Bayesian Genome-wide TWAS Method to Leverage both Cis- and Trans- eQTL

J. Yang[1], J. Luningham[1], J. Chen[1], S. Tang[1], P. De Jager[2], D.A. Bennett[3], A.S. Buchman[3]

1) Emory University School of Medicine. 2) Columbia University Irving Medical Center, 3) Rush University Medical Center

Method Overview

Motivation: a) Leverage both cis- and trans- eQTL information in TWAS; **b)** Efficient computation; **c)** Illustrate molecular mechanisms of Alzheimer's Dementia (AD) associations

Methods: Novel Bayesian Genome-wide TWAS (BGW-TWAS) method based on Bayesian Variable Selection Regression (BVSR) model **[1]**, with gene expression $\mathcal{E}_g$, genotype matrix $\boldsymbol{X}$,

$$\mathcal{E}_g = \boldsymbol{X}_{cis}\boldsymbol{w}_{cis} + \boldsymbol{X}_{trans}\boldsymbol{w}_{trans} + \boldsymbol{\epsilon}, \qquad \epsilon_i \sim N(0, \sigma_\epsilon^2)$$

$$w_{q,i} \sim \pi_q N\left(0, \sigma_q^2 \sigma_\epsilon^2\right) + \left(1 - \pi_q\right)\delta_0\left(w_{q,i}\right), \qquad q \in (cis, trans)$$

Summary statistics of single variant eQTL analysis and pre-calculated LD matrix are used in adapted EM-MCMC algorithm for computation efficiency.

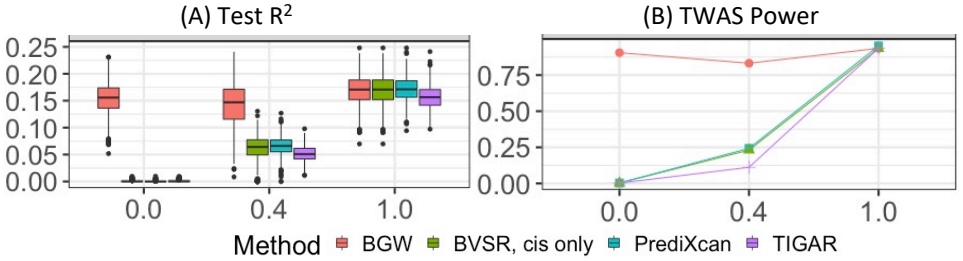
Results: Significant genes driving by trans-eQTL were identified by BGW-TWAS. Tool Available from https://github.com/yanglab-emory/BGW-TWAS .

Note: Postdoc position is available now in YangLab (https://yanglab-emory.github.io/) for studying multi-omics data of complex diseases.

Simulation Results

Considered 1,269 cis and 21,372 trans SNPs, with 499 training and 1,209 test sample size, gene expression heritability 0.2, and various proportion of contribution from cis-eQTL (0, 0.4, 1.0) as shown in X-axis in **Figure 1**.

Figure 1. Compare prediction accuracy of gene expression by Test R^2 (A) and TWAS power (B) to alternative TWAS methods [2, 3] using only cis-eQTL.



Application Studies of AD

- Estimated cis- and trans- eQTL effect sizes of 14K genes for postmortem brain tissues.
- Meta-analysis with BGW-TWAS p-value per study for ROS/MAP and MCADGS individual-level GWAS data of AD.
- BGW-TWAS using IGAP summary-level GWAS data of AD.

Application Results

- Different posterior distribution was identified for the probabilities of being a cis- or trans- eQTL (**Figure 2**).
- Gene ZC3H12B on Chromosome X (**Table 1**) driven completely by trans-eQTL near APOC1 (**Table 2**) was identified by BGW-TWAS.
- Thirteen significant genes including ZC3H12B were identified by BGW-TWAS with IGAP summary-level GWAS data (**Table 3**).

Figure 2. Distribution of Log_{10} estimates of Prob($\pi_q \neq 0$), $q \in (cis, trans)$.

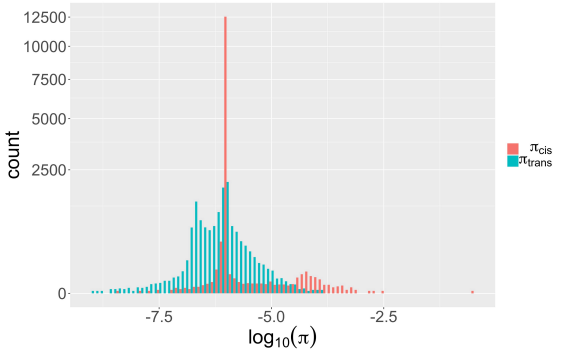


Table 1. Significant genes identified by BGW-TWAS using individual-level GWAS data of AD from ROS/MAP and MCADGS.

Gene	CHR	Position	Train R^2	p-value	Effect size (SD)	Phenotype
ZC3H12B	X	64,708,614	0.24	5.42×10^{-13}	0.265 (0.037)	AD
ZC3H12B	X	64,708,614	0.24	9.59×10^{-7}	0.142 (0.029)	Global AD pathology
ZC3H12B	X	64,708,614	0.24	1.89×10^{-6}	0.138 (0.029)	Tangles
KCTD12	13	77,454,311	0.09	3.44×10^{-8}	0.143 (0.026)	β-Amyloid

Table 2. Trans-SNPs with top five $\pi_{trans} > 0.003$ for gene *ZC3H12B*.

CHR	POS	rsID	Function	MAF	PP	w	p-value
1	159,135,282	*rs3026946*	Intergenic	0.213	0.0147	-0.071	6.25×10^{-7}
19	45,422,160	*rs12721051*	3' UTR (APOC1)	0.161	0.0031	0.071	3.94×10^{-6}
19	45,422,846	*rs56131196*	Downstream (APOC1)	0.173	0.0048	0.069	1.75×10^{-6}
19	45,422,946	*rs4420638*	Downstream (APOC1)	0.173	0.0051	0.068	1.77×10^{-6}
19	45,424,514	*rs157592*	Regulatory Region (APOC1)	0.181	0.0056	0.075	1.43×10^{-6}

Table 3. Significant genes identified by BGW-TWAS using IGAP GWAS summary statistics of AD.

Gene	CHR	Position	TWAS P-VALUE			
			BGW-TWAS	BVSR cis-eQTL	PrediXcan	TIGAR
GPX1[a]	3	49,394,608	2.45×10^{-98}	2.45×10^{-98}	-	3.15×10^{-1}
FAM86DP	3	75,484,261	1.55×10^{-13}	4.81×10^{-1}	5.38×10^{-1}	9.63×10^{-1}
BTN3A2[a]	6	26,378,546	1.59×10^{-26}	1.56×10^{-26}	3.17×10^{-1}	5.04×10^{-1}
ZNF192[a]	6	28,124,089	1.26×10^{-32}	1.25×10^{-32}	8.56×10^{-2}	2.07×10^{-1}
AL022393.7[a]	6	28,144,452	3.25×10^{-178}	2.24×10^{-178}	1.50×10^{-1}	8.36×10^{-2}
HLA-DRB1[ab]	**6**	**32,557,625**	$\mathbf{1.02 \times 10^{-12}}$	$\mathbf{8.99 \times 10^{-13}}$	$\mathbf{2.06 \times 10^{-6}}$	-
AEBP1	7	44,154,161	5.55×10^{-220}	8.62×10^{-1}	6.69×10^{-1}	4.19×10^{-1}
BUB3	10	124,924,886	6.64×10^{-18}	1.05×10^{-2}	-	4.76×10^{-1}
FBXO3	11	33,796,089	1.48×10^{-9}	6.88×10^{-1}	-	1.13×10^{-1}
CEACAM19[abc]	**19**	**45,187,631**	$\mathbf{4.7 \times 10^{-13}}$	$\mathbf{2.54 \times 10^{-13}}$	$\mathbf{3.60 \times 10^{-12}}$	$\mathbf{2.83 \times 10^{-16}}$
APOC1[a]	19	45,422,606	8.9×10^{-11}	1.11×10^{-10}	3.18×10^{-6}	7.2×10^{-3}
ZC3H12B	**X**	**64,727,767**	$\mathbf{2.08 \times 10^{-37}}$	-	-	-
CXorf56	X	118,699,397	6.02×10^{-07}	-	-	-

a. Genes that were also identified as significant by using BVSR cis-eQTL estimates.
b. Genes that were also identified by PrediXcan.
c. Genes that were also identified by TIGAR.

References

[1] Guan, Y.T. et al. (2011). Bayesian Variable Selection Regression for Genome-Wide Association Studies and Other Large-Scale Problems. Annals of Applied Statistics 5, 1780-1815.
[2] Gamazon, E.R. et al. (2015). A gene-based association method for mapping traits using reference transcriptome data. Nature genetics 47, 1091-1098.
[3] Nagpal, S. et al. (2019). TIGAR: An Improved Bayesian Tool for Transcriptomic Data Imputation Enhances Gene Mapping of Complex Traits. AJHG 105, 258-266.