

# Application of Transcriptome-Wide Association Studies for Identifying Genes Associated with Inflammatory Bowel Disease

Esha Desai

Jacqueline Lee  
jacquelineklee@yahoo.com

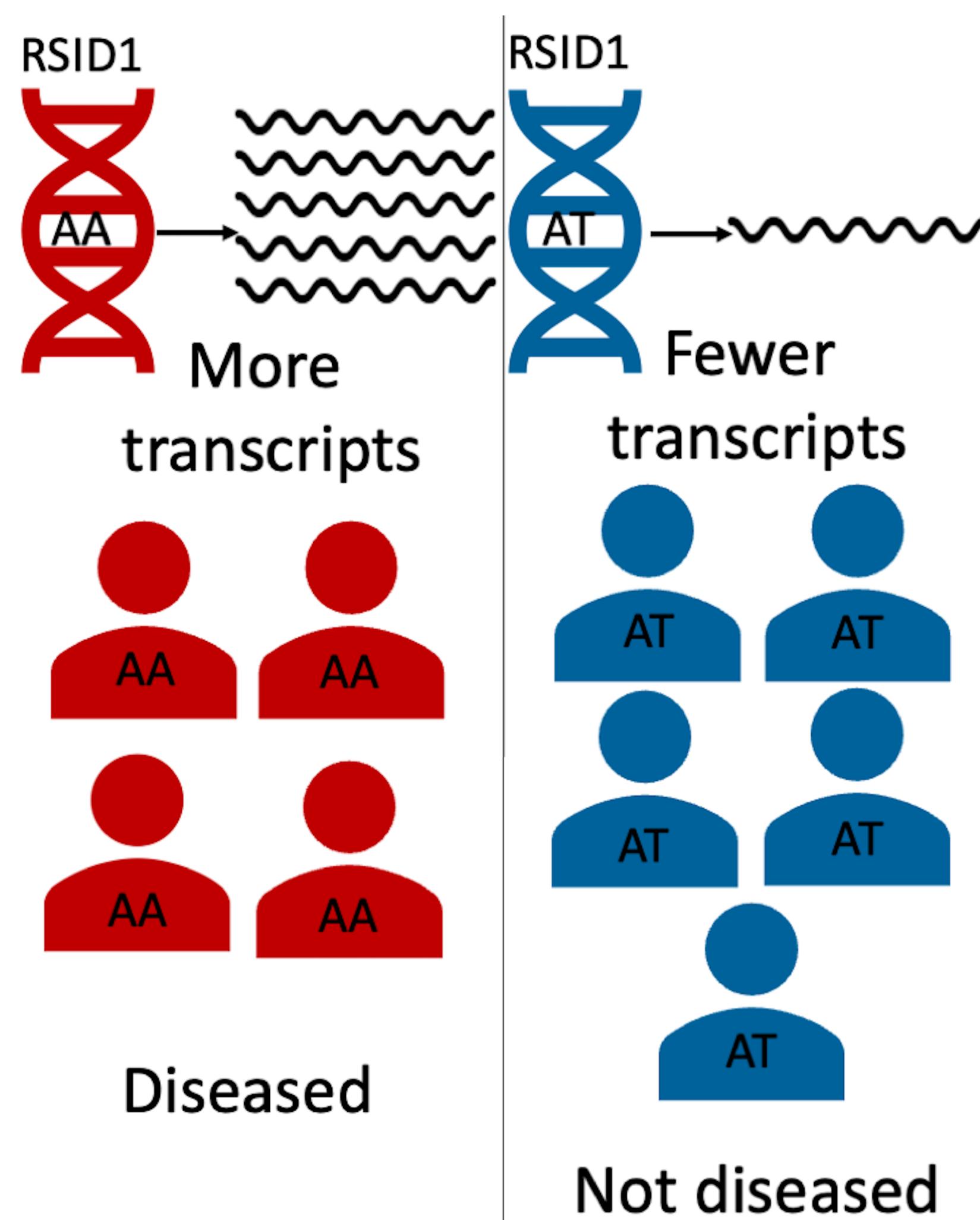
Moksha Poladi

Samuel Zhou

Tiffany  
Amariuta-Bartell

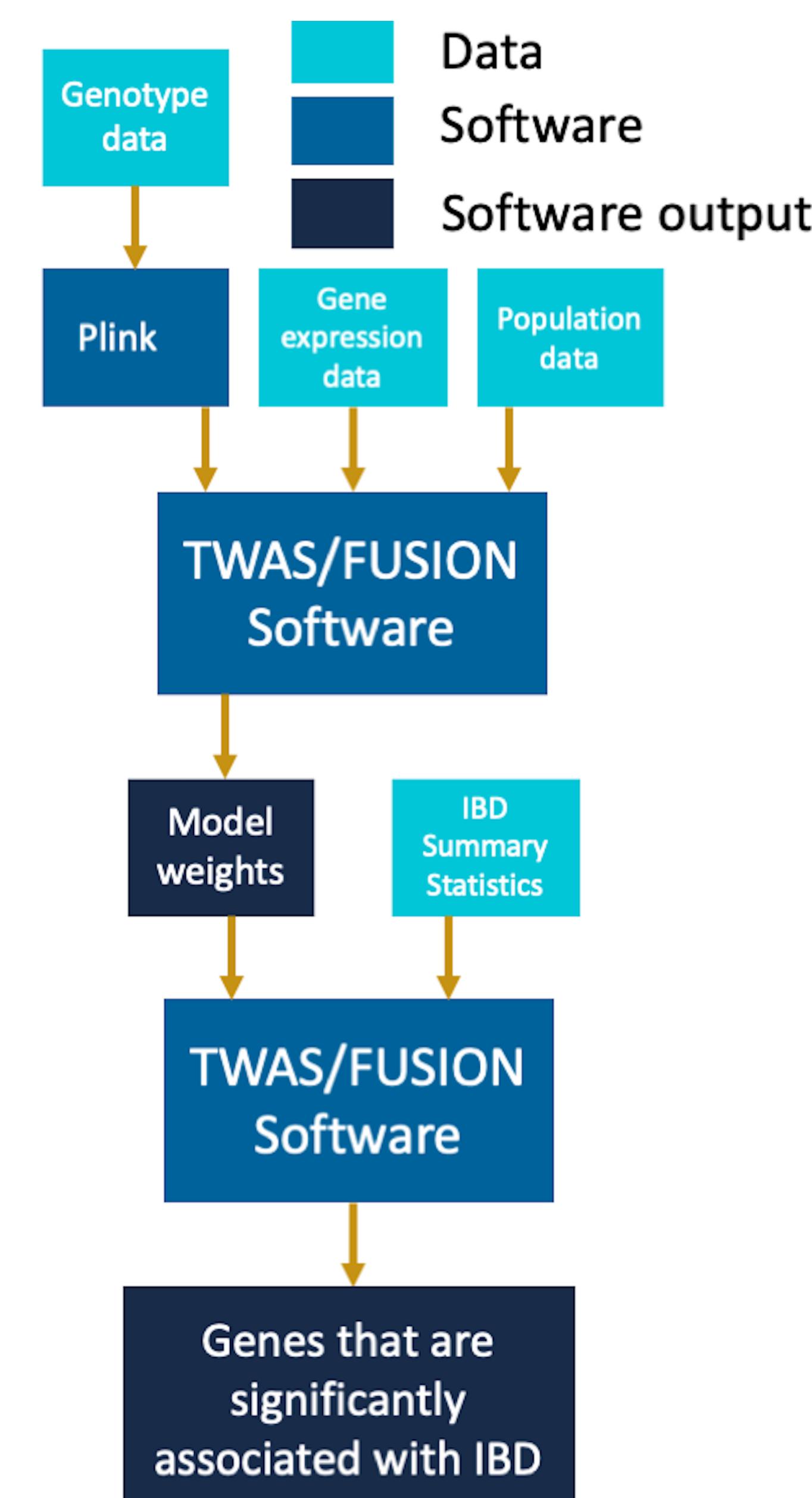
Halıcıoğlu Data Science Institute at UC San Diego

## Introduction



- Understanding how genetic variation impacts gene expression can help us identify gene-based mechanisms of disease risk.
- Previous approaches to identifying the gene-based mechanisms of disease risk, like GWAS, find thousands of associated variants, but many of those lie in noncoding sequences of the DNA and are thus difficult to understand<sup>1</sup>.
- TWAS leverages a cohort of individuals with measured gene expression and genotyping to discover gene-trait associations from GWAS summary statistics<sup>2</sup>.
- TWAS is powerful in finding disease-associated genes because it aggregates gene expression-modifying effects of all nearby variants<sup>3</sup>.
- Inflammatory Bowel Disease (IBD) is a highly heritable disease that causes a chronic inflammation of tissues in an individual's digestive tract and affects 3.1 million adults in the United States<sup>4,5</sup>.

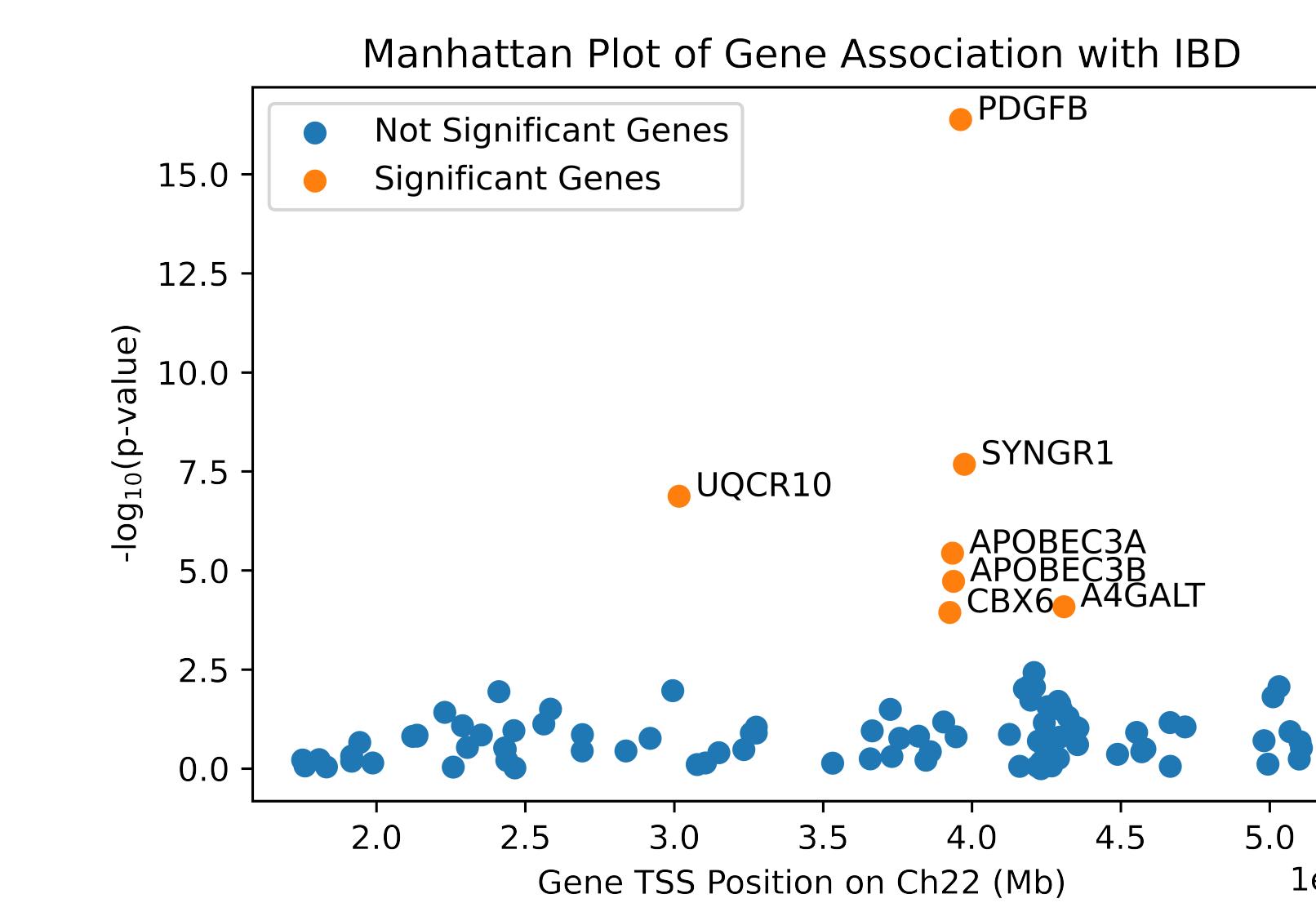
## Methods and Data



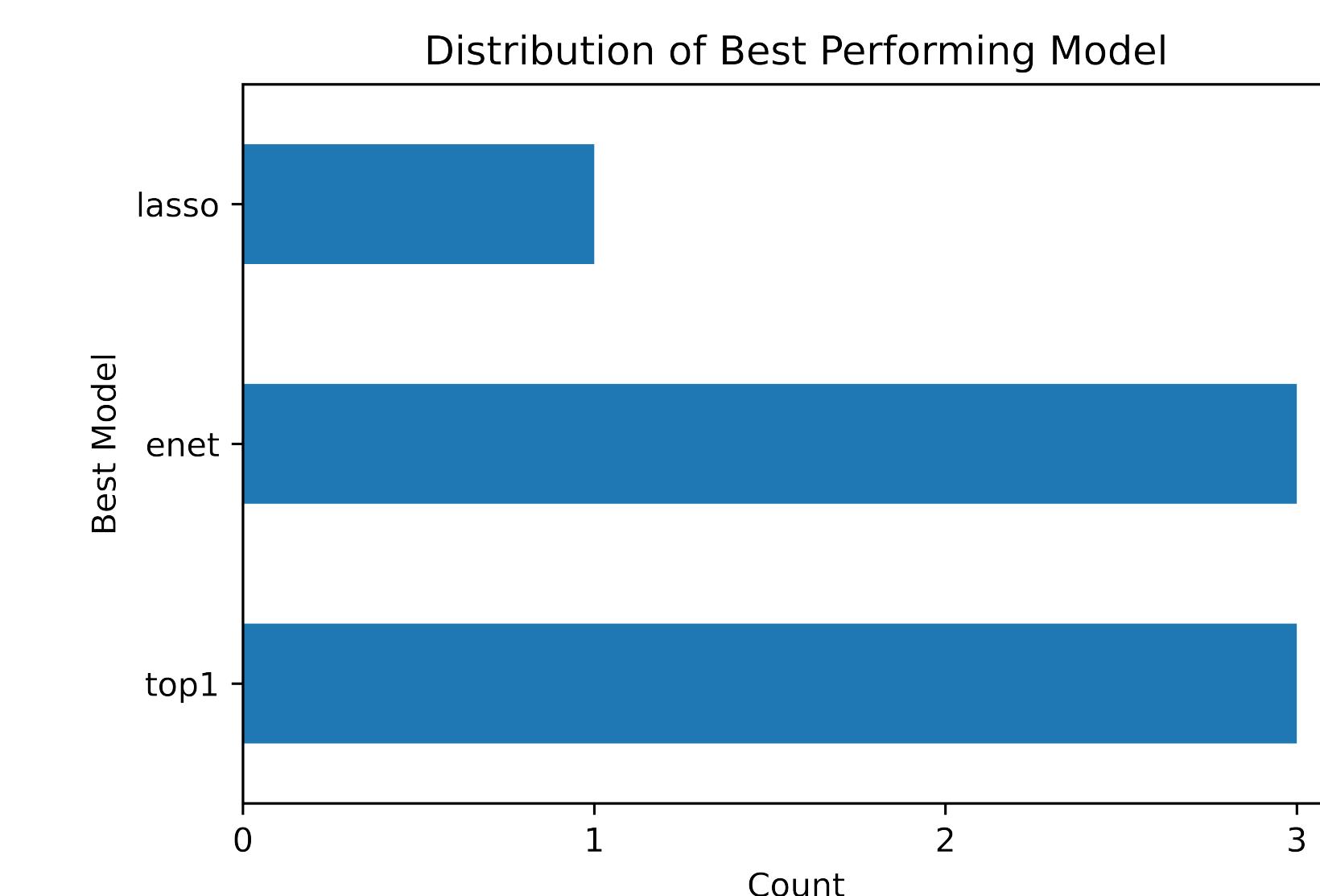
- The genotype data comes from the 1000 Genomes Project and contains information on millions of single nucleotide polymorphisms (SNPs), or individual variants of DNA, on chromosome 22. Chromosome 22 was chosen because its small size fits our limited computation capacity.
- The gene expression data comes from the Geuvadis RNA-sequencing project.
- Combined, these sources contain genetic data for 344 individuals across four populations: people of Northern and Western European Ancestry from Utah (CEU), Fins (FIN), British (GBR), and Toscani (YRI).
- The genome wide association study (GWAS) summary statistics for IBD were obtained from the analysis carried out in the paper Finucane, et al. 2015 Nature Genetics<sup>6</sup>.

## Results

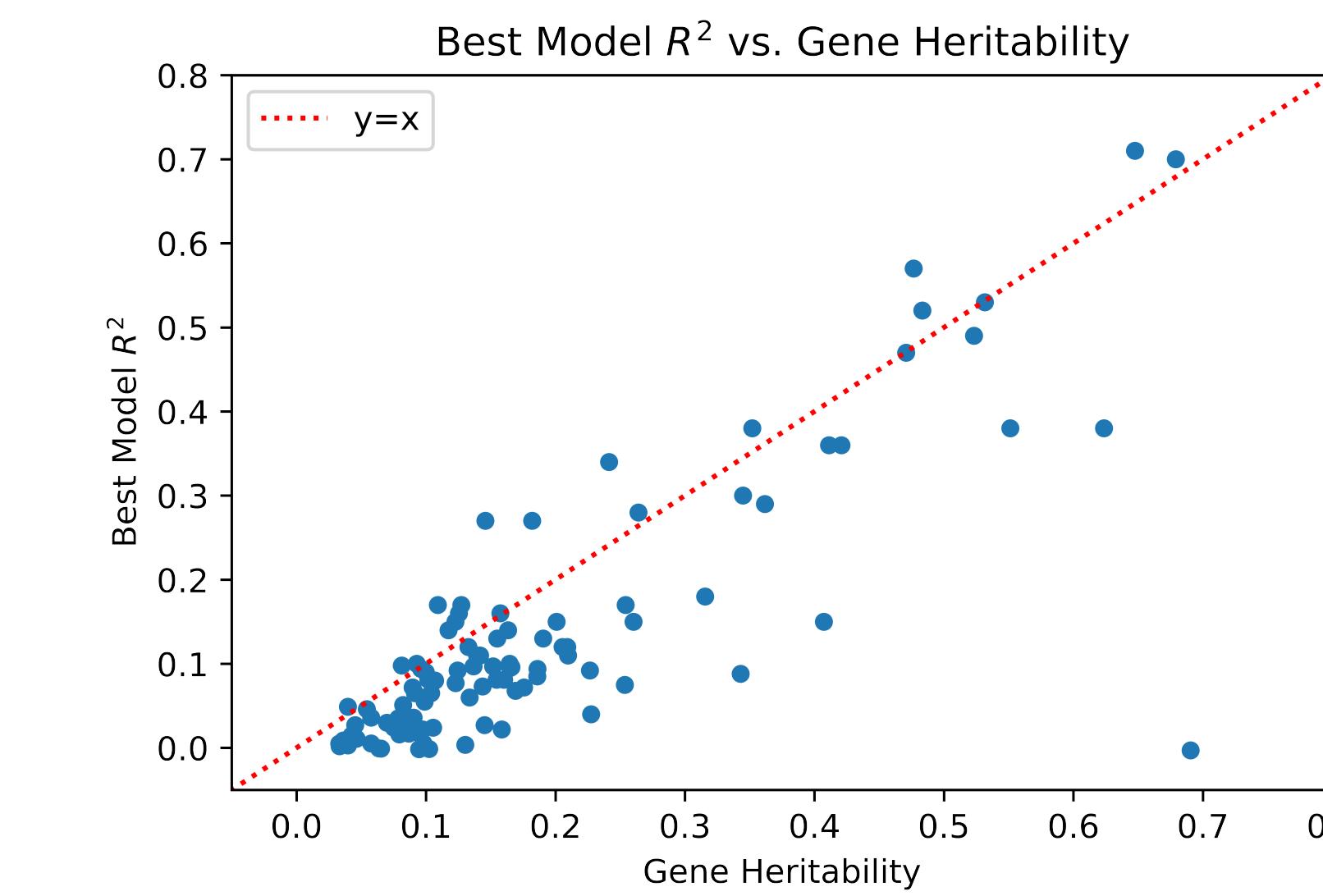
- Identified 100 highly heritable genes, 7 of which had significant association with IBD.



- Of the seven genes, 5 of them were very close together and shared the same best GWAS SNP, which might suggest that they contribute to the same signal.
- PDGFB had the most significant association with a p-value of  $4.17 \times 10^{-17}$ .
- The single variant most highly associated with gene expression and elastic net each performed the best for 3 genes.



- $R^2$  was roughly bounded by heritability, as expected.



## Conclusion

- TWAS was ultimately able to identify 7 genes that are associated with IBD.
- One of the genes, A4GALT, did not have any genome-wide significant GWAS SNPs nearby. This highlights that TWAS is able to find relevant genes even if that locus is not genome-wide significant in the GWAS analysis.
- Our findings can be verified by other sources. For example, Marigorta et al. (2017) showed that there was an association between SYNGR1 and IBD, which our analysis also identified.
- Identification of these genes can be useful for disease prevention and early detection.

## Website



## References

- Gene Expression. *Nature News*. Nature Publishing Group, <https://www.nature.com/scitable/topicpage/gene-expression-14121669/>
- Michael Wainberg. *Nature Genetics*, 51(4):592-599, March 2019
- Molly West. *Human Genomics*, 13(1), August 2019.
- Warren Strober. *Journal of Clinical Investigation*, 117(3):514-521, March 2007.
- Fang Xu. *MMWR. Morbidity and Mortality Weekly Report*, 67(6):190-195, February 2018.
- Hilary K Finucane. *Nature Genetics*, 47(11):1128-1235, September 2015.