



Comparison of Methods of Association Mapping in Latinos

Andrew Boyd, Abra Brisbin

boydar@uwec.edu

Department of Mathematics, University of Wisconsin Eau Claire, Eau Claire, WI 54956

1 Introduction

- Admixed individuals have genomic regions of ancestry from each ancestral population.
- We can identify these regions using methods found in previous research. [1] (Figure 3)
- We are testing and comparing methods [2] that correlate phenotypes with genotypes and their associated ancestry.
- This can be used to find what genes can increase the chance of a disease.

2 Method Design

- Generate a hypothetical genotype and ancestry for haploid individuals and combine them to make a genotype for hypothetical children.
- Generate hypothetical phenotypes with a normal distribution to attach to the generated genotypes and ancestries (Figure 2a).
- Test the methods on this null association to see how many false positives we find at various confidence levels.
- Add various degrees of association (Figure 2b) between the phenotype and genotype and then test the methods to see how many true positives we find at various confidence levels. (Figure 4)

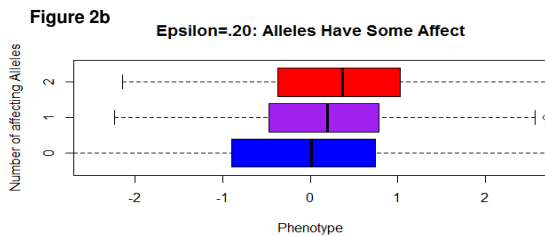
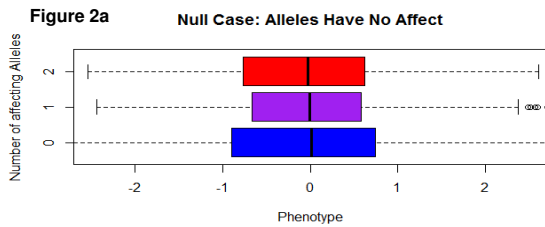


Figure 2. Phenotype boxplot with and without association. The different colors show the number of affecting alleles they have: 0 (blue), 1 (purple), 2 (red). In Figure 2a, the allele has no effect, and in Figure 2b, the allele has an effect of $\epsilon=.20$.

Key
■ two affecting alleles
■ one affecting allele
■ zero affecting alleles

Figure 1

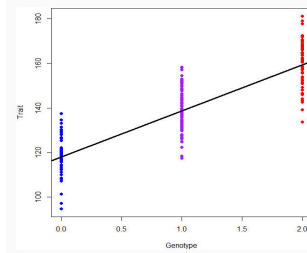


Figure 1. Testing for Association

- In a graph of trait versus genotype, if the slope is nonzero then it is said that there is an association between that trait and the genotype.
- A False Positive:** When the slope in the population is zero, but the method detects it to be significantly different from zero in the sample.
- A True Positive:** When the slope in the population is nonzero, and the method detects it to be significantly different from zero in the sample.

3 Simulations

- Extracted HGDP French, Yoruba, and Mayan chromosomes to simulate individuals with admixed ancestry with French, Yoruba, and Mayan sections.

Figure 3

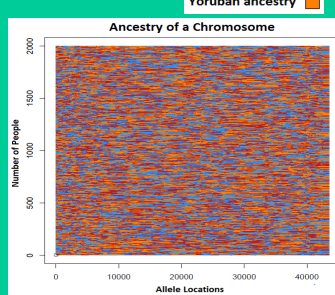


Figure 3. Generated Phenotype. Each horizontal line shows the ancestry of a chromosome which is randomly generated using sections of real chromosomes. A section of neighboring alleles are often from one particular ancestry. The previous research allows us to look at a chromosome and identify what section of alleles belongs to which ancestry. We use the ancestries in addition to the genotype in order to find an association between a genotype and a phenotype.

Figure 4

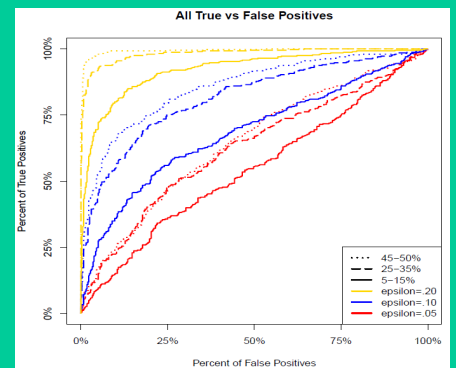


Figure 4. ROC curves for one method.

A ROC curve is a curve comparing the number of true positives to false positives. The x axis, or percent of false positives, is the frequency of associations found in the null simulations at different P-values. The y axis, or percent of true positives, is the frequency of associations found in the simulations with a given effect size (epsilon). Solid, dashed, and dotted lines represent the frequency of the less common genetic variant. As the two variants get closer to equal frequencies (when the smaller amount is closer to 50%), the methods are better at finding true positives and avoiding false positives (resulting in higher ROC curves).

4 Potential Applications

- Can be used to help identify affecting alleles in diseases.
- We are going to be using this to test for cardiovascular disease in Brazilians.
- Once affecting alleles are found, this can possibly help in discovering a cure or treatment.
- Also, as testing DNA becomes less expensive, people can be tested to see what affecting alleles for various diseases they might have, and thus, how often they should be screened for the diseases for which they are at higher risk.

5 Discussion

We tested three methods and found that the QATT method most accurately estimates when a genotype affects a quantitative phenotype. QATT compares the phenotype with the genotype and the average ancestry in the DNA. We are currently working to find the best method for associating genotypes and ancestry with binary phenotypes, such as Huntington's Disease.

Literature Cited

- Brisbin, A. et al. 2012. PCAdmix: Principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. 84: 343-364
- Pasaniuc, B. et al. 2011. Enhanced statistical tests for GWAS in admixed populations: Assessment using African Americans from CARE and a breast cancer consortium. 7: e1001371

Acknowledgements

This research was supported by the University of Wisconsin Eau Claire Small Research Grants Program.