# Development and Validation of Multi-Ancestry Polygenic Risk Scores for Coronary Artery Disease

**George B. Busby[1,*], Alessandro Bolli[1], Jen Kintzle[1], Paolo Di Domenico[1], Giordano Bottà[1,*]**

[1] Allelica Inc, 447 Broadway, New York, 10013, USA

## ABSTRACT

A wide range of Polygenic Risk Scores (PRS) have been developed for Coronary Artery Disease (CAD) but these remain to be benchmarked in populations with different genetic ancestries. At the same time, several novel methods have been developed to leverage ancestry-specific genome-wide association study (GWAS) summary statistics and linkage disequilibrium (LD) maps to better characterize the association between genetic variants and disease risk across populations.

In this report we outline how Allelica's *DISCOVER* software was used to develop, validate and calibrate multiple ancestry-specific PRSs for CAD. *DISCOVER* implements modified versions of multiple PRS development algorithms to identify the best performing panel for specific populations. We show that by incorporating summary statistics from ancestry-specific GWAS PRS can be developed with improved performance relative to scores developed on single GWAS generated using individuals with largely European ancestry. Importantly, across datasets we are able to identify at least one in ten individuals at greater than 2 fold increased risk of CAD compared to the rest of the population. Our analyses provide a foundation for the application of PRS as a Risk Enhancing Factor in cardiovascular disease risk assessment by identifying individuals at high genetic risk of CAD, regardless of their genetic ancestry.

**KEY WORDS** Polygenic Risk Scores; Coronary Artery Disease; ancestry; genomics

## I. INTRODUCTION

Polygenic Risk Scores, or PRS, have shown promise as tools to identify significant proportions of the population at high genetic risk of disease (*1–4*). For example, Khera and colleagues showed that individuals in the top 8 percentiles of the distribution of a PRS for Coronary Artery Disease (CAD) had a three fold increase in risk of disease compared to the remainder of the distribution, a risk equivalent to carriers of rare mutations in Familial Hypercholesterolemia (FH) associated genes. (*1*). However, carriers of FH mutations are rare at the population level (~0.4% of the population), meaning that PRS can identify roughly 20 times more people at high genetic risk than traditional approaches based on gene panels.

"*" corresponding authors,
george@allelica.com
giordano@allelica.com

Published: 25th July 2022

Recent work has also highlighted the potential of PRS to be used to aid clinical management of cardiovascular disease (CVD). Current approaches to identify individuals at risk of disease are based on applying algorithms such as the Pooled Cohort Equations (PCE) to clinical risk factors to calculate an individual's 10 year risk of disease (*5*). Aragam and colleagues (*3*) demonstrated that a significant proportion of individuals at heightened risk of disease can be identified by PRS and, crucially, that methods such as the PCE are unable to capture these at risk individuals. Clinical guidelines for CVD management allow for so-called Risk Enhancing Factors to be used to reclassify individuals into higher risk strata despite intermediate or borderline clinical risk. Given their ability to identify individuals who are at least two times the risk of the population, and their demonstrated cost-effectiveness (*6*), it is increasingly clear that PRS satisfy the criteria to be considered a Risk Enhancing Factor for CVD.

Efforts are now underway to translate these research findings into clinical practice. In order to do so however, it is essential that PRS are able to be applied to individuals of all ethnicities and ancestries. This has so far been challenging because the majority of genetic and clinical data available to build PRS come from individuals of western European ancestry (*7*, *8*). When these scores are applied to cohorts containing non-European ancestry individuals, their performance is attenuated (*9*). Whilst this is expected due to differences in Linkage Disequilibrium (LD) and allele frequencies in different populations as a result of population separation since our ancestors left African 100-200,000 years ago, the non-transferability of PRS across different populations needs to be addressed before their potential as clinical tools can be realised.

Approaches to improve the utility of PRS in different populations fall into two main camps. The first is to build a PRS using GWAS summary statistics and datasets (both validation and testing) from the Target ancestry of interest. We will refer to this as the single GWAS approach. Although PRS can be optimised in different populations by utilising ancestry-specific datasets throughout the pipeline, the main drawback to single GWAS approaches is their requirement for well powered, large and deep GWAS summary statistics from the Target population. As discussed above, these rarely exist in practice for non-European ancestry datasets.

Novel methodology is available that can utilise multiple sets of GWAS summary statistics in PRS development (*10*–*12*). These multiple GWAS approaches can leverage several ancestry specific GWAS and combine them with ancestry-specific LD maps (*13*). Crucially, when combining these smaller ancestry-specific GWAS with larger, well powered (typically European) GWAS, such approaches have the potential to better estimate both the variants causal to disease and their effect sizes (*14*), making them particularly appropriate to researchers interested in improving the performance of scores across different ancestry groups.

In this short report, we outline analyses to develop PRS for CAD across different ancestries using this second multiple GWAS approach. The aim of this work is to provide robust evidence of the performance of PRS in CAD across ancestries that can be used to build a foundation for their use in clinical tools for the prevention of CVD.

# II. METHODS

## A. Association Data

### II.A.1 GWAS

We identified three large GWAS from the literature that did not include any individuals in the downstream Validation and Testing datasets that we used to develop multi-ancestry PRS. These were the CardioGRAMplusC4D GWAS (*15*), the Japanese 161k and the Japanese 52k studies (*16*, *17*).

In the absence of publicly available South Asian and African specific GWAS summary statistics, we performed GWAS on separate subsets of individuals from both ancestries in the UK Biobank using *fastGWA* (*19*). The South Asian GWAS comprised 932 cases and 4,043 controls and the African GWAS comprised 91 cases and 3,230 controls. *fastGWA* was used due to the computational efficiency of the generalized linear mixed model (GLMM)-based method that it implements, which has improved statistical properties when applied to binary traits. In addition we also performed a secondary GWAS on a subset of individuals from the UK Biobank with European ancestry (26,135 cases and 433,131 controls).

**Table 1.** Datasets used in this study. References starting with phs refer to datasets downloaded from DBGaP.

| Dataset/Cohort | Ancestry | Cases | Controls | Reference |
|---|---|---|---|---|
| ***Genome Wide Association Studies*** | | | | |
| CardioGRAMplusC4D | European,South Asian | 60,801 | 123,504 | Nikpay et al 2015 (*15*) |
| Japanese 161k | East Asian | 14,992 | 146,214 | Sakaue et al 2021 (*16*) |
| Japanese 52k | East Asian | 15,302 | 36,140 | Matsunaga et al 2020 (*17*) |
| ***Genomics and Clinical Datasets*** | | | | |
| MultiEthnic Study of Atherosclerosis (MESA) | African American | 76 | 1,247 | phs000209.v13.p3 |
| | Hispanic | 79 | 1,155 | |
| | Chinese | 30 | 615 | |
| | European | 162 | 1,950 | |
| UK Biobank | African | 227 | 2,329 | Bycroft et al 2018 (*18*) |
| | South Asian | 2,004 | 8,482 | |
| | East Asian | 78 | 1,425 | |
| | European | 26,135 | 433,131 | |

### II.A.2 Finemapping

GWAS provide an assessment of the size and significance of the association of alleles and disease at millions of sites across the genome. Because not all potentially causal alleles will have been genotyped or imputed in the original GWAS, the associated variants from GWAS may only tag those causing disease rather than be the actual disease causing variants themselves. To identify putatively causal variants, we applied a finemapping approach, implemented by *POLYFUN* (*12*, *20*) (*15*). This method uses ancestry-specific LD scores and additional data on the putative function of variants across the genome to generate a set of finemapped summary statistics. We used the CardioGRAMplusC4D summary statistics and functional information from

**Table 2.** Labels for the GWAS used in this study. Seven main GWAS were used in this study but were variously filtered based on Minor Allele Frequency (MAF) in relevant the Thousand Genomes Project superpopulation and/or for the top 2 million variants by *P*-value

| GWAS Name | Source/Ancestry | No. Inds | Filters |
|---|---|---|---|
| EUR-1 | CardioGRAMplusCD4 | 184,305 | MAF<0.01 |
| EUR-1a | CardioGRAMplusCD4 | 184,305 | MAF<0.01; top 2M |
| EUR-1b | CardioGRAMplusCD4 | 184,305 | MAF<0.00001 |
| EUR-2 | EUR-1, finemapped | 184,305 | MAF<0.01 |
| EUR-2a | EUR-1, finemapped | 184,305 | MAF<0.01; top 2M |
| EUR-2b | EUR-1, finemapped | 184,305 | MAF<0.00001 |
| EUR-3 | UK Biobank European | 459,266 | MAF<0.01 |
| EUR-3a | UK Biobank European | 459,266 | MAF<0.01; top 2M |
| EUR-3b | UK Biobank European | 459,266 | MAF<0.00001 |
| EUR-3c | UK Biobank European | 459,266 | MAF<0.01† |
| AFR-1 | UK Biobank African | 3,321 | MAF<0.01 |
| EAS-1 | Biobank Japan | 161,206 | MAF<0.01 |
| EAS-2 | OACIS (Japan) | 52,442 | MAF<0.01 |
| SAS-1 | UK Biobank South Asian | 4,966 | MAF<0.01 |

†*The MAF in the summary statistics, rather than the TGP European superpopulation, was used to filter these variants.*

**Table 3.** Datasets used to validate and test ancestry specific PRS. Where possible, independent datasets were used as Validation and Testing datasets. If only one dataset was available, this was split to generate approximately equal sized Validation and Testing datasets. An initial analysis using a limited set of South Asians identified a best performing score that did not include the GWAS from the South Asian ancestry individuals (SAS-1) from the UK Biobank. We therefore used all South Asian ancestry individuals from the UK Biobank as Validation and Testing datasets.

| Ancestry | Stage | Cases | Controls | Dataset / Cohort |
|---|---|---|---|---|
| African American | Validation | 372 | 3,924 | UK Biobank |
| | Testing | 76 | 1,247 | MESA |
| Hispanic/Latino | Validation | 38 | 579 | MESA |
| | Testing | 41 | 576 | MESA |
| East Asian | Validation | 62 | 1,222 | UK Biobank |
| | Testing | 30 | 615 | MESA |
| South Asian | Validation | 725 | 3,093 | UK Biobank |
| | Testing | 726 | 3,094 | UK Biobank |
| European | Validation | 81 | 975 | MESA |
| | Testing | 81 | 975 | MESA |

**Table 4.** Details of the three external PRS that were used for benchmarking the multi-ancestry PRS.

| Authors | PRS Method | Num. variants | Reference |
|---|---|---|---|
| Bolli et al (2021) | SCT/metaPRS | 1,926,521 | (*4*) |
| Khera et al (2018) | LDPred | 6,630,150 | (*1*) |
| Inouye et al (2021) | metaPRS | 1,745,179 | (*2*) |

Gazal et al 2018 (*21*). The combination of GWAS and finemapping datasets resulted in a final list of 7 different sets of summary statistics for downstream analyses (Table 2).

### II.A.3 Variant filtering

We applied a range of filters to the summary statistics before running the main PRS development pipeline (Table 2). For all non-European GWAS we removed variants with a Minor Allele Frequency (MAF) of <0.01 in the One Thousand Genomes Project (TGP) (*22*) superpopulation matching the relevant GWAS Discovery population. For the European GWAS, we generated additional filtered datasets that contained the top 2 million most significant results, as well as a filtered set that filtered only the rarest variants (Table 2).

## B. Datasets

We compiled and harmonised a joint dataset of individuals from publicly available datasets (Tables 1 and 3). These comprised the MultiEthnic Study of Atherosclerosis (MESA), the Institute of Personalised Medicine (IPM) subset of the eMerge study, and the UK Biobank (*18*).

## C. DISCOVER

We used a modified version of *PRS-CSx* (*11*, *23*) within Allelica's *DISCOVER* software to build multiple PRS with different combinations of GWAS summary statistics. We ran *PRS-CSx* with a range of values for the global shrinkage parameter $\phi$. If $\phi$ is not specified, it is learnt from the data using a fully Bayesian approach. We performed a small-scale grid search (e.g., $\phi$=1e-6, 1e-4, 1e-2, 1) to find the optimal $\phi$ value in the Validation datasets.

## D. Combining GWAS-specific PRS panels

To build ancestry-specific PRS, we explored a range of different combinations of between 2 and 4 sets of the ancestry-specific GWAS summary statistics outlined above. The output of *PRS-CSx* is

**Table 5.** The best performing multi-ancestry PRS in 5 different genetic ancestries. For each ancestry, we show the source of the GWASs used to build the score, the number of variants in the score, the Odds Ratio per Standard Deviation (ORxSTD) of the score in the Testing dataset, the proportion of the population at two fold risk risk (2X%) and the 95% CI around the Odds Ratio for the proportion at 2X risk.

| Allelica PRS Name | GWAS Source | Ancestry | Num Variants | ORxSTD (95% CI) | 2X (%) | 2X (95% CI) |
|---|---|---|---|---|---|---|
| Allelica_CAD_HIS_2022 | AFR-1;EUR-2b | Hispanic | 202,975 | 1.59 (1.25-1.87) | 11 | (1.95-4.03) |
| Allelica_CAD_AFR_2022 | AFR-1;EUR-3c | African American | 260,855 | 1.37 (1.05-1.79) | 16 | (1.27-3.92) |
| Allelica_CAD_SAS_2022 | EAS-2;EUR-1 EAS-2;EUR-1 | South Asian | 123,622 | 1.45 (1.33-1.59) | 17 | (1.45-2.47) |
| Allelica_CAD_EAS_2022 | EAS-1;EUR-3 | East Asian | 260,536 | 1.60 (1.12-2.29) | 18 | (1.09-5.66) |
| Allelica_CAD_EUR_2022 | EUR-1;EUR-2 EUR-3 | European | 285,231 | 1.57 (1.29-2.00) | 25 | (1.70-3.30) |

a separate PRS panel for each GWAS summary statistic dataset used. To combine these into a single panel for downstream validation and testing, we combined the constituent variant effect sizes across panels using an inverse-variance-weighted meta-analysis of the panel specific posterior effect sizes.

Each resulting meta-PRS is labelled with the GWAS datasets used to generate it (Table 2). For example, the PRS labeled **EUR-1_AFR-1** is the result of running *PRS-CSx* with the EUR-1 (CardioGRAMplusCD4) and AFR-1 (UK Biobank African) GWAS and combining the PRS using the method outlined above.

## E. PRS Validation and Testing

We used separate independent Validation and Testing datasets to assess the performance of the PRS across different genetic ancestries (Table 3). For each PRS, we assessed the association with CAD in ancestry-specific datasets by computing the Odds Ratio per Standard Deviation (ORxSTD) from a logistic regression using phenotype as the dependent variable and standardised PRS, age, sex, family history, and the first four principal components of variation as independent variables.

The best performing PRS in each ancestry-specific Validation dataset was then applied to a separate ancestry-specific Testing population to compute an independent assessment of its performance. We used logistic regression to identify the proportion of each ancestry group at two fold risk, adjusting for age, sex and 4 principal components of ancestry (*3*). PRS were benchmarked in the Testing population using the external datasets outlined in Table 4.

To further assess the performance of the PRSs in the Testing datasets, we additionally computed Brier scores (*24*). Similar to the mean squared error, Brier scores are a test of predictive model calibration and comparisons of two Brier scores is achieved by assessing the Brier skill score. This statistic will be positive when the model to be tested is better calibrated than the reference model. We tested the calibration of the five ancestry specific scores against each of the three external PRSs.

# III. RESULTS AND DISCUSSION

We developed 108 PRS for CAD using 21 different combinations of ancestry-specific GWAS with four different values of the $\phi$ shrinkage parameter. These were applied to five different ancestry-specific Validation datasets (Table 3) to identify the best performing score in each ancestry. The results are shown in Figure 1. The best performing scores in the Validation dataset were applied to independent Testing datasets to define a final assessment of their predictive performance. The new ancestry-specific scores were benchmarked against three previously published scores and, with the exception of South Asians, showed an increased performance across ancestries (Table 4 and Figure 1). Area Under the Receiver Operator Curve (AUCs) for the PRSs are shown in Figure 2 and Table 7.

The predictive performance (ORxSTD) of the best scores ranged from 1.37 (1.05-1.79) in African

**Table 6.** Brier scores for the best multi-ancestry scores and the three external PRS for comparison. A Brier score of 0 equates to perfectly calibrated model and 1 is perfectly uncalibrated. In each ancestry, the best multi-ancestry PRS (listed in the Best column and detailed in Table 5) has a smaller Brier score than the external PRS. Table 7 shows the results of comparisons of these scores.

| Ancestry | Best | Bolli et al (2021) | Inouye et al (2018) | Khera et al (2018) |
|---|---|---|---|---|
| Hispanic | 0.0565 | 0.0566 | 0.0566 | 0.0566 |
| African-American | 0.0523 | 0.0529 | 0.0528 | 0.0528 |
| South Asian | 0.1318 | 0.1326 | 0.1333 | 0.1333 |
| East Asian | 0.0420 | 0.0425 | 0.0423 | 0.0424 |
| European | 0.0669 | 0.0675 | 0.0671 | 0.0676 |

**Table 7.** Allelica's multi-ancestry PRS are better calibrated risk models than three published PRSs. For each multi-ancestry PRS, we show the Area Under the Receiver Operator Curve (AUC) for the PRS in a model adjusted for sex, the first four principal components of ancestry and family history, where available, together with and comparisons of Brier scores (Brier skills scores from Table 6), a calibration metric, between each ancestry specific score and three external datasets (Table 4). Positive values indicate that the test model, in this case the Allelica score, is better calibrated than the reference score.

| Allelica PRS Name | Ancestry | AUC (95% CI) | Bolli et al (2021) | Inouye et al (2018) | Khera et al (2018) |
|---|---|---|---|---|---|
| Allelica_CAD_HIS_2022 | Hispanic | 0.77 (0.72-0.81) | 0.171 | 0.075 | 0.14 |
| Allelica_CAD_AFR_2022 | African American | 0.69 (0.63-0.75) | 1.128 | 1.101 | 1.099 |
| Allelica_CAD_SAS_2022 | South Asian | 0.77 (0.74-0.79) | 0.586 | 1.11 | 1.15 |
| Allelica_CAD_EAS_2022 | East Asian | 0.78 (0.71-0.86) | 1.117 | 0.862 | 0.912 |
| Allelica_CAD_EUR_2022 | European | 0.71 (0.67-0.75) | 0.773 | 0.260 | 1.041 |

American ancestry dataset to 1.60 (1.12-2.29) in East Asians. Across ancestry groups the new multi-ancestry scores outperformed published scores (Table 5). The ORxSTD for previously published scores presented here may differ from those reported elsewhere. However, these comparisons are based on benchmarking the PRSs on Testing datasets comprising the same individuals with the same phenotype definition, and the relative performance represents a fair comparison within an ancestry.

Our analysis of calibration using Brier scores (Tables 6 and 7) demonstrated that all of the five multi-ancestry PRS were better calibrated than any of the external PRS. We hypothesise that this increase in calibration performance was due to the additional information from the ancestry-specific effect sizes used in the Discovery GWAS of the current approach.

We identified the proportion of each ancestry at two fold increased risk compared to the the remainder (Table 5). The analysis identified between 11% and 25% of individuals at two fold risk across the different ancestry groups. Following the approach of Aragam et al (*3*) we also show the 95% CI for the estimate of the Odds Ratio of 2. The 2X threshold represents that of well established risk factors such as family history and Mendelian-inherited genetic variants (*25*), as well other risk enhancing factors such as diabetes and ethnicity which are currently used to reclassify individuals in the PCE (*3*, *26*).

Adjusting PRS for subtle population structure is an important step in assessing performance (*27*, *28*). In Figure 3A we show the effect of ancestry adjustment on the relative risk distributions generated for the multi-ancestry scores. We note that in small, relatively genetically homogeneous popula-

tions the effect of the adjustment on the relative risk distributions is small. To visualise genetic diversity, we projected individuals from each analysis cohorts onto principal components from the Thousand Genomes Project (TGP, Figure 3B). Out of the cohorts used in this analysis, the Hispanic ancestry group showed the greatest diversity with individuals spread across PC1 and PC2. This is also shown by the effect of the ancestry adjustment, which reduces the stratification of risk in the Hispanic groups, potentially because of inflation due to population structure rather than true signal. In the absence of ancestry adjustment, the population structure present in this group could lead to inflation of the predictive performance. Throughout, including in the calibration analysis we used the ancestry-adjusted PRSs.

Finally, we note that we have used the point estimate for the ORxSTD in the Validation datasets to identify the best multi-ancestry score. Whilst other statistics are available (e.g. the AUCs shown in Figure 2), the ORxSTD has direct clinical relevance. Moreover, testing these scores in the independent Testing datasets gives an unbiased assessment of their performance against the external PRS. Whilst the confidence intervals overlap between scores, the results of the calibration analysis (Tables 6 and 7) demonstrate the potential of adding ancestry specific GWAS to build stronger ancestry-specific PRS.

## IV. CONCLUSION

Identifying individuals at high risk of disease is the first important step in the clinical management and prevention of common disease. Here we show that by using a CAD PRS we are able to identify significant proportions of the population at increased risk of disease. Importantly, these results have been validated and calibrated on multiple US populations and can be considered ready for use in clinical management of CVD.

## ACKNOWLEDGEMENTS

We thank all individuals who contributed data to the publicly available datasets utilised in this study. UK Biobank data was accessed under application ID 40692. This study was funded by Allelica Inc.

**Figure 1.** Validation of 21 PRS in multiple ancestries. For each of the 21 multi-ancestry PRS, we chose the best performing (highest ORxSTD) in each Validation population from each of the four replicate runs with different $\phi$. The best performing multi-ancestry PRS in each dataset was identified and then compared against three published PRS.

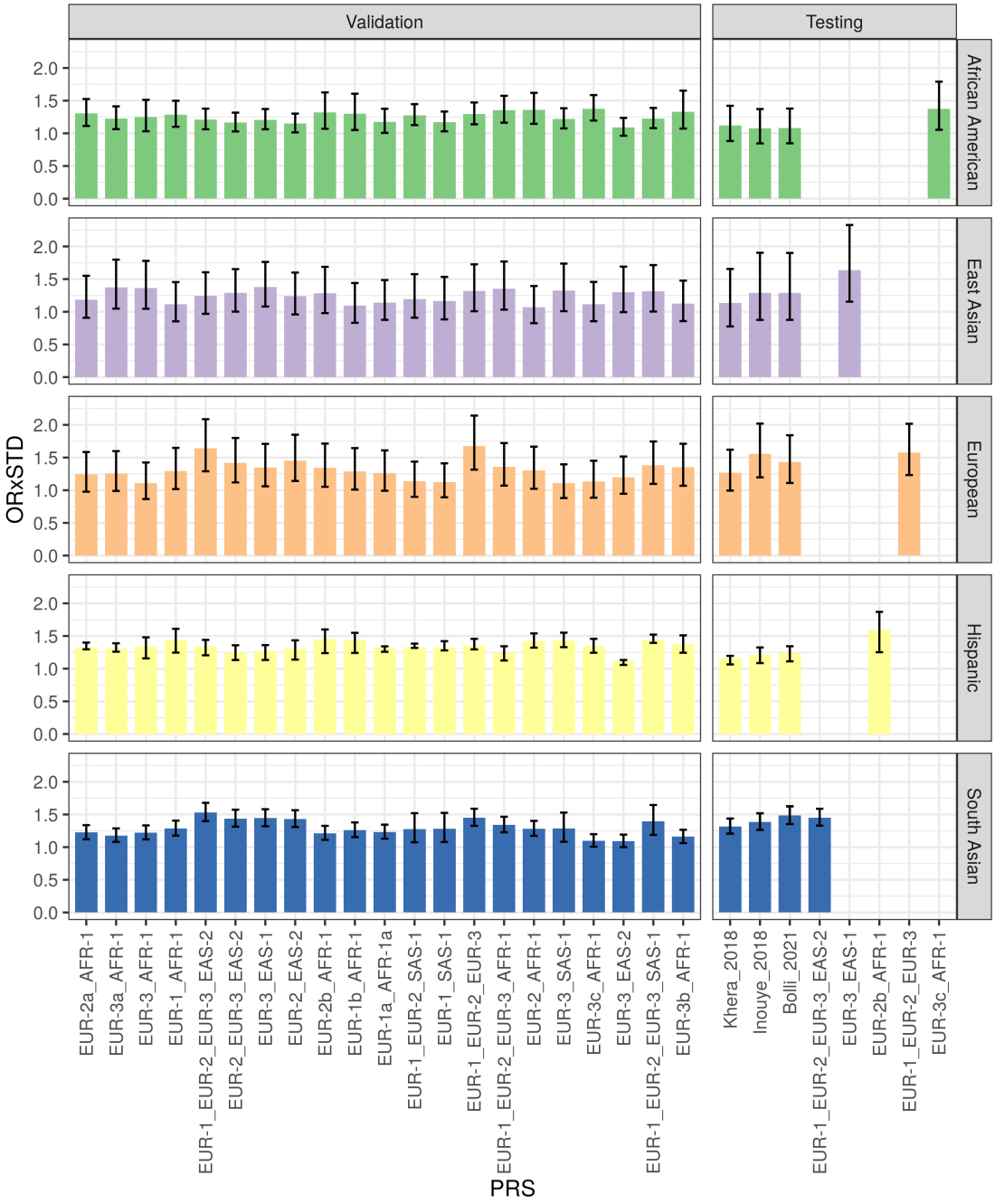


**Figure 2.** Validation of 21 PRS in multiple ancestries. For each of the 21 multi-ancestry PRS, we chose the best performing (highest ORxSTD) in each Validation population from each of the four replicate runs with different $\phi$. This figure shows the Area Under the Receiver Operator Curve (AUC) for the PRSs. AUC was computed from a model comprising PRS, four PCs, sex, age and family history of CAD, where available.

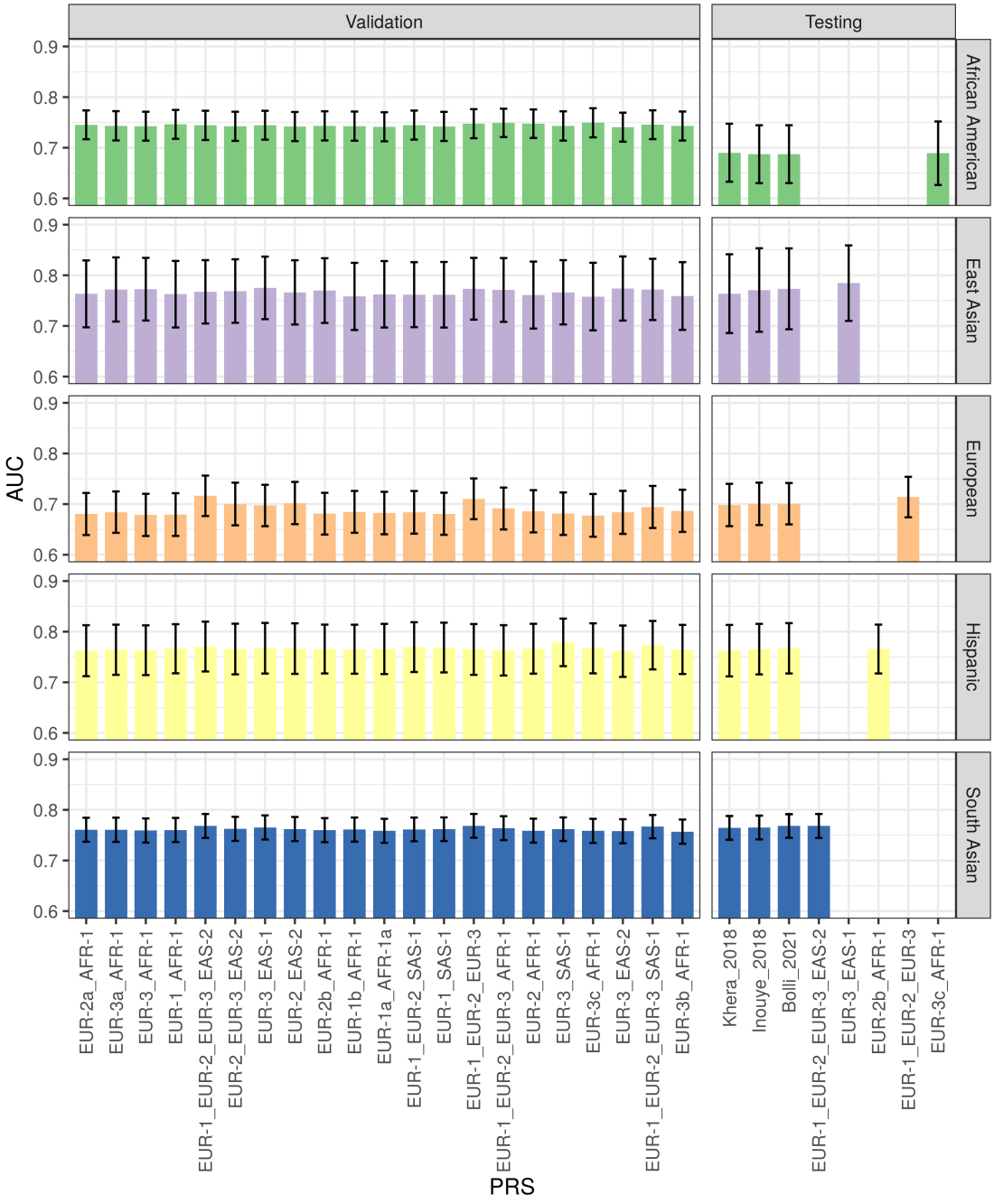
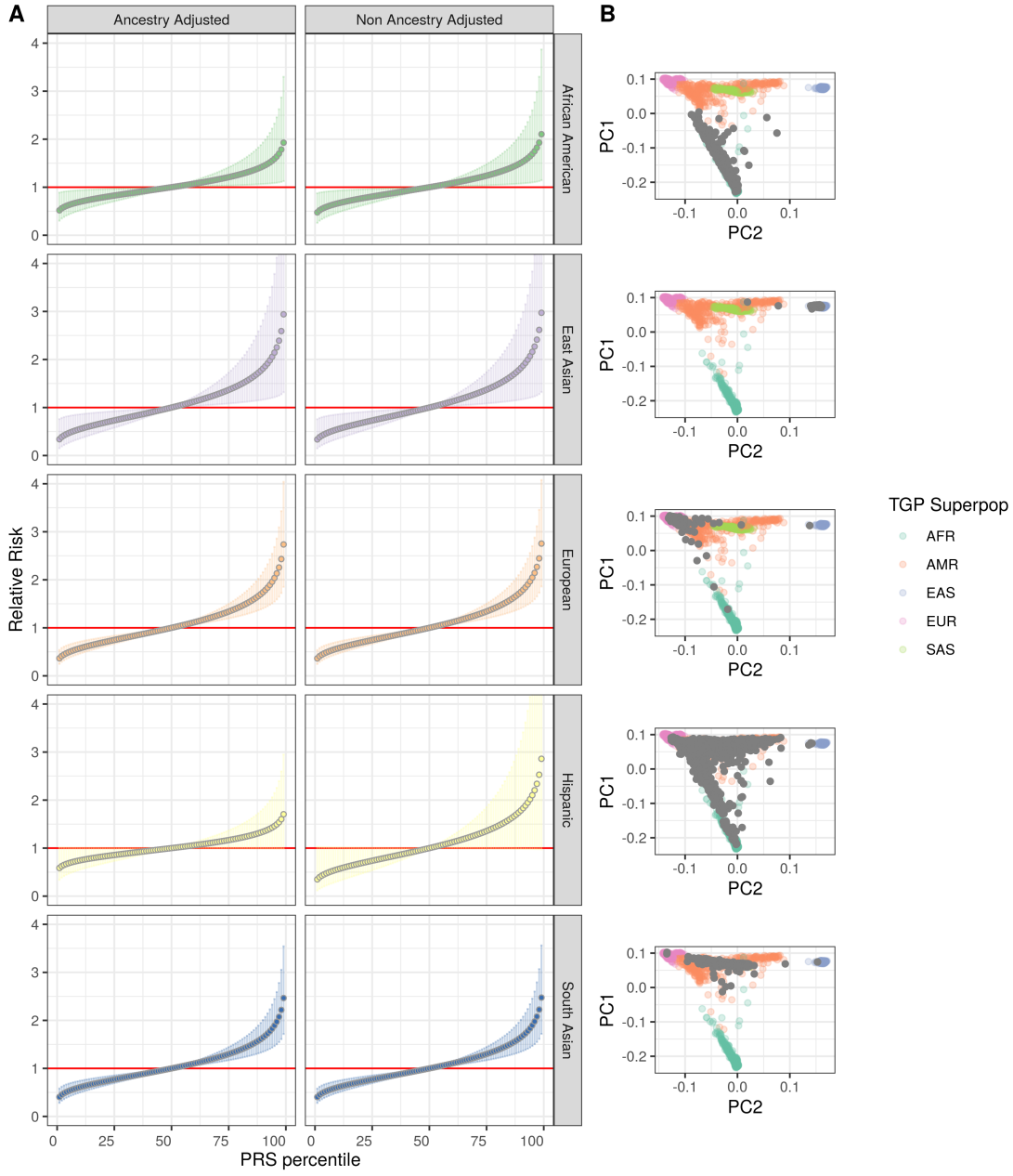


**Figure 3.** (A) Assessment of the relative risk conferred by each percentile of the PRS distribution using the final multi-ancestry PRSs. In each ancestry we show the distribution of relative risks for both ancestry adjusted and un-adjusted PRS. Ancestry adjustment was performed using principal components as outlined in (*29*). The horizontal red line denotes a relative risk of 1. (B) Visualisation of population structure in the ancestry specific cohorts. Individuals from the Testing datasets (in grey) projected on the first two principal components of ancestry from the Thousand Genomes Project (TGP). This analysis demonstrates the heterogeneity of self-reported ancestry and the diverse genetic ancestry of the Hispanic ancestry group.



## REFERENCES


1. A. V. Khera *et al.*, eng, *Nature Genetics* **50**, 1219–1224, ISSN: 1546-1718 (2018).
2. M. Inouye *et al.*, eng, *Journal of the American College of Cardiology* **72**, 1883–1893, ISSN: 1558-3597 (Oct. 2018).
3. Aragam Krishna G. *et al.*, *Journal of the American College of Cardiology* **75**, Publisher: American College of Cardiology Foundation, 2769–2780, (2021; https://www.jacc.org/doi/full/10.1016/j.jacc.2020.04.027) (June 2020).
4. Bolli Alessandro, Di Domenico Paolo, Pastorino Roberta, Busby George B., Bottà Giordano, *Circulation* **143**, Publisher: American Heart Association, (2021; https://www.ahajournals.org/doi/abs/10.1161/CIRCULATIONAHA.120.051843) (2021).
5. S. M. Grundy *et al.*, en, *Journal of the American College of Cardiology* **73**, 3168–3209, ISSN: 0735-1097, (2021; https://www.sciencedirect.com/science/article/pii/S0735109718390338) (June 2019).
6. D. Mujwara *et al.*, *Journal of the American Heart Association* **11**, Publisher: American Heart Association, e025236, (2022; https://www.ahajournals.org/doi/10.1161/JAHA.121.025236) (June 2022).
7. D. B. Goldstein, *The New England Journal of Medicine* (2009).
8. S. Fatumo *et al.*, en, *Nature Medicine* **28**, Number: 2 Publisher: Nature Publishing Group, 243–250, ISSN: 1546-170X, (2022; https://www.nature.com/articles/s41591-021-01672-4) (Feb. 2022).
9. A. R. Martin, M. J. Daly, E. B. Robinson, S. E. Hyman, B. M. Neale, *Biological Psychiatry*, Molecular Mechanisms of Affective Disturbance **86**, 97–109, ISSN: 0006-3223, (2019; http://www.sciencedirect.com/science/article/pii/S000632231832119X) (July 2019).
10. C. Márquez-Luna, P.-R. Loh, South Asian Type 2 Diabetes (SAT2D) Consortium, SIGMA Type 2 Diabetes Consortium, A. L. Price, eng, *Genetic Epidemiology* **41**, 811–823, ISSN: 1098-2272 (Dec. 2017).
11. Y. Ruan *et al.*, en, *Nature Genetics* **54**, Number: 5 Publisher: Nature Publishing Group, 573–580, ISSN: 1546-1718, (2022; https://www.nature.com/articles/s41588-022-01054-7) (May 2022).
12. O. Weissbrod *et al.*, en, *Nature Genetics* **54**, Number: 4 Publisher: Nature Publishing Group, 450–458, ISSN: 1546-1718, (2022; https://www.nature.com/articles/s41588-022-01036-9) (Apr. 2022).
13. B. K. Bulik-Sullivan *et al.*, eng, *Nature Genetics* **47**, 291–295, ISSN: 1546-1718 (Mar. 2015).
14. T. B. Cavazos, J. S. Witte, eng, *HGG advances* **2**, 100017, ISSN: 2666-2477 (Jan. 2021).
15. M. Nikpay *et al.*, eng, *Nature Genetics* **47**, 1121–1130, ISSN: 1546-1718 (Oct. 2015).
16. S. Sakaue *et al.*, eng, *Nature Genetics* **53**, 1415–1424, ISSN: 1546-1718 (Oct. 2021).
17. H. Matsunaga *et al.*, eng, *Circulation. Genomic and Precision Medicine* **13**, e002670, ISSN: 2574-8300 (June 2020).
18. C. Bycroft *et al.*, eng, *Nature* **562**, 203–209, ISSN: 1476-4687 (2018).
19. L. Jiang, Z. Zheng, H. Fang, J. Yang, eng, *Nature Genetics* **53**, 1616–1621, ISSN: 1546-1718 (Nov. 2021).
20. O. Weissbrod *et al.*, en, *Nature Genetics* **52**, Number: 12 Publisher: Nature Publishing Group, 1355–1363, ISSN: 1546-1718, (2022; https://www.nature.com/articles/s41588-020-00735-5) (Dec. 2020).
21. S. Gazal *et al.*, en, *Nature Genetics* **50**, Number: 11 Publisher: Nature Publishing Group, 1600–1607, ISSN: 1546-1718, (2022; https://www.nature.com/articles/s41588-018-0231-8) (Nov. 2018).
22. T. 1. G. P. Consortium, en, *Nature* **491**, 56–65, ISSN: 0028-0836, (2013; http://www.nature.com/nature/journal/v491/n7422/full/nature11632.html) (Nov. 2012).
23. T. Ge, C.-Y. Chen, Y. Ni, Y.-C. A. Feng, J. W. Smoller, en, *Nature Communications* **10**, Number: 1 Publisher: Nature Publishing Group, 1776, ISSN: 2041-1723, (2021; https://www.nature.com/articles/s41467-019-09718-5) (Apr. 2019).
24. G. W. Brier, en, *Monthly Weather Review* **78**, 3 (1950).
25. O. Senol-Cosar *et al.*, eng, *Genetics in Medicine: Official Journal of the American College of Medical Genetics* **21**, 2765–2773, ISSN: 1530-0366 (Dec. 2019).
26. D. C. Goff *et al.*, eng, *Circulation* **129**, S49–73, ISSN: 1524-4539 (June 2014).
27. A. V. Khera *et al.*, eng, *Circulation* **139**, 1593–1602, ISSN: 1524-4539 (Mar. 2019).

*28*. L. Hao *et al.*, eng, *Nature Medicine* **28**, 1006–1013, ISSN: 1546-170X (May 2022).

*29*. A. V. Khera *et al.*, eng, *Cell* **177**, 587–596.e9, ISSN: 1097-4172 (Apr. 2019).