

Cell Genomics, Volume 4

Supplemental information

**Integrative polygenic risk score improves
the prediction accuracy of complex traits
and diseases**

Buu Truong, Leland E. Hull, Yunfeng Ruan, Qin Qin Huang, Whitney Hornsby, Hilary Martin, David A. van Heel, Ying Wang, Alicia R. Martin, S. Hong Lee, and Pradeep Natarajan

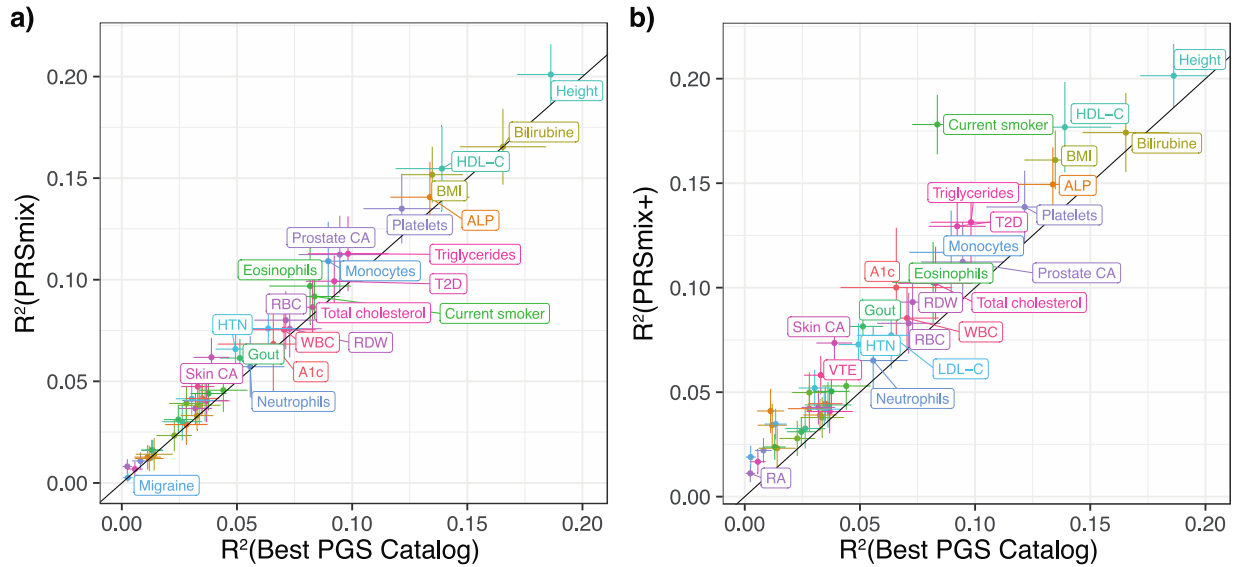
Supplementary Information

**Integrative polygenic risk score improves prediction accuracy of
complex traits and diseases**

Buu Truong et al.

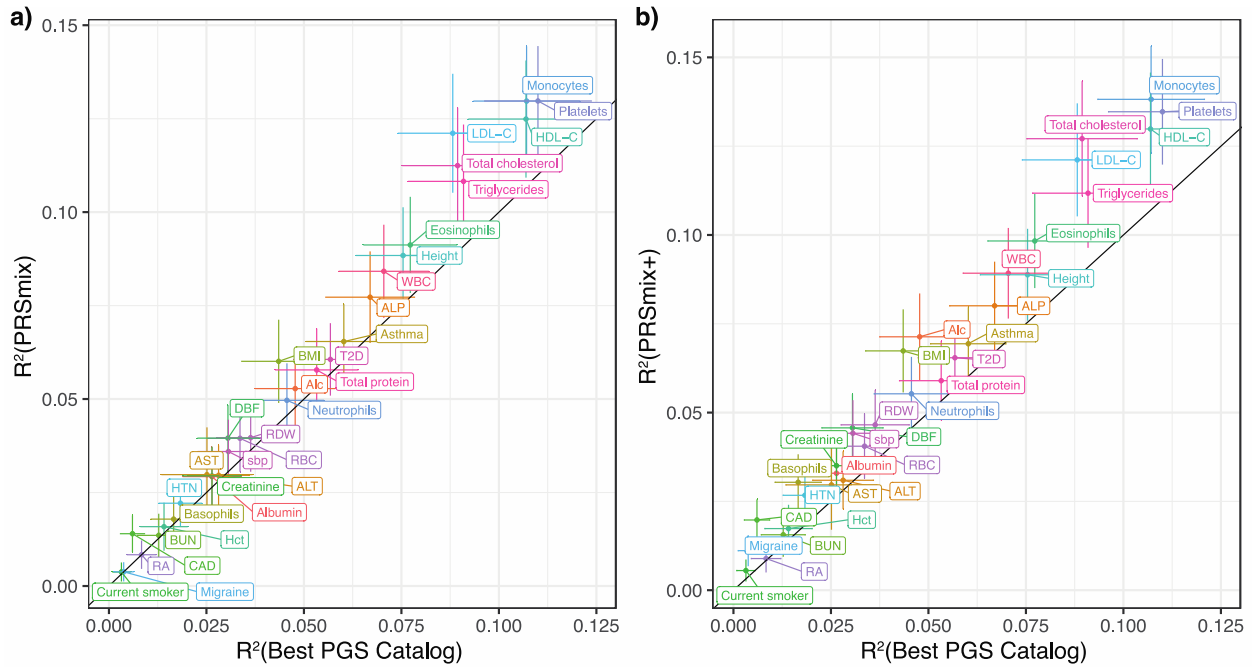
Supplementary Figure 1. Prediction accuracy of the best PGS catalog versus PRSmix and PRSmix+ for 47 traits in the European ancestry, related to Figure 3.

The prediction accuracies of the best PGS Catalog are displayed on the x-axis versus PRSmix (a) and PRSmix+ (b) on the y-axis. The prediction accuracies are estimated as incremental R2 and liability R2 for continuous traits and binary traits, respectively. The points represent the mean prediction and the whiskers demonstrate 95% confidence intervals. The black line represents the reference that the prediction accuracy of PGS Catalog is equal to PRSmix and PRSmix+.



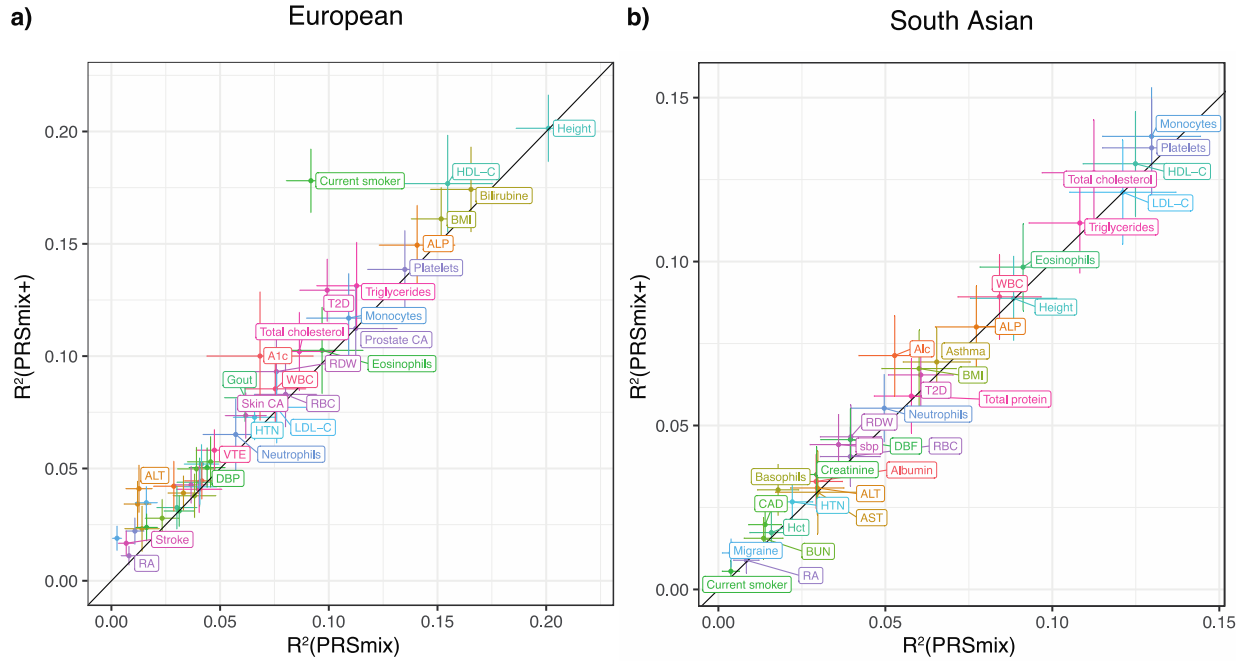
Supplementary Figure 2. Prediction accuracy of the best PGS catalog versus PRSmix and PRSmix+ for 32 traits in the South Asian ancestry, related to Figure 3.

The prediction accuracies of the best PGS Catalog are displayed on the x-axis versus PRSmix (a) and PRSmix+ (b) on the y-axis. The prediction accuracies are estimated as incremental R2 and liability R2 for continuous traits and binary traits, respectively. The points represent the mean prediction and the whiskers demonstrate 95% confidence intervals. The black line represents the reference that the prediction accuracy of PGS Catalog is equal to PRSmix and PRSmix+.



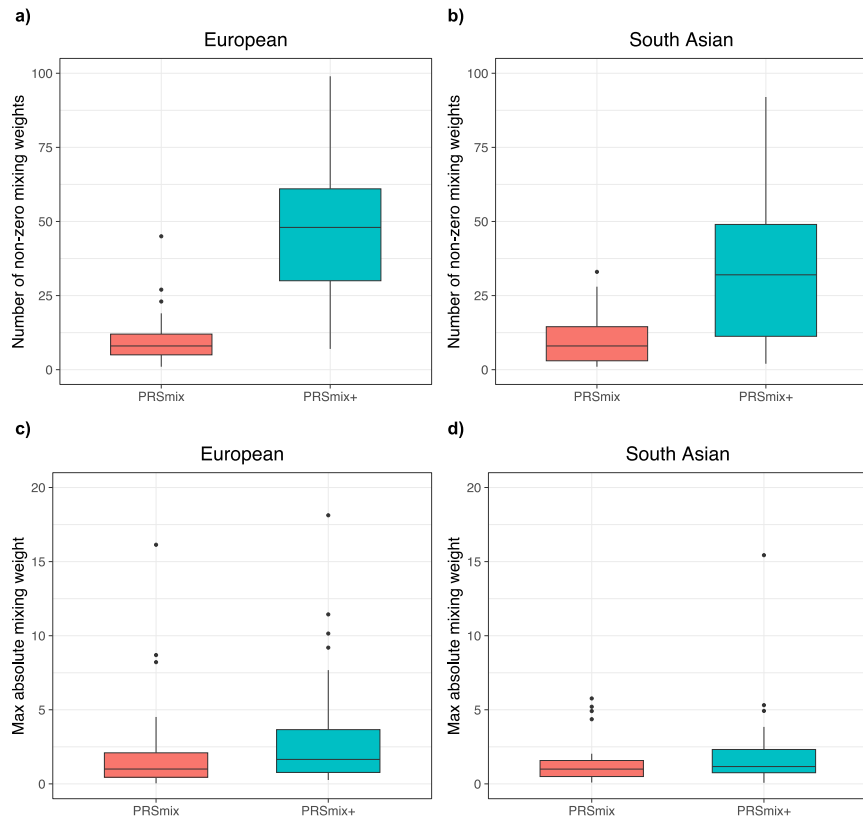
Supplementary Figure 3. Prediction accuracy of the PRSmix versus PRSmix+ in the European and South Asian ancestries, related to Figure 3.

The prediction accuracies of the PRSmix are displayed on the x-axis versus PRSmix+ on the y-axis in **a)** European ancestry and **b)** South Asian ancestry. The prediction accuracies are estimated as incremental R2 and liability R2 for continuous traits and binary traits, respectively. The points represent the mean prediction and the whiskers demonstrate 95% confidence intervals. The black line represents the reference that the prediction accuracy of PGS Catalog is equal to PRSmix and PRSmix+.



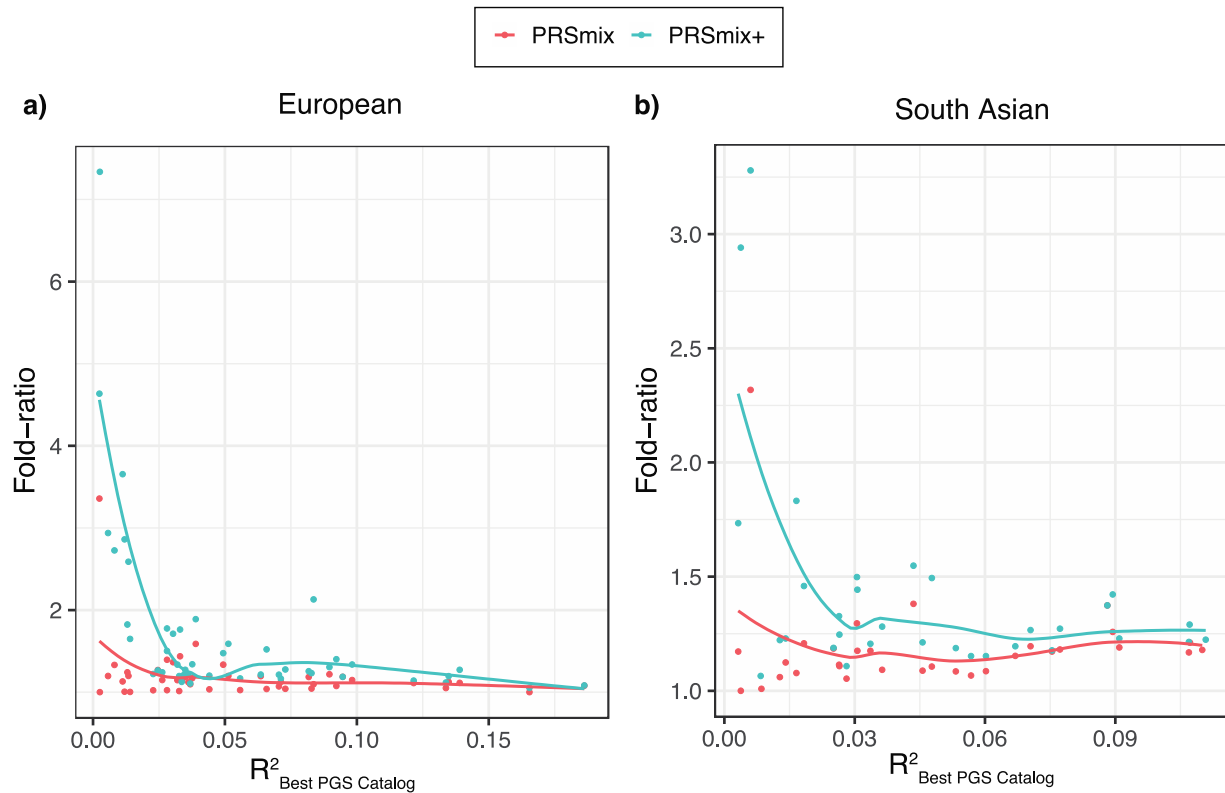
Supplementary Figure 4. The number of non-zero mixing weight and the max absolute mixing weights for PRSmix and PRSmix+ in European and South Asian ancestries, related to Figure 3.

The number of non-zero mixing weights are displayed for 47 traits in European ancestry (a) and 32 traits in South Asian ancestry (b). The max absolute mixing weights are showed for European ancestry (c) and South Asian ancestry (d). The boxplots show the first to the third quartile of prediction accuracies for 47 traits and 32 traits in European and South Asian ancestries, respectively. The whiskers reflect the maximum and minimum values within the $1.5 \times$ interquartile range for each group



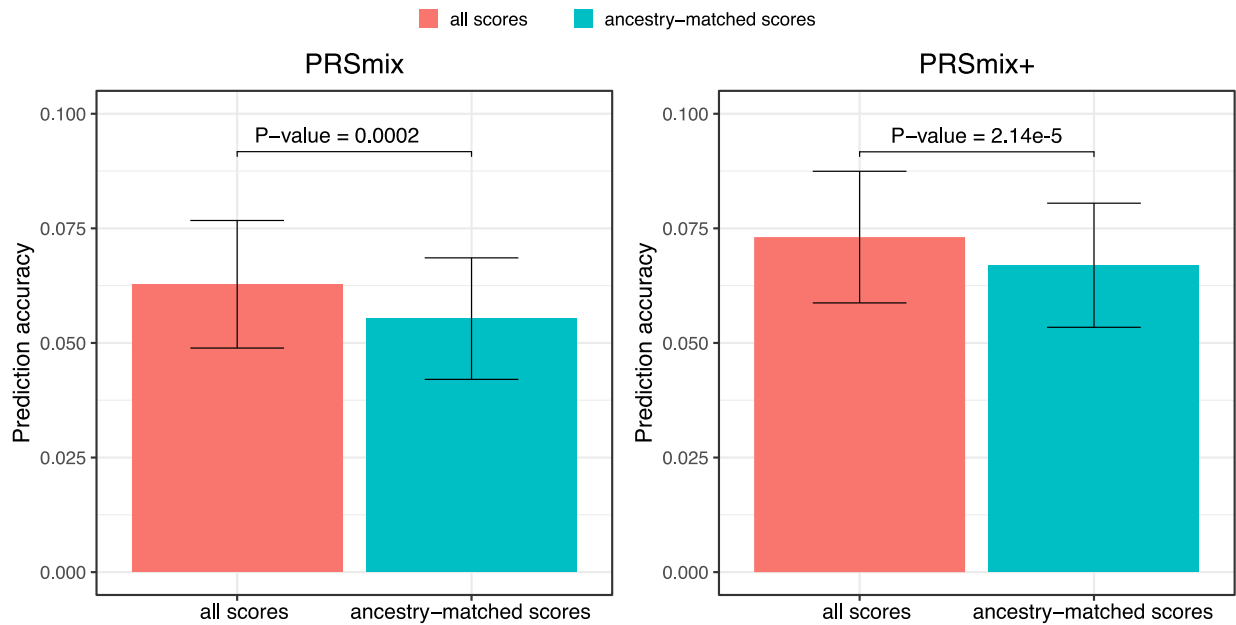
Supplementary Figure 5. Predictive improvement of PRSmix and PRSmix+ in European and South Asian ancestries, related to Figure 3.

The prediction accuracies of the best PGS Catalog are displayed on the x-axis versus ratio of R^2_{PRSmix} and $R^2_{\text{PRSmix+}}$ on the y-axis. The prediction accuracies are estimated as incremental R^2 and liability R^2 for continuous traits and binary traits, respectively. The red and blue line represents the local regression line for improvement of PRSmix and PRSmix+ over the best PGS Catalog, respectively.



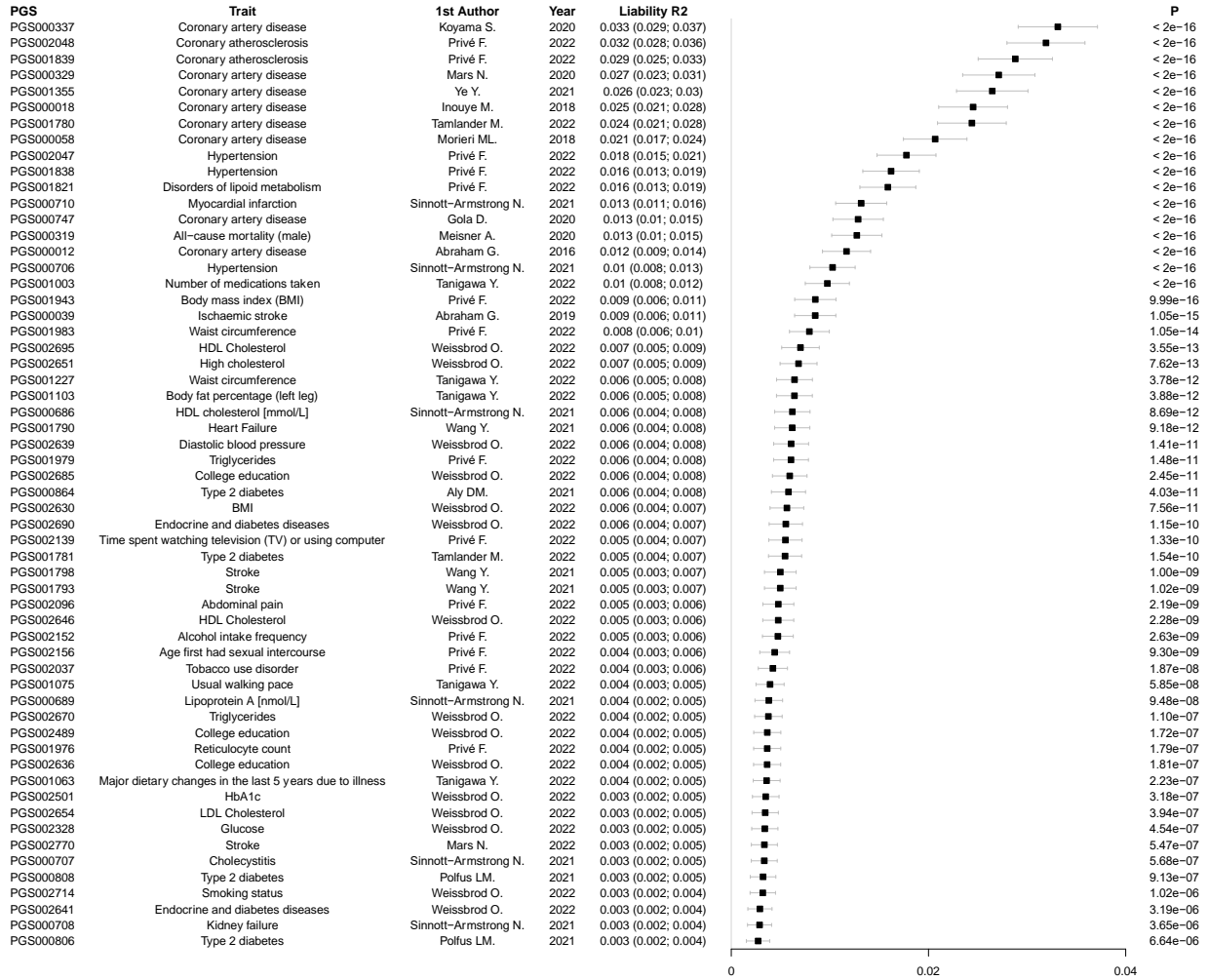
Supplementary Figure 6. Prediction accuracy of PRSmix and PRSmix+ in European using all scores or ancestry-matched scores to the target population, related to Figure 3.

The prediction accuracies using all scores and ancestry-matched scores are displayed on the x-axis. The ancestry-matched scores were chosen by scores with percentage of ancestry 100% matched with the target ancestry. Due to the limited number of 100%-South Asian-scores in the PGS Catalog, we only conducted the analyses with European ancestry. The prediction accuracies are estimated as incremental R2 and liability R2 for continuous traits and binary traits, respectively. The bars represent the mean prediction accuracy across 47 traits and the whiskers indicate 95% confidence interval.



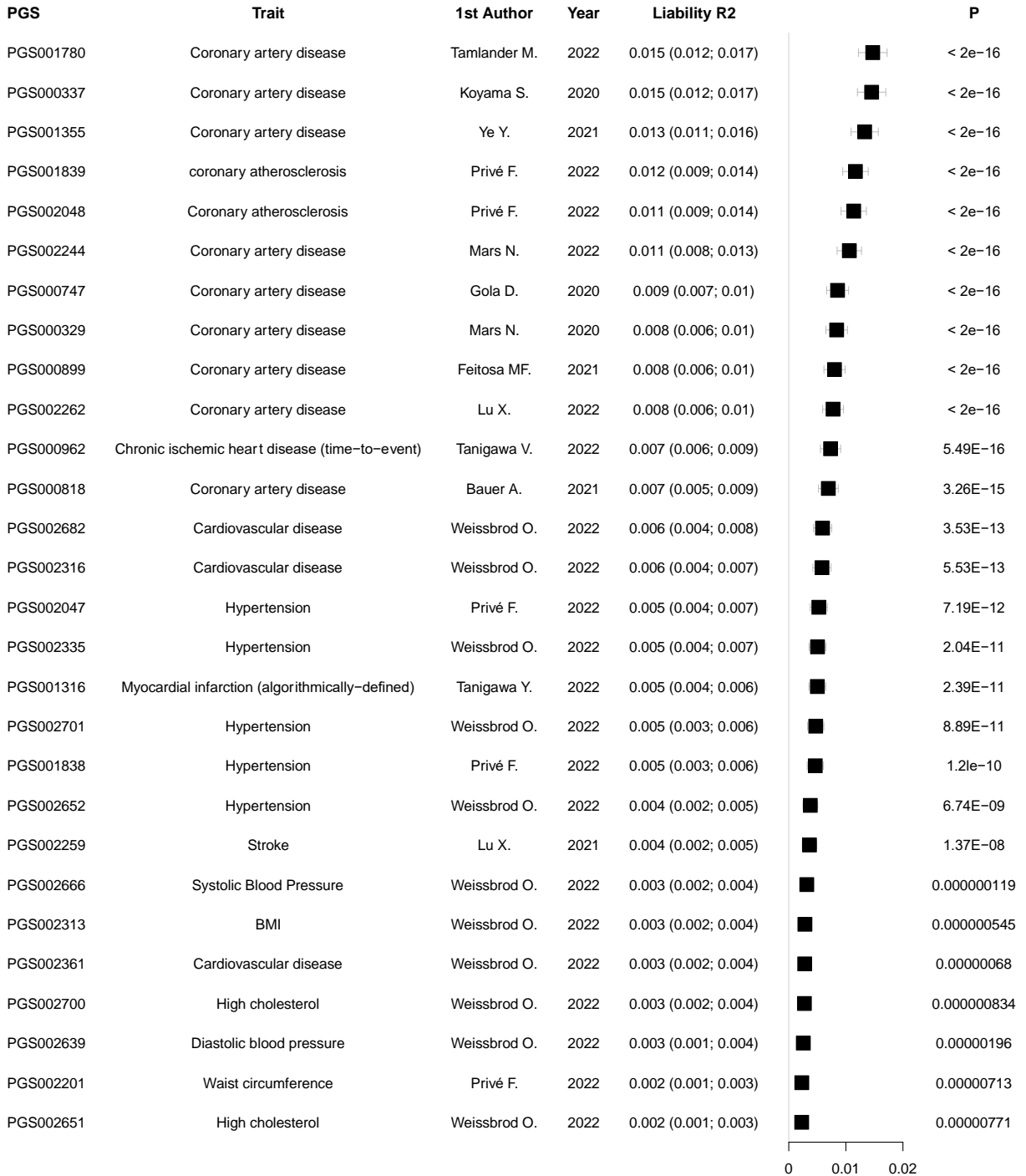
Supplementary Figure 7. Prediction accuracy in the training set of the contributing PGSs to PRSmix+ of coronary artery disease in the European ancestry, related to Figure 6.

We used Elastic Net to estimate the mixing weights of multiple PGS and report the contributing PGS with non-zero effects. Prediction accuracy was estimated with liability R². The boxes represent the mean prediction accuracy across the traits in that group and the whiskers demonstrate 95% confidence intervals.



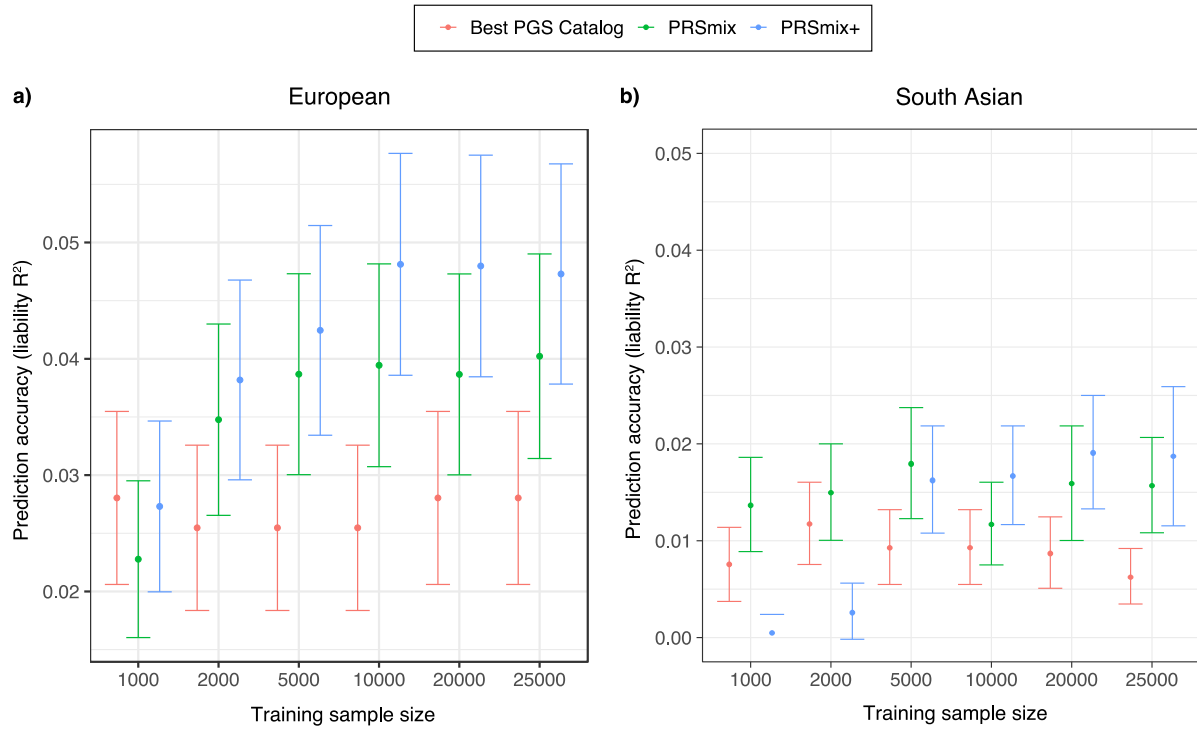
Supplementary Figure 8. Prediction accuracy in the training set of the contributing PGSs to PRSmix+ of coronary artery disease in the South Asian ancestry, related to Figure 6.

We used Elastic Net to estimate the mixing weights of multiple PGS and report the contributing PGS with non-zero effects. Prediction accuracy was estimated with liability R2. The boxes represent the mean prediction accuracy across the traits in that group and the whiskers demonstrate 95% confidence intervals.



Supplementary Figure 9. Empirical prediction accuracy with various training sample sizes for linear combination for coronary artery disease in the European and South Asian ancestries, related to Figure 6 and 7.

We assessed the performance of PRSmix and PRSmix+ with various training sample sizes. The prediction accuracy was estimated as liability R². The points represent the mean prediction accuracy and the whiskers demonstrate 95% confidence intervals.



Supplementary Figure 10. Contributing traits to PRSmix+ of stroke in the European ancestry, related to Figure 6 and 7.

We used Elastic Net to estimate the mixing weights of multiple PGS and report the top 80 contributing PGS with non-zero effects. Prediction accuracy was estimated with liability R2. The boxes represent the mean prediction accuracy across the traits in that group and the whiskers demonstrate 95% confidence intervals.

