

Reproducibility in Identified eQTLs (SNV-gene pairs)

Protocol

- 1) Filter SNVs from GTEx (remove SNVs w/ MAF<5%, imputation score < 0.4, etc)
- 2) GTEx reports ~7 million SNVs (the deviation from our SNV dataset is ~ 1.5%)
- 3) Select a random sample of 100000 SNVs from each of the tissue lists
- 4) Use matrix-eQTL to identify eQTLs within each tissue using a linear regression model
- 5) Remove all eQTLs with and FDR exceeding 5% (per GTEx pilot Supp)
- 6) Determine which fraction of in-house identified eQTLs match eQTLs provided by GTEx

Results

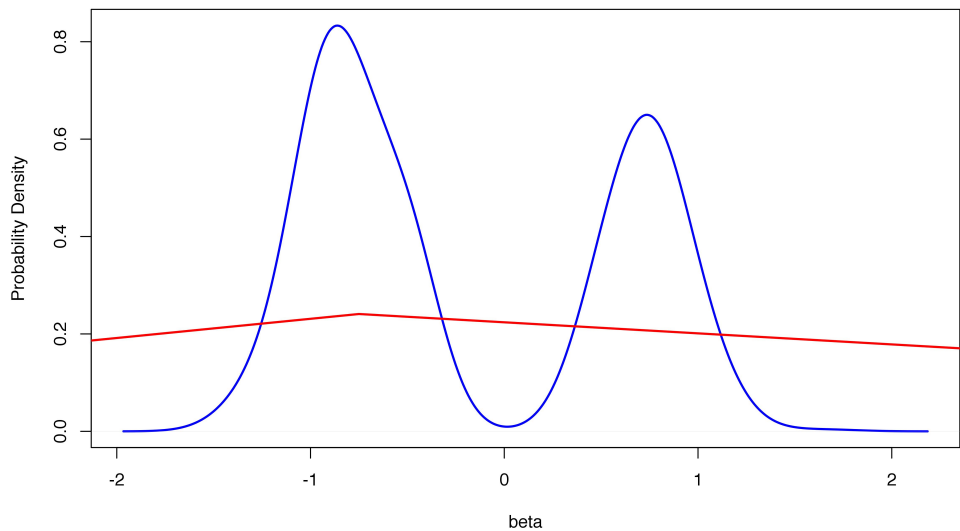
<u>brain_region</u>	<u>fract_matching_eQTLs</u>
Caudate_basal_ganglia	0.918580376
Cerebellar_Hemisphere	0.922613065
Cerebellum	0.938430311
Anterior_cingulate_cortex_BA24	0.708029197
Cortex	0.888831967
Frontal_Cortex_BA9	0.922551253
Hippocampus	0.970394737
Hypothalamus	0.859872611
Nucleus_accumbens_basal_ganglia	0.855907781
Putamen_basal_ganglia	0.898044693

Brain_Frontal_Cortex

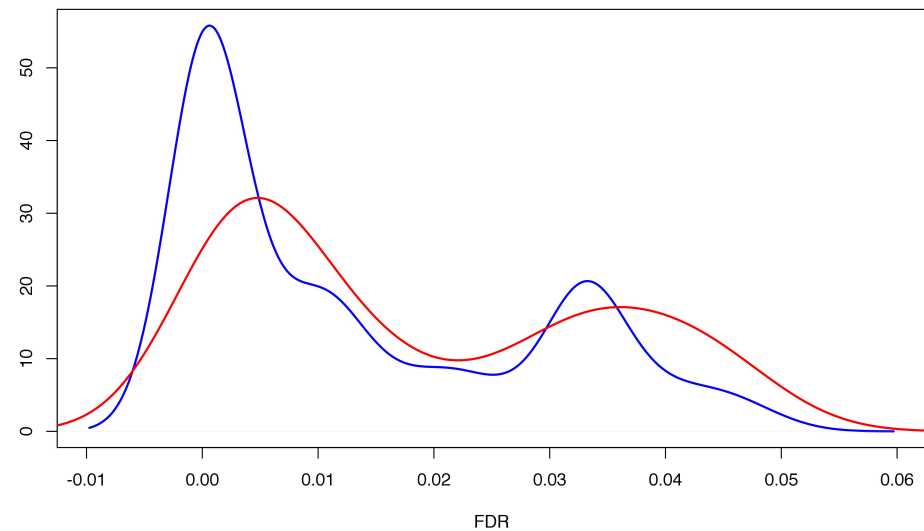
— True positives

— False positives

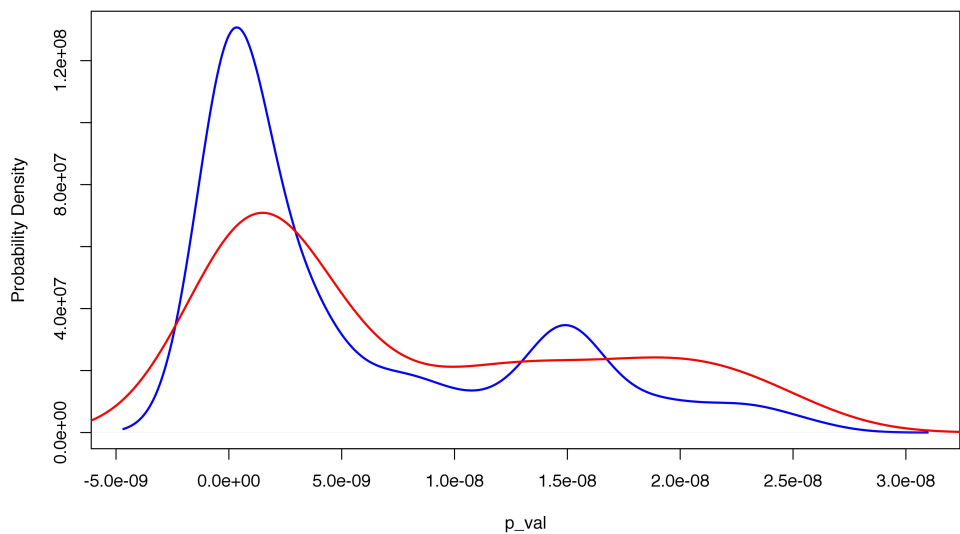
beta values



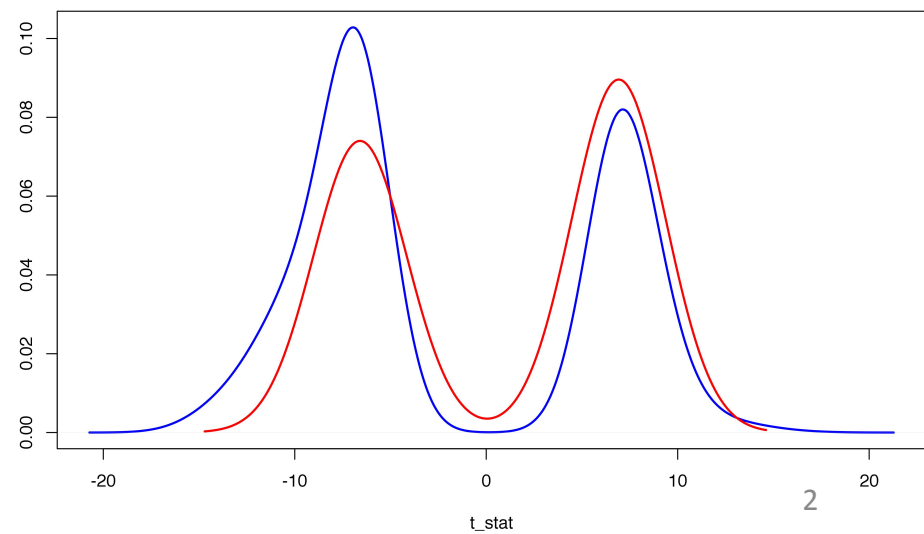
FDR values



p_val values



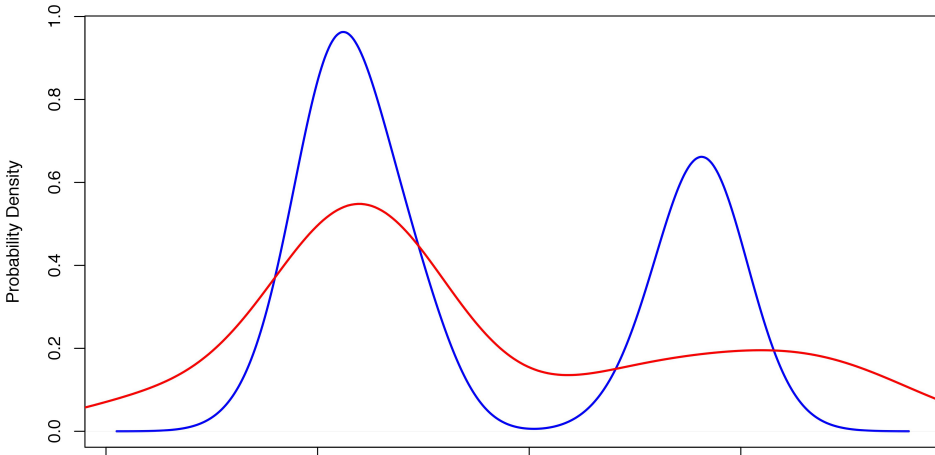
t_stat values



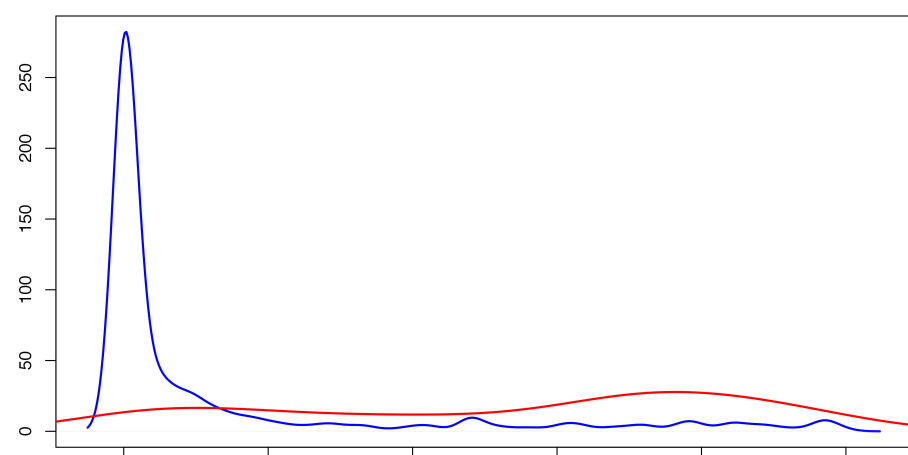
Brain_Caudate_basal_ganglia

— True positives
— False positives

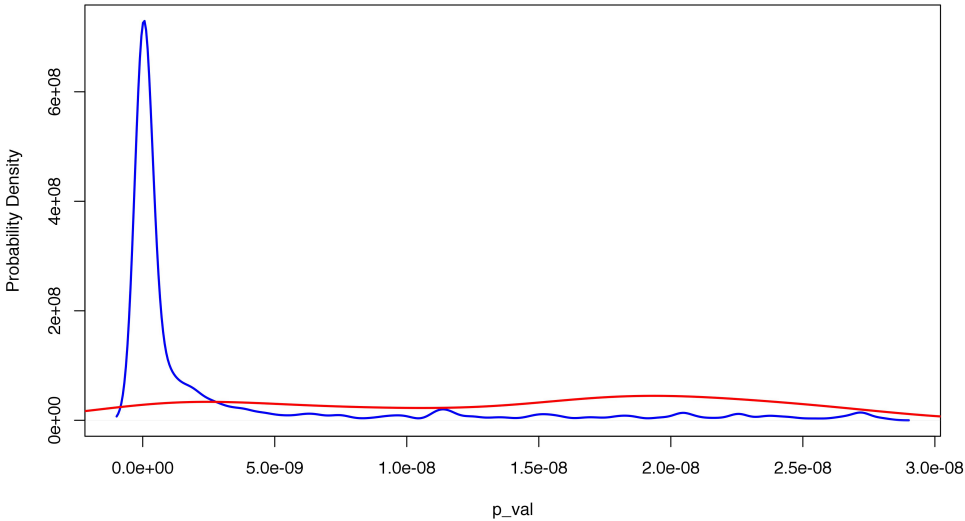
Brain_Caudate_basal_ganglia
beta values



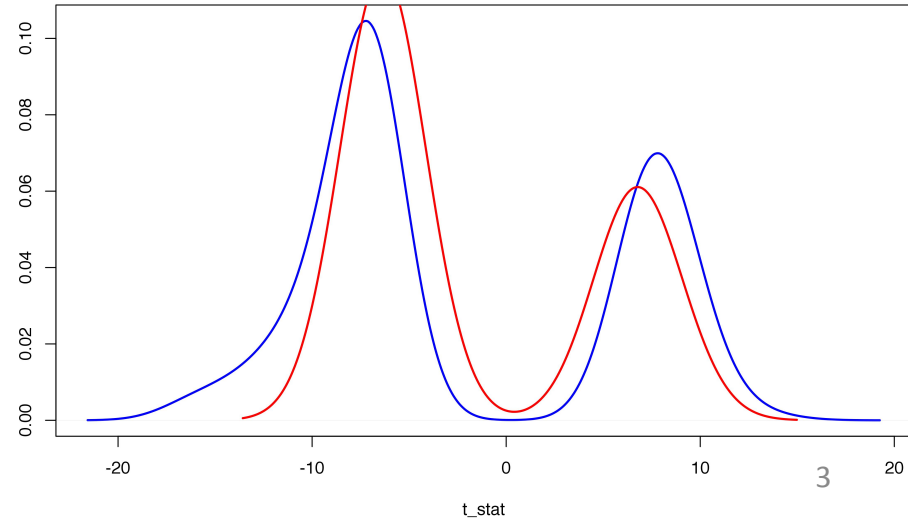
Brain_Caudate_basal_ganglia
FDR values



Brain_Caudate_basal_ganglia
p_val values



Brain_Caudate_basal_ganglia
t_stat values



Reproducibility in Identified eQTLs (SNV-gene pairs)

Potential causes of disparities

- 1) Investigate the 1.5% difference in SNV sets (unlikely a major issue)
- 2) cis/trans eQTLs? – “The cis window was defined as 1 megabase up- and down-stream of the transcriptional start site ($\pm 1\text{Mb}$)”
- 3) “For a given tissue, genes having at least 0.1 RPKM in 2 or more individuals were retained.”
- 4) FDR cutoffs

Reproducibility in Identified eQTLs (SNV-gene pairs)

Potential causes of disparities

- 1) Investigate the 1.5% difference in SNV sets (unlikely a major issue)
- 2) cis/trans eQTLs? – “The cis window was defined as 1 megabase up- and down-stream of the transcriptional start site ($\pm 1\text{Mb}$)”
- 3) “For a given tissue, genes having at least 0.1 RPKM in 2 or more individuals were retained.”
- 4) ~~FDR cutoffs~~

Reproducibility in Identified eQTLs (SNV-gene pairs)

Potential causes of disparities

- 1) Investigate the 1.5% difference in SNV sets (unlikely a major issue)
- 2) cis/trans eQTLs? – “The cis window was defined as 1 megabase up- and down-stream of the transcriptional start site ($\pm 1\text{Mb}$)”
- 3) “For a given tissue, genes having at least 0.1 RPKM in 2 or more individuals were retained.”
- 4) ~~FDR cutoffs~~

Misc notes

UCSC human genome release version hg19 (Genome Reference Consortium GRCh3

currently running cis (grace keeps crashing

extremely good results (but some currently running, and p-vals may be too stringent)