

# WHI Spurious Association Simulations

September 21, 2017

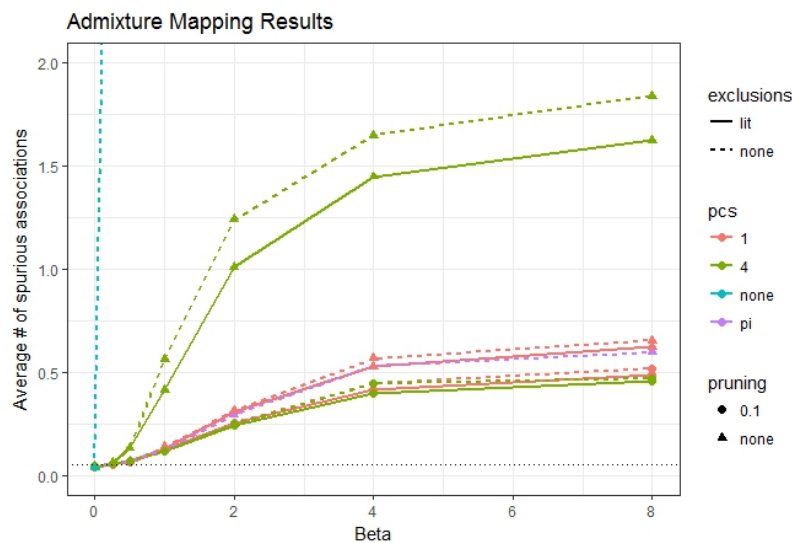
## Reminders from Last Week

### Simulation Setup

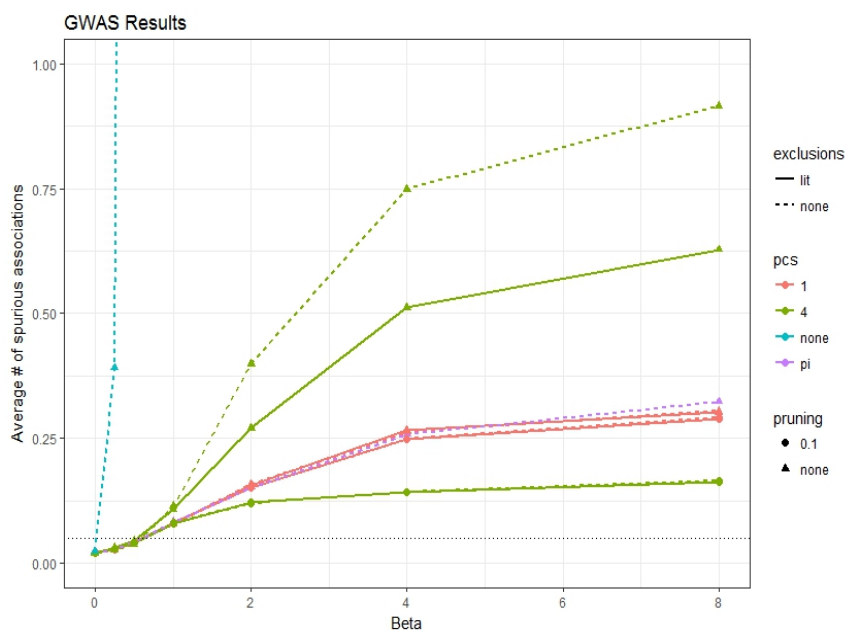
We use real genotypes from WHI and simulate traits with a single causal SNP ( $Y_i = \beta X_{ij} + \epsilon_i$ ), and we quantify spurious associations by counting the number of other chromosomes with at least one p-value below the designated significance threshold.

### Overall Admixture Mapping and GWAS Results

**Figure 1.** Visual depiction of admixture mapping results in Table 1.



**Figure 2.** Visual depiction of GWAS results in Table 2.



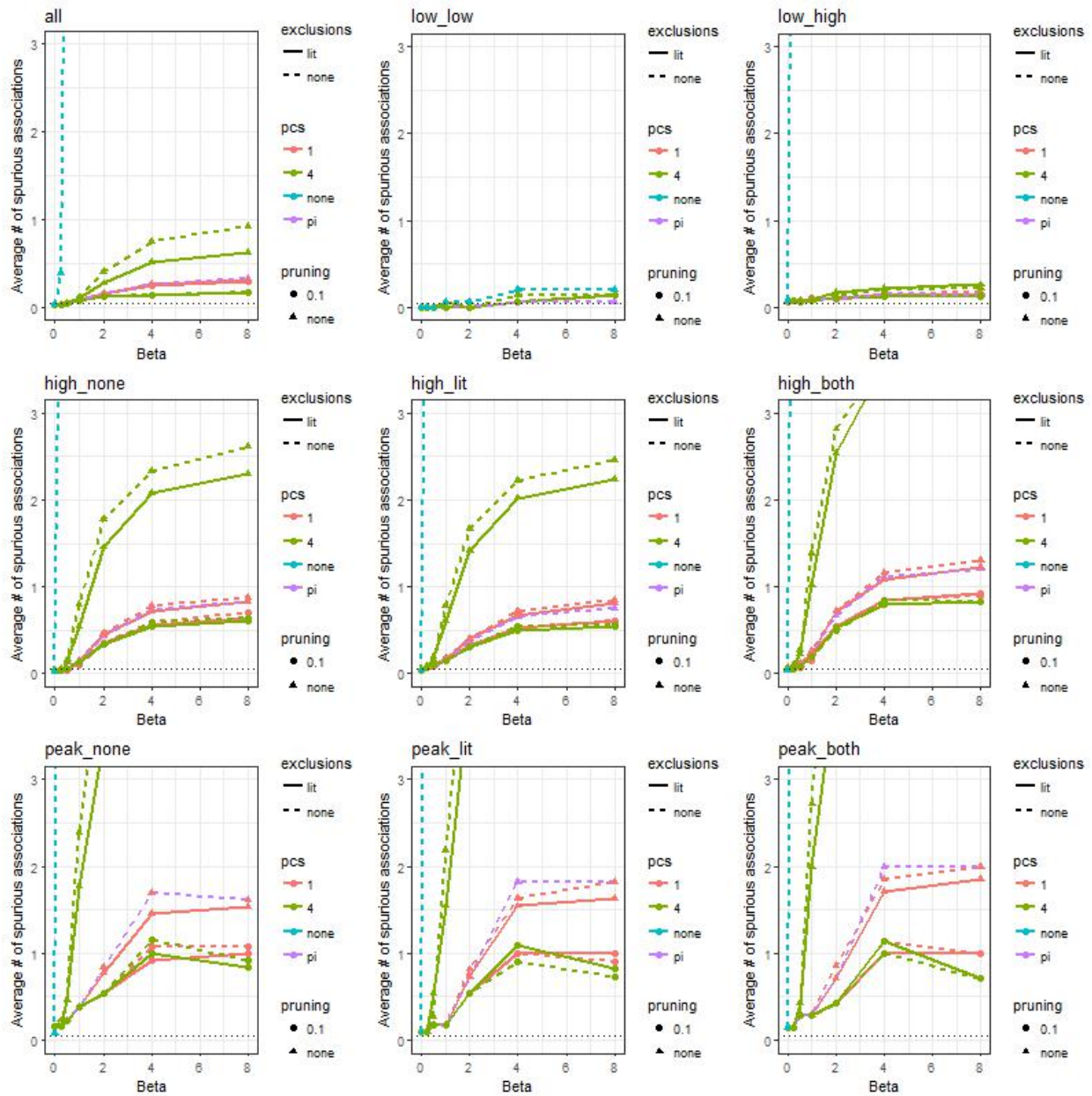
## New Results: Breakdown of Spurious Associations by Simulation Setting

### Types of Simulation Settings

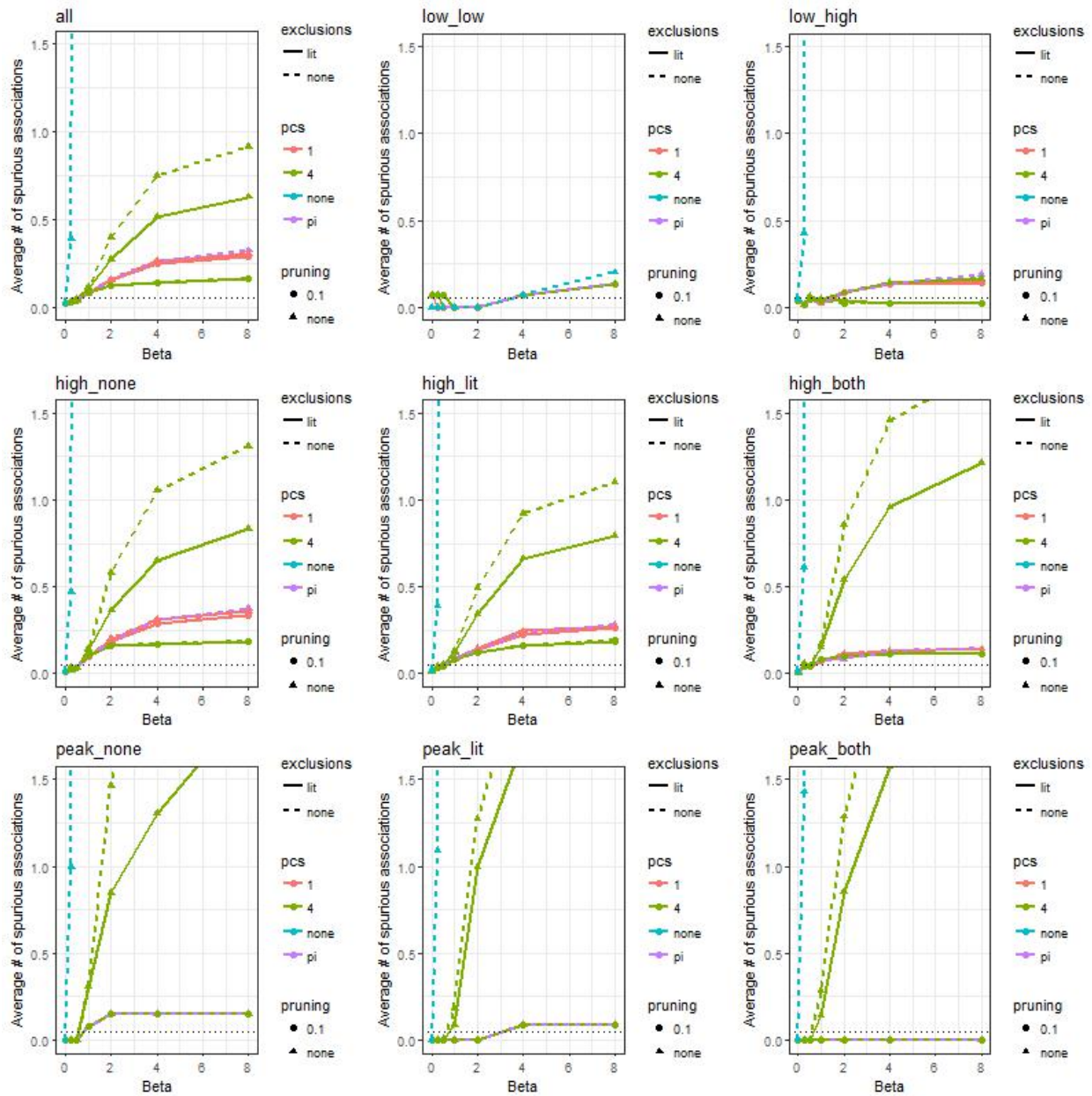
In the figures below, I present simulation results broken down according to nine categories:

- *all* = all 473 settings
- *low\_low* = low SNP loadings, low difference in ancestral allele frequencies (15 settings)
- *low\_high* = low SNP loadings, high difference in ancestral allele frequencies (85 settings)
- *high\_none* = high SNP loadings across any of PCs 2-4, based on PCs with no exclusions (220)
- *high\_lit* = high SNP loadings across any of PCs 2-4, based on PCs with lit based exclusions (220)
- *high\_both* = high SNP loadings across any of PCs 2-4, based on both sets of PCs (70)
- *peak\_none* = highest SNP loading for one of PCs 2-4, based on PCs with no exclusions (13)
- *peak\_lit* = highest SNP loading for one of PCs 2-4, based on PCs with lit based exclusions (11)
- *peak\_both* = highest SNP loading for one of PCs 2-4, based on both sets of PCs (7)

## Admixture Mapping by Setting



## GWAS by Setting



## Conclusions

### Overall Conclusions

1. The number of spurious associations increases as we increase the effect size of the causal SNP, for both GWAS and admixture mapping, and for all choices of covariate adjustment
2. The best approaches: estimated global ancestry, 1 PC (w/ or w/o LD pruning), 4 PCs w/ pruning
3. The worst approach: no covariates
  - a. We also see poor performance if we use 4 PCs w/o LD pruning
4. Patterns are generally similar for admixture mapping and GWAS, with a few slight differences:
  - a. GWAS shows slight advantage to using 4 PCs w/ LD pruning over estimated global ancestry or 1 PC
  - b. Admixture mapping showed slight advantage to using 4 PCs or 1 PC w/ LD pruning over estimated global ancestry or 1 PC w/o pruning
  - c. We see fewer spurious associations with GWAS, but this is probably because the threshold is too conservative

### By Setting Conclusions

1. When the causal SNP has high loadings: we see more spurious associations for all approaches
2. When the causal SNP has low loadings: all approaches (except “none”) control the average number of spurious associations pretty close to 0

## Next Steps

1. Write up results into a paper
  - a. Goal: have a draft ready before we go to ASHG
2. Start preparing ASHG poster
  - a. Goal: have a draft ready by October 5
3. Eventually:
  - a. Consider other approaches for adjusting for population structure (e.g., ADMIXTURE, mixed model based approaches)
  - b. Run a similar simulation study using simulated genotype data

## Administrative

### Fall quarter meeting times

I'll be TAing for BOST 310 this quarter, which has a discussion section on Thursdays from 1:30-2:20pm. I'm not sure yet if I'll have to go to that discussion section, but if I do then I won't be able to meet at 2pm on Thursdays. Some other options:

- Thursday, 2:30-3:30pm
- Friday, any time after 1pm

### Upcoming lab meetings

I will be out of town October 9, which I think is the week I was scheduled to present?