

# REGIONAL ADMIXTURE MAPPING AND STRUCTURED ASSOCIATION TESTING: CONCEPTUAL UNIFICATION AND AN EXTENSIBLE GENERALIZED LINEAR MODEL

T. M. Beasley<sup>1</sup>, D. T. Redden<sup>1,3</sup>, J. Divers<sup>1</sup>, L. K. Vaughan<sup>1</sup>, H. K. Tiwari<sup>1,4</sup>, J. R. Fernandez<sup>1,2,3</sup>, R. P. Kimberly<sup>5</sup>, R. Feng<sup>1</sup>, M. Padilla<sup>1</sup>, N. Liu<sup>1</sup>, M. B. Miller<sup>6</sup>, & D. B. Allison<sup>1,2,3</sup>

<sup>1</sup>*Department of Biostatistics, Section on Statistical Genetics*, <sup>2</sup>*Department of Nutrition Sciences*,  
<sup>3</sup>*Clinical Nutrition Research Center*, <sup>4</sup>*Department of Pathology*, <sup>5</sup>*Division of Clinical Immunology and Rheumatology*, *Department of Medicine*,  
*University of Alabama at Birmingham, Birmingham, Alabama*  
<sup>6</sup>*Division of Epidemiology, University of Minnesota, Minneapolis, Minnesota*

Email: [mbeasley@uab.edu](mailto:mbeasley@uab.edu)

Individual genetic admixture estimates, determined both across the genome and at specific genomic regions, have been proposed for use in identifying specific genomic regions harboring loci influencing phenotypes in regional admixture mapping (RAM). Estimates of individual ancestry can be used in structured association tests (SAT) to reduce confounding induced by various forms of population substructure. Although presented as two distinct approaches, we provide a conceptual framework in which both RAM and SAT are special cases of a more general linear model. We clarify which variables are sufficient to condition upon in order to prevent spurious associations and also provide a simple closed form "semi-parametric" method of evaluating the reliability of individual admixture estimates. The reliability of individual admixture estimates is required to make an inherent errors-in-variables problem tractable. Casting RAM and SAT methods as a generalized linear model offers enormous flexibility enabling application to a rich set of phenotypes, populations, covariates and situations including interaction terms and multi-locus models. This approach should allow far wider use of RAM and SAT, often using standard software, in addressing admixture as either a confounder of association studies or a tool for finding loci influencing complex phenotypes in species as diverse as plants, humans, and non-human animals.