

**Comparative analysis of the Predictive Utility of Integrated and Single-Source Polygenic
Risk Scores for SAT Traits
Group 4: PRS Avengers**

Group Members

Fredrick E. Kakembo

Caitlin Grace Short

Jacob Apibilla Ayembilla

Benson R. Kidenya

Kingsley Osei

Emmanuella E. Matumamboh

1. Introduction

Subcutaneous adipose tissue (SAT) plays a crucial role in energy storage, thermoregulation, and metabolism of biomolecules (hormones, glucose, and lipids) (Ziadlou et al., 2024). Variations in SAT distribution and volume are strongly influenced by genetic factors and are implicated in a range of cardiometabolic outcomes, including obesity, insulin resistance, and cardiovascular disease (Sanghera et al., 2019; Zhang et al., 2025). Understanding the genetic architecture underlying the SAT is therefore vital for both biological insight and risk stratification in metabolic disorders.

Although individually significant markers in genome-wide association scans (GWAS) explain limited heritability of complex traits, evidence has been accruing that a considerable proportion of phenotypic variation can be explained by the ensemble of markers not achieving significance (Dudbridge, 2013). Polygenic score analysis has recently generated much interest in assessing the explanatory power of an ensemble of markers. Polygenic risk scores (PRS) have emerged as a powerful tool for quantifying inherited susceptibility to complex traits by aggregating the effects of multiple genetic variants. Traditional single-source PRS construction methods typically rely on summary statistics from genome-wide association studies (GWAS) and apply fixed or heuristic thresholds (e.g., clumping and p-value thresholding). While these methods are straightforward, they may fail to capture the full polygenic signal or the correlation structure of the genome, potentially limiting their predictive utility (Truong et al., 2024).

In contrast, integrative approaches such as elastic net regression and Bayesian mixture models (e.g., PRS-CSx, PRSmixture) allow for the simultaneous modeling of correlated variants and multiple information sources, potentially enhancing predictive performance (Truong et al., 2024). These methods can incorporate shrinkage, variable selection, and cross-trait or multi-cohort information, offering a more nuanced representation of genetic architecture.

Despite the growing availability and sophistication of PRS tools, limited work has systematically evaluated whether integrative methods provide meaningful improvements over single-source approaches in predicting SAT traits. Given the central role of SAT in metabolic health, this study seeks to fill that gap by conducting a comprehensive comparison of integrative and traditional PRS models. Here, we evaluate the predictive performance of pre-constructed polygenic risk scores for subcutaneous adipose tissue (SAT), comparing single-method approaches (e.g., clumping + thresholding, LDpred2) with integrative models such as elastic net regression and PRSmixture. We assess the predictive utility of each method in independent validation cohorts and quantify their performance using standard metrics such as AUC, R^2 , and mean squared error. Our goal is to determine whether integration confers added predictive value and to guide the optimal construction of PRS for adiposity traits. We hypothesize that PRSmix is not a better predictor of SAT levels in the African population than the individual PGSs.

2.0 Materials and methods

2.1 Dataset

The dataset for this analysis is a polygenic score (PGS) for platelet count traits for predicting the SAT levels among Africans, which was obtained from the Polygenic Scores (PGS) catalog and imputed by the course facilitators. The dataset contains 24 PRS scores with 10 principal components (PCA). It includes age and sex as covariates, with each participant having a unique identifier.

2.2 Study design

In this study, we adopted the PRSmix approach to evaluate the predictive performance of single-source and integrative PGS for SAT in African populations. The dataset consists of PGS scores for different traits obtained from the PGS catalog. The data was first cleaned to remove outliers, missing values, etc, and explored to understand the data. Summary statistics for the data set were

generated and reported in tables and graphs. The data was split into a train (80%) and test (20%) dataset. The elastic net algorithm was used to train the data using train the data. The test data was used to evaluate the model's performance. The MSE, AUC, and R² values were used for the model evaluation.

2.3 Data Preparation

We began with a dataset containing: the SAT the continuous target variable, Multiple PGS columns (e.g., PGS000109, PGS000186, etc.), and Covariates including sex, age, and genetic ancestry principal components (PC1–PC10). We then loaded the data into a Python Jupyter notebook using NumPy and Pandas. We then generated some descriptive statistics of the data.

2.4 Covariate Adjustment (Residualization)

To isolate the genetic signal, we first regressed SAT scores on age, sex, and PCs (PC1–PC10) using ordinary least squares (OLS). This step removed the non-genetic confounders. The residuals from this model represented the SAT variation not explained by demographics or ancestry the portion.

2.5. PRSmix Construction via Elastic Net

We used Elastic Net Regression (via ElasticNetCV) to model the residual SAT score as a linear combination of all available PGS scores. This approach balances selection (via Lasso) and shrinkage (via Ridge), automatically selects and weights the most predictive PGSs, and avoids overfitting even in the presence of multicollinearity. We split the data (80/20) and trained the Elastic Net model on the training set.

2.5. Performance Evaluation of PRSmix

On the test set, we evaluated Prediction accuracy using R² (coefficient of determination). This R² was interpreted as the variance in SAT explained by PRSmix.

2.6. Benchmarking Individual PGS Scores

We repeated the same prediction process separately for each PGS, fit a univariate linear model using each PGS as the sole predictor, and computed test-set R² for each model. This gave us a

baseline to compare how well each single PGS predicts SAT relative to the integrative PRSmix model.

2. 7 Analysis Pipeline



2.8 Statistical analysis

The continuous variables were reported as mean \pm SD or median \pm IQR, depending on the distribution of the variables. Categorical variables were reported as count (%). Statistically significant differences between the groups were determined using ANOVA and Student's test. *P-values* < 0.05 were considered statistically significant. The data were analyzed in Python with numpy, pandas, seaborn, sklearn, and scipy packages.

3.0 Results

3.1 Summary statistics

We used the Elastic net regression to examine the predictive performance of single-source PRS and integrative. Table 1 shows the summary statistics of the data used for this study. The data is made up of 1624 participants, consisting of 700 males and 924 females. The mean age of the males of 32.78 ± 17.95 , while that of the females was 34.22 ± 17.7 , which was not a significant

difference. The mean subcutaneous adipose tissue (SAT) was statistically significantly different between males and females (209.80 ± 79.39 vs. 227.00 ± 76.50). The following PRS were different statistically among males and females: PGS000109, PGS000186, PGS000186, PGS0002464, PGS0002513, PGS0002562, PGS0002611. Except PCA2, all the PCAs did not statistically significant differ between males and females.

Table 1: Summary statistics of the subcutaneous adipose tissue polygenic risk scores data

	Ancestry	Male	Female	p
n		700	924	
age (mean (SD))		32.76 (18.95)	34.22 (17.77)	0.109
SAT (mean (SD))		209.80 (79.39)	227.00 (76.50)	<0.001
PGS000109 (mean (SD))		0.00 (0.00)	0.00 (0.00)	0.036
PGS000186 (mean (SD))		0.02 (0.00)	0.02 (0.00)	<0.001
PGS001238 (mean (SD))		0.00 (0.00)	0.00 (0.00)	0.985
PGS001973 (mean (SD))		0.00 (0.00)	0.00 (0.00)	0.386
PGS002191 (mean (SD))		0.00 (0.00)	0.00 (0.00)	0.367
PGS002343 (mean (SD))		0.00 (0.00)	0.00 (0.00)	0.422
PGS002373 (mean (SD))		0.00 (0.00)	0.00 (0.00)	0.643
PGS002415 (mean (SD))		0.01 (0.00)	0.01 (0.00)	0.12
PGS002464 (mean (SD))		0.01 (0.00)	0.01 (0.00)	0.014
PGS002513 (mean (SD))		0.01 (0.00)	0.01 (0.00)	0.034
PGS002562 (mean (SD))		0.00 (0.00)	0.00 (0.00)	0.007
PGS002611 (mean (SD))		0.00 (0.00)	0.00 (0.00)	0.003
PGS002660 (mean (SD))		0.00 (0.00)	0.00 (0.00)	0.089
PGS002709 (mean (SD))		0.00 (0.00)	0.00 (0.00)	0.109
PGS003546 (mean (SD))		0.00 (0.00)	0.00 (0.00)	0.278
PGS003932 (mean (SD))		0.00 (0.00)	0.00 (0.00)	0.418
PGS004352 (mean (SD))		0.00 (0.00)	0.00 (0.00)	0.127
PGS004582 (mean (SD))		0.00 (0.00)	0.00 (0.00)	0.587
PGS004583 (mean (SD))		0.00 (0.00)	0.00 (0.00)	0.917
PGS004584 (mean (SD))		0.03 (0.01)	0.03 (0.01)	0.025
PGS004811 (mean (SD))		0.00 (0.00)	0.00 (0.00)	0.935
PGS004812 (mean (SD))		0.00 (0.00)	0.00 (0.00)	0.732
PGS004813 (mean (SD))		0.00 (0.00)	0.00 (0.00)	0.767
PGS004814 (mean (SD))		0.00 (0.00)	0.00 (0.00)	0.732
PC1 (mean (SD))		0.00 (0.02)	0.00 (0.01)	0.517
PC2 (mean (SD))		0.00 (0.02)	0.00 (0.01)	0.029
PC3 (mean (SD))		0.00 (0.01)	0.00 (0.01)	0.844
PC4 (mean (SD))		0.00 (0.01)	0.00 (0.01)	0.558
PC5 (mean (SD))		0.00 (0.01)	0.00 (0.01)	0.496

PC6 (mean (SD))	0.00 (0.02)	0.00 (0.02)	0.452
PC7 (mean (SD))	0.00 (0.01)	0.00 (0.01)	0.495
PC8 (mean (SD))	0.00 (0.02)	0.00 (0.01)	0.443
PC9 (mean (SD))	0.00 (0.01)	0.00 (0.01)	0.737
PC10 (mean (SD))	0.00 (0.01)	0.00 (0.01)	0.811

Distribution of SAT levels among males and females by age groups

Figure 1 shows the distribution of SAT levels among males and females. The results that the mean SAT was higher among all the ages than females.

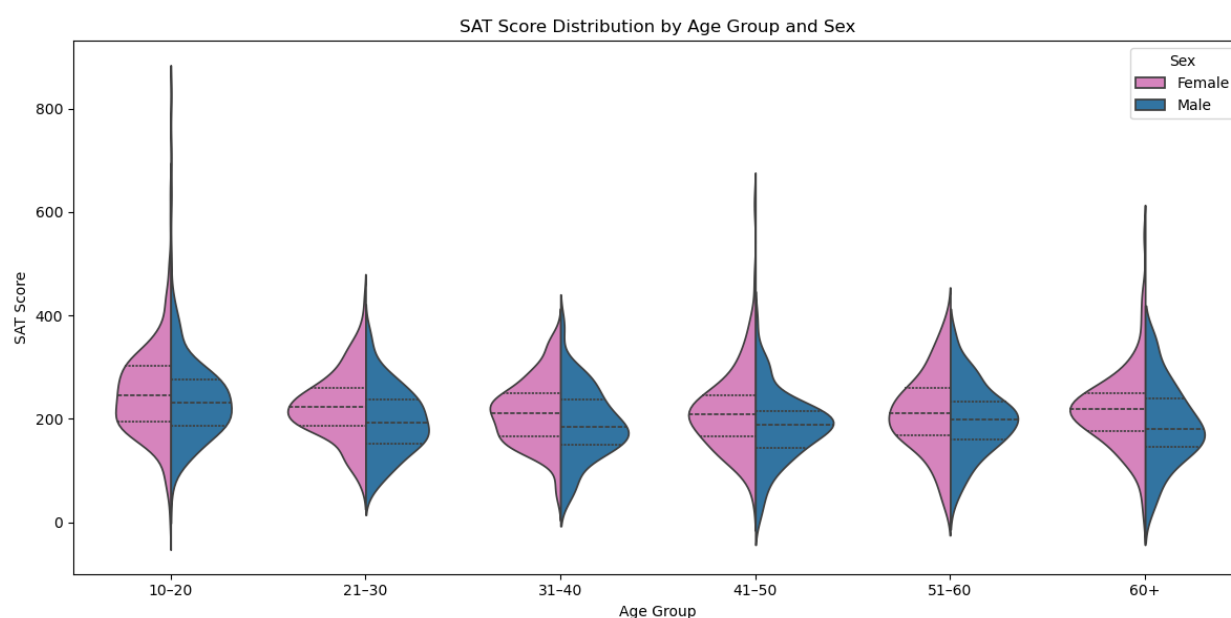


Fig. 1: Distribution of SAT levels by age groups and sex.

3.2 Comparative analysis of the predictive performance of single-sourced PRS and

Figure 2 shows the predictive performance of PRSmix and the single-sourced PGSs. The data showed that PGS002660 recorded the highest performance with an R2 OF 0.057 and the PRSmix -0.0029

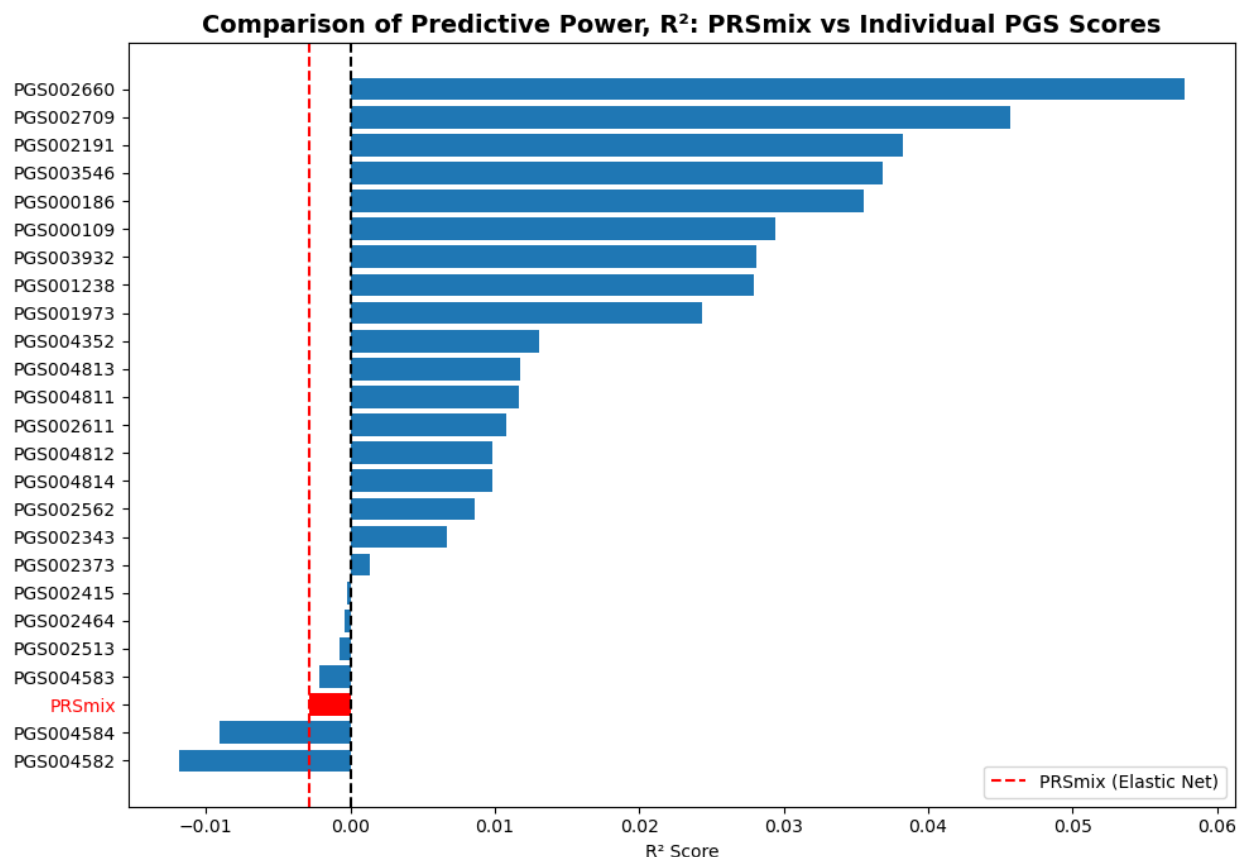


Fig. 2: Comparative analysis of predictive performance of single-sourced PRS and integrative PRS using the Elastic net regression model.

Correlation analysis for the PGS scores

Fig. shows the correlation heatmap of the top 10 PGS scores used for the PRS mix analysis. The heatmap revealed that the PGSs were poorly correlated with SAT. PGS002660 recorded the highest correlation with SAT, with a correlation score of 0.19, and PGS004583 recorded the poorest correlation coefficient (0.01).

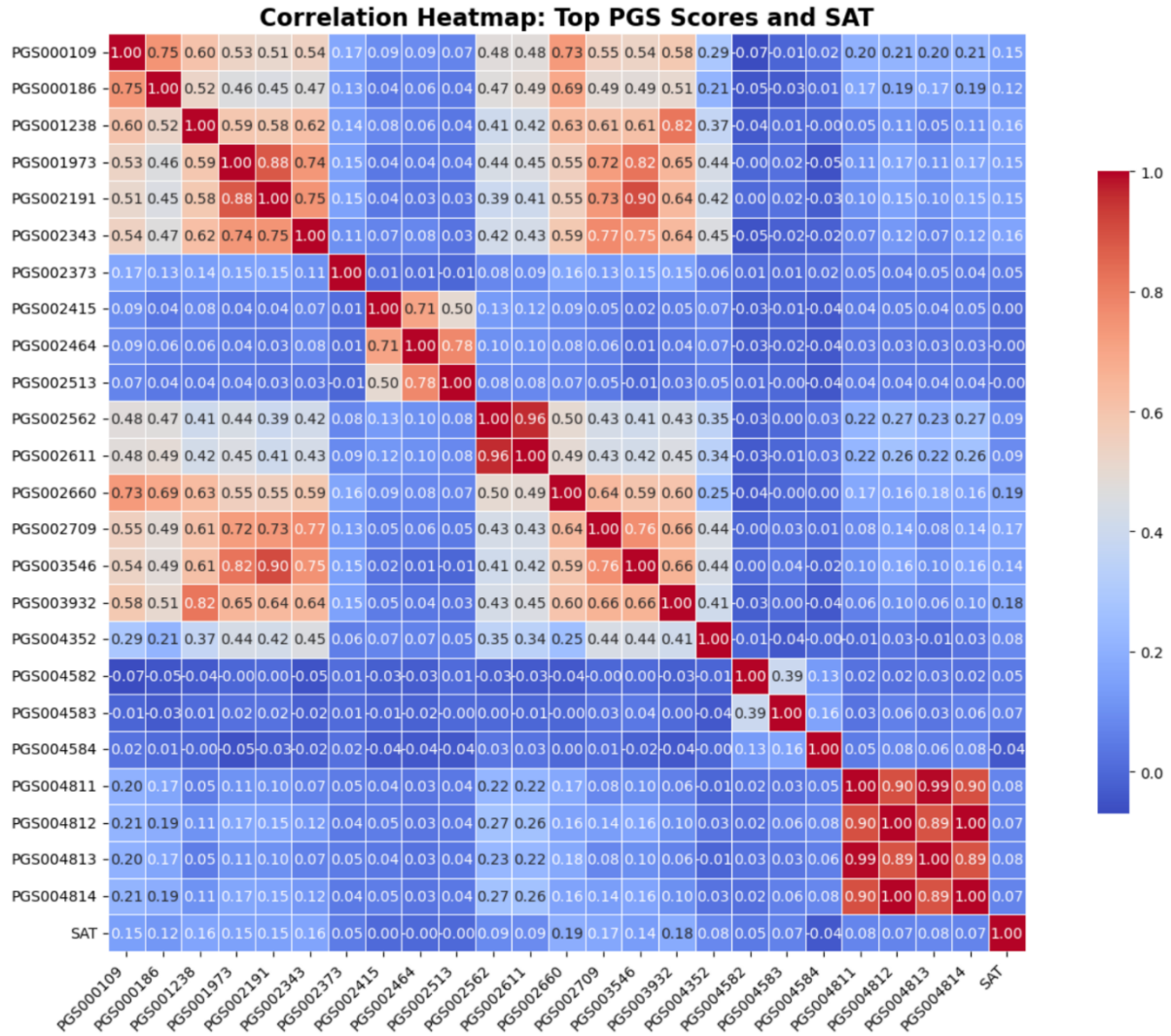


Fig. 3: Comparative analysis of predictive performance of single-sourced PRS and integrative PRS using the Elastic net regression model.

Predictive weights of PRSmix and the individual PGS

Figure 4 shows the predictive weights of the performance of the individual PRS. While all the PGS had a positive weight, PGS004584 recorded a negative weight.

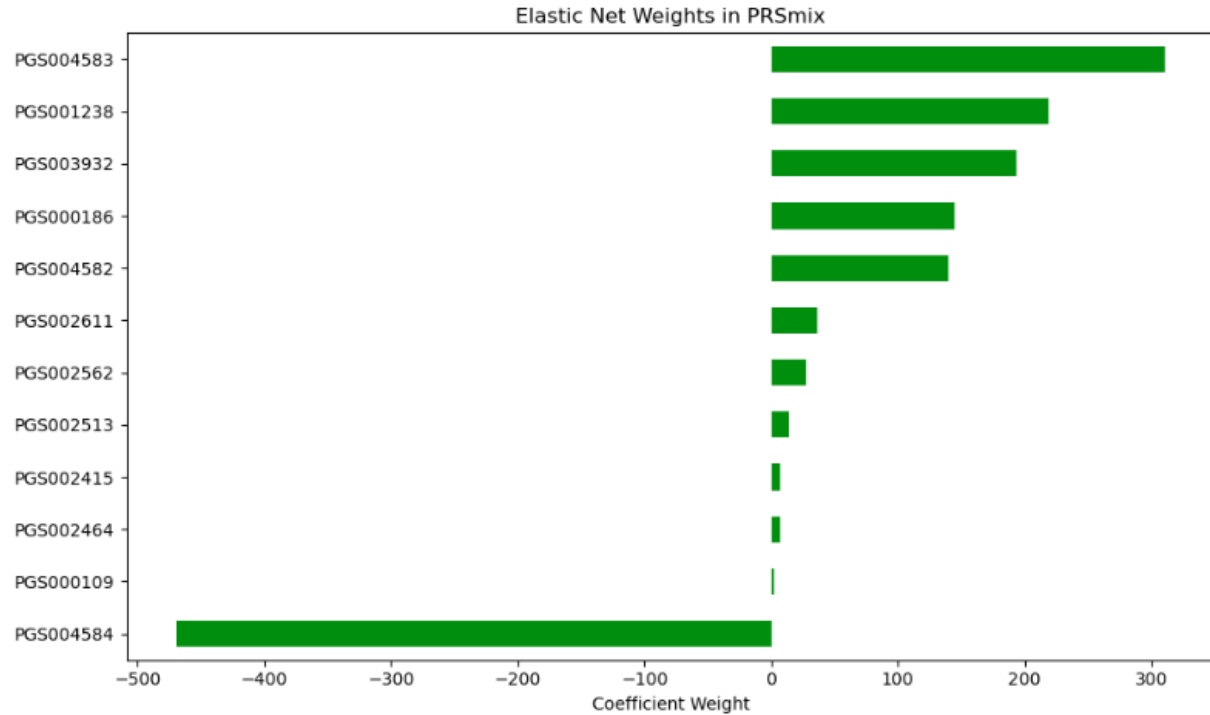


Fig. 4: Comparative analysis of predictive weights of single-sourced PRS and integrative PRS using the Elastic net regression model.

Discussion

We set out to examine the predictive performance of integrative PRS and single-sourced PRS in predicting SAT levels in an African population. Variations in SAT distribution and volume are strongly influenced by genetic factors and are implicated in a range of cardiometabolic outcomes, including obesity, insulin resistance, and cardiovascular disease (Sanghera et al., 2019; Zhang et al., 2025). Understanding the genetic architecture underlying the SAT is therefore vital for both biological insight and risk stratification in metabolic disorders. Despite the growing availability and sophistication of PRS tools, limited work has systematically evaluated whether integrative methods provide meaningful improvements over single-source approaches in predicting SAT traits. This study found that the PGS scores were poorly correlated with the SAT level. This is contrary to the reported literature. In a cross-sectional study by Han et al., (2018) both platelet count (PLT) and platelet crit PCT were associated with body fat, and the associations depended on fat location. Because PCT is mainly influenced by PLT, these associations suggest that platelet quantity is more closely related to body fat mass and fat distribution.

The analysis revealed that the PRSmix using the elastic net regression for SAT in the African population was poorly performing compared to the individuals' PGSs. This finding confirms the correlation results, which showed that the individual PGS were poorly correlated with SAT.

Limitations

PRS construction relies on publicly available GWAS summary statistics, which may be biased, are heterogeneous, and have different phenotypic definitions, which may affect the accuracy and comparability of the derived scores. Subcutaneous adipose tissue measurements employ imaging-based techniques (MRI, DXA, CT) and measurement using different protocols, which can introduce heterogeneity that impacts PRS performance.

Recommendation

More research is needed to fully elucidate the molecular mechanisms underlying the interaction between platelets and body fat, particularly subcutaneous fat. African-specific PRS are required to study the portability, applicability, and clinical utility of PRS. Longitudinal studies incorporating platelet polygenic scores are needed to understand the causal relationships between platelet function, fat accumulation, and disease development.

Conclusions

Genetic predisposition to SAT was found to be significantly associated with SAT levels among Africans. However, PRSmix predisposing Africans was poorly associated with SAT levels. Better understanding the mechanisms of these relationships may allow tailored intervention in obesity or early selection of populations at risk of metabolic disease.

Address to Questions

Q1.

Statement to support alternative hypothesis

From previous studies, the PRSmix approach is likely to show an improvement in predictive performance in continental African populations compared to using a single PRS primarily derived from European-ancestry data. This is because the core mechanism of PRSmix integrates information from various PRSs and potentially leverages genetically correlated traits, theoretically

capturing a broader spectrum of genetic effects, which is likely to offer some benefits even when the foundational data has biases (Truong et al., 2024).

However, continental African populations possess the greatest genetic diversity globally, leading to significant variations in genetic markers and their effects across different ethnic groups. This inherent diversity makes it difficult for any Polygenic Risk Score (PRS), including combined approaches like PRSmix, to transfer effectively across the continent without sufficient local genetic data. African populations equally exhibit shorter blocks of linkage disequilibrium compared to Europeans, which affects how genetic variants are inherited and how well SNPs can identify causal variants (Segun Fatumo et al., 2023).

Given that most existing PRSs and their foundational Genome-Wide Association Studies (GWAS) are based on European genetic patterns, applying these directly or combined (even with PRSmix) often results in reduced accuracy for African populations. The primary obstacle remains the severe underrepresentation of African populations in global genomic research; as of recent data, approximately 86% of GWAS data is derived from individuals of European descent (Segun et al. 2022). This Eurocentric bias means that the individual PRSs within PRSmix are largely built on European genetic information. While PRSmix aims to improve generalizability, it cannot fully overcome the fundamental lack of GWAS conducted in diverse African ancestries, which is crucial since PRS accuracy significantly improves with ancestry-matched discovery cohorts.

Besides the limited African specific GWAS available, PRSmix+ considers genetically correlated traits. The specific genetic architecture of complex traits can differ across populations due to varying allele frequencies and population-specific causal variants. If large-effect, ancestry-enriched variants important for a trait in African populations are not identified in primarily European-based GWAS, the PRSs and consequently PRSmix will miss these vital genetic signals, hence its likely poor performance (Lappalainen et al., 2024).

References

- Dudbridge, F. (2013). Power and Predictive Accuracy of Polygenic Risk Scores. *PLoS Genetics*, 9(3). <https://doi.org/10.1371/journal.pgen.1003348>
- Han, S., Gan, D., Wang, G., Ru, Y., Huang, C., Lin, J., Zhang, L., Meng, Z., & Zhu, S. (2018). Associations of Platelet Indices with Body Fat Mass and Fat Distribution. *Obesity*, 26(10), 1637–1643. <https://doi.org/10.1002/oby.22294>

- Fatumo, S., Chikowore, T., Choudhury, A., Ayub, M., Martin, A. R., & Kuchenbaecker, K. (2022). Diversity in Genomics Studies: A roadmap to increase diversity in genomic studies. *Nature Medicine*, 28(2), 243–250. <https://doi.org/10.1038/s41591-021-01672-4>.
- Lappalainen, T., Li, Y. I., Ramachandran, S., & Gusev, A. (2024). Genetic and molecular architecture of complex traits. *Cell*, 187(5), 1059–1075. <https://doi.org/10.1016/j.cell.2024.01.023>
- Fatumo, S., Sathan, Chaimae Samtal, Itunuoluwa Isewon, Tsaone Tamuhla, Chisom Soremekun, Jafali, J., Sumir Panji, Tiffin, N., & Yasmina Jaufeerally Fakim (2023). Polygenic risk scores for disease risk prediction in Africa: current challenges and future directions. *Genome Medicine*, 15(1). <https://doi.org/10.1186/s13073-023-01245-9>
- Sanghera, K. D., Bejar, C., Sharma, S., Gupta, R., & Blackett, R. P. (2019). Obesity genetics and cardiometabolic health: potential for risk prediction. *Physiology & Behavior*, 21(5), 1088–1100. <https://doi.org/10.1111/dom.13641>
- Truong, B., Hull, L. E., Ruan, Y., Huang, Q. Q., Hornsby, W., Martin, H., van Heel, D. A., Wang, Y., Martin, A. R., Lee, S. H., & Natarajan, P. (2024). Integrative polygenic risk score improves the prediction accuracy of complex traits and diseases. *Cell Genomics*, 4(4), 1–27. <https://doi.org/10.1016/j.xgen.2024.100523>
- Zhang, X., Xiao, J., Jiang, M., Phillips, C. J. C., & Shi, B. (2025). Thermogenesis and Energy Metabolism in Brown Adipose Tissue in Animals Experiencing Cold Stress. *International Journal of Molecular Sciences*, 26(7), 1–23. <https://doi.org/10.3390/ijms26073233>
- Ziadlou, R., Pandian, G. N., Hafner, J., Akdis, C. A., Stingl, G., Maverakis, E., & Brüggen, M. C. (2024). Subcutaneous adipose tissue: Implications in dermatological diseases and beyond. *Allergy: European Journal of Allergy and Clinical Immunology*, February, 3310–3325. <https://doi.org/10.1111/all.16295>