An Improved Approach to Model Epistatic Interactions Among Multiple Genetic Variants Using Pleiotropy

Tian-Shun Allan Jiang

Stanford University,
NC School of Science and Mathematics

Gregory Carter

The Jackson Laboratory

The Jackson Lab Summer Student Program

July 31, 2014

*Todo list

Summarize what you've done with little cross dataset, like in Carter 2013 2
write about the data... 2
In our data, these eigentraits captured x% of the variance. 3
Figure: Show plotSVD that has percent variance plots 3

Introduction

One area of focus in systems biology is the use of quantitative and statistical genetics to understand how genetic variants combine to influence complex traits. Advances in high-density genotyping and multidimensional phenotyping can produce highly detailed views of biological systems for this purpose. However, extracting meaningful biological models for human health and disease from these large datasets will require the creation of new analytical methods and software.

One promising approach to use this data is in the systematic study of genetic interactions as a way to map genetic networks [2]. These networks can then be used to infer functional relationships in molecular biology such as activation, repression, and pathway ordering [1]. Previous work by Carter et al. [3, 4] created mathematical models to infer genetic networks from epistatic interactions using various fly and yeast crosses exhibiting partial pleiotropy. These models were then used on some well-studied biological systems, such as the reproductive pathway in yeast [4] to confirm wet-lab results found in literature.

However, interactions detected by existing models only analyse two genetic variants at a time and thus may not provide a complete model of epistasis. In this paper, we create a more complete model of epistasis by extending the previous model to incorporate multiple genetic variants at a time, which can help uncover novel interactions and detect indirect variant-to-variant influences.

Summarize what you've done with little cross dataset, like in Carter 2013

Methods

Data Source

Data were obtained from a study of

write about the data...

Genetic Interaction Model

Since our model extends upon previous data analysis techniques by Carter et al. which are described at length in previous publications [3, 4], we summarize the known analysis procedures here, and cover novel techniques and ideas in higher detail. All of the previously known analysis techniques were performed using the software package `R/cape` [5].

1. Singular Value Decomposition

First, we performed singular value decomposition (SVD) on our mean-centered, standard-deviation-normalized phenotype data to maximize orthogonality. This allows us to fully exploit the complementary phenotype data by reducing linear dependence. Then we chose the first five left singular vectors for analysis. For convenience, we hereafter refer to these left singular vectors as "eigentraits".

In our data, these eigentraits captured x% of the variance.

Show plotSVD that has percent variance plots

2. Singlescan to find covariates

Then, we performed a linear regression for each of the 100 genes to identify strong-effect genetic variants to be treated as additive covariates in subsequent pairwise regressions. For each locus we performed the regression:

$$U_i^j = \beta_0^j + \beta_1^j x_i + \epsilon_i^j \tag{1}$$

where i is looped over all individuals in the population, j is looped over all eigentraits,

U_i^j is the value of eigentrait j for individual i, β^j is the effect of the effect of the locus on eigentrait j, and ϵ_i^j is the residual error. In our data, each of the x_i genotype values were either $0, 1$ or 2, corresponding to the AA, AB, and BB genotypes. Genotypes are encoded this way by the additive assumption of variants on phenotypes. Then, strong-effect knockdowns are identified as significant β_1^j coefficients, which are then included as covariates for the associated phenotype.

Here is where our analysis extends upon methodology by Carter et al. [3, 4]. Previous methodology conducted a "pairscan", shown below for sample variants 1 and 2.

$$U_i^j = \beta_0^j + \underbrace{\sum_c x_{c,i}\beta_c^j}_{\text{Covariates}} + \underbrace{\beta_1^j x_{1i} + \beta_2^j x_{2i}}_{\text{Main Effects}} + \underbrace{\beta_2^j x_{1i} x_{2i}}_{\text{Interaction}} + \epsilon_i^j \tag{2}$$

The variables are the same as those in Equation 1, with the additional interaction term β_{12}^j and the addition of strong-effect knockdowns as covariates. This interaction term can be reparameterized into variant-to-variant influences, but is limited by the fact that it can only model interactions between two variants at a time.

3. Extending the model to n variants

Here, we extend the model to the analysis of n variants as follows:

$$U_i^j = \beta_0^j + \underbrace{\sum_c x_{c,i}\beta_c^j}_{\text{Covariates}} + \underbrace{\sum_{a=1}^{n} \beta_1^j x_{ai}}_{\text{Main Effects}} + \underbrace{\sum_{a,b} \beta_{ab}^j x_{ai} x_{bi}}_{\text{all pairs } a,b} + \underbrace{\sum_{a,b,c} \beta_{abc}^j x_{ai} x_{bi} x_{bi}}_{\text{all triples } a,b,c} + \ldots + \epsilon_i^j \tag{3}$$

This regression models combined variant interaction effects among n variants, as it includes terms for any combination of variants under analysis. For example, the β_{abc}^j coefficient represents the strength of the interaction among variants a, b and c in explaining the variance of eigentrait j.

To resolve these combined interaction coefficients into variant-to-variant influences, we first reparameterize the interaction coefficients in terms of variables δ_{ab} and δ_{ba} for all com-

binations of variants a and b from the n variants of interest.

The value of δ_{ab} represents the change in genotype value in variant b when the perturbation in variant a is present (i.e. when variant a is not homozygous for the reference strain).

Results

Discussion

References

[1] Leon Avery and Steven Wasserman. Ordering gene function: the interpretation of epistasis in regulatory hierarchies. *Trends in genetics*, 8(9):312–316, 1992.

[2] Charles Boone, Howard Bussey, and Brenda J Andrews. Exploring genetic interactions and networks with yeast. *Nature Reviews Genetics*, 8(6):437–449, 2007.

[3] Gregory W Carter. Inferring gene function and network organization in drosophila signaling by combined analysis of pleiotropy and epistasis. *G3: Genes— Genomes— Genetics*, 3(5):807–814, 2013.

[4] Gregory W Carter, Michelle Hays, Amir Sherman, and Timothy Galitski. Use of pleiotropy to model genetic interactions in a population. *PLoS genetics*, 8(10):e1003010, 2012.

[5] Anna L Tyler, Wei Lu, Justin J Hendrick, Vivek M Philip, and Gregory W Carter. Cape: An r package for combined analysis of pleiotropy and epistasis. *PLoS computational biology*, 9(10):e1003270, 2013.