

Applied population genetics

Visualisation and understanding

BIO392 practical session

Qingyao Huang

Data sets (from 1000 Genomes)

- All populations together (2504 individuals)
 - 661 of them: African (AFR)
 - 103 of them: Han Chinese (CHB)
- Only chromosome 20, small chunk 1- 3.2Mb

Create file for subset populations

- Use “igsr_samples.tsv” and find the IDs of desired population subgroups → create a file of IDs
- Use **--keep-fam** from plink and generate new files
 - HAN prefixed files
 - AFR prefixed files
- Use **--maf** to filter file
 - HAN.maf prefixed files
 - AFR.maf prefixed files

Histogram of minor allele frequency

- Calculate frequency for all 3 populations
- Use the xxx.frq file and plot histogram in R
- How do they differ?

LD decay

- Generate pairwise correlation with **--r2**
- Additional parameters: **--ld-window-r2 0** and **--ld-window 300**
- Make use of the R function - plotDecayLD

Pairwise LD plot

- Generate pairwise correlation with **--r2**
- Additional parameters: **--ld-window-r2 0**
- Use the region 2 - 3Mbase
- Make use of the R function - plotPairwiseLD
- How do the LD plots differ?

Sum-up

- Minor allele frequency (MAF) - proportion rare variants
- LD decay - rate shows diversity
- Pair-wise LD plot - linked regions