Allele-specific method for testing the association between molecular quantitative traits and phenotype-genotype interaction

Qi Yan[1], Erick Forno[2], Juan C. Celedón[2], Wei Chen[2,3,4] and Daniel E. Weeks[3,4]

[1]Department of Obstetrics and Gynecology, Columbia University Irving Medical Center; [2]Division of Pediatric Pulmonary Medicine, UPMC Children's Hospital of Pittsburgh, University of Pittsburgh; [3]Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh; [4]Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh.

Background

- One advantage of next generation sequencing data is that it can be used to identify allele-specific differences in molecular traits.
- Several molecular quantitative trait locus (QTL) methods, such as WASP [1] and RASQUAL [2], integrated the allele-specific information into statistical models, and obtained new insights that are not obtainable by standard QTL approaches.
- The variation of molecular quantitative traits (e.g., gene expression) could result from the joint effect of genotypes and phenotypes.
- We assume that integrating allele-specific information can potentially boost power for testing phenotype-genotype interactions.
- We propose a novel statistical method, the combined haplotype interaction test (CHIT), that tests the association between molecular quantitative traits and phenotype-genotype interactions by modeling the total read counts and allele-specific reads in a target region.

Method

Modeling the total reads in a *feature region*

$$\boldsymbol{y_{il}} \sim \mathbf{BNB}(\boldsymbol{\Omega_i}, \boldsymbol{a_{il}}, \boldsymbol{b_{il}}):\ y_{il}|p \sim \mathrm{NB}(\Omega_i, p);\ p \sim \mathrm{Beta}(a_{il}, b_{il})$$

where y_{il} is the observed number of reads for individual *i* in *feature region l*, Ω_i is an overdispersion parameter for each individual, and a_{il} and b_{il} are predefined shape parameters.

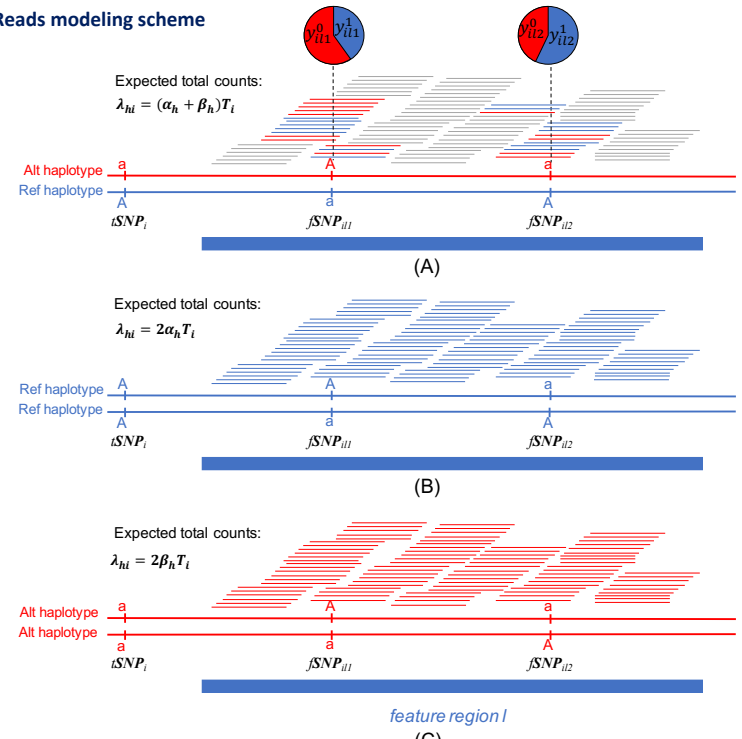
Modeling the allele specific reads in a *feature region*

$$\boldsymbol{y^1_{ilj}} \sim \mathbf{BB}\left(\boldsymbol{y_{ilj}}, \boldsymbol{c_{il}}, \boldsymbol{d_{il}}\right):\ y^1_{ilj}|p \sim \mathrm{Binomial}\left(y_{ilj}, p\right);\ p \sim \mathrm{Beta}(c_{il}, d_{il})$$

where y^1_{ilj} is the number of allele-specific reads from the reference haplotype for the *j*-th *fSNP* in *feature region l* in individual *i*, y_{ilj} is the total number of allele-specific reads for the *j*-th *fSNP*, and c_{il} and d_{il} are predefined shape parameters.

Full likelihood: $L = \prod_i \left\{ \Pr_{\mathrm{BNB}}(y_{il}|\alpha_l, \beta_l, \Omega_i, \phi_l, T_i) \prod_j \Pr_{\mathrm{BB}}\left(y^1_{ilj}|y_{ilj}, \alpha_l, \beta_l, \Upsilon_i\right) \right\}$

Reads modeling scheme



The reference and alternative haplotypes are determined by the *test-SNP* (*tSNP*). The *feature SNPs* (*fSNPs*) are the heterozygous SNPs inside the *feature region.* (A) When the *tSNP* is heterozygous *Aa*, the allele-specific read counts $y_{ilj} = \left(y^0_{ilj}, y^1_{ilj}\right)$ can be observed at the *j*-th *fSNP*. The blue reads map to the reference haplotype, the red reads map to the alternative haplotype, and the grey reads could be mapped to either haplotype. (B) When the *tSNP* is homozygous *AA*, all reads are mapped to the reference haplotype. (C) When the *tSNP* is homozygous *aa*, all reads are mapped to the alternative haplotype. Thus, it is non-informative to count allele-specific reads in (B) or (C).

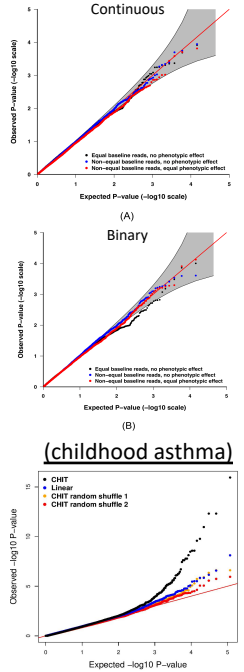
Testing the association between molecular quantitative trait and phenotype-genotype interaction (Adding phenotype in the model)

$$\lambda_{il} = \begin{cases} 2\alpha_l e^{u_l \cdot Z_i} T_i, & if\ tSNP = AA \\ (\alpha_l e^{u_l \cdot Z_i} + \beta_l e^{v_l \cdot Z_i}) T_i, & if\ tSNP = Aa \\ 2\beta_l e^{v_l \cdot Z_i} T_i, & if\ tSNP = aa \end{cases}$$

where Z_i is the phenotype for *i-th* subject, u_l and v_l are the coefficients of the phenotypic effect for the reference and alternative haplotypes.

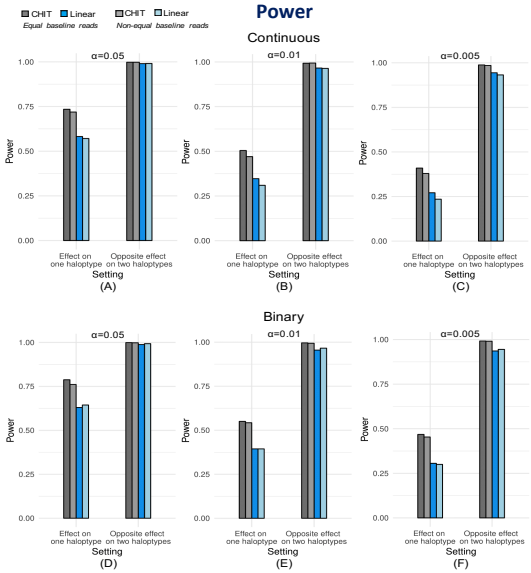
Results (simulations)

Type I error rate



(childhood asthma)

Power



References

1. 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 Methods 12, 1061-1063.
2. Kumasaka, N., Knights, A.J., and Gaffney, D.J. (2016). Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. Nat Genet 48, 206-213.