# Final Project

Inferring relationships from genetic data is of popular intrigue, as evidenced by companies like 23andMe and Ancestry as well as by large online communities of genealogists. For biomedical research, knowing relationships is also important to appropriately design experiments and perform analyses. There are many algorithms to infer familial relationships. These split into (1) methods based on allele frequencies, (2) methods based on identity-by-descent segments, and (3) methods aware of admixed ancestry. Ramstetter et al. (2017) survey some of these methods applied to a sample of Hispanic families in the San Antonio area. Here we study a subset of these methods applied to other admixed populations from the International HapMap Project.

We will setup much of the software and dependencies during a future meetup. Below is the summary of the software. This project will provide practice running command line programs and analyzing outputs in `R`. Interpretation of the data analyses will rely on what we have learned about kinship coefficients, IBD state probability, and degrees of relatedness. Various text references will be placed in the STATGEN DRP Dropbox.

- **KinInbcoef**: A simple but dated command line program to calculate kinship based on known pedigrees
- **PLINK**: A software suite enabling a host of efficient genetic analyses based of SNP (single nucleotide polymorphism) marker data
- BEAGLE: A program to infer maternal and paternal origins of alleles, a prerequisite for IBD detection methods
- hap-ibd: A program to detect segments of IBD
- **IBDkin**: A program to estimate kinships and IBD probabilities given IBD segments
- KING: Another program to estimate kinships
- PC-Relate: Another program to estimate kinships, with special attention to admixture

# Quarter Project

1. Rerun KinInbcoef, PLINK, and IBDkin analyses for the entire YRI (Nigerians) and ASW (admixed African-Americans) HapMap cohorts.
2. Analyses in `R`
   (a) Combine KinInbcoef and PLINK outputs into a single table. Combine KinInbcoef and IBDkin outputs into a single table.
   (b) Plot inferred kinship against pedigree kinship using base `R` graphics or `ggplot2`.
   (c) Plot inferred kinship against IBD0 probability.

(d) Interpret these plots. Compare and contrast PLINK versus IBDkin. Which program estimates kinship more accurately? Which program is easier to run?

3. Write and deliver a presentation. Some things to highlight:

   (a) The path-counting formula and an example calculation

   (b) Overview of IBD: a definition, how we detect IBD segments

   (c) Data analyses: summarize the software, show plots, interpret plots

# Extended Project

If mentee wants to present research at the undergraduate symposium, the mentor suggests extending these data analyses further. Here is how the mentor suggests extending things.

- Select 1 or 2 more admixed cohorts from the HapMap Project and download data.
- Self-guided reading about KING and/or PC-Relate. Mentor can provide appropriate references.
- Run kinship analysis using KING (easier) or PC-Relate (challenging) or both.
- Make more plots. Compare (3) methods aware of admixed ancestry against (2) methods based on identity-by-descent.

The driving question of this research is: how well do IBD detection methods perform in inferring relationships among admixed individuals? This impacts the quality of downstream analyses for an applied project in the Alzheimer's Disease Sequencing Project Follow-Up Study (ADSP FUS).