

February 3, 2021

This repository contains the SAS code to complete the genetic pleiotropy analysis described in Lutz et al., "Shared genetic etiology underlying late-onset Alzheimer's disease and posttraumatic stress syndrome" (1). Details of the statistical analysis method are described in the paper.

The SAS code is used for the steps necessary to calculate conditional probabilities for Alzheimer's disease conditional with PTSD at different significance levels. The code also produces the plots included in the manuscript. Note that the `Pleio_calc.sas` programs run the primary conditional probability analysis for each direction,  $Q(AD|PTSD)$  and  $Q(PTSD|AD)$ . The programs that produce the plots and tables must be modified to reflect the specific dataset names used for each direction and therefore run twice, once for  $AD|PTSD$  and once for  $PTSD|AD$ .

For studies that are using multiple phenotypes, Mark Brown at Wake Forest University School of Medicine created a version of these programs that work with macros and has kindly shared this code.

### **Alzheimer's disease and PTSD genetic pleiotropy**

**Pleio\_calc\_AD\_primary.sas:** calculates the conditional probabilities for Alzheimer's disease conditional with PTSD. This program uses the GWAS p values from the Alzheimer's disease and PTSD GWAS and produces an output SAS dataset that contains the conditional probabilities  $Q(AD|PTSD)$  for different p value thresholds. The SAS statement with the comment "secondary phenotype inclusion threshold" should be changed to run the program at different p value thresholds. The program also produces a fold enrichment plot.

**Pleio\_calc\_PTSD\_primary.sas:** calculates the conditional probabilities for PTSD conditional with Alzheimer's disease. This program uses the GWAS p values from the Alzheimer's disease and PTSD GWAS and produces an output SAS dataset that contains the conditional probabilities  $Q(PTSD|AD)$  for different p value thresholds. The SAS statement with the comment "secondary phenotype inclusion threshold" should be changed to run the program at different p value thresholds. The program also produces a fold enrichment plot.

**AD PTSD plots for paper.sas:** This program produces the fold enrichment plots and conditional qq plots presented in Lutz et al. (1). The program takes the individual conditional probability sas datasets produced in the first step (`pleio_calc` programs), concatenates the files into a "fold\_all" dataset and makes the plots. The program also produces a SAS dataset with the name "fold\_subs" that is used with the next program to produce the Table of FDR conditional statistics. The program should be updated to reflect the specific dataset names and labeling/scaling on the plots.

**AD\_PTSD\_fdr\_calc.sas:** produces the Table of conditional probabilities for the statistical analysis. This program uses the datasets from the AD PTSD plots for paper program (`fold_subs` dataset) and produces the tables of conditional probabilities reported in the paper. The program should be modified to use the specific dataset names and to set any thresholds for FDR level inclusion in the table. The final table is a SAS dataset with a name of `_fintab`.

### **Details:**

The main calculation programs “pleio\_calc” runs through the conditional FDR calculations and produces the graphics for the fold enrichment plots and conditional qq plots. There are a few code/parameter changes that need to be made for each level of significance that you want to use with the secondary phenotype. 3 lines of code get modified when this level/threshold is changed.

This statement sets the inclusion criteria/threshold for the secondary phenotype. The threshold is expressed as  $-\log_{10}(p)$

```
/* secondary phenotype inclusion threshold */  
if lp_crp >= 2.5 then output;
```

These statements at the end of the program (corresponding to the threshold above) put the results for the specific subset in a permanent sas dataset. The group = statement just sets the name of the group/threshold for plotting. The name of the dataset (like crp2p5) corresponds to the threshold. The p in 2p5 is just to signify the decimal in the threshold

```
data dsd.crp2p5;  
    set dsd.pleiocomb;  
    group = 2.5;
```

Prior threshold values were  $-\log_{10}(p)$  of 2, 2.5, 3, 3.5 and 4. Alternatively, you can look at distributions of the  $-\log_{10}(p)$  values for the secondary phenotypes and see what reasonable thresholds would be.

1) After you have run the pleio\_calc.sas code for the different thresholds, you will have a set of sas datafiles that can be merged to produce all of the results tables and plots. The first step is to make the fold change and qq plots for all of the groups. The sas code for this step is in the file AD PTSD plots for paper.sas. You will see that the first few lines are PROC appends the subset datasets into one file. This file will be used for all downstream analysis. Substitute whatever file names you use for the datasets. Then, be sure to comment out these lines if you rerun the program to adjust axis limits or titles, etc. You may want to change some of the axis limits and text, but they are probably fairly close to what will be needed.

2) The false discovery rate calculations are made in the file named AD\_PTSD fdr calc.sas. The important point about this file, though, is that it also brings in GWAS data from the primary studies since it would get summarized in the Tables in the paper. You will need to modify this code for the specific file names that you are using. The output is a permanent sas dataset named “fintab” and this Table contains most of the data needed to make the tables for the paper.

### Macro version of code

**Pleio\_calc\_macro\_primary.sas:** macro version of the forward direction genetic pleiotropy calculations. This file makes the calculations in the “pleio\_calc” code, however, it is structured as a macro that is called for each level of significance to be used.

**Pleio\_calc\_macro\_secondary.sas:** macro version of the reverse direction genetic pleiotropy calculations. This file makes the calculations in the “pleio\_calc” code, however, it is structured as a macro that is called for each level of significance to be used.

**Crp\_plots\_paper\_macro.sas:** macro version of the code to make plots for the paper. This program contains macros that produce the plot for each phenotype with a macro call.

**Crp\_fdr\_calc\_macro.sas:** macro version of the code to make the table of conditional probabilities for the statistical analysis. The program contains macros that produce the table for each phenotype with a macro call.

1. Lutz MW, Luo S, Williamson DE, Chiba-Falek O. Shared genetic etiology underlying late-onset Alzheimer's disease and posttraumatic stress syndrome. *Alzheimers Dement*. 2020 Sep;16(9):1280-1292. doi: 10.1002/alz.12128. Epub 2020 Jun 26. PMID: 32588970; PMCID: PMC7769164.