507–7512. doi:10.1073/pnas.0810916106

12. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017;8(1):1826. doi:10.1038/s41467-017-01261-5

13. Barbeira AN, Dickinson SP, Bonazzola R, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat Commun. 2018;9(1):1825. doi:10.1038/s41467-018-03621-1

14. Yao DW, O’Connor LJ, Price AL, Gusev A. Quantifying genetic effects on disease mediated by assayed gene expression levels. Nat Genet. 2020;52(6):626–633. doi:10.1038/s41588-020-0625-2

15. Liu X, Li YI, Pritchard JK. Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. Cell. 2019;177(4):1022–1034.e6. doi:10.1016/j.cell.2019.04.014

16. Tashman KC, Cui R, O’Connor LJ, Neale BM, Finucane HK. Significance Testing for Small Annotations in Stratified LD-Score Regression; 2021:2021.03.13.21249938. doi:10.1101/2021.03.13.21249938

17. Watanabe K, Stringer S, Frei O, et al. A global overview of pleiotropy and genetic architecture in complex traits. Nat Genet. 2019;51(9):1339–1348. doi:10.1038/s41588-019-0481-0

18. Lonsdale J, Thomas J, Salvatore M, et al. The genotype-tissue expression (GTEx) project. Nat Genet. 2013;45(6):580–585.

19. Finucane HK, Bulik-Sullivan B, Gusev A, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet. 2015;47(11):1228–1235.

20. Akbarian S, Liu C, Knowles JA, et al. The PsychENCODE project. Nat Neurosci. 2015;18(12):1707–1712. doi:10.1038/nn.4156

21. Brodie A, Azaria JR, Ofran Y. How far from the SNP may the causative genes be? Nucleic Acids Res. 2016;44(13):6046–6054. doi:10.1093/nar/gkw500

22. Liu M, Jiang Y, Wedow R, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nat Genet. 2019;51(2):237–244.

23. Walters RK, Polimanti R, Johnson EC, et al. Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. Nat Neurosci. 2018;21(12):1656–1669. doi:10.1038/s41593-018-0275-1

24. Jansen IE, Savage JE, Watanabe K, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer’s disease risk. Nat Genet. 2019;51(3):404–413. doi:10.1038/s41588-018-0311-9
25. Watson HJ, Yilmaz Z, Thornton LM, et al. Genome-wide Association Study Identifies Eight Risk Loci and Implicates Metabo-Psychiatric Origins for Anorexia Nervosa. Nat Genet. 2019;51(8):1207–1214. doi:10.1038/s41588-019-0439-2

26. Otowa T, Hek K, Lee M, et al. Meta-analysis of genome-wide association studies of anxiety disorders. Mol Psychiatry. 2016;21(10):1391–1399. doi:10.1038/mp.2015.197

27. Mullins N, Forstner AJ, O’Connell KS, et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. Nat Genet. 2021;53(6):817–829. doi:10.1038/s41588-021-00857-4

28. Pulit SL, Stoneman C, Morris AP, et al. Meta-analysis of genome-wide association studies for body fat distribution in 694,649 individuals of European ancestry. Hum Mol Genet. 2019;28(1):166–174. doi:10.1093/hmg/ddy327

29. Jansen PR, Nagel M, Watanabe K, et al. Genome-wide meta-analysis of brain volume identifies genomic loci and genes shared with intelligence. Nat Commun. 2020;11(1):5606. doi:10.1038/s41467-020-19378-5

30. Johnson EC, Demontis D, Thorgeirsson TE, et al. A large-scale genome-wide association study meta-analysis of cannabis use disorder. Lancet Psychiatry. 2020;7(12):1032–1045. doi:10.1016/S2215-0366(20)30339-4

31. Lee JJ, Wedow R, Okbay A, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet. 2018;50(8):1112–1121. doi:10.1038/s41588-018-0147-3

32. Jansen PR, Watanabe K, Stringer S, et al. Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. Nat Genet. 2019;51(3):394–403. doi:10.1038/s41588-018-0333-3

33. Okbay A, Baselmans BML, De Neve JE, et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. Nat Genet. 2016;48(6):624–633. doi:10.1038/ng.3552

34. Yengo L, Sidorenko J, Kemper KE, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700,000 individuals of European ancestry. Hum Mol Genet. 2018;27(20):3641–3649. doi:10.1093/hmg/ddy271

35. Savage JE, Jansen PR, Stringer S, et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. Nat Genet. 2018;50(7):912–919. doi:10.1038/s41588-018-0152-6

36. Wray NR, Ripke S, Mattheisen M, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat Genet. 2018;50(5):668–681. doi:10.1038/s41588-018-0090-3

37. Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. Mol Psychiatry. 2018;23(5):1181–1188. doi:10.1038/mp.2017.154
38. Nievergelt CM, Maihofer AX, Klengel T, et al. International meta-analysis of PTSD genome-wide association studies identifies sex- and ancestry-specific genetic risk loci. Nat Commun. 2019;10(1):4558. doi:10.1038/s41467-019-12576-w

39. Yu D, Sul JH, Tsetsos F, et al. Interrogating the genetic determinants of Tourette syndrome and other tic disorders through genome-wide association studies. Am J Psychiatry. 2019;176(3):217–227. doi:10.1176/appi.ajp.2018.18070857

40. Ripke S, Neale BM, Corvin A, et al. Biological Insights From 108 Schizophrenia-Associated Genetic Loci. Nature. 2014;511(7510):421–427. doi:10.1038/nature13595

41. Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. Mol Autism. 2017;8:21. doi:10.1186/s13229-017-0137-9

42. Pardiñas AF, Holmans P, Pocklington AJ, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat Genet. 2018;50(3):381–389. doi:10.1038/s41588-018-0059-2

43. Badre D, Nee DE. Frontal Cortex and the Hierarchical Control of Behavior. Trends Cogn Sci. 2018;22(2):170–188. doi:10.1016/j.tics.2017.11.005

Figures

Figure 1. Gene-SNP Annotation Workflow

- Specifically expressed gene sets
- Physical proximity annotation
- Expression-based annotation
- GWAS summary statistics
- LDSC baseline annotations
- Cell-type-specific stratified LDSC
- LDSC enrichment coefficients
- All genes in expression dataset
- All genes with expression weight

Figure 1
Overview of S-LDSC physical proximity and expression-based annotation methods used to compare estimates of LDSC enrichment coefficients across tissues.

**Figure 2. Random Annotation Permutation Workflow**

Overview of permutation procedure to test whether a SNP annotation based on putative expression effects provides stronger evidence of cell type or tissue enrichment than random sets of nearby SNPs.
Figure 3

Regression coefficients of LDSC specifically expressed gene set analyses utilizing a physical proximity (grey) and expression-based (black) annotation across 13 brain regions for schizophrenia using the Clozuk GWAS. The dashed line represents an FDR < 5%.
Figure 4

Distribution of 1,000 permuted S-LDSC coefficients when using a random annotation procedure for schizophrenia in the frontal cortex within-brain analyses. Solid line represents mean S-LDSC coefficient. Dashed line represents expression-based annotation coefficient.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supptablesEBA.xlsx
- suppfiguresEBA.pdf