Baylor College of Medicine
CENTER FOR STATISTICAL GENETICS

Suzanne M. Leal, Ph.D.
Professor
Director, Center for Statistical Genetics
Department of Molecular and Human Genetics
One Baylor Plaza, Suite 700D
Houston, Texas 77030-3498
TEL: (713) 798-4011
FAX: (713) 798-4373
E-mail: sleal@bcm.edu

January 28, 2014

Prof. Gert-Jan van Ommen
Editor in Chief, European Journal of Human Genetics

Re: Manuscript 699-14-EJHG

Dear Gert,

Thank you for the opportunity to resubmit our manuscript *Collapsed Haplotype Pattern Method for Linkage Analysis of Next-Generation Sequence Data*. This revision presents a more thorough investigation on the type I error of our method. As suggested by reviewer 3, we used sufficiently many null replicates to match the scale of genome-level applications and compared the statistics with the expected distribution by generating QQ plots. This led to the addition of Figure S1 with slight modification in the wording of the article (text in red). The results and conclusions of our study remain unchanged.

Below we provide detailed response to reviewer 3's question regarding type I error. Please do not hesitate to contact me if there are any additional concerns.

Sincerely,

Suzanne. M. Leal, Ph.D.
Professor
Department of Molecular and Human Genetics

REVIEWER 3:

The authors have made alterations to the manuscript in response to the 3 critiques. These changes have improved the manuscript by clarifying the method and providing additional information about its performance. The new figure is nice. The approach likely will find practical usage, which in my view is the most important aspect of the paper.

We thank the reviewer for the positive comments.

My main reservation continues to be the relatively superficial assessment of the performance of the method under the null hypothesis. The authors have evaluated this issue now, but there is only one sentence in the manuscript: "Empirical type I error estimates are constantly zero for all tested scenarios, assuring that there is no inflation of the test statistic in the presence of within-gene recombination, strong inter-marker LD or missing genotype data." This statement is unlikely to be true. What is likely the case is that the authors did not conduct sufficiently many replicates to estimate the false positive rate reliably. Yes, at a LOD of 3.3, when conducting 500 null replicates it is likely that not a single significant test will be observed, leading to a numerical estimate of 0 for the false positive rate.

We agree with the reviewer that many more null replicates should be used. In the revised manuscript we increased the replicates to 2,000,000 to evaluate type I error of our method. We also improved the modeling of recombination events by using the recombination rates obtained from the *Hapmap recombination rates and hotspots* database. The update led to a more reliable numerical estimate of type I error rate ($\hat{\alpha} = 2.8\times10^{-5}$, 95% CI: $2.11\times10^{-5} \le \alpha \le 3.63\times10^{-5}$), demonstrating that type I error is well controlled, and even conservative, at a required α level (HLOD of 3.6 corresponds to a p-value of 4.7×10^{-5}). Additionally we have included Figure S1 showing the distribution of HLOD under the null as well as QQ plots under various scenarios. Changes to the article have been made on lines 150 – 155.

The problem with leaving it at that is that in reality it will be necessary to conduct genome-wide analysis, with its concomitant multitude of tests, and hence it is not trivial whether the false positive rate is along expectation or elevated. At the present time, we do not know the answer. At the very least, the authors could compare the obtained p-values under the null hypothesis to expectation, e.g. by generating QQ plots.

We have generated QQ plots of test statistic under the null and we confirm that in the presence of recombination events and missing data the false positive rates are controlled. The plots are added to Figure S1.

Table S1 is useless and could/should be omitted.

We have removed Table S1 from the supplemental material.