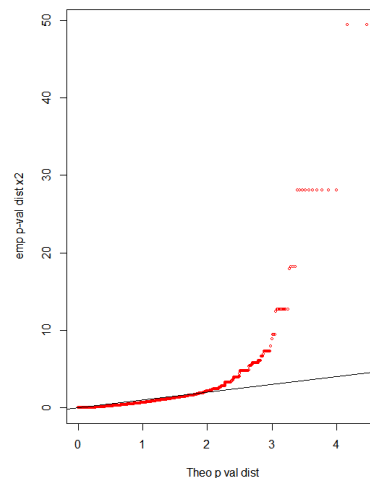1. With the pan bed object you have created in the previous practical, calculate for each locus its inbreeding coefficient; use it to test whether the loci are in Hardy Weinberg Equilibrium; plot −log10 of these p-values against their expectations under the null hypothesis.

Is it what you would have expected?
**Yes, given that the data contains subpopulations that are likely in different HW equilibriums.**

2. redo the same but for loci with minor allele count of at least 10 (maf ≥0.01; this filtering "rule" is often used in genomic analysis). Identify, using e.g. a different color, the loci that are not in HWE according to this test on the plot of observed heterozygosity against allele frequencies.
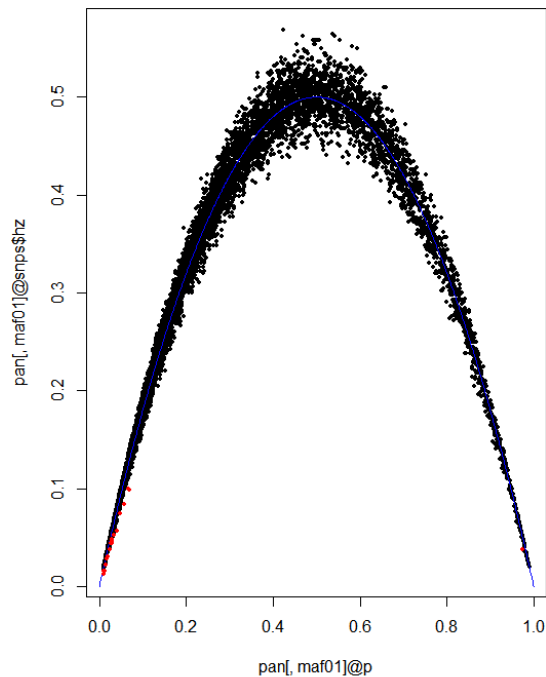
Do you see a pattern?
**There is a pattern of those loci not in HWE are at the extremes (o and 1) of the MAF distribution.**

a. A rule often stated for the validity of a χ2-test is The minimum expected number per cell is 1 (5) [the proportion of cells with expected counts lower than 5 should not exceed 20%]
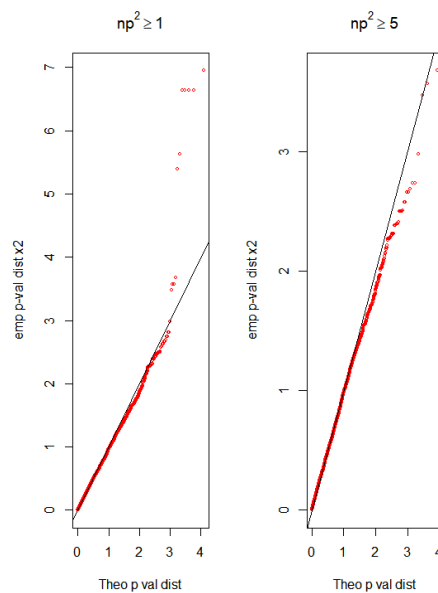

Is this rule working here?
Yes, since we have no more outliers.

3. Now load (install it if not done yet) the HardyWeinberg library, and use the function HWExactStats to obtain the exact p-values for these loci, using first the argument midp set to FALSE and then set to TRUE.
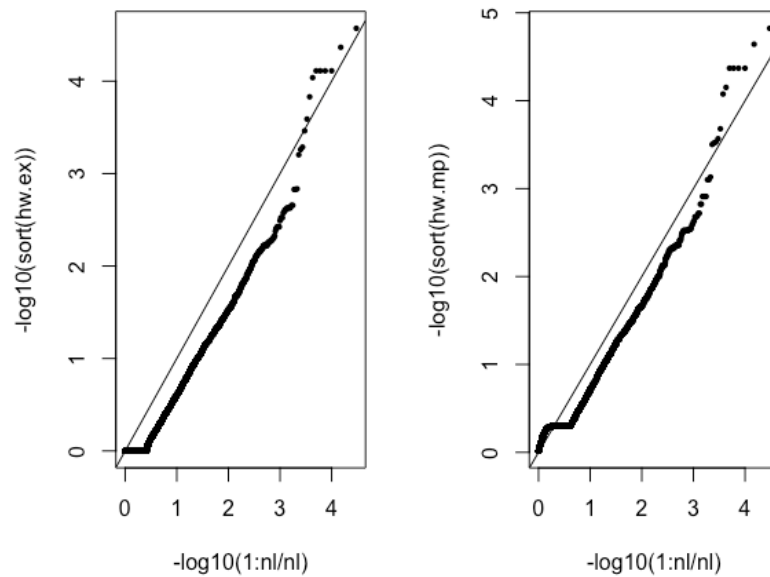

4. using the East Asian samples from the 1000 genome project, plot the SNPs heterozygosity against the frequency of the alternate allele for chr22:0-20M. Add to the plot a line showing the expected heterozygosity under Hardy Weinberg Equilibrium;
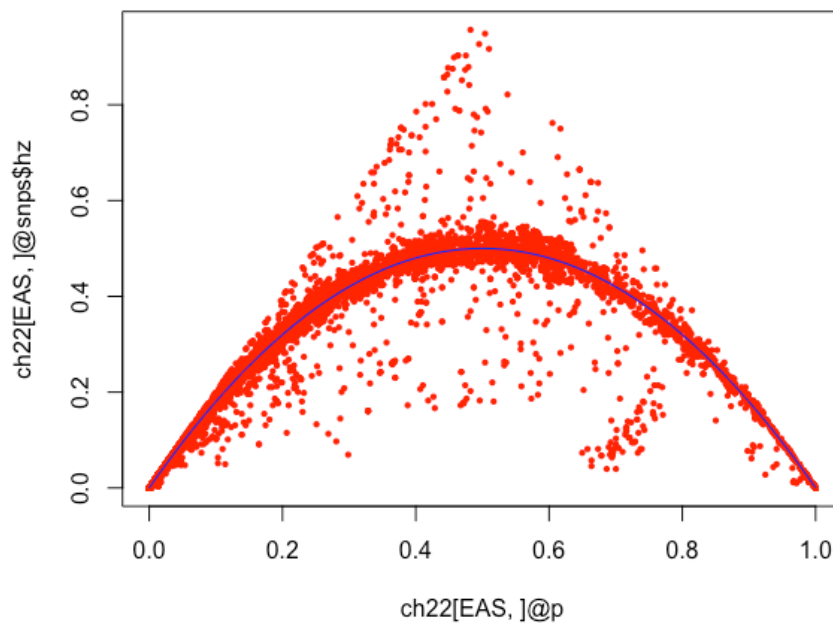
Discuss the resulting figures, compared to what we saw with the simulated data. (reference allele frequencies on x, heterozygosity on y)

We see a lot more spread from the expected heterozygosity. There are loci at the limits of heterozygosity. These loci might be a problem for demographic inference, but are interesting biologically. The loci under the curve have lower than expected heterozygosity (heterozygote deficiency), and are likely homozygous.

5. Test for Hardy Weinberg using the exact mid-pvalue test for all these loci.
6. plot the p-values of the previous tests against their expectation under the null (Rather than the p-values, -log10 of the p-values is more informative).
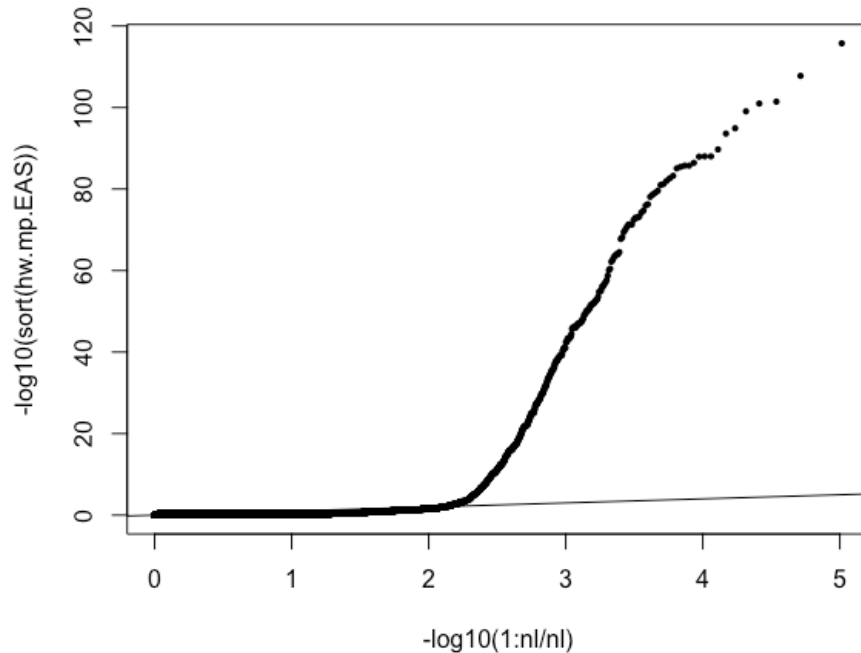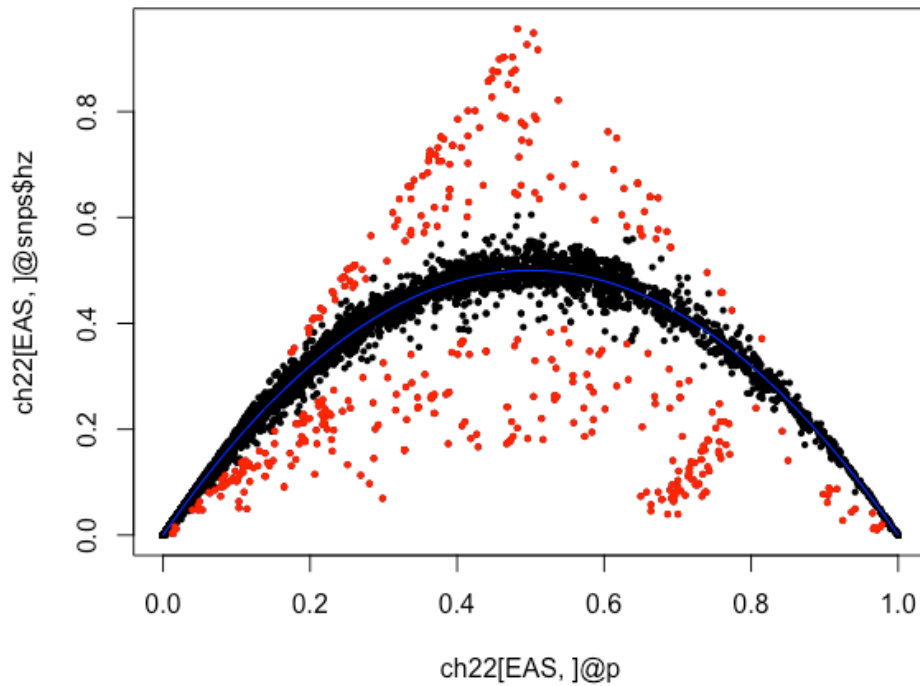
Identify on the plot of heterozygosity against allele frequencies the loci not conforming to HWE, and discuss the results.

7. **[optional]** How likely are we to detect departure from HW if f=0.125 with a sample of 100 individuals?

8. **[optional]** How many individuals would be needed to detect departures from HW when f=0.125 with a power of 0.8?

Good working with you all! -Tim
Same! -Elizabeth