**2021/04/10 Group meeting**

Questions

Q2. What are “p values”? Statistical test p values or frequencies of alternative allele?

Google doc link

<https://docs.google.com/document/d/1b5lXxmQfwsy30Z6bV-NkHo4IfkD7EExhvjyvMhtYcGw/edit?usp=sharing>

Google folder:

<https://drive.google.com/drive/folders/1OQE7PiZvGxOWcV4uY46UZhPVL1rqEF6j>

**2021/04/22 Bingshan’s office hour**

Presentation:

~30 minutes

1. A team leader collects results from every team member together, compare within and cross groups

2. Everyone can present, take turns

3. Do not need to show the code. Just need to present visualizations (histogram, etc.)

**2021/04/29**

Bingshan’s feedback.

Q1. AF

* If we have infinite number samples, what does the first bin might look like?
  + When zoom in we can see the AF is skewed to the left in the first bin
  + (Why)

Q2. HWE

* Why do you see deviation from HWE? (Mixed population)
* HW test is conservative. So if everything in perfect HWE, then in the QQ pot the observed values will be below the expected line
* QQ plot: looks like p values follow uniform distribution, if nothing deviates from HWE
  + We use log because we care more about variants with small p values (double check???)
  + If you only see deviation from expected line at the end of the tail, then

Q3. Why there are a lot of large D’?

* A lot of SNP has small allele frequencies, so you are likely to not observe al hapotypes (cannot observe enough recombination, not due to sample size in this data)
  + Bingshan mentioned if sample size is too small, all D’ values will be affected. (This make sense, in Maggie’s slide it says “Large upward bias for small sample sizes”)
* If you remove SNPs with small frequencies, plot and see what D’ you get
* D does not have much information. D’ and R2 are different ways to normalize D.
  + D’ measures recombination rates
  + When R2 is 1, D’ is always 1, but not vice versa

Q4. PCA

* (My) PCAs from different packages might be different, due to sign of coordinate system