

Evaluating the effectiveness of a PCOS-mediator effect polygenic risk score in predicting Type 2 diabetes in women



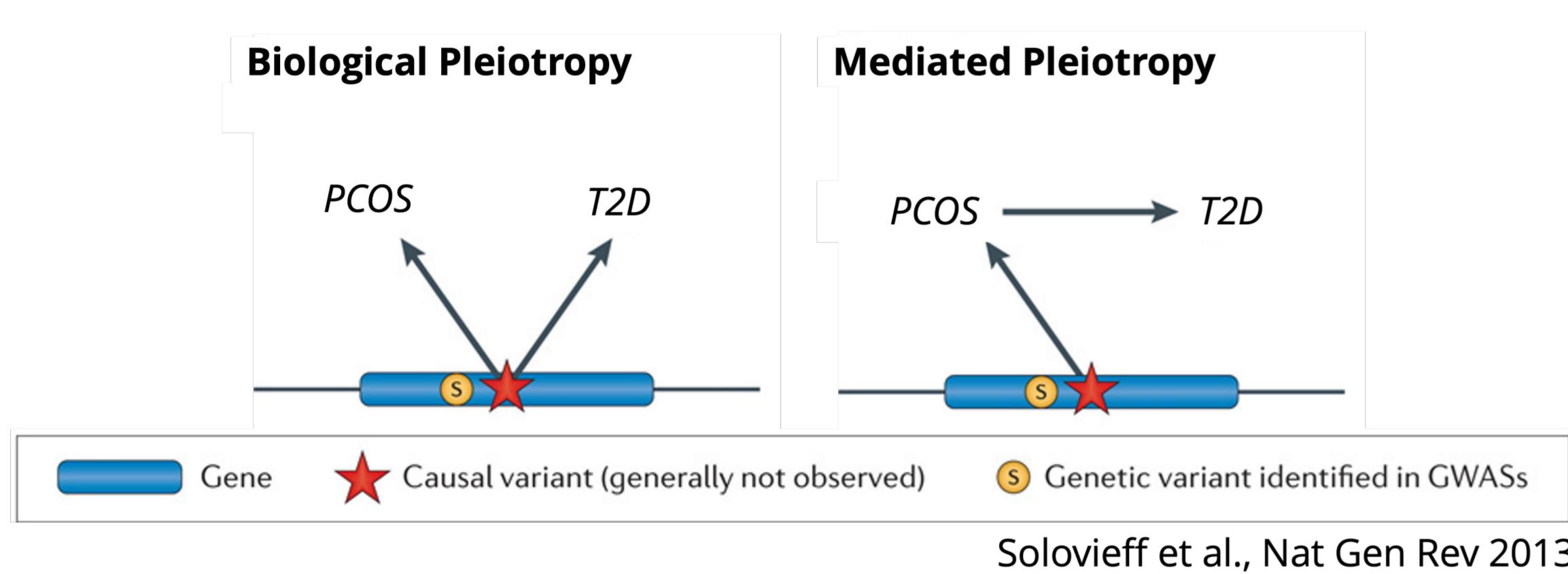
Ananya Rajagopalan¹, Lannawill Caruth¹, Jagyashila Das¹, Lindsay Guare¹, Stephanie Mohammed¹, Marijana Vujković², Shefali Setia-Verma¹

¹Department of Pathology and Laboratory Medicine, Perelman School of Medicine, Philadelphia, PA, United States

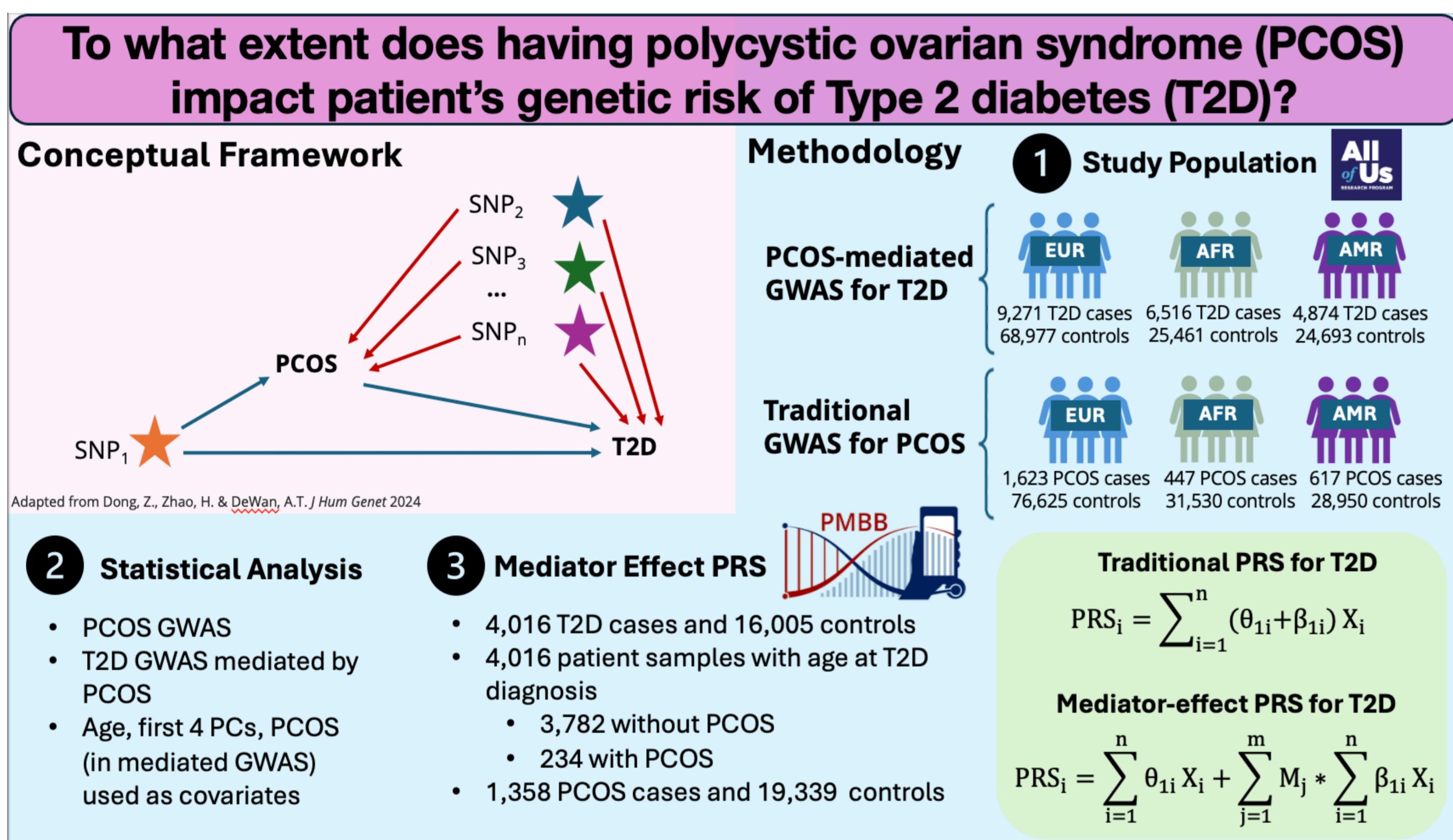
²Division of Translational Medicine and Human Genetics, Perelman School of Medicine, Philadelphia, PA, United States

Background

- Type 2 diabetes (T2D) is a complex metabolic disorder influenced by multiple genetic, environmental, and lifestyle factors.
- Individuals of reproductive age with polycystic ovary syndrome (PCOS) are at significantly greater risk of later developing T2D.
- Previous studies have established plausible genetic correlation between T2D and PCOS.
- Current polygenic risk scores (PRS) often assess direct genetic contributions to diabetes risk, overlooking the indirect effect of potential mediating pathways.
- Determining whether an individual's risk for T2D is absolute or if it comes from PCOS mediation is critical for shifting the goal of clinical interventions for women to address this underlying pathophysiology.**



Graphical Abstract



Methods

Phenotyping

To obtain women from the All of Us (AOU) dataset who have been diagnosed with T2D and/or PCOS, we used both the 9th and 10th edition of the International Classification of Disease (ICD) codes corresponding to these phenotypes. The number of patients in each cohort stratified by phenotype and ancestry are shown in the table below.

Ancestry	T2D controls	T2D cases	PCOS controls	PCOS cases
EUR	68,977	9,271	76,625	1,623
AFR	25,461	6,516	31,530	447
AMR	24,693	4,874	28,950	617

Genome-Wide Association Tests (GWAS)

We leveraged microarray genotyping array data and short-read whole genome sequencing data from AOU, which were provided as the input genotype data in SAIGE to run both traditional and mediated GWAS. Age and the first four principal components were used as covariates. When a mediating GWAS was conducted, we added the mediating phenotype (PCOS) as a covariate. PLINK was used for meta-analysis.

Mediator Effect Polygenic Risk Score (PRS)

We selected genetic variants for the mediator-effect PRS weights by leveraging 372 SNPs that had been robustly identified in large-scale GWAS studies as significantly associated with either T2D or PCOS. We used Penn Medicine BioBank (PMBB) data as the testing set and we calculated the Area Under the Receiver Operating Characteristic (AUC-ROC) curve to empirically compare the predictive accuracy of the traditional PRS versus the mediator effect PRS for T2D.

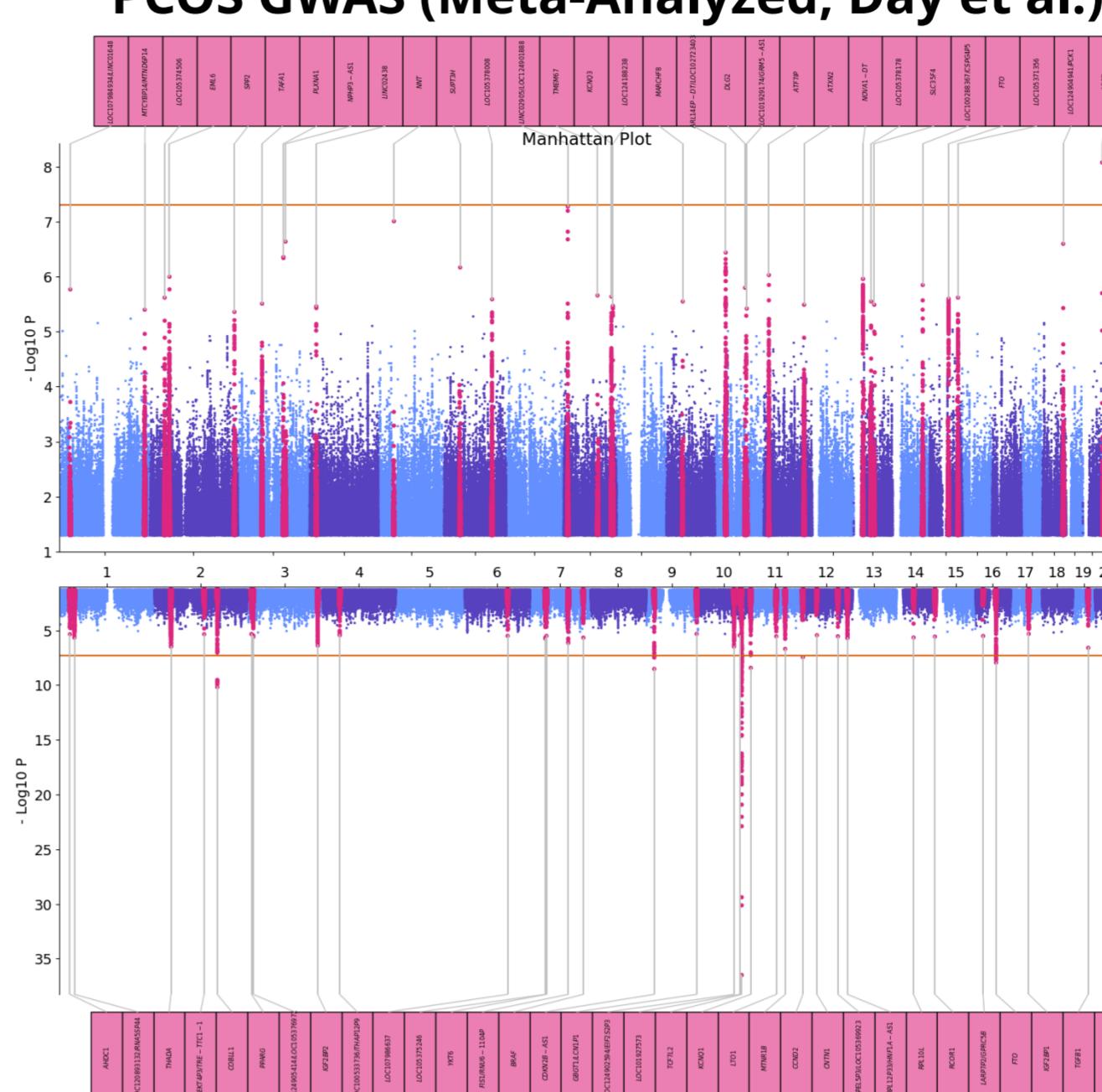
$$\text{PRS}_i = \sum_{i=1}^n \theta_{1i} X_i + \sum_{j=1}^m M_j * \sum_{i=1}^n \beta_{1i} X_i$$

This mediator-effect PRS is calculated by incorporating the effect size of each variant from the meta-analyzed PCOS-mediated T2D GWAS (θ_{1i}) and the PCOS GWAS (β_{1i}), individual-level genotype data for each variant (X_i), and individual-level sample data indicating a diagnosis of PCOS (M_j).

Results

Hudson plots below illustrate the difference between direct genetic effects on T2D (biological pleiotropy) and indirect effects mediated through PCOS (mediated pleiotropy).

PCOS GWAS (Meta-Analyzed, Day et al.)



T2D Mediated by PCOS GWAS (Meta-Analyzed)

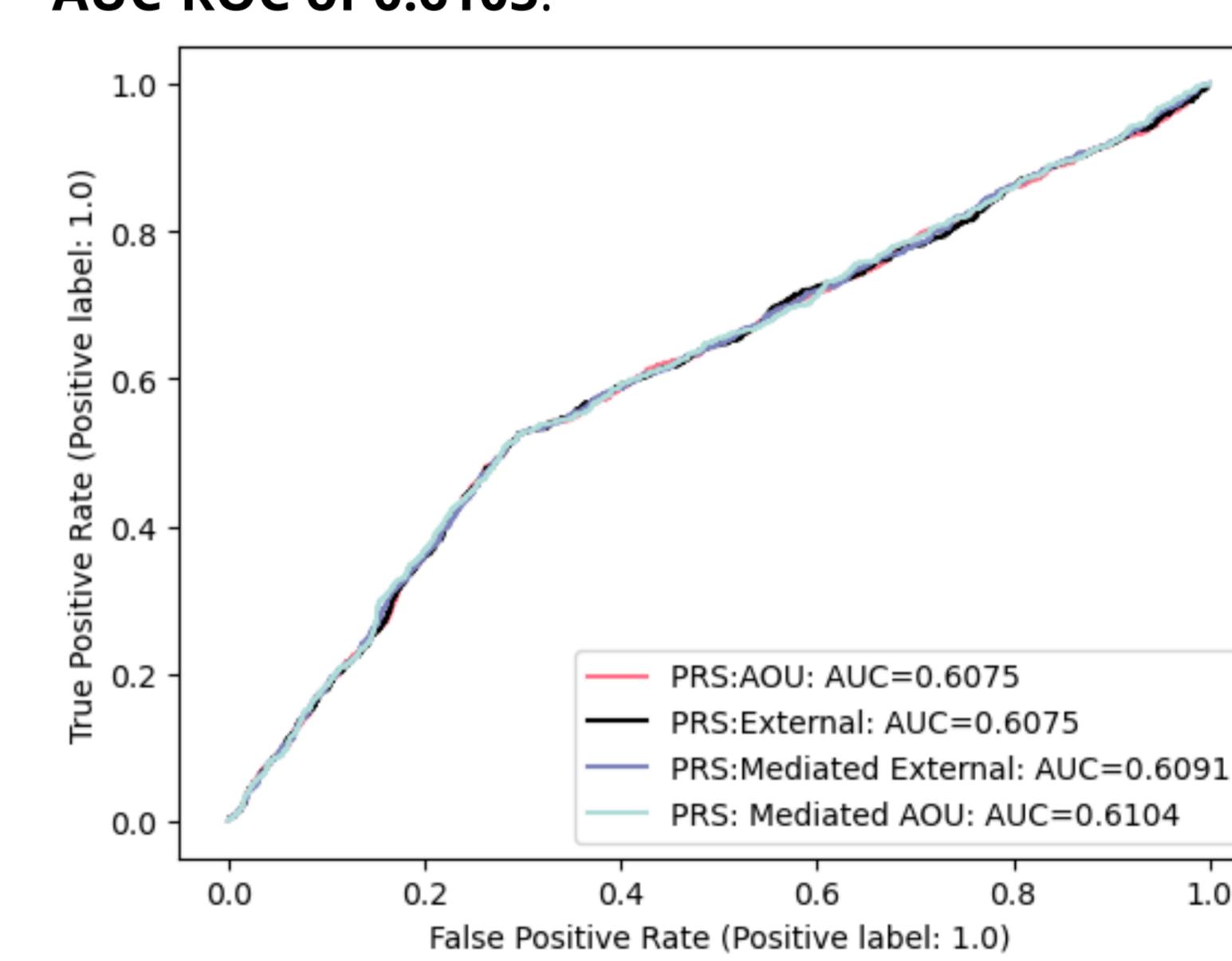
The *FTO* gene (Chromosome 16) is shared between both GWAS, indicating that variants on that gene have a direct biological effect on T2D and PCOS (this supports discussions in existing literature).

T2D GWAS (Meta-Analyzed)

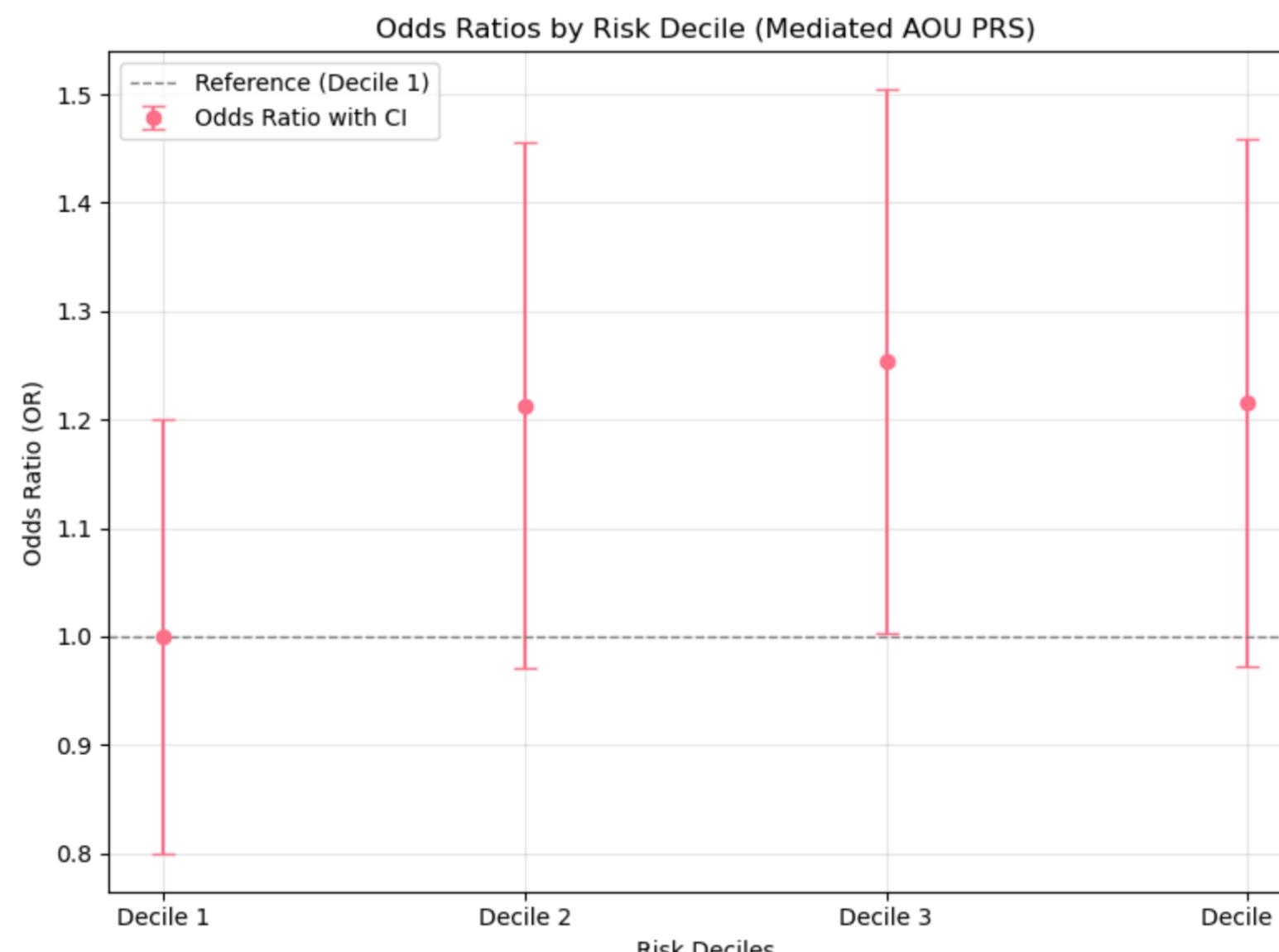
The *LOC107986105* gene (Chromosome 3) disappears from the T2D GWAS to the mediated GWAS, suggesting that the effect of variants in this gene mediate T2D indirectly through PCOS (instead of directly).

PRS Performance

Using PMBB data as the testing set, the traditional T2D PRS demonstrated modest predictive accuracy, with an **AUC-ROC of 0.6075**. Our mediator-based PRS achieved an **AUC-ROC of 0.6105**.



The mediated PRS achieves an AUC comparable to the unmediated PRS for both sets of weights considered with minimal improvement. These AUC values indicate that the mediated PRS is at least capable of capturing the same level of risk as the unmediated score and may include a level of granularity not present in the unmediated score.



The decile plot on the left shows that as the risk decile increases (higher PRS value for an individual) due to patients having PCOS, the odds ratio for developing T2D also increases. This indicates that our mediator effect PRS is indeed capable of capturing the heightened risk of developing T2D for individuals with PCOS, compared to individuals without PCOS.

Conclusions

- In this work, we developed an innovative approach to risk prediction that distinguishes between direct genetic effects on T2D, and indirect effects mediated through PCOS.
- This approach provides a more nuanced understanding of diabetes risk by acknowledging that genetic variants may impact disease risk through complex, interconnected biological mechanisms rather than through direct, linear effects.
- This approach not only enables more accurate disease prediction, but preventive action as well since clinical interventions can be adapted to the underlying, mediating phenotype.

Future Work

- Limited sample size in the GWAS dataset is likely contributing to the reduced predictive performance of the mediator-effect PRS, so we hope to test this approach using data with a larger number of patients with PCOS (incorporating data from the UK Biobank, eMerge, and Million Veterans Program).

Acknowledgements

- Research reported in this poster was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health under award number R01HD110567-01A0.
- We gratefully acknowledge All of Us participants for their contributions, without whom this research would not have been possible. We also thank the National Institutes of Health's All of Us Research Program for making available the participant data examined in this study.
- We acknowledge the Penn Medicine BioBank (PMBB) for providing data and thank the patient-participants of Penn Medicine who consented to participate in this research program. We would also like to thank the Penn Medicine BioBank team for providing genetic variant data for analysis. The PMBB is approved under IRB protocol #813913 and supported by Perelman School of Medicine at University of Pennsylvania, a gift from the Smilow family, and the National Center for Advancing Translational Sciences of the National Institutes of Health under CTSA award number UL1TR001878.