Deep learning models of gene regulation and the application in Cross-ancestral PRS in Alzheimer's Disease

Jui-Shan T. Lin[1], Anjali Das[1,3], Chirag M. Lakhani[1], Towfique Raj[2], David A. Knowles[1,3]

[1]New York Genome Center [2]Icahn School of Medicine at Mount Sinai [3]Columbia University

Abstract

Recent Genome-Wide Association Studies (GWAS) have identified several loci related to Alzheimer's Disease (AD). However, most GWAS predominantly focus on European populations, which limits their transferability across diverse ancestral groups due to differences in linkage disequilibrium (LD) structures. Hence, we developed functional annotations incorporating omics data and deep learning (DL) models. These annotations were applied in PolyFun[1], a functionally informed fine-mapping tool designed to enhance the identification of causal SNPs. Additionally, we examined the impact of various annotations on fine-mapping outcomes. Using calibrated effect sizes, we calculated polygenic risk scores (PRS) and benchmarked our results against standard Clumping and p-value Thresholding (C+PT)[3] and PRS-CS[4], highlighting the potential for improved cross-ancestry risk prediction.

Datasets

Summary Statistics

- Bellenguez et al. (2022)[2] with 387,715 samples, and 14,432,485 SNPs

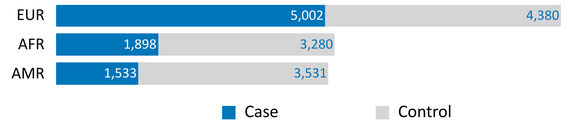
Target Data (From ADSP)

(1) Phenotype:

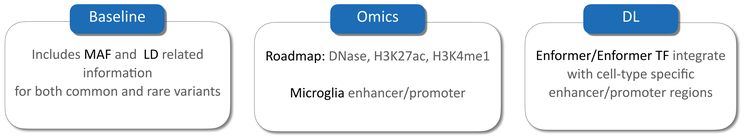
- Ethnicity was defined by 1K genome PCA projection
- Kept the none identical by descent (IBD) samples with AD diagnosis and age information

(2) Genotype:

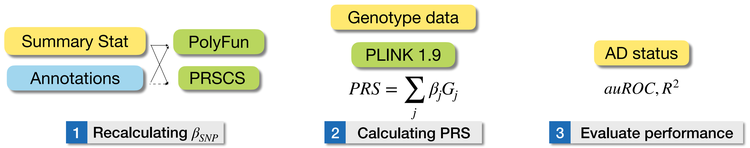
- Remove SNPs with > 10% missingness in all subject
- Remove individuals with > 10% genotype missingness



Annotations



Method



Create LD reference panel

PIP value thresholding

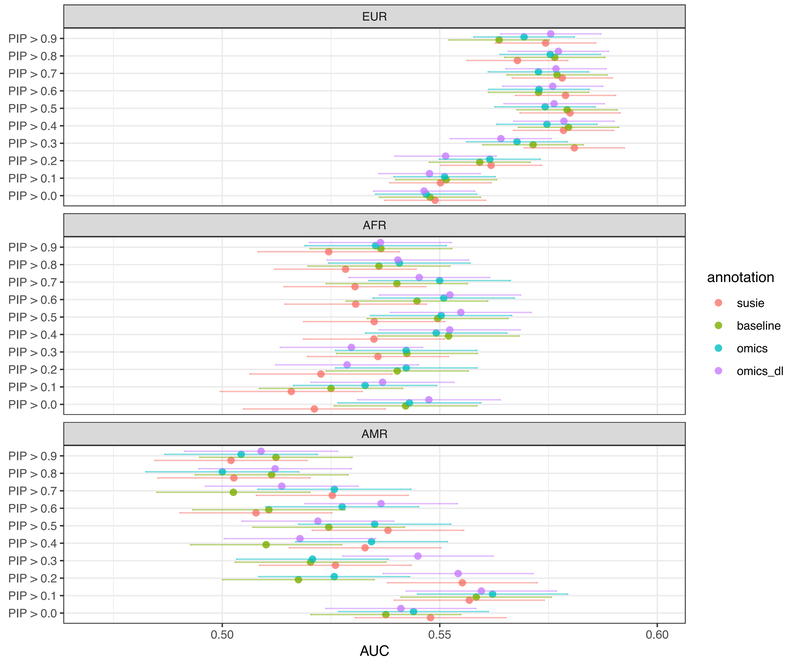
- Only retain the region with at least one SNPs that passes the threshold

auROC calculation

- We only use PRS for AD status prediction. Age, sex, and other information is not included in the model
- The confidence interval is calculated with DeLong's test

Results

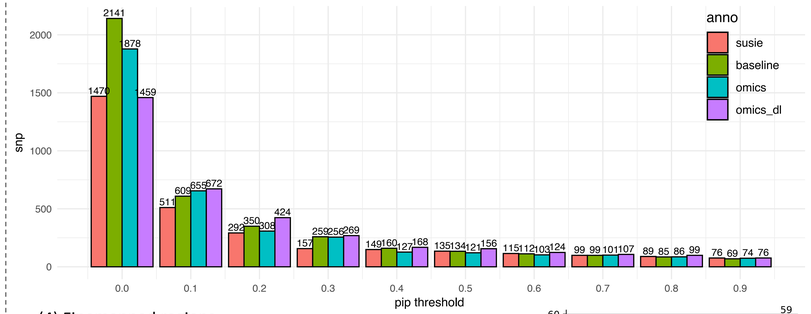
(1) PRS performance with different PIP threshold and annotation set



(2) PRS auROC in other methods

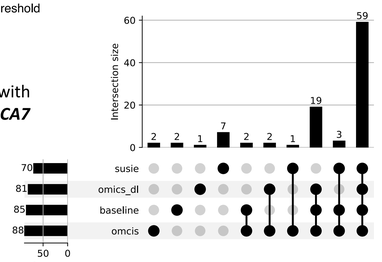
Clumping + P threshold			PRSCS		
EUR	AFR	AMR	EUR	AFR	AMR
0.58	0.56	0.59	0.58	0.51	0.54

(3) SNP counts when using different PIP threshold

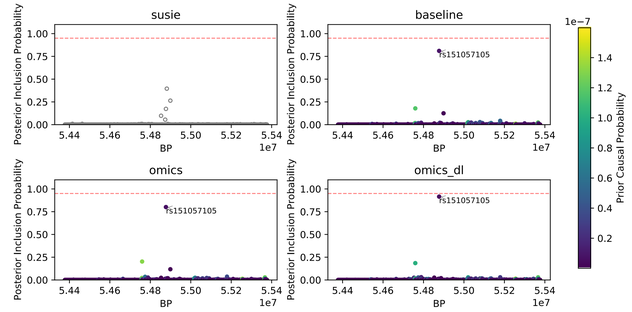


(4) Finemapped regions

Functional fine-mapping found SNPs associated with AD-linked genes like ***TREM2***, ***SORL1***, ***CLU***, and ***ABCA7***



(5) Functional annotations help find one potential causal SNP in CLU



Conclusion

- Cross-ancestry PRS performance shows significant gains for both AMR and AFR populations
- Careful selection of annotations is essential, as simply increasing the number does not necessarily improve results
- Applying PIP threshold is key for noise reduction, leading to more accurate outcomes

Future work

- Validate fine-mapping signal with microglia eQTL data
- Incorporate method with PRSCS + annotation for additional testing

Contact
Jui-Shan (Teresa) Lin
tlin@nygenome.org

Reference

1.Weissbrod *et al.* Functionally informed fine-mapping and polygenic localization of complex trait heritability. *Nat Genet*. (2020). https://doi.org/10.1038/s41588-020-00735-5
2.Bellenguez *et al.* New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet* (2022). https://doi.org/10.1038/s41588-022-01024-z
3.Purcell, Shaun, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *AJHG* (2007). http://pngu.mgh.harvard.edu/purcell/plink/
4.Ge *et al.* Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun* (2019). https://doi.org/10.1038/s41467-019-09718-5

Funding
U01 AG068880-02
Learning the Regulatory Code of Alzheimer's Disease Genomes