

A Comparison of GMAC and ADDIS

Jarred Kvamme, University of Idaho

9/22/2021

Overview

Both MRPC-LOND and MRPC-ADDIS techniques inferred a large number of trans mediated trios. The trans-mediation model has been previously identified, but is not the commonly acknowledged mode of mediation. Since this result is surprising relative to the existing literature, we sought to apply another method for inferring mediation on a subset of GTEx trios analyzed herein by MRPC. The Genomic Mediation analysis with Adaptive Confounding (GMAC) algorithm allows for a unique selection of a subset of potential confounders, \mathbf{X}_{ij} from a larger covariate pool, \mathbf{H} , for each trio. By taking advantage of the Principle of Mendelian Randomization, the authors filter \mathbf{H} by removing common child and intermediate confounding variables (e.g variables associated with the eQTL as well as the cis/trans genes). Post-filtering, GMAC performs a mediation test on the edge between the cis gene and trans gene via the regression of the trans-gene T_j on the cis-eQTL L_i , cis-gene C_i , and the set of adaptively selected confounders \mathbf{X}_{ij} :

$$T_j = \beta_0 + \beta_1 C_i + \beta_2 L_i + \mathbf{\Gamma} \mathbf{X}_{ij} + \epsilon$$

The mediation statistic is the observed t -value of the cis-gene coefficient β_1 . A null distribution for no-mediation is constructed by iteratively permuting the values of the cis-transcript conditioned on each genotype and repeating the above regression. The authors argue that the permutation of the cis-transcript within the genotypes of the cis-eQTL removes the association between the cis and trans gene transcripts while preserving the higher order associations with the cis-eQTL. The resulting mediation test compares the observed relationship between the trans and cis gene to a null distribution constructed from a model with no association and assuming that possible confounding has been well adjusted via the selected covariates.

It is important to note that the above mediation test describes only the association between cis-gene and trans-gene transcripts ($C_i \leftrightarrow T_j$) and does not consider possible effects between the cis-eQTL and the cis-gene transcript ($L_i \rightarrow C_i$), or the cis-eQTL and trans-gene transcript ($L_i \rightarrow T_j$).

2. Methods

To compare the GMAC and MRPC algorithms, we applied the GMAC algorithm to the top five GTEx tissues by sample size. Following with the creators of GMAC, we used the full set of principle components retained from the PCA of the expression matrix as the covariate pool, and three additional known confounders: the PCR used, the platform used, and sex of the individual in each sample (Yang et al. 2017).

Consistent with Yang et al. (2017), the analysis was performed using a common child and intermediate variable filtering FDR of 10% and a confounder selection FDR of 5% for each trio. Each trio supplied to GMAC consisted of the cis-QTL and the PEER normalized cis and trans gene transcripts with the highest association to the eQTL. To mitigate missing values in the eQTL matrix, multiple imputation of the matrix of unique cis-eQTLs was performed via multiple correspondence analysis (MCA) prior to its use in the

GMAC (Josse, Husson, and others 2016). The analysis was preformed twice on each trio, first with the cis gene as the mediator and second with the trans gene as the mediator.

After applying GMAC to each tissue, the false discovery rate among the retained mediation p-values was controlled at the more liberal rate of 10% (Yang et al. 2017). Each trio determined to have significant mediation after FDR filtering was compared with the regulatory network type inferred by MRPC-ADDIS. MRPC-ADDIS can infer three types of regulatory networks that contain an edge between the cis and trans gene (M1, M2, or M4). Since GMAC considers only the presence of the edge and not its direction, trios inferred to be one of M1, M2, or M4 under ADDIS that were also significant under GMAC were considered consistent (e.g $C_i \rightarrow T_j$; $T_j \rightarrow C_i$; $C_i \leftrightarrow T_j$ are synonymous under GMAC).

Table 1: Descriptive statistics for the distribution of missing values across the eQTL’s for each tissue used in GMAC

	Adipose Subcutaneous	Artery Tibial	Muscle Skeletal	Skin Sun Exposed	Whole Blood
Min.	0.000000	0.000000	0.000000	0.000000	0.000000
1st Qu.	0.000000	0.000000	0.000000	0.000000	0.000000
Median	0.000000	0.000000	0.000000	0.000000	0.000000
Mean	0.006365	0.006625	0.006560	0.006701	0.006103
3rd Qu.	0.003442	0.003425	0.002833	0.003306	0.002985
Max.	0.156627	0.159247	0.158640	0.160331	0.155224

3. Results

At the 10% false discovery rate, GMAC identified 2,243 trios with cis/trans mediation out of 55,446 total trios tested across the five tissues: Adipose subcutaneous, Tibial artery, Muscle skeletal, Sun exposed skin, and Whole blood. 653 of the trios were identified as cis mediation, 245 as trans mediation and 1,345 as both (29.1%, 10.9%, and 60% respectively).

Among the significant trios retained, we refer to the unique set of trios as the union of the non-disjoint sets of trios identified as cis-gene mediated or trans-gene mediated (see **Table 3**). As can be seen from **Table 3**, the consistency in inferred mediation edges between the two methods varied between 33% and 46% of the trios across tissues. We observed two primary differences in the inference of the two algorithms; i) First, let \mathbf{Z}_{ij} denote the set of confounding PC’s used under ADDIS such that the columns $\{z_{ij}\} \subset \{x_{ij}\}$. The inclusion of a larger set of confounding variables as seen in GMAC could change the significance of the mediation p-value (prior to permutation) ii)

Table 2: The breakdown of unique trios with inferred significant cis or trans mediation under GMAC across their respective ADDIS inferred regulatory networks. The column “Percentage In Common” is the proportion of significant trios that also contained a mediation edge in the regulatory network inferred under ADDIS

Tissue	M0	M1	M2	M3	M4	Other	Total GMAC Inferred	Percentage In Common
AdiposeSubcutaneous	109	43	7	126	132	3	420	0.4333
ArteryTibial	89	25	5	139	110	0	368	0.3804
MuscleSkeletal	126	31	7	183	118	3	468	0.3333
SkinSunExposed	106	33	12	185	138	0	474	0.3861
WholeBlood	122	55	22	152	162	0	513	0.4659

Table 3: Breakdown of trios with inferred mediation under GMAC across both cis and trans mediation types. The column “Unique Both” represents the intersect of columns 2 and 3

Tissue	Total Trios	Total Cis Mediated	Total Trans Mediated	Unique Cis Only	Unique Trans Only	Unique Both	Unique Total
AdiposeSubcutaneous	11485	375	298	122	45	253	420
ArteryTibial	11471	341	228	140	27	201	368
MuscleSkeletal	10257	412	321	147	56	265	468
SkinSunExposed	13045	430	332	142	44	288	474
WholeBlood	8823	440	411	102	73	338	513

References

- Josse, Julie, François Husson, and others. 2016. “MissMDA: A Package for Handling Missing Values in Multivariate Data Analysis.” *Journal of Statistical Software* 70 (1): 1–31.
- Yang, Fan, Jiebiao Wang, Brandon L Pierce, Lin S Chen, François Aguet, Kristin G Ardlie, Beryl B Cummings, et al. 2017. “Identifying Cis-Mediators for Trans-eQTLs Across Many Human Tissues Using Genomic Mediation Analysis.” *Genome Research* 27 (11): 1859–71.