Supplementary Materials

1 GTEx Data Preprocessing and Analysis

Tissue	Obs.	Genes (<10% zero)	SNPs	Pairs (<10% zero)
Muscle - Skeletal	361	23,948	9,991,147	166,070,588
Whole Blood	338	23,973	9,878,498	164,903,296
Lung	278	27,974	9,036,719	176,028,082
Thyroid	278	27,735	9,173,566	176,881,830

Table S1: Description of the GTEx data in four tissues.

The data used for the analyses described in this manuscript are the GTEx v6 data obtained from dbGaP with accession number phs000424.v6.p1. In total there are 449 individuals with genotype information. We focus on 4 tissues with large sample sizes. The gene expression data and the genotype data are preprocessed in the same way as in The GTEx Consortium (2015). In addition, we remove genes with more than 10% zero read count, as in such case the Gaussian assumption in linear regression is violated and our preliminary analyses also found that the existing variance eQTL method CLS had largely inflated type I error. The preprocessing details are as follows:

For the RNA sequencing data in each tissue, the genes with RPKM values greater or equal to 0.1 in less than 10 individuals are removed. In addition, the genes with more than 10% of observations have zero RPKM values are also removed. The expression data of each remaining gene are quantile normalized to have a standard normal distribution across samples. The genotype data for all the individuals are first imputed using 1000 Genomes Project Phase I, version 3. Then the data are separated into several data files, each corresponding to the samples in one tissue. The SNPs with minor allele frequency smaller than 1% are removed in each tissue. As a result, for each tissue, we obtain a normalized gene expression data matrix and an imputed genotype data matrix on the matched samples.

To correct for the confounding effect, we obtain the top 3 principal components from the genotype data, 35 PEER factors from the normalized expression data, and sex and platform index as covariates. In total, we have 40 covariates in each tissue. In particular, the principal components of the genotype data capture the ancestry information of the individuals. The covariates serve as confounding factors to be taken into account in the eQTL analysis.

We apply different methods for eQTL analysis in each tissue separately. In particular, we focus on the *cis* gene-SNP pairs where the SNP locates within 1 megabase of the transcription starting site of the gene. As a result, there are about 170 million gene-SNP pairs in each tissue. Each method is applied to one gene-SNP pair in one tissue at a time to get a p-value. The p-values from all gene-SNP pairs are further converted to the q-values (Storey and Tibshirani, 2003) to control the false discovery rate (FDR).

We remark that the proposed Q-rank method does not have any distributional assumption on the gene expression data. It can be easily applied to the RPKM data or the raw read counts. However, the competing methods such as the linear regression method rely on the normal assumption and the CLS method has inflated type I errors under the existence of excess zeros in gene expressions. Thus, for fair comparison, we apply different methods to the same preprocessed data to get the p-values as described above.

2 Venn diagrams of SNP-gene pairs

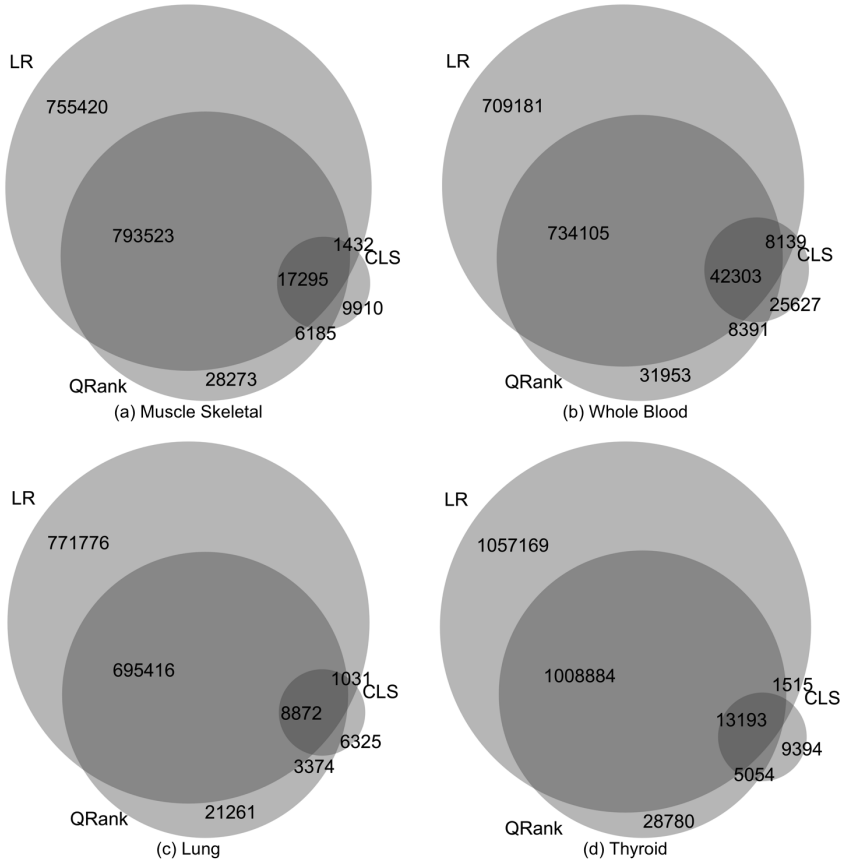

Figure S1: Venn diagrams depicting overlap among SNP-gene pairs identified by LR, CLS and QRank controlling FDR at $= 0.05$

3 Pairwise SNP-gene pairs sharing

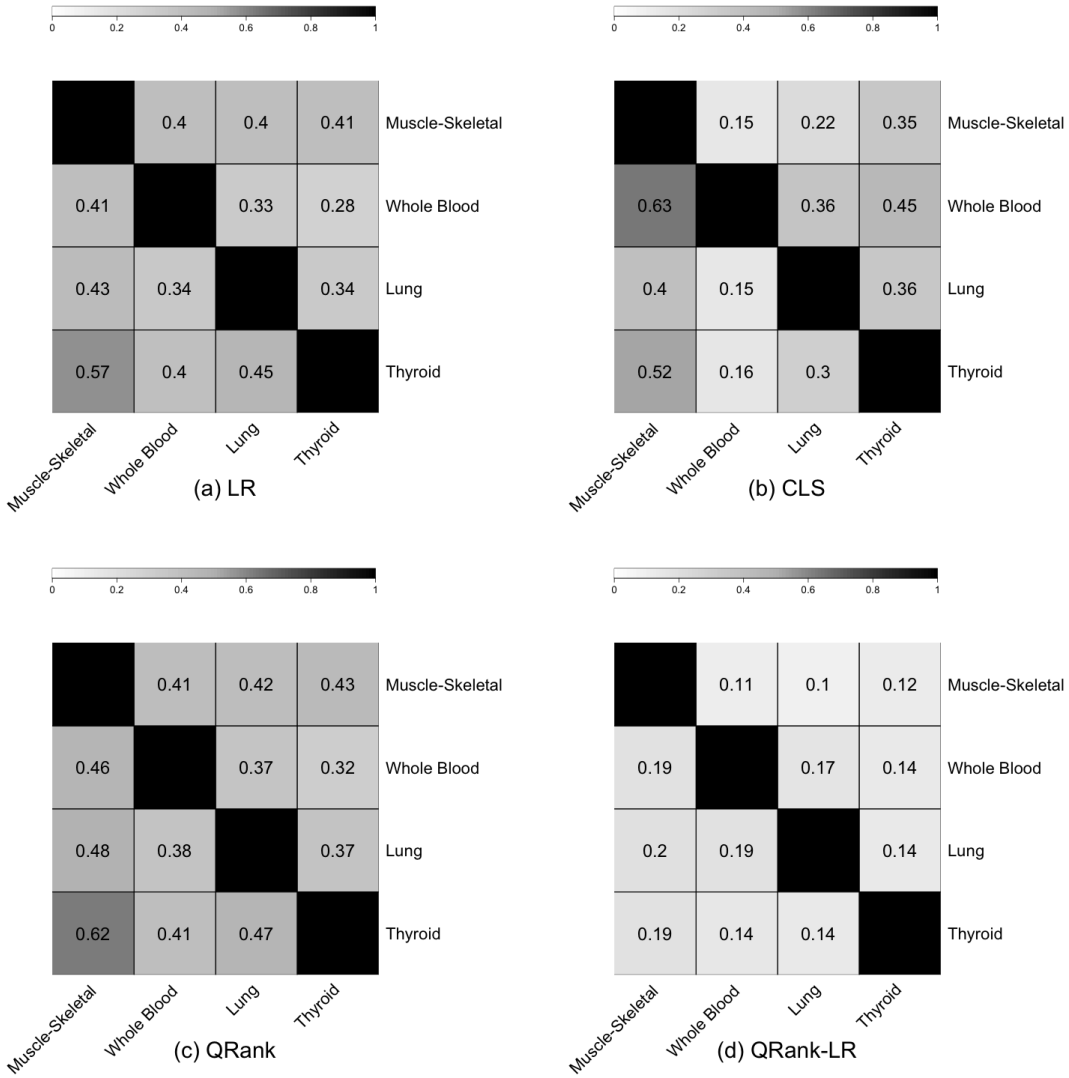

Figure S2: Cross-tissue sharing of SNP-gene pairs. The entry in row i and column j is an estimate of τ_{ij} =Pr(eQTL in tissue i | eQTL in tissue j).

4 GWAS enrichment of identified eQTLs and validation of QRank-LR tissue-specific eQTLs

FDR		LR	CLS	QRank	QRank-LR	LR tissue specific	Qrank-LR tissue specific
	No. of eQTLs	2,008,166	82,943	1,126,388	114,867	1,708,487	98,291
5E-2	No. in GWAS	5,959	452	3,868	357	5,027	264
	RR	Ref	1.84	1.16	1.05	0.99	0.91
	p-value	Ref	$<$ 2.2E-16	1.3E-12	3.9E-1	6.6E-1	1.1E-1
	95% CI	Ref	(1.67,2.02)	(1.11, 1.20)	(0.94, 1.17)	(0.96, 1.03)	(0.80, 1.02)
	No. of eQTLs	1,391,053	49,350	853,689	47,598	1,121,689	35,772
1E-2	No. in GWAS	4,611	311	3,154	213	3,787	140
	RR	Ref	1.90	1.11	1.35	1.02	1.18
	p-value	Ref	$<$ 2.2E-16	2.5E-6	1.7E-5	4.0E-1	5.2E-2
	95% CI	Ref	(1.67, 2.13)	(1.07, 1.17)	(1.18 ,1.55)	(0.98, 1.06)	(1.00, 1.40)
	No. of eQTLs	1,016,882	31,823	649,850	23,927	792,215	16,536
1E-3	No. in GWAS	3,651	199	2,536	125	2,947	80
	RR	Ref	1.74	1.09	1.46	1.04	1.35
	p-value	Ref	1.0E-14	1.2E-3	3.3E-5	1.5E-1	8.0E-3
	95% CI	Ref	(1.51, 2.01)	(1.03, 1.14)	(1.22, 1.74)	(0.99, 1.09)	(1.08, 1.68)
	No. of eQTLs	804,881	22,893	519,992	14,727	617,103	9,682
1E-4	No. in GWAS	3,018	147	2,068	89	2,445	54
	RR	Ref	1.71	1.06	1.61	1.06	1.49
	p-value	Ref	1.1E-10	3.9E-2	7.2E-6	4.2E-2	3.5E-3
	95% CI	Ref	(1.45, 2.02)	(1.00, 1.12)	(1.31, 1.99)	(1.00, 1.11)	(1.14, 1.95)
	No. of eQTLs	661,469	17012	426,322	10,740	501,919	7,446
1E-5	No. in GWAS	2,560	113	1,729	72	2,021	51
	RR	Ref	1.72	1.05	1.73	1.04	1.77
	p-value	Ref	1.2E-8	1.3E-1	3.1E-6	1.8E-1	4.1E-5
	95% CI	Ref	(1.42, 2.07)	(0.99, 1.11)	(1.37, 2.19)	(0.98, 1.10)	(1.34, 2.33)

Table S2: The enrichment of identified eQTLs in GWAS catalog at FDR =$(0.05, 10^{-5})$ in four tissues.

5 Type I error of QRank by the number and location of τ's

	Nominal p-value	$\tau = 0.5$	$\tau = (0.25, 0.5, 0.75)$	$\tau = (0.15, 0.25, 0.5, 0.75, 0.85)$
No covariates	5E-02	4.9E-02	4.8E-02	4.6E-02
	1E-02	9.3E-03	9.0E-03	9.1E-03
	1E-03	8.2E-04	8.4E-04	9.6E-04
	1E-04	7.9E-05	7.6E-05	1.1E-04
40 covariates	5E-02	5.6E-02	4.3E-02	3.0E-02
	1E-02	1.2E-02	7.9E-03	4.8E-03
	1E-03	1.2E-03	6.9E-04	4.0E-04
	1E-04	1.1E-04	6.9E-05	3.0E-05

Table S3: Type I error of QRank by the number and location of τ's. When there are no covaraites, the type I error remains constant regardless of the number of τ's. When a large number of nuisance covariates exist, the more quantiles we combine, the more conservative QRank becomes. A recommended number of quantile levels for QRank is 3-5.

References

Storey, J. D. and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, **100**(16), 9440–9445.

The GTEx Consortium (2015). The genotype-tissue expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science*, **348**(6235), 648–660.