Write-Up ADDIS Re-run and Verification

University of Idaho Department of Statistical Science

Jarred Kvamme

2/25/2021

Operations in January and February

Re-analysis of GTEx Using MRPC-ADDIS

DETAILED DESCRIPTION:

Implementation of the ADDIS improvement to the MRPC algorithm: This involved adapting the existing scripts developed by M. Badsha to rerun the GTEx data using the ADDIS version of the FDR control. The distribution of model types (M0, M1, . . . M4, Other) relative to each tissue was retained from the analyses for both **MRPC-LOND** and **MRPC-ADDIS** for comparison. Additionally, a host of programs were developed to identify the specific classification of each trio analyzed in each tissue. These programs were further adapted to identify the number of trios for each tissue classified as cis or trans mediated (M1 type 1 or M1 type 2). Such information was again compared with results from **MRPC LOND** with the intention of understanding the differences in inferred networks between the two FDR control methods.

In General, **MRPC-ADDIS** loosens the rejection threshold such that more edges/directions are inferred. There was typically a reshuffling of Model types with some specific trends such as most M1's whose graph classification was changed was converted to M2 or M4. M0's were converted to either M1, M2, M3, or M4 (with a large number in each tissue transferring to M3). The path by which M1 was converted to M2 or M4 was generally a direction flip between the cis and trans leading to a dependence structure of the cis gene dependent on both the trans gene and variant. The other method was the inference of an direct edge between the variant and both genes and a general edge between the cis and trans gene.

MAIN POINTS AND FILE LOCATIONS:

• Re-run of GTEx data with **MRPC** using the **ADDIS** FDR control
→ Done using a slightly modified version of M. Badsha's scripts.
→ All associated output data files are stored at:

$$/mnt/ceph/jarredk/AddisReRunFiles$$

→ The script used to Re-run **MRPC** using **ADDIS** is located at:

$$/mnt/ceph/jarredk/ADDIS\_verify/ADDIS\_ReRun.R$$

• Summarizing of differences between **LOND** and **ADDIS** trio classifications:
→ quantified the distribution of trio classifications for each tissue as counts of (M0, M1, . . . M4, Other) →
table can be accessed through the cluster at:

$$/mnt/ceph/jarredk/ADDIS\_verify/Trios\_analysis\_GTEx\_v8\_allPCs\_V3\_ordered\_updated.xlsx$$

• followed up on FDR methods by investigating specific trios and their classification differences between FDR methods:

GENERAL STEPS:

1. read in SNP, PC, and final output files generated from $ADDIS\_rerun()$ (trio–>model information)

2. catalog and store each trios **LOND** and **ADDIS** model classification

3. count/summarize trios between methods or within tissues (generally both)

4. rerun MRPC on trio using both **LOND** and **ADDIS** methods and compare graph and correlation matrices

→ R-functions developed to assist in tracking trios are in:

$$/mnt/ceph/jarredk/ADDIS\_verify/ADDIS\_Post\_Analysis\_processing.R$$

• followed up and compared mediation model trios of Type-1 (cis mediated) and Type-2 (trans-mediated) for both **ADDIS** and **LOND**
GENERAL STEPS:

1. read final output files generated from $ADDIS\_rerun()$ for both FDR methods (trio–>model information)

2. use **MRPC** to re-classify each trio identified as M1 to obtain causal graph matrix for next step

3. match trio to appropriate M1-type (1 or 2) by comparing causal graph matrix to type-1 and type-2 expected matrix

4. store M1-type classification information in a tiered list and save

→ function to classify M1-type is called $ADDIS.M1.Check()$ and is part of the R-script:

$$/mnt/ceph/jarredk/ADDIS\_verify/ADDIS\_Post\_Analysis\_processing.R$$

→ tables summarizing the counts and relative proportions of cis/trans mediated M1 models can be found at:

$$/mnt/ceph/jarredk/ADDIS\_verify/$$

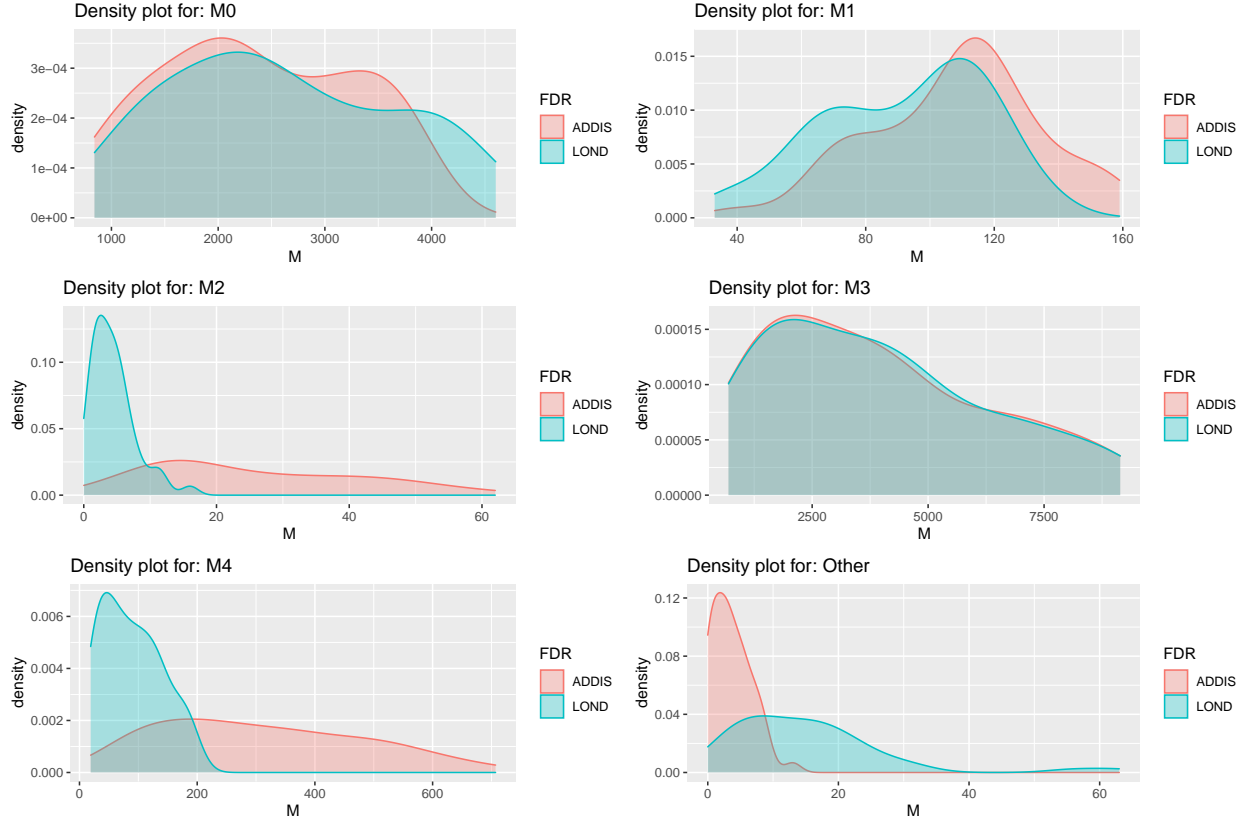# Summarizing Information For LOND/ADDIS Comparisons

Figure 1: Density plots showing the distribution of each graph type under both ADDIS and LOND

Table 1: table of the average difference in count and percentage between ADDIS and LOND and their standard errors for the differences

	M0	M1	M2	M3	M4	Other
mean.change.ct	-235.06250	16.47917	21.41667	-13.83333	222.45833	-11.45833
SD.change.ct	187.55248	12.39078	13.72191	127.75214	121.25373	10.03814
mean.change%	-0.01742	0.00116	0.00166	-0.00107	0.01653	-0.00085
SD.change%	0.02104	0.00200	0.00186	0.01165	0.01706	0.00125

Table 2: Table of confidence intervals for the average count of each classified graph for all tissues under the LOND FDR control

	lower.limit	mean	upper.limit
M0	2594.695	2603.733	2612.772
M1	91.557	91.779	92.001
M2	4.439	4.468	4.498
M3	3899.157	3919.076	3938.995
M4	88.827	89.278	89.730
Other	15.102	15.213	15.323

Table 3: Table of confidence intervals for the average count of each classified graph for all tissues under the ADDIS FDR control

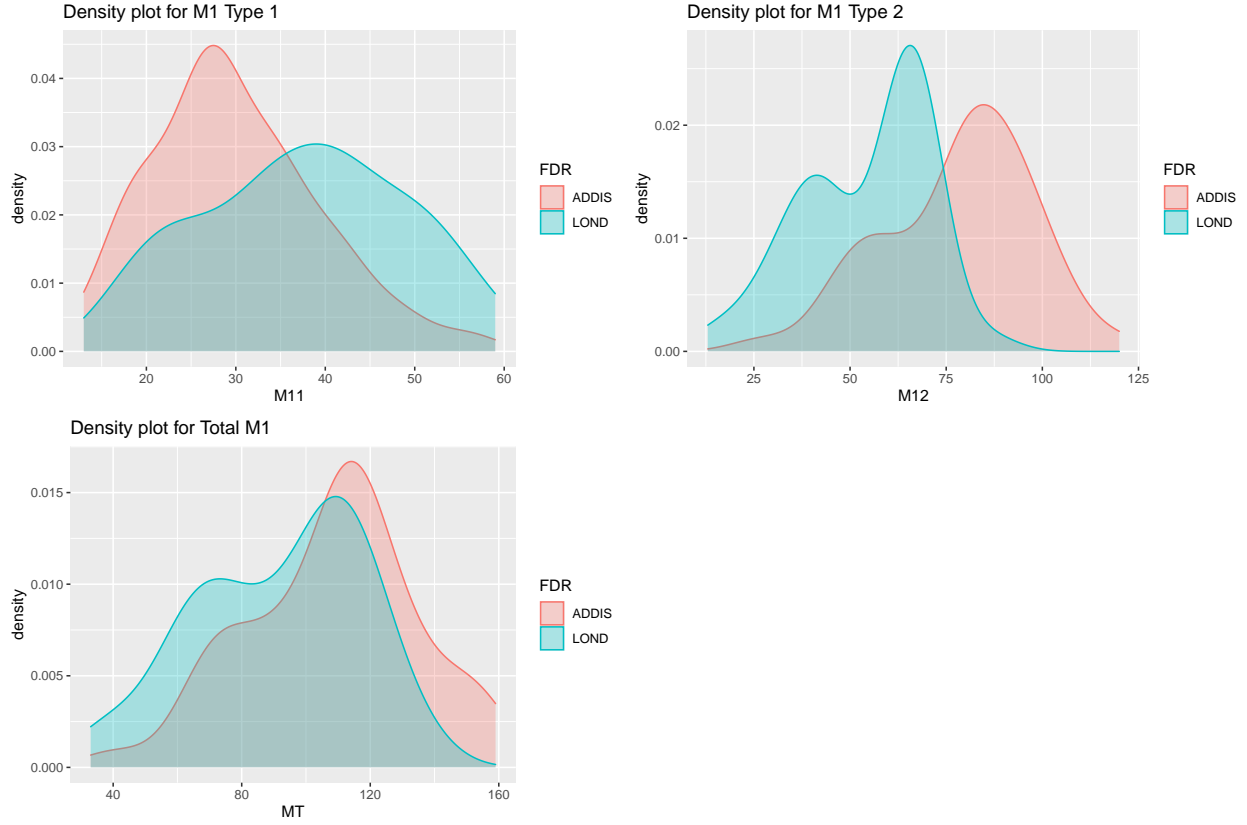|	lower.limit	mean	upper.limit
M0	2365.036	2373.098	2381.160
M1	107.884	108.117	108.349
M2	25.733	25.874	26.015
M3	3884.267	3905.882	3927.498
M4	310.577	312.077	313.577
Other	3.639	3.665	3.691

Figure 2: Density plots summarizing the differences between the relative counts for model 1 type 1 (M11) and Model 2 type 2 (M12) as well as the total distribution of Model 1's (MT)