The Jackson Laboratory
Leading the search for tomorrow's cures

Gregory W. Carter
Associate Professor
207.288.6025 *t* | 207.288.6150 *f* | greg.carter@jax.org

August 30, 2017

Dear Editors,

My co-authors and I submit our manuscript entitled "A Bayesian framework for generalized linear mixed modeling identifies new candidate loci for late-onset Alzheimer's disease", as an investigation in *Genetics*.

Recent methodological and technical advances have greatly expanded genetic association studies in humans. The advent of low-cost whole-genome sequencing facilitates high-resolution variant identification, and the development of linear mixed models (LMM) allows improved identification of putatively causal variants. While essential for correcting false associations due to population structure, LMMs have generally been restricted to numerical variables. However, phenotypic traits in association studies are often categorical, coded as binary case-control or ordered variables describing disease stage or probability. To address these common study designs, we have created a generalized linear mixed model (*Bayes-GLMM*) approach. To efficiently address genome-scale variation, we have implemented this model using a Bayesian approach in the Stan programming environment.

As a pilot study, we applied our *Bayes-GLMM* to the whole-genome sequencing cohort in the Alzheimer's Disease Sequencing Project (ADSP). This study contains 570 individuals distributed across 111 families, each with Alzheimer's disease diagnosed at four confidence levels. The profound population structure in these data required a mixed model approach, and the categorical trait necessitated a generalized model. We applied our analysis to these data and found four novel candidate variants at three loci that achieved standard genome-wide significance. All four reside in intergenic regions and correspond to potential regulatory variants. Two of these variants lie between genes expressed in the neurovascular unit of the brain, which is an emerging source of pathogenicity in Alzheimer's disease studies, and transcripts of one have been associated with AD-related neuropathology in the ROS/MAP cohort. None of these loci achieved genome-wide significance without inference based on the categorical model we implemented. Therefore, while we recognize the clear limitations of inference in a small sample size, we believe this particular data set highlights the power of our *Bayes-GLMM* approach to provide novel candidate loci for further study.

To ease reproducibility and possible extensions of our analysis, we have used whole-genome variant calls determined by the ADSP consortium. While there is current

uncertainty in the consistency of calls within the ADSP whole-exome data set, our discussions with ADSP personnel have convinced us that the whole-genome data are unlikely to be affected by this quality issue. To provide further confidence, we performed our own independent variant calls from raw sequence data obtained from ADSP and found complete genotype concordance for all significantly associated variants. We also observed extremely high concordance across all variants, with observed differences due to slight differences in quality filter parameters. We are therefore confident in the data and the results of our analysis, with the abovementioned caveat regarding the small sample size.

This work provides the first implementation of a flexible, generalized mixed model approach. Furthermore, our paper would be one of the first to describe genome-wide analysis of the ADSP whole-genome sequencing cohort, providing a critical resource for the broader Alzheimer's disease community. Although our analysis alone may be insufficient to identify causal variants with high confidence, we feel that the work provides a framework for a broad range of genetic studies and demonstrates its effectiveness on an important data resource.

Thank you for your considering our manuscript.

Sincerely,

Gregory W. Carter, PhD Gareth Howell, PhD