Overview of the Anthropometric Traits and CHD Bivariate Scan Project

**R Markdown file** refers to Voight\_Lab\_Notebook.Rmd

All outputs for this work can be found on the LPC at /project/voight\_GWAS/wbone/

**Versions of the bivariate scan script used in this project:**

**Generic** – This is the original version provided to me by Katie Siewert. Uses independent SNPs via a plink command to estimate the χ2 parameters for the bivariate scan H0 distribution. Each SNP that was in both trait summary statistic files is then tested to see if the trait/genotype distribution is significantly different from the expected H0 distribution. This χ22 goodness of fit test returns a final output all SNPs that are have a genome wide significant bivariate p-value or near genome wide significant bivariate p-value. It filters away any SNPs that have been previously associated with (or are in LD with an associated SNP) with either of the trait. This script can be found on the LPC at /project/voight\_GWAS/wbone/bivariate\_scan\_project/code/Bivar\_Script\_generalized.R. The exact code run for each analysis can be found in the analysis directory.

**All\_SNPs** – This version of the bivariate scan script does not filter out any SNPs (other than the requirement that each SNP is in both single trait GWAS summary statistic files) to estimate the χ2 parameters for the bivariate scan H0 distribution. This script can be found on the LPC at /project/voight\_GWAS/wbone/bivariate\_scan\_project/code/Bivar\_Script\_generalized\_all\_SNPs\_for\_var.R. The exact code run for each analysis can be found in the analysis directory.

Edits to Bivar\_Script\_generalized.R to get Bivar\_Script\_generalized\_all\_SNPs\_for\_var.R are documented on *Line 1267 R Markdown file*

**Var\_harm\_out\_is1** – This version of the bivariate scan script replaces the variance of the outcome trait used for harmonization step with a 1 in the variance/covariance matrix (for our purposes this was always the anthropometric trait, but would do the same for any trait put in the outcome trait position in the config.R file). NOTE: The covariance of the two traits *WILL NOT* be adjusted to reflect the variance being adjusted to 1. This script can be found on the LPC at /project/voight\_GWAS/wbone/bivariate\_scan\_project/code/Bivar\_Script\_generalized\_var\_trait1\_is1.R. The exact code run for each analysis can be found in the analysis directory.

Edits to Bivar\_Script\_generalized.R to get Bivar\_Script\_generalized\_var\_trait1\_is1.R are documented in *Lines 1864 – 1876 R Markdown file*

**Transform\_var\_harm\_out** – This version of the bivariate scan script transforms the outcome trait used for harmonization step to make sure the z-scores of the trait are standard normal. This makes it so the variance in the variance/covariance matrix of the trait is 1 and the covariance has been updated to reflect this trait’s Z-scores being standardized. This script can be found on the LPC at /project/voight\_GWAS/wbone/bivariate\_scan\_project/code/Bivar\_Script\_generalized\_transform\_var\_harm\_out.R. The exact code run for each analysis can be found in the analysis directory.

Edits to Bivar\_Script\_generalized.R to get Bivar\_Script\_generalized\_transform\_var\_harm\_out.R are documented in *Lines 1999 – 2004 R Markdown file*

**Transform\_var\_both** – This version of the bivariate scan script transforms the both traits used to make sure the z-scores of the trait are standard normal. This makes the variance in the variance/covariance matrix of both traits 1 and the covariance has been updated to reflect the traits’ Z-scores being standardized. This script can be found on the LPC at /project/voight\_GWAS/wbone/bivariate\_scan\_project/code/Bivar\_Script\_generalized\_transform\_both\_var.R. The exact code run for each analysis can be found in the analysis directory.

**March 30, 2018**

Downloaded Anthropometric trait GWAS summary statistic data –

**BMI data** from: GIANT, Yengo L. et al. 2018

[url:https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT\_consortium\_data\_files#WHRadjBMI\_.28download\_GZIP.29](url:%20https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files#WHRadjBMI_.28download_GZIP.29)

file name: Meta-analysis\_Locke\_et\_al+UKBiobank\_2018.txt.gz

Uploaded this file to /project/voight\_datasets/GWAS/04\_giant/

**Body Fat % data** from: LD hub, Lu Y. et al. 2016

[url:https://walker05.u.hpc.mssm.edu/](url:%20https://walker05.u.hpc.mssm.edu/)

file name: body\_fat\_percentage\_GWAS\_PLUS\_MC\_ALL\_ancestry\_se\_Sex\_combined\_for\_locus\_zoom\_plot.TBL.txt

Uploaded this files to /project/voight\_datasets/GWAS/23\_BodyFat2016/

**April 6, 2018**

**Check MRbase harmonization method on a few SNPs**

**Description:** Running a few SNPs through MRbase harmonization to make sure I understand the process and to make sure that it is handling palindromic SNPs appropriately**.** *Lines 62 -96 of R Markdown file*

***Results:* MRbase worked fine on these SNPs.**

**Bivariate Scan using generic bivariate scan script of HDL and CHD**

***Description:*** An outline of the analysis being performed by the generic bivariate scan can be seen at the beginning of this overview. Purpose of running this analysis was to confirm with Katie that I was in fact running the bivariate scan scripts correctly.*Lines 25 -60; 198 -209 of R Markdown file*

**April 12, 2018**

Checked that my HDL CHD bivariate p-value results perfectly correlated with Katie’s results.  *Lines 346 -409 of R Markdown file*

***Results:* Got the same mus, and variance/ covariance matrix as Katie. Several SNP p-values output by the HDL\_CHD\_bivar\_scan are the same as the ones that Katie got when she ran the code.**

**April 9, 2018**

**Bivariate Scan using generic bivariate scan script of BMI and CHD**

***Description:*** An outline of the analysis being performed by the generic bivariate scan can be seen at the beginning of this overview.

File formatting through the command to run the bivariate scan are all documented in: *Lines 99 -196 of R Markdown file*

**Input GWAS summary statistic files:**

/project/voight\_datasets/GWAS/21\_CADMeta/CAD\_META

/project/voight\_GWAS/wbone/bivariate\_scan\_project/BMI\_CHD\_bivar\_scan/BMI\_input\_files/Meta-analysis\_Locke\_et\_al+UKBiobank\_2018.txt

**Command:**

Bsub file submitted:

/project/voight\_GWAS/wbone/bivariate\_scan\_project/BMI\_CHD\_bivar\_scan/BMI\_CHD\_bivar\_scan\_04\_09\_2018.bsub

bsub < BMI\_CHD\_bivar\_scan\_04\_09\_2018.bsub -q voight\_normal

Job <13737521> is submitted to queue <voight\_normal>.

**Location of the output files on LPC:**

All outputs and the config.R file used for this run of bivariate scan are now located:

/project/voight\_GWAS/wbone/bivariate\_scan\_project/BMI\_CHD\_bivar\_scan/

**April 12, 2018**

Looking into high variation seen in the BMI data from Yengo L. et al. 2018: *Lines 412 -487 of R Markdown file*

***Results:*** **BMI had very high variance in the data. Upon review of the Yengo L. et al. 2018 paper and looking at the distribution of the betas for BMI in these data it was clear that these data had inherently high variability. Decided to try using MTAG for our multivariate scans to see if this could handle the variance in the Yengo L. et al. 2018 BMI data better.**

**April 9, 2018**

**Bivariate Scan using generic bivariate scan script of WHRBMI and CHD**

***Description:*** An outline of the analysis being performed by the generic bivariate scan can be seen at the beginning of this overview.

File formatting through the command to run the bivariate scan are all documented in: *Lines 212 - 331 of R Markdown file*

**Input GWAS summary statistic files:**

/project/voight\_datasets/GWAS/21\_CADMeta/CAD\_META

/project/voight\_GWAS/wbone/bivariate\_scan\_project/WHRadjBMI\_CHD\_bivar\_scan/WHRadjBMI\_input\_files/WHRadjBMI\_COMBINED\_EUR\_alleles\_and\_hg19pos\_noRowNames.txt

**Command:**

/project/voight\_GWAS/wbone/bivariate\_scan\_project/WHRadjBMI\_CHD\_bivar\_scan/ WHRadjBMI\_CHD\_bivar\_scan\_04\_09\_2018.bsub

bsub < WHRadjBMI\_CHD\_bivar\_scan\_04\_09\_2018.bsub -q voight\_normal

Job <13738618> is submitted to queue <voight\_normal>.

**Location of the output files on LPC:**

All outputs and the config.R file used for this run of bivariate scan are now located:

/project/voight\_GWAS/wbone/bivariate\_scan\_project/WHRadjBMI\_CHD\_bivar\_scan/WHRadjBMI\_04\_09\_2018\_output/

**April 12, 2018**

Review of WHRadjBMI and CHD bivariate scan results: *Lines 489 – 515 of R Markdown file*

***Results:* Got a variance less than 1 for WHRadjBMI which will likely give overly confident z-scores, however saw 4 hits rs17046742, rs6739187, rs7492628, rs9859406. Three of these are near SNPs associated with BMI and Diabetes.**

**April 12, 2018 – April 20, 2018**

**Testing MTAG Bivariate Scans on lipid traits and CHD**

***Description:*** Rerunning the analyses Katie performed using the bivariate scan code on the lipid data (High-density lipoprotein, Low-density lipoprotein, Triglycerides, and Total Cholesterol) and CHD data to see how MTAG performs provided metabolic traits for multivariate scans.

Input files for these analyses were reformatted (for MTAG compatibility) versions of the files Katie used for her lipid and CHD bivariate scans. /project/voight\_datasets/GWAS/14\_lipids/jointGwasMc\_HDL.txt, /project/voight\_datasets/GWAS/14\_lipids/jointGwasMc\_LDL.txt, /project/voight\_datasets/GWAS/14\_lipids/jointGwasMc\_TC.txt, and /project/voight\_datasets/GWAS/14\_lipids/jointGwasMc\_TG.txt.

**HDL and CHD MTAG Bivariate Scan**

File formatting HDL and CHD GWAS summary statistics files to go into MTAG and the commands to run a Multivariate scan in MTAG are documented in *Lines 532 – 941 R Markdown file*

**Input GWAS summary statistic files:**

/project/voight\_GWAS/wbone/MTAG\_multivariate\_scans/HDL\_CHD\_MTAG\_scan/HDL\_CHD\_sumstats\_inputs/jointGwasMc\_HDL\_formatted\_HDL\_dataDF\_PosNegBs\_no0or1freqs\_MTAG.txt

/project/voight\_GWAS/wbone/MTAG\_multivariate\_scans/HDL\_CHD\_MTAG\_scan/HDL\_CHD\_sumstats\_inputs/CAD\_META\_Ns\_Zs\_MTAG.txt

**Command to run MTAG:**

python /appl/mtag-20180412/mtag.py --sumstats /project/voight\_GWAS/wbone/MTAG\_multivariate\_scans/HDL\_CHD\_MTAG\_scan/HDL\_CHD\_sumstats\_inputs/jointGwasMc\_HDL\_formatted\_HDL\_dataDF\_PosNegBs\_no0or1freqs\_MTAG.txt,/project/voight\_GWAS/wbone/MTAG\_multivariate\_scans/HDL\_CHD\_MTAG\_scan/HDL\_CHD\_sumstats\_inputs/CAD\_META\_Ns\_Zs\_MTAG.txt --out ./MTAG\_HDL\_CHD\_scan --stream\_stdout --n\_min 0.0

**Location of the output files on LPC:**

project/voight\_GWAS/wbone/MTAG\_multivariate\_scans/HDL\_CHD\_MTAG\_scan

NOTE: HDL was the only lipid trait that was not converted to MTAG compatible format using /project/voight\_GWAS/wbone/MTAG\_multivariate\_scans/code/lipid\_GWAS\_MTAG\_parser.py

**LDL and CHD MTAG Bivariate Scan**

Used the same formatted CHD GWAS summary statistics file as were used for the HDL data to go into MTAG and used the script /project/voight\_GWAS/wbone/MTAG\_multivariate\_scans/lipid\_GWAS\_MTAG\_parser.py to format the LDL GWAS summary statistics data for MTAG. the commands to run a Multivariate scan with MTAG are documented in: *Lines 943 – 959 R Markdown file*

**Input GWAS summary statistic files:**

/project/voight\_GWAS/wbone/MTAG\_multivariate\_scans/TC\_CHD\_MTAG\_scan/jointGwasMc\_TC\_MTAG.txt

/project/voight\_GWAS/wbone/MTAG\_multivariate\_scans/LDL\_CHD\_MTAG\_scan/LDL\_CHD\_sumstats\_inputs/CAD\_META\_Ns\_Zs\_MTAG.txt

**Command to run MTAG:**

python /appl/mtag-20180412/mtag.py --sumstats /project/voight\_GWAS/wbone/MTAG\_multivariate\_scans/TC\_CHD\_MTAG\_scan/jointGwasMc\_TC\_MTAG.txt,/project/voight\_GWAS/wbone/MTAG\_multivariate\_scans/LDL\_CHD\_MTAG\_scan/LDL\_CHD\_sumstats\_inputs/CAD\_META\_Ns\_Zs\_MTAG.txt --out ./MTAG\_TC\_CHD\_scan --stream\_stdout --n\_min 0.0

**Location of the output files on LPC:**

/project/voight\_GWAS/wbone/MTAG\_multivariate\_scans/LDL\_CHD\_MTAG\_scan

**TC and CHD MTAG Bivariate Scan**

Used the same formatted CHD GWAS summary statistics file as were used for the HDL data to go into MTAG and used the script /project/voight\_GWAS/wbone/MTAG\_multivariate\_scans/lipid\_GWAS\_MTAG\_parser.py to format the TC GWAS summary statistics data for MTAG. the commands to run a Multivariate scan with MTAG are documented in: *Lines 961 – 967 R Markdown file*

**Input GWAS summary statistic files:**

/project/voight\_GWAS/wbone/MTAG\_multivariate\_scans/TC\_CHD\_MTAG\_scan/jointGwasMc\_TC\_MTAG.txt

/project/voight\_GWAS/wbone/MTAG\_multivariate\_scans/LDL\_CHD\_MTAG\_scan/LDL\_CHD\_sumstats\_inputs/CAD\_META\_Ns\_Zs\_MTAG.txt

**Command to run MTAG:**

python /appl/mtag-20180412/mtag.py --sumstats /project/voight\_GWAS/wbone/MTAG\_multivariate\_scans/TC\_CHD\_MTAG\_scan/jointGwasMc\_TC\_MTAG.txt,/project/voight\_GWAS/wbone/MTAG\_multivariate\_scans/LDL\_CHD\_MTAG\_scan/LDL\_CHD\_sumstats\_inputs/CAD\_META\_Ns\_Zs\_MTAG.txt --out ./MTAG\_TC\_CHD\_scan --stream\_stdout --n\_min 0.0

**Location of the output files on LPC:**

/project/voight\_GWAS/wbone/MTAG\_multivariate\_scans/TC\_CHD\_MTAG\_scan/

**TG and CHD MTAG Bivariate Scan**

Used the same formatted CHD GWAS summary statistics file as were used for the HDL data to go into MTAG and used the script /project/voight\_GWAS/wbone/MTAG\_multivariate\_scans/lipid\_GWAS\_MTAG\_parser.py to format the TG GWAS summary statistics data for MTAG. the commands to run a Multivariate scan with MTAG are documented in: *Lines 973 – 983 R Markdown file*

**Input GWAS summary statistic files:**

/project/voight\_GWAS/wbone/MTAG\_multivariate\_scans/TG\_CHD\_MTAG\_scan/jointGwasMc\_TG\_MTAG.txt

/project/voight\_GWAS/wbone/MTAG\_multivariate\_scans/LDL\_CHD\_MTAG\_scan/LDL\_CHD\_sumstats\_inputs/CAD\_META\_Ns\_Zs\_MTAG.txt

**Command to run MTAG:**

python /appl/mtag-20180412/mtag.py --sumstats /project/voight\_GWAS/wbone/MTAG\_multivariate\_scans/TG\_CHD\_MTAG\_scan/jointGwasMc\_TG\_MTAG.txt,/project/voight\_GWAS/wbone/MTAG\_multivariate\_scans/LDL\_CHD\_MTAG\_scan/LDL\_CHD\_sumstats\_inputs/CAD\_META\_Ns\_Zs\_MTAG.txt --out ./MTAG\_TG\_CHD\_scan --stream\_stdout --n\_min 0.0

**Location of the output files on LPC:**

/project/voight\_GWAS/wbone/MTAG\_multivariate\_scans/TG\_CHD\_MTAG\_scan

**April 23, 2018**

Review of MTAG lipid CHD bivariate scan results, including runs using the FDR function of MTAG for LDL CHD and HDL CHD bivariate scans: *Lines 985 - 1096 R Markdown file*

**Results:** **We found that MTAG was returning the SNPs that Katie is reporting in her lipids CHD bivariate scan manuscript, but that there are a lot of false positives particularly where a combination of nearly significant and not at all significant p-values were combined to move SNPs into the significant category. After reviewing these results, we decided to return to the bivariate scan code for our analyses.**

**May 3, 2018**

**Bivariate Scan generic bivariate scan script of using Locke et al. 2015 BMI data and CHD**

***Description:*** We wanted to run bivariate scan on the older BMI data hoping that without the meta-analysis that includes the UK Biobank data the λgc value and the variance of the BMI data would be reduced.

File formatting Locke et al. 2015 BMI and CHD GWAS summary statistics files to go into MTAG and the commands to run a Multivariate scan in MTAG are documented in *Lines 1105– 1184 R Markdown file*

**NOTE:** I have double checked that /project/voight\_GWAS/wbone/bivariate\_scan\_project/BMI\_Locke\_et\_al\_CHD\_bivar\_scan/BMI\_Locke\_et\_al\_CHD\_results\_05\_07\_2018 is the correct output data directory for the Locke\_2015\_BMI\_CHD\_bivar\_scan\_04\_09\_2018.bsub run. The unmatched dates are due to unfortunate typos.

**Input GWAS summary statistic files:**

/project/voight\_datasets/GWAS/21\_CADMeta/CAD\_META

/project/voight\_GWAS/wbone/bivariate\_scan\_project/BMI\_Locke\_et\_al\_CHD\_bivar\_scan/BMI\_input\_files/Locke\_GIANT\_2015\_BMI\_EUR\_withhg19pos.txt

**Command:**

Path to bsub:

/project/voight\_GWAS/wbone/bivariate\_scan\_project/WHRadjBMI\_CHD\_bivar\_scan/ WHRadjBMI\_CHD\_bivar\_scan\_04\_09\_2018.bsub

bsub < Locke\_2015\_BMI\_CHD\_bivar\_scan\_04\_09\_2018.bsub -q voight\_normal

Job < 13938353> is submitted to queue <voight\_normal>.

**Location of the output files on LPC:**

All outputs used for this run of bivariate scan are now located (accidentally overwrote config.R file for this run):

/project/voight\_GWAS/wbone/bivariate\_scan\_project/BMI\_Locke\_et\_al\_CHD\_bivar\_scan/BMI\_Locke\_et\_al\_CHD\_results\_05\_07\_2018

**NOTE:** I have double checked that /project/voight\_GWAS/wbone/bivariate\_scan\_project/BMI\_Locke\_et\_al\_CHD\_bivar\_scan/BMI\_Locke\_et\_al\_CHD\_results\_05\_07\_2018 is the correct output data directory for the Locke\_2015\_BMI\_CHD\_bivar\_scan\_04\_09\_2018.bsub run. The unmatched dates are due to unfortunate typos.

**May 4, 2018**

The commands run to review the Locke et al. CHD bivar scan data are in *Lines 1207 – 1237 R Markdown file*

***Result:* Using the Locke et al. BMI data gave us a BMI variance of 0.919 which is not idea since this means our z-scores will be overly confident when we calculate them. This is due to the level of genomic correction done in the original analysis.**

**May 7, 2018**

**Bivariate Scan All\_SNPs bivariate scan script of using WHRadjBMI data and CHD**

***Description:*** We wanted to try to alter the variance of WHRadjBMI to be 1 or greater to get more appropriate (more conservative) z-scores. We first tried to do this by including all SNPs in that were in both GWAS summary statistic files in the estimation of the genome wide variance of the WHRadjBMI trait. *Lines 1189 – 1204 R Markdown file*

**Input GWAS summary statistic files:**

/project/voight\_datasets/GWAS/21\_CADMeta/CAD\_META

/project/voight\_GWAS/wbone/bivariate\_scan\_project/WHRadjBMI\_CHD\_bivar\_scan/WHRadjBMI\_input\_files/WHRadjBMI\_COMBINED\_EUR\_alleles\_and\_hg19pos\_noRowNames.txt

**Command:**

Path to bsub:

project/voight\_GWAS/wbone/bivariate\_scan\_project/WHRadjBMI\_CHD\_bivar\_scan/ WHRadjBMI\_CHD\_bivar\_scan\_05\_03\_2018\_all\_vars\_for\_variance.bsub

bsub < WHRadjBMI\_CHD\_bivar\_scan\_05\_03\_2018\_all\_vars\_for\_variance.bsub -q voight\_normal

Job < 13948265> is submitted to queue <voight\_normal>.

**Location of the output files on LPC:**

**NOTE:** The original command was written on May 3rd, but the final command used was run on May 7th, thus why the bsub is dated 05/03/2018 and the data is saved in a directory dated 05/07/2018

Files moved to here in *Lines 1920 – 1940 R Markdown file*

/project/voight\_GWAS/wbone/bivariate\_scan\_project/WHRadjBMI\_CHD\_bivar\_scan/WHRadjBMI\_05\_07\_2018\_all\_vars\_var\_output

**May 9, 2018**

Comparison of significant hits from the generic bivariate scan on WHRadjBMI and CHD results from April 9th in the all\_SNPs bivariate scan results from May 7th. *Lines 1275 – 1315 R Markdown file*

***Results:* When using all variants to estimate variance in the WHRadjBMI data we got a variance of 1.08, but there were also no genome wide significant hits when using this increased variance. Previously genome wide significant hits were now just below significant.**

**May 9, 2018**

**Bivariate Scan All\_SNPs bivariate scan script of using Locke et al. BMI data and CHD**

***Description:*** We wanted to try to alter the variance of Locke et al. BMI to be 1 or greater to get more appropriate (more conservative) z-scores like we did with WHRadjBMI *Lines 1317 – 1351 R Markdown file*

See beginning of this document for description of how All\_SNPs bivariate works and an overview of how it works.

Edits to Bivar\_Script\_generalized.R to get Bivar\_Script\_generalized\_all\_SNPs\_for\_var.R are documented on *Line 1267 R Markdown file*

**Input GWAS summary statistic files:**

/project/voight\_datasets/GWAS/21\_CADMeta/CAD\_META

/project/voight\_GWAS/wbone/bivariate\_scan\_project/WHRadjBMI\_CHD\_bivar\_scan/WHRadjBMI\_input\_files/WHRadjBMI\_COMBINED\_EUR\_alleles\_and\_hg19pos\_noRowNames.txt

**Command:**

Path to bsub:

/project/voight\_GWAS/wbone/bivariate\_scan\_project/BMI\_Locke\_et\_al\_CHD\_bivar\_scan/ Locke\_2015\_BMI\_CHD\_bivar\_scan\_05\_09\_2018\_all\_vars\_var.bsub

bsub < Locke\_2015\_BMI\_CHD\_bivar\_scan\_05\_09\_2018\_all\_vars\_var.bsub -q voight\_normal

Job < 13960511-0 > is submitted to queue <voight\_normal>.

**Location of the output files on LPC:**

/project/voight\_GWAS/wbone/bivariate\_scan\_project/BMI\_Locke\_et\_al\_CHD\_bivar\_scan/BMI\_Locke\_et\_al\_CHD\_results\_05\_09\_2018

Output files moved to this directory May 14, 2018 commands to move these file: *Lines 1842 – 1862 R Markdown*

Comparison of significant hits from the generic bivariate scan on Locke et al. BMI and CHD results from May 7th in the all\_SNPs bivariate scan results. *Lines 1351 – 1416 R Markdown file*

***Results:* When using all variants to estimate variance in the Locke et al. BMI data we got a variance greater than 1, but there were also no genome wide significant hits when using this increased variance. Previously genome wide significant hits were now just below significant.**

**Bivariate Scan All\_SNPs bivariate scan script of using LDL and CHD**

***Description:*** To have a positive control of whether the all\_SNPs version of the bivariate scan script was estimating the variance of the outcome trait appropriately, we ran all\_SNPs on the LDL and CHD data that Katie had already identified and vetted real bivariate scan hits. *Lines 1416 – 1461 R Markdown file*

See beginning of this document for description of how All\_SNPs bivariate works and an overview of how it works.

**Input GWAS summary statistic files:**

/project/voight\_datasets/GWAS/21\_CADMeta/CAD\_META

/project/voight\_selscan/ksiewert/CardioMetaAnalysis/BivarAnnot/CADMeta\_ genvertest/HDL/jointGwasMc\_LDL\_formatted.txt

**Command:**

Path to bsub:

/project/voight\_GWAS/wbone/bivariate\_scan\_project/LDL\_CHD\_bivar\_all\_vars\_var/LDL\_CHD\_bivar\_scan\_05\_09\_2018.bsub

bsub < LDL\_CHD\_bivar\_scan\_05\_09\_2018.bsub -q voight\_normal

Job <13961261> is submitted to queue <voight\_normal>.

**Location of the output files on LPC:**

/project/voight\_GWAS/wbone/bivariate\_scan\_project/LDL\_CHD\_bivar\_all\_vars\_var/

The review of the All\_SNPs bivariate scan using LDL and CHD arein *Lines**1445 – 1461 R Markdown file*

***Results:* No SNPs were genome wide significant in the LDL CHD bivariate scan when using all variants to estimate the variance. We decided that we will need to try a new method for estimating the parameters of the null χ2 that considers the λgc’s of these data. This lead to the plan to run bivariate generic script on each anthropometric trait with CHD and if there is a variance less than 1 due to genomic correction we will conservatively adjust the variance to be 1.**

**May 9, 2018, May 14, 2018**

**Bivariate Scan generic bivariate scan script of using Body Fat % data and CHD**

***Description:*** We knew that not all traits had estimated variances less than 1 when using the bivariate scan generic code, so we wanted to see if the generic version of the code would produce a variance of 1. If the variance was 1 or greater we would use these results as they are and if the variance was less than 1 we will conservatively set the variance to 1 and reperform the bivariate scan. *Lines**1463 – 1493; 1736 – 1781 R Markdown file*

**Input GWAS summary statistic files:**

/project/voight\_datasets/GWAS/21\_CADMeta/CAD\_META

/project/voight\_GWAS/wbone/bivariate\_scan\_project/BodyFatPer\_CHD\_bivars can/BFP\_CHD\_input\_data/BFP\_chr\_pos\_from\_CHD.txt

**Command:**

path to bsub:

/project/voight\_GWAS/wbone/bivariate\_scan\_project/BodyFatPer\_CHD\_bivarscan/BFP\_CHD\_bivar\_scan\_05\_14\_2018.bsub

bsub < BFP\_CHD\_bivar\_scan\_05\_14\_2018.bsub -q voight\_normal

Job < 14026689> is submitted to queue <voight\_normal>.

**Location of the output files on LPC:**

Output files were moved here May 16, 2018 *Lines 1956 – 1977 R Markdown file*

/project/voight\_GWAS/wbone/bivariate\_scan\_project/BodyFatPer\_CHD\_bivarscan/BFP\_05\_14\_2018\_bivar\_scan\_output

***Results:* The variance of Body Fat % was estimated to be 0.969, so our plan was to conservatively correct the estimated variance to be 1.**

**May 14, 2018, May 15, 2018**

**Bivariate Scan Var\_harm\_out\_is1** **bivariate scan script of using Locke et al. BMI data and CHD**

***Description:*** We changed the bivariate scan code to change the variance of Locke et al. BMI to be 1 to get more appropriate (more conservative) z-scores but did not edit the covariance. We did not want to edit the covariance in case this inappropriately increased the power of the bivariate scan. *Lines 1864 – 1876; 1887 - 1912 R Markdown file*

See beginning of this document for description of how Var\_harm\_out bivariate works.

**Input GWAS summary statistic files:**

/project/voight\_datasets/GWAS/21\_CADMeta/CAD\_META

/project/voight\_GWAS/wbone/bivariate\_scan\_project/BMI\_Locke\_et\_al\_CHD\_bivar\_scan/BMI\_input\_files/Locke\_GIANT\_2015\_BMI\_EUR\_withhg19pos.txt

**Command:**

Path to bsub:

/project/voight\_GWAS/wbone/bivariate\_scan\_project/BMI\_Locke\_et\_al\_CHD\_bivar\_scan/ Locke\_2015\_BMI\_CHD\_bivar\_scan\_05\_14\_2018\_var\_trait1\_is1.bsub

bsub < Locke\_2015\_BMI\_CHD\_bivar\_scan\_05\_14\_2018\_var\_trait1\_is1.bsub -q voight\_normal

Job < 144027166 > is submitted to queue <voight\_normal>.

**Location of the output files on LPC:**

/project/voight\_GWAS/wbone/bivariate\_scan\_project/BMI\_Locke\_et\_al\_CHD\_bivar\_scan/BMI\_Locke\_et\_al\_05\_14\_2018\_var\_trait1\_is1\_bivar\_scan\_output

Review of the results of this run of bivar scan *Lines 1887 - 1912 R Markdown file*

***Results:* Running bivariate scan with the variance set to 1 for the BMI data seemed to give us a good middle ground between using every SNP in the data to estimate variance (too high) and using the generic bivariate scan method of estimating variance on these high genomic corrected data sets (too low). It would be preferable to have a more complete correction of the χ2 parameters, so we decided to standardize the genomic controlled z-scores for our anthropometric traits in order to get a variance of 1 and update the covariance with CHD accordingly.**

**May 16, 2018**

**Bivariate Scan Var\_harm\_out\_is1** **bivariate scan script of using WHRadj BMI data and CHD**

***Description:*** We changed the bivariate scan code to change the variance of WHRadjBMI to be 1 to get more appropriate (more conservative) z-scores but did not edit the covariance. We want to be able to compare the output of this method to standardizing the genomic corrected z-scores, which will update the covariance of the two traits as well. *Lines 1918 – 1953 R Markdown file*

See beginning of this document for description of how Var\_harm\_out bivariate works.

**Input GWAS summary statistic files:**

project/voight\_datasets/GWAS/21\_CADMeta/CAD\_META

/project/voight\_GWAS/wbone/bivariate\_scan\_project/WHRadjBMI\_CHD\_bivar\_scan/WHRadjBMI\_input\_files/WHRadjBMI\_COMBINED\_EUR\_alleles\_and\_hg19pos\_noRowNames.txt

**Command:**

Path to bsub:

/project/voight\_GWAS/wbone/bivariate\_scan\_project/WHRadjBMI\_CHD\_bivar\_scan/ WHRadjBMI\_CHD\_bivar\_scan\_05\_16\_2018\_var\_trait1\_is1.bsub

Job <14087061> is submitted to queue <voight\_normal>.

**Location of the output files on LPC:**

/project/voight\_GWAS/wbone/bivariate\_scan\_project/WHRadjBMI\_CHD\_bivar\_scan/

***Results:* Similar to the Locke et al. BMI data we saw that using a variance of 1 gave some genome wide significant hits, without using the overly liberal variance of less than 1 due to the data being genomic controlled.**

**Bivariate Scan Var\_harm\_out\_is1** **bivariate scan script of using Body Fat % data and CHD**

***Description:*** We changed the bivariate scan code to change the variance of Body Fat % to be 1 to get more appropriate (more conservative) z-scores but did not edit the covariance. We want to be able to compare the output of this method to standardizing the genomic corrected z-scores, which will update the covariance of the two traits as well. *Lines 1956 – 1997 R Markdown file*

See beginning of this document for description of how Var\_harm\_out bivariate works.

**Input GWAS summary statistic files:**

project/voight\_datasets/GWAS/21\_CADMeta/CAD\_META

/project/voight\_GWAS/wbone/bivariate\_scan\_project/BodyFatPer\_CHD\_bivarscan/BFP\_CHD\_input\_data/BFP\_chr\_pos\_from\_CHD.txt

**Command:**

Path to bsub:

/project/voight\_GWAS/wbone/bivariate\_scan\_project/BodyFatPer\_CHD\_bivarscan/BFP\_CHD\_bivar\_scan\_05\_16\_2018\_var\_trait1\_is1.bsub

bsub < BFP\_CHD\_bivar\_scan\_05\_16\_2018\_var\_trait1\_is1.bsub -q voight\_normal

Job <14087133> is submitted to queue <voight\_normal>.

**Location of the output files on LPC:**

Output files were moved to this location on May 16th 2018. The commands to move these files are on *Lines 2145 – 2163 R Markdown file*

/project/voight\_GWAS/wbone/bivariate\_scan\_project/BodyFatPer\_CHD\_bivarscan/BFP\_05\_16\_2018\_var\_trait1\_is1\_bivar\_scan\_output/

***Results:* Similar to the Locke et al. BMI and WHRadjBMIdata we saw that using a variance of 1 gave some genome wide significant hits, without using the overly liberal variance of less than 1 due to the data being genomic controlled.**

**Bivariate Scan Transform\_var\_both bivariate scan script of using Locke et al. BMI data and CHD**

***Description:*** We changed the bivariate scan code to change the variance of Locke et al. BMI to be 1 by standardizing the BMI z-scores before calculating their variance and covariance with CHD to get a more appropriate z-scores for our bivariate scans. We will compare these results to our bivariate scans where we simply edited the variance to be 1 to be sure that the update to the covariance is not giving us more significant bivariate z-scores than are appropriate. *Lines 2006 – 2105 R Markdown file*

**Input GWAS summary statistic files:**

/project/voight\_datasets/GWAS/21\_CADMeta/CAD\_META

/project/voight\_GWAS/wbone/bivariate\_scan\_project/BMI\_Locke\_et\_al\_CHD\_b ivar\_scan/BMI\_input\_files/Locke\_GIANT\_2015\_BMI\_EUR\_withhg19pos.txt

**Command:**

Path to bsub file:

/project/voight\_GWAS/wbone/bivariate\_scan\_project/BMI\_Locke\_et\_al\_CHD\_bivar\_scan /BFP\_CHD\_bivar\_scan\_05\_16\_2018\_transform\_var\_harm\_out.bsub

bsub < Locke\_2015\_BMI\_CHD\_bivar\_scan\_05\_16\_2018\_transform\_var\_harm\_out.bsub -q voight\_normal

Job < 14111845> is submitted to queue <voight\_normal>.

**Location of the output files on LPC:**

/project/voight\_GWAS/wbone/bivariate\_scan\_project/BMI\_Locke\_et\_al\_CHD\_bivar\_scan/

***Results:* The results of this analysis were very similar to the variance changed to 1, this confirmed that the recalibrated covariance associated with the standardized z-scores was not going to give overly confident bivariate scan results. We decided to perform this version of the bivariate scan on the other traits as well.**

**Bivariate Scan Transform\_var\_both bivariate scan script of using WHRadjBMI data and CHD**

***Description:*** We changed the bivariate scan code to change the variance of WHRadjBMI to be 1 by standardizing the WHRadjBMI z-scores before calculating their variance and covariance with CHD to get a more appropriate z-scores for our bivariate scans. We will compare these results to our bivariate scans where we simply edited the variance to be 1 to be sure that the update to the covariance is not giving us more significant bivariate z-scores than are appropriate. *Lines 2108 – 2142 R Markdown file*

**Input GWAS summary statistic files:**

/project/voight\_datasets/GWAS/21\_CADMeta/CAD\_META

/project/voight\_GWAS/wbone/bivariate\_scan\_project/WHRadjBMI\_CHD\_bivar\_scan/WHRadjBMI\_input\_files/WHRadjBMI\_COMBINED\_EUR\_alleles\_and\_hg19pos\_noRowNames.txt

**Command:**

Path to bsub file:

/project/voight\_GWAS/wbone/bivariate\_scan\_project/WHRadjBMI\_CHD\_bivar\_scan/ WHRadjBMI\_CHD\_bivar\_scan\_05\_16\_2018\_transform\_var\_harm\_out.bsub

bsub < WHRadjBMI\_CHD\_bivar\_scan\_05\_16\_2018\_transform\_var\_harm\_out.bsub -q voight\_normal

Job <14088400> is submitted to queue <voight\_normal>.

**Location of the output files on LPC:**

/project/voight\_GWAS/wbone/bivariate\_scan\_project/WHRadjBMI\_CHD\_bivar\_scan/

***Results:* The results of this analysis were very similar to the variance changed to 1, this confirmed that the recalibrated covariance associated with the standardized z-scores was not going to give overly confident bivariate scan results. We decided to perform this version of the bivariate scan on the other traits as well.**

**Bivariate Scan Transform\_var\_both bivariate scan script of using Body Fat % data and CHD**

***Description:*** We changed the bivariate scan code to change the variance of Body Fat % to be 1 by standardizing the Body Fat % z-scores before calculating their variance and covariance with CHD to get a more appropriate z-scores for our bivariate scans. We will compare these results to our bivariate scans where we simply edited the variance to be 1 to be sure that the update to the covariance is not giving us more significant bivariate z-scores than are appropriate. *Lines 2144 – 2174 R Markdown file*

**Input GWAS summary statistic files:**

/project/voight\_datasets/GWAS/21\_CADMeta/CAD\_META

/project/voight\_GWAS/wbone/bivariate\_scan\_project/BodyFatPer\_CHD\_bivarscan/BFP\_CHD\_input\_data/BFP\_chr\_pos\_from\_CHD.txt

**Command:**

Path to bsub file:

/project/voight\_GWAS/wbone/bivariate\_scan\_project/BodyFatPer\_CHD\_bivarscan /BFP\_CHD\_bivar\_scan\_05\_16\_2018\_transform\_var\_harm\_out.bsub

bsub < BFP\_CHD\_bivar\_scan\_05\_16\_2018\_transform\_var\_harm\_out.bsub -q voight\_normal

Job <14111845> is submitted to queue <voight\_normal>.

**Location of the output files on LPC:**

/project/voight\_GWAS/wbone/bivariate\_scan\_project/BodyFatPer\_CHD\_bivarscan/

***Results:* The results of this analysis were very similar to the variance changed to 1, this confirmed that the recalibrated covariance associated with the standardized z-scores was not going to give overly confident bivariate scan results. We decided to perform this version of the bivariate scan on the other traits as well.**

**May 17-18, 2018**

**Review of all three Anthropometric traits and CHD Bivariate Scan Results**

***Description:*** Reviewed the genome wide significant and near genome wide significant SNPs for each trait in the Transform\_var\_both results. I reviewed the genomic region of each SNP by making LocusZoom plots. *Lines 2176 – 2476 R Markdown file*

***Results:* For each set of bivariate scans there were 3 genome wide significant SNPs that I wanted to investigate further and a fourth near genome wide significant SNP that had a univariate p-value of 1.0e-3 or less for both trait that I thought was worthwhile to investigate further.**

**Check if these SNPs were reported in the most recent univariate GWAS’s for these traits**

***Description:*** Checked to see if there were SNPs within 1MB of our bivariate scan hits for either trait from univariate scans in the most in the most recent papers or the available summary statistics for the most recent publications on the univariate traits (based on the tables from the publication or the summary statistics that were made available). The publications we reviewed were: van der Harst et al. 2018 (CHD), Yengo et al. 2018 (BMI), Shungin et al. 2016 (WHRadjBMI), and Lu et al. 2016 (Body Fat %) *Lines 2492 – 2660 R Markdown file*

***Results:* I only found reported SNPs near rs889398 which was a near genome wide significant hit for BMI and CHD, and this SNPs is actually found to be genome wide significantly associated with BMI in the Yengo et al. BMI meta-analysis. I also found that rs556335 had a p-value of 0.8 in the Yengo et al. BMI meta-analysis and had the opposite direction of effect in the Yengo et al. data than in our bivariate scan which lead us to believe that this association is likely noise.**

**Vetted each SNP individual with a combiniation of colocalization experiments (using coloc), conditional analyses (using gcta’s COJO).**

**May 22, 2018**

**Vetting rs556335 BMI and CHD**

***Description:*** Even though this SNP was found to have a p-value of 0.8 and had the opposite direction of effect in the Yengo et al. BMI meta-analysis, we wanted to see what the coloc scores for the BMI and CHD peaks were. We performed conditional analysis on this SNP, since there was a second peak at the locus that seemed to be associated with the traits, but was only in low linkage with rs556335. *Lines 2811 – 3031 R Markdown file*

***Results:*** **Conditional Analysis showed that the second peak was very correlated with rs556335 and the coloc scores were PP.H4 = 0.21, PP4 / PP3 + PP4 = 0.91. These are strong coloc scores for a bivariate candidate, but is not strong enough evidence to consider this SNP a great candidate due to the Yengo et al. BMI meta-analysis p-value.**

**May 23- May 25, 2018**

**Vetting rs6739187 for WHRadjBMI and CHD**

***Description:*** Performed conditional analyses on peaks for both traits that were nearby. The peaks nearby had not been associated with WHRadjBMI before, but they had been associated with CHD in van der Harrst et al. 2018 and directly published to the NHGRI GWAS Catalog. *Lines 3075 – 3031; 4040 – 4145 R Markdown file*

***Results:*** **Conditional analysis showed that rs6739187’s association with CHD is only nominally effected by conditioning on 1550155. The other highly associated (but not genome wide significant) WHRadjSNPs were in linkage with rs6739187. Due to already beign sub-genome wide significant p-value (9.89e-8) before conditioning on this 1550155. The coloc results gave a PP.H0 of 0.67, which is fairly high for bivariate scans and suggest that this association could be noise specifically due to nearby CHD associations. The PP4 / PP3 + PP4 = 0.936**

**Vetting rs9859406 for WHRadjBMI and CHD**

***Description:*** Performed conditional analysis on a nearby WHRadjBMI sub-genome wide significant peak to make sure it is independent. Also performed and did a manual review of the NHGRI GWAS Catalog. *Lines 3260 – 3360 ; 4148 – 4177 R Markdown file*

***Results:*** **Conditional analysis showed that this peak was independent of the nearby sub-genome wide significant peak. I did find that this SNP is completely linked with a SNP associated with CHD directly reported to the NHGRI GWAS Catalog by van der Harrst et al. 2018, which our scripts missed due to the version of our download of the catalog. So, this SNP seems to be a real bivariate scan hit, but it has been previously reported.**

**Vetting rs10938397 for Body Fat % and CHD**

***Description:*** There were no peaks nearby for Body Fat percentage or CHD, so this region only required colocalization analysis. *Lines 3362 – 3498 R Markdown file*

***Results:*** **coloc results gave a PP2.H of 0.556 indicating that it does not trust the CHD peak. I think this conclusion is overly conservative particularly considering the that the PP4/ PP3+PP4 = 0.991. Assuming that both peaks are real, very high probability of colocalization.**

**Vetting rs1993709 for Body Fat % and CHD**

***Description:***We performed conditional analysis on a nearby CHD associated peak (sub-genome wide significance) to confirm that this peak was independent of it. We also performed colocalization analysis of the Body Fat % and CHD peaks, and did a manual review of the region in the NHGRI GWAS Catalog. *Lines 3498 – 3692 R Markdown file*

***Results:*** **Conditional analysis showed that the CHD association of rs1993709 was only marginally linked with the nearby CHD sub-genome wide significant peak. Coloc results gave a PP.H2 of 0.679, showing that it does not have high confidence in the CHD association. The PP4/ PP3+PP4 = 0.728 which is on the low end of the bivariate scan SNPs that I investigated during these analyses. This region only had BMI associations in it (~+/- 500 KB ) but this SNP was not associated with BMI before.**

**Vetting rs2127821 for BMI and CHD**

***Description:*** Performed COJO on nearby CHD associated SNPs and a NHGRI GWAS Catalog manual look up. *Lines 3694 – 3718; 3941 – 4037 R Markdown file*

***Results:*** **Conditional analysis showed that the CHD association of rs2127821 had its p-value reduced by 1000X, from 8.336E-08 to 1.52529e-05when conditioning on the nearby previously reported CHD associated SNPs. This makes it so that there is some evidence to support this SNP is allelic series for the CHD association, but it is not clear how to move forward with that hypothesis at this point. Also given the weaker association with BMI this SNP is definitely not genome wide significantly associated for the bivariate scan when conditioning on the nearby reported CHD SNPs.**

**Vetting rs11170468 for BMI and CHD**

***Description:*** There were no peaks nearby for BMI or CHD, so this region only required colocalization analysis. *Lines 4179 – 4293 R Markdown file*

***Results:*** **Coloc gave a PP.H2 of 0.740 and a particularly low PP.H4 of 0.131, but a decent PP4/ PP3+PP4 = 0.965. This suggests that this bivariate association could be due to noise, since the CHD association is particularly weak, and the BMI association is fairly strong.**

**Vetting rs17046742 for WHRadjBMI and CHD**

***Description:*** There were recent van der Harst et al. 2018 CHD SNPs that are near this SNP as well as some near genome wide significant WHRadjBMI SNPs to perform conditional analysis on. *Lines 4295 – 4457 R Markdown file*

***Results:* Conditional analysis showed that the WHRadjBMI association seems to be correlated with the sub-genome wide significant SNPs nearby. The CHD association is also somewhat correlated to rs1550115 which was reported by van der Harst et al. Due to this association and the LocusZoom plot that shows that rs17046742 there is a weak association to CHD particularly compared to the other SNPs at the locus, this bivariate scan hit is not a good candidate to further follow up.**

**May 29 -May 31, 2018**

**Vetting rs7492628 for WHRadjBMI and CHD**

***Description:*** We performed conditional analysis on a nearby sub-genome wide significant peak to confirm that rs7492628 was independent of it. We performed colocalization analysis between the WHRadjBMI association and the CHD association. I manually looked up the region in the NHGRI GWAS Catalog. We also did colocalization of the WHRadjBMI association with the eQTL peak for RPS6KA5 in thyroid, transformed fibroblasts, and skeletal muscle, since rs7492628 was called as a significant eQTL in the GTEx analyses. *Lines 4459 – 4639; 4718 – 5094 R Markdown file*

***Results:* Conditional analysis showed that rs7492628 is independent of the nearby sub-genome wide significant peak in CHD. Coloc gave a PP2.H2 of 0.616, a PP.H4 of 0.25, and PP4/ PP3+PP4 = 0.954, which are solid bivariate scan coloc values, and suggest that this is a real bivariate scan hit. The region is previously associated with waiste circumference, but not WHRadjBMI. The eQTL analysis show that there was stronger colocalization between WHRadyBMI and expression of RPS6KA5 in thyroid tissue in GTEx. Associations with skeletal muscle were poor enough that it is possible this is not a real colocalization. Overall it seem that this SNP is potentially a real bivariate scan hit.**