GWAS Protocol

Notebook: GWAS
Created: 2016-02-29 오후 6:13 **Updated:** 2016-03-10 오후 7:30
Author: Harry Yang

1. Download data
2. Pylmm (get it from https://github.com/michaelbilow/pylmm_zarlab)
    1. python pylmmKinship.py --bfile (base of the binary plink files) outfile.kinship
        1. INPUT - Binary PLINK - *.bed *.bim *.fam all in same base - i.e.
        2. OUTPUT - Kinship matrix -- I prefer *.kinship with same base as the binary plink files
    2. Extract pheno file from fam file
        1. cat base.fam | awk '{print $1, $2, $6}' >> base.pheno
    3. python pylmmGWAS.py -- bfile (base of the binary plink file) --kfile (the kinship including the extension) --phenofile (pheno file generated in step 2) outfile (*.gwas)
        1. INPUT - Plink Binary files, Kinship, Pheno
        2. OUTPUT - *.gwas file
        3. GWAS file
            1. SNP ID
            2. Beta value
            3. Stdev of Beta
            4. Z-score
            5. P-value
3. Manhattan Plot from GWAS
    1. Get map file from the Binary PLINK files
        1. plink --bfile (Base of the binary plink) --out (Base of the binary plink file)
            1. INPUT - Binary Plink
            2. OUTPUT - Regular Plink files (*.ped, *.map) -- the map file is required
    2. Extract the coordinates and p value for manhattan plot
        1. extract_data_for_manhattan_plot.py
            1. INPUT
                1. GWAS file from pylmm
                2. MAP file from plink
                3. OUTPUT file name
            2.
    3. R script for Manhattan Plot
        1. library(qqman)
           f <- read.table('THE MANHATTAN PLOT DATA FILE FROM STEP 2')
           colnames(f)[1]='CHR'
           colnames(f)[2]='BP'
           colnames(f)[3]='SNP'
           colnames(f)[4]='P'
           manhattan(f, main = "THE NAME OF THE STUDY", col = c("blue4", "orange3"))
4. Metasoft Analysis
    1. GET SNP_list and GET metasoft_input
        1. meta_analysis_prep.py
            1. INPUT == list of *.gwas files from pylmmGWAS
    2. metasoft - http://genetics.cs.ucla.edu/meta/
        1. java -jar (Metasoft.jar) -input (OUTPUT from meta_analysis_prep) -output

(DESIRED OUTPUT NAME) -mvalue

3. ForestPMPLOT
    1. #python pmplot.py u4c/new_metasoft_input.txt u4c/meta_result_with_m.txt u4c/study_names.txt u4c/study_order.txt rs3013451 eh test_three.pdf
    2. python pmplot.py (INPUT FOR METASOFT) (OUTPUT FROM METASOFT) (list of study names) (list of order - 1 \n 2 \n 3\n 4\n ... is fine) (GENE NAME of the SNP) (OUTPUT.pdf)

5. CAVIAR
    1. ### SPECIFY THIS PART ###
6. Data Handling - if there are discrepancies in data
    1. If the indexing is different
        1. It is in chr:coordinate, rather than rsid
            1. change_coordinate_gwas_file.py
                1. GWAS FILE
                2. INDEX FILE
                3. OUTPUT FILE
                    1. outputs the gwas file with rsid
    2. if map file does not have coordinates
        1. extract_data_for_manhattan_plot_without_map.py
            1. GWAS FILE
            2. INDEX FILE
            3. OUTPUT FILE
                1. outputs the manhattan plot data