BS859: Final Write Up
Option 2
Sydney Sorbello
Due: 5/7

Task 1: Clean Data and Perform PCA

In this project, we will explore genetic data to investigate its relationship to Rheumatoid Arthritis (RA). In order to perform a proper GWAS analysis, we must first clean the data. There are multiple measurements used to filter data. We first filter for a minor allele frequency greater than 0.1. This eliminates unreliable statistical estimates and reduces genotyping error. The second factor we filter on is a genotype missing rate of fewer than 5%. On the control samples we perform hardy-weinberg equilibrium. This is a test that measures the heterozygosity for bi-allelic variants. To ensure random mating on the genotypes, we set the threshold to 1e-6. Finally we check on an individual basis to ensure every sample has a genotype missingness of less than 5%. During this analysis, 0 people were removed due to missing genotypes. There were, however, 18402 variants removed due to missing genotype data. The Hardy Weinberg test removed 663 variants. Finally, 22907 variants were removed due to the minor allele frequency. After this step 502304 variants and 2062 individuals remained for analysis.

A sex check was performed to ensure the phenotypic assignment of sex matches that of the genotype. This analysis revealed 7 individuals that failed the sex check. These individuals were then removed from the dataset for quality control leaving 2055 individuals.

The individuals were then tested for relatedness. The instructions for this assignment noted that this step was unnecessary, but for quality assurance reasons, we performed it regardless. As expected, no individuals were related in the dataset.

Before we can move forward to perform PCA, we must first LD prune the variants. This step will remove markers with low correlation. During this step in the analysis we utilized a 10000 kb window size, a variant shift of 1 and a maximum r^2 correlation of 0.2. Once the variants are pruned, we extract those that are left in for our PCA. During this step 394662 of 502304 variants were removed leaving 107642 variants.

The PCA analysis produced 10 principal components using the default normalization method. The resulting PCA featured an ANOVA test for population differences along each principal component. This test yielded 4 significant principal components where the p-value threshold is set at 0.01. PC1 had a p-value of 1.88738e-15, PC2 had a p-value of 1.33227e-15, PC4 yielded a p-value of 3.33067e-16, and PC8 yields a p-value of 0.00229878.

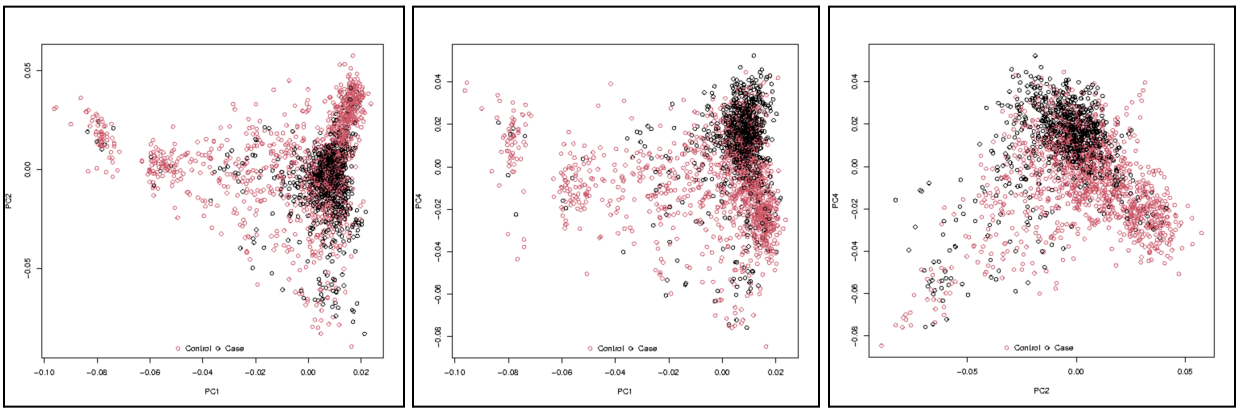

Fig 1. Principal Component plots for three most significant principal components. The plots depict PC1 v. PC2, PC1 v. PC4, and PC2 v. PC4 from left to right respectively. In each of the plots, the red circles indicate individuals in the control group and the black circles indicate individuals in the case group.

With respect to outliers, there is only one cluster in PC1 that raises concern. There is a clear separation in the cluster of individuals with PC1 measurement less than about -0.065 (Fig.1 furthest right). For this reason we will decide to remove them from the dataset to avoid unnecessary noise. This cluster featured 66 individuals, 6 of which were in the case group. The resulting number of individuals is 1989 where 860 are cases and 1129 are controls.

Although there were 4 significant PCs (PC1, PC2, PC4, PC8), 3 of the PCs were highly significant in their detection of differences between each PC. PC8 was moderately significant so, to reduce noise during the analysis, we recommend moving forward with PC1, PC2, and PC4 only for further analysis.

Task 2: Sex-Stratified Analysis

A: GWAS Analysis

It is well known that females are affected by RA much more frequently than males. For this reason we will perform GWAS analysis on the respective populations individually. We expect there to be a slightly stronger signal from the female population, but roughly similar overall signals. To begin we add the principal components PC1, PC2, and PC4 to the covariate file so that we can call on them during the analysis.

We stratify the covariate file by sex by separating the samples based on their assigned sex. In this dataset a 1 indicates male and a 2 indicates female. The result is two covariate files exclusive to

the sex. We now perform GWAS on the male and female datasets using the same parameters. We cite the principal components as covariates and the confidence interval as 0.95.

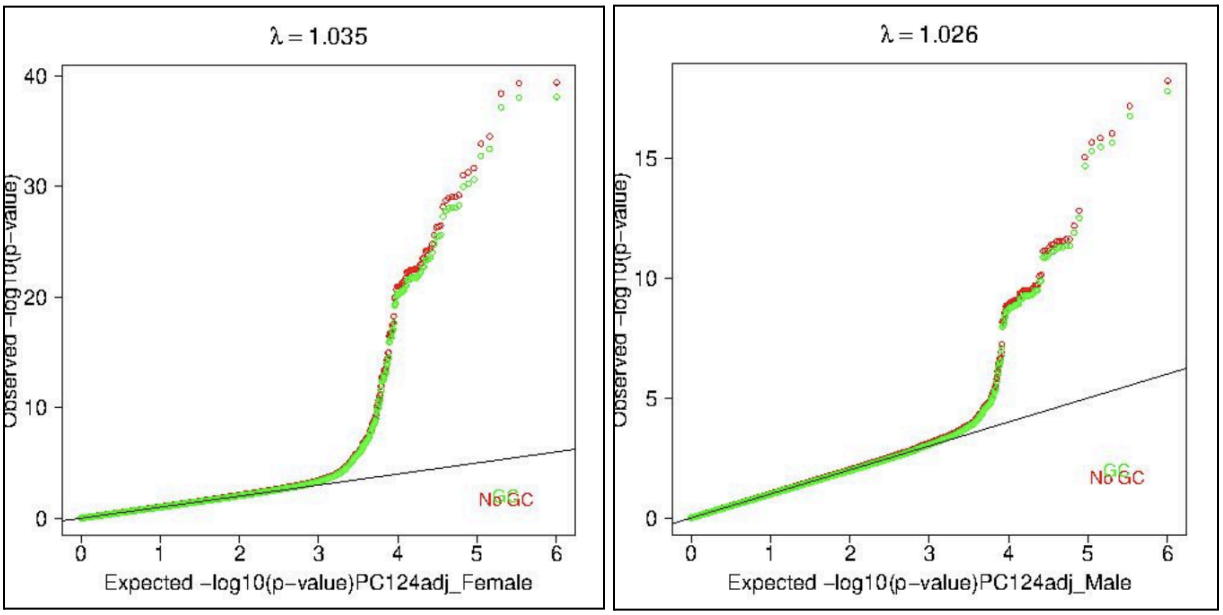

Fig 2. QQ Plot representing the sex-stratified GWAS analysis. The female GWAS yielded a lambda of 1.035 and the male a lambda of 1.026 pictured above from left to right respectively.

In order to visualize the results of the GWAS analysis, we make a QQ plot using the provided script from class. The QQ plot visualizes the quantiles of the data against the theoretical quantiles of the distribution. The QQ plots produced from both GWAS analyses show a very strong signal (Fig. 2). The actual p-values were much higher than expected as reflected in the lambda measurements. The female dataset received a lambda value of 1.035 and the male dataset 1.026. Both measurements are above 1 indicating that the signal is likely significant. As expected, the female dataset has a slightly larger lambda value reflecting the known fact that women tend to have RA at a higher frequency than males.

Moving forward in the analysis, we visualized a Manhattan plot for both analyses (Figure 3). This plot allows us to visualize and summarize the significant variants across the genome and help place their location. The p-values for each variant are plotted against the genomic locus. Variants are deemed significant when their p-value is less than 0.00000001. In both populations all significant variants are located on chromosome 6. This is consistent with previous findings showing the chromosome 6 region 6p21.31 to be linked to RA. The female population saw significant SNPs on the order of 1e-40. This is an incredibly significant finding as there are many SNPs that are also highly significant. The male population also saw highly significant SNPs. However, they were on the order of 1e-19. Both findings are highly interesting are the signal is very clean.

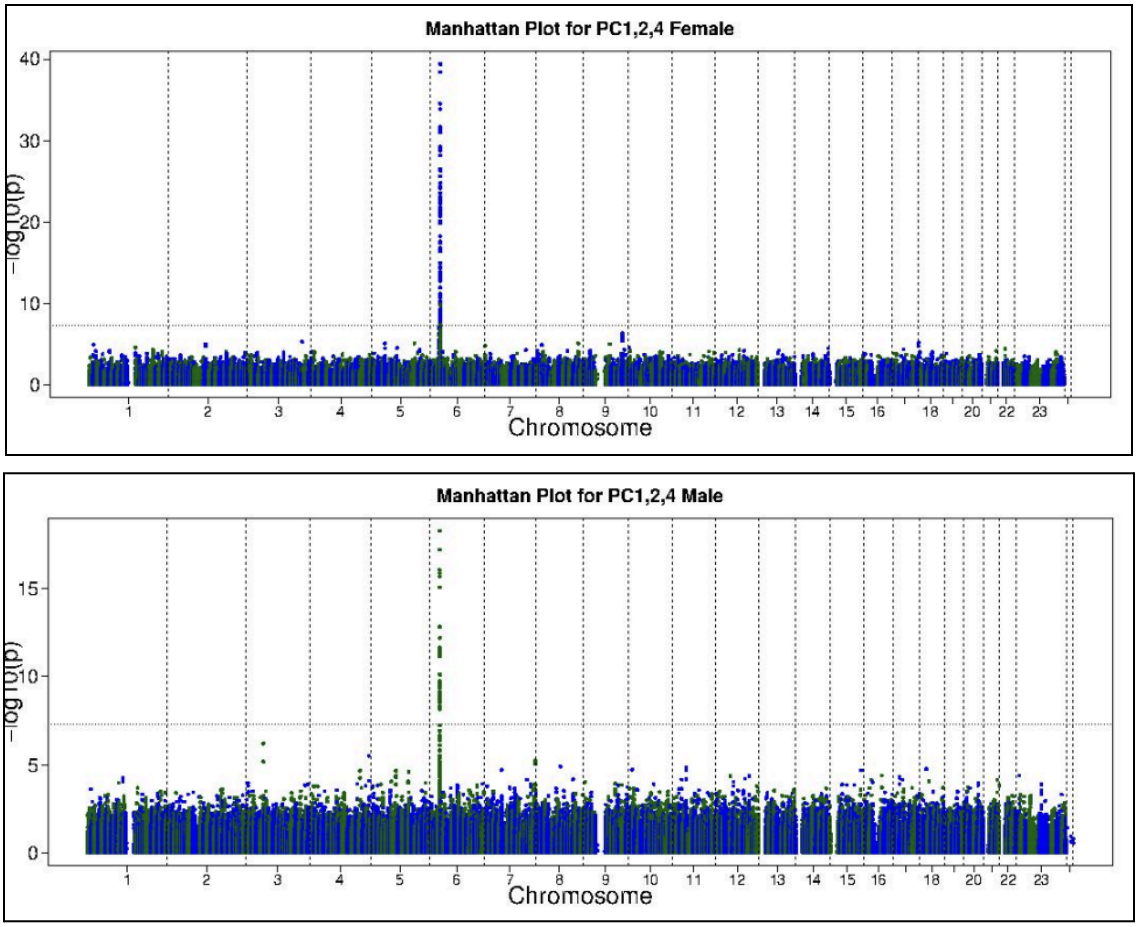

Fig 3. Manhattan plots representing the GWAS sex-stratified analysis. The female population and the male population are depicted above from top to bottom respectively.

To validate the results, we take a look at the 5 most significant SNPs from both populations (Table 1). 3 significant SNPs are shared between the populations. Among those SNPs, rs660895 was the most significant in both populations with a p-value of 3.849e-40 and 5.905e-19 for the female and male populations respectively. As part of further validation we query SNPedia for the biological underpinning of the variant. Rs660895 has been associated with RA such that the minor allele is associated with an increased risk (https://www.snpedia.com/index.php/Rs660895). It is reported that individuals with the minor allele homozygosity are estimated to be at 6x higher risk for the disease.

Female		
CHR	SNP	P-value
6	rs660895	3.849e-40
6	rs9275224	4.638e-40

Male		
CHR	SNP	P-value
6	rs660895	5.905e-19
6	rs2395175	6.843e-18

6	rs6457617	3.909e-39	6	rs2395163	9.482e-17
6	rs2395175	3.051e-35	6	rs3763312	1.454e-16
6	rs2395163	1.344e-34	6	rs3763309	2.181e-16

Table 1. Summary table for the 5 most significant SNPs found during GWAS sex stratified analysis.

B: Meta Analysis Recombining Sexes

In order to compare the results from the two populations and further investigate significant variants, we will perform a meta analysis. First we reformat our files so that they can be read by the required program ‘metal’. During the analysis we provide the individual results from GWAS. The meta analysis yielded promising results as the top 10 most significant genes all agreed in effect direction between the two samples (Table 2). Here again, the top SNP is rs660895 followed by other variants located on chromosome 6. Due to the significance of the variants in addition to the agreement in direction, the meta analysis returned more significant p-value. The lowest p-value from rs660895 was 2.564e-55. These results are very promising and corroborate the expected findings within the 6p21.31 region.

CHR	SNP	A1	A2	Z score	P value	Direction
6	rs660895	A	G	-15.666	2.564e-55	-/-
6	rs2395175	A	G	14.853	6.668e-50	+/+
6	rs2395163	A	G	-14.552	5.699e-48	-/-
6	rs6457617	A	G	14.473	1.797e-47	+/+
6	rs9275224	A	G	-14.450	2.501e-47	-/-
6	rs3763309	A	G	14.185	1.132e-45	+/+
6	rs3763312	A	C	14.166	1.485e-45	+/+
6	rs6910071	A	G	-13.972	2.297e-44	-/-
6	rs2395185	A	G	12.733	3.894e-37	+/+

Table 2. Top 10 most significant SNPs from the meta analysis.

Once again, we visualized the data by creating a QQ plot and a manhattan plot (Figure 4). The QQ plot revealed a stronger lambda score than either of the initial GWAS analyses with a measurement of 1.043. This indicates that the significant findings saw a higher statistical power when combined into one analysis. Similarly in the Manhattan plot, we see an even stronger signal from chromosome 6 with the most significant SNP being on the order of 1e-55.

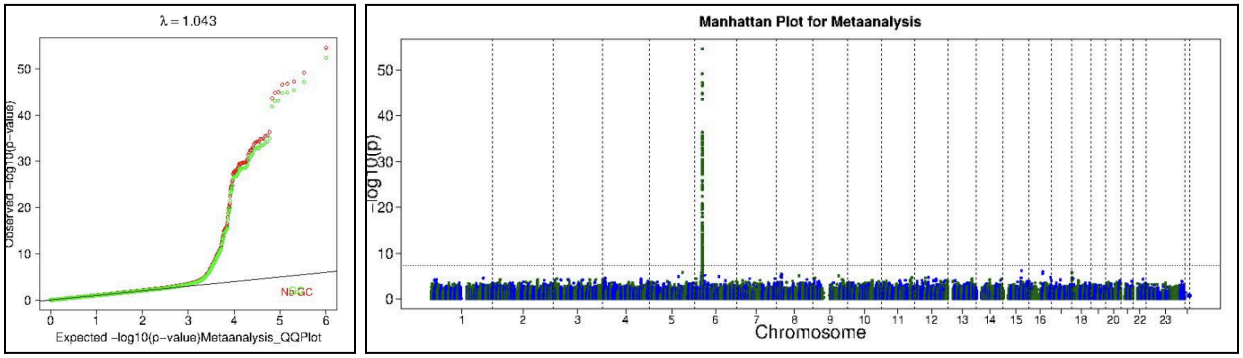

Fig 4. Diagnostic QQ plot and Manhattan plot visualizing meta analysis results depicted above from left to right respectively.

Task 3: LD Score Regression

In order to investigate the heritability of RA we perform LD score regression. LD score regression utilizes statistical methods to estimate heritability and genetic correlations. We began analyzing the data collected from both European and Asian populations. This data was gathered from Okada et. al. who performed a study in 2014. We first process the datafile so that it is in the right format for measurement using sumstats.py. After the correct file format is obtained, we perform LD score regression using LDSC. An important note is that the UK BioBank SNP information was provided as reference from the European population. This decision was made as roughly 75% of the dataset were from the European population.

The combined population revealed an estimated inheritance of 11%. A mean chi square of 1.1703 indicates a strong polygenic trait and the intercept of below 1 shows little to no inflation from confounder. Overall this was a clean analysis (Table 3).

The previous steps were repeated for the European and Asian populations individually. The analysis for the European population was provided the UK BioBank reference for EUR and the Asian population analysis was provided the UK BioBank reference for EAS. The individual analyses revealed similar results to the initial combined test. The Eurpoean population saw a much greater inheritance compared to the Asian population at 10.3% compared to only 4.9%. Both mean chi square statistics were above 1 indicating a strong polygenic trait. This

measurement was higher in European populations, consistent with the inheritance findings (1.1359 compared to 1.0623). Both intercepts were below 1 indicating clean data with little effect from confounders (Table 3).

Population	h^2	Lambda GC	Mean Chi	Intercept
ALL	0.1155 (0.0172)	1.0466	1.1703	0.9737 (0.011)
European	0.1025 (0.0172)	1.0466	1.1359	0.9621 (0.01)
Asian	0.0491 (0.0148)	1.0466	1.0623	0.979 (0.0089)

Table 3. LD score regression results for the combined, european and asian population analyses.

The various analyses throughout this project reflected previously published literature with respect to genetic association of RA. The genetic analysis studies yielded very strong results that would typically cause a mild warning sign. However, consistent validation of QC measurements as well as cross checking of variants reassured proper analysis. The genetic association and inheritance of RA was further supported through LD score regression, revealing subtle differences between populations self assigned to geographic locations. Further research may reveal more about polygenic risk factors or variants associated with RA at other locusts of the genome.

Work Cited:

Okada, Y., Wu, D., Trynka, G. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 506, 376–381 (2014). https://doi-org.ezproxy.bu.edu/10.1038/nature12873

Code:

```
## FINAL PROJECT
# Sydney Sorbello
# Due: 4/7

# let set the path to the NARAC data
BASE=/projectnb/bs859/data/RheumatoidArthritis/final_project/narac_hg19

# lets load some key packages
module load R
module load plink/1.90b6.27
module load eigensoft
```

```
module load metal

# 1: Preprocessing and PCA analysis
# First, let's filter the dataset based on the following requirements
# MAF 0.01
# Fewer than 5% missing genotypes
# hwe p-value greater than 0.000001
# individuals with less than 5% genotype misssingness
plink --bfile $BASE --maf 0.01 --geno 0.05 --hwe 1e-6 --mind 0.05 --make-bed --out narac_filtered

# here we can check the sex assignmnet aligns with the genotypes
plink --bfile narac_filtered --check-sex --out narac_sexcheck
awk '$5 == "PROBLEM"' narac_sexcheck.sexcheck > failed_sexcheck.txt

# the sex check revealed 7 individuals with problems during the sex check, lets remove them from the dataset
cut -f1,2 failed_sexcheck.txt > remove_sexcheck.txt
plink --bfile narac_filtered --remove remove_sexcheck.txt --make-bed --out narac_sexcheck_filtered

# now we check for relatedness. the data description noted that the participants should not be related, but we check
regardless
plink --bfile narac_sexcheck_filtered --genome --out narac_ibd
awk '$10 > 0.185' narac_ibd.genome > related_pairs.txt
wc -l related_pairs.txt

# now we prune LD from the cleaned data
plink --bfile narac_sexcheck_filtered --indep-pairwise 10000kb 1 0.2 --out narac_pruned

# finally we extract the pruned snps
plink --bfile narac_sexcheck_filtered --extract narac_pruned.prune.in --make-bed --out narac_pca

# now we perform PCA
smartpca -p narac_pca.par > narac_pca.log

# lets plot the first two principal components
Rscript --vanilla plotPCs.R narac_pca.evec 1 2 10
Rscript --vanilla plotPCs.R narac_pca.evec 1 4 10
Rscript --vanilla plotPCs.R narac_pca.evec 2 4 10

# lets find the outliers in the PC1 cluster < -0.067
awk '$2 < -0.065 { split($1, id, ":"); print id[1], id[2] }' narac_pca.evec > pca_outliers.txt

# and remove them from the dataset
plink --bfile narac_sexcheck_filtered --remove pca_outliers.txt --make-bed --out narac_pcaout_filtered
plink --bfile narac_pca --remove pca_outliers.txt --make-bed --out narac_pcaout_filtered

# Part 2A: Sex Stratified Analysis
# the covariate file for the dataset has already been supplied
# however, we need to align this file with the individuals we are still working with in the dataset
# this awk command with create a new covariate file consistent with the current individuals
(head -n 1 narac.cov && awk 'NR==FNR {keep[$1,$2]; next} ($1,$2) in keep' narac_pcaout_filtered.fam narac.cov) >
narac_cov_synced.cov
# we pick out PC1, PC2, and PC4 so that we can add them to the covariate file
awk '{ split($1, id, ":"); print id[1], id[2], $2, $3, $5 }' narac_pca.evec > pca_selected.txt

# save and remove header
## ONLY WORKS directly in terminal
head -n 1 narac_cov_synced.cov > cov_header.txt
tail -n +2 narac_cov_synced.cov > narac_cov_noheader.cov

# make sure the files are formatted similarly
sed -i 's/\r$//' pca_selected.txt
sed -i 's/\r$//' narac_cov_noheader.cov

# and merge the files
{ echo -e "FID IID RA sex SEN antiCCP IgM PC1 PC2 PC4"; awk 'NR==FNR && $1 !~ /^#/ {a[$1]=$3"\t"$4"\t"$5; next} ($1 in
a) {print $0 "\t" a[$1]}' pca_selected.txt narac_cov_noheader.cov; } > merged_output.txt
```

```
# in order to continue with the sex-stratified analysis, we split the covariate file by the sex
# Female-only covariates
awk '$4 == 2' merged_output.txt > narac_females_cov.cov

# Male-only covariates
awk '$4 == 1' merged_output.txt > narac_males_cov.cov

# set the new header for the file
header="FID IID RA sex SEN antiCCP IgM PC1 PC2 PC4"

# Add to female covariates
{ echo -e "$header"; cat narac_females_cov.cov; } > narac_females_cov_with_header.cov

# Add to male covariates
{ echo -e "$header"; cat narac_males_cov.cov; } > narac_males_cov_with_header.cov

# now lets perform GWAS using the selected PCs for females
plink --bfile narac_pcaout_filtered --covar narac_females_cov_with_header.cov --covar-name PC1,PC2,PC4 --logistic beta
hide-covar --ci .95 --out narac_female_gwas

# and the same for the male population
plink --bfile narac_pcaout_filtered --covar narac_males_cov_with_header.cov --covar-name PC1,PC2,PC4 --logistic beta
hide-covar --ci .95 --out narac_male_gwas

# Lets show the qqplot for bth analyses
Rscript --vanilla qqplot.R narac_female_gwas.assoc.logistic PC124adj_Female ADD
Rscript --vanilla qqplot.R narac_male_gwas.assoc.logistic PC124adj_Male ADD

# lets take a look at significant SNPs
awk '$9 < 5e-8' narac_female_gwas.assoc.logistic > female_sig_snps.txt
awk '$9 < 5e-8' narac_male_gwas.assoc.logistic > male_sig_snps.txt

# and check the size of these files
wc female_sig_snps.txt
wc male_sig_snps.txt

# lets make a manhattan plot
Rscript --vanilla gwaplot_pgen.R narac_female_gwas.assoc.logistic "Manhattan Plot for PC1,2,4 Female"
manhattan_pc124_female2
Rscript --vanilla gwaplot_pgen.R narac_male_gwas.assoc.logistic "Manhattan Plot for PC1,2,4 Male" manhattan_pc124_male

# PART 2B: Meta analysis
# first we have to reformat the plink files for METAL
awk 'BEGIN {print "A2"} {print $6}' narac_pcaout_filtered.bim > A2.txt

# lets add A2 to the female data
awk 'NR==FNR {a[NR]=$1; next} {print $1, $2, $3, $4, a[FNR], $5, $6, $7, $8, $9, $10, $11, $12}' A2.txt
narac_female_gwas.assoc.logistic > narac_female_final.assoc.logistic

# and to the male dataset
awk 'NR==FNR {a[NR]=$1; next} {print $1, $2, $3, $4, a[FNR], $5, $6, $7, $8, $9, $10, $11, $12}' A2.txt
narac_male_gwas.assoc.logistic > narac_male_final.assoc.logistic

# and perform the meta analysis
metal metal.txt > metal.log

# lets take a look at the most significant gene and compare between the GWAS analyses
awk 'NR==1||$1=="rs660895"{print $0}' METAANALYSIS1.TBL
awk 'NR==1||$2=="rs660895"{print $0}' narac_female_gwas.assoc.logistic
awk 'NR==1||$2=="rs660895"{print $0}' narac_male_gwas.assoc.logistic

# lets take a look at the 9 most significant SNPs for all analyses
(head -n 1 METAANALYSIS1.TBL && tail -n +2 METAANALYSIS1.TBL | sort -k6,6g) | head
(head -n 1 narac_female_gwas.assoc.logistic && tail -n +2 narac_female_gwas.assoc.logistic | awk '$12 != "NA"' | sort
-k12,12g) | head
(head -n 1 narac_male_gwas.assoc.logistic && tail -n +2 narac_male_gwas.assoc.logistic | awk '$12 != "NA"' | sort
-k12,12g) | head
```

```
# Lets create a file for the meta data that can be used to visualize the data
awk 'NR==1 {print "SNP  P"} {print $1, $6}' METAANALYSIS1.TBL > meta_p.txt
awk 'NR != 2' meta_p.txt > tmp && mv tmp meta_p.txt

# remove the P value column so we can insert the P from the meta analysis
awk '{
    for (i = 1; i <= NF; i++) {
        if (i != 12) {
            printf "%s%s", $i, (i == NF || i == 11 ? ORS : OFS)
        }
    }
}' narac_female_gwas.assoc.logistic > meta_without_p.txt

# and add the p value fromt he meta analysis
awk '
    NR==FNR && FNR > 1 { pval[$1] = $2; next }
    FNR==1 { print $0, "P"; next }
    { print $0, (pval[$2] ? pval[$2] : "NA") }
' meta_p.txt meta_without_p.txt > meta_joined.txt

# remove the first row
awk 'NR > 1' meta_joined.txt > metaanalysis.txt

# finaize the file
sed -i 's/\r$//' metaanalysis.txt

# make a qqplot
Rscript --vanilla qqplot.R metaanalysis.txt Metaanalysis_QQPlot ADD

# and a manhattan plot
Rscript --vanilla gwaplot_pgen.R metaanalysis.txt "Manhattan Plot for Metaanalysis" manhattan_metaanalysis

### Task 3C: LD Score Regression

# first lets initilialize the path to the final folder
OKADA=/projectnb/bs859/data/RheumatoidArthritis/final_project

# and import the necessary packages
module load python2
module load ldsc

### ALL
# lets take a look at the data from both populations
zcat $OKADA/RA_GWASmeta_TransEthnic_v2.txt.gz|head
zcat $OKADA/RA_GWASmeta_TransEthnic_v2.txt.gz|wc

# format OKADA summary statistics for ldsc
# the reformatting process is repeated for the three analyses
zcat $OKADA/RA_GWASmeta_TransEthnic_v2.txt.gz | awk 'BEGIN {OFS="\t"} {
    if ($1 ~ /:/) {
        split($1, a, ":");
        $1 = "rs" a[2]
    }
    print
}' > RA_ALL_RS.txt

# add a column for BETA
gawk 'BEGIN {OFS="\t"}
NR==1 {print $0, "Beta"; next}
{print $0, log($6)}' RA_ALL_RS.txt > RA_ALL_BETA.txt

# and the sample size
awk 'BEGIN {OFS="\t"}
NR==1 {print $0, "N"; next}
{print $0, 80799}' RA_ALL_BETA.txt > RA_ready.txt
```

```
# rename the column OR_A1 as just OR
awk 'BEGIN {OFS="\t"} NR==1 { $6 = "OR" } { print }' RA_ready.txt > RA_ALL.txt

# intialize the path to the LD files
export LDDIR='/projectnb/bs859/data/ldscore_files'

# SNPID   Chr     Position_hg19  A1      A2      OR_A1  OR_95CIlow     OR_95CIup      P-val
# reformat the file for ldsc with munge sumstats
munge_sumstats.py \
  --sumstats RA_ALL.txt \
  --snp SNPID \
  --a1 A1 \
  --a2 A2 \
  --p P-val \
  --signed-sumstats Beta,0 \
  --N-col N \
  --merge-alleles $LDDIR/w_hm3.snplist \
  --out RA_ALL_DONE

# now we can run ldsc
ldsc.py \
--h2 RA_ALL_DONE.sumstats.gz \
--ref-ld $LDDIR/UKBB.ALL.ldscore/UKBB.EUR.rsid \
--w-ld $LDDIR/UKBB.ALL.ldscore/UKBB.EUR.rsid \
--out RA_ALL_LDSC

#### EUR
# lets take a look at the data from both populations
head $OKADA/RA_GWASmeta_European_v2.txt
wc $OKADA/RA_GWASmeta_European_v2.txt

# format OKADA summary statistics for ldsc
awk 'BEGIN {OFS="\t"} {
    if ($1 ~ /:/) {
        split($1, a, ":");
        $1 = "rs" a[2]
    }
    print
}' $OKADA/RA_GWASmeta_European_v2.txt > RA_EUR_RS.txt

gawk 'BEGIN {OFS="\t"}
NR==1 {print $0, "Beta"; next}
{print $0, log($6)}' RA_EUR_RS.txt > RA_EUR_BETA.txt

awk 'BEGIN {OFS="\t"}
NR==1 {print $0, "N"; next}
{print $0, 80799}' RA_EUR_BETA.txt > RA_EUR_ready.txt

awk 'BEGIN {OFS="\t"} NR==1 { $6 = "OR" } { print }' RA_EUR_ready.txt > RA_EUR.txt

export LDDIR='/projectnb/bs859/data/ldscore_files'

# SNPID   Chr     Position_hg19  A1      A2      OR_A1  OR_95CIlow     OR_95CIup      P-val
munge_sumstats.py \
  --sumstats RA_EUR.txt \
  --snp SNPID \
  --a1 A1 \
  --a2 A2 \
  --p P-val \
  --signed-sumstats Beta,0 \
  --N-col N \
  --merge-alleles $LDDIR/w_hm3.snplist \
  --out RA_EUR_DONE

# now we can run ldsc
ldsc.py \
--h2 RA_EUR_DONE.sumstats.gz \
```

```
--ref-ld $LDDIR/UKBB.ALL.ldscore/UKBB.EUR.rsid \
--w-ld $LDDIR/UKBB.ALL.ldscore/UKBB.EUR.rsid \
--out RA_EUR_LDSC

#### EAS
# lets take a look at the data from both populations
zcat $OKADA/RA_GWASmeta_Asian_v2.txt.gz|head
zcat $OKADA/RA_GWASmeta_Asian_v2.txt.gz|wc

# format OKADA summary statistics for ldsc
zcat $OKADA/RA_GWASmeta_Asian_v2.txt.gz | awk 'BEGIN {OFS="\t"} {
    if ($1 ~ /:/) {
        split($1, a, ":");
        $1 = "rs" a[2]
    }
    print
}' > RA_EAS_RS.txt

gawk 'BEGIN {OFS="\t"}
NR==1 {print $0, "Beta"; next}
{print $0, log($6)}' RA_EAS_RS.txt > RA_EAS_BETA.txt

awk 'BEGIN {OFS="\t"}
NR==1 {print $0, "N"; next}
{print $0, 80799}' RA_EAS_BETA.txt > RA_ready_EAS.txt

awk 'BEGIN {OFS="\t"} NR==1 { $6 = "OR" } { print }' RA_ready_EAS.txt > RA_EAS.txt

export LDDIR='/projectnb/bs859/data/ldscore_files'

munge_sumstats.py \
  --sumstats RA_EAS.txt \
  --snp SNPID \
  --a1 A1 \
  --a2 A2 \
  --p P-val \
  --signed-sumstats Beta,0 \
  --N-col N \
  --merge-alleles $LDDIR/w_hm3.snplist \
  --out RA_EAS_DONE

# now we can run ldsc
ldsc.py \
--h2 RA_EAS_DONE.sumstats.gz \
--ref-ld $LDDIR/UKBB.ALL.ldscore/UKBB.EAS.rsid \
--w-ld $LDDIR/UKBB.ALL.ldscore/UKBB.EAS.rsid \
--out RA_EAS_LDSC
```