

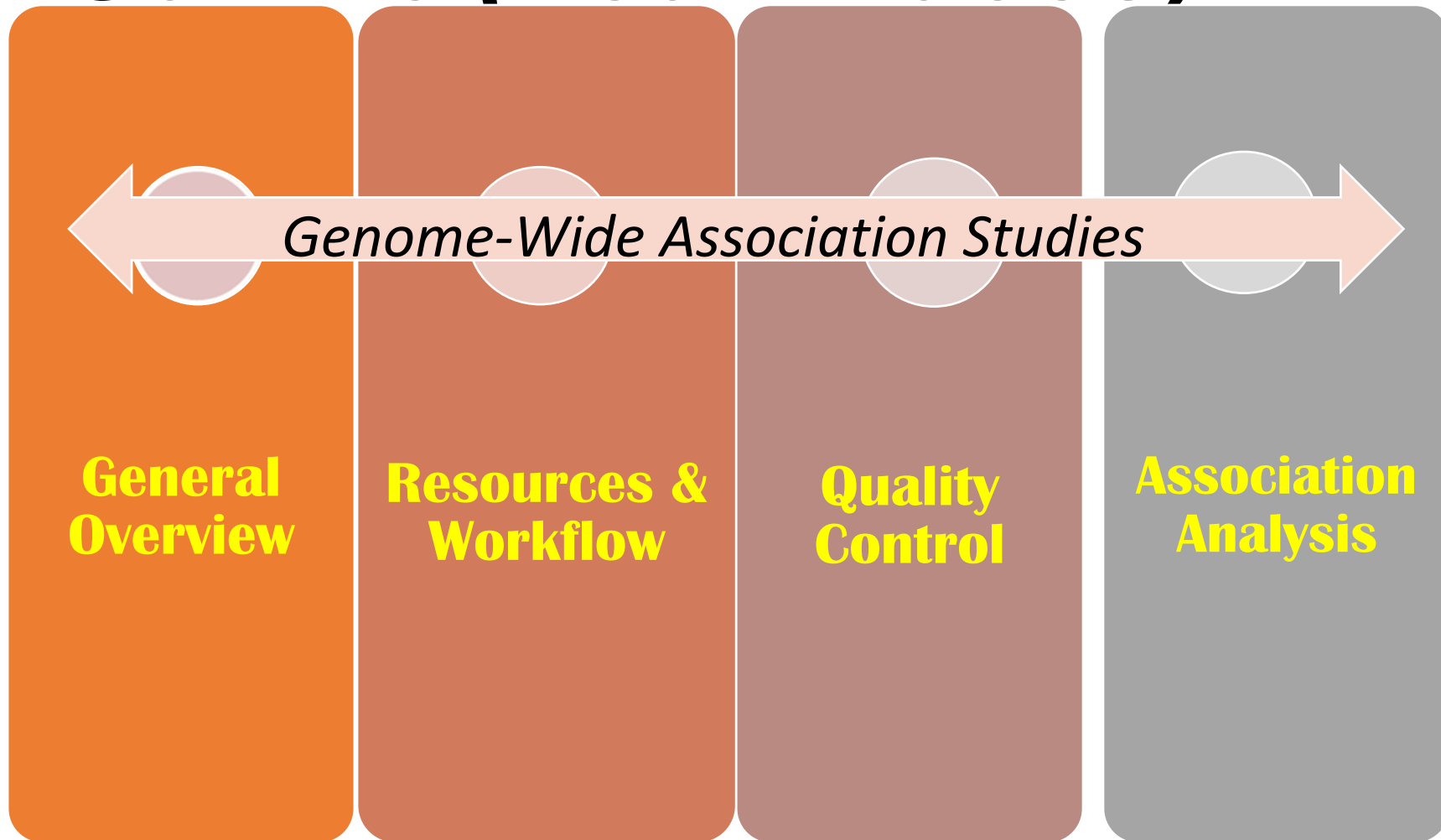
**H3ABioNet**

# ***Genome-Wide Association Study (GWAS): Association Analysis***

**Segun Fatumo PhD**

*Associate Professor, Department of Non-Communicable Disease Epidemiology, LSHTM, UK  
Group Leader, The African Computational Genomics Group, MRC/UVRI and LSHTM, Uganda*

# Outline (Four Videos)

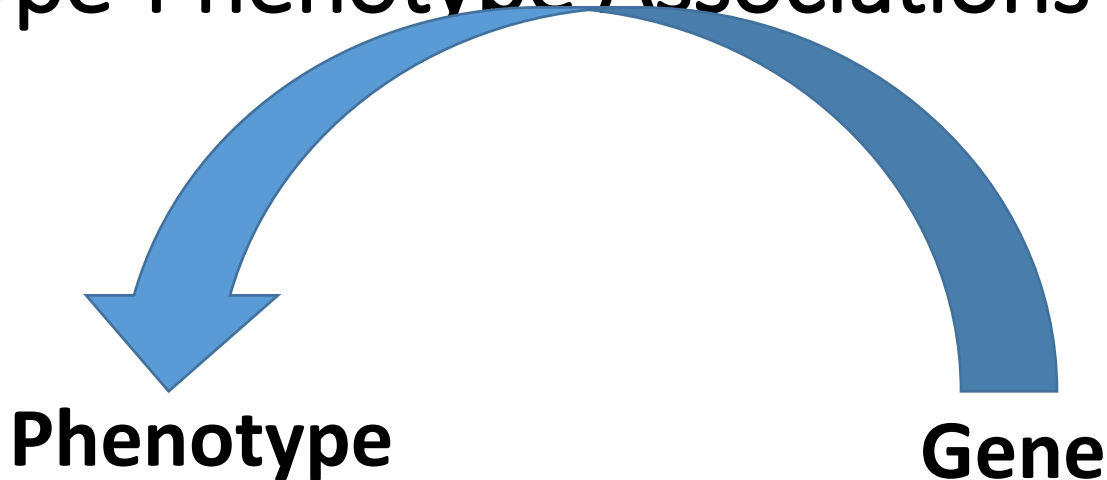


# Intended Learning Outcomes

By the end of the session, students will be able to

- Describe the statistical models for identifying phenotype-genotype associations.
- Identify appropriate model for adjusting for potential confounding when performing genetic association study.
- Interpret plots from genetics association study

# Genotype-Phenotype Associations



## • Case-control GWAS

- Identify SNPs where one allele is significantly more common in cases than controls
- The SNP is *associated* with disease
- Logistic regression

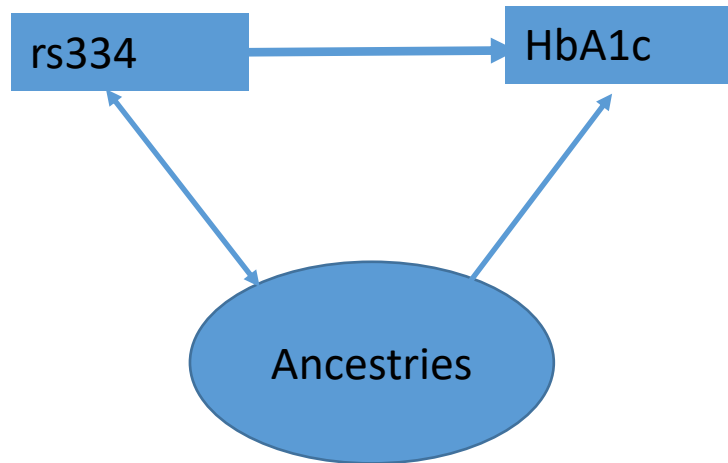


## Continuous trait GWAS

- Fit **linear regression** model for each (SNP, phenotype) pair



# Mixed Model



- A major concern in GWAS is the need to account for the complicated dependence-structure of the data both between loci as well as between individuals. (Adjust for the confounding variable)
- For example, ethnic groups (and subgroups) often share distinct dietary habits and other lifestyle characteristics that leads to many traits of interest being correlated with ancestry and/or ethnicity
- The linear mixed model (LMM) is a powerful way to control covariation arising from complex correlation

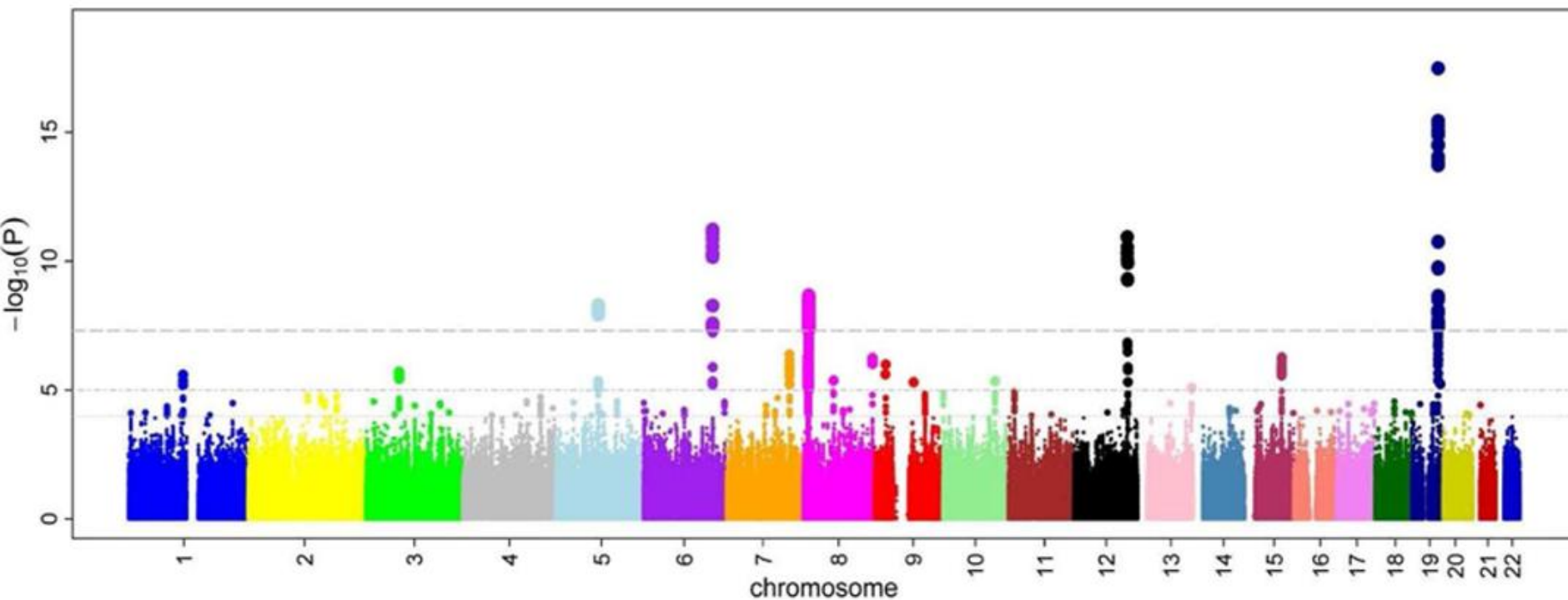
# Testing for association

- All methods produce a **test statistic** and a  **$p$  value** indicating how significant the association/correlation appears to be
  - ◆ i.e. how likely it was to have occurred by chance
- In GWAS, we require **stringent significance levels** (e.g.  $p = 5 \times 10^{-8}$ ) to overcome the multiple testing problem incurred when we test many SNPs throughout the genome
  - ◆ If testing 1 million SNPs using  $p = 0.05$ , we would obtain 50,000 'significant' results just by chance!
  - ◆ Bonferroni correction

# Useful software

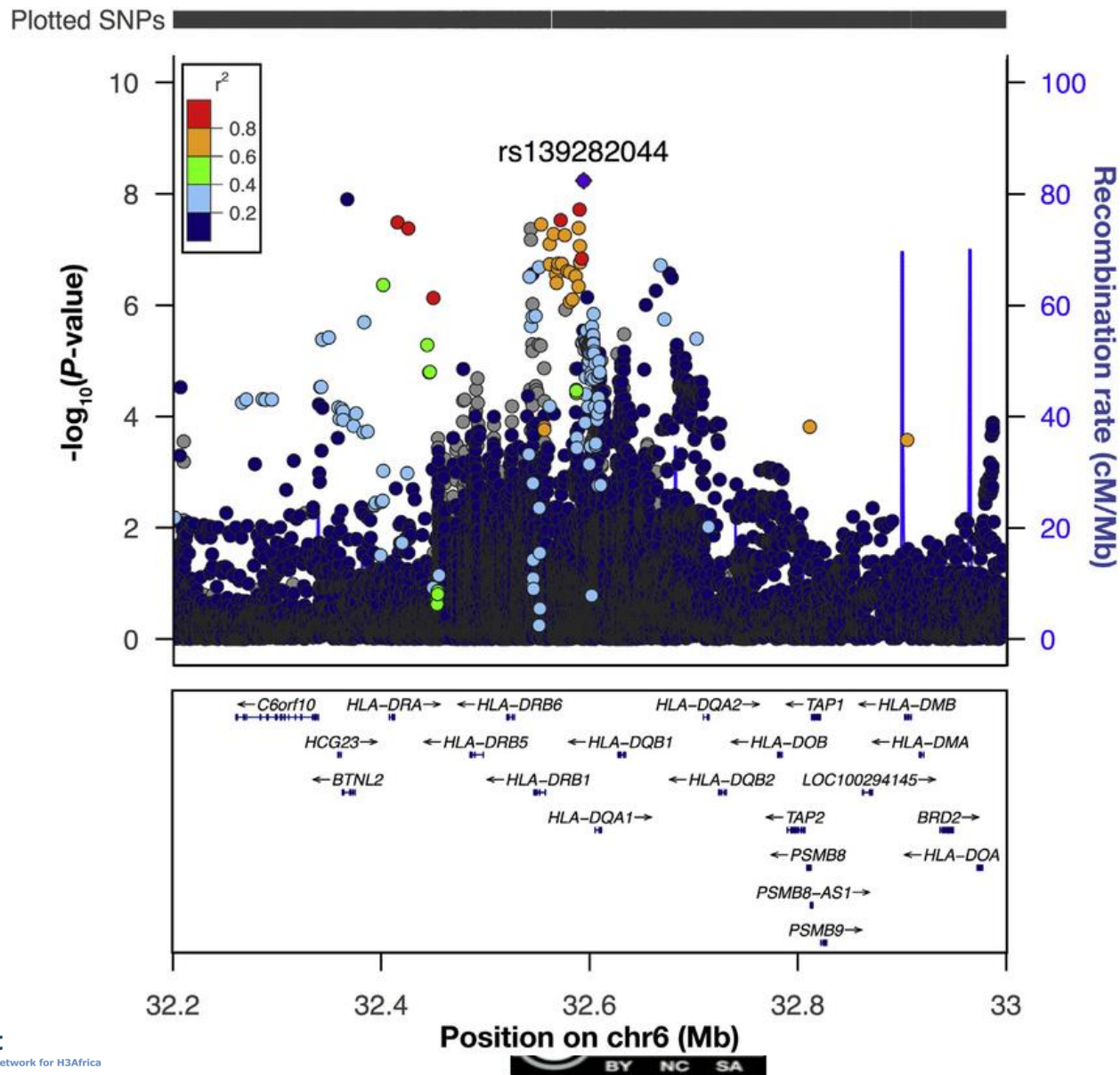
- **PLINK (Standard tool for QC and association analysis)**
  - <http://pngu.mgh.harvard.edu/~purcell/plink>
- **GEMMA (Association) – Implemented Mixed Model**
  - <http://home.uchicago.edu/xz7/software>
- **SNPTEST (Association)**
  - [https://mathgen.stats.ox.ac.uk/genetics\\_software/snptest/snptest.html](https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html)

# Manhattan Plot

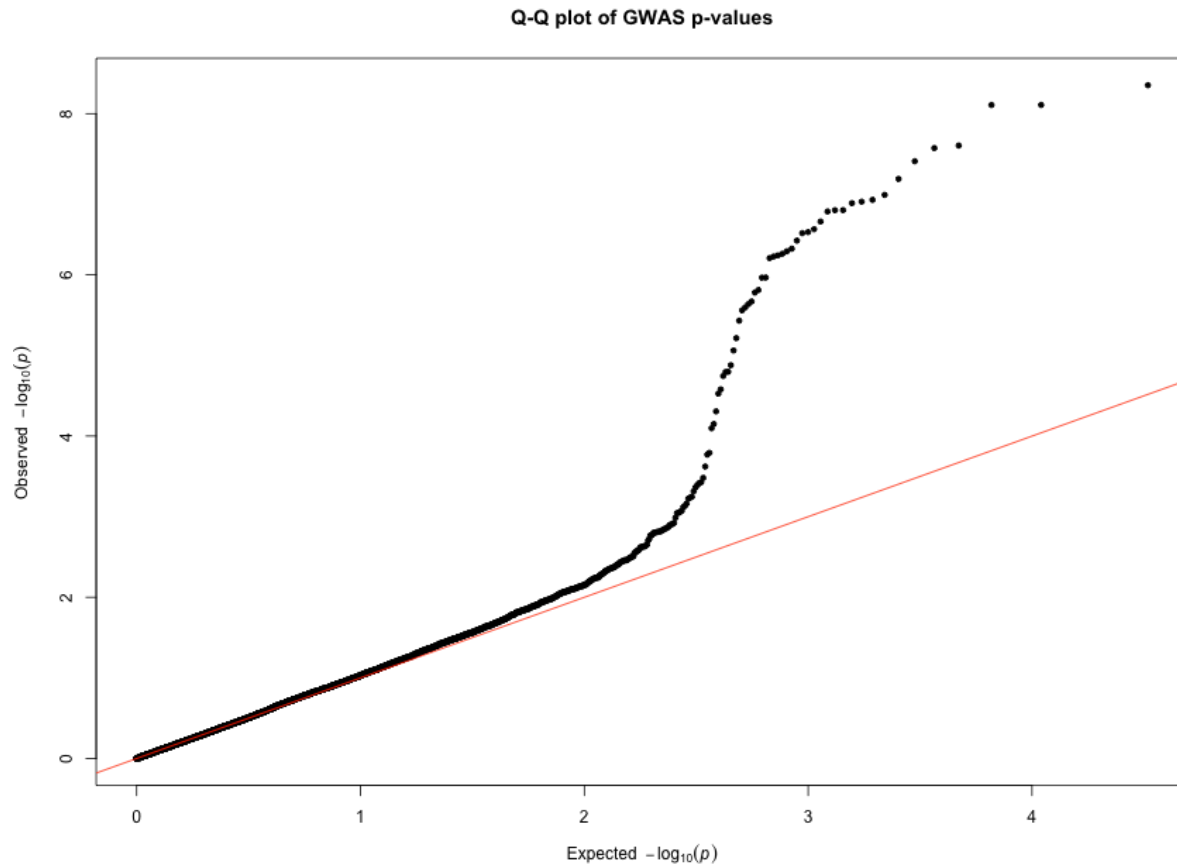




# Close-up of hit region



# Quantile-Quantile (QQ) Plot



# Conclusions

- **In designing GWAS, the selection of the appropriate association test depends on a number of factors:**
  - type of phenotype,
  - control for covariates eg population structure.
  - etc
- **Visualisation of GWAS output is key**
  - Manhattan
  - QQ Plot
  - Regional plot

# Further Reading

Pearson, T.A. and Manolio, T.A., 2008. How to interpret a genome-wide association study. *Jama*, 299(11), pp.1335-1344.

<http://www.biostat.jhsph.edu/~iruczins/teaching/misc/gwas/papers/pearson2008.pdf>



# H3ABioNet