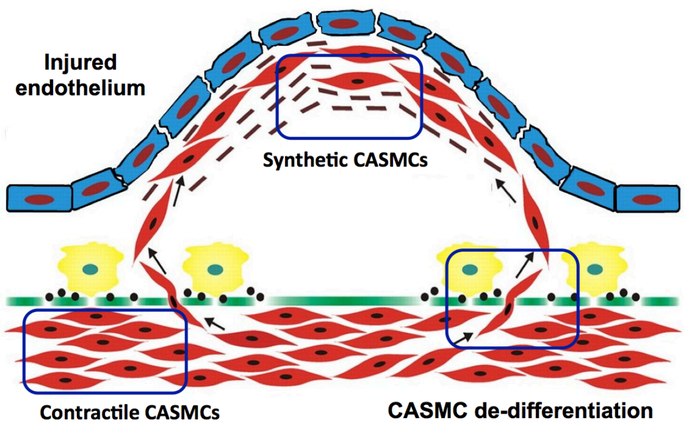
Detect Gene-by-Environment Interactions in coronary artery disease

Boxiang Liu[1,2,3], Robert S. Kirby[4]

Department of Pathology[1], Biology[2], Statistics[3], and Electrical Engineering[4], Stanford University

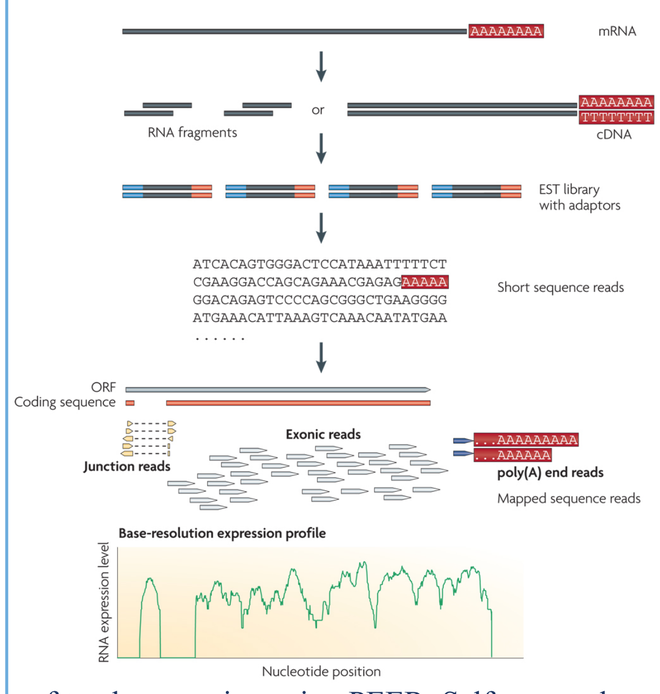
I. Introduction



Coronary artery disease (CAD) is a complex disease influenced by both genetic and environmental factors. It is the leading cause of morbidity and mortality, leading to every 1 in 4 deaths in the US. The majority of coronary artery disease results from atherosclerosis, or formation of plaque. Cell lineage studies have shown that more than 80% of plaque mass constitutes of coronary artery smooth muscle cells (CASMC).

In this study, we use CASMCs to model coronary artery in normal and disease phenotype. CASMCs switch between the synthetic phenotype when stimulated with serum, approximating disease-state coronary artery, and the contractile phenotype when resting. Previous studies have focused on either genetic or environmental risk factor for CAD but few have jointly analyzed gene-by-environment interaction (GxE) effects. In this study, we aim to discovery genes and associated genetic variants that interact with serum treatment using a small cohort.

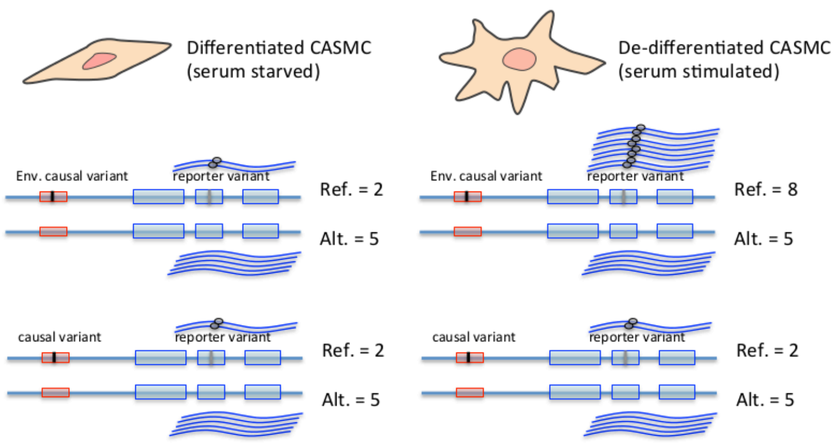
II. Dataset and Preprocessing



We obtained 21 human coronary artery smooth muscle cell (HCASMC) samples. Eleven samples are cultured in serum-free growth media to emulate the contractile phenotype, the other ten samples are treated with fetal-bovine serum (FBS) to induce synthetic phenotypes. We performed RNA sequencing (RNAseq) and whole-genome sequencing (WGS) on all samples. We aligned RNAseq data to the human reference genome v19 using STAR, and WGS data using BWA. We called WGS variants using iSAAC and quantified RNAseq total expression using HTSeq and allele-specific expression using custom scripts. We correct known (age, sex, ancestry, batch) and latent covariate of total expression using PEER. Self-reported ancestry is confirmed using PCA. We imputed missing variants and phased haplotypes using impute2 with 1000 Genomes reference panel.

RNAseq pipeline in above figure is from Wang, Z. et al (2009) Nature Reviews Genetics.

III. Methods



We hypothesize that GxE interaction manifest as differential allele-specific expression (left figure). We use two different approaches to detect interaction effects. In the first approach, we locate significant genes using a lenient FDR of 0.2. We subsequently map interaction eQTLs using both allele-specific and total read count. Hypothesis tests are carried out using likelihood ratio tests. In the second approach, we use WASP to map expression QTLs separately for two conditions. In brief, WASP jointly models the allele-specific and total reads and tests for significant eQTL using likelihood ratio tests. In simulation study, we simulated total reads as a overdispersed Poisson random variable, and the allele-specific read count as an overdispersed binomial random variable.

Generalized Linear Mixed Model (EAGLE)

$$H_0: y^a \mid n, \varepsilon; \mu \sim Binom(n, \sigma(\mu + \varepsilon))$$

$$H_1: y^a \mid n, \varepsilon; \mu, \beta^e \sim Binom(n, \sigma(\beta^e e + \mu + \varepsilon))$$

Interaction eQTL testing

$$H_0: n = \beta^g g + \beta^e e + \mu + \varepsilon$$

$$H_1: n = \beta^g g + \beta^e e + \beta^{g \times e} ge + \mu + \varepsilon$$

Simulation

$$n \mid \lambda \sim Poisson(\lambda); \lambda \sim Gamma(2,2)$$

$$p = \sigma(\mu + \beta^e e + \beta^h h + \beta^{e \times h} eh + \varepsilon)$$

$$\varepsilon \sim N(0, IG(1.13, 0.0122))$$

$$y \mid n, p \sim Binomial(n, p);$$

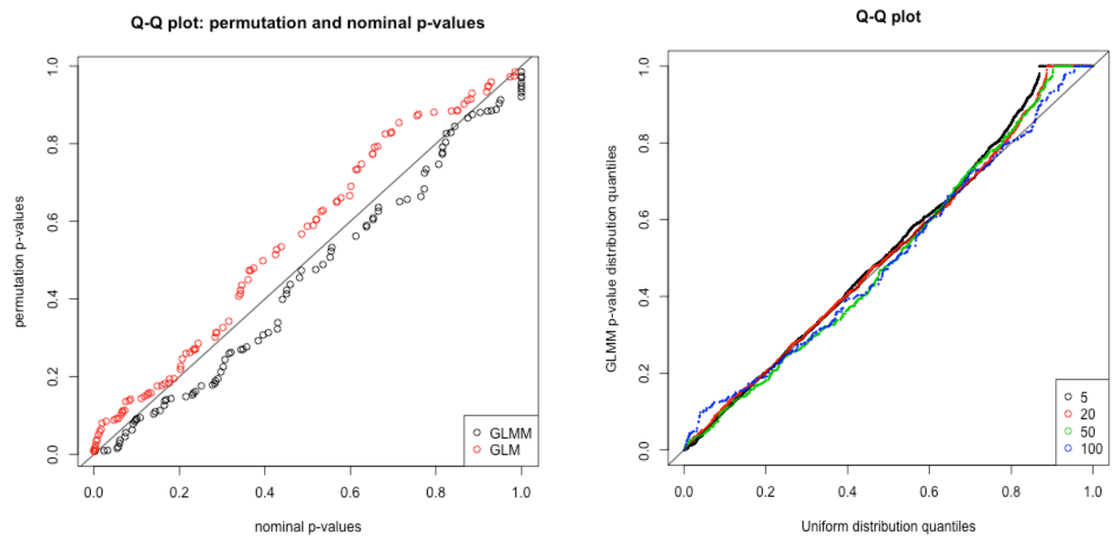
eQTL mapping (WASP)

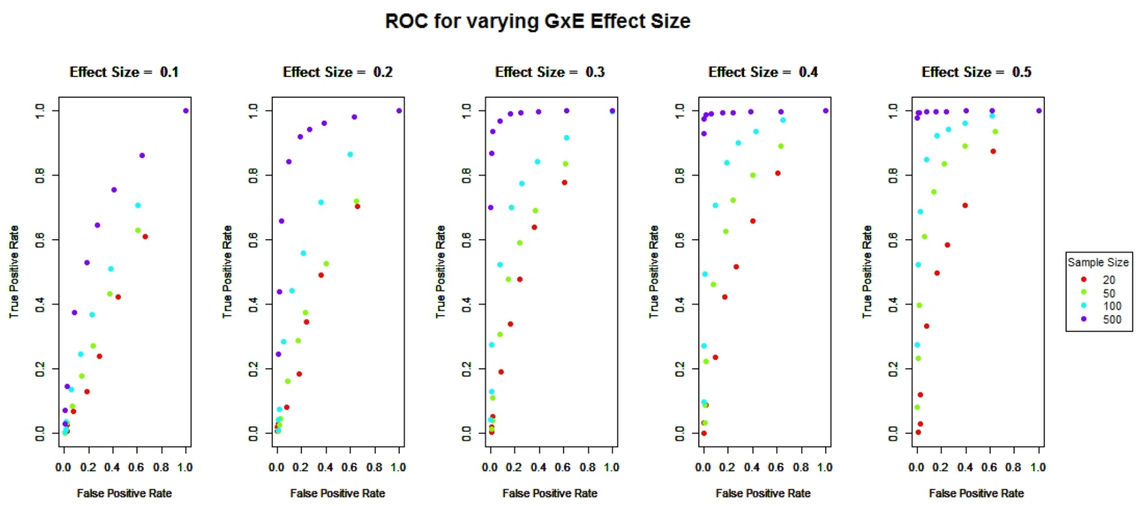
$$N \sim BetaNegativeBinom(\lambda, \eta, \phi)$$

$$y \sim BetaBinom(p, N, \psi)$$

IV.A Results: Generalized Linear Mixed Modeling

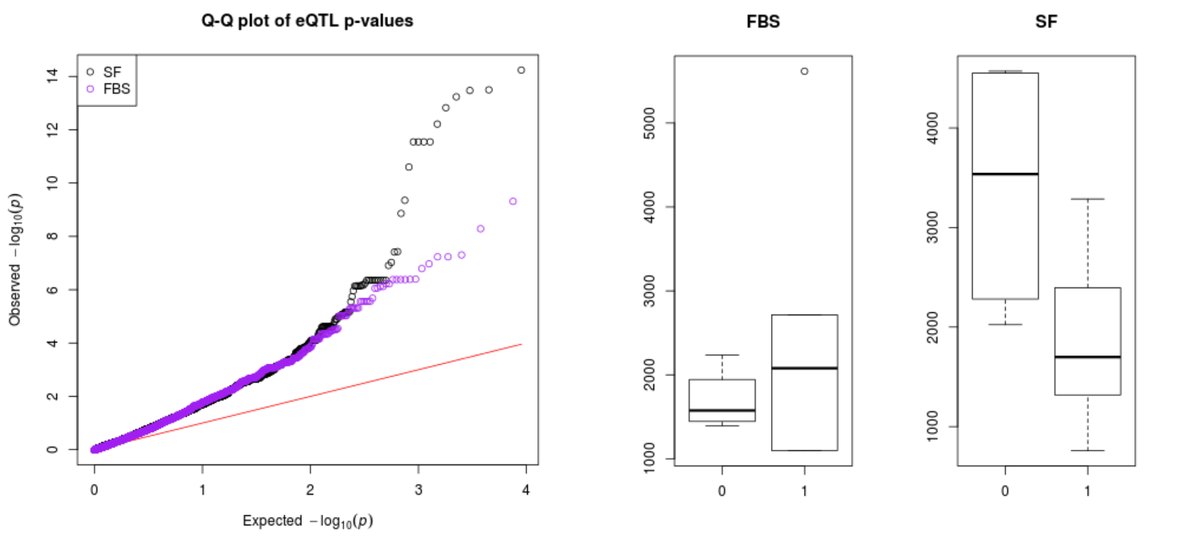
The generalized linear model with logit link assumes the allele-specific read count follows a binomial distribution. We tested the validity of this assumption by comparing BGLM and permutation p-values (top-left). The upward shifts of red dots indicate overdispersion in the allele-specific counts and anti-conservativeness of the GLM. We relaxed the GLM assumption by decomposing the mean effect into a fixed effect μ and a random effect ε. The resultant mixed model captured overdispersion and is conservative (black dots on the top-left panel). Unfortunately, we failed to discover significant hits using the GLMM. The top-right panel shows that p-values approximates the uniform distribution under different minimum read count threshold. We used simulated data to investigate the lack-of-power issue (bottom). Using 20 samples, a type I error rate (or false positive rate) of 0.05 allows a recovery of less than 5% of true positives hits, leading to zero discovery after multiple hypothesis correction.





IV.B Results: interaction QTL mapping

Using an alternative approach, we mapped eQTL for two conditions separately and compare the differences. We mapped eQTLs using WASP by van de Geijn et al. Even with small sample size of 10 and 11, the eQTL tests are well-powered (left). To compare the difference between two conditions, we use a two-step FDR cutoff suggested by Barreiroa et al. In the first step, we use a stringent FDR of 1% to select significant eQTLs in either condition. We subsequently relax the FDR to 50% for the other condition. This approach gives a conservative list of interaction eQTLs. The right-hand panel shows one example of such discovered interaction QTL.



Future Directions

1. Extend analysis from chr22 to whole genome
2. Annotate discovered interaction eQTL
3. Perform weighted tests by incorporating epigenomic information

References

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics *10*, 57–63.

van de Geijn, B., McVicker, G., Gilad, Y., and Pritchard, J.K. (2015). WASP: allele-specific software for robust molecular quantitative trait locus discovery. Nat Meth *12*, 1061–1063.

Barreiro, L.B., Tailleux, L., Pai, A.A., Gicquel, B., Marioni, J.C., and Gilad, Y. (2012). Deciphering the genetic architecture of variation in the immune response to Mycobacterium tuberculosis infection. Proc. Natl. Acad. Sci. U.S.a. *109*, 1204–1209.

Knowles, D.A., Davis, J.R., Raj, A., Zhu, X., Potash, J.B., Weissman, M.M., Shi, J., Levinson, D., Mostafavi, S., Montgomery, S.B., et al. (2015). Allele-specific expression reveals interactions between genetic variation and environment. bioRxiv 025874.