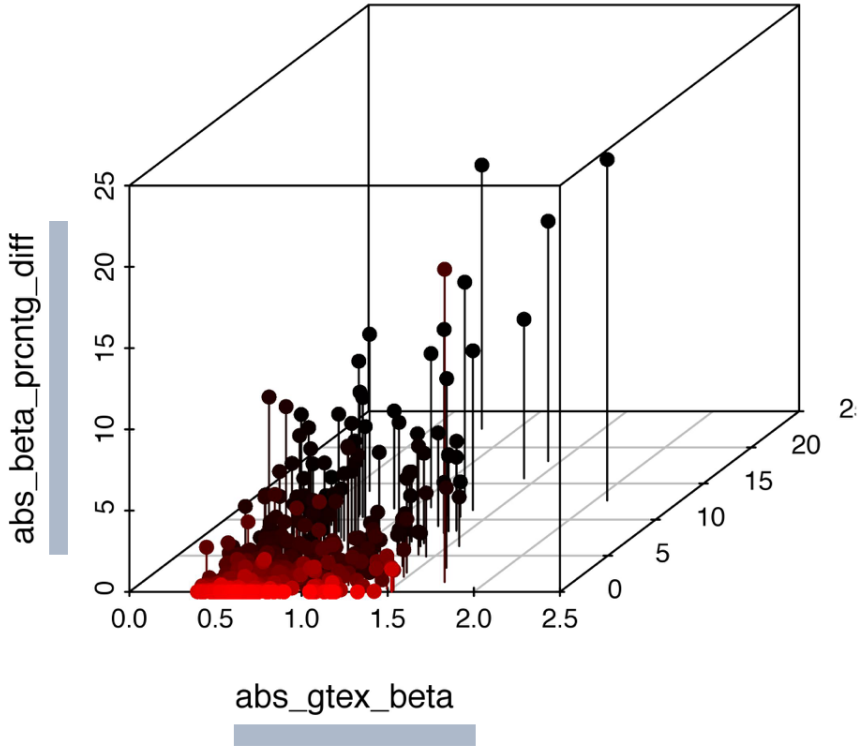
1) phased vs. non-phased genotypes in detecting eQTLs (GTEx uses non-phased genotype data)



← **errors from using phased genotype**

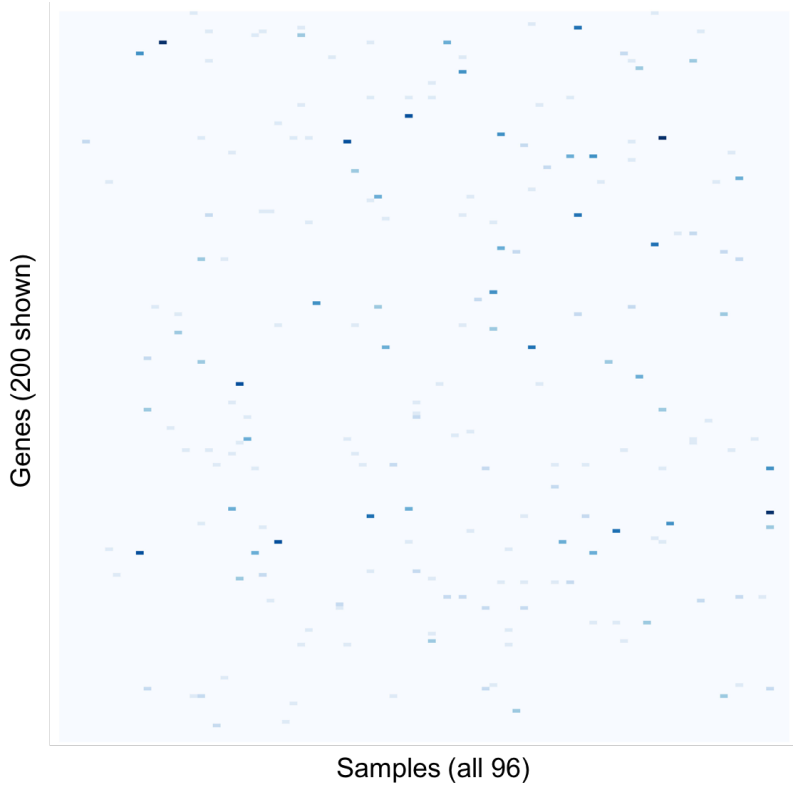
2) eQTLs & permutations

"To identify eQTL-containing genes (eGenes), a permutation procedure was applied, correcting for the multiple hypothesis effect of many SNPs in LD for a given gene. The minimal p-value per gene (min(p)) was used as the test statistic. Permutations were performed by randomizing sample labels for the expression data. The same random indexes were applied to the PEER factors and gender covariates. Genotypes and Genotyping PCs were not randomized. A minimum of 1000 permutations and a maximum of 10,000 permutations were performed, with an exit criteria in between this range of having at least 15 permuted min(p) values less than the nominal min(p). Having derived an empirical p value for each gene, q-values were calculated using the Storey approach (65) and a q-value threshold of <= 0.05 was applied…"

-- Supp to GTEx Pilot Paper

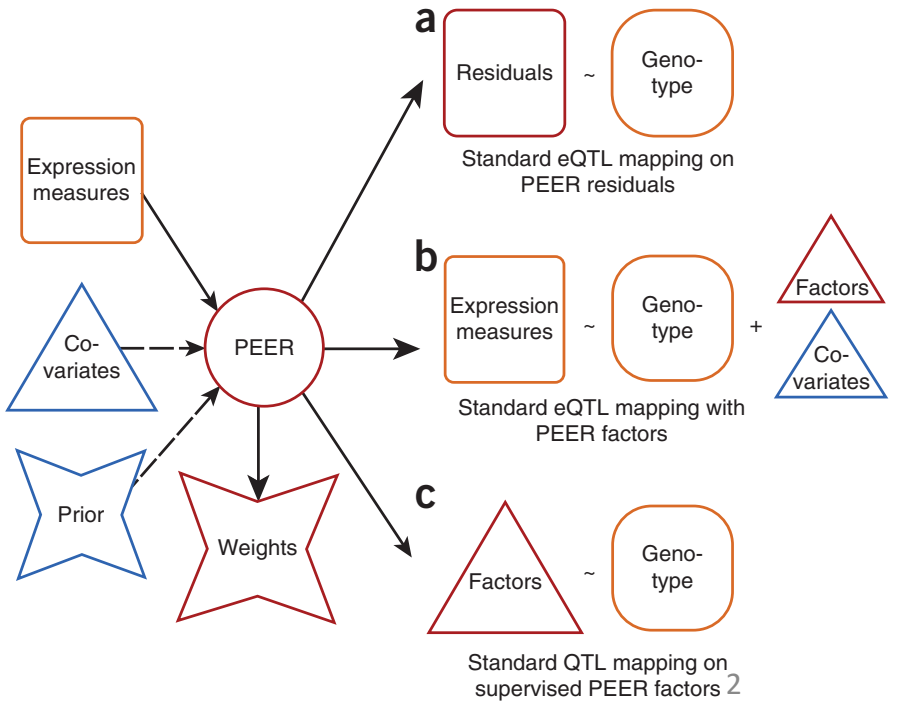
3) gene expression normalization (variables include mean vs. median quantile normalization, log2, and at which step log transformations are performed, etc)

+ why were mitochondrial genes not excluded in expression normalization?
+ filters applied across all samples or only within each tissue?



4) t-SNE-related

+ why do our peer factors not match their factors exactly (see heatmap)?
+ w.r.t. peer -- why do 2 methods of including peer not match?
+ additional covariates added per population?



Mtg on Jun 28:

- key contacts
- v8 = final release (when and and what to expect)?
- Though analysis likely to continue for several months thereafter (Knowles)
- useful to obtain slides again (as w/previous July)?
- additional questions from others (SL, DW, etc)?