Reply to Celia’s comments:

1. **Line 74** This does not flow logically for me. I think you should first define notation. Then the question becomes why are GESAT and iSKAT introduced first? (moved the notation up here) Perhaps the logic should be: suppose one wants to perform a GxE test with measured and observed E and multiple SNPs. (added the reasoning here) Then one would proceed as follows - as implemented in GESAT and iSKAT. The section would then be titled "Interaction tests with measured $E$". Then another section "Interaction tests when $E$ is not available. I added explanation on why GESAT and iSKAT are called direct tests and our method are called indirect. I'm not sure why GESAT and iSKAT are being introduced first.

I followed other GxE test paper. They all introduced the existing methods, that will be compared with the proposed method first. And the connection would be, existing method cannot solve certain problem, or has certain defects.

1. **Line 123** This seems like a result, not something you would say in Methods unless it follows from the formulas. Can you express this differently? Yes I move the power problem into discussion and change the expression here by: In the simulation, we will study how the assumption affect the performance of mvgS with absolute residual regarding its type 1 error and power.
2. **Line 128** p suddenly appears undefined. What is being transformed here? the residual?

**Line 131** paragraph needs a lot of work. You have the issue of power, type 1 error (edited these part) and transformations all floating around. Levene’s test first compute , subtracting the group mean from each observed phenotype (similar to removing main effect by regression) then compare the between group variation vs the within group variation. The power transformation is to compute and then apply the rest of Levene’s test. It is not easy to explain what is here without mentioning the formula’s of Levene’s test. I marked this part blue for further editing.

(k=4 is it generalizable to larger k)

Try to reorganize to be more straightforward. Make sure to say that there will be advantages and disadvantages, that the best choice is unknown and that is why you do simulations in section 3. (added)

1. **Line 132** After completing stage 1 and choosing the residual transformation (The choice of second stage statistics is not independent from the choice or residual, I edited as: After completing stage 1 with a chosen form of residual transformation), we would like to choose a stage 2 test that can capture an increase in variability associated with the genotypes at J SNPs. In gS as well as in SKAT and SKAT-O, an F-test from an analysis of variance was used, and we take this choice also here.

(combination -> correct T1E -> filter out -> power , ealier research has already demonstrated the absolute value has correct t1e control. For completeness we also investigate the squared residual.)

1. **Line 152** not true, see the paper by Jianping Sun (having trouble finding this paper) candidate SNPs for bone mineral density, not from the same region. Gene based LD based.
2. **Line 160** put into a table not into sentences. (Edited) Also these numbers appear to come from thin air. Why these numbers? Could you generate beta from a distribution that changes each time? Yes these numbers are generated from a product of a normal random variable and a binomial random variable. I generated these before and fixed them.

Enrich the range of betas

1. Get real data diff Gene 0 diff LD, diff pattern. Larger n. (exome sequenced variance from the 3500 individuals. Subset of genes 5-50 that are not )

beta1<-matrix(rnorm(p,mean=0.01,sd=0.02)\*rbinom(p,size=1,p=0.6),nrow=p)

beta3<-matrix(rnorm(p,mean=0.1,sd=0.2) \*rbinom(p,size=1,p=0.4),nrow=p)

I tried to follow the coefficients in iSKAT paper, they use the estimated coefficient from a previous data analysis paper and J=11. But in that case, mvgS barely has any power. That’s why I generated a new set of coefficients. The mvgS and the direct tests are comparable with data generated from our set of coefficients.

Common or rare – 3000 acknowledge that the

1. The structure of Section 3 Simulation is changed. Do I need to add another paragraph above the subsections to introduce what are included and why they are include? Or I just start each subsection with introductions?