Controlling and Testing Horizontal Pleiotropy with Probabilistic Mendelian Regression for Transcriptome-wide Association Studies

Xiang Zhou

John G. Searle Assistant Professor
Department of Biostatistics
University of Michigan
Transcriptome-wide Association Studies

• Genome-wide association studies (GWASs) have identified many genetic variants associated with diseases and complex traits.

• Expression quantitative trait loci (eQTL) mapping studies have also identified enabled accurate measurements of gene expression levels.

• Integrative analysis of GWASs and eQTL mapping studies has the potential to yield insight into the causal relationship between genes and complex traits.
Existing Integrative Approaches: PrediXcan

GReX
Genetically regulated expression

Other factors

Trait-altered component

Gene expression decomposition

PrediXcan

Trait

Gamazon et al., Nature Genetics, 2015
Existing Integrative Approaches: PrediXcan

“SNP aggregation approach”

GReX
Genetically regulated expression

Other factors

Trait-altered component

Gene expression decomposition

Gamazon et al., Nature Genetics, 2015
Existing Integrative Approaches: PrediXcan

“SNP aggregation approach”

Step 1: Construct a genetic predictor of gene expression using ElasticNet

GReX
Genetically regulated expression

Other factors

Trait-altered component

Gene expression decomposition

Gamazon et al., Nature Genetics, 2015
Existing Integrative Approaches: PrediXcan

"SNP aggregation approach"

Step 1: Construct a genetic predictor of gene expression using ElasticNet

Step 2: Test the association between genetic predictor of expression and trait

Gamazon et al., Nature Genetics, 2015
Existing Integrative Approaches: TWAS

Gusev et al., Nature Genetics, 2016
Existing Integrative Approaches: TWAS

"expression-trait associations"

Gusev et al., Nature Genetics, 2016
Existing Integrative Approaches: TWAS

“expression-trait associations”

Step 1: Construct a genetic predictor of gene expression using BSLMM

Gusev et al., Nature Genetics, 2016
Existing Integrative Approaches: TWAS

"expression-trait associations"

Step 1: Construct a genetic predictor of gene expression using **BSLMM**

Step 2: Test the association between genetic predictor of expression and trait

Gusev et al., Nature Genetics, 2016
Existing Integrative Approaches: SMR

Zhu et al., Nature Genetics, 2016
Existing Integrative Approaches: SMR

“identify causal genes”

Zhu et al., Nature Genetics, 2016
Existing Integrative Approaches: SMR

Step 1: Construct a genetic predictor of gene expression using linear regression with one SNP

“identify causal genes”

Zhu et al., Nature Genetics, 2016
Existing Integrative Approaches: SMR

Step 1: Construct a genetic predictor of gene expression using linear regression with one SNP

Step 2: Test the association between genetic predictor of expression and trait

“identify causal genes”

Zhu et al., Nature Genetics, 2016
Mendelian Randomization

• These existing approaches can all be thought of as a two-stage regression version of Mendelian randomization (MR) analysis.

• MR is a form of instrumental variable analysis with SNPs serving as instruments.

• MR is a powerful statistical tool to determine causal relationship between an exposure variable (in this case, gene expression) and an outcome variable (in this case, complex trait) in observational studies.
Mendelian Randomization

\[ \text{Gene Expression} \xrightarrow{\alpha} \text{Trait} \]
Mendelian Randomization

Gene Expression

\[ x \]

\[ \alpha \]

Trait

\[ y \]

Confounders

\[ U \]
Mendelian Randomization

SNPs \( Z \) \( \rightarrow \beta \) Gene Expression \( x \) \( \rightarrow \alpha \) Trait \( y \)

Confounders \( U \)
Mendelian Randomization

Horizontal Pleiotropy: $\gamma$

SNPs $Z$ $\xrightarrow{\beta}$ Gene Expression $x$ $\xrightarrow{\alpha}$ Trait $\gamma$

Confounders $U$
Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases

Marie Verbanck, Chia-Yen Chen, Benjamin Neale, and Ron Do
MR with Horizontal Pleiotropy Accounted for

• Sample I, the observed gene expression data:

\[ x = \mu_x + Z_x \beta + \epsilon_x \]  \hspace{1cm} (1)
MR with Horizontal Pleiotropy Accounted for

• Sample I, the observed gene expression data:

\[ x = \mu_x + Z_x \beta + \varepsilon_x \]  

(1)

• Sample II, the unobserved gene expression data:
MR with Horizontal Pleiotropy Accounted for

• Sample I, the observed gene expression data:

\[ x = \mu_x + Z_x \beta + \varepsilon_x \]  \hspace{1cm} (1)

• Sample II, the unobserved gene expression data:

\[ \tilde{x} = \mu_x + Z_y \beta + \varepsilon_y \]  \hspace{1cm} (2)
MR with Horizontal Pleiotropy Accounted for

• Sample I, the observed gene expression data:

\[ x = \mu_x + Z_x \beta + \varepsilon_x \]  

(1)

• Sample II, the unobserved gene expression data:

\[ \tilde{x} = \mu_x + Z_y \beta + \varepsilon_y \]  

(2)

• Sample II, the observed GWAS data:
MR with Horizontal Pleiotropy Accounted for

• Sample I, the observed gene expression data:
  \[ x = \mu_x + Z_x \beta + \varepsilon_x \quad (1) \]

• Sample II, the unobserved gene expression data:
  \[ \tilde{x} = \mu_x + Z_y \beta + \varepsilon_y \quad (2) \]

• Sample II, the observed GWAS data:
  \[ y = \mu_y + \tilde{x} \alpha + Z_y \gamma + \varepsilon \quad (3) \]
Additional Modeling Assumptions

• Because the number of SNPs ($p$) is often larger than the sample size ($n$), we need to make additional modeling assumption for model identifiability.

• For $\bm{\beta}$, we follow standard polygenic models to assume $\beta_j \sim N(0, \sigma_\beta^2)$.

• For $\bm{\gamma}$, we follow Egger regression to assume $\gamma_1 = \cdots = \gamma_p = \gamma$
Probabilistic Mendelian Randomization

• Instead of the usual two-stage regression procedure, we rely on the maximum likelihood estimation procedure for inference.

• We develop a computationally efficient fitting algorithm, based on a parameter expansion version of the expectation maximization algorithm (PX-EM).

• We test causal effect \( H_0: \alpha = 0 \) through LRT.

• We test horizontal pleiotropic effect \( H_0: \gamma = 0 \) through LRT.

• We refer to our method as PMR-Egger.
Simulations

- We extracted $p = 556$ cis-SNPs of a gene from the GEUVADIS data ($n_1 = 465$) and simulated gene expression.

- We extricated the same SNPs from 2,000 controls in the Wellcome trust case control consortium (WTCCC) and simulated trait.

- We examined various scenarios, with 10,000 replicates for each scenario.
Compared Methods: Testing $\alpha$

- PrediXcan: Elastic Net prior on $\beta$; no $\gamma$; two-stage inference
- TWAS: BSLMM prior on $\beta$; no $\gamma$; two-stage inference
- SMR: Single $\beta$; no $\gamma$; two-stage inference
- CoMM: Normal prior on $\beta$; no $\gamma$; maximum likelihood inference
- LDA MR Egger: Fixed effects of $\beta$; Egger assumption on $\gamma$; two-stage inference
- PMR-Egger: Normal prior on $\beta$; Egger assumption on $\gamma$; maximum likelihood inference
Testing Causal Effect $\alpha$ under the Null

$\gamma = 0$

$\gamma = 0.001$

- Observed ($-\log_{10}$ p-value)
- Expected ($-\log_{10}$ p-value)
Violation of the Polygenic $\beta$ Assumption

$\gamma = 0.001$

$\gamma = 0.002$
Violation of the Homogeneous $\gamma$ Assumption

$\gamma = 0.001$

$\gamma = 0.002$
Power of Testing $\alpha$ under Alternative
Compared Methods: Testing $\gamma$

- LDA MR Egger: Fixed effects of $\beta$; Egger assumption on $\gamma$; two-stage inference.

- MR-PRESSO: Permutation based approach; assumes independent instruments.

- PMR-Egger: Normal prior on $\beta$; Egger assumption on $\gamma$; maximum likelihood inference.
Testing Horizontal Pleiotropy $\gamma$ under the Null

PVE_{zy} = 0.004

PVE_{zy} = 0.006
Power of Testing $\gamma$ under the Alternative
Real Data Applications

• GEUVADIS Expression Data \( (n_1 = 465) \), with \( \sim 15,000 \) genes.

• WTCCC: Seven common diseases \( (n_2 = \sim 5,000) \).

• UK Biobank: Ten quantitative traits \( (n_2 = \sim 300,000) \).
WTCCC: Testing Causal Effects
WTCCC: Testing Horizontal Pleiotropy
UK Biobank: Testing Causal Effects
UK Biobank: Testing Horizontal Pleiotropy
Summary

• We have presented an MR framework that unifies many existing integrative transcriptome wide association analysis method.

• Our method PMR-Egger effectively controls for horizontal pleiotropy through a maximum likelihood/probabilistic inference framework.

• We have demonstrated the effectiveness of PMR-Egger through simulations and real data applications.

• PMR-Egger is implemented in the PMR R package, to be available on www.xzlab.org
Acknowledgements

• Zhongshang Yuan

• Collaborators: Ping Zeng (Xuzhou Medical University), Can Yang (HKUST) and Jin Liu (Duke-NUS Medical School)

• NIH R01HG009124 and NSF DMS1712933