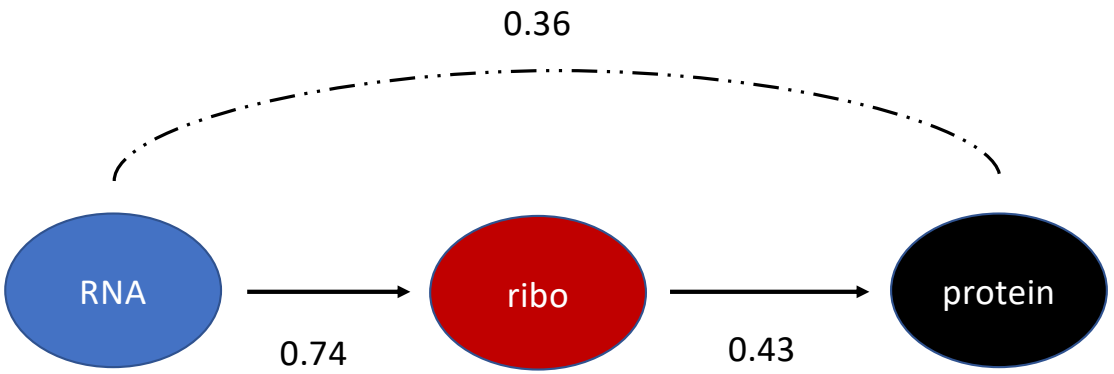
BrainGVEX

01/18/21

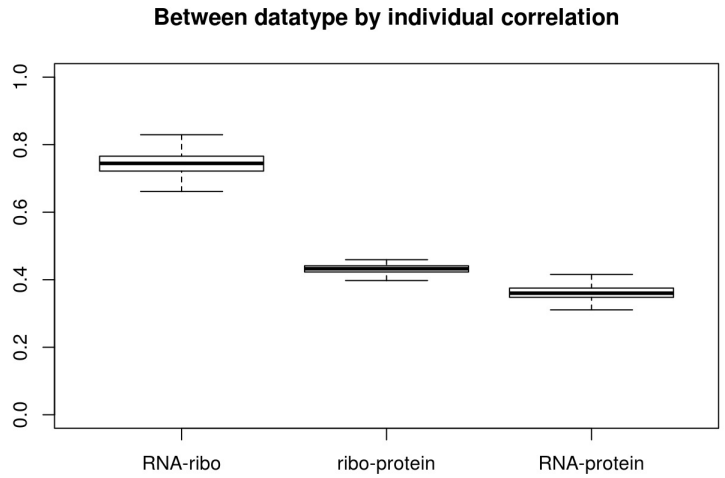
Design

- Prefrontal cortex sample
- 190 individuals (subset of BrainGvex) have matching **RNA-seq**, **ribo-seq** and **quantitative mass spec data**
- Corresponding genotype data generated from psychchip, affyarray, and low pass whole genome sequencing

Overall correlation between datatypes



Reported values are average by sample Spearman's rho



cisQTL mapping

- After excluding related samples, the final dataset for cisQTL mapping has 187 samples
- Genotype data: exclude SNPs in ENCODE black listed regions
- Phenotype (trait) data: keep only autosomal genes that are considered sufficiently quantified in all three datatype : 7,566 genes
- Phenotype data were first normalized for expression level (TMM for NGS, mass spec normalized using house keeping genes) and then data from each gene transformed to standard normal before QTL mapping.

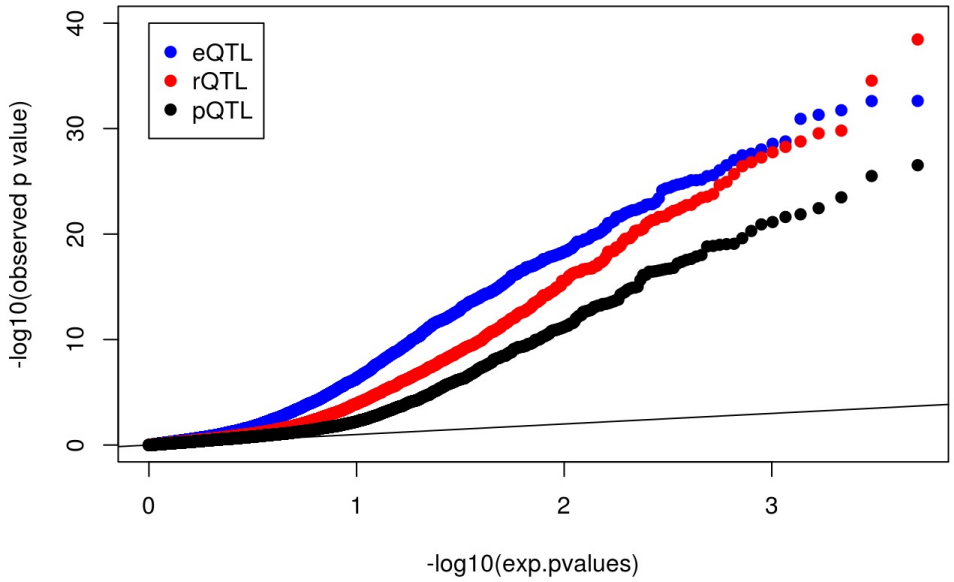
cisQTL mapping

- Used QTLtools, 1Mb cis-window +/- TSS, 10K permutation, beta approximation to estimate empirical p value
- Select best SNP to represent QTL for each gene (eGENE)
- Further adjust gene level FDR using qvalue
- Remove PC until maximizing number of eGENEs identified
  - eQTL: 53 PC removed
  - rQTL: 50 PC removed
  - pQTL: 56 PC removed

cisQTL mapping

- At 10% FDR, out of 7,566 genes deemed quantified, we identified

  - 3,787 eQTLs
  - 2,118 rQTLs
  - 1,046 pQTLs



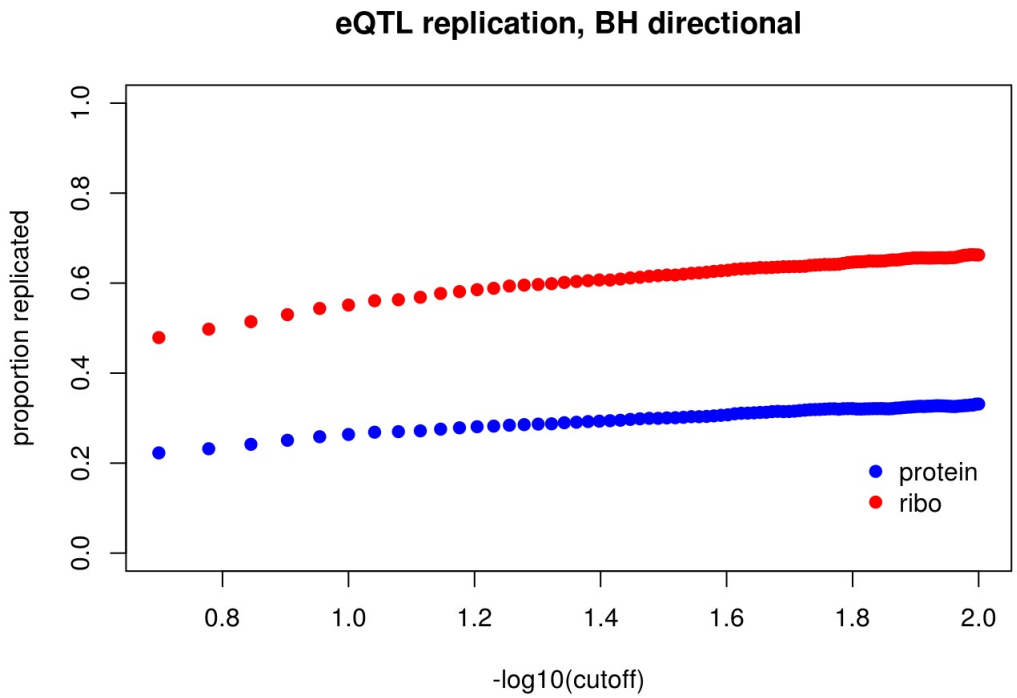
How often are effects of QTL found in one datatype replicated in other datatypes?

Replication phenotype

Discovery phenotype	RNA	ribo	protein
eQTL		55%	26%
rQTL	72%		43%
pQTL	53%	62%	

- Neighboring phenotypes have higher replication
- Upstream phenotypes have higher replication

Proportion eQTL replicated across eQTL FDR



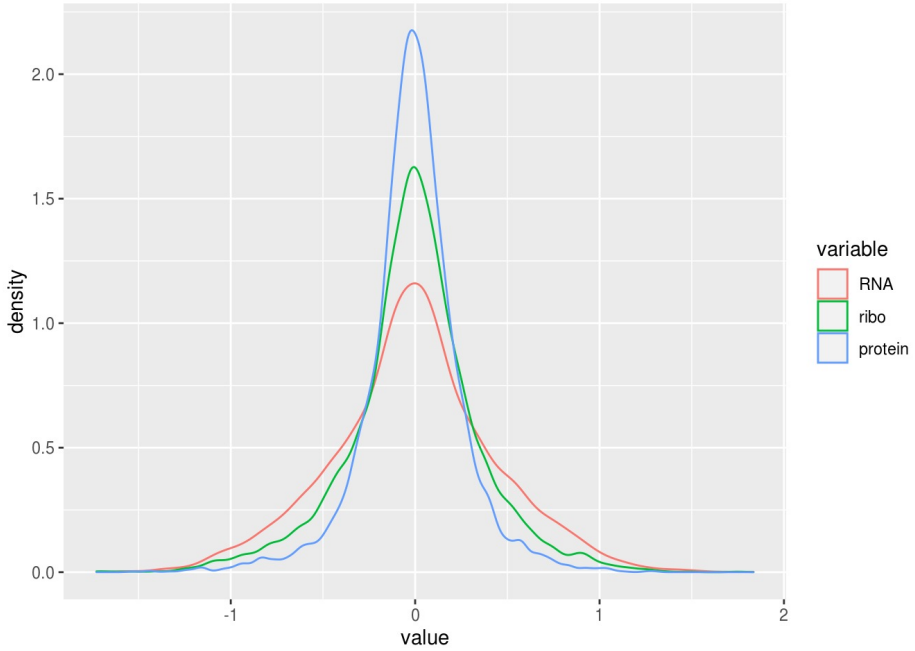
Fewer QTL were found in downstream phenotypes with low replication rates

Are eQTL effects often attenuated (buffered) downstream or just reflecting power issue?

- Check effect size, noisy data will decrease power but will not bias slope (effect)
- However, need eQTL identified from independent data. Because by focusing on eQTL identified from our dataset, our sampling will be biased towards SNPs with larger effect size in RNA-seq data.
- Use common mind consortium (CMC) eQTL

Effect size of CMC eQTL in brainGVEX data

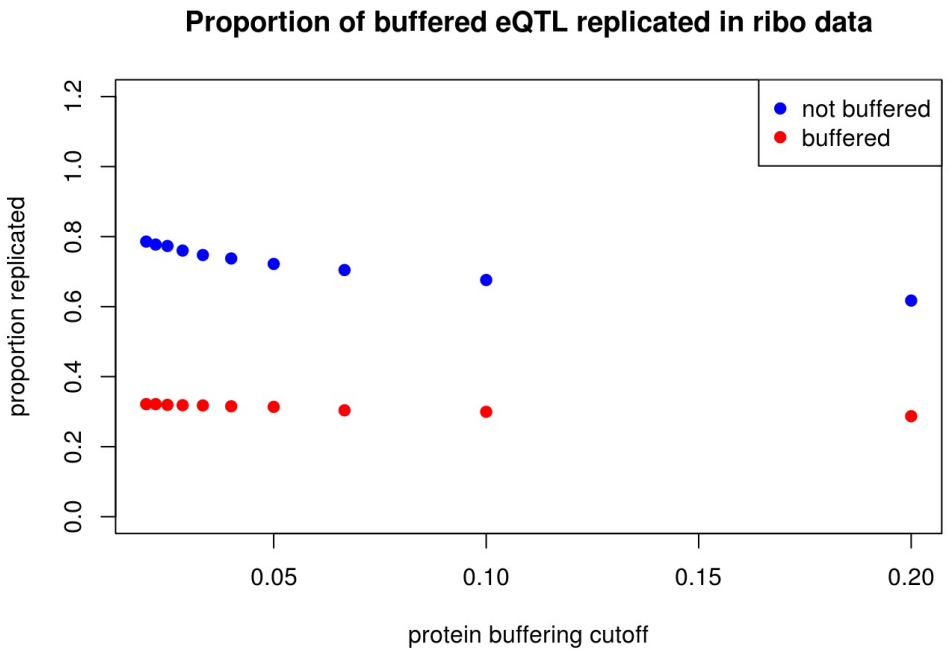
- Using 5% FDR to define CMC eQTL, 4804 eGENEs were quantified in brainGVEX dataset.

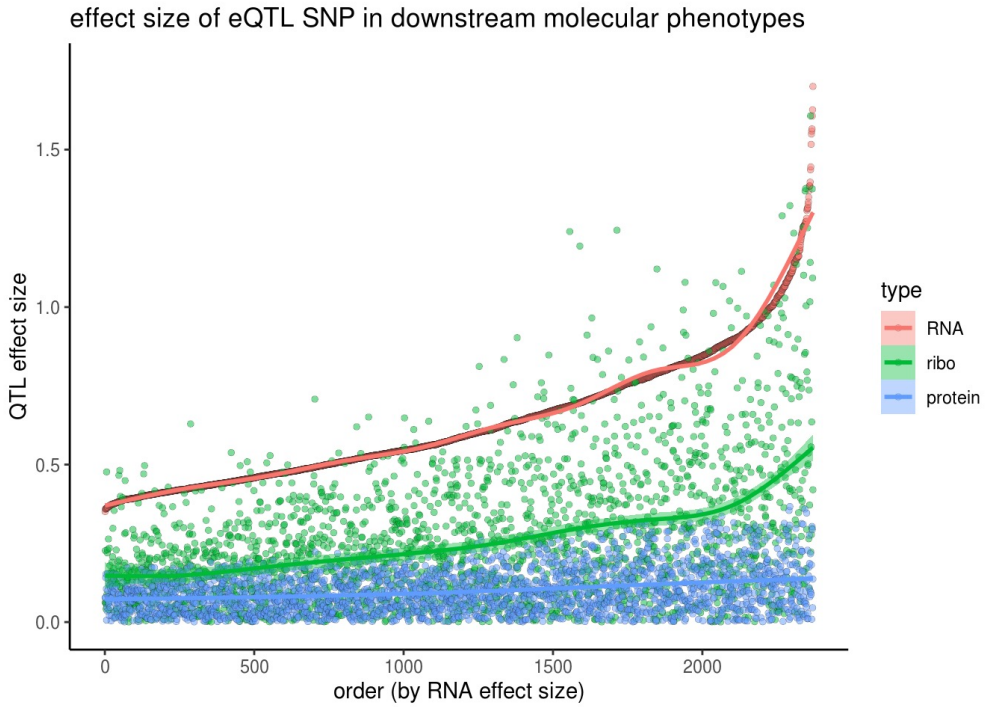


Effect size of CMC eQTL in brainGVEX data

- Effect of CMC eQTL variant in brainGVEX gene expression are attenuated (buffered) downstream at the translation level and further attenuated at the protein level.
- Compared to our previous work on posttranslational buffering of gene expression in LCL, here we found an additional layer (translational) of gene expression buffering.

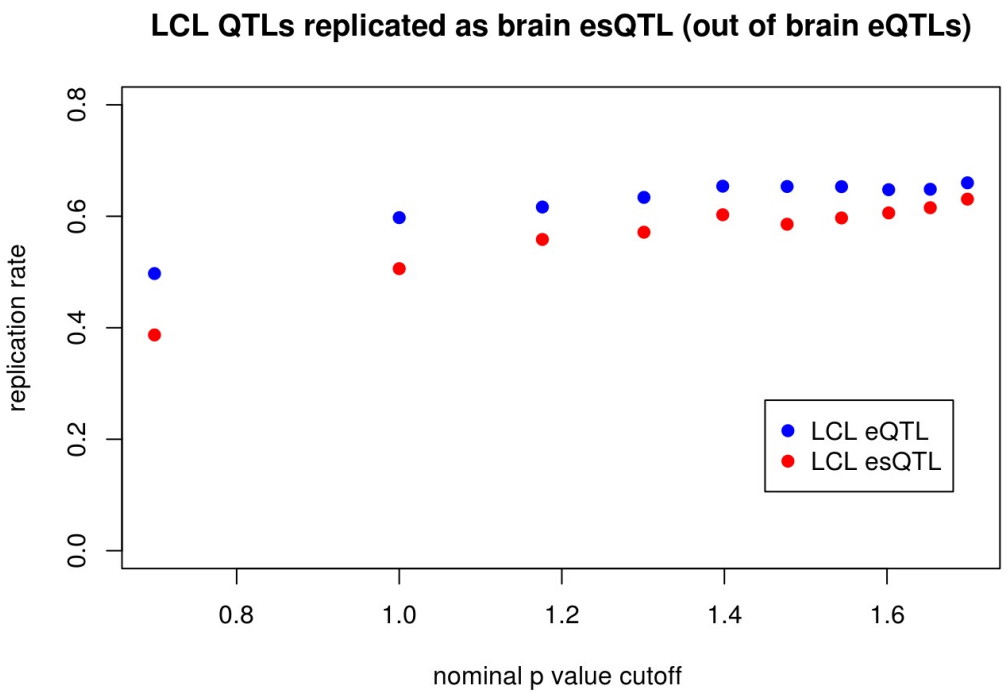
Translational vs. post-translational buffering





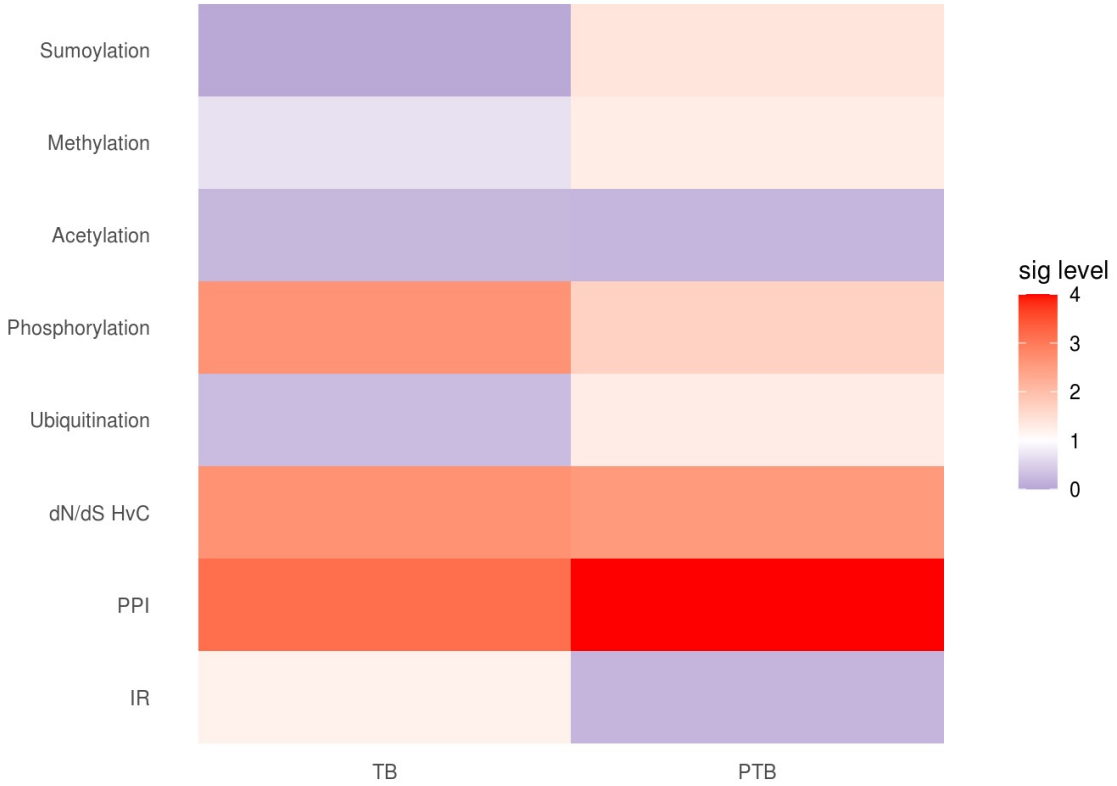
Translational buffering appears to be more prominent than post-translational buffering in brainGVEX dataset

eQTLs are often shared between cell types, are gene expression buffering also shared?



Buffered LCL eQTLs are not more likely to be buffered in the brain than typical LCL eQTLs

Enrichment of Intron retention and PTM distinguishes two types of Buffered eGENEs



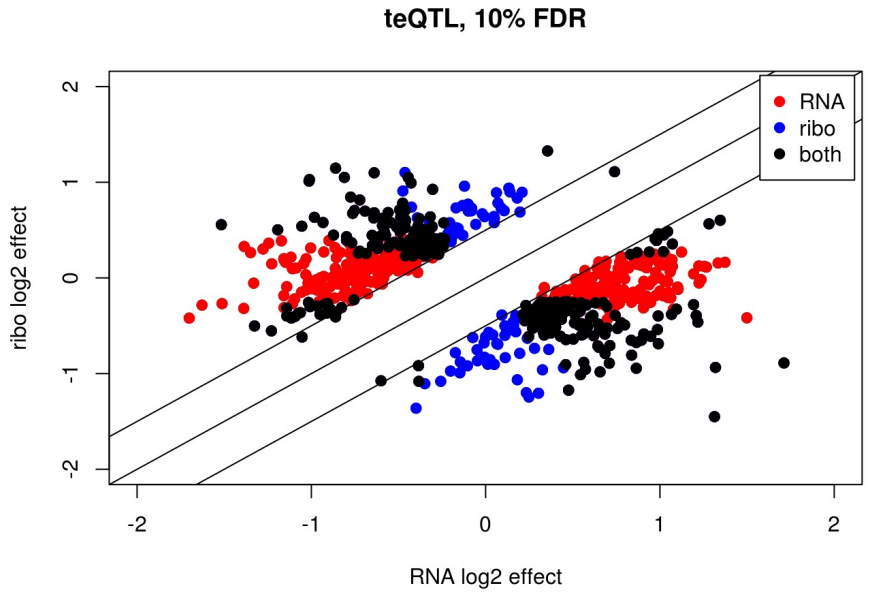
Translation efficiency QTL mapping

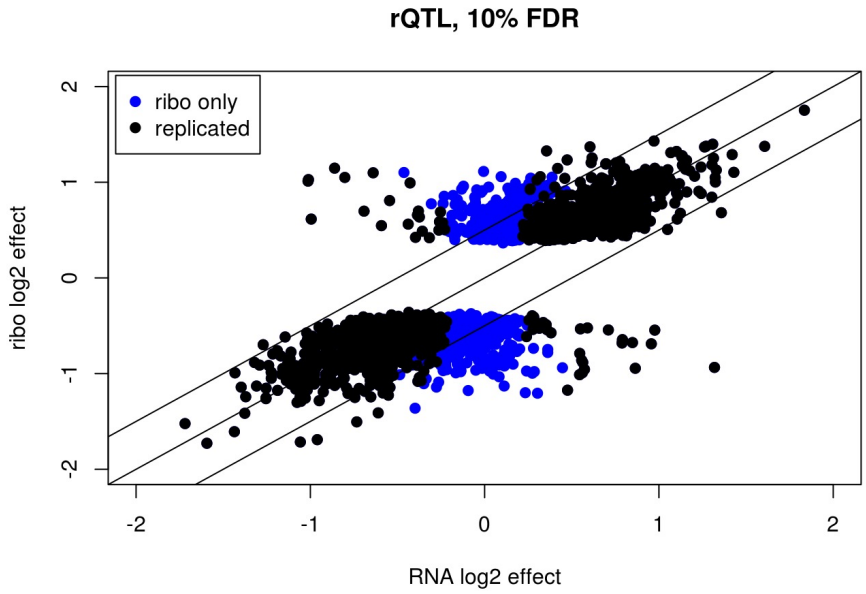
- Pervasive translational buffering of eQTL indicates genetic regulation of translation efficiency.
- By jointly modeling RNA-seq and ribo-seq data, we test for genotype by datatype interaction to identify genotype association with translation level conditional on the transcript level.
- Used the same parameters as eQTL/rQTL mapping, but implemented teQTL mapping in Matrix eQTL, which has built-in interaction model.

Expression = B1 Genotype + B2 Datatype + B3 Genotype * Datatype
H0: B3 = 0

cis-teQTL mapping

- At 10% FDR, out of 7,566 genes deemed quantified, we identified 691 teQTL





cis-teQTL features

- Genomic feature enrichment (uORF, Kozak, miRNA)