A framework for predicting tissue-specific effects of rare genetic variants

Farhan N. Damani[1], Yungil Kim[1], Xin Li[2], Emily K. Tsang[3], Joe R. Davis[4], Colby Chiang[5], Zachary Zappala[4], Benjamin J. Strober[6], Alexandra J. Scott[5], Ira M. Hall[5], GTEx Consortium , Stephen B. Montgomery[2,4], and Alexis Battle[1]

[1]**Department of Computer Science, Johns Hopkins University, Baltimore, MD.**
[2]**Department of Pathology, Stanford University, Stanford, CA.**
[3]**Biomedical Informatics Program, Stanford University, Stanford, CA.**
[4]**Department of Genetics, Stanford University, Stanford, CA.**
[5]**McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO.**
[6]**Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD.**

Corresponding author:
Alexis Battle[1]

Email address: ajbattle@cs.jhu.edu

ABSTRACT

Despite the abundance of rare genetic variants—variants carried by less than one percent of the population—in human genomes, the impact of these variants on specific tissues has been largely uncharacterized. Population-level test statistics, while effective in understanding the impact of common variants—variants carried by at least five percent of the population, have had limited success in characterizing the effect of rare variants mainly due to limited statistical power. In addition, the effect of each rare variant can vary greatly between specific tissues. This heterogeneity coupled with limited sample sizes and a lack of known disease-causing rare variants makes predicting tissue-specific cellular consequences of rare variants a difficult task. To make these predictions, we propose a new method called SPEER (SPecific tissuE variant Effect predictoR): a hierarchical Bayesian model that uses transfer learning, allowing separate predictions in each tissue while flexibly sharing signal across tissues to improve power. Our probabilistic model capitalizes on a growing body of rich epigenetic annotations to inform the consequences of a variant in specific tissues. These annotations are integrated with tissue-specific RNA expression levels and common variants. We show our method improves prediction accuracy in simulations and in genomic data from the Genotype-Tissue Expression (GTEx) project.

INTRODUCTION

Recent advances in genomic technologies provide us with a unique opportunity to study the contribution of genetic variation to disease risk. Genome-wide association studies (GWAS) have been largely successful over the past decade in identifying statistical associations between common genetic variants—those carried by at least five percent of the population, and complex traits and diseases including height, diabetes and heart disease. However, these statistical techniques do not generalize well to analyzing rare variants—variants carried by less than one percent of the population—due to low sample size (Uricchio et al., 2016). Because rare variants have been shown to be implicated in disease risk and shown to be potentially more deleterious than common variants (Tennessen et al., 2012; Nelson et al., 2012), developing methods that can effectively characterize these variants remains essential.

Several tools have been developed to understand the functional consequences of rare variants. Kircher et al. (2014) developed CADD, a supervised learning approach that used functional annotations of the genome to predict deleteriousness. Quang et al. (2015) built on the success of CADD with a deep learning approach also for predicting deleteriousness. Li et al. (2016) introduced RIVER, an unsupervised learning method that integrates genomic annotations with gene expression data from the same individual to prioritize deleterious variants. They showed that genomic annotations are enriched for variants nearby genes with extreme expression levels. Building on this knowledge, RIVER used gene expression outliers—samples with extreme over or under expression—across diverse tissues to prioritize deleterious variants. They were better able to identify deleterious variants with global effects compared to models that exclusively used genomic annotations.

While these methods have made significant strides in understanding the global impact of genetic variants, their usefulness in understanding the tissue-specific consequences of genetic variants is somewhat limited. Recent work by Backenroth et al. (2016) integrated tissue-specific regulatory elements with GWAS summary statistics in order to understand these effects. Despite providing unique insights about the sharing of genetic variants within known physiological tissue groups Aguet et al. (2016), these methods do not apply to rare variant analysis due to a lack of known pathogenic tissue-specific rare variants and a scarcity of samples.

Transfer learning, a framework that allows sharing of knowledge across learning tasks, has been shown to be effective in low-resource settings with complex structure (Thrun, 1996; McCallum et al., 1998). In the hierarchical Bayes framework, parameters for each task are dependent on each other through a Bayesian prior (Raina et al., 2006). Here we propose SPEER (SPecific tissuE variant Effect predictoR), a hierarchical Bayes model that uses transfer learning to predict the tissue-specific functional consequences of rare variants. Each task here translates to understanding the effects of rare variants in a specific tissue and the sharing across tissues captures global effects. By using transfer learning to share information across tissues, SPEER learns reliable parameters and prioritizes rare variants in a tissue-specific manner. SPEER has three parts. First, a per-

sample component models the effect of both genomic annotations and gene expression on the presence of rare regulatory variation. Second, a tissue-specific component models the influence of genomic annotations on individual tissues. Third, a global component models the shared impact of genomic annotations across tissues.

We apply our method to simulated data and data from the Genotype-Tissue Expression (GTEx) project and show that SPEER performs better than state-of-the-art baselines. The methods developed in this paper are available at https://github.com/farhand7/speer.

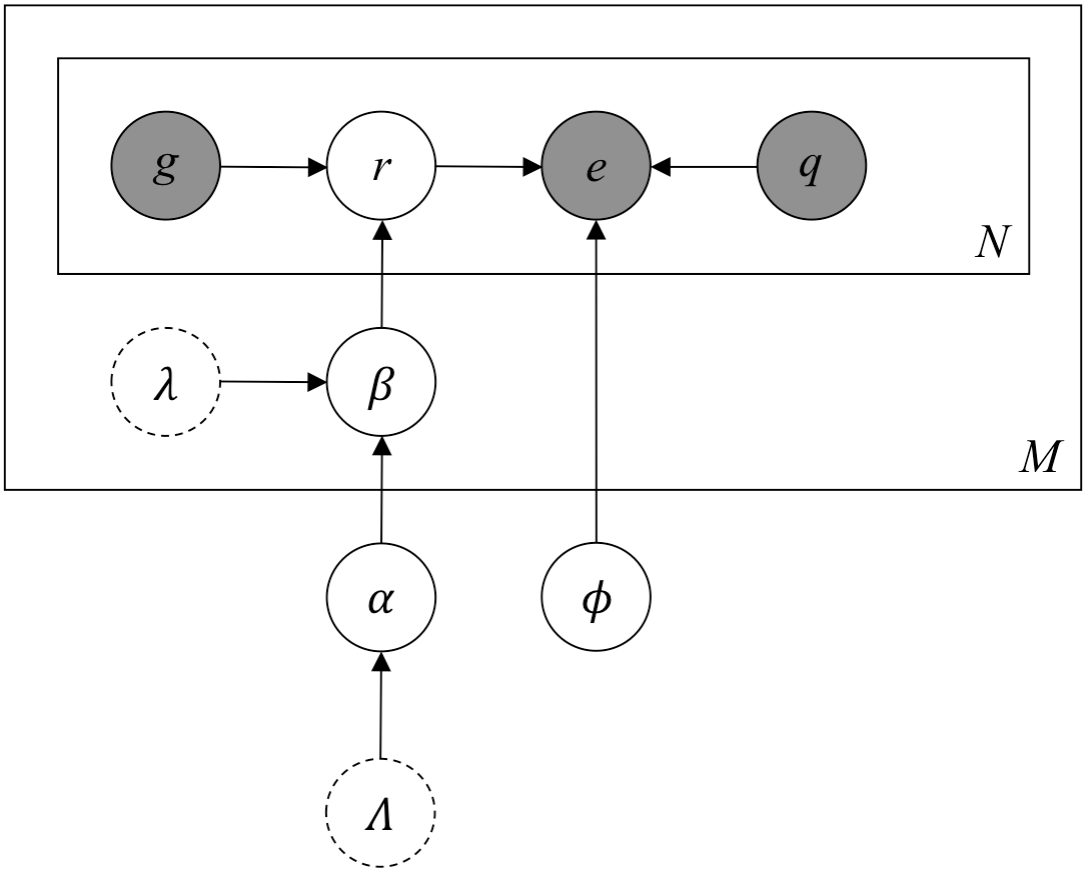

Figure 1. Graphical representation of our model. The outer plate represents tissues, while the inner plate represents individuals and genes within a tissue. Shaded circles represent observed variables; white circles represent hidden variables; dotted edged circles represent hyperparameters.

METHODOLOGY

SPEER is a probabilistic model for inferring the functional consequences of rare variants in M individual tissues. For each tissue c, we have N_c samples, each representing a single individual for a single gene. For each sample i within tissue c, X posits that the presence of a rare regulatory variant r_{ci} can be inferred by integrating measured tissue-specific gene expression e_{ci}, significant common variants q_{ci} nearby sample i, and genomic annotations g_{ci} describing the rare variants nearby sample i, which is a function of both tissue-specific $\{\beta_c, \lambda_c\}$ and shared tissue parameters $\{\alpha, \Lambda\}$. The graphical model is shown in Figure 1.

SPEER infers the presence of a rare regulatory variant nearby a sample by optimizing a joint objective function. The objective has three components: a global component, a tissue-specific component, and a sample-level component.

$$\log p(e,g,r,q,\beta,\lambda,\alpha,\Lambda,\phi) = \underbrace{\log p(\alpha|\Lambda)}_{\text{(A) global component}} + \underbrace{\sum_{c=1}^{M}\left(\sum_{j=1}^{L}\log p(\beta_{cj}|\alpha_j,\lambda_c\right)}_{\text{(B) tissue-specific component}} + \underbrace{\sum_{i=1}^{N_c}\log\sum_{r_{ci}}^{S}p(e_{ci}|r_{ci},q_{ci},\phi)p(r_{ci}|g_{ci},\beta_c)}_{\text{(C) per-sample component}} \quad (1)$$

Per-sample component. Each individual by gene sample is assumed to belong to one of S latent groups (functional variant classes). The random variable $r_{ci} \in \{1,\ldots,S\}$ encodes functional variant class membership. We infer the membership of each sample by integrating genomic annotations, tissue-specific gene expression, and significant common variants. $g_{ci} \in \mathbb{R}^L$ is a vector of L genomic annotations describing the set of rare variants nearby sample i, and $\beta_c \in \mathbb{R}^L$ is a vector of L weights. Formally, we model the effects of g_{ci} on r_{ci} as:

$$r_{ci}|g_{ci},\beta_c \sim Bern(\psi)$$

$$\psi = \frac{1}{1+e^{-\beta_c^T g_{ci}}}$$

We expect functional variants to cause disruption at a cellular level potentially evident by individual molecular phenotypes. Similar to Li et al., we hypothesize that extreme gene expression levels can inform effects of rare variants even at low frequencies. Therefore, we use tissue-specific gene expression outliers denoted by $e_{ci} \in \{0,1\}$, which identifies the outlier status of sample i within tissue c. We compute outliers by evaluating whether the absolute z-score of a sample's gene expression is greater than a predefined threshold. $q_{ci} \in \{0,1\}$ denotes the presence of a significant common variant nearby the gene in sample i. Together we model the effects of r_{ci}, q_{ci}, on e_{ci} as:

$$e_{ci}|r_{ci},q_{ci},\phi \sim NoisyOr(\phi)$$

ϕ controls the rate of functional rare variants to expression outliers and is the same across tissues.

Tissue-specific component. Genomic annotations g_{ci} are assumed to inform both global and tissue-specific effects of genetic variants. For each tissue c, $\beta_c \in \mathbb{R}^L$ is a random variable that deviates from the global effects parameter $\alpha \in \mathbb{R}^L$ with a tissue-specific transfer factor $\lambda_c \in \mathbb{R}$. λ_c is shared across features. For the jth feature, we have:

$$\beta_{cj}|\alpha_j,\lambda_c \sim \mathcal{N}(\alpha_j,\lambda_c^{-1})$$

We exclusively model transferable effects between tissues, not between tissue-specific features. This allows our model to scale well with a large number of annotations.

Global component. The shared tissue level captures global effects across tissues. For the jth feature, the global genomic annotations coefficients $\alpha_j \in \mathbb{R}^L$ is distributed as $\alpha_j|\Lambda \sim \mathcal{N}(\vec{0}, \Lambda^{-1})$.

Learning

We want to learn the parameters of our model $\Theta = \{\beta_{1_1:M_G}, \phi, \alpha\}$ and our hyperparameters $\{\lambda_{1:M}, \Lambda\}$.

We use the empirical Bayes bootstrap estimation procedure described in Efron and Tibshirani (1994) to estimate the transfer factors $\{\lambda_{1:M}, \Lambda\}$. Let $\delta_{j,c}$ = β_j^c - α_j. For i = {1,...,K} randomly sampled with replacement datasets, we compute the maximum likelihood estimation (with regularization) for β_c and α. With these estimates, we compute the empirical variance of $\delta_{j,c}$ across K datasets:

$$\lambda_c{}^{-1} = \frac{\sum_{i=1}^{K}\sum_{j=1}^{L}(\beta_{cj}^{(i)} - \alpha_j^{(i)})^2}{(K-1)L}$$

After estimating our hyperparameters, we compute MAP estimates of Θ by optimizing the log of the joint distribution in Eq. (1) with respect to Θ. Because latent variables make optimization non-convex, we use expectation maximization (EM) to maximize the observed data log likelihood.

Expectation step. We compute the posterior distribution over the set of latent variables r by conditioning on the observed data and our model parameters. Assuming each sample is i.i.d, compute:

$$q_{ci}(r_{ci}) = p(r_{ci} = 1|e_{ci}, g_{ci}, q_{ci}, \beta_c, \lambda_c, \alpha, \Lambda, \phi) = \frac{p(r_{ci} = 1|g_{ci}, \beta_c)p(e_{ci}|r_{ci} = 1, q_{ci}, \phi)}{\sum_{r_{ci}} p(r_{ci}|g_{ci}, \beta_c)p(e_{ci}|r_{ci}, q_{ci}, \phi)} \tag{2}$$

Maximization step. The expectation of the complete data log likelihood with respect to $p(\mathbf{r}|...)$ is:

$$\begin{aligned}\underset{\beta_{1_1:M_L}, \alpha, \phi}{\arg\max}\ \log p(\alpha|\Lambda^{-1}) + \sum_{c=1}^{M}\left(\sum_{j=1}^{L}\log p(\beta_{cj}|\alpha_j, \lambda_c^{-1})\right) + \\ \sum_{i=1}^{N_c}\sum_{z_{ci}} q(r_{ci})\log\left[p(r_{ci}|g_{ci}, \beta_c)p(e_{ci}|r_{ci}, q_{ci}, \phi)\right]\end{aligned} \tag{3}$$

We use blocked coordinate gradient descent to estimate β_c and α, iterating between updating $\alpha_j = \frac{\sum_{c=1}^{M}\lambda_c\beta_{cj}}{\Lambda + \sum_{c=1}^{M}\lambda_c}$ and $\beta_{cj}^{t+1} = \beta_{cj}^t - \nabla f(\beta_{cj}^t, \alpha_j^t, q_{ci}, g_{ci})$, where $\nabla f = \frac{\partial f}{\partial \beta_{cj}^t} =$

$-\lambda_c(\beta_{cj}^t - \alpha_j^t) + \sum_{i=1}^{N_c} -g_{cij}(q_{ci}(r_{ci}) - h(\beta_c, g_{ci}))$ where h is the inverse logit function. ϕ is updated using a NoisyOR MAP estimation procedure with soft assignments to $\mathbf{r}$ as weights.

RESULTS

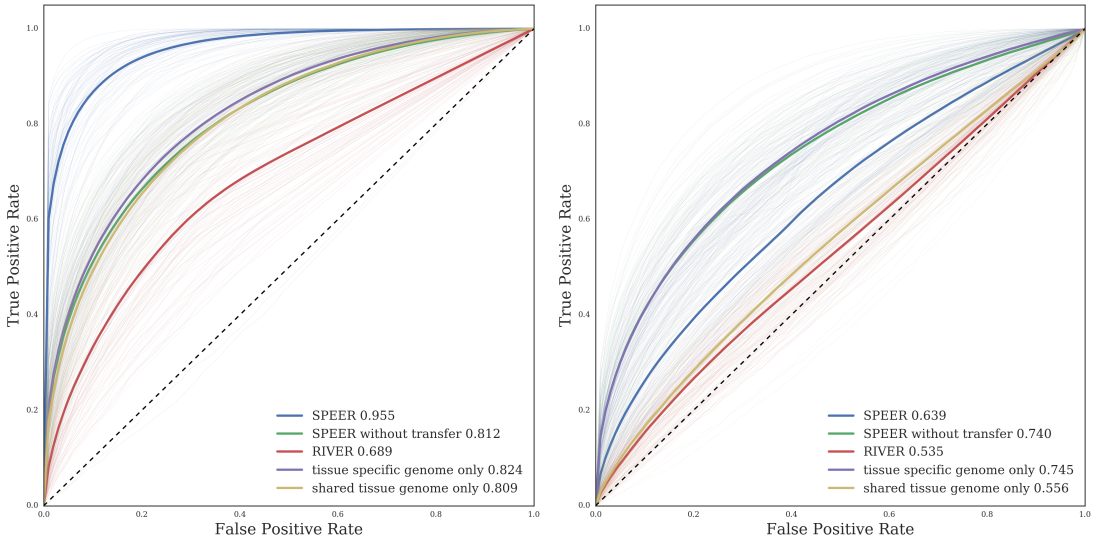

Figure 2. Tissue-specific receiver operating characteristic (ROC) curve averaged across five tissue groups using the stronger effects parameter setting evaluated in the tied tissue (A) and independent tissue (B) simulations. The darker lines represent average ROC curves across 75 simulated runs. Area under the curve (AUC) scores are reported in the legend. There are four benchmarks described here: SPEER w/o transfer was trained on the same data as SPEER but assumes parameter independence; RIVER integrates genomic annotations with shared tissue expression outlier status in an unsupervised setting; shared tissue genome only is a supervised model trained on exclusively genomic annotations using shared tissue expression outlier status as labels; tissue-specific genome only is also a supervised model trained on exclusively genomic annotations using tissue-specific expression outlier status as labels.

Simulation Results.

To highlight the intuition behind SPEER, we performed two simulations: one involving tied tissues and the other involving independent tissues. The tied tissue simulation used transfer learning to generate data. The independent tissue simulation generated data for each tissue independently. Because none of the other approaches considered here include common variants, we excluded q in order to evaluate the usefulness of tissue sharing in simulation. Therefore, tissue-specific gene expression is only conditioned on r, so we used a categorical distribution with parameter ϕ to model this dependency and used a Beta prior on ϕ with hyperparameters $\mu_{r_{ci}}$ and $\sigma^2_{r_{ci}}$ to generate the rate of functional variants to expression outliers. Formally, for sample i within tissue c we have:

$$e_{ci}|r_{ci}, \phi \sim Cat(\phi)$$

$$\phi_{r_{ci}} \sim Beta(\mu_{r_{ci}}, \sigma^2_{r_{ci}})$$

We re-parameterized the Beta distribution using a mean and variance (Ferrari and Cribari-Neto, 2004) to allow for better interpretability of the parameter settings described in our simulation. The simulations were crafted to mimic scenarios with strong effects from genomic annotations coupled with noisy gene expression data. Besides the caveat described above, the simulated data for the tied tissue setting was generated by sampling from the joint distribution assumed by SPEER, as described in Eq. (1). The independent tissue setting followed a similar procedure except each β_{cj} was sampled independently from $\mathcal{N}(0, \lambda_c)$. Tables 1 and 2 describe three scenarios that were tested.

Table 1. Tied tissue simulation.

Parameter	stronger effects	equal effects	weaker effects
Λ	0.01	0.01	0.1
λ_c	$\{2, \ldots, 6\}$	$\{2, \ldots, 6\}$	$\{2, \ldots, 6\}$
$\phi_{e\|z=0} \sim Beta(\mu, \sigma^2)$	(0.4, 1e-4)	(0.3, 1e-4)	(0.4, 1e-4)
$\phi_{e\|z=1} \sim Beta(\mu, \sigma^2)$	(0.6, 1e-4)	(0.7, 1e-4)	(0.6, 1e-4)

Table 2. Independent tissue simulation.

Parameter	stronger effects	equal effects	weaker effects
λ_c	0.01	0.01	0.1
$\phi_{e\|z=0} \sim Beta(\mu, \sigma^2)$	(0.4, 1e-4)	(0.3, 1e-4)	(0.4, 1e-4)
$\phi_{e\|z=1} \sim Beta(\mu, \sigma^2)$	(0.6, 1e-4)	(0.7, 1e-4)	(0.6, 1e-4)

For each setting, we measured the simulation uncertainty by performing each experiment 75 times. The stronger effects scenario underlined strong influence of genomic annotations coupled with noisy expression labels. The tied tissue simulation (Table 1) highlighted genomic annotations with strong functional effects combined with correlated influences across tissues. The independent tissue simulation showed similarly strong functional consequences from genomic annotations but independent influences across tissues. SPEER performed significantly better than all baselines at predicting held-out tissue-specific labels in the tied simulation (Fig. 2A). Even with limited training data, SPEER provided a significant performance boost (Fig. 5). In the independent tissue simulation, SPEER performed worse than the other two tissue-specific models—SPEER without transfer and tissue specific genome only (Fig. 2B). In this simulation, the data generation process was independent for each tissue, so encouraging tissue similarity would rightly hurt performance.

The equal effects scenario mimicked strong influence of genomic annotations similar to the previous simulation but with highly predictive gene expression labels. We observed a significant boost in AUC scores across all benchmarks when using expression data that is more predictive of regulatory status (Fig. 6A). SPEER scores remained highly predictive of the regulatory status of rare variants when switching to expression data with a stronger signal (AUC of 0.988 vs 0.955). We observed significant performance boosts in all other models using this parameter setting, implying that highly

predictive expression data might be critical to the performance of these other models.

The weaker effects scenario highlights weaker influence of genomic annotations. In the tied case, we observed a lower AUC score for SPEER compared to the stronger effects scenario (Fig. 7A). Despite a lower AUC score, the predictive performance of SPEER remains significantly better than all other models. In the independent tissue simulation (Fig. 7B), we predictably observed SPEER without transfer performing better than SPEER. Given the weaker influence of genomic annotations, the general performance across all models is worse.

Results from GTEx data.

We applied our method to data from the Genotype-Tissue Expression (GTEx V6p) project. We included whole genome sequence data from 113 donors with European ancestry and 5574 RNA-sequence samples from 27 tissues. We defined a rare variant using a minor allele frequency (MAF) below 1% within the GTEx cohort and within the European panel of the 1000 Genomes project (Consortium, 2015). We restricted our analysis to rare single nucleotide variants (SNVs), which are polymorphisms occurring at specific positions in the genome. We generated a set of genomic features describing each rare SNV. This included describing the location of the rare variant with respect to regulatory elements, the conservation status, and summary statistics from genome only variant predictor tools including CADD and DANN. We also separately generated a set of binary tissue-specific annotations that described whether each rare SNV was present in any of the cell-type specific promoter or enhancer regions from ROADMAP Epigenomics and ENCODE projects (Consortium, 2012; Kundaje et al., 2015) using summary statistics from ChromImpute developed by Ernst and Kellis (2015). We then mapped these annotations to one of the 27 GTEx tissues considered here. We then aggregated all rare SNVs within 10 kb of the transcription start site (TSS) to generate gene-level summary statistics by computing the maximum of each annotation across all nearby rare SNVs. Next, we removed technical and environmental confounders from each tissue's gene expression using PEER estimates (Stegle et al., 2012). We then computed gene expression outliers using the z-score across all subjects and genes for each tissue. We refer interested readers to Li et al. (2016) for a complete description of the genomic annotations used, the processing of RNA-expression data and the subsequent gene expression outlier calls. Finally, we identified the top significant common variant nearby each gene using the methods described in Aguet et al. (2016) and used this data to denote the presence of a significant common variant for each sample.

We measured the sensitivity of our results to the threshold used to call tissue-specific expression outliers in the supplement (Fig. 8). The remaining results used a 1.5 z-score threshold. Because single tissue gene expression outliers are too noisy, we identified clusters of tissues that shared similar patterns of gene expression. We used five tissue groups—brain, digestive, epithelial, artery and fats together, and muscles—as input to our model. We used prior experiments to choose tissue groups by evaluating the pairwise-similarity between individual tissues. A list of tissues in each tissue group is available in the supplement.

Allele-specific expression is known to present strong evidence of a causal cis-regulatory effect, which often arises from a non-coding variant (Zhang et al., 2009; Yan et al., 2002). Because the majority of the rare variants in GTEx are non-coding and heterozygous, measuring tissue-specific allele-specific expression allowed us to evaluate SPEER at prioritizing functional variants. We measured allelic imbalance as a function of reference and alternate allele expression read counts, which is computed using an allelic ratio = $|\frac{ref}{ref-alt}| - 0.5$. Higher values here imply greater allelic imbalance.

We computed the statistical association between SPEER's predictions and measured tissue-specific allelic imbalance for all genes in each tissue using Fisher's Exact Test and observed significantly greater predictive power using SPEER compared to all benchmarks (Fig. 3A). We also measured the effect for each tissue individually (Fig. 9). In addition, we investigated the SPEER posteriors for samples with strong allelic imbalance (defined using 90 percentile cut-off) and limited to at least one model having a posterior greater than 0.5 (Fig. 3B). Among samples with observed allelic imbalance, SPEER identified 120 samples with all predictions greater than 0.85. Genome only tissue-specific model identified 3 samples; and the shared tissue genome only model identified 2 samples.

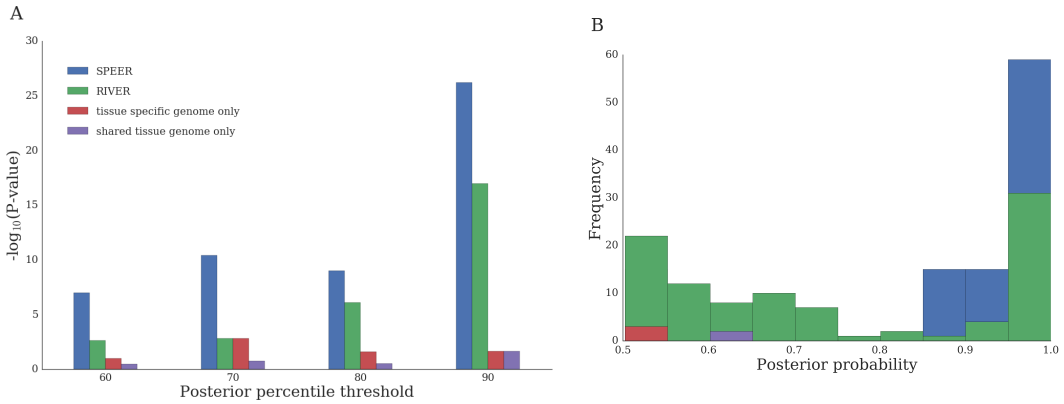

Figure 3. A) Using SPEER scores to predict tissue-specific allelic imbalance. Allelic imbalance was defined by the 90th percentile of allelic ratios. A deleterious SPEER score was defined using four percentile thresholds. We computed p-values for each of the four settings using Fisher's exact test and compared our results to two benchmarks. B) Histogram of SPEER scores for samples with allelic imbalance limited to samples with at least one of the four models having a posterior greater than 0.5.

Comparing SPEER to RIVER.

SPEER is a probabilistic model that uses transfer learning to infer the *tissue-specific* regulatory impact of each rare SNV. RIVER is a general method to infer the *global* regulatory impact of each rare SNV across diverse tissues. SPEER integrates genomic annotations with tissue-specific expression labels across M tissues. RIVER integrates genomic annotations with a shared tissue expression label. We compared the two

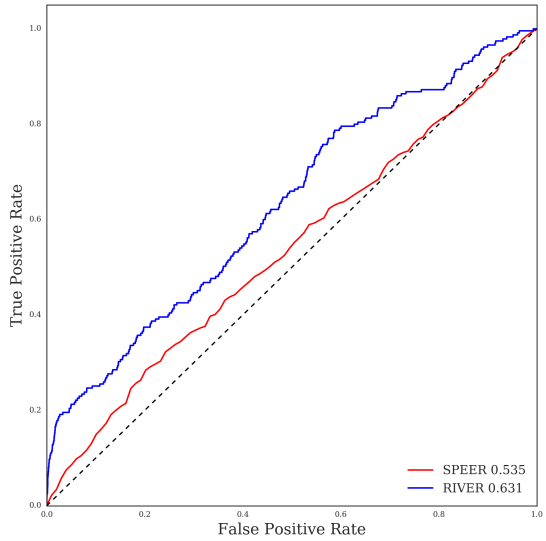

Figure 4. ROC curve comparing predictive performance of SPEER to RIVER on their respective tasks (predicting tissue-specific expression labels versus shared tissue expression labels respectively) using held-out pairs of individuals with identical variants nearby a specific gene.

methods at their respective tasks, predicting tissue-specific held-out expression labels and shared tissue held-out expression labels (Fig. 4). For evaluation, we followed a similar approach to Li et al. (2016) by holding out pairs of individuals that share the same rare variants nearby a specific gene. After training SPEER and RIVER on the remaining data, we computed SPEER and RIVER scores for the first individual and compared these scores to the held-out expression labels for the second individual. We observed significant performance boosts at predicting held-out shared tissue expression labels using RIVER compared to held-out tissue-specific expression labels using SPEER. These results show that tissue-specific expression labels are noisier and simply harder to predict. We investigated this further by computing the correlation between the gene expression labels across all pairs of individuals with the same rare variants. We observed a 5x increase in correlation when using shared tissue expression labels (Kendall's tau rank correlation, $\rho = 0.144$, p-value $<$ 1.33-124) instead of tissue-specific expression labels ($\rho = 0.033$, p-value $<$ 3.27e-12).

CONCLUSION.

Rare variant prediction is an important problem for understanding the heritability of a large number of diseases. Understanding the functional consequences of these variants is a critical hurdle in our efforts towards personalized genomics. Because most diseases are known to have tissue-specific molecular consequences, the development of variant prediction tools that use tissue and cell-type specific context remain essential. Here we have developed a probabilistic model that provides tissue-specific functional predictions for rare variants. Our method shares information across tissues in order to make reliable predictions.

Using our method, we observe significant performance boosts in predicting tissue-specific allele-specific expression compared to the state-of-the-art, including genome only prediction tools such as CADD and VEP and integrative methods like RIVER. We also highlight the model's predictive power using simulated data. The simulation highlights SPEER's particular usefulness with low resources across diverse tissues.

A future direction for this work is to leverage the information sharing across tissues in order to make single tissue functional predictions. This will be a necessary step forward given the large number of datasets with limited resources. However, predicting the molecular consequences in single tissues remains a difficult problem for learning reliable parameters and evaluating model performance due to noisy transcriptomic reads.

The primary application of SPEER described here involves the use of tissue-specific gene expression. However, this method may also be useful for predicting alternative splicing using isoform ratios or allelic imbalance using allele-specific expression by direct integration of these data sources.

Data availability The GTEx V6 release genotypes and allele-specific expression data are available on dbGaP (study accession phs000424.v6.p1; http://www.ncbi.nlm.nih.gov/projects/gap/cgi-574bin/study.cgi?study_id=phs000424.v6.p1). GTEx V6p release expression data is available on the GTEx portal (http://www.gtexportal.org

REFERENCES

Aguet, F., Brown, A. A., Castel, S., Davis, J. R., Mohammadi, P., Segre, A. V., Zappala, Z., Abell, N. S., Fresard, L., Gamazon, E. R., Gelfand, E., Gloudemans, M. J., He, Y., Hormozdiari, F., Li, X., Li, X., Liu, B., Garrido-Martin, D., Ongen, H., Palowitch, J. J., Park, Y., Peterson, C. B., Quon, G., Ripke, S., Shabalin, A. A., Shimko, T. C., Strober, B. J., Sullivan, T. J., Teran, N. A., Tsang, E. K., Zhang, H., Zhou, Y.-H., Battle, A., Bustamonte, C. D., Cox, N. J., Engelhardt, B. E., Eskin, E., Getz, G., Kellis, M., Li, G., MacArthur, D. G., Nobel, A. B., Sabbati, C., Wen, X., Wright, F. A., Lappalainen, T., Ardlie, K. G., Dermitzakis, E. T., Brown, C. D., and Montgomery, S. B. (2016). Local genetic effects on gene expression across 44 human tissues. *bioRxiv*.

Backenroth, D., Kiryluk, K., Xu, B., Pethukova, L., Vardarajan, B., Khurana, E., Christiano, A., Buxbaum, J., and Ionita-Laza, I. (2016). Tissue-specific functional effect prediction of genetic variation and applications to complex trait genetics. *bioRxiv*.

Consortium, . G. P. (2015). A global reference for human genetic variation. *Nature*, 526(7571):68–74.

Consortium, E. P. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414):57–74.

Efron, B. and Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.

Ernst, J. and Kellis, M. (2015). Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nat Biotech*, 33(4):364–376.

Ferrari, S. and Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31(7):799–815.

Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nature genetics*, 46(3):310–315.

Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., and Ziller, M. J. (2015). Integrative analysis of 111 reference human epigenomes. *Nature*, 518(7539):317–330.

Li, X., Kim, Y., Tsang, E. K., Davis, J. R., Damani, F. N., Chiang, C., Zappala, Z., Strober, B. J., Scott, A. J., Ganna, A., Merker, J., Hall, I. M., Battle, A., and Montgomery, S. B. (2016). The impact of rare variation on gene expression across tissues. *bioRxiv*.

McCallum, A. K., Rosenfeld, R., Mitchell, T. M., and Ng, A. Y. (1998). Improving text classification by shrinkage in a hierarchy of classes. *International Conference on Machine Learning (ICML)*, pages 359–367.

Nelson, M. R., Wegmann, D., Ehm, M. G., Kessner, D., St. Jean, P., Verzilli, C., Shen, J., Tang, Z., Bacanu, S.-A., Fraser, D., Warren, L., Aponte, J., Zawistowski, M., Liu, X., Zhang, H., Zhang, Y., Li, J., Li, Y., Li, L., Woollard, P., Topp, S., Hall, M. D., Nangle, K., Wang, J., Abecasis, G., Cardon, L. R., Zöllner, S., Whittaker, J. C., Chissoe, S. L., Novembre, J., and Mooser, V. (2012). An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science (New York, N.Y.)*, 337(6090):100–104.

Quang, D., Chen, Y., and Xie, X. (2015). DANN: A deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics*, 31(5):761–763.

Raina, R., Ng, A., and Koller, D. (2006). Constructing informative priors using transfer learning. *Proceedings of the 23rd international conference on Machine learning*, page 713–720.

Stegle, O., Parts, L., Piipari, M., Winn, J., and Durbin, R. (2012). Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nature protocols*, 7(3):500–7.

Tennessen, J. A., Bigham, A. W., O'Connor, T. D., Fu, W., Kenny, E. E., Gravel, S., McGee, S., Do, R., Liu, X., Jun, G., Kang, H. M., Jordan, D., Leal, S. M., Gabriel, S., Rieder, M. J., Abecasis, G., Altshuler, D., Nickerson, D. A., Boerwinkle, E., Sunyaev, S., Bustamante, C. D., Bamshad, M. J., Akey, J. M., GO, B., GO, S., and Project, o. b. o. t. N. E. S. (2012). Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes. *Science*, 337(6090):64–69.

Thrun, S. (1996). Is learning the n-th thing any easier than learning the first? *Advances in neural information processing systems*, pages 640–646.

Uricchio, L. H., Zaitlen, N. A., Ye, C. J., Witte, J. S., and Hernandez, R. D. (2016). Selection and explosive growth alter genetic architecture and hamper the detection of causal rare variants. *Genome Research*, 26(7):863–873.

Yan, H., Yuan, W., Velculescu, V. E., Vogelstein, B., and Kinzler, K. W. (2002). Allelic variation in human gene expression. *Science*, 297(5584):1143.

Zhang, K., Li, J. B., Gao, Y., Egli, D., Xie, B., Deng, J., Li, Z., Lee, J.-H., Aach, J., and Leproust, E. M. (2009). Digital RNA allelotyping reveals tissue-specific and allele-specific gene expression in human. *Nature methods*, 6(8):613–618.

SUPPLEMENT.

Tissue groups. We evaluated the pairwise similarity between gene expression patterns across tissues and identified the following list of tissue groups used in the GTEx results section. GTEx ids are listed below:

Brain Brain Caudate basal ganglia, Brain Nucleus accumbens basal ganglia, Brain Putamen basal ganglia, Brain Anterior cingulate cortex BA24, Brain Cortex, Brain Frontal Cortex BA9

Artery and Fat Artery Coronary, Artery Aorta, Artery Tibial, Esophagus Muscularis, Esophagus Gastroesophageal Junction, Colon Sigmoid, Adipose Subcutaneous, Adipose Visceral Omentum, Breast Mammary Tissue

Muscle Muscle Skeletal, Heart Atrial Appendage, Heart Left Ventricle

Epithelial Skin Not Sun Exposed Suprapubic, Skin Sun Exposed Lower leg, Esophagus Mucosa, Vagina

Digestive Stomach, Colon Transverse, Lung, Thyroid, Prostate

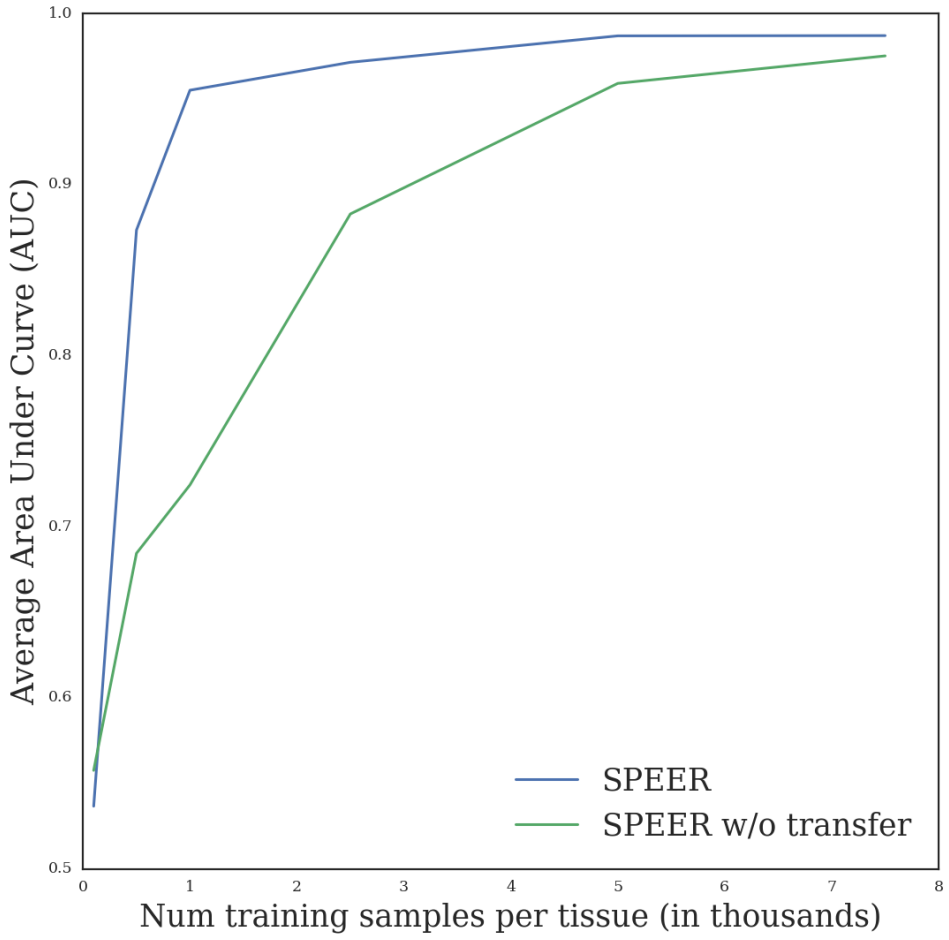

Figure 5. Area under curve (AUC) averaged across five tissue groups for different number of training samples using simulated data.

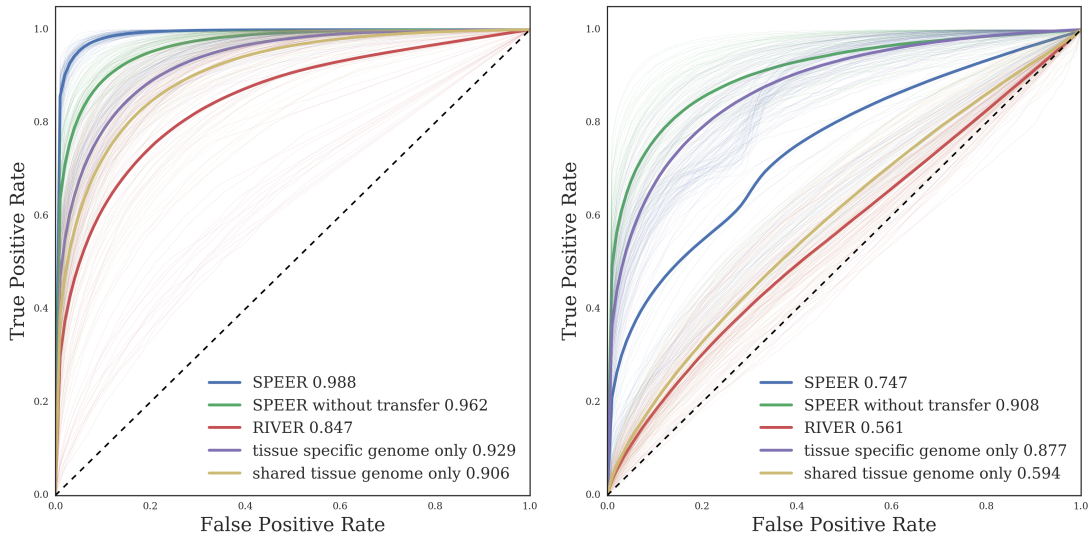

Figure 6. ROC curves for equal effects setting comparing SPEER to four benchmarks in the tied tissue simulation (left) and the independent tissue simulation (right).

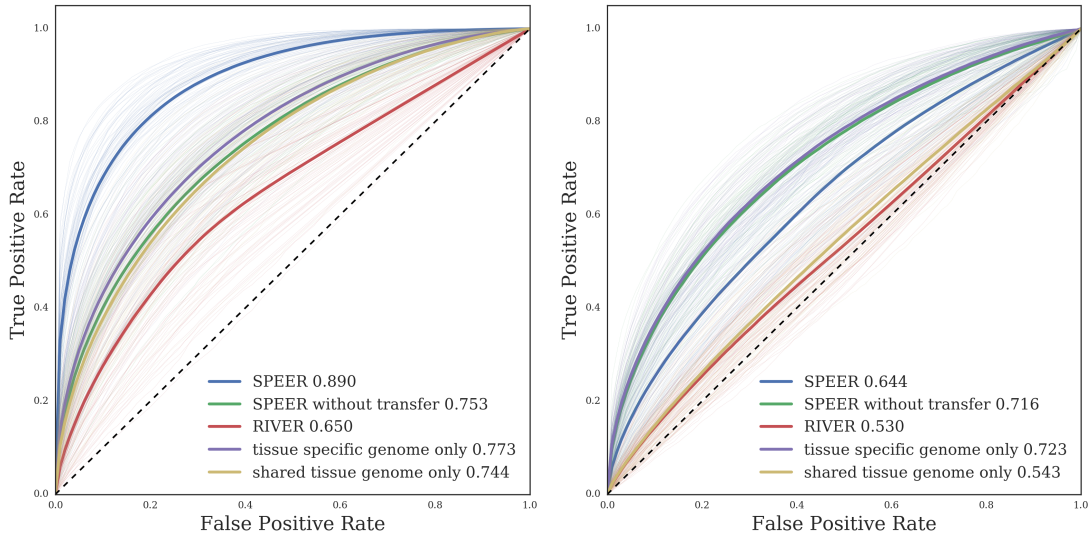

Figure 7. ROC curves for weaker effects setting comparing SPEER to four benchmarks in the tied tissue simulation (left) and the independent tissue simulation (right).

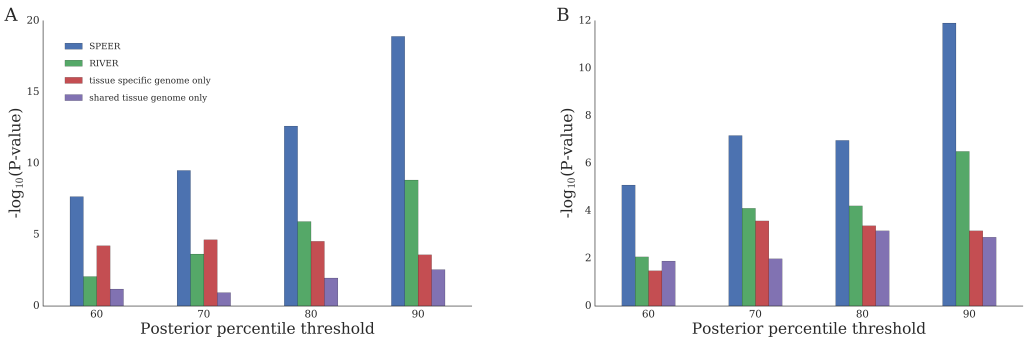

Figure 8. SPEER scores compared to tissue-specific allelic imbalance using z-score expression outlier thresholds of 1.75 (left) and 2.0 (right) in GTEx data.

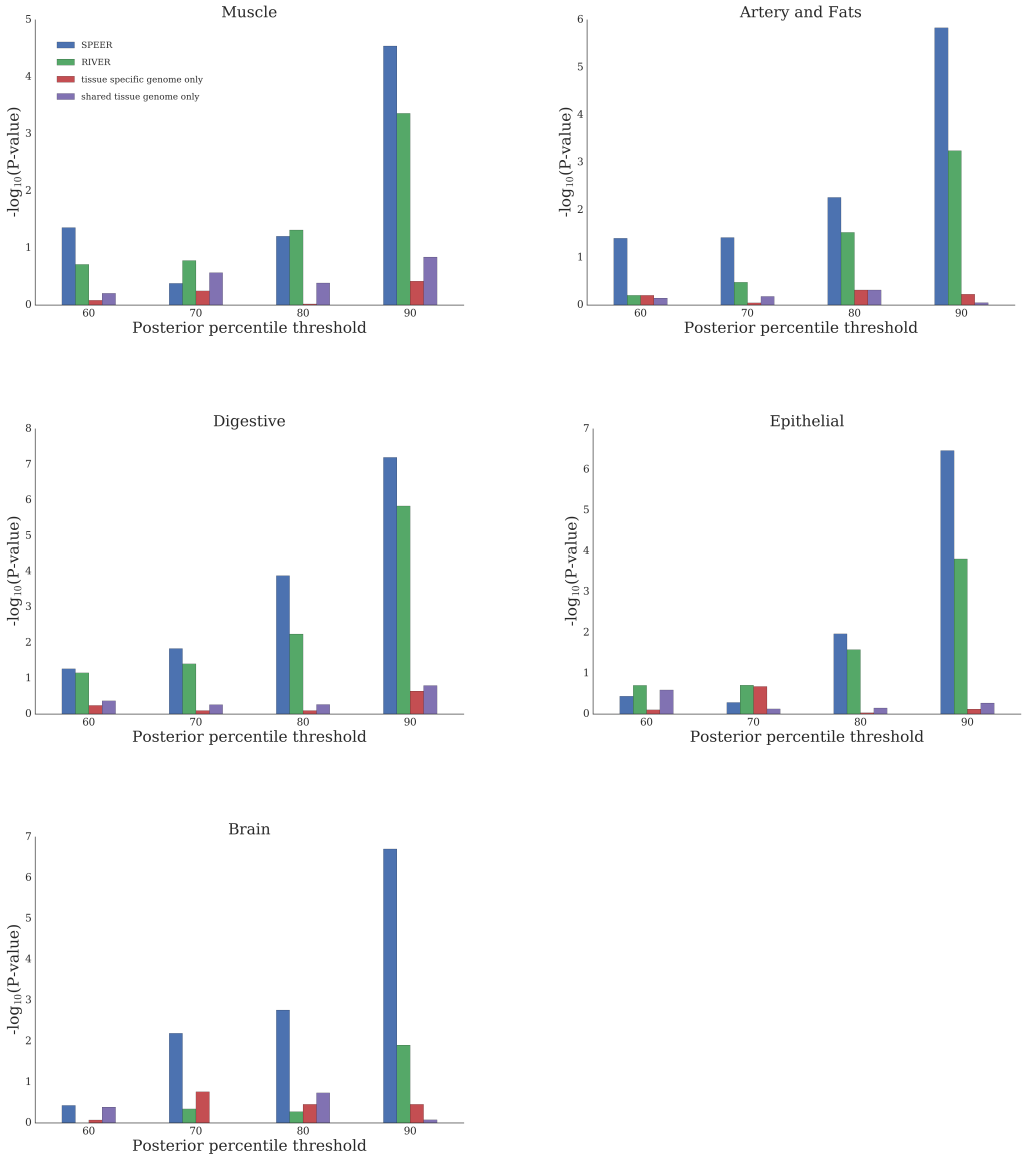

Figure 9. SPEER scores compared to allelic imbalance in the five tissue groups using GTEx data.