

Draft: Investigation of High Heterozygosity Variants in Freeze2

AACHueva

July 14, 2016

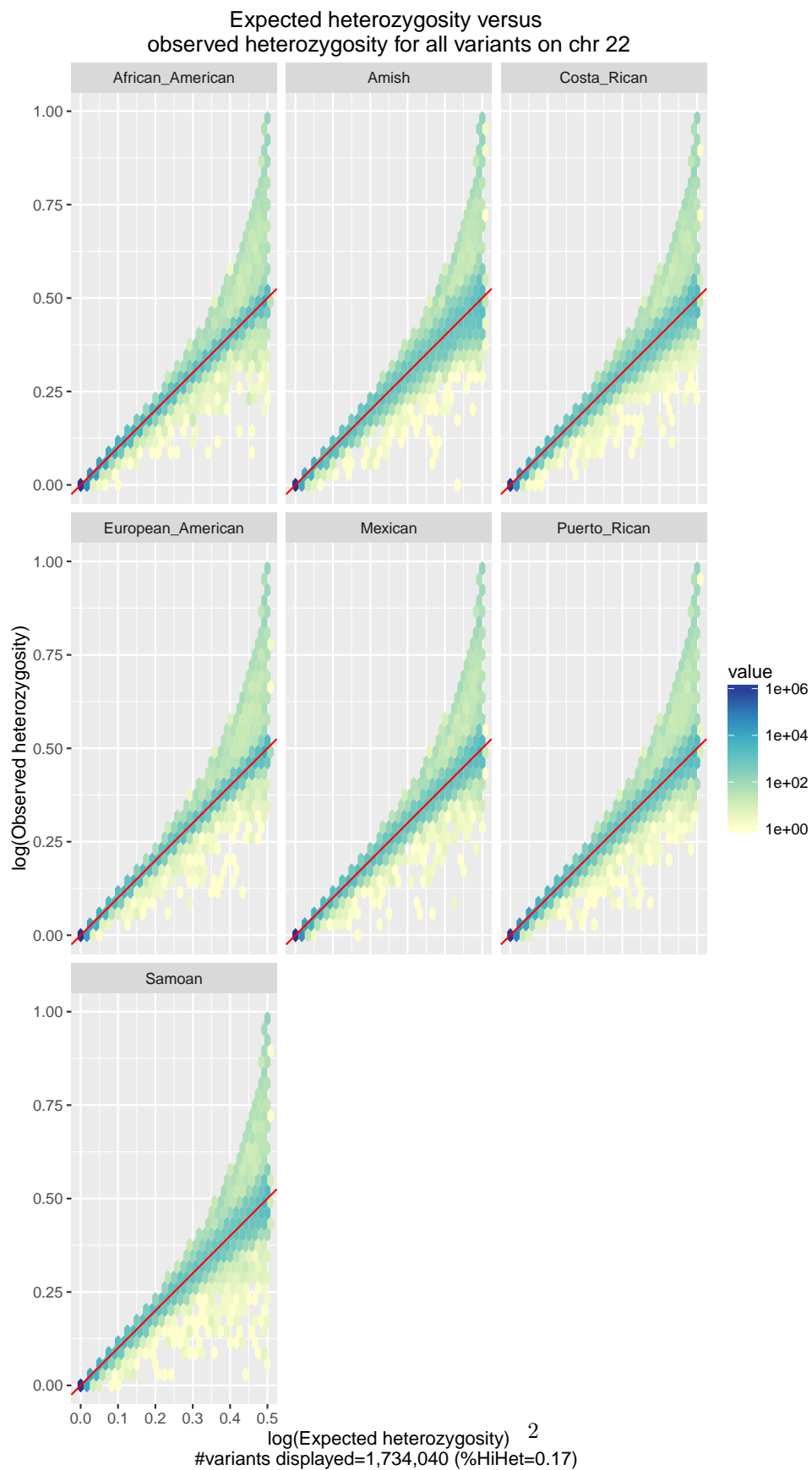
Files that were used in the analysis:

1. File with TOPMed InDel Annotations (courtesy of Xiaoming Liu)
2. Feeeze 2 GDS GT only (includes all chromosomes)
3. HWE results for each ancestry (courtesy of Stephanie G.)

Pre-work that was done:

- * Created a dataframe with variant.id, chr, pos, ref, alt, MAP20 and MAP35 fields.
- * Extracted variants only for chromosome 22.

Plot 1.Expected heterozygosity versus observed heterozygosity for all variants on chr 22



Fraction of high hets over all chromosomes

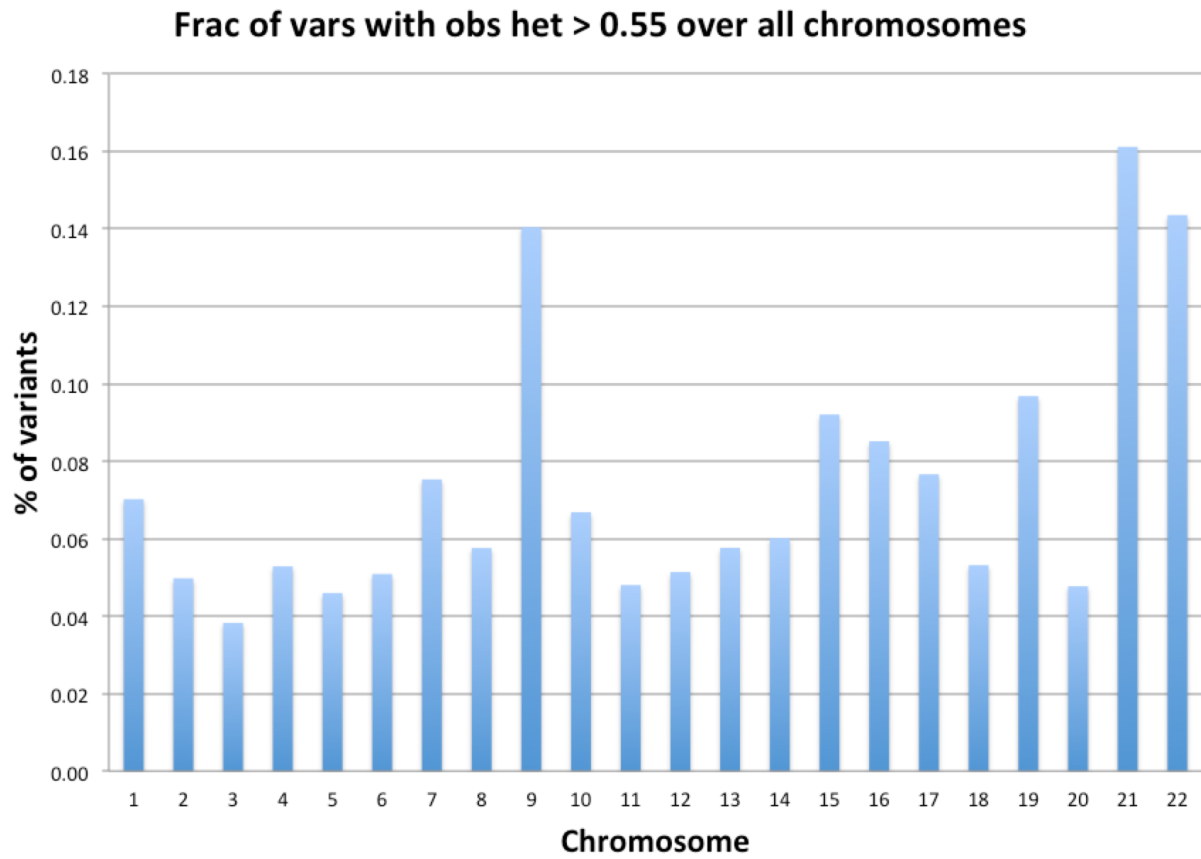
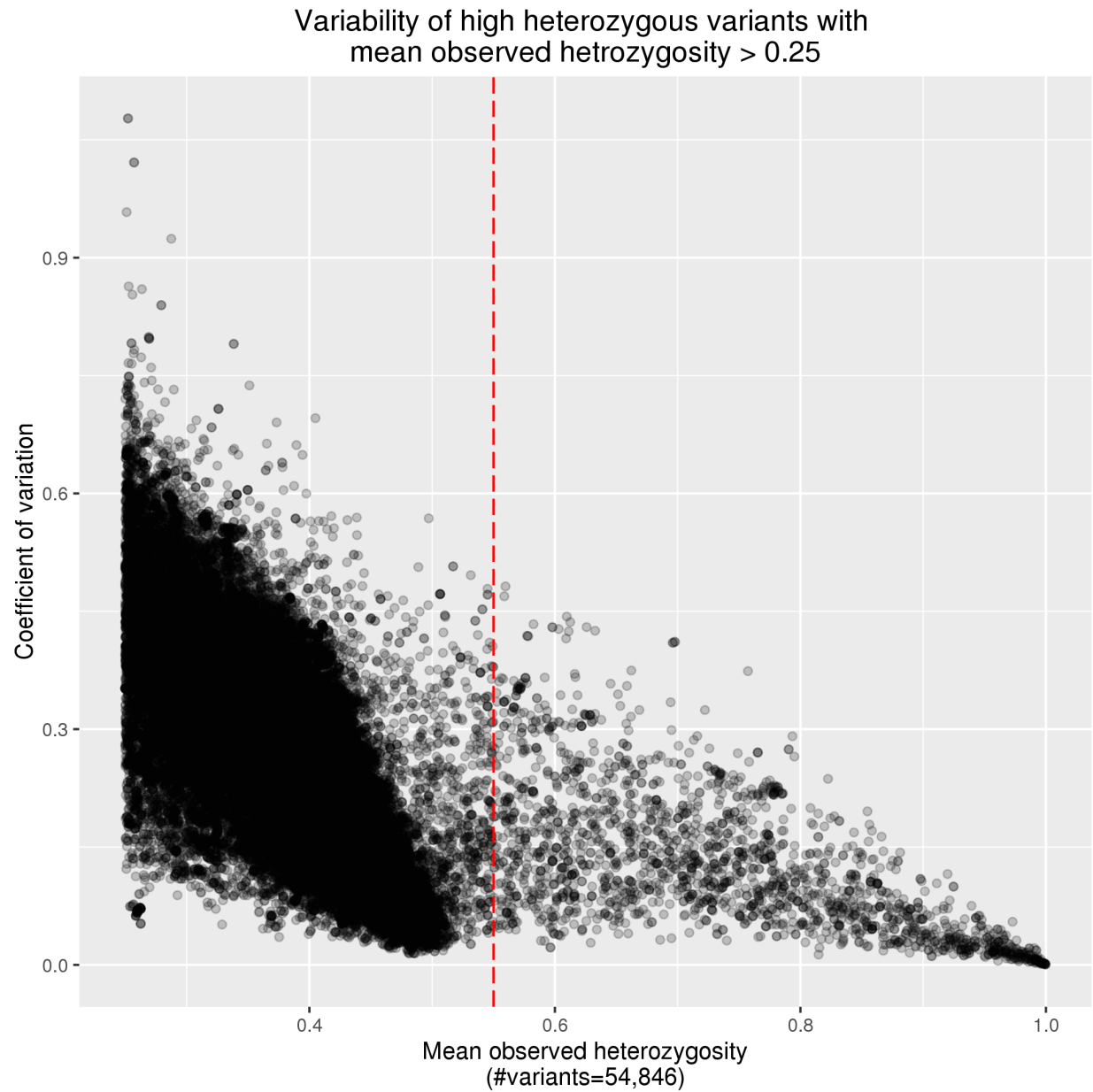
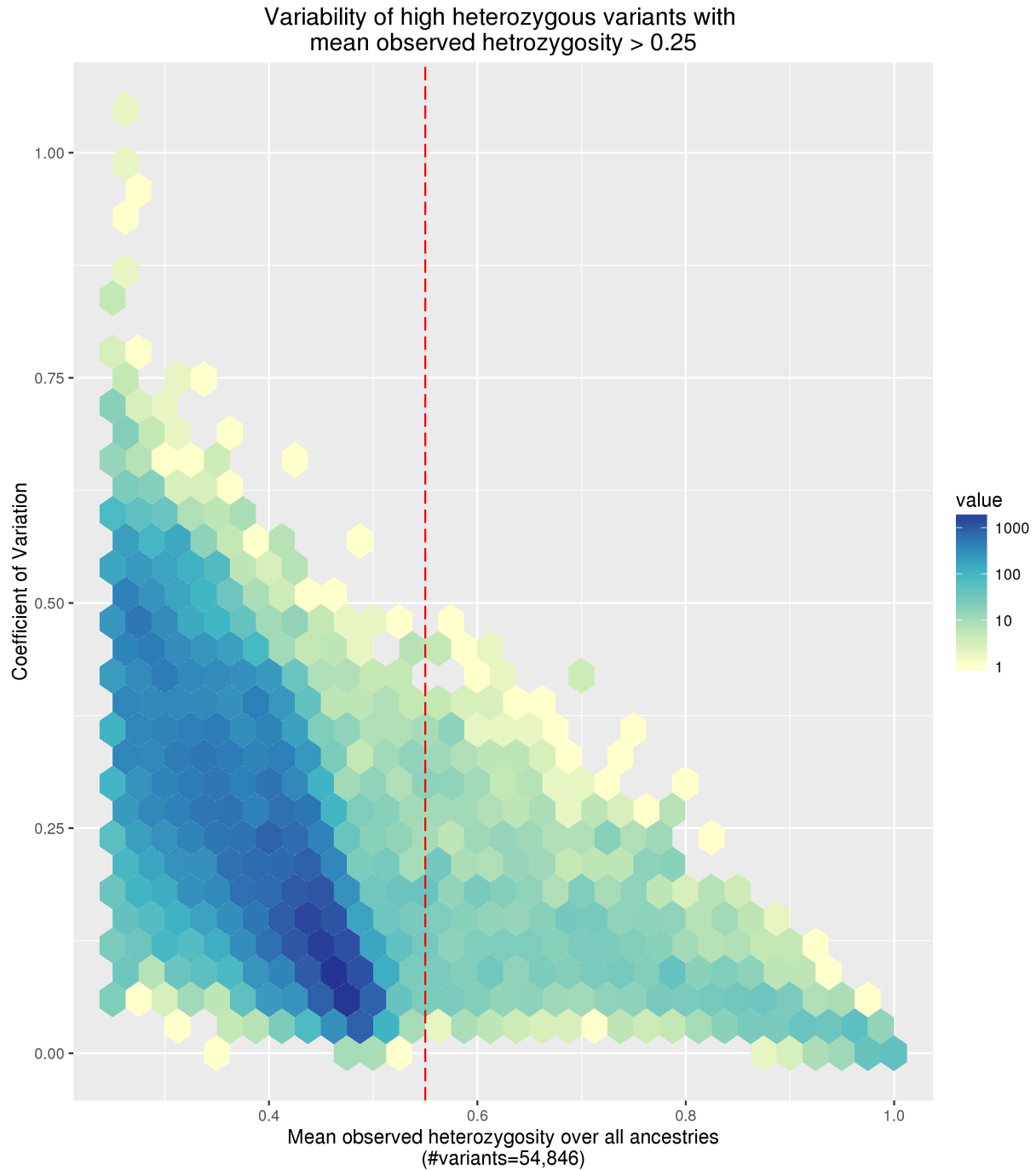


Figure 1:

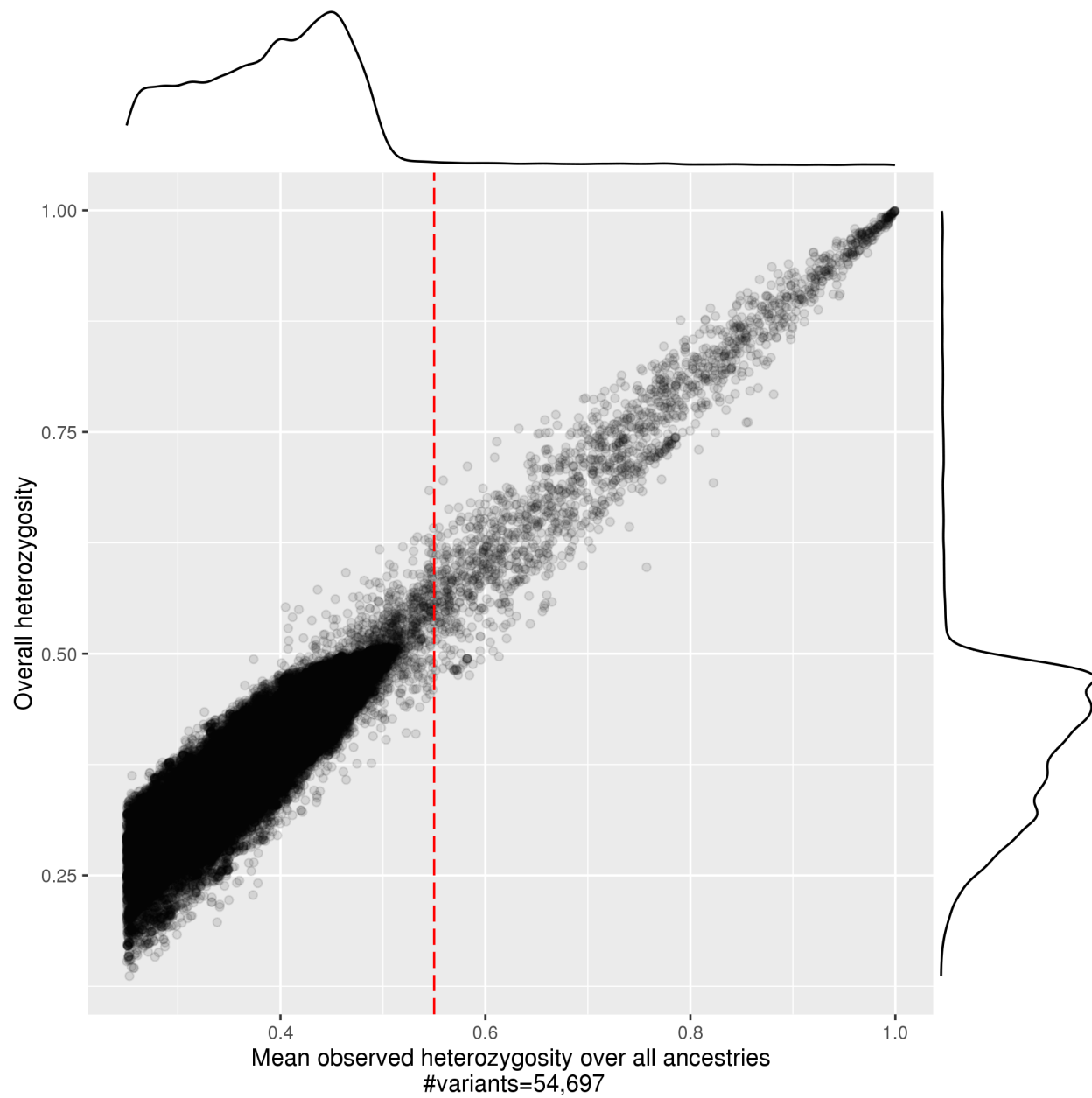
Plot 2. Mean observed heterozygosity versus coefficient of variation (among 7 ancestry groups) for all variants on chr 22 with observed heterozygosity > 0.25



Plot 3. Mean observed heterozygosity versus coefficient of variation (among 7 ancestry groups) for all variants on chr 22 with observed heterozygosity > 0.25 (Density)

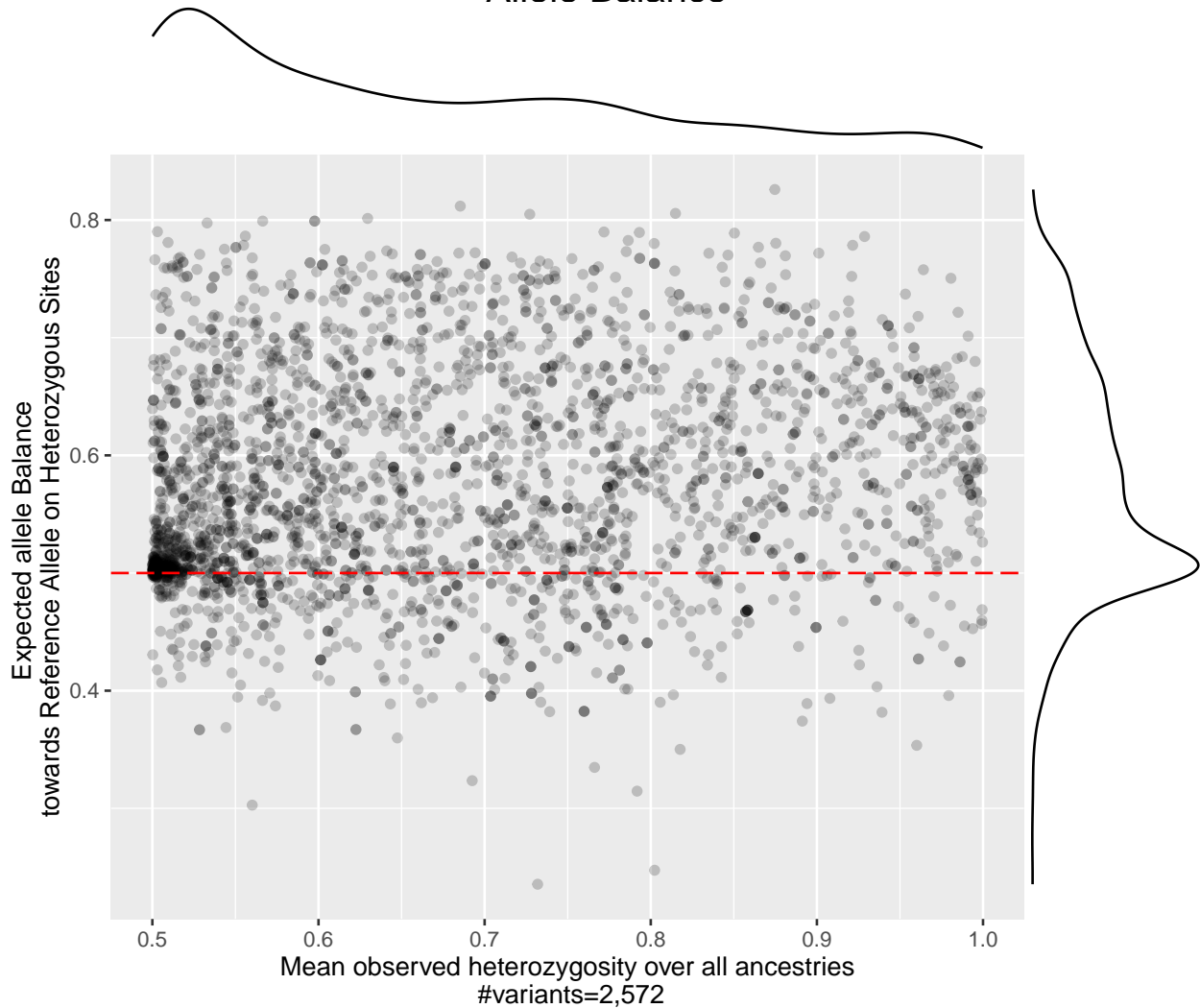


Plot 4. Overall observed heterozygosity versus mean observed heterozygosity (over 7 ancestry groups) for all variants on chr 22 with mean observed heterozygosity > 0.25



Plot 5. Allele Balance versus mean heterozygosity

Variability of high heterozygous
variants with mean observed heterozygosity > 0.5:
Mean observed heterozygosity vs.
Allele Balance



ABE is (reference allele count)/(reference allele count + alternate allele count), averaged over heterozygous genotypes