**PAGE data description**

1. Primary/index SNPs

PAGE\_Lipids\_Replication\_Table\_updated.csv

Created from table in PAGE\_Lipids\_Replication\_updated.docx

54 rows, each represents a significant SNP for a given trait

Trait distribution - HDL: 18, LDL: 8, TC: 14, TG: 14

51 unique SNPs, since 3 SNPs significant for both LDL and TC

12:623129 (rs35882350), 2:17472023 (rs57520889), 7:137559799 (rs73729083)

2. Proxy SNPs

PAGE\_Lipids\_Replication\_Proxies\_updated.csv

Created from PAGE\_Lipids\_Replication\_Proxies\_updated.xlsx

716 rows, each represents a proxy SNP for the 51 primary/index SNPs

Trait distribution - HDL: 415, LDL: 28, TC: 64, TC\_LDL: 45, TG: 164

Proxy number distribution per index SNP

0 (only self): 9, 1-15 proxies: 35, 29-66 proxies: 6, 220 proxies: 1

**MVP data description**

1. Transethnic

Downloaded from Box folder:

MVP-CAP other local/Manuscripts/MVP lipids GWAS/Summary\_Statistics/

4 files, 1 per trait

MVP.te.HDL.gwas.tsv.gz

MVP.te.LDL.gwas.tsv.gz

MVP.te.TC.gwas.tsv.gz

MVP.te.TG.gwas.tsv.gz

Each row is a SNP, 9 column names are:

MarkerName, Allele1, Allele2, Freq1, FreqSE, Effect, StdErr, P.value, Direction

All files are ~570mb and contain ~29million SNPs

2. Race/ethnicity-specific

Downloaded from Box folder:

MVP GWAS/Lipids GWAS/1000G\_GWAS\_OUT\_GENESIS\_KLARIN/

12 files, 1 per trait per race/ethnicity

allchr.AFR.HDL.results\_10\_31\_17.gz

allchr.AFR.LDL.results\_10\_31\_17.gz

allchr.AFR.TC.results\_10\_31\_17.gz

allchr.AFR.TG.results\_10\_31\_17.gz

allchr.EUR.HDL.results\_10\_31\_17.gz

allchr..EUR.LDL.results\_10\_21\_17.gz (2 dots after allchr not a typo)

allchr.EUR.TC.results\_10\_31\_17.gz

allchr.EUR.TG.results\_10\_31\_17.gz

allchr.HIS.HDL.results\_10\_31\_17.gz

allchr.HIS.LDL.results\_10\_31\_17.gz

allchr.HIS.TC.results\_10\_31\_17.gz

allchr.HIS.TG.results\_10\_31\_17.gz

Each row is a SNP, 16 column names are:

rsid, chromosome, position, alleleA, alleleB, average\_maximum\_posterior\_call, info, all\_AA, all\_AB, all\_BB, all\_total, missing\_data\_proportion, cohort\_1\_hwe, frequentist\_add\_pvalue, frequentist\_add\_beta\_1, frequentist\_add\_se\_1

AFR and HIS files are ~1.4gb each and contain ~32million SNPs

EUR files are ~850-900mb each and contain ~20million SNPs

**Results**

Scripts backed up at <https://github.com/AustinTHilliard/MVP/tree/master/lipids/>

lipids\_replication\_transethnic\_updated.R

lipids\_replication\_ethnic-specific\_updated.R

Note that nothing can be reproduced with just the scripts, nor do they contain any results or sensitive information. They’re on github solely for backup purposes and are just pure R code with references to files on my local machine.

**Transethnic MVP**

PAGE\_Lipids\_Replication\_MVP.transethnic\_updated.xlsx

Sheet 1: all\_PAGE\_index\_SNPs

Same as PAGE\_Lipids\_Replication\_Table\_updated.csv, but with columns added containing the MVP results from the 4 transethnic files.

Sheet 2: all\_proxies\_for\_missing\_SNPs

Each row is a proxy for one of the missing primary/index SNPs. Columns 1-3 list the trait, index SNP, and proxy SNP, and remaining columns contain results for the proxy SNPs from the appropriate MVP transethnic file.

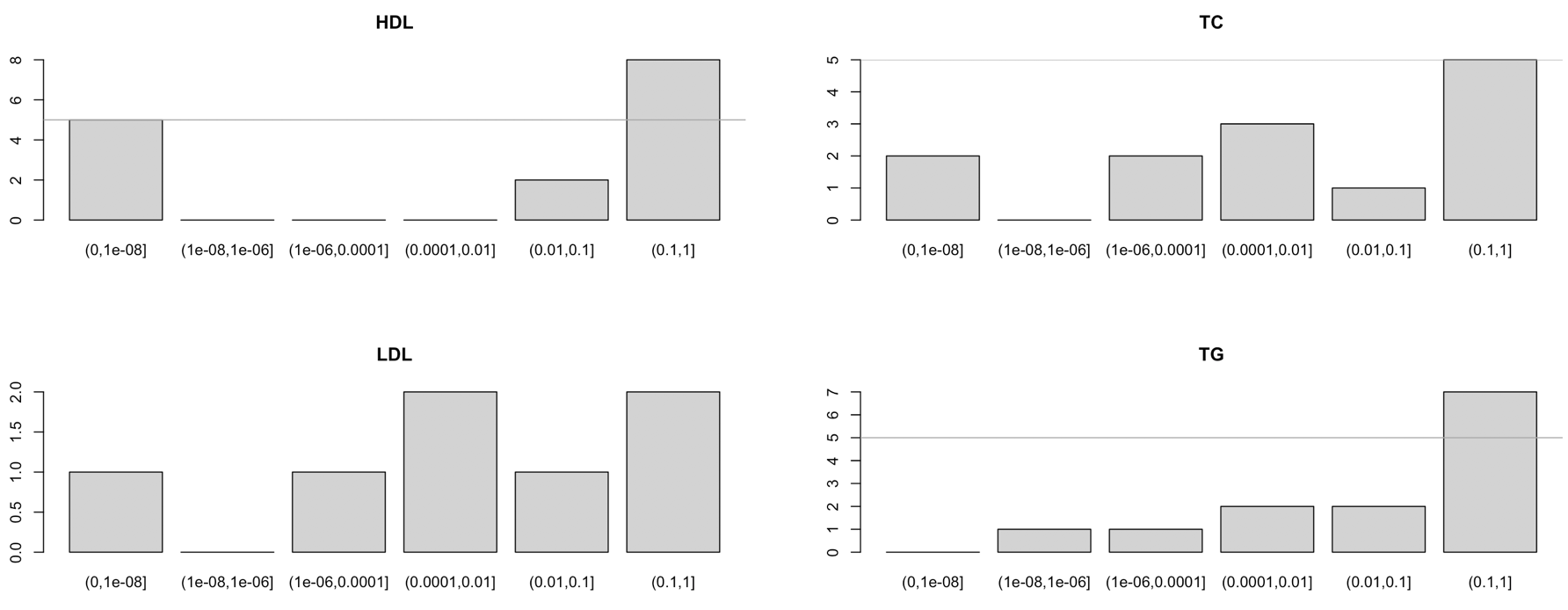
1. Primary PAGE SNPs

PAGE\_Lipids\_Replication\_MVP.transethnic\_updated.xlsx – Sheet 1

6 were not found in the MVP results

3 from HDL, and 1 each from LDL, TC, and TG

Rows with NA in all\_PAGE\_index\_SNPs

 MVP p-values for other primary PAGE SNPs in all\_PAGE\_index\_SNPs

y-axis is number of SNPs in each p-value range

Note: Although I haven’t systematically explored this, from spot-checking proxy SNPs it looks like proxies sometimes have p-values that are more significant than those of their primary/index SNPs.

2. Proxy SNPs for primary PAGE SNPs missing from MVP data

PAGE\_Lipids\_Replication\_MVP.transethnic\_updated.xlsx – Sheet 2

3 missing primary SNPs were listed as their own proxy, i.e. they didn’t have a proxy.

1 missing primary SNP had 3 proxies but none were found in MVP.

2 had at least 1 proxy SNP with results in MVP (1 had 1/2 proxies, 1 had 2/12)

**Race/ethnicity-specific MVP**

PAGE\_Lipids\_Replication\_MVP.ethnic-specific\_updated.xlsx

Sheets are named and mostly organized the same way as transethnic results

3x the number of rows as PAGE\_Lipids\_Replication\_Table\_updated.csv.

Each primary PAGE SNP was searched in each of 3 MVP race/ethnicity groups.

Columns 1-5 are the same as PAGE\_Lipids\_Replication\_Table\_updated.csv

Other columns contain results from the correct race/ethnicity-specific MVP file.

1. Primary PAGE SNPs

PAGE\_Lipids\_Replication\_MVP.ethnic-specific\_updated.xlsx – Sheet 1

Distribution by trait of primary PAGE SNPs missing from MVP data

HDL – 10 SNPs missing from 1 or more race/ethnicities

1 from AFR only

3 from EUR only

1 from EUR and HIS

5 from AFR, EUR, and HIS

LDL – 3 SNPs missing from 1 or more race/ethnicities

2 from EUR only\*

1 from AFR, EUR, and HIS

TC – 6 SNPs missing from 1 or more race/ethnicities

2 from EUR only\*

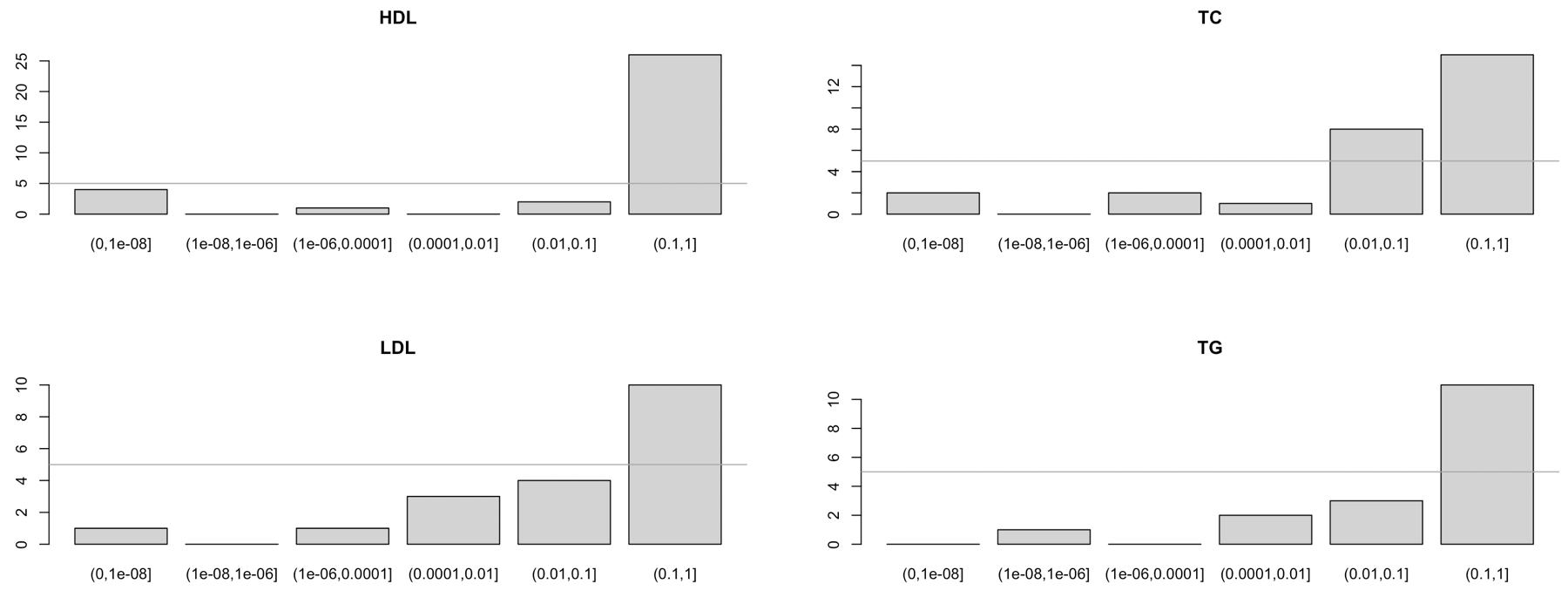
4 from AFR, EUR, and HIS

TG – 10 SNPs missing from 1 or more race/ethnicities

2 from EUR only

8 from AFR, EUR, and HIS

\*1 is same SNP, i.e. it’s missing in EUR from both LDL and TC

MVP p-values for other primary PAGE SNPs by trait

y-axis is number of SNPs in each p-value range

2. Proxy SNPs for primary PAGE SNPs missing from MVP data

PAGE\_Lipids\_Replication\_MVP.ethnic-specific \_updated.xlsx – Sheet 2

Out of 377 proxy SNPs, only 5 were found in the relevant MVP results

These 5 represented 2 missing primary/index SNPs

This seems strange, but I’ve spot-checked 10-20 of them and they were indeed missing