Integrating 3D Genomic and Epigenomic Data to Enhance Target Gene Discovery and Drug Repurposing in Transcriptome-Wide Association Study

Chachrit Khunsriraksakul[1,2], Daniel McGuire[3], Renan Sauteraud[3], Fang Chen[3], Lina Yang[3], Rosa Kim[2], Jordan Hughey[1], Scott Eckert[1], James Weissenkampen[2], Laura Carrel[4], Bibo Jiang[2], Dajiang Liu[2,3]

[1]Department of Bioinformatics and Genomics, Pennsylvania State University College of Medicine, Hershey, PA; [2]Institute for Personalized Medicine, Pennsylvania State University College of Medicine, Hershey, PA; [3]Department of Public Health Sciences, Pennsylvania State University College of Medicine, Hershey, PA; [4]Department of Biochemistry and Molecular Biology, Pennsylvania State University College of Medicine, Hershey, PA

BACKGROUND

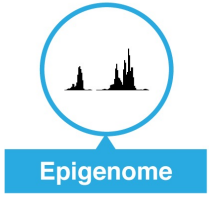
Transcriptome-wide association studies (TWAS) is an emerging gene-based association method.

- Gene expression prediction models are created from reference dataset with matched genotypes and expression data.
- Models are then applied to GWAS summary statistics to identify trait-associated genes.

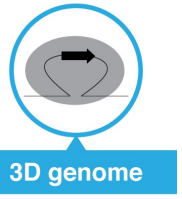
Gene expression prediction models are usually created without the integration of functional annotation.

- EpiXcan is the first TWAS method that integrates epigenomic annotations into the model building step; however, the framework is computationally expensive and not flexible.

OBJECTIVE



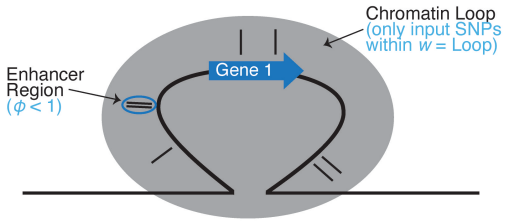
Use epigenomic and 3D genomic data to improve the accuracy of gene expression prediction models.



MODEL DEVELOPMENT and SIMULATIONS

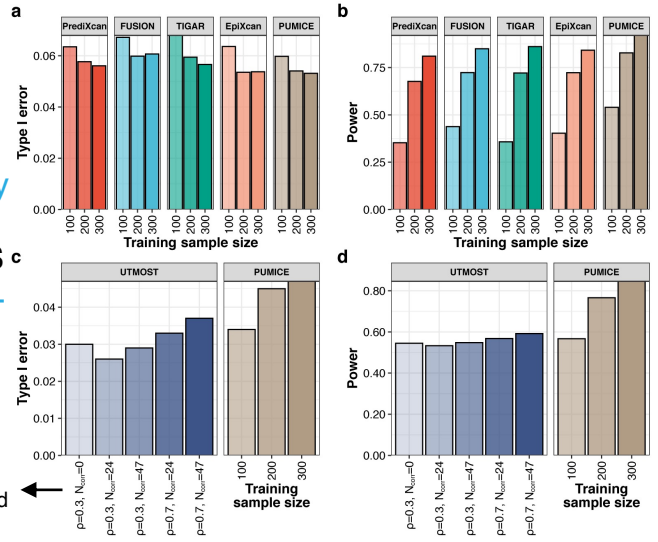
PUMICE utilizes multistep elastic net framework to tune for:

- Best penalty factor ($\emptyset$) according to the epigenomic data
- Best window size (w) according to the 3D-genomic data



We tuned model across $\emptyset \in [0,1]$ and $w \in \{\pm 250kb, \pm 1Mb, Loop, Domain, TAD, pcHiC\}$. Best model was selected based on the lowest mean cross-validated error.

Extensive simulation studies illustrated that PUMICE performed robustly in comparison to all previous TWAS methods with well-controlled type I error.



ρ = Genetic correlation between the causal tissue and correlated tissues
N_{corr} = Number of correlated tissues

EXTENSION and APPLICATION

PUMICE+ combines single-tissue and multi-tissue TWAS methods by Cauchy combination test.

Applying TWAS models to 79 complex traits, PUMICE+ identified

- Highest number of novel gene counts.
- Largest average chi-square value at MAGMA-prioritized genes.
- Putative target genes that are most consistent with target genes of approved drugs.

* PUMICE is the second-best method.

CONCLUSION

Integration of publicly available epigenomic and 3D genomic data can further improve the power of TWAS method and associated downstream analyses.

Future Directions:

- Integrate transcription factor annotations
- Enhance PUMICE framework to train model using only eQTL data

@ChachritK

cxk502@psu.edu