**Conditional Analysis Cakewalk**

Linkage disequilibrium (LD), the nonrandom association of alleles at different loci (Maruki & Lynch, 2014), is an important consideration when performing association analysis as it can bias the results. LD can cause some variants to appear significantly associated with a trait that conditional analysis will reveal to have no association. Therefore, ignoring LD can lead to a misinterpretation of genome wide association studies (GWAS), impacting conclusions or future studies derived from the results (Christoforou et al., 2012). As high throughput sequencing becomes more widely used, the challenge of managing the LD is growing. Conditional analysis can identify which variants are independently associated with an outcome (Knight et al., 2012), but running sequential conditional analyses can be tedious and time-consuming. To meet this challenge, we have developed the Conditional Analysis Cakewalk tool for automating this important process.

The Conditional Analysis Cakewalk begins with a section requiring input from the user. The user will enter:

* the desired trait
* chromosome number
* start/end positions for the region to be scanned
* covariates
* desired suffix for output files.
* four thresholds that can be adjusted:
  + maximum p-values for including variants in both the initial and subsequent LD calculations
  + the p-value for establishing significance of variants as the termination condition
  + a p-value for limiting which variants are included in each association analysis, which was included to improve processing times.
* The y-axis range and significance lines within the LocusZoom plot can also be specified.

Finally, the results will be routed to a user-specified directory with an option for removing all other files to conserve storage space.

The Conditional Analysis Cakewalk utilizes MMAP to conduct the initial association analysis between the specified trait and genomic region (O’Connell). The output file is then sorted and processed into a marker list for LD calculations, which are also conducted using MMAP. The LD results are again sorted and processed into files formatted specifically for LocusZoom. The LocusZoom plot includes all variants in the specified region, with a green vertical line in the location of the top variant. The Conditional Analysis Cakewalk will continue to iterate through the MMAP analysis, LD calculations, and LocusZoom plotting until the p-value of the top variant is greater than the cutoff p-value specified in the beginning of the script.

**References**

Christoforou, A., Dondrup, M., Mattingsdal, M., Mattheisen, M., Giddaluru, S., Nothen, M.M., Rietschel, M., Cichon, S., Djurovic, S., Andreassen, O.A., Jonassen, I., Steen, V.M., Puntervoll, P., & Le Hellard, S. (2012). Linkage-disequilibrium-based binning affects the interpretation of GWASs. *American Journal of Human Genetics, 90*(4), 727-733.

Knight, J., Spain, S.L., Capon, F., Hayday, A., Nestle, F.O., Clop, A., Wellcome Trust Case Control Consortium, Genetic Analysis of Psoriasis Consortium, I-chip for Psoriasis Consortium, Barker, J.N., Weale, M.E., & Trembath, R.C. (2012). Conditional analysis identifies three novel major histocompatibility complex loci associated with psoriasis. *Human Molecular Genetics, 21*(23), 5185-5192.

Maruki, T., & Lynch, M. (2014). Genome e-wide estimation of linkage disequilibrium from population-level high-throughput sequencing data. *Genetics, 197*(4), 1303-1313.

O’Connell, J.R. *MMAP: Mixed Model Analysis for Pedigrees and Populations.* Available from: https://mmap.github.io/