bioRxiv preprint doi: https://doi.org/10.1101/107482; this version posted June 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Stratified Linkage Disequilibrium Score Regression reveals enrichment of eQTL effects on complex traits is not tissue specific

Authors

Hill F. Ip,[1]* Rick Jansen,[2] Abdel Abdellaoui,[1] Meike Bartels,[1,3,4] , UK Brain Expression Consortium, Dorret I. Boomsma,[1,3,4] & Michel G. Nivard[1]**

Affiliations

[1] Department of Biological Psychology, Vrije Universiteit Amsterdam, the Netherlands

[2] Department of Psychiatry, VU University Medical Center, Amsterdam, the Netherlands

[3] Amsterdam Public Health Research Institute, Amsterdam, the Netherlands

[4] Neuroscience Amsterdam, the Netherlands

Corresponding authors

* h.f.ip@vu.nl

** m.g.nivard@vu.nl

bioRxiv preprint doi: https://doi.org/10.1101/107482; this version posted June 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Both gene expression levels and eQTLs (expression quantitative trait loci) are partially tissue-specific, complicating the detection of eQTLs in tissues with limited sample availability, such as the brain. However, eQTL overlap between tissues might be non-trivial, allowing for inference of eQTL functioning in the brain via eQTLs measured in readily accessible tissues, e.g. whole blood. Using Stratified Linkage Disequilibrium Score Regression (SLDSR), we quantify the enrichment in GWAS signal of blood and brain eQTLs in genome-wide association study (GWAS) for three brain-related (schizophrenia, BMI, and educational attainment), three immune-related traits (Crohn's disease, rheumatoid arthritis, and ulcerative colitis), and five traits not strongly associated with either tissue (age at menarche, coronary artery disease, height, LDL levels, and smoking behavior). Our analyses established significant enrichment of blood and brain eQTLs in their effects across all traits. As we do not know the true number of causal eQTLs, it is difficult to determine the precise magnitude of enrichment. We found no evidence for tissue-specific enrichment in GWAS signal for either eQTLs uniquely seen in the brain or whole blood. To follow up on our findings, we tested tissue-specific enrichment of eQTLs discovered in 44 tissues by the Genotype-Tissue Expression (GTEx) consortium, and, again, found no tissue-specific eQTL effects. Finally, we integrated the GTEx eQTLs with SNPs associated with tissue-specific histone modifiers, and interrogated its effect on rheumatoid arthritis and schizophrenia. We observed substantially enriched effects on schizophrenia, though again not tissue-specific. We conclude that, while eQTLs are strongly enriched in GWAS signal, the enrichment is not specific to the tissue used in eQTL discovery. Therefore, working with relatively accessible tissues, such as whole blood, as proxy for eQTL discovery is sensible and restricting lookups for GWAS hits to a specific tissue might not be advisable.

Key words

bioRxiv preprint doi: https://doi.org/10.1101/107482; this version posted June 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

eQTL; gene expression; tissue-specificity; complex traits; enrichment; genome-wide; brain; whole blood; stratified linkage disequilibrium score regression; SLDSR; eQTL discovery

Introduction

The aim of genome-wide association studies (GWAS) is to detect statistically significant associations between genetic variants, such as single nucleotide polymorphisms (SNPs), and a trait of interest (Hirschhorn and Daly 2005). GWAS have identified many genetic variants and thereby provided insights into the genetic architecture of complex traits (Hirschhorn and Daly 2005; Visscher et al. 2012). However, a large number of variants identified through GWAS are located outside of coding regions and specific knowledge of regulatory elements is limited (Lowe and Reddy 2015). Therefore, uncovering a relationship between GWAS hits and biological function has proven to be complicated (Lowe and Reddy 2015). Expression quantitative trait loci (eQTLs) contain SNPs that influence gene expression, and are not necessarily located in coding regions. eQTLs may aid functional annotation of SNPs that have been identified in a GWAS and are located outside of coding regions (Morley et al. 2004; Lowe and Reddy 2015). Previous work has found substantial enrichment of eQTLs among GWAS hits (Manolio et al. 2009; Nicolae et al. 2010; Torres et al. 2014) and an enrichment in their genome-wide effect on complex traits (Davis et al. 2013). Therefore, eQTLs are viewed as an important tool in moving from genome-wide association to biological interpretation.

As a result of differences in gene expression between cells originating from different tissues, eQTLs are potentially tissue-specific (Hernandez et al. 2012; GTEx Consortium 2015). Tissue-specificity poses no problem if the tissue of interest is readily available for research, such as whole blood. However, discovery of eQTLs gets complicated when measurement of expression levels in a tissue is limited by ethical and practical considerations, for example in brain tissue. Several studies have shown that the overlap between eQTLs from different tissues might actually be larger than initially assumed (Ding et al. 2010; Nica et al. 2011). The Genotype-Tissue Expression (GTEx) consortium identified eQTLs in a wide range of human tissues and showed that 54-90% of the eQTLs

bioRxiv preprint doi: https://doi.org/10.1101/107482; this version posted June 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

identified in one tissue are also designated as an eQTL in at least one other tissue (GTEx Consortium 2015; Aguet et al. 2016). In another study, Liu *et al* (2016) found a high average pairwise genetic correlation (r_g=0.738) of local gene expression between tissues. Nevertheless, small differences in terms of eQTL effect may be of considerable importance in terms of the effect an eQTL might have on complex traits related to specific tissues. It is, therefore, worthwhile to investigate the specific utility of tissue-specific eQTLs in their effect on complex traits, as studied in GWAS, as the discovery of eQTLs for tissues such as the brain might be advanced by eQTLs discovered in more accessible tissues, such as whole blood. The use of accessible tissues, though, depends on a substantial degree of similarity of eQTL effect across tissue, and to what extend eQTL differences between tissues are important in complex trait etiology.

Stratified Linkage Disequilibrium Score Regression (SLDSR) is a technique that estimates the SNP-heritability (h^2_{SNP}) of a trait based on GWAS summary statistics (Bulik-Sullivan et al. 2015; Finucane et al. 2015). By simultaneously analyzing multiple categories of SNPs (annotations), SLDSR can partition h^2_{SNP} by annotation (h^2_{annot}) and thereby provides a way to jointly quantify the enrichment in GWAS signal of several annotations. Here, we extend SLDSR by including annotations containing *cis*-eQTLs, i.e. eQTLs located closely to the gene with which they associate (Brem et al. 2002; Ramasamy et al. 2014), discovered in multiple tissues. To this end, we perform analyses based on representative eQTL resources, and consider a variety of traits as outcomes.

Firstly, we selected all eQTLs per gene discovered in large samples of RNA expression levels assessed in whole blood (N=4896)(Wright et al. 2014; Jansen et al. 2017) and in brain tissues (N=134) (Ramasamy et al. 2014), and quantified the contribution of these blood and brain eQTLs to the genetic variance in complex traits captured in GWAS. We then estimated tissue-specific eQTL effects on complex traits by quantifying the enrichments of eQTLs uniquely found in whole blood or uniquely found in brain, conditional on the enrichment of the complete blood eQTL annotation or complete brain eQTL annotation, respectively. We considered the effect of eQTLs on three brain-related phenotypes: schizophrenia, BMI, and educational attainment; three immune disorders:

bioRxiv preprint doi: https://doi.org/10.1101/107482; this version posted June 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Crohn's disease, rheumatoid arthritis, and ulcerative colitis; and five other complex traits and disorders: age at menarche, coronary artery disease, height, LDL levels, and smoking behavior.

Secondly, we retrieved all eQTLs identified in any of the 44 tissues from the GTEx consortium (N=70-361, median=126.5)(GTEx Consortium 2015; Aguet et al. 2016). We considered the enrichment in GWAS signal of the union of all GTEx eQTLs, and, additionally, the enrichment of tissue-specific eQTL effects on top of the union of all GTEx eQTLs. We expected to observe tissue-specific enrichment of eQTLs in their effects on complex traits related to the tissue in question, e.g. eQTLs discovered in immune-related tissues are expected to show higher enrichments in their effect on immune-related traits compared to eQTLs found in skin tissue. We considered tissue-specific enrichment of *cis*-eQTLs in their effect on schizophrenia (a disorder where there is strong prior evidence for the involvement of processes in the brain) and rheumatoid arthritis (a disease with strong prior evidence for the involvement of processes in the immune system) as GWAS for these traits are well powered for extended LD-score-based analyses. We further considered the enrichment of the intersection of *cis*-eQTLs discovered in any tissue, and histone modification in a specific tissue (i.e. tissue-specific epigenetically changed chromatin states in regulatory regions).

Our analyses were designed to elucidate the relation between eQTLs and complex traits, and to quantify the extent to which this relation is dependent on the tissue used in eQTL discovery. Our analysis further considered the enrichment of genomic regions related to gene expression and epigenetically modified in specific tissues.

Material and Methods

SLDSR method

A measure of linkage disequilibrium (LD) for each SNP, called an "LD score", can be computed by taking the sum of correlations between that SNP and all neighboring SNPs (Bulik-Sullivan et al. 2015; Finucane et al. 2015). Under a polygenic model, LD scores are expected to show a linear relationship with GWAS test statistics of corresponding SNPs, where the slope is proportional to h^2_{SNP}. For SLDSR,

bioRxiv preprint doi: https://doi.org/10.1101/107482; this version posted June 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

LD scores are based on only (functional) parts of the genome and used as predictors in a multiple linear regression (Finucane et al. 2015). In this manner, SLDSR is able to partition h^2_{SNP} into parts that are explained by these parts of the genome (i.e. h^2_{annot}), while accounting for influences of the remaining annotations in the model. The enrichment of an annotation is then obtained by taking the ratio of h^2_{annot} over the proportion of SNPs that fall within that annotation. For eQTLs, the denominator, i.e. the number of SNPs in the annotation, is a complicated quantity: not all significant eQTLs are likely causal; whereas including only lead, or putative causal, eQTLs may result in very small annotations located near genes and other regulatory elements, which presents a risk of inflated estimates of the enrichment in GWAS signal. Because of these issues, we consider all significant *cis*-eQTLs as an annotation, and retain additional gene-centric and regulatory annotations in the model.

Target traits

As outcome for SLDSR, we used summary statistics of GWAS on Crohn's disease (Jostins et al. 2012), rheumatoid arthritis (Okada et al. 2014), ulcerative colitis (Jostins et al. 2012), BMI (Speliotes et al. 2010), educational attainment (Rietveld et al. 2013), schizophrenia (Ripke et al. 2014), age at menarche (Perry et al. 2014), coronary artery disease (Schunkert et al. 2011), height (Allen et al. 2010), LDL levels (Teslovich et al. 2010), and smoking behavior (Furberg et al. 2010). The first three traits were chosen because these are related to the immune system and are therefore expected to reveal considerable enrichment of blood eQTL signal (Jostins et al. 2012; Okada et al. 2014). Similarly, brain eQTLs are expected to show substantial enriched effects due to previous reports on the involvement of the central nervous system (CNS) in schizophrenia (Ripke et al. 2014), educational attainment (Rietveld et al. 2013), and BMI (Vimaleswaran et al. 2012). Of course, these traits do not perfectly align with either tissue, e.g. the immune system has been implicated in the etiology of schizophrenia (Andreassen et al. 2015) and BMI (Karalis et al. 2009), and might therefore also be enriched in their effects for the other eQTL set. However, this is expected to occur at lower rates.

bioRxiv preprint doi: https://doi.org/10.1101/107482; this version posted June 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Enrichment of blood and brain eQTL effects on the remaining traits was calculated to contrast the results with traits for which we do not have a strong *a priori* expectation of the relationship between trait and tissue.

The discovery sample for detection of blood eQTLs (Wright et al. 2014; Jansen et al. 2017) included participants from the Netherlands Twin Register (NTR)(Boomsma et al. 2008; Ripke et al. 2013) and Netherlands Study of Depression and Anxiety (NESDA)(Penninx et al. 2008). Subjects from these studies, not necessarily the same ones, also participated in the GWAS for some of the traits examined (Allen et al. 2010; Speliotes et al. 2010; Teslovich et al. 2010; Furberg et al. 2010; Rietveld et al. 2013; Perry et al. 2014). To ensure that the discovery sample did not affect estimates of enrichments of eQTL effects in the various GWAS signals, we looked at trait-specific enrichment of blood and brain eQTL signal in GWAS signal for educational attainment and smoking behavior. We compared the results from using publicly available datasets with using summary statistics based on the same sample without subjects from the NTR or NESDA. The results did not reveal appreciable differences between the respective datasets for educational attainment, but did show substantial differences for smoking behavior (S1 Figure). This latter finding could conceivably be a function of relatively strong effects of smoking behavior on gene expression levels (Vink et al. 2015). Therefore, the remaining analyses for smoking behavior were performed using the summary statistics without participants from NTR or NESDA, whereas analyses for the other traits (age at menarche, BMI, educational attainment, height, and LDL level) were run using publicly available summary statistics.

Blood and brain eQTL enrichment

A catalog of whole blood *cis*-eQTLs was obtained from Jansen *et al* (2017; Wright et al. 2014), where all eQTLs significantly associated with gene expression in whole blood for each probe set were selected for inclusion in our whole blood eQTL annotation. A list of brain eQTLs was obtained from the UK brain expression consortium (UKBEC), for which the analyses are described in Ramasamy *et al* (2014). We based the brain eQTL annotation on SNPs that were significantly associated with the

bioRxiv preprint doi: https://doi.org/10.1101/107482; this version posted June 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

average gene expression across 12 brain regions. SLDSR annotations were constructed as per the instructions in Bulik-Sullivan et al. and Finucane et al. (2015). To guard against upward bias in the eQTL enrichment signal, two extra annotations containing SNPs within a 500 base pair (bp) and 100bp window around any eQTL were constructed for each eQTL set (Finucane et al. 2015). Finally, to ensure that the enrichment of eQTL effects in GWAS signal was not in fact caused by their proximity to the genes they influence, an additional gene centric annotation was computed, which contained all genes for which eQTLs were included.

Tissue-specific eQTL enrichment

To distinguish between the effects of blood- and brain-specific eQTLs, we split each annotation into two sets based on the overlap in genes that were tagged by eQTLs from both tissue. That is, the brain eQTL annotation was split into an annotation of brain eQTLs which regulate genes for which also at least one blood eQTL was found, and a second annotation of eQTLs that tagged genes for which only brain eQTLs were found. Likewise, the blood eQTL annotation was split into an annotation containing only eQTLs that tagged genes for which eQTLs from both tissue was found, and an annotation consisting of blood eQTLs that tagged genes for which only eQTLs have been found in blood.

Enrichment of eQTLs from 44 tissues

There are limitations to above mentioned analyses of tissue-specific enrichments of eQTL effects in GWAS signal. The eQTLs are obtained from two different projects, which vary in terms of sample size and their definition of an eQTL. To mitigate the heterogeneity between studies, and to extend to additional tissues. We performed additional analyses using eQTLs obtained by a common pipeline from 44 tissues (see S2 Table) and based on a broader eQTL locus definition (GTEx Consortium 2015; Aguet et al. 2016). For each of the 44 tissues, we created annotations for analysis in SLDSR following the previously described procedure. Analogous to the procedure of Finucane *et al* (2015) for cell-

bioRxiv preprint doi: https://doi.org/10.1101/107482; this version posted June 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

type-specific analysis using SLDSR, we additionally created an annotation that contained all GTEx eQTLs, i.e. a SNP would be included in this annotation if it was designated as part of at least one of the 44 tissue-specific GTEx annotations, and added a 100bp and 500bp window. No windows were specified for the tissue-specific GTEx annotations. Using GWAS summary statistics for schizophrenia, we then ran one SLDSR model containing only the baseline categories and the union of GTEx eQTLs, and 44 additional models with the two previous annotations and one of the tissue-specific GTEx annotations at a time. We repeated the procedure using summary statistics for rheumatoid arthritis.

GTEx has relative small sample sizes for the brain eQTL discovery (mean=89 sample size, range=72-103) compared to other tissues (mean=160 sample size, range=70-361) (GTEx Consortium 2015; Aguet et al. 2016). To investigate the effect of differences in sample size on estimates of enrichments in GWAS signal, we collapsed the union of individual brain eQTL annotations into a shared brain eQTL annotation (i.e. an eQTL found in at least one of the GTEx brain annotations was included in the shared brain eQTL annotation). This annotation was then analyzed as an additional GTEx eQTL annotation. We further tested the relationship between tissue sample size and tissue eQTL enrichment.

Enrichment of the intersection between eQTLs and histone marks

The availability of annotations based on tissue-specific histone marks made it possible to create an annotation that represents the intersection between eQTLs and this type of epigenetic modification related to enhancers and promoters of actively transcribed genes. We obtained LD score annotations of SNPs in regions that bare histone marks in cells from the CNS or immune system from Finucane *et al* (2015). Out of the 220 cell-type-specific histone mark that were available, 101 were found in the CNS or immune tissues. For each of the 101 annotations of SNPs in cell-type-specific histone marks, we extracted its intersection with the union of GTEx eQTLs and made a new annotation of eQTLs which intersected with histone marks (i.e. SNPs found in both annotations). We then analyzed each of the intersection annotations individually in a model together with the baseline

bioRxiv preprint doi: https://doi.org/10.1101/107482; this version posted June 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

categories, the union of GTEx eQTLs, and the corresponding cell-type-specific histone marks. Enrichments in GWAS signal of the intersection should be interpreted as enrichment of genome-wide SNP effects on a complex trait beyond the additive effects which work on all SNPs that are a *cis*-eQTL and histone mark in question. In fact, we test whether the interaction between tissue-specific chromatin state and eQTLs are enriched in their genome-wide effect on complex traits.

Results

Blood and brain eQTL enrichment

We fitted an SLDSR model containing the baseline categories; the complete annotation for both brain and blood eQTL tissues, their 100 and 500bp windows, and gene-centric annotations to all traits (Crohn's disease, rheumatoid arthritis, ulcerative colitis, BMI, educational attainment, schizophrenia, age at menarche, coronary artery disease, height, LDL levels, and smoking behavior). We found significant effects of brain eQTLs on educational attainment, rheumatoid arthritis, smoking behavior, and schizophrenia, and significant effect of blood eQTLs on height and smoking behavior (see S3 Table). We then meta-analyzed the results for all annotations, both in the baseline model, and those associated with eQTLs across the 11 traits. Our analyses revealed significant effect of both blood ($p < 0.001$) and brain ($p < 0.001$) eQTL effects on all traits (Figure 1, S4 Table), exceeding, in terms of significance, all the baseline categories considered by Finucane *et al* (2015) but conserved genomic regions. The gene-centric annotation for both blood and brain eQTLs showed no effect on any trait. We further observed no evidence for depletion of blood-specific eQTLs (relative to all blood eQTLs) on brain-related traits, nor do we find significant depletion of effect on immune-related traits of eQTLs associated with genes for which eQTLs were solely identified in brain tissue (Table I).

Enrichment of eQTLs from 44 tissues in GTEx

bioRxiv preprint doi: https://doi.org/10.1101/107482; this version posted June 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We interrogated the enrichment of the union of GTEx eQTLs and 44 tissue-specific GTEx annotations in their effect on schizophrenia and rheumatoid arthritis. Figure 2 shows the coefficient Z-scores of the 45 GTEx annotations, sorted from largest to smallest. In both cases, the union of GTEx eQTLs had a substantial Z-score (Z=5.501 and Z=3.802 for schizophrenia and rheumatoid arthritis, respectively, both $p<0.001$, S5 Table), indicating that eQTLs were significantly enriched in their effects on complex traits. The tissue-specific annotations, however, performed notably worse and in some cases even suggested depletion of genome-wide effects of tissue-specific eQTLs on schizophrenia and rheumatoid arthritis. For rheumatoid arthritis, the coefficient Z-scores of the whole blood annotation reached nominal significance (Z=2.036, $p=0.021$), but failed correction for multiple testing. None of the other annotations reached nominal significance. The union of all GTEx brain annotations did not contribute significantly to explaining h^2_{SNP} (Z=0.147, $p=0.441$). Sample size in the eQTL discovery phase appears to be a strong determinant of tissue-specific enrichment in GWAS signal. The correlation coefficients between the coefficient Z-scores and sample sizes were 0.6453 ($p=2.253*10^{-6}$) and 0.4247 ($p=0.004$) for schizophrenia and rheumatoid arthritis, respectively.

Enrichment of the intersection between eQTLs and histone marks

We interrogated the intersection of eQTLs and histone marks found in specific CNS and immune cells, and estimated the enrichment of the intersection in its effect on rheumatoid arthritis and schizophrenia. We found significant enrichment in GWAS signal for eQTLs that intersect with histones that bare modification H3K4me1, a modification thought to be present in the enhancer of actively transcribed genes (Zhou et al. 2011; Allis and Jenuwein 2016), in CNS cells for schizophrenia (see Figure 3). There was some evidence for significant enrichment of eQTLs that intersected with genomic regions in immune cells baring the H3K4me1 mark in their effect on schizophrenia, but not on rheumatoid arthritis. Specifically, none of the intersecting annotations showed evidence of enrichment for rheumatoid arthritis. For the separate annotations, we found significant enrichment in GWAS signal across all histone marks found in CNS cells and three significant immune cell-types

bioRxiv preprint doi: https://doi.org/10.1101/107482; this version posted June 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

that bare the H3K4me3 modification, a modification associated with transcriptional start sites and promoters of actively transcribed genes (Zhou et al. 2011; Allis and Jenuwein 2016), for schizophrenia (S6 Figure). The opposite picture was seen for rheumatoid arthritis: a wide variety of immune-cell specific histone marks showed significant enrichments in GWAS signal, while all marks found in CNS cells were below zero. The union of GTEx eQTLs reached statistical significance for all models (S6 Figure).

Discussion

Stratified Linkage Disequilibrium Score Regression provides a way to partition h^2_{SNP} into parts explained by (functional) parts of the genome (Finucane et al. 2015). A "full baseline model" containing 24 non-cell-type-specific annotations of SNPs, such as SNPs located in promoters or coding regions, was developed previously for analysis using SLDSR. Here, we added annotations containing eQTLs derived from whole blood and brain tissue into the model, and showed that eQTLs were substantially stronger enriched in their effect on complex traits compared to all categories considered by Finucane *et al* (2015). The complete brain eQTL annotation was significantly enriched in GWAS signal for educational attainment, rheumatoid arthritis, smoking behavior, and schizophrenia. This finding is consistent with previous estimates of eQTL effect enrichment (Davis et al. 2013). Considerable enrichment for eQTLs, even for traits not apparently linked to the brain or immune system (e.g. smoking behavior), suggested that non-trivial eQTL overlap across tissues might be present.

Inclusion of both brain and blood eQTLs into the SLDSR model did not separate the signal into tissue-specific effects. In general, we are not able to clearly identify tissue-specific eQTL signals using these datasets and SLDSR. Our second analysis of eQTL enrichment based on 44 tissue-specific *cis*-eQTL sets, obtained from the GTEx consortium (2015; Aguet et al. 2016), confirms the lack of tissue-specific eQTL enrichment. While an annotation containing all eQTLs identified in GTEx is significantly enriched in its effect on schizophrenia and rheumatoid arthritis (Z=5.501 and Z=3.802,

bioRxiv preprint doi: https://doi.org/10.1101/107482; this version posted June 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

respectively, both $p<0.001$), none of the analyzed brain tissues are enriched beyond all eQTLs in their effect on schizophrenia. Similarly, whole blood eQTLs are not significantly enriched beyond all GTEx eQTLs taken together in their effect on rheumatoid arthritis. Again, these findings are not consistent with the hypothesis of abundant tissue-specific *cis*-eQTLs with effects on complex traits related to the specific tissue in question. Especially, when contrasted with tissue-specific gene expression levels and tissue-specific histone modifications (Liu et al. 2016; Finucane et al. 2017), tissue-specific eQTLs are of limited value in relating complex traits to a tissue. Our conclusions are limited to *cis*-eQTLs and it is not unlikely that *trans*-eQTLs behave differently in terms of tissue-specificity. We do, find evidence for possible enrichment for eQTLs that intersect with tissue-specific H3K4me1 histone marks in the brain, but also immune cells, in their effect on schizophrenia but not rheumatoid arthritis. This means that eQTLs in H3k4me1 marks are enriched in their effect on schizophrenia above the expected enrichment based on the fact that these SNPs are both eQTLs and located in H3K4me1 histone marks. What is of substantial interest is that the enrichment in GWAS signal appears specific to H3K4me1 marks, and no other histone marks, suggesting that these marks specifically can aid in prioritizing genomic regions in which tissue-specific eQTLs may reside. Though, again, the totality of evidence is inconclusive on the relevance of tissue-specific eQTLs to variation in complex traits.

Our results are consistent with, and complimentary to, a study investigating the genetic correlation between gene expression levels across 15 tissues (Liu et al. 2016). This study revealed substantial correlations between *cis*-genetic effects on gene expression across 15 tissues (Liu et al. 2016). Our analyses confirmed the value of using whole blood as discovery tissue for detection of cis-eQTLs and further demonstrated the usefulness of techniques that use *cis*-eQTLs discovered in whole blood to study the etiology of complex traits related to different tissues (Gamazon et al. 2015; Gusev et al. 2016). The results presented here highlight the overlap of *cis*-eQTL effects across tissues on a genome-wide level. However, the effect of a *cis*-eQTL might vary substantially across tissues for individual genes (Grundberg et al. 2012). Our conclusions are based on genome-wide enrichments

bioRxiv preprint doi: https://doi.org/10.1101/107482; this version posted June 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and therefore should not be interpreted as limited evidence for tissue-specific eQTL effects for individual genes. Therefore, eQTL discovery in the tissue most relevant to a specific trait or disorder remains important to further our understanding of the genetic regulation of tissue-specific gene expression. What is also clear is that, to discover those tissue-specific eQTLs that are of relevance to the interpretation of GWAS of complex traits, tissue-specific eQTL discovery needs to be refined. The practice of, as a post-hoc analysis to GWAS, performing eQTL lookup in a specific tissue linked to a trait, when larger dataset for other accessible tissues are available, may be suboptimal. In fact, one may prefer to perform a lookup in the overlap between histone modifications in a relevant tissue and eQTLs regardless of tissue. One can further consider utilizing eQTLs to link GWAS findings to a gene, and subsequently consider the differential expression of a gene to identify the tissue in which the gene is most likely to act in effecting the trait. Tissue-specific differential gene expression vastly outperforms eQTLs in tagging regions of the genome enriched in their effect on complex traits (Finucane et al. 2017).

It is also evident that a limited dichotomous definition of eQTL/no-eQTL may be insufficient to identify tissue-specific eQTL effects. An evident improvement would be to compute the *difference* in eQTL effect on expression of the gene between tissues, and perform inference based on this difference in effect. eQTLs are strongly enriched SNPs, with clear biological function and utility for the translation of GWAS findings, though tissue-specific eQTL mechanisms remain elusive. The discovery of tissue-specific eQTL effects, which can aid in linking complex trait to tissue, may require novel research strategies.

Supplemental Data

Supplemental Data includes 2 figures and 4 tables

Compliance with Ethical Standards

bioRxiv preprint doi: https://doi.org/10.1101/107482; this version posted June 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Funding

MGN is supported by the Royal Netherlands Academy of Science Professor Award (PAH/6635) to DIB. HFI is supported by the "Aggression in Children: Unraveling gene-environment interplay to inform Treatment and InterventiON strategies" (ACTION) project. ACTION receives funding from the European Union Seventh Framework Program (FP7/2007-2013) under grant agreement no 602768. MB is supported by a University Research Chair of the Vrije Universiteit. The discovery of blood eQTL was funded by the US National Institute of Mental Health (RC2 MH089951, principal Investigator PFS) as part of the American Recovery and Reinvestment Act of 2009. We thank T. Lehner (National Institute of Mental Health) for his support. We acknowledge Hillary Finucane, Raymond Walters and Benjamin Neale for critical comments on our methods, design and manuscript.

Conflict of Interest

The authors declare that they have no conflict of interest

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Web Resources

Age at menarche summary statistics, www.reprogen.org/data_download.html

Blood eQTLs, https://eqtl.onderzoek.io/

Brain eQTLs, http://www.braineac.org/

Coronary artery disease summary statistics, www.cardiogramplusc4d.org/data-downloads/

Crohn's disease and ulcerative colitis summary statistics, www.ibdgenetics.org/downloads.html

Educational attainment summary statistics, http://www.thessgac.org/data

bioRxiv preprint doi: https://doi.org/10.1101/107482; this version posted June 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Full baseline model LD scores, http://data.broadinstitute.org/alkesgroup/LDSCORE/

GTEx dataset, http://www.gtexportal.org/home/datasets

Height and BMI summary statistics,

www.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files

LDL levels summary statistics, www.broadinstitute.org/mpg/pubs/lipids2010/

Rheumatoid arthritis summary statistics, http://plaza.umin.ac.jp/yokada/datasource/software.htm

Schizophrenia and smoking behavior summary statistics, www.med.unc.edu/pgc/results-and-downloads

SLDSR software, https://github.com/bulik/ldsc/

bioRxiv preprint doi: https://doi.org/10.1101/107482; this version posted June 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

Aguet F, Brown AA, Castel S, et al (2016) Local genetic effects on gene expression across 44 human tissues. bioRxiv. doi: https://doi.org/10.1101/074450

Allen HL, Estrada K, Lettre G, et al (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 467:832–838.

Allis CD, Jenuwein T (2016) The molecular hallmarks of epigenetic control. Nat Rev Genet 17:487–500.

Andreassen OA, Harbo HF, Wang Y, et al (2015) Genetic pleiotropy between multiple sclerosis and schizophrenia but not bipolar disorder: differential involvement of immune-related gene loci. Mol Psychiatry 20:207–14. doi: 10.1038/mp.2013.195

Boomsma DI, Willemsen G, Sullivan PF, et al (2008) Genome-wide association of major depression: description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. Eur J Hum Genet 16:335–342. doi: 10.1038/sj.ejhg.5201979

Brem RB, Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. Science 296:752–5. doi: 10.1126/science.1069516

Bulik-Sullivan BK, Finucane HK (2015) LD Score Estimation Tutorial. https://github.com/bulik/ldsc/wiki/LD-Score-Estimation-Tutorial. Accessed 21 Nov 2015

Bulik-Sullivan BK, Loh P-R, Finucane HK, et al (2015) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 47:291–295. doi: 10.1038/ng.3211

Davis LK, Yu D, Keenan CL, et al (2013) Partitioning the heritability of Tourette syndrome and obsessive compulsive disorder reveals differences in genetic architecture. PLoS Genet 9:e1003864. doi: 10.1371/journal.pgen.1003864

Ding J, Gudjonsson JE, Liang L, et al (2010) Gene expression in skin and lymphoblastoid cells: Refined statistical method reveals extensive overlap in cis-eQTL signals. Am J Hum Genet 87:779–89. doi: 10.1016/j.ajhg.2010.10.024

bioRxiv preprint doi: https://doi.org/10.1101/107482; this version posted June 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Finucane H, Reshef Y, Anttila V, et al (2017) Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types.

Finucane HK, Bulik-Sullivan B, Gusev A, et al (2015) Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet 47:1228–1235. doi: 10.1038/ng.3404

Furberg H, Kim Y, Dackor J, et al (2010) Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nat Genet 42:441–7. doi: 10.1038/ng.571

Gamazon ER, Wheeler HE, Shah KP, et al (2015) A gene-based association method for mapping traits using reference transcriptome data. Nat Genet 47:1091–1098. doi: 10.1038/ng.3367

Grundberg E, Small KS, Hedman ÅK, et al (2012) Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat Genet 44:1084–9. doi: 10.1038/ng.2394

GTEx Consortium (2015) The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science (80-) 348:648–660.

Gusev A, Ko A, Shi H, et al (2016) Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet 48:245–252. doi: 10.1038/ng.3506

Hernandez DG, Nalls MA, Moore M, et al (2012) Integration of GWAS SNPs and tissue specific expression profiling reveal discrete eQTLs for human traits in blood and brain. Neurobiol Dis 47:20–8. doi: 10.1016/j.nbd.2012.03.020

Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 6:95–108. doi: 10.1038/nrg1521

Jansen R, Hottenga J-J, Nivard MG, et al (2017) Conditional eQTL Analysis Reveals Allelic Heterogeneity of Gene Expression. Hum Mol Genet. doi: 10.1093/hmg/ddx043

Jostins L, Ripke S, Weersma RK, et al (2012) Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 491:119–124.

Karalis KP, Giannogonas P, Kodela E, et al (2009) Mechanisms of obesity and related pathology: linking immune responses to metabolic stress. FEBS J 276:5747–5754.

bioRxiv preprint doi: https://doi.org/10.1101/107482; this version posted June 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Liu X, Finucane HK, Gusev A, et al (2016) Functional partitioning of local and distal gene expression regulation in multiple human tissues. bioRxiv 46383. doi: 10.1101/046383

Lowe WL, Reddy TE (2015) Genomic approaches for understanding the genetics of complex disease. Genome Res 25:1432–1441. doi: 10.1101/gr.190603.115

Manolio TA, Collins FS, Cox NJ, et al (2009) Finding the missing heritability of complex diseases. Nature 461:747–53. doi: 10.1038/nature08494

Morley M, Molony CM, Weber TM, et al (2004) Genetic analysis of genome-wide variation in human gene expression. Nature 430:743–7. doi: 10.1038/nature02797

Nica AC, Parts L, Glass D, et al (2011) The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. PLoS Genet 7:e1002003. doi: 10.1371/journal.pgen.1002003

Nicolae DL, Gamazon E, Zhang W, et al (2010) Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet 6:e1000888. doi: 10.1371/journal.pgen.1000888

Okada Y, Wu D, Trynka G, et al (2014) Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 506:376–381.

Penninx BWJH, Beekman ATF, Smit JH, et al (2008) The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. Int J Methods Psychiatr Res 17:121–140.

Perry JRB, Day F, Elks CE, et al (2014) Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. Nature 514:92–97.

Ramasamy A, Trabzuni D, Guelfi S, et al (2014) Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat Neurosci 17:1418–28. doi: 10.1038/nn.3801

Rietveld CA, Medland SE, Derringer J, et al (2013) GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. Science (80-) 340:1467–1471.

Ripke S, Neale BM, Corvin A, et al (2014) Biological insights from 108 schizophrenia-associated genetic loci. Nature 511:421–427. doi: 10.1038/nature13595

bioRxiv preprint doi: https://doi.org/10.1101/107482; this version posted June 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Ripke S, Wray NR, Lewis CM, et al (2013) A mega-analysis of genome-wide association studies for major depressive disorder. Mol Psychiatry 18:497–511. doi: 10.1038/mp.2012.21

Schunkert H, König IR, Kathiresan S, et al (2011) Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nat Genet 43:333–338.

Speliotes EK, Willer CJ, Berndt SI, et al (2010) Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet 42:937–948.

Teslovich TM, Musunuru K, Smith A V, et al (2010) Biological, clinical and population relevance of 95 loci for blood lipids. Nature 466:707–713.

Torres JM, Gamazon ER, Parra EJ, et al (2014) Cross-tissue and tissue-specific eQTLs: partitioning the heritability of a complex trait. Am J Hum Genet 95:521–34. doi: 10.1016/j.ajhg.2014.10.001

Vimaleswaran KS, Tachmazidou I, Zhao JH, et al (2012) Candidate genes for obesity-susceptibility show enriched association within a large genome-wide association study for BMI. Hum Mol Genet 21:4537–42. doi: 10.1093/hmg/dds283

Vink JM, Jansen R, Brooks A, et al (2015) Differential gene expression patterns between smokers and non-smokers: cause or consequence?

Visscher PM, Brown MA, McCarthy MI, Yang J (2012) Five Years of GWAS Discovery. Am J Hum Genet 90:7–24. doi: 10.1016/j.ajhg.2011.11.029

Wright FA, Sullivan PF, Brooks AI, et al (2014) Heritability and genomics of gene expression in peripheral blood. Nat Genet 46:430–437. doi: 10.1038/ng.2951

Zhou VW, Goren A, Bernstein BE (2011) Charting histone modifications and the functional organization of mammalian genomes. Nat Rev Genet 12:7–18.

Figure Titles and Legends

Figure 1. Average enrichment in GWAS signal of the 24 baseline annotations, 4 brain eQTL annotations and 4 blood eQTL annotations.

bioRxiv preprint doi: https://doi.org/10.1101/107482; this version posted June 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Bar plot of the average enrichment in GWAS signal across all traits for the 24 main baseline annotations and 8 main eQTL annotations. Grey beans represent the baseline categories. Blue beans represent eQTLs. Black bars indicate average enrichment. Boxes show upper- and lower-bounds of the 95% confidence interval of the mean. Red dots show enrichments for immune-related traits. Horizontal red line indicates enrichment of 1, i.e. no enrichment.

Figure 2. Coefficient Z-scores of the 45 GTEx annotations

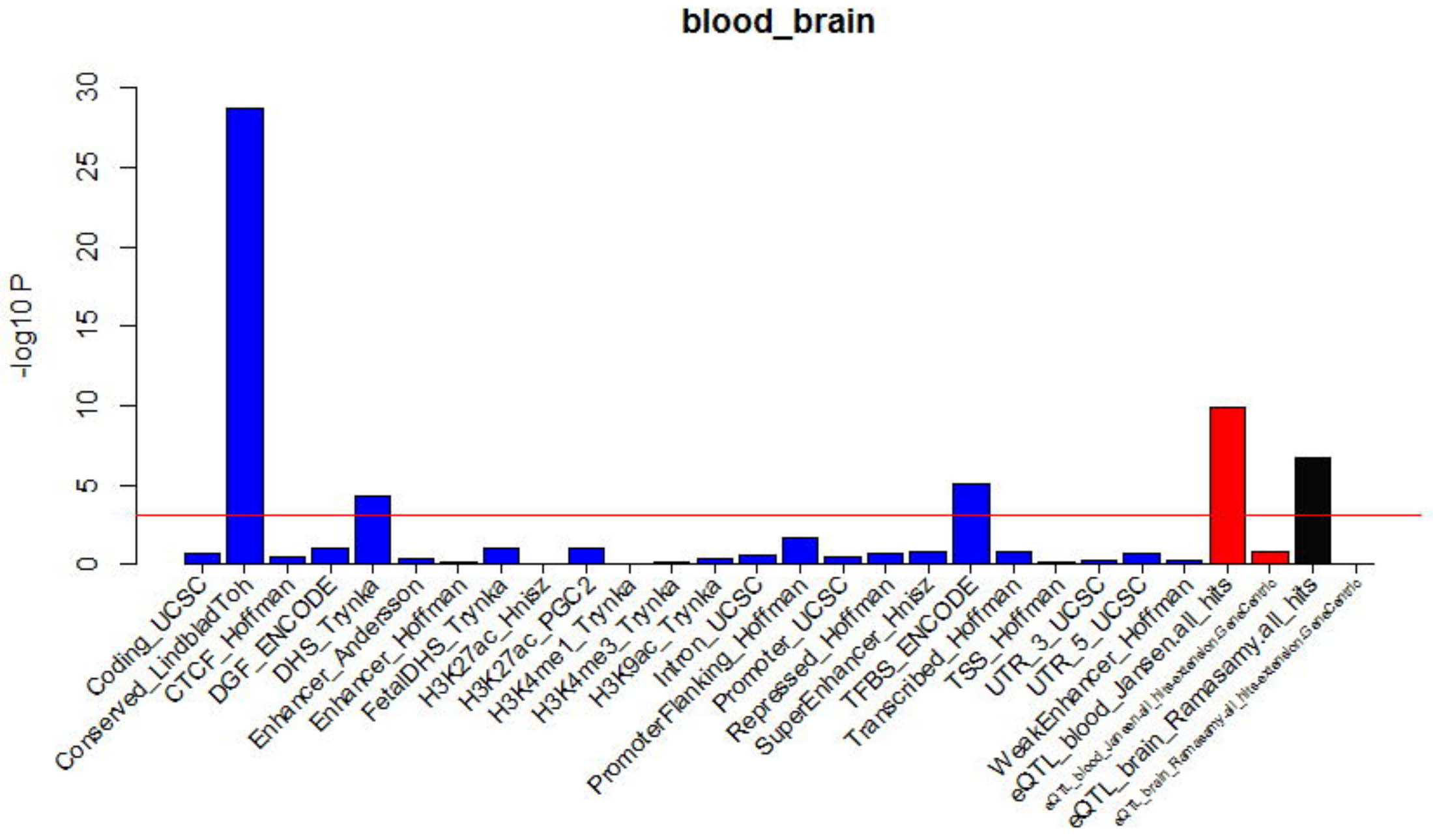
Barplot of coefficient z-scores for all GTEx annotations for schizophrenia (grey) and rheumatoid arthritis (red). Bars are sorted from highest to lowest based on the results from schizophrenia. Horizontal dotted line indicates Bonferroni threshold for 45 tests. Two asterisks indicate bars passing Bonferroni correction for multiple testing.

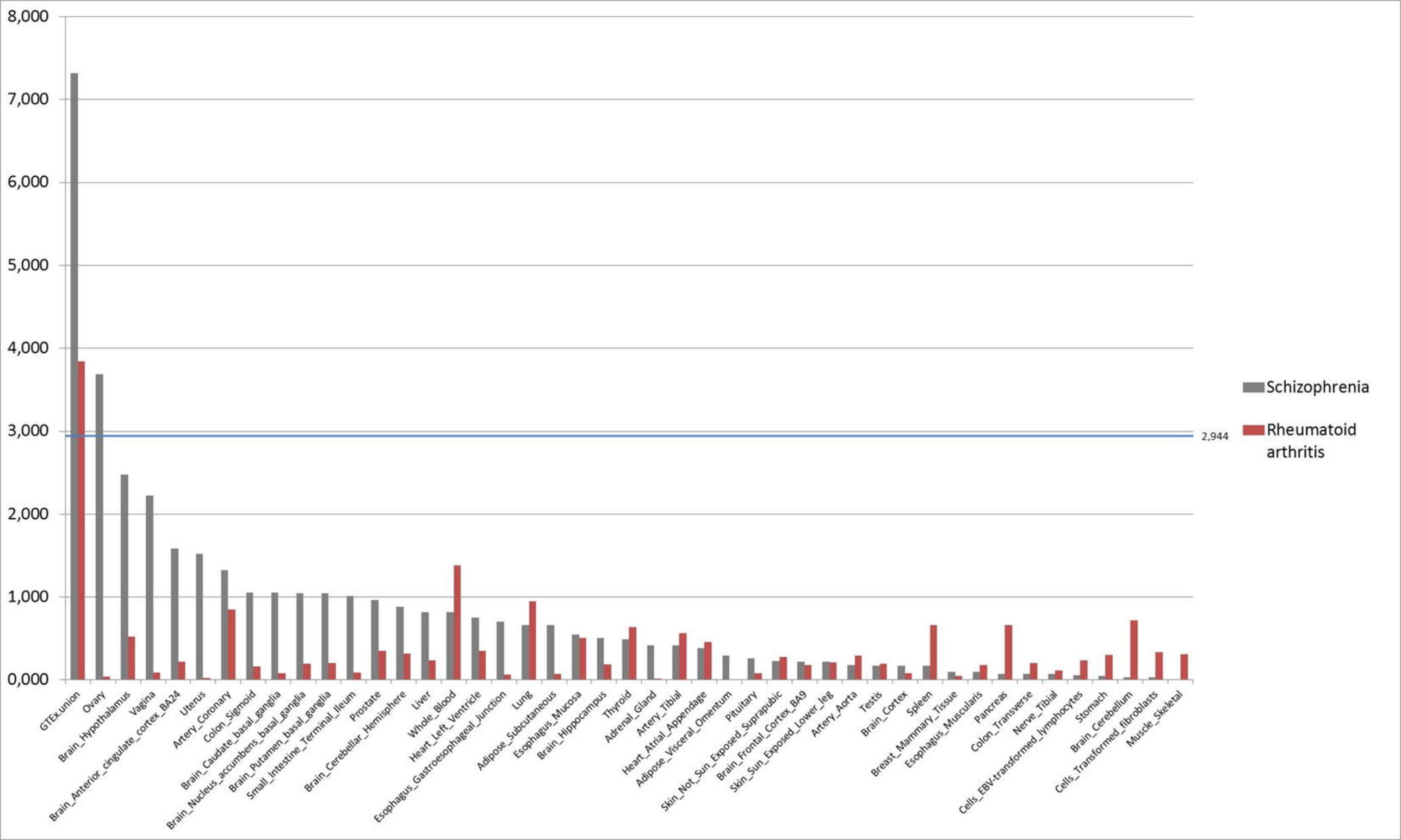
Figure 3. Coefficient Z-score of intersection between union of GTEx eQTLs and cell-type-specific histone marks

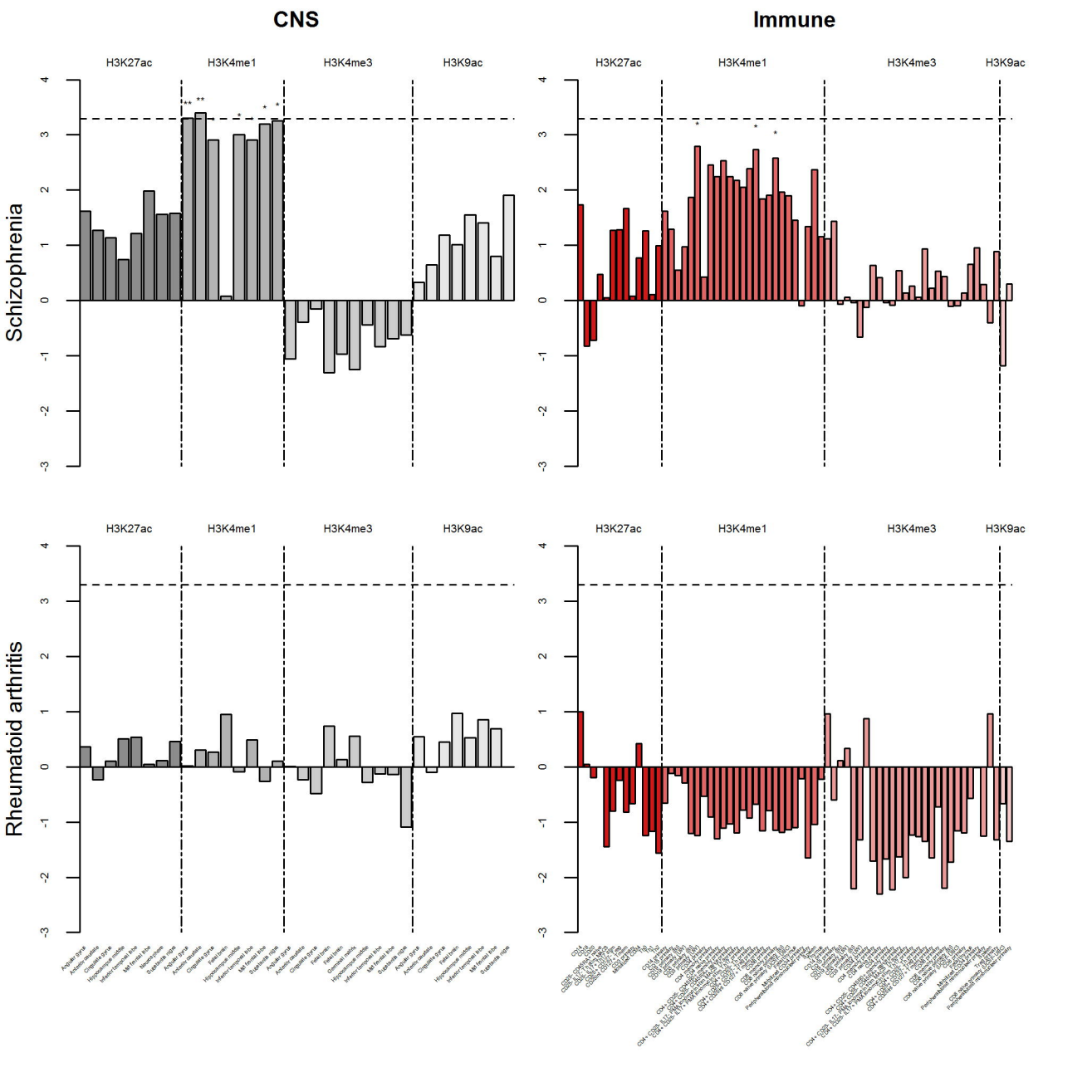
Top two graphs show coefficient Z-scores for schizophrenia. Bottom two graphs show the same for rheumatoid arthritis. Grey bars indicate histone marks found in cells from the central nervous system. Red bars represent histone marks found in cells from the immune system. From dark to light, shades of the bars indicate histone marks H3K27ac, H3K4me1, H3K4me3, and H3K9ac. Vertical dotted lines indicate separation between histone marks. One asterisk above the bars indicate annotations passing FDR correction for multiple testing. Two asterisks indicate bars passing Bonferroni correction for multiple testing. Horizontal dotted line indicates Bonferroni threshold for 101 tests.

Tables

Table I. (see attachments)


blood_brain
-log10 P
0
5
10
15
20
25
30
Coding_UCSC
Conserved_LindbladToh
CTCF_Hoffman
DGF_ENCODE
DHS_Trynka
Enhancer_Andersson
Enhancer_Hoffman
FetalDHS_Trynka
H3K27ac_Hnisz
H3K27ac_PGC2
H3K4me1_Trynka
H3K4me3_Trynka
H3K9ac_Trynka
Intron_UCSC
PromoterFlanking_Hoffman
Promoter_UCSC
Repressed_Hoffman
SuperEnhancer_Hnisz
TFBS_ENCODE
Transcribed_Hoffman
TSS_Hoffman
UTR_3_UCSC
UTR_5_UCSC
WeakEnhancer_Hoffman
eQTL_blood_Jansen.all_hits
eQTL_blood_Jansen.all_hits.extension.GeneCentric
eQTL_brain_Ramasamy.all_hits
eQTL_brain_Ramasamy.all_hits.extension.GeneCentric


8,000
7,000
6,000
5,000
4,000
3,000
2,000
1,000
0,000
2,944
Schizophrenia
Rheumatoid arthritis
GTEx.union
Ovary
Brain_Hypothalamus
Vagina
Brain_Anterior_cingulate_cortex_BA24
Uterus
Artery_Coronary
Colon_Sigmoid
Brain_Caudate_basal_ganglia
Brain_Nucleus_accumbens_basal_ganglia
Brain_Putamen_basal_ganglia
Small_Intestine_Terminal_Ileum
Prostate
Brain_Cerebellar_Hemisphere
Liver
Whole_Blood
Heart_Left_Ventricle
Esophagus_Gastroesophageal_Junction
Lung
Adipose_Subcutaneous
Esophagus_Mucosa
Brain_Hippocampus
Thyroid
Adrenal_Gland
Artery_Tibial
Heart_Atrial_Appendage
Adipose_Visceral_Omentum
Pituitary
Skin_Not_Sun_Exposed_Suprapubic
Brain_Frontal_Cortex_BA9
Skin_Sun_Exposed_Lower_leg
Artery_Aorta
Testis
Brain_Cortex
Spleen
Breast_Mammary_Tissue
Esophagus_Muscularis
Pancreas
Colon_Transverse
Nerve_Tibial
Cells_EBV-transformed_lymphocytes
Stomach
Brain_Cerebellum
Cells_Transformed_fibroblasts
Muscle_Skeletal


CNS
Immune
Schizophrenia
Rheumatoid arthritis
H3K27ac
H3K4me1
H3K4me3
H3K9ac
4
3
2
1
0
-1
-2
-3