**Readme – Conduct a GWAS using the 1000 Genome Project Data**

**Learning GWAS**

* Background reading from NIH: <https://www.genome.gov/about-genomics/fact-sheets/Genome-Wide-Association-Studies-Fact-Sheet>
* A tutorial: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6001694/pdf/MPR-27-e1608.pdf>

<https://github.com/MareesAT/GWA_tutorial/>

**Learning PLINK**

* Homepage: <http://zzz.bwh.harvard.edu/plink/>; updated version is here: https://www.cog-genomics.org/plink2
* Good to start with the Tutorial: <http://zzz.bwh.harvard.edu/plink/tutorial.shtml>

**Implementing a GWAS**

* The 1000 Genome Project (1kGP) background: <https://www.internationalgenome.org/>
* A cleaned set of 1000 genome project data: <http://tcag.ca/tools/1000genomes.html>
  + Important to read: [report](http://tcag.ca/documents/tools/omni25_qcReport.pdf) on the quality steps performed
  + Good to repeat the QC steps after learning PLINK
* Conduct and report GWAS as a practice
  + The cleaned dataset has 1736 unrelated/indepedent individuals and ~2M high-qualtiy SNPs. For your analysis, you might want to have another QC based on the genotype group size so that you don’t run into the hatP(Y=1|G)=0 or 1 issue.
  + Randomly assign an individual to case or control, then conduct the apporopriate GWAS, including the classical logisitcal regression for the main effects, and the proposed method for indirect interaction effects.
  + The random assignment is important: there are clear population stratifications based on PCA analysis; see page 10 of the [report](http://tcag.ca/documents/tools/omni25_qcReport.pdf) on the QC steps.
  + Results should include the Manhanttan plot, the QQ-plot, and the histogram.