Novel Variance-Component TWAS Method for Studying Alzheimer's Disease Dementia

Jingjing Yang, PhD
Assistant Professor

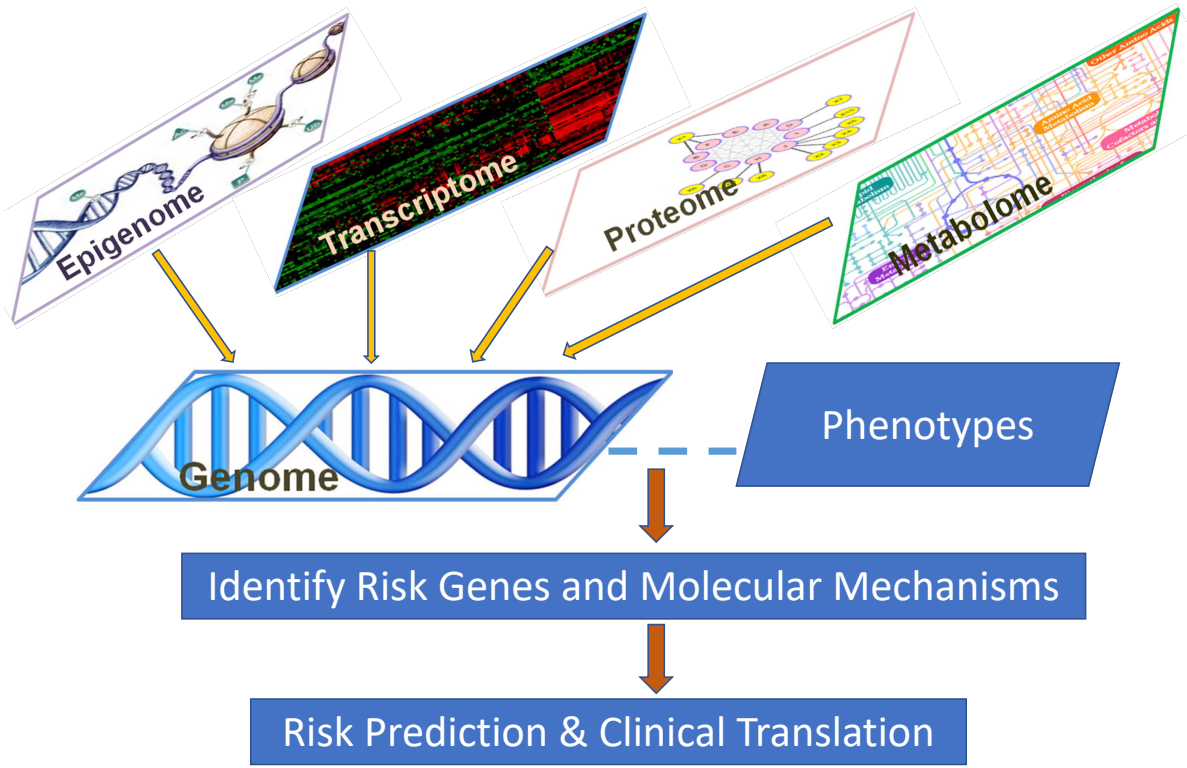
CCQG
CENTER FOR COMPUTATIONAL AND QUANTITATIVE GENETICS

EMORY UNIVERSITY SCHOOL OF MEDICINE

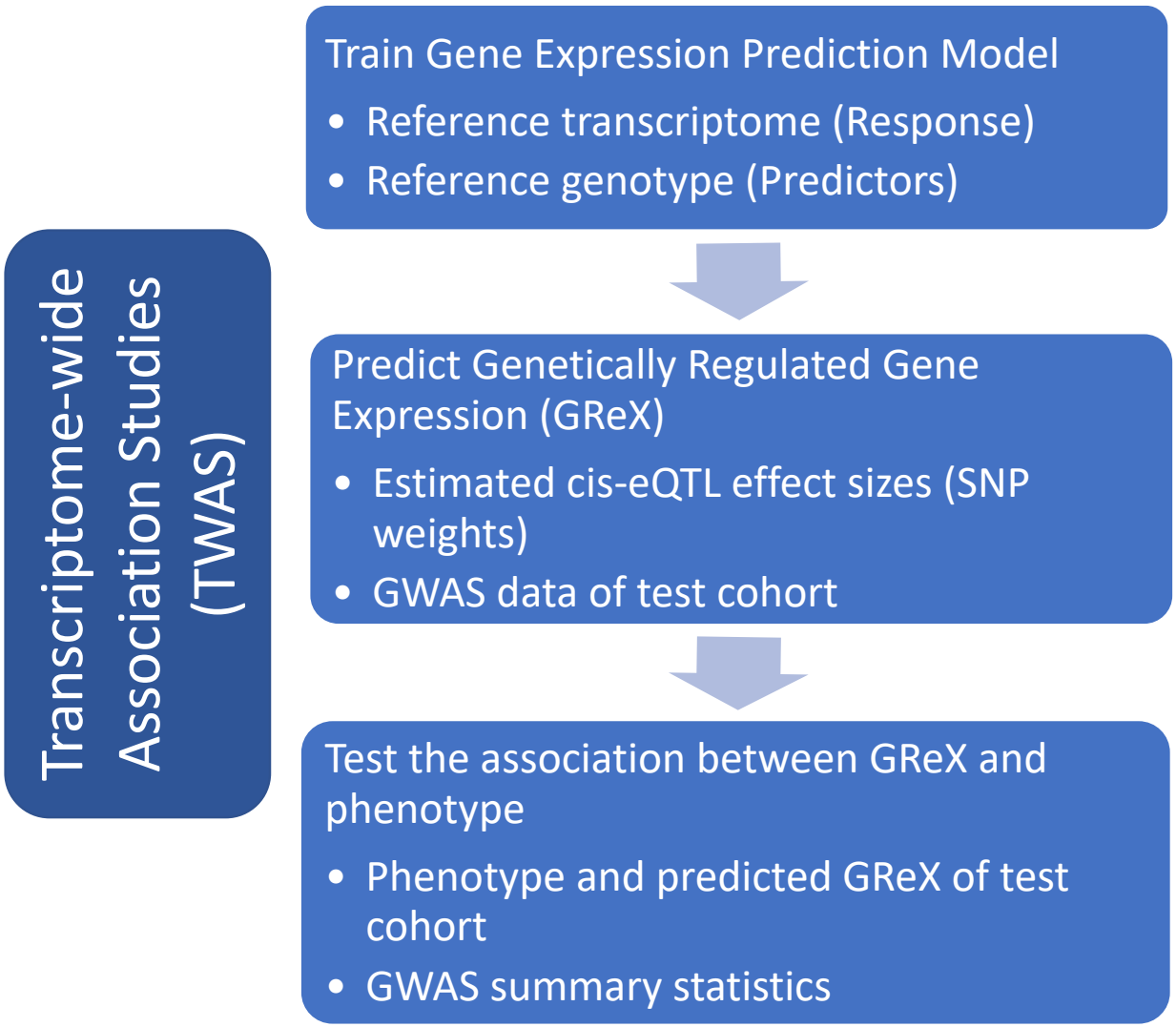
Outline

Genomic Etiology of Alzheimer's Disease Dementia

- Polygenic with low penetrance by individual genes
- Affected by multi-layers of Omics data
- Largely unknown genomic etiology



Integrate Transcriptomic Data in GWAS by TWAS



Standard TWAS Framework

- Quantitative gene expression trait $\boldsymbol{T}_g$
- Genotype data of all cis-SNPs $\boldsymbol{X}$
- SNP effect sizes on transcriptome (eQTL effect sizes) w_i

$$\mathbf{T}_g = \mathbf{X}\mathbf{w} + \boldsymbol{\varepsilon}$$

$$\widehat{GReX}_g = \sum_{i=1}^{m} \widehat{w}_i x_i^*$$

$$E[g(\mathbf{Y}_{pheno}|\mathbf{X}^*, \widehat{\mathbf{w}})] = \gamma\widehat{GReX}_g = \gamma\left(\sum_{i=1}^{m} \widehat{w}_i x_i^*\right) = \sum_{i=1}^{m} (\gamma\widehat{w}_i)\, x_i^*$$

- Existing TWAS tools all assume SNP effect sizes on phenotype $\beta_i = \gamma\widehat{w}_i,\ i = 1, \cdots, m.$
- Burden test: $H_0 : \gamma = 0$

Variance-Component TWAS (VC-TWAS)

- Tests if the phenotype variance component due to $GReX_g$ is non-zero:

$$E[g(\mathbf{Y}_{pheno}|\mathbf{X}^*, \widehat{\mathbf{w}})] = \sum_{i=1}^{m} (\gamma \widehat{w}_i)\, x_i^* = \mathbf{X}^* \beta,$$

$$\beta_i \sim N(0, \widehat{w}_i{}^2 \tau)$$

- Estimated eQTL effect sizes $\widehat{w}_i$ from the reference panel will be taken as SNP weights
- Variance Component test: $H_0 : \tau = 0$

VC-TWAS Test Statistic

- $\widehat{\mu}$ denotes the phenotype mean under the null model

$$Q = (\mathbf{Y} - \widehat{\mu})\mathbf{K}(\mathbf{Y} - \widehat{\mu}),\ \mathbf{K} = \mathbf{X}\mathbf{W}\mathbf{X}'$$

$$\mathbf{W} = diag(\widehat{w_1}^2, \cdots, \widehat{w_m}^2)$$

- Test statistic Q follows a mixture of chi-square distributions under the null hypothesis
- P-value can be easily calculated by Davies exact method as used by SKAT
- P-value calculation is also derived for using GWAS summary statistics

Simulation Study Design

- Use the real genotype data of gene *ABCA7* with 2799 cis-SNPs with MAF > 5% and HWP $> 10^{-5}$
- Training sample size (499), test sample size $(400, 800, 1200)$
- Consider $p_{causal} = 0.2$ for the proportion of causal SNPs for gene expression
- Consider target gene expression heritability and phenotype heritability: (h_e^2, h_p^2)
- Simulate quantitative gene expression traits:

$$\mathbf{E_g} = \mathbf{Xw} + \boldsymbol{\varepsilon}_\mathbf{E},\ \boldsymbol{\varepsilon}_\mathbf{E} \sim N(0, (1 - h_e^2)\mathbf{I})$$

Consider Two Phenotype Models

1.

$$\mathbf{Y} = \gamma \mathbf{E_g} + \boldsymbol{\varepsilon}_\mathbf{Y} = \gamma(\mathbf{X}\mathbf{w} + \boldsymbol{\varepsilon}_\mathbf{E}) + \boldsymbol{\varepsilon}_\mathbf{Y};\ \boldsymbol{\varepsilon}_\mathbf{Y} \sim N(0, (1 - h_p^2)\mathbf{I})$$

- γ is chosen to ensure phenotype heritability h_p^2 due to gene expression

2.

$$\mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\varepsilon}_\mathbf{Y};\ \beta_i \sim N(0, \gamma w_i^2),\ \boldsymbol{\varepsilon}_\mathbf{Y} \sim N(0, (1 - h_p^2)\mathbf{I})$$

- w_i denotes the true eQTL effect size used to simulate gene expression
- γ is chosen to ensure phenotype heritability h_p^2 due to genotypes

Compare TWAS Power

eQTL effect sizes w_i were estimated from training cohort

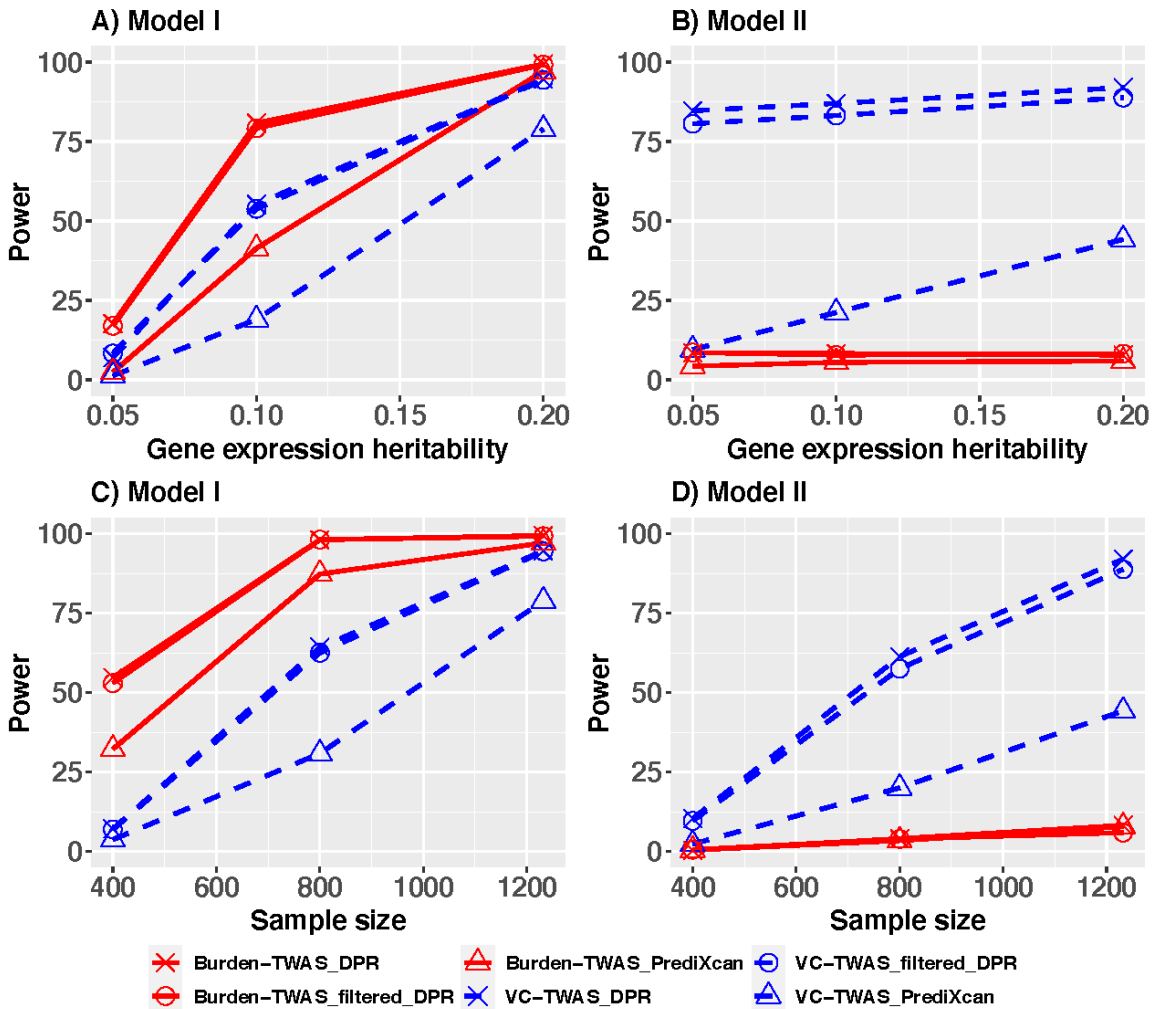
- Estimated by TIGAR using the nonparametric Bayesian Dirichlet Process Regression (DPR) model
- Or by PrediXcan using Elastic-net penalized regression model

Compare TWAS power

- Standard Burden TWAS vs. VC-TWAS
- VC-TWAS using individual-level vs. summary-level GWAS data

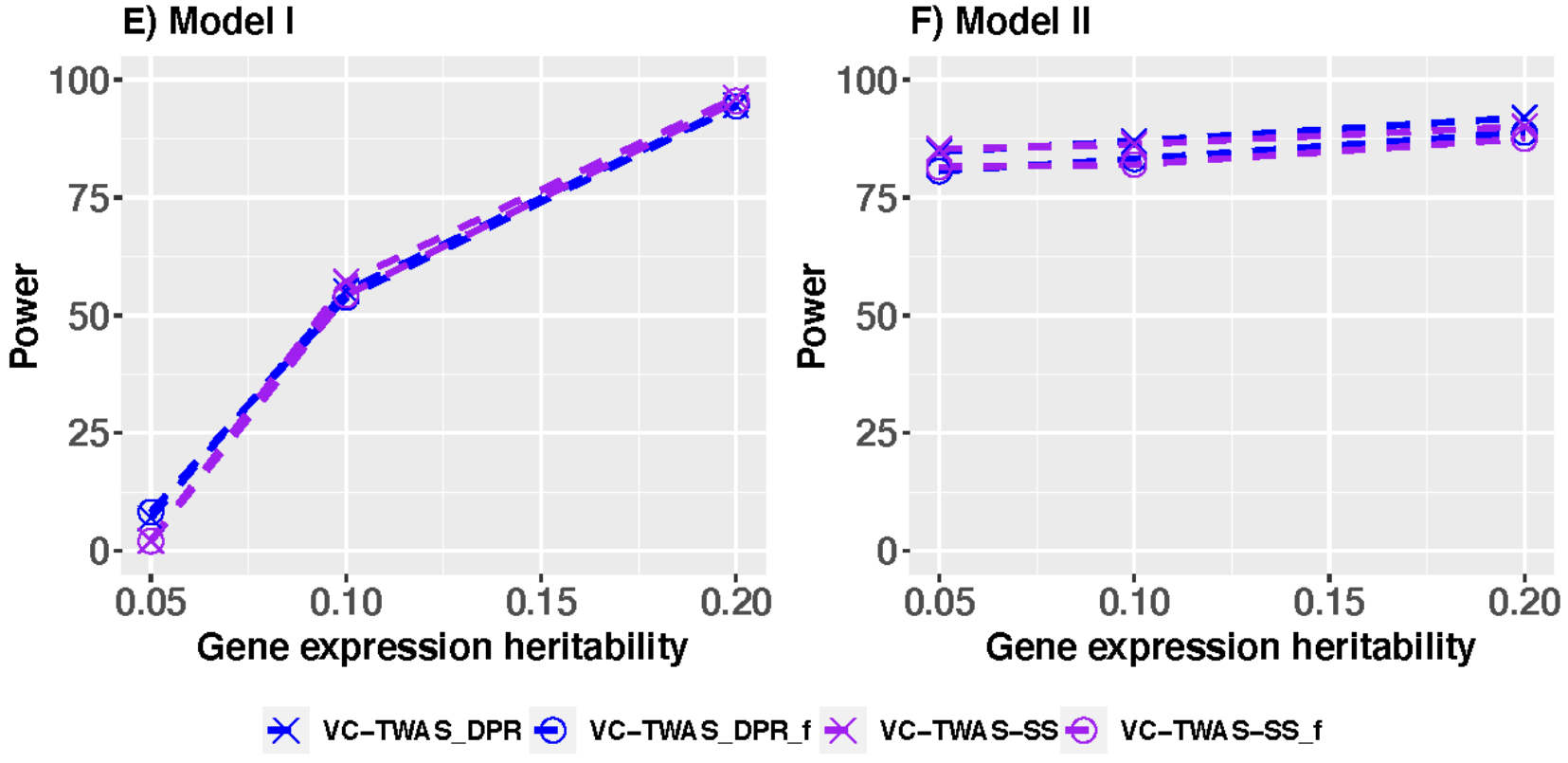
Compare TWAS power between standard Burden TWAS and VC-TWAS

$$p_{causal} = 0.2$$



Compare VC-TWAS power between using individual-level and summary-level GWAS data

$p_{causal} = 0.2$ and w_i estimated by DPR



Application Studies of Alzheimer's Disease Dementia (AD)

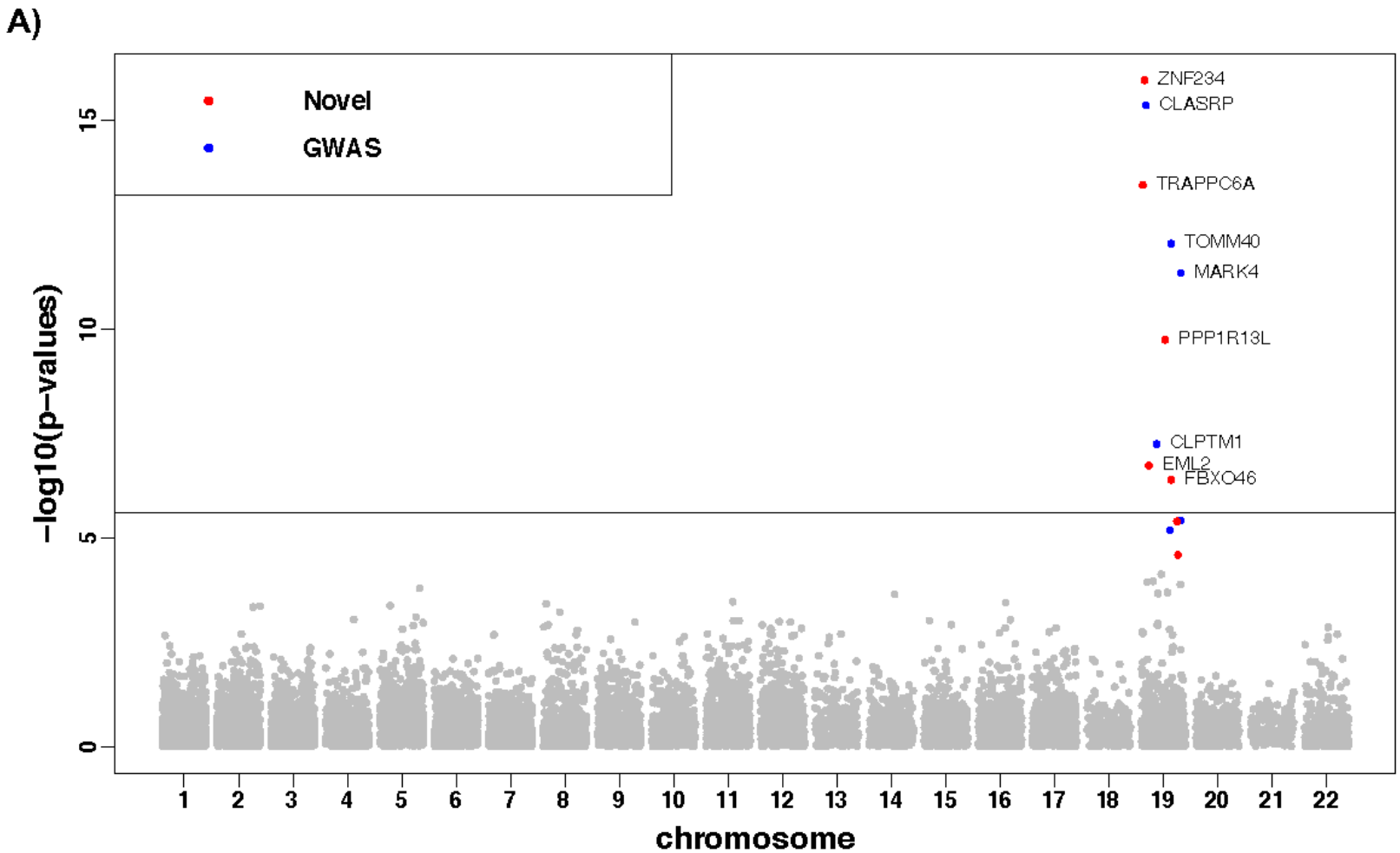
ROS/MAP

- Training data: 499 samples with both genotype and transcriptomic data (14,156 genes)
- Individual GWAS data of 2,093 samples
- Study AD clinical diagnosis, Global AD pathology
- Adjusted for covariates age at death, sex, smoking, ROS or MAP study, education level, top 3 PCs

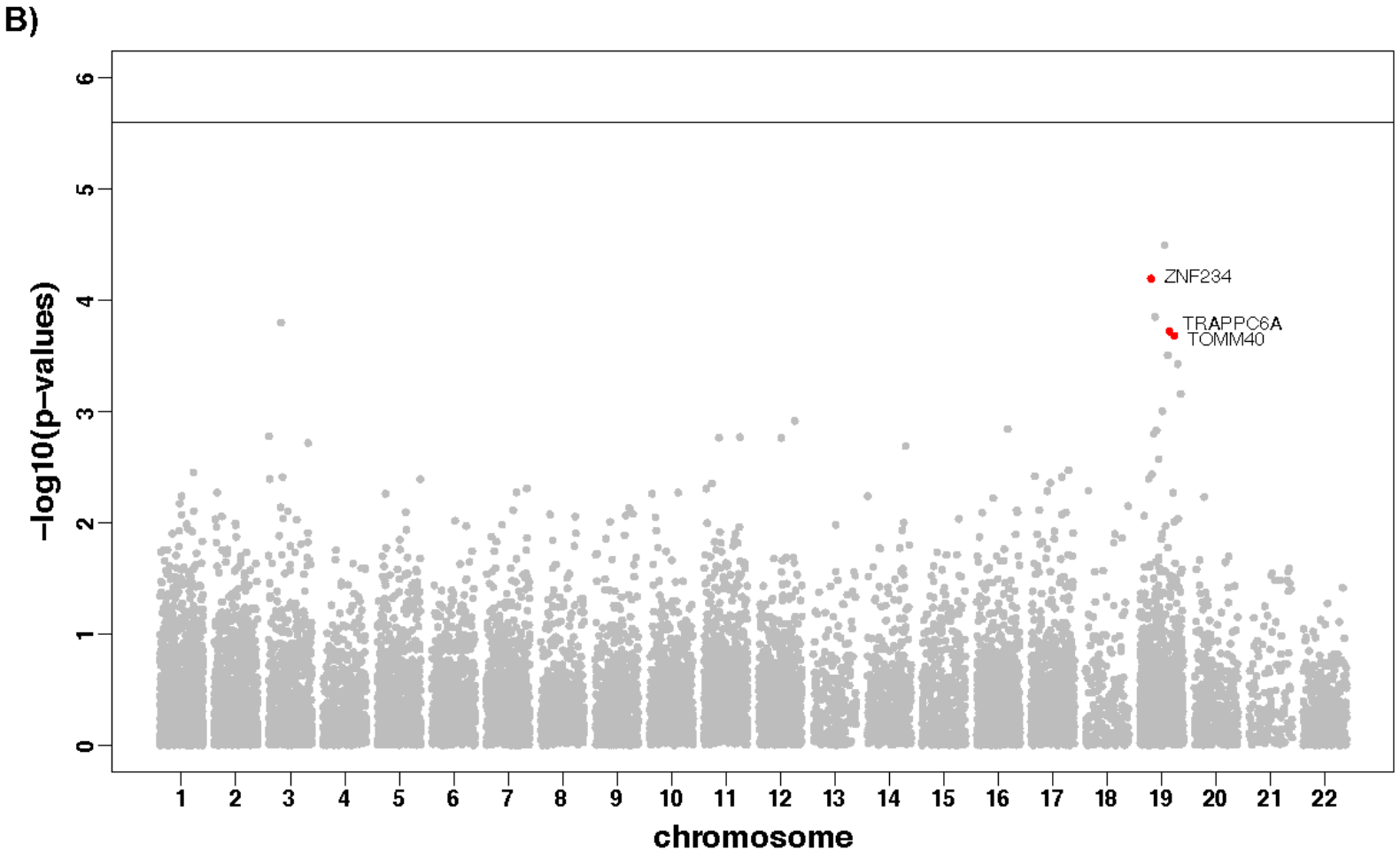
Mayo Clinic LOAD GWAS Data

- Individual GWAS data of 2,099 samples
- Study AD clinical diagnosis
- Adjusted for covariates age, sex, top 3 PCs

VC-TWAS of AD Clinical Diagnosis



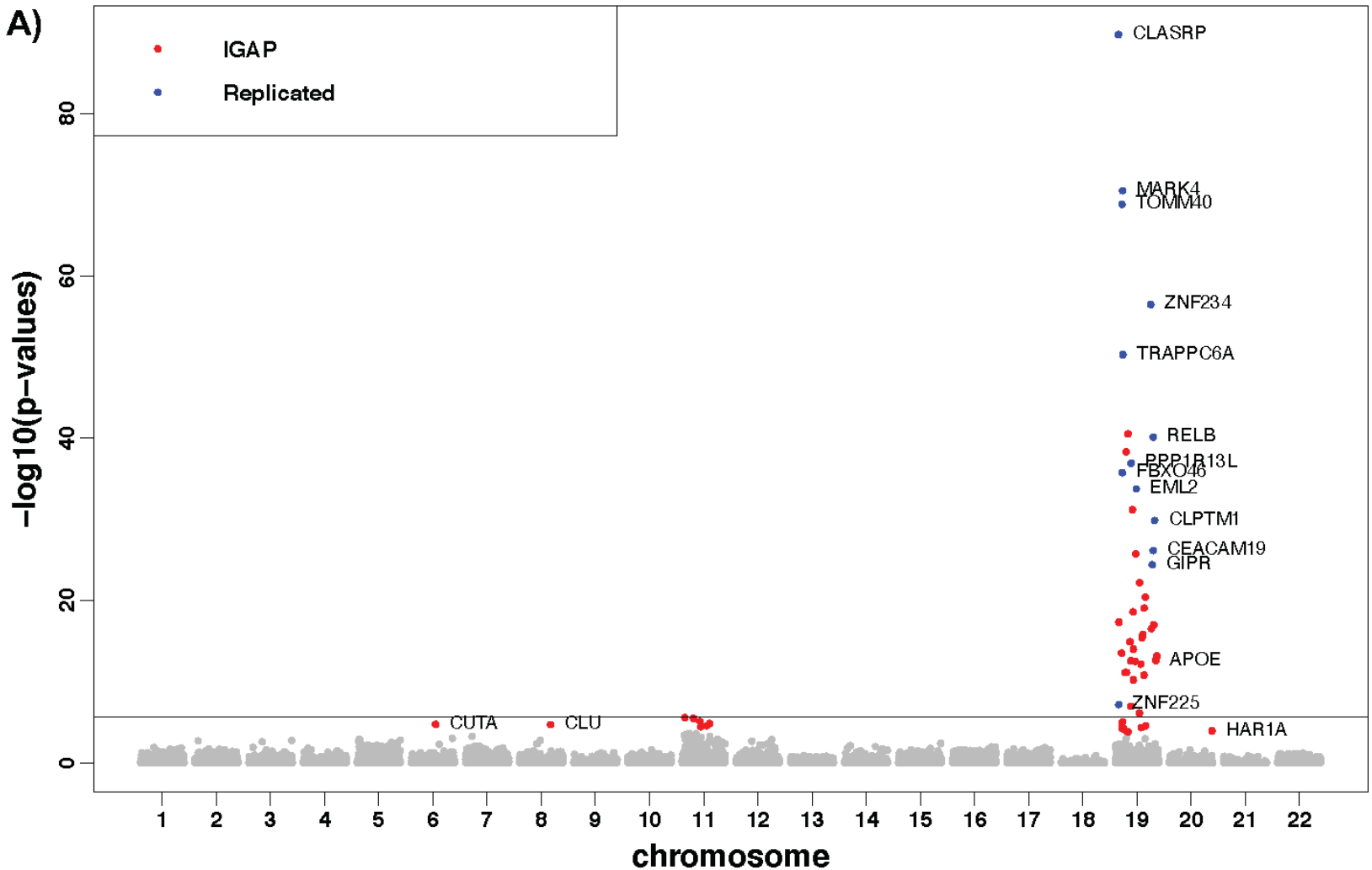
VC-TWAS of Global AD Pathology



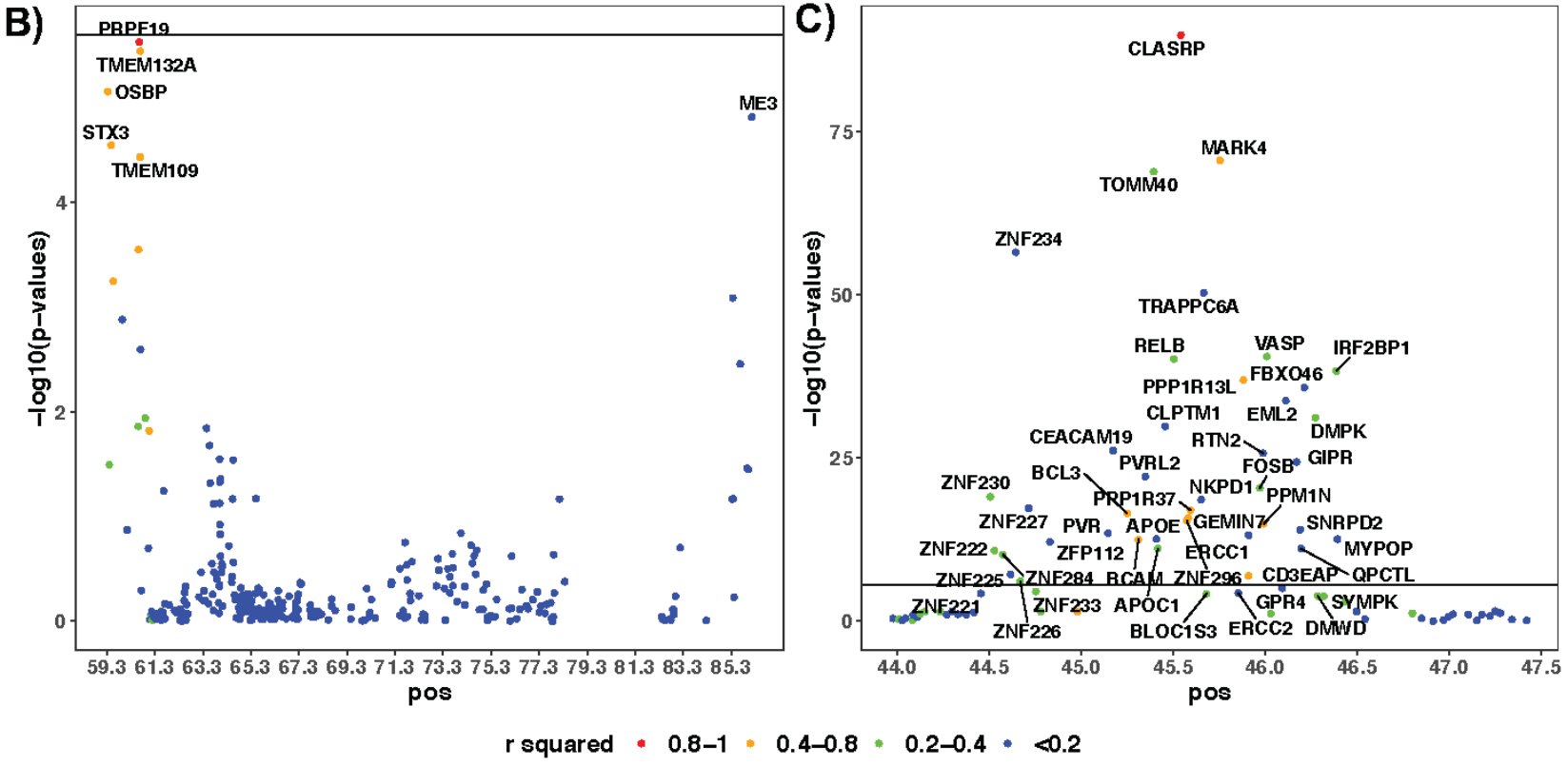
IGAP summary-level GWAS data

- GWAS summary statistics for studying AD by International Genomics of Alzheimer's Project (IGAP)
- Generated by meta-analysis of four consortia ($\sim$ 17K cases and $\sim$ 37K controls; European)
  - Alzheimer's Disease Genetic Consortium (ADGC)
  - Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium
  - European Alzheimer's Disease Initiative (EADI)
  - Genetic and Environmental Risk in Alzheimer's Disease (GERAD) Consortium

VC-TWAS results with IGAP summary-level GWAS data



Locus Zoom plot for loci on Chr 11 and Chr 19



Using both cis- and trans- eQTL effect sizes estimated by Bayesian Genome-wide TWAS method

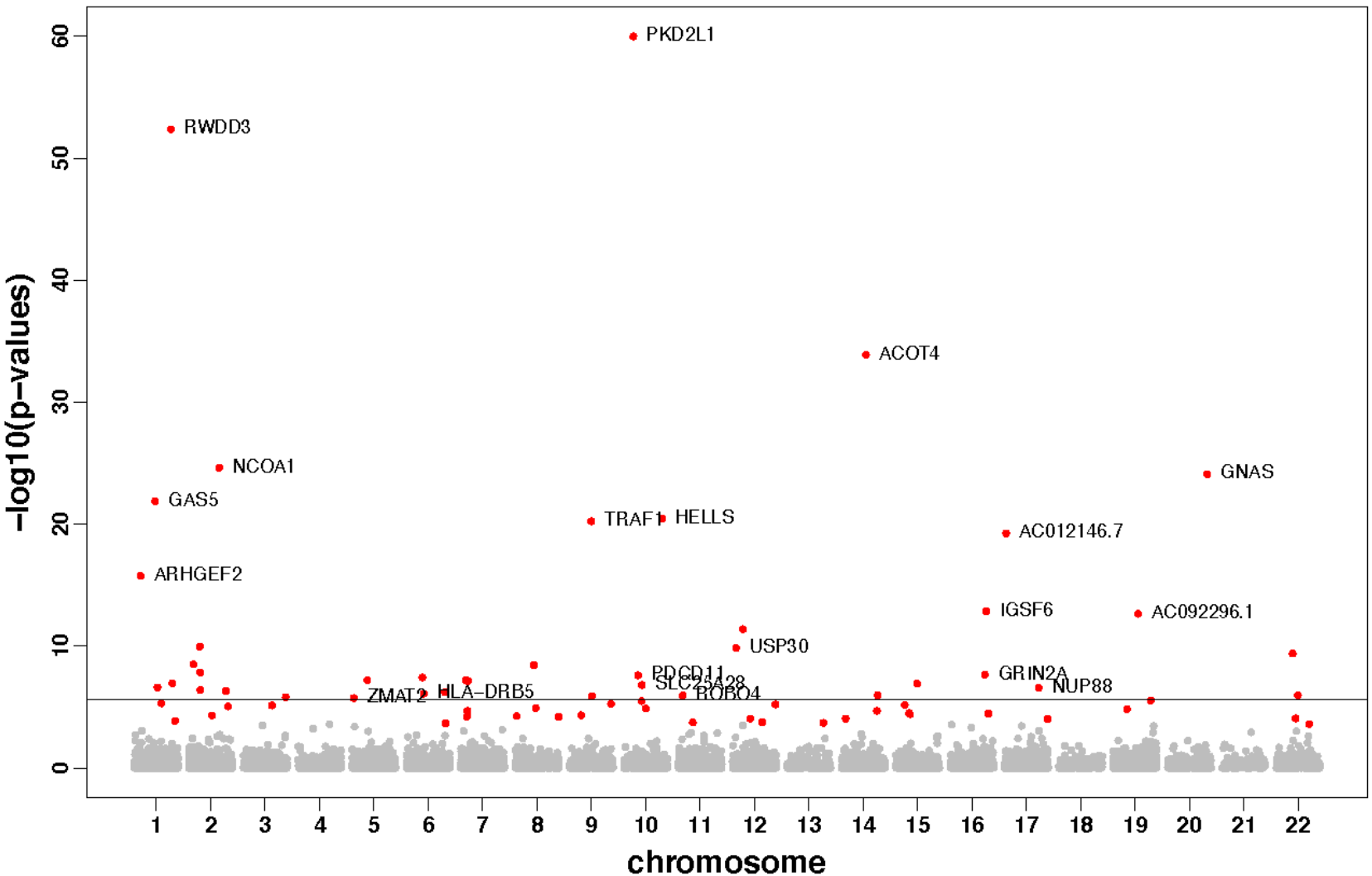


Table 3. Significant genes identified by VC-TWAS using IGAP summary statistics data with BGW cis- and trans- eQTL weights, which are either known GWAS loci or shown to be related with AD or other neurological diseases by previous studies. AD risk genes identified by previous GWAS are shaded in grey.

Gene Name	CHROM	Start	End	P-value	FDR
ARHGEF2[a]	1	155,916,644	155,966,129	1.75×10^{-16}	2.46×10^{-13}
GAS5 [a]	1	173,833,037	173,838,020	1.33×10^{-22}	3.11×10^{-19}
NCOA1 [a]	2	24,714,782	24,993,571	2.42×10^{-25}	8.51×10^{-22}
CENPO [a]	2	25,016,004	25,045,245	4.07×10^{-7}	1.85×10^{-4}
SLC4A10 [a]	2	162,280,842	162,841,792	8.80×10^{-6}	2.58×10^{-3}
MAN2A1	5	109,025,066	109,205,326	6.30×10^{-8}	3.70×10^{-5}
ZMAT2 [a]	5	140,079,746	140,086,266	1.83×10^{-6}	6.43×10^{-4}
HLA-DRB5 [a]	6	32,485,119	32,498,064	7.81×10^{-7}	3.23×10^{-4}
AHR [a]	7	17,338,245	17,385,776	2.11×10^{-5}	5.62×10^{-3}
JAZF1	7	27,870,191	28,220,362	7.49×10^{-8}	4.22×10^{-5}
ACHE [a]	7	100,487,614	100,494,594	5.93×10^{-5}	1.39×10^{-2}
DLGAP2 [a]	8	1,449,530	1,656,642	1.18×10^{-5}	3.38×10^{-3}
RFX3 [a]	9	3,218,296	3,526,004	4.69×10^{-5}	1.16×10^{-2}
SLC25A28 [a]	10	101,370,281	101,380,535	1.55×10^{-7}	7.79×10^{-5}
PDCD11 [a]	10	105,156,404	105,206,049	2.52×10^{-8}	1.69×10^{-5}
ROBO4 [a]	11	124,753,586	124,768,396	1.13×10^{-6}	4.30×10^{-4}
USP30 [a]	12	109,460,893	109,525,831	1.50×10^{-10}	1.41×10^{-7}
PPP1R3E [a]	14	23,765,111	23,772,057	2.11×10^{-5}	5.62×10^{-3}
MTFMT [a]	15	65,294,844	65,321,977	3.10×10^{-5}	8.09×10^{-3}
PARP6	15	72,533,521	72,565,340	6.77×10^{-6}	2.07×10^{-3}
GRIN2A [a]	16	9,852,375	10,276,611	2.22×10^{-8}	1.56×10^{-5}
SETD6 [a]	16	58,549,382	58,554,431	3.44×10^{-5}	8.81×10^{-3}
NUP88 [a]	17	5,264,257	5,323,480	2.65×10^{-7}	1.24×10^{-4}
NOS2 [a]	17	26,083,791	26,127,555	9.55×10^{-5}	2.07×10^{-2}
CEACAM19	19	45,174,723	45,187,631	1.46×10^{-5}	4.03×10^{-3}
APOC1 [a]	19	45,417,920	45,422,606	2.80×10^{-6}	9.63×10^{-4}

a: Genes shown to be related with AD or other neurological diseases by previous studies

https://doi.org/10.1371/journal.pgen.1009482.t003

Summary

- VC-TWAS relaxes the assumption of a linear relationship between SNP effect sizes on phenotype and one transcriptome
- Higher TWAS power is achieved in practice
- VC-TWAS can be applied to using both individual-level and summary-level GWAS data
- Different methods, e.g., PrediXcan/Elastic-Net, TIGAR/DPR, BGW-TWAS/BVSR, can be used to estimate the eQTL effect sizes that will be taken as SNP weights by VC-TWAS
- Implemented in our TIGAR tool
  `https://github.com/yanglab-emory/TIGAR`

Paper on PLOS Genetics

RESEARCH ARTICLE

Novel Variance-Component TWAS method for studying complex human diseases with applications to Alzheimer's dementia

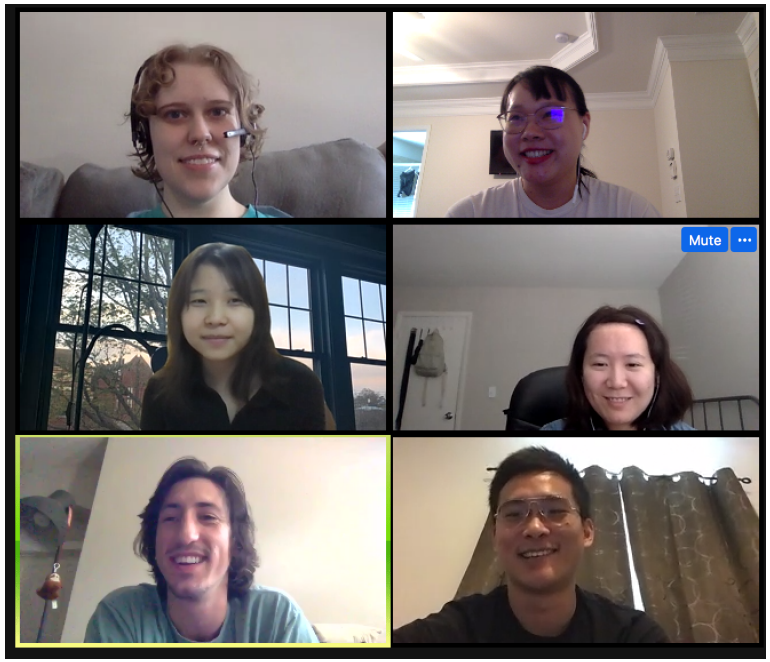
Shizhen Tang[1,2], Aron S. Buchman[3], Philip L. De Jager[4], David A. Bennett[3], Michael P. Epstein[1], Jingjing Yang[1]*

1 Center for Computational and Quantitative Genetics, Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia, United States of America, **2** Department of Biostatistics and Bioinformatics, Emory University School of Public Health, Atlanta, Georgia, United States of America, **3** Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, Illinois, United States of America, **4** Center for Translational and Computational Neuroimmunology, Department of Neurology and Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University Irving Medical Center, New York, New York, United States of America

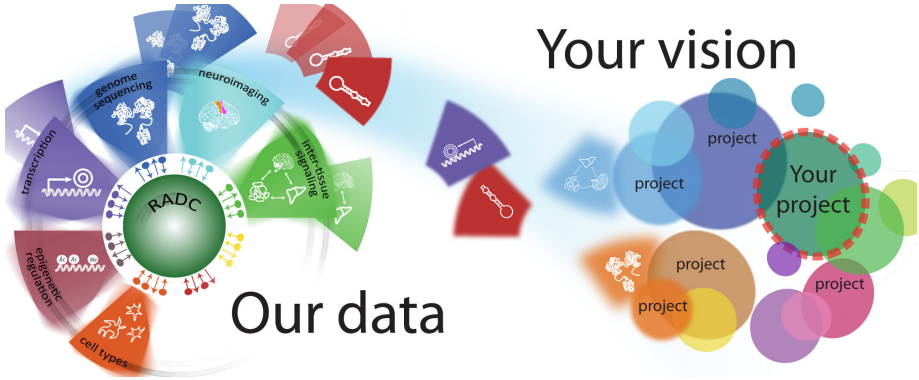
* jingjing.yang@emory.edu

Acknowledgement

Yang Lab @ Emory



Rush Alzheimer's Disease Center



Supported by R35GM138313

NIH National Institute of General Medical Sciences

Mayo Clinic LOAD GWAS

AMP-AD Knowledge Portal