**Anderson protocol using PLINK on Colt-cohort 1 array data and subsequent GCTA association testing.**

Plink version: 1.9

GCTA version: 1.94

Protocol as per following paper in Nature protocols 2010: Data quality control in genetic case-control association studies. Carl A Anderson et al.

Commands entered in command prompt.

Step 1: Create bed files

plink --file GSA2016-49-025\_12jan2017 --make-bed --out Cohort1

Step 2: Calculate sample missing genotype and heterozygosity rates.

plink --bfile Cohort1 --missing --out Cohort1

plink --bfile Cohort1 --het –out Cohort1

The output is data file containing the number of missing SNPs and heterozygosity rate of SNPs per sample. These are then plotted on a graph. Individuals with a genotype failure rate of ≥0.03 (n=2) and a heterozygosity rate of +/- 3 standard deviations are deemed poor-quality and are flagged for removal.

Step 3: Identify same individuals

Firstly, create list of SNPs to use to run IBD analysis

plink --file GSA2016-49-025\_12jan2017 --indep-pairwise 50 5 0.2 --out Cohort1

Creates a data.prune.in file

Next step: Create list of all pairs with IBS >0.1.

plink --file GSA2016-49-025\_12jan2017 --extract Cohort1.prune.in --genome --min 0.1 --out Cohort1

This creates an IBS matrix of proportion of shared alleles between each cohort participant. Cohort1.genome, which can be opened in excel.

Z0 = Proportion of alleles which are different. Z1=proportion of single allelle they share. Z2= proportion of both alleles shared

Individuals with shared alleles ~1 suggest a duplicate sample and both should be removed.

Step 4: Removal of Ids

Create text file of all ffid and id of individuals to be removed ‘IDstoremove.txt’

plink --bfile Cohort1 --remove IDstoremove.txt --make-bed --out Cohort1-IDclean

Step 5: QC of SNPs

Removal of SNPS which have MAF<0.01, call rate of <0.025 and HWE p-value <0.0001

plink --bfile Cohort1-IDclean --maf 0.01 --geno 0.025 --hwe 0.0001 --make-bed --out Cohort1-complete

Step 6: Update sex info of cohort

Create .txt file with ffid, id and sex (as numbers) of cohort and make into file ‘sex-info.txt’

plink --bfile Cohort1-complete --update-sex sex-info.txt --make-bed --out FinalGTCAcohort1 --noweb

Step 7: GWAS of Markov defined cases  
gcta64 --mlma-loco --bfile FinalGTCAcohort1 --pheno Cohort1-MMcase-phenotype.phen --out Cases-result --thread-num 10

Step 8: GWAS of linear eGFR decline

gcta64 --mlma-loco --bfile FinalGTCAcohort1 --pheno Cohort1-linear-eGFR-pheno.phen --out lineareGFR-result --thread-num 10