Reviewer #1 (Comments for the Authors (Required)):   
  
In this paper the authors develop a Bayesian generalized linear mixed model for categorical trait GWAS, and propose a novel way to incorporate the prior knowledge about SNP effect into the analysis to increase power. Their model is overall statistically valid and the parameter inference takes advantages of the most advanced MCMC techniques in Stan, the platform they used to implement the model. This method is applied to the Alzheimer's disease (AD) using a whole-genome sequencing cohort. They showed by a mouse experiment that two putative genes identified from their analysis are functionally relevant to vascular dysfunction, which provides supportive evidence to their findings. Although the sample size is relatively small (n=570), the ability to better model the categorical trait for GWAS is shown. Both their modelling and findings could potentially be valuable contributions to the methodology of GWAS and the understanding of AD etiology. Overall this paper is well written, but clarifications are needed to better present the method. Please find my specific comments below.   
  
The presentation of linear mixed model is confusing (lines 125-140), with several typos. In line 130, "e~N(0,1)", but in line 139, it is said "σe followed inverse gamma distribution". If e~N(0,1) is assumed, does it result from the assumption that y is standardized to have variance one? If so, it needs to be stated. Another typo is that beta\_0 ~ 𝑁(0,1) but later they have sigma\_0 ~ inv\_gamma(2,1). Also, the authors assign a standard normal to beta, but contemporary methods commonly use a flat prior to avoid introducing shrinkage on the fixed effects. These need to be clarified. Xulong

Choosing priors are tricky. A Gaussian on effect size is a common practice (refer to Gelman’s BDA book). I was not aware of the phenomena that a Gaussian prior would introduce shrinkage on the fixed effects. I am running some simulations to understand this in more details.

The authors use logistic link in the GLMM. I wonder if they have investigated the effect of using a different link function, such as probit link function, which is equivalent to the well-known liability threshold model in animal breeding. Xulong

I have not test the probit link. Logistic link was what used in specifying ordered categorical models. I will run simulation test on this, to answer this question better.  
  
There is miscommunication in the ordered-GLMM part. Line 165, the minus sign should be plus. They said "theta = 10\*cumsum(theta\_0)". Why is ten? Is this coefficient data dependent? This needs to be explained. I wonder if it is equivalent to move the coefficient to be the parameter of the dirichlet prior for theta\_0. Xulong

There was a history on this. Dirichlet distribution ranges 0-1. I used cumulative sum of dirichlet to specify a vector of ever-growing components. Then, the cut values of an ordered categorical distribution of our model were an order larger than 0-1, which led to the 10 times configuration. Not sure whether this is data dependent. My gut is this is not data-dependent, but rather model dependent. Have to run simulations to investigate in more details.   
  
To incorporate the prior information of SNP effects, the authors model beta\_0 ~N(t\*sigma\_0, sigma\_0) but do not explain their choice to model the mean of beta\_0 in this way. I think they multiply t by sigma\_0 in order to standardize the prior knowledge (t) of the SNP effect to be in the same scale (i.e. sd) of beta\_0. If so, should it be t\*sqrt(sigma\_0)? Because for beta\_0 / t = sd(beta\_0)/sd(t) it follows beta\_0 = t\*sd(beta\_0) = t\*sqrt(sigma\_0) and E(beta\_0) = beta\_0 for just one beta\_0. Moreover, from a Bayesian perspective, sigma\_0 measures the uncertainty of one's belief on the SNP effect at the value of t\*sigma\_0. So linking the mean with variance could potentially be problematic, as it implies that the SNP with higher uncertainty on the effect tends to have larger effect size. Although it might not be a big issue in practice as sigma\_0 is predominated by the prior (which has mean 1) in their model, it should worth a clarification/discussion. Xulong

In both R and Stan, sigma means standard deviation, not variance. Might not be clear in the statements. It is interesting of the reviewer’s saying that “SNP with higher uncertainty on the effect tends to have larger effect size”. Think more on this.  
  
Line 194, L-BFGS is not defined. In the equation below phi is not defined. Xulong

I will borrow some words in Stan Manual to define L-BFGS.   
  
Line 200, "In MCMC sampling, SE(β0) was computed directly from the samples. A standardized z value was computed as β0 / SE(β0), which led to a P-value that quantified the probability of obtaining the β0 by chance." Is the standard deviation of MCMC samples used as the SE of the estimate? Note that the standard deviation of posterior samples may not have the Frequentist property of sampling variance with repeated data. So, the Frequentist interpretation for the p-value calculated from the posterior SE may not hold. Although the asymptotic normality of the posterior mean is approximated by invoking central limit theory, the consistency between the variance of posterior samples and the sampling variance needs to be justified by simulations based on their data. Xulong

Reviewer is an expert in Bayesian statistics. Bayesian p-value is what we computed. There are some papers on this which I will compile together. Bayesian p-value is essentially the tail probability of the parameter’s tail probability. Definitely needs a good paragraph to reply this comment.  
  
Line 220, any reference to this equation? Without term (1− 𝑔𝑚,𝑖)∗(1− 𝑔𝑚,𝑗), it is the VanRaden et al. (2008)'s G matrix if g is centred. Xulong

Right, I will pull out the literature on this.  
  
Line 312. The statement about negative selection is not convincing. It is not clear whether the larger effect sizes of rare variants are simply due to sampling, since larger sampling variance for rare variants is expected in GWAS. The authors could overlay a power curve as in Figure 1 of Marouli et al. (2017) to show the excess of the effects of rare variants in contrast to sampling. In addition, if there is negative selection, the alleles that are deleterious and therefore kept at low frequency must be risk-increasing alleles. The authors should plot the effects of the minor alleles against their frequencies, and examine if most (or the mean effect) of the minor alleles with low frequencies have positive effects on the AD risk (see Figure 4 in Yang et al. 2015 for an example). Greg

This is a very interesting topic, Greg.   
  
Are the identified genes overlap with any approved drug targets (e.g. those in Drugbank and Therapeutic Targets Database)? This might be useful for drug repositioning. Greg  
  
Line 356, "In GWAS, prior information of a variant can be implemented with multiple strategies, each allowing posterior estimations to carry different weights of the priors." What are the strategies? Any reference? Xulong Greg

I did some tests trying different methods. Basic idea is to put different weights on prior and data. Variances was the major thing, because posterior variance will always be smaller than the prior variances, so a GWAS results of large sample size (small variance) will always make the posterior variance very small. I will put together more words, and some simulation results on this.  
  
It is not clear what data is used as prior information in the study. To demonstrate any advantage, it would be good to show the GWAS results with and without incorporating the prior information. Xulong Greg  
It was the Lambert paper. We did not show that because the Lambert paper did not agree with the current ADSP data so much…

For a rigorous analysis, more details about the MCMC implementation should be shown, including the length of chain, burn-in, and assessment of convergence. Xulong

Yes, I will put more details on this in the Method part.   
  
Line 437, the second "possible" should be "probably". Greg  
  
The authors mention GMMAT is much faster than their methods but can only fit binary data. I wonder how different the results would be between their binary-GLMM and GMMAT. Xulong – have you done this or is this easy? This is basically the same as using Bayes-GLMM with a binarization, which I think maybe you did? We could refer to that instead.

Yes. I believe we did both. I will pull out some results on this.

Line 478, should be Figure 7. Xulong

Correct.  
  
In Figure 2, are the shown the 95% HPD of the posterior distribution? It is also not clear how they estimate the cut points. Xulong

I believe it was 95%. Cut points were only parameters of the models, so those were estimated together with other parameters.  
  
The authors report the significant SNPs at a genome-wide significant threshold of 5\*10^-8. However, given the sequence data (~10M SNPs), rigorously speaking, the genome-wide significant level should be at 5\*10^-9. I suggest the authors to report both. Greg  
  
  
  
  
Reviewer #2 (Comments for the Authors (Required)):   
  
This manuscript applies a novel genetic association analysis approach to a whole-genome sequencing Alzheimer's disease cohort. The manuscript is very well written and clear. For this review I have been asked to review the Alzheimer's disease aspects.   
  
The Alzheimer's Disease Sequencing Project (ADSP) data is used as a test cohort, where the association method must overcome categorical disease variables, sample relatedness, population substructure and prior knowledge integration. The categorical disease variables of AD diagnosis (including no, possible, probable and definite) are generally collapsed into case-control status, and this work shows the importance of taking these groupings into account. This work looks to overcome major challenges relevant to GWAS analysis as a whole, therefore the study is of great general interest in the wider genetics community.   
The authors give a short but comprehensive overview of the genetics of LOAD including referencing all recent large scale GWAS studies and the exome sequencing studies which identified rare variants in TREM2.   
Using this whole genome sequencing data, four novel non-coding variants, in three loci associated with AD were identified using the Bayes-GLMM methods (P<5x10-8). 28 loci were associated with p<1x10-6. Many variants mapped to genes in biologically plausible disease associated pathways, including two previously AD associated genes SLC24A4 and GABRG3.   
The associated PRKAT1B and PDGFA gene regions were followed up by investigating brain expression in mice using immunoflurescence, and in humans using post-mortem RNA sequence data. Expression was correlated with plaque burden in a key brain region. This work adds to the increasing evidence that vascular dysfunction is a critical component of AD pathogenesis. They highlight the potential for the identification of novel disease mechanisms and therapeutic targets.   
  
In terms of novel findings for AD, this result highlights potential novel pathways in AD pathogenesis, but this small sample size requires replication in larger datasets. This would be possible though available whole genome sequencing data from ADNI.

Greg I’ll have another look into the ADNI data, but snp-based GWAS usually don’t report on this region. We’ll need to wait until there is more whole-genome seq data on AD.