

An Integrative Polygenic and Epigenetic Risk Score for Overweight-related Hypertension in Chinese Population

Yaning Zhang (张亚宁) ^{1,#}, Qiwen Zheng (郑启文) ^{2,3,#}, Qili Qian (钱其漂) ^{4,5,#},
Na Yuan (苑娜) ^{2,3}, Tianzi Liu (刘天资) ⁴, Xingjian Gao (高行健) ^{2,3,6}, Xiu Fan (凡秀) ^{2,3,5},
Youkun Bi (毕友坤) ⁷, Guangju Ji (姬广聚) ¹, Peilin Jia (贾佩林) ^{2,3,5}, Sijia Wang (汪思佳) ^{4,5,*},
Fan Liu (刘凡) ^{8,*}, Changqing Zeng (曾长青) ^{1,2,3,5,*}

¹Henan Academy of Sciences, Zhengzhou 450046, China

²China National Center for Bioinformation, Beijing 100101, China

³Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China

⁴CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences, Shanghai 200031, China

⁵University of Chinese Academy of Sciences, Beijing 100049, China

⁶National Clinical Research Center of Kidney Diseases, Jinling Hospital, Nanjing 210002, China

⁷Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China

⁸Department of Forensic Sciences, College of Criminal Justice, Naif Arab University for Security Sciences, Riyadh 12271, Saudi Arabia

*Corresponding authors: czeng@big.ac.cn (Zeng C), fliu@nauss.edu.sa (Liu F), wangsjia@sinh.ac.cn (Wang S).

#Equal contribution.

Handling Editor: Yajie Zhao

Abstract

Overweight-related hypertension (OrH), defined by the coexistence of excess body weight and hypertension (HTN), is an increasing health concern elevating cardiovascular disease risks. In this study, we evaluated the prediction performance of polygenic risk scores (PRSs) and methylation risk scores (MRSs) for OrH in 7605 Chinese participants from two cohorts: the Chinese Academy of Sciences (CAS) and the National Survey of Physical Traits (NSPT). In the CAS cohort, which predominantly consists of academics, males showed significantly higher prevalence of obesity, HTN, and OrH, along with worse metabolic syndrome indicators, compared to females. This disparity was less pronounced in the NSPT cohort and in broader Chinese epidemiological studies. Among ten PRS methods, PRS-CSx was the most effective, enhancing prediction accuracy for obesity [area under the curve (AUC) = 0.75], HTN (AUC = 0.74), and OrH (AUC = 0.75), compared to baseline models using only age and sex (AUC = 0.55–0.71). Similarly, least absolute shrinkage and selection operator (LASSO)-based MRS models improved prediction accuracy for obesity (AUC = 0.70), HTN (AUC = 0.73), and OrH (AUC = 0.78). Combining PRS and MRS further boosted prediction accuracy, achieving AUC values of 0.77, 0.76, and 0.80 for obesity, HTN, and OrH, respectively. These models stratified individuals into high (> 0.6) or low (< 0.1) risk categories, covering 59.95% for obesity, 31.75% for HTN, and 43.89% for OrH. Our findings highlight a higher OrH risk among male academics, emphasize the influence of metabolic and lifestyle factors on MRS predictions, and highlight the value of multi-omics approaches in enhancing risk stratification.

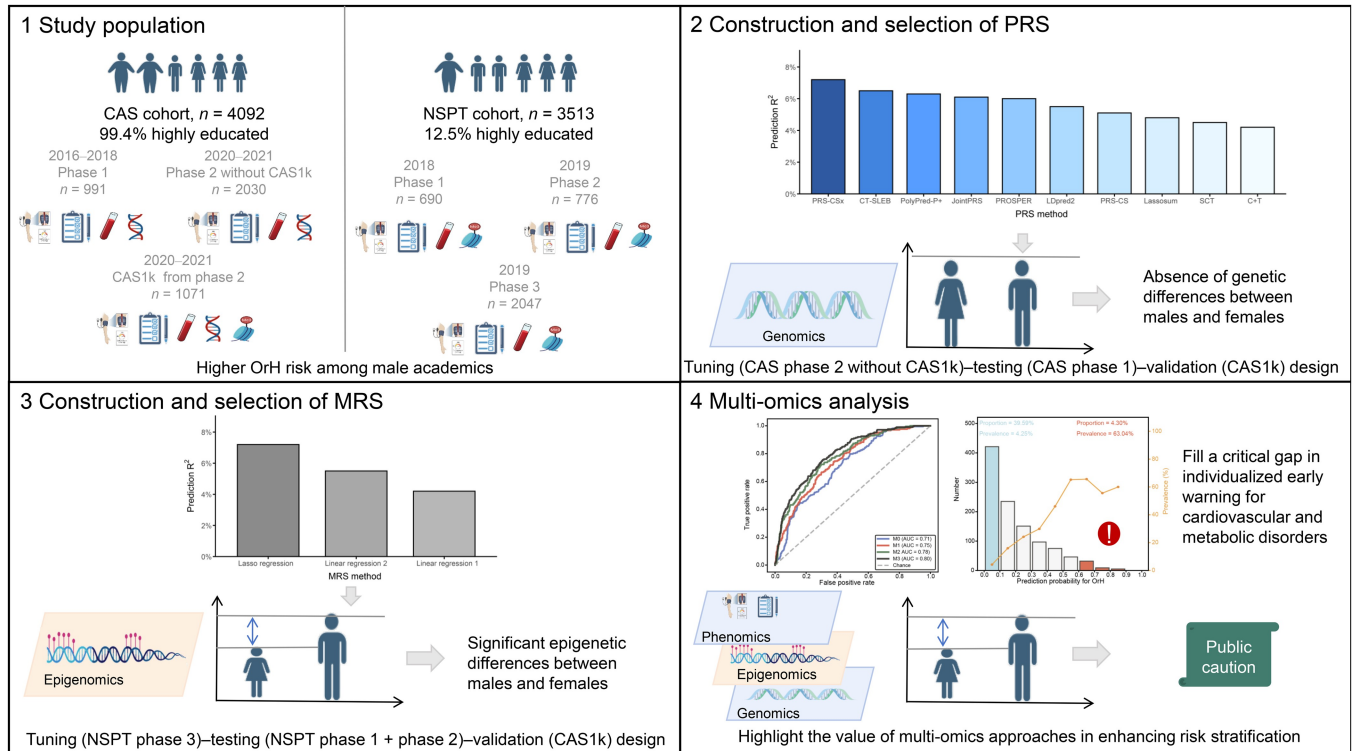
Key words: Overweight-related hypertension; Polygenic risk score; Methylation risk score; Multi-omics prediction; Academic.

Received: 19 May 2024; Revised: 7 March 2025; Accepted: 18 April 2025.

© The Author(s) 2025. Published by Oxford University Press and Science Press on behalf of the Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Graphical abstract



Introduction

Overweight-related hypertension (OrH) is a distinct clinical condition characterized by the concurrent disorders of both body weight and blood pressure [1–3]. Its global prevalence has largely increased over the past two decades, linking to the rising risks of cardiovascular and cerebrovascular diseases [4,5]. Recent genome-wide association studies (GWAS) [6] and epigenome-wide association studies (EWAS) [7] have uncovered numerous genetic and epigenetic factors associated with body weight and blood pressure. To date, the National Human Genome Research Institute - European Bioinformatics Institute (NHGRI-EBI) GWAS catalog [8,9] has cataloged 4263 single nucleotide polymorphisms (SNPs) from 54 studies that are significantly associated with body mass index (BMI), spanning 1252 genes. Meanwhile, 36 studies have identified 2853 SNPs across 862 genes significantly associated with diastolic blood pressure (DBP) and systolic blood pressure (SBP) (Table S1). Furthermore, EWAS have identified 1581 CpG sites across 855 genes that are associated with BMI [10–15] along with 150 CpG sites from 85 genes associated with blood pressure (Table S2) [16–22].

With the continuous discovery of a large number of genetic and epigenetic risk factors, polygenic risk scores (PRSs) [23,24] and methylation risk scores (MRSs) [18,25] have emerged as pivotal tools for profiling the risk landscape of OrH. For instance, a meta-analysis involving 700,000 European individuals constructed a PRS using 941 SNPs, which explained approximately 6% of the variance in BMI [26]. A stratification analysis from the Korean Genome and Epidemiology Study (KoGES) showed that participants in the highest PRS quartile had a two-

fold increased risk of obesity and hypertension (HTN) compared to those in the lowest quartile [27]. Similarly, using EWAS data of nearly 5000 Europeans and Africans, a MRS constructed from 33 CpG loci accounted for 3.31% and 3.99% of the variance in SBP and DBP, respectively [28]. Additionally, a MRS based on 435 CpG sites, derived from penalized regression of methylation data from 2562 unrelated participants in Generation Scotland, explained around 10% of BMI variance, with each standard deviation (SD) increase in MRS associated with a 37% higher risk of obesity [29]. Furthermore, by contrasting MRS with PRS, a recent review emphasized the importance of integrating genetic and epigenetic data for improved trait prediction [25]. Indeed, studies combining PRS and MRS have demonstrated an increase in the explained variance in BMI, up to 14% [30] and 19% [10], highlighting the potential of multi-omics approaches.

Despite these achievements, the prediction of OrH, a comorbidity with various disorders, remains underexplored. The construction of PRS and MRS for OrH faces several challenges, especially in the Chinese population. One major limitation is the reduced efficacy of PRS and MRS when developed in one ancestry group and applied to others. To date, well-powered GWAS and EWAS have predominantly focused on individuals of European ancestry, limiting their applicability to other populations [31,32]. Moreover, cultural and environmental factors unique to the Chinese population may influence how genetic variations and epigenetic modifications contribute to disease risk [7,33].

On the other hand, multiple approaches have been developed for constructing PRS and MRS. In addition to the

classic clumping and thresholding (C+T) method [34], shrinkage methods, such as Stacked C+T (SCT) [35], PRS-CS [36], LDpred2 [37,38], and lassosum [39], adjust the weight of SNPs based on linkage disequilibrium (LD) information. Other methods, such as PRS-CSx [40], CT-SLEB [41], PolyPred-P+ [42], JointPRS [43], and Polygenic Risk scOres based on an enSemble Penalized Regression (PROSPER) [44], are specifically designed for deployment across multiple ancestries, enhancing generalizability across diverse populations. For MRS construction, common computational approaches include C+T [30] and penalized linear regression [11]. Furthermore, effectively integrating PRS and MRS could offer a promising opportunity to improve risk prediction for OrH.

In an attempt to accurately assess OrH in the Chinese population, this study aimed to construct an integrative multi-omics model. Using data from 3021 individuals in the Chinese Academy of Sciences (CAS) cohort, we evaluated various PRS methodologies based on GWAS statistics from the BioBank Japan (BBJ) and the UK Biobank (UKB). Simultaneously, we analyzed several MRS models based on prior EWAS findings using data from 3513 individuals in the National Survey of Physical Traits (NSPT) cohort. The performance of both PRS and MRS in predicting OrH risk was further validated in a separate dataset of 1071 individuals from the CAS cohort.

Results

Higher obesity, HTN, and OrH risk in male academics

Our study included a total of 7605 Chinese individuals from two cohorts: 991 participants of phase 1 and 3101 of phase 2 from the CAS cohort with phenotypic and genomic data, and 3513 participants from the NSPT cohort with phenotypic and DNA methylation data. Additionally, methylation data were available for 1071 samples in phase 2 of the CAS cohort. According to the baseline data in **Table 1**, both the CAS and NSPT cohorts are middle-aged (average age: 39.42 ± 10.11 years and 50.21 ± 12.75 years, respectively), with slightly fewer males (47.8% and 37.1%, respectively). A notable feature of the CAS cohort is the high proportion of participants with higher education (99.4% compared to 12.5% in the NSPT cohort).

A distinct pattern was observed in the CAS cohort, where obesity, HTN, and OrH exhibited notably higher male-to-female prevalence ratios (M/F ratios) compared to both the NSPT cohort and broader Chinese epidemiological studies. In the CAS cohort, the proportion of males was 77.72% in obesity cases, 72.88% in HTN cases, and 80.96% in OrH cases (**Table 1**). This corresponded to M/F ratios of 3.8 (14.63% vs. 3.84%, $P = 3.60E-33$) for obesity, 2.9 (29.00% vs. 9.87%, $P = 2.52E-72$) for HTN, and a particularly concerning 4.7 (21.53% vs. 4.63%, $P = 8.50E-59$) for OrH. Notably, this gender disparity persisted across all age groups (**Figure 1**). In contrast, the proportions of males in obesity, HTN, and OrH cases were all lower in the NSPT cohort (44.20%, 43.45%, and 45.14%, respectively), with the gender disparity (M/F ratio) being much less pronounced (1.34, 1.30, and 1.40, respectively). The M/F ratio of the

NSPT cohort was similar to that observed in the national survey on obesity and HTN (**Table S3**) [45–50].

Additionally, compared to the NSPT cohort, gender disparities in metabolic health were more pronounced in the CAS cohort. CAS males exhibited significantly worse levels of multiple metabolic syndrome indicators, including total cholesterol (TC), triglycerides (TG), low-density lipoprotein (LDL), and high-density lipoprotein (HDL) (**Table S4**). In contrast, CAS females demonstrated significantly better indicators, including TG, HDL, LDL, and fasting blood glucose (FBG), compared to NSPT females. These findings highlight a notable gender difference in the health conditions of academics in China.

Furthermore, we assessed healthy lifestyle scores in the CAS cohort, calculated as the sum of six binary indicators (**Table S5**). As demonstrated in **Figure 1**, as high as 67.1% of women had a score of ≥ 4 , compared to only 30.8% of men who reached this threshold ($P = 2.54E-31$). Normal BMI ($18.5-23.9 \text{ kg/m}^2$) and waist circumference (WC; $< 85 \text{ cm}$ for men and $< 80 \text{ cm}$ for women) showed the most notable gender differences: normal BMI (34.43% in men vs. 66.20% in women, $P = 3.12E-24$) and normal WC (35.52% in men vs. 74.31% in women, $P = 2.84E-35$). Significant differences were also observed in the absence of current smoking (81.06% in men vs. 99.31% in women, $P = 1.36E-19$) and the absence of excessive drinking (92.64% in men vs. 98.84% in women, $P = 7.30E-06$).

PRS-CSx outperforms other PRS methods

In this analysis, we utilized subsets from the CAS cohort for PRS tuning, testing, and validation (see Materials and methods), including: the PRS tuning set ($n = 2030$, phase 2 without CAS1k), the PRS testing set ($n = 991$, phase 1), and the validation set ($n = 1071$, CAS1k). For quantitative traits including BMI, DBP, and SBP, 10 PRS methods (C+T, SCT, PRS-CS, LDpred2, lassosum, PRS-CSx, CT-SLEB, PolyPred-P+, JointPRS, and PROSPER) were trained on the PRS tuning set using GWAS summary statistics from the UKB ($n \approx 450,000$, European) and the BBJ ($n \approx 150,000$, Japanese) (**Figures S1 and S2**; **Table S6**). The latter five methods (PRS-CSx, CT-SLEB, PolyPred-P+, JointPRS, and PROSPER) represent multi-ancestry PRSs, where weights were derived by integrating UKB and BBJ GWAS data.

Overall, throughout the tuning-testing-validation process (**Tables S7–S12**), the PRS generated by PRS-CSx demonstrated robust performance, showing relatively strong prediction ability for the residuals of quantitative phenotypes (including BMI, SBP, and DBP) after regressing out age, sex, and six genomic principal components. Specifically, PRS-CSx achieved an R^2 of 2.40%–9.81% in predicting residual variance, with an average of 4.76% in the testing set (slightly lower than 4.87% for PROSPER) and 5.54% in the validation set (higher than 4.98% for PROSPER) (**Figure 2A**; **Table S12**).

When compared to the PRSs from the Polygenic Score (PGS) Catalog, the PRS-CSx method showed strong performance across multiple traits (BMI, SBP, and DBP) in both testing and validation sets (**Table S13**). The exceptions were for DBP in the testing set, where PRS-CSx ranked second, slightly lower than PGS003964 (3.75% vs. 4.90%), and for SBP, where PRS-CSx ranked second in the testing set and third in the validation set, with marginal differences from the

Table 1 Baseline characteristics of participants in the study

Characteristic	Total		Obesity (BMI ≥ 28)		HTN (SBP ≥ 140 or DBP ≥ 90)		OrH (BMI ≥ 24 with HTN)	
	Control	Case	Control	Case	Control	Case	Control	Case
CAS cohort	4092	3724	3314	778	3314	778	3572	520
Age (year): mean ± SD	39.42 ± 10.11	39.18 ± 10.10	38.05 ± 9.35	45.27 ± 11.10	38.05 ± 9.35	45.27 ± 11.10	38.42 ± 9.63	46.27 ± 10.69
Male: number (%)	1955 (47.78%)	1669 (44.82%)	1388 (41.88%)	567 (72.88%)	1388 (41.88%)	567 (72.88%)	1534 (42.95%)	421 (80.96%)
Higher education: number (%)	4067 (99.39%)	3700 (99.36%)	3303 (99.67%)	764 (98.20%)	3303 (99.67%)	764 (98.20%)	3554 (99.50%)	513 (98.65%)
SBP (mmHg): mean ± SD	118.17 ± 14.26	117.10 ± 13.69	114.03 ± 10.78	135.84 ± 13.78	114.03 ± 10.78	135.84 ± 13.78	115.34 ± 11.99	137.63 ± 13.41
DBP (mmHg): mean ± SD	76.57 ± 10.95	75.83 ± 10.56	73.35 ± 8.48	90.31 ± 9.59	73.35 ± 8.48	90.31 ± 9.59	74.43 ± 9.44	91.26 ± 9.24
BMI (kg/m ²): mean ± SD	23.80 ± 3.26	23.14 ± 2.44	23.34 ± 3.05	25.75 ± 3.40	23.34 ± 3.05	25.75 ± 3.40	23.27 ± 2.98	27.42 ± 2.78
Number	3513	2970	2314	1199	2314	1199	2741	782
Age (year): mean ± SD	50.21 ± 12.75	50.34 ± 13.01	47.26 ± 13.03	55.89 ± 9.99	47.26 ± 13.03	55.89 ± 9.99	50.03 ± 12.95	52.40 ± 9.75
Male: number (%)	1304 (37.12%)	1064 (35.82%)	783 (33.84%)	521 (43.45%)	783 (33.84%)	521 (43.45%)	957 (34.91%)	353 (45.14%)
Higher education: number (%)	439 (12.49%)	372 (12.53%)	391 (16.90%)	48 (4.00%)	391 (16.90%)	48 (4.00%)	403 (12.42%)	37 (4.73%)
SBP (mmHg): mean ± SD	129.47 ± 20.41	128.24 ± 20.43	118.16 ± 11.26	151.29 ± 15.91	118.16 ± 11.26	151.29 ± 15.91	127.71 ± 19.81	150.88 ± 14.70
DBP (mmHg): mean ± SD	80.65 ± 11.79	79.68 ± 11.50	75.02 ± 8.27	91.54 ± 9.77	75.02 ± 8.27	91.54 ± 9.77	79.57 ± 11.22	93.87 ± 10.47
BMI (kg/m ²): mean ± SD	24.49 ± 3.54	23.44 ± 2.62	23.97 ± 3.41	25.50 ± 3.59	23.97 ± 3.41	25.50 ± 3.59	24.00 ± 3.16	30.52 ± 2.11

Note: Characteristics of participants in the CAS cohort (n = 4092) and the NSPT cohort (n = 3513), stratified by HTN, obesity, and OrH, are presented. The table includes the distribution of higher education, age, sex, SBP, DBP, and BMI for both control and case groups. Data are presented as mean ± SD and number (%) for characteristics. Statistical significance (P) for differences between cases and controls is indicated. The self-reported clinically diagnosed HTN or the use of antihypertensive medication were included as HTN cases. BMI, body mass index; DBP, diastolic blood pressure; SBP, systolic blood pressure; HTN, hypertension; OrH, overweight-related hypertension; CAS cohort, Chinese Academy of Sciences cohort; NSPT cohort, National Survey of Physical Traits cohort; SD, standard deviation.

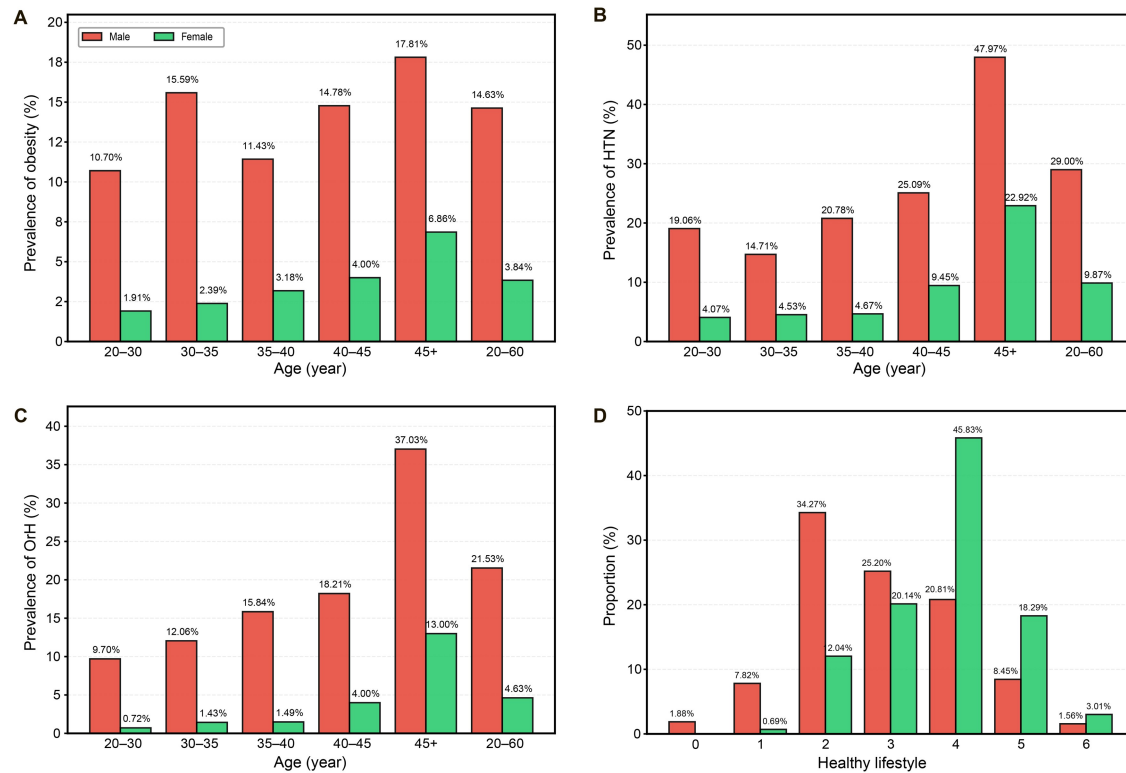


Figure 1 Prevalence of obesity, HTN, and OrH across age groups and healthy lifestyle scores by gender

A–C. Prevalence of obesity (A), HTN (B), and OrH (C) in males (red) and females (green) across different age groups in the CAS cohort ($n = 4092$). **D.** Distribution of healthy lifestyle scores by gender in the CAS1k cohort ($n = 1071$). The healthy lifestyle score, ranging from 0 (least healthy) to 6 (most healthy), is calculated as the sum of six binary criteria: non-smoking, no excessive alcohol consumption, healthy diet (daily intake of fruits and vegetables), regular physical activity, normal BMI ($18.5\text{--}23.9\text{ kg/m}^2$), and normal WC ($< 85\text{ cm}$ for males and $< 80\text{ cm}$ for females). BMI, body mass index; HTN, hypertension; OrH, overweight-related hypertension; WC, waist circumference; CAS cohort, Chinese Academy of Sciences cohort.

top-performing PRS. These results underscore the broad applicability and robustness of PRS-CSx in capturing the genetic architecture of complex traits, leading us to choose PRS-CSx for subsequent analyses.

Prediction accuracy of PRS-CSx for quantitative and binary traits

After comparison and selection of PRS-CSx, we further assessed its prediction performance for quantitative traits (BMI, SBP, and DBP) and binary disease outcomes (obesity, HTN, and OrH) in the validation set ($n = 1071$, CAS1k). The PRSs were approximately normally distributed (Kolmogorov–Smirnov normality test with Bonferroni correction for multiple comparisons, $P > 0.05$).

For quantitative traits, the baseline models, which incorporated sex and age only, yielded R^2 values of 18.28% for BMI, 17.86% for DBP, and 20.47% for SBP. When these models were augmented with respective PRS-CSx, there was a significant increase in accuracy, evidenced by R^2 values of 26.54% for BMI, 21.21% for DBP, and 22.71% for SBP (Table 2).

Similarly, for binary disease statuses including obesity, HTN, and OrH, models integrated with PRS-CSx also demonstrated improved accuracy compared to the baseline models. PRS_BMI distinctively segregated obesity from the non-obesity group ($P = 4.35\text{E}\text{--}11$, Figure 2B), with an odds ratio (OR) for obesity increased by 2.13 for each SD increment (OR/SD in PRS_BMI [95% confidence interval (CI): 1.77–2.57, $P = 1.70\text{E}\text{--}11$] (Table S14). In a five-quantile

schema, the OR for obesity rose progressively across quintiles, reaching 12.63 in the highest quintile (Figure 2C; Table S14). This inclusion of PRS_BMI notably enhanced the model's accuracy for obesity prediction, increasing the area under the curve (AUC) from 0.55 to 0.75 (Table 2). For HTN, the combined PRS_SBP and PRS_DBP (PRS_HTN) significantly differentiated between the HTN and non-HTN groups ($P = 2.24\text{E}\text{--}08$) (Figure 2D), with an OR/SD of 1.68 in PRS_HTN (95% CI: 1.47–1.91, $P = 1.04\text{E}\text{--}10$) (Table S14). The OR for HTN increased gradually across the five quintiles, peaking at 2.93 in the highest quintile (Figure 2E; Table S14). The inclusion of PRS_HTN in the model resulted in an improvement in prediction accuracy for HTN, increasing AUC from 0.70 to 0.74 (Table 2). Lastly, the model for predicting OrH showed a significant improvement in accuracy when including both PRS_BMI and PRS_HTN as predictors, with the AUC increasing from 0.71 to 0.75 (Table 2). Both PRS_BMI and PRS_HTN significantly differentiated between the OrH and non-OrH groups ($P = 7.72\text{E}\text{--}05$ and $P = 5.16\text{E}\text{--}04$, respectively) (Figure 2F), with an OR/SD of 1.42 in PRS_BMI (95% CI: 1.23–1.63, $P = 3.88\text{E}\text{--}05$) (Table S14) and 1.45 in PRS_HTN (95% CI: 1.26–1.66, $P = 1.63\text{E}\text{--}05$) (Table S14). In the highest quintile, the ORs for OrH peaked at 2.36 and 2.02 for PRS_BMI and PRS_HTN, respectively (Figure 2G; Table S14).

We further compared gender disparities in aforementioned analyses. No statistically significant differences were detected

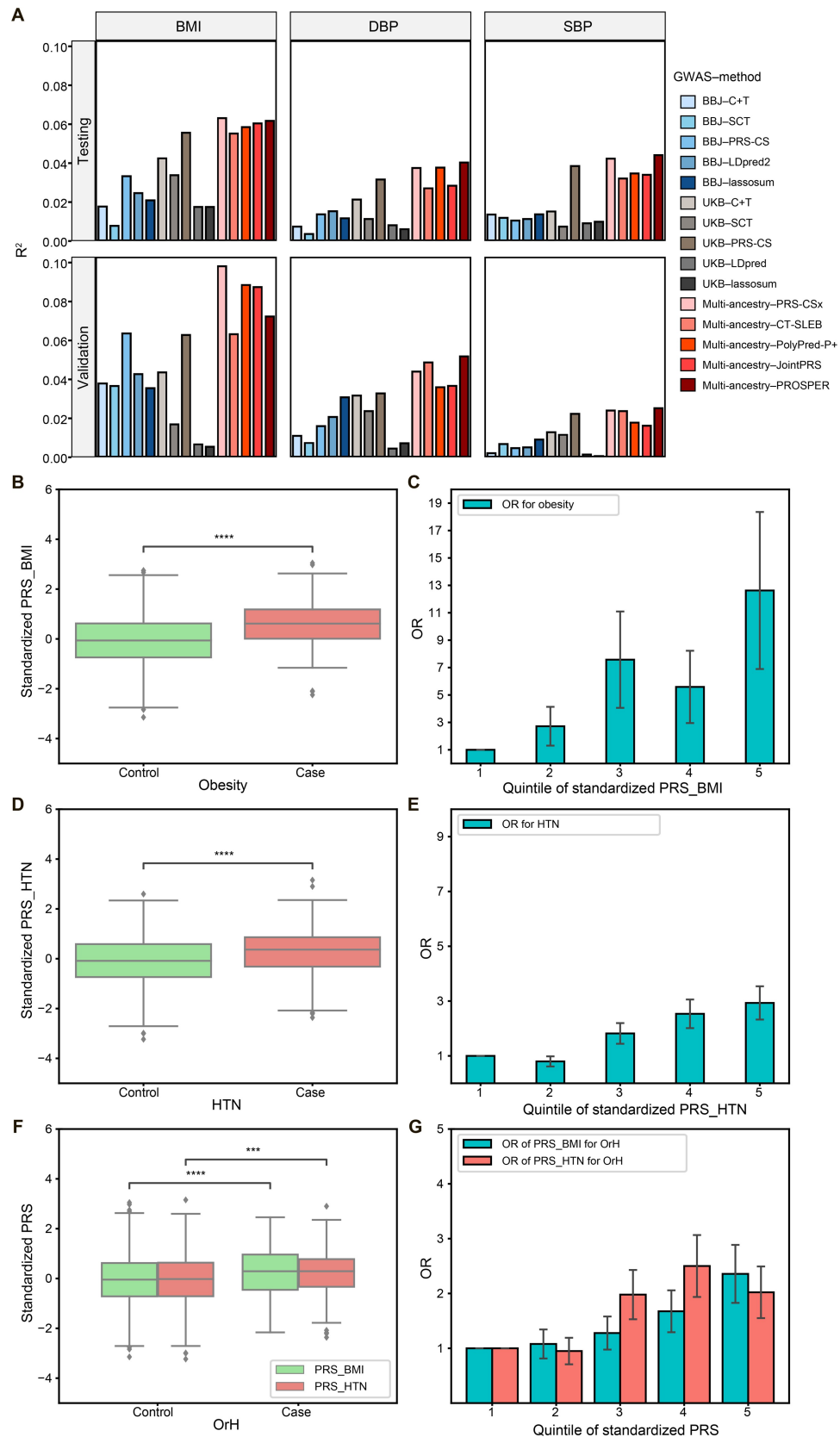


Figure 2 Performance of PRSs for BMI, DBP, and SBP as well as their associations with obesity, HTN, and OrH

A. R^2 values (Y-axis) represent the proportion of phenotypic variance explained by the PRSs for BMI, DBP, and SBP across different GWAS-method combinations (X-axis) in the testing set ($n = 991$, upper) and the validation set ($n = 1071$, lower). Each bar corresponds to a specific method, as indicated by the color-coded legend. The phenotype was regressed on age, sex, and six principal components, and the residual from this regression was used as the dependent variable in the PRS modeling analyses. **B.** Boxplot of PRS for BMI (PRS_{BMI}) in control vs. obesity case groups. **C.** OR for obesity across quintiles of PRS_{BMI}, with the lowest 20% quintile serving as the reference group. **D.** Boxplot of PRS for HTN (PRS_{HTN}) in control vs. HTN case groups. **E.** OR for HTN across quintiles of PRS_{HTN}, with the lowest 20% quintile serving as the reference group. **F.** Boxplot of PRS for BMI (PRS_{BMI}) and HTN (PRS_{HTN}) in control vs. OrH case groups. **G.** OR for OrH across quintiles of PRS_{BMI} and PRS_{HTN}, with the lowest 20% quintile serving as the reference group. ****, $P < 0.0001$; ***, $P < 0.001$ (Wald test). DBP, diastolic blood pressure; SBP, systolic blood pressure; PRS, polygenic risk score; OR, odds ratio; GWAS, genome-wide association study; C+T, clumping and thresholding; SCT, Stacked C+T; PROSPER, Polygenic Risk sCores based on an enSemble PEnalized Regression; UKB, UK Biobank; BBJ, BioBank Japan.

Table 2 Performance of different models for obesity, HTN, and OrH in the validation set

Model	R ² (95% CI)			AUC (95% CI)		
	BMI	SBP	DBP	Obesity	HTN	OrH
Baseline model	18.28% (13.79%–22.48%)	20.47% (16.02%–24.63%)	17.86% (13.69%–21.84%)	0.55 (0.51–0.59)	0.70 (0.66–0.73)	0.71 (0.68–0.75)
PRS model	26.54% (21.71%–31.01%)	22.71% (18.10%–26.81%)	21.21% (16.51%–25.31%)	0.75 (0.70–0.79)	0.74 (0.71–0.77)	0.75 (0.72–0.78)
MRS model	26.86% (22.52%–31.00%)	25.72% (20.87%–29.89%)	23.15% (18.42%–27.42%)	0.70 (0.65–0.74)	0.73 (0.69–0.76)	0.78 (0.74–0.81)
Multi-omics model	33.98% (29.12%–38.52%)	27.58% (22.76%–31.61%)	26.02% (21.42%–30.61%)	0.77 (0.72–0.81)	0.76 (0.73–0.79)	0.80 (0.77–0.83)

Note: The performance of baseline model (only with age and sex), PRS model (with age, sex, and PRS), MRS model (with age, sex, and PRS), and multi-omics model (with age, sex, PRS, and MRS) was evaluated using R² for BMI, DBP, and SBP and AUC for obesity, HTN, and OrH. All metrics and 95% CIs were determined using a five-fold cross-validation approach in the validation set ($n = 1071$). PRS, polygenic risk score; MRS, methylation risk score; AUC, area under the curve; CI, confidence interval.

between males and females (t -test with Bonferroni correction for multiple comparisons, $P > 0.05$) (Figure S3A and B), suggesting little genetic influence on the higher risk in men as observed above.

LASSO-based MRS outperforms linear models and contributes to OrH risk profiling

Three MRS methods [LASSO, linear regression 1, and linear regression 2] were compared using distinct subsets of the NSPT cohort for MRS tuning ($n = 2047$, NSPT phase 3) and testing ($n = 1466$, NSPT phase 1 + phase 2), as well as the validation set ($n = 1071$, CAS1k) for final evaluation (see Materials and methods). We focused on 1506 CpG sites for BMI, 77 for DBP, and 107 for SBP by reviewing prior EWAS results (Table S11) [10–22]. All MRSs were normally distributed (Kolmogorov–Smirnov normality test with Bonferroni correction for multiple comparisons, $P > 0.05$). Among the three methods, LASSO achieved the best performance across all phenotypes in the MRS testing set ($R^2_{\text{BMI}} = 8.48\%$, $R^2_{\text{DBP}} = 1.61\%$, and $R^2_{\text{SBP}} = 3.02\%$) and the validation set ($R^2_{\text{BMI}} = 10.03\%$, $R^2_{\text{DBP}} = 4.68\%$, and $R^2_{\text{SBP}} = 3.70\%$) (Figure 3A; Table S15).

When compared to the baseline models that included sex and age as predictors, the LASSO MRS showcased enhanced accuracy in the validation set. Specifically, R² values increased from 18.28% to 26.86% for BMI, from 17.86% to 23.15% for DBP, and from 20.47% to 25.72% for SBP (Table 2). Notably, MRS_{BMI} exhibited a substantial difference between obesity and non-obesity individuals ($P = 6.09\text{E}^{-11}$) (Figure 3B) and an OR/SD of 1.88 for obesity (95% CI: 1.57–2.24, $P = 4.94\text{E}^{-09}$) (Table S14). In a five-quintile schema, the OR for obesity rose progressively across quintiles, reaching 5.77 in the highest quintile (Figure 3C; Table S14). This integration improved the AUC for obesity from 0.55 to 0.70 (Table 2). Similarly, MRS_{HTN} revealed a clear differentiation between the HTN and non-HTN groups ($P = 6.87\text{E}^{-21}$) (Figure 3D), with an OR/SD of 1.65 (95% CI: 1.43–1.89, $P = 3.85\text{E}^{-09}$) (Table S14). The ORs for HTN increased gradually across the five quintiles, peaking at 4.02 in the highest quintile (Figure 3E; Table S14). This integration improved the AUC for HTN from 0.70 to 0.73 (Table 2). When considering both MRS_{BMI} and MRS_{HTN} in the OrH model, there was an increase in AUC from 0.71 to 0.78 (Table 2). Both MRS_{BMI} and MRS_{HTN} demonstrated significant differentiation between the OrH and non-OrH groups ($P = 7.17\text{E}^{-16}$ and $P = 3.83\text{E}^{-23}$, respectively) (Figure 3F), with an OR/SD of 1.52 in MRS_{BMI} (95% CI: 1.31–1.77, $P = 4.42\text{E}^{-06}$) (Table S14) and 1.72 in MRS_{HTN} (95% CI: 1.46–2.02, $P = 4.99\text{E}^{-08}$) (Table S14). In the highest quintile, the ORs for OrH peaked at 6.03 and 6.52 for MRS_{BMI} and MRS_{HTN}, respectively (Figure 3G; Table S14).

Notably, males exhibited significantly higher MRSs than females for BMI and blood pressure ($1.20\text{E}^{-41} < P < 5.08\text{E}^{-04}$) (Figure S3). These disparities may suggest the distinct influence of life styles and environmental exposures between genders in this cohort.

Impact of metabolic and lifestyle factors on MRS predictions

We conducted a grouping analysis in the validation set based on MRS prediction errors to assess whether other metabolic and lifestyle factors were associated with discrepancies

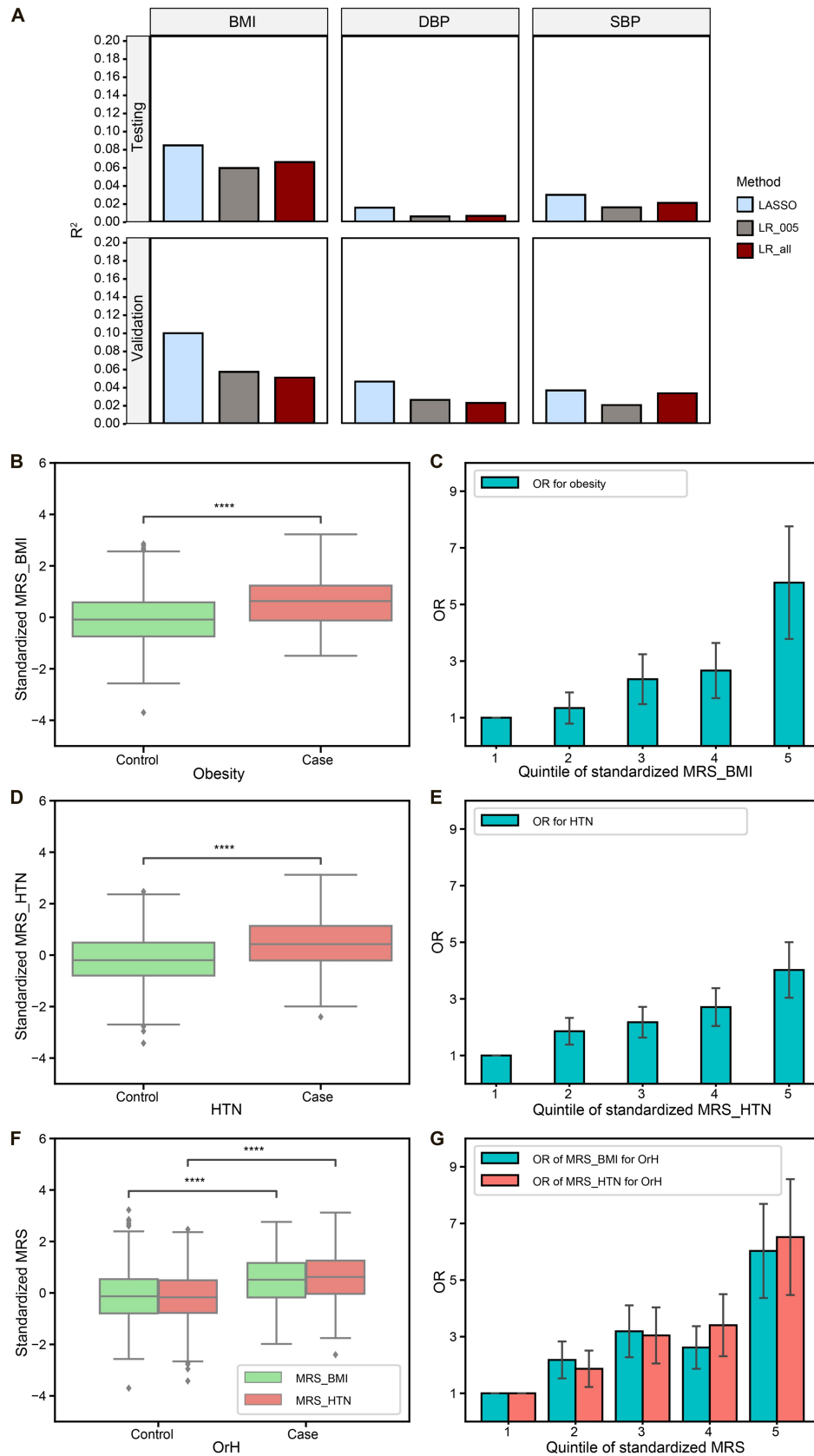


Figure 3 Performance of MRSs for BMI, DBP, and SBP as well as their associations with obesity, HTN, and OrH

A. R^2 values (Y-axis) represent the proportion of phenotypic variance explained by the MRSs for BMI, DBP, and SBP across different methods (X-axis). Results are shown for the testing set ($n = 1466$, upper) and the validation set ($n = 1071$, lower). Each bar corresponds to a specific method, as indicated by the color-coded legend. The phenotype was regressed on age, sex, and cell composition, and the residual from this regression was used as the dependent variable in the MRS modeling analyses. **B.** Boxplot of MRS for BMI (MRS_BMI) in control vs. obesity case groups. **C.** OR for obesity across quintiles of MRS_BMI, with the lowest 20% quintile serving as the reference group. **D.** Boxplot of MRS for HTN (MRS_HTN) in control vs. HTN case groups. **E.** OR for HTN across quintiles of MRS_HTN, with the lowest 20% quintile serving as the reference group. **F.** Boxplot of MRS for BMI (MRS_BMI) and HTN (MRS_HTN) in control vs. OrH case groups. **G.** OR for OrH across quintiles of MRS_BMI and MRS_HTN, with the lowest 20% quintile serving as the reference group. ****, $P < 0.0001$ (Wald test). MRS, methylation risk score; LASSO, least absolute shrinkage and selection operator; LR_005, linear regression using only significant CpG sites with $P < 0.05$; LR_all, linear regression using all available CpG sites.

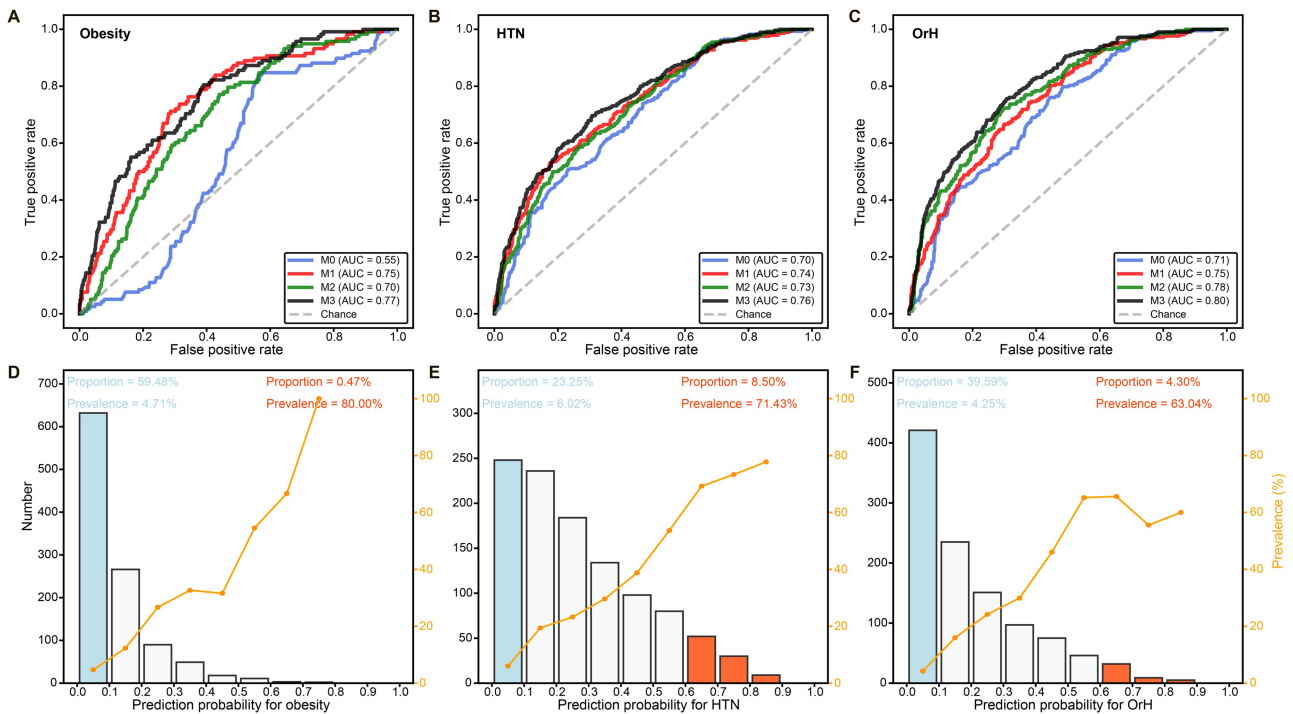


Figure 4 Prediction performance of multi-omics models for obesity, HTN, and OrH in the validation set

A–C. Assessment of AUC values for obesity (A), HTN (B), and OrH (C) using a 5-fold cross-validated logistic regression model. Predictors included age and sex (M0), age, sex, and PRS (M1), age, sex, and MRS (M2), and age, sex, PRS, and MRS (M3) in the validation set ($n = 1071$). **D–F.** In the multi-omics prediction model (M3), the analyses further focused on individuals with extreme prediction probabilities for obesity (D), HTN (E), and OrH (F). Bars represent the number of participants within specific prediction probability intervals, with blue bars indicating low risk (< 0.10) and red bars indicating high risk (> 0.60). The orange line represents the prevalence in each interval. AUC, area under the curve.

between MRS-predicted and observed values of BMI, DBP, and SBP (Table S16). The results indicate that discrepancies between the predicted and observed values are indeed associated with specific metabolic and lifestyle factors. For BMI, participants in the underestimated group (the lowest 10% quantile of prediction errors) tended to have healthier lipid profiles (higher HDL, lower TG, and lower LDL) and better lifestyle scores. Conversely, the overestimated group (the highest 10% quantile) was associated with less favorable lipid profiles, higher FBG, and poorer lifestyle scores. Similar patterns were observed for DBP and SBP, where overestimation by the MRS was linked to higher levels of TC, LDL, and FBG, as well as lower lifestyle scores. These results indicate that metabolic health and lifestyle behaviors may influence the accuracy of MRS predictions for these cardiovascular risk factors.

Multi-omics model improves OrH risk profiling

We further integrated MRS and PRS into a multi-omics score and assessed its performance in predicting the risk of obesity, HTN, and OrH in the validation set according to a five-fold cross-validation design. For obesity, integrating MRS_BMI into the PRS model improved the AUC from 0.75 to 0.77 (Figure 4A; Table 2). This multi-omics score fairly classified 0.47% of the population as high risk (prediction probability > 0.6), who indeed showed a high prevalence of 80.00% (Figure 4D); meanwhile, it effectively identified 59.48% of the population as low risk (prediction probability < 0.1), who in fact had a low prevalence of 4.71%. Consequently,

this indicates that our model is informative for 59.95% of the population in obesity risk profiling. For HTN, integrating MRS_SBP and MRS_DBP into the PRS model improved the AUC from 0.74 to 0.76 (Figure 4B; Table 2). This model is informative for 31.75% of the population in HTN risk profiling, effectively classifying the high-risk group (8.50% with prediction probability > 0.6 and a prevalence of 71.43%) and low-risk group (23.25% with prediction probability < 0.1 and a prevalence of 6.02%) (Figure 4E). In particular, for OrH, integrating MRS_BMI, MRS_SBP, and MRS_DBP into the PRS model boosted the AUC from 0.75 to 0.80 (Figure 4C; Table 2). This is informative for 43.89% of the population in OrH risk profiling, effectively classifying the high-risk group (4.30% with prediction probability > 0.6 and a prevalence of 63.04%) and low-risk group (39.59% with prediction probability < 0.1 and a prevalence of 4.25%) (Figure 4F).

Discussion

In this study, we aimed to develop an accurate and effective prediction model for OrH by analyzing data from two general population cohorts, CAS and NSPT, totaling 7605 individuals. We assessed the performance of ten methods for PRSs and three strategies for MRSs using a tuning-testing-validation approach. Additionally, we developed a multi-omics model to enhance prediction accuracy. Throughout our analysis, we also found distinct population characteristics among academics in our study.

Factors related to gender disparity in disease prevalence among Chinese academics

One unexpected observation in our study was the notably high prevalence of obesity, HTN, and OrH among males in Chinese academics. Therefore, we reviewed epidemiological data from the past three decades, which show significant changes in the prevalence of these conditions [45–50], most likely driven by China's rapid economic growth and various risk factors, including economic status, gender, age, education, smoking, drinking, and inhabiting regions [51–56]. The M/F ratio for obesity has fluctuated between 0.6 and 1.3 in both urban and rural areas, with a marked disparity in urban areas (M/F ratio = 1.9) and even more pronounced among academics (M/F ratio = 3.8) (Table S3). For HTN, the M/F ratio in the CAS cohort reaches 2.9, significantly higher than in the NSPT cohort and broader Chinese epidemiological studies, where it ranges from 1.1 to 1.3.

The gender disparity in disease prevalence observed among Chinese academics aligns with findings from several studies. Research from the China Health and Nutrition Survey and the Chinese Center for Disease Control and Prevention has indicated that women with higher education levels tend to have a lower BMI and reduced odds of being overweight, while men with higher education levels exhibit a higher BMI and increased odds of being overweight in China [57,58]. Similar patterns have been observed in studies from Brazil, Russia, India, China, South Africa (BRICS) economies [59] and Southern European countries [60,61], further highlighting the association between education levels and obesity, particularly among women. Some studies have also examined the impact of education on HTN, indicating that individuals with higher education levels generally have healthier blood pressure. A Mendelian randomization study using data from FinnGen and the UKB suggests a causal relationship between education level and HTN. For each SD increase in genetically predicted higher education, the risk of HTN decreases by 44% [62]. Additionally, a study involving approximately 1.28 million adults from the China Health Evaluation And Risk Reduction through nationwide Teamwork (ChinaHEART) project has found that as education level increases, there is a significant downward trend in SBP [63].

To explore the potential reasons for the significant gender disparity, we first examined the genetic possibility and found no significant genetic differences in the performance of PRS. However, notable epigenetic differences were observed in MRS for the corresponding diseases between males and females. Consistent with previous studies linking higher MRS to poorer metabolic health [11,64], our analysis showed that male academics exhibited higher MRS on BMI, DBP, and SBP, along with their relatively unhealthy lifestyles and metabolic syndrome traits. These findings suggest that the observed gender disparity is likely influenced by the combined effects of metabolic health and epigenetic factors.

We then briefly explored potential epigenetic explanations for the observed gender differences in disease prevalence. First, when examining regular physical activity, only minimal differences between sexes were found (Table S5), which aligns with data of a 15-year national survey [65]. Thus, physical activity does not provide a compelling explanation for the gender disparities observed in the CAS cohort. However, for other lifestyle factors and metabolic syndrome profiles, male academics showed significantly poorer

parameters compared to their female counterparts (Tables S4 and S5). These differences are likely related to the more frequent social gatherings in males nationwide, such as dinners and drinking events, which are commonly associated with higher calorie expenditure and alcohol consumption [66]. Additionally, men may experience great social pressure as the primary bread-winners for their families, which is also likely associated with unhealthy lifestyle choices and an increased risk of metabolic-related diseases [67,68]. Considering the dominant male composition in academia (especially in full professors), one possible reason for the notably high M/F ratio (3.8) in the CAS cohort could be the intense academic pressure and heavy workload in males, which may largely boost unhealthy lifestyle habits. On the other hand, cultural attitudes in China tend to favor slimmer figures for women [47], and female academics may possess greater self-discipline regarding their health as well as more resources and opportunities to maintain their body shapes [55], which further accentuate the gender disparity observed in our cohort.

Leveraging a multi-omics approach to enhance OrH risk prediction

OrH, as a common comorbidity pattern, would exacerbate cardiovascular and cerebrovascular damage more aggressively than simple obesity and HTN. Especially, its prevalence has shown a significant upward trend globally in the past 20 years. Although the integration of PRS and MRS demonstrates improved utility in various diseases, multi-omics prediction models for OrH remain limited. In this study, we developed such an approach for OrH prediction using both genomic and epigenomic signals, achieving an AUC as high as 0.80.

We first developed effective PRSs for the CAS cohort to predict BMI, DBP, and SBP by benchmarking five single-ancestry approaches (C+T, SCT, PRS-CS, LDpred2, and lassosum) and five multi-ancestry approaches (PRS-CSx, CT-SLEB, PolyPred-P+, JointPRS, and PROSPER). Among all methods, the top 3 with the highest accuracy are multi-ancestry methods across all traits. For BMI, the R^2 values for the top 3 multi-ancestry methods ranged from 8.74% to 9.81%, compared to the best single-ancestry model, which achieved 6.36% in validation analysis (Table S12). Similarly, for DBP, the highest R^2 of multi-ancestry models was 5.18%, surpassing the best single-ancestry model of 3.28%. Our results well confirm the outperformance of multi-ancestry PRS approaches over single-ancestry, and further demonstrate the enhanced generalizability of multi-ancestry PRS by leveraging shared genetic effects across different ancestries.

Among multi-ancestry PRS models, PRS-CSx consistently exhibited robust performance, achieving the highest R^2 for BMI and consistently ranking among the top 3 models for both DBP and SBP in the validation set. This superior performance of complex traits across ancestries may be attributed to PRS-CSx's advantage of Bayesian continuous shrinkage. Except for being only slightly (< 0.4%) behind PGS003882 and PGS005015 for SBP in the validation set, the PRS-CSx model developed in our study notably outperforms many published models in the PGS Catalog (116 PRSs for BMI, 72 for SBP, and 51 for DBP). These results further emphasize the potential of our PRS profiling for the Chinese population as a robust and reliable tool for genetic risk prediction and precision medicine applications.

Unlike genetic models, methylation data provide a real-time snapshot of an individual's risk profile by capturing the epigenetic landscape, which reflects not only genetic susceptibility but also modifiable influences that contribute to disease progression, as reported in numerous studies [69,70]. Therefore, we aimed to use MRS for potential disease prediction based on currently available baseline data. After feature selection and optimization across various methylation models, LASSO-based MRS demonstrated the best performance in predicting BMI, DBP, and SBP, achieving similar or better R^2 values compared to previous studies (e.g., 10.03% for BMI *vs.* 10.00% reported [29], and 4.68% for blood pressure *vs.* 3.99% reported [28]). Considering the environmental or lifestyle factors, these MRS models, especially with longitudinal data in the future, may provide valuable insights into an individual's health status and potentially serve as early warnings for unhealthy conditions.

Moreover, combining MRS with PRS enhances risk prediction by linking genetic susceptibility with current epigenetic states. Indeed, compared to the AUC of 0.75 for PRS and 0.78 for MRS, we observed an integrated AUC of 0.80 for OrH risk profiling, further confirming a shared molecular mechanism in obesity and HTN. This profiling may also fill a critical gap in individualized early warning for cardiovascular and metabolic disorders. By identifying high-risk individuals (risk score > 0.6) using multi-omics models, such as 4.30% for OrH in the CAS cohort, healthcare providers can implement more targeted preventive measures and treatment strategies to improve their health status.

Conclusion

This research reveals a notably high prevalence of obesity, HTN, and OrH among males but significantly lower prevalence among females in Chinese academics with characterizations of research career and higher education. These results considerably diverge from common patterns observed in Chinese epidemiological investigations. Additional analysis indicates such large gender disparities are primarily associated to the complex interplay among epigenetic factors, lifestyle, and metabolic health, raising concerns about notably higher risks for males within Chinese academics. In omics analysis, PRS-CSx and LASSO-based MRS demonstrate high potential as robust tools for risk assessment of obesity, HTN, and OrH. The integration of PRS and MRS further enhance the accuracy of the risk profiling, suggesting the effectiveness of multi-omics approaches for improved personalized risk assessment strategies especially for OrH high-risk populations.

Materials and methods

Study population

The CAS cohort

This study involved 4092 Chinese participants from the CAS cohort, which was established in 2015 to target employees of the CAS in Beijing, China. The cohort was highly educated, with 99.4% holding at least a university degree. Participants first completed an online questionnaire that gathered information on factors such as gender, smoking status, alcohol consumption, tea intake, and sleep duration. Subsequently, they underwent clinical assessments at designated hospitals, including anthropometric, physical, blood, urine, and

imaging exams, with 8 ml of blood collected from each participant.

Recruitment occurred in two phases. Phase 1 (2016–2018) included 991 participants; their DNA samples were analyzed using 30× whole-genome sequencing (WGS), and all phenotypic data were collected at the General Hospital of Aviation Industry Corporation of China. Phase 2 (2020–2021) added 3101 participants; their DNA samples were analyzed using Illumina genotyping microarrays, and all phenotypic data were collected at Beijing Zhongguancun Hospital. In phase 2, a subset of 1071 individuals (the CAS1k subgroup) was designed to provide multi-omics data, with their samples further analyzed using Illumina methylation microarrays.

The NSPT cohort

The NSPT cohort is a population-based prospective cohort study consisting of 3523 Han Chinese individuals from multiple regions in China, including Taizhou, Nanning, and Zhengzhou (1310 males and 2213 females; aged from 18 to 83 years, mean \pm SD = 50.21 \pm 12.75). After quality control, 3513 participants were retained for subsequent analysis. Recruitment occurred in three phases: phase 1 ($n = 690$) in 2018, phase 2 ($n = 776$) in 2019, and phase 3 ($n = 2047$) in 2019. DNA methylation was assessed using Illumina methylation microarrays on blood samples.

Definitions of overweight, obesity, HTN, OrH, healthy lifestyle, and higher education

Overweight was defined as a BMI between 24.0 and 27.9 kg/m², while obesity was defined as BMI \geq 28.0 kg/m², according to China's guidelines [71]. HTN was defined as SBP \geq 140 mmHg, DBP \geq 90 mmHg, self-reported HTN diagnosis, or use of antihypertensive medications. Individuals with both BMI \geq 24.0 kg/m² and HTN were categorized as having OrH [72]. It is important to note that thresholds for defining obesity or HTN may vary across populations [73,74], and comparisons with studies using different criteria should be interpreted with caution.

Healthy lifestyle factors were defined based on the China Kadoorie Biobank (CKB) criteria [75], which include non-smoking, no excessive alcohol consumption, healthy diet (daily fruit and vegetable intake), regular physical activity, normal BMI (18.5–23.9 kg/m²), and normal WC (< 85 cm for males and < 80 cm for females). Participants earned a score of 1 for each criterion that they met and 0 for each one that they did not, resulting in a total score ranging from 0 to 6, representing their overall healthy lifestyle. Higher education was defined as having any college or university degree.

WGS and microarray genotyping in the CAS cohort

WGS was performed at 30× coverage on the Illumina HiSeq 3000 platform (Illumina, San Diego, CA), and sequencing reads were aligned to the hg19 reference genome [76]. Variants were called using Genome Analysis Toolkit (GATK) [77] and annotated using ANNOVAR [78], with detailed methods for sample and library preparation reported previously [79].

Microarray genotyping was conducted on the Infinium Asian Screening Array + MultiDisease-24 BeadChip (Illumina, San Diego, CA). SNP genotypes were phased and imputed using IMPUTE2 [80] based on the East Asian population in the 1000 Genomes Project [81,82].

Quality control included removing individuals with gender mismatch, low genotyping call rate (< 97%), or abnormal

heterozygosity (outside the mean \pm 3 SD range). For SNPs, we excluded those with imputation score $<$ 0.6 (in the CAS phase 2 cohort), Hardy–Weinberg equilibrium $P <$ $1E-04$, genotyping call rate $<$ 98%, and minor allele frequency (MAF) $<$ 1%. After these steps, 3,169,262 SNPs and 4092 individuals were retained for subsequent analysis.

Methylation microarray of the CAS1k and NSPT cohorts

DNA methylation profiling for both the CAS1k and NSPT cohorts was performed on the Illumina Infinium MethylationEPIC BeadChip (Illumina, San Diego, CA). The raw array data were processed using the ChAMP package [83] in R to compute β values for methylation levels. Probes were filtered based on Illumina quality thresholds (bead count $<$ 3 in $>$ 5% of samples and 1% of samples with a detection P value $>$ 0.05). Batch effects were corrected using ComBat [84,85], and cell-type heterogeneity was adjusted using EpiDISH [86]. After quality control, 751,015 CpG sites were retained for the CAS1k cohort, and 811,876 CpG sites were retained for the NSPT cohort.

Construction and selection of PRSs

PRS construction and selection followed a tuning–testing–validation design using the CAS cohort. The tuning set included 2030 participants from phase 2 (excluding CAS1k), the testing set had 991 participants from phase 1, and the validation set consisted of 1071 participants from phase 2 (*i.e.*, the CAS1k cohort).

PRSs for BMI, DBP, and SBP were constructed using GWAS summary statistics from the UKB (\sim 450,000 Europeans) [87] and the BBJ (\sim 150,000 Japanese) [88,89]. Ten PRS methods were applied: C+T [34], SCT [35], PRS-CS [36], LDpred2 [37,38], lassosum [39], PRS-CSx [40], CT-SLEB [41], PolyPred-P+ [42], JointPRS [43], and PROSPER [44]. Hyperparameters were fine-tuned in the tuning set and evaluated in the testing and validation sets, with performance assessed by R^2 and 95% CI using bootstrap resampling ($k = 10,000$, detailed in File S1).

To identify the optimal PRS for East Asians, the PRSs constructed by the ten methods were compared to existing scores in the PGS Catalog, selecting 48, 4, and 20 PRSs for BMI, DBP, and SBP, respectively, based on the required information. The best-performing PRS was then used for multi-omics prediction analysis in the validation set. In all PRS analyses, phenotypes were regressed on age, sex, and six genomic principal components, with the residuals used for PRS modeling to calculate adjusted R^2 values reflecting variance explained beyond potential confounders.

Construction and selection of MRSs

MRS construction and selection followed a tuning–testing–validation design using both the NSPT and CAS1k cohorts. The MRS tuning set included 2047 participants from phase 3 of the NSPT cohort, while the MRS testing set consisted of 1466 participants from the phase 1 and phase 2 of the NSPT cohort. The validation set was the CAS1k cohort ($n = 1071$). MRSs for BMI, DBP, and SBP were derived from findings in previous studies [10–22], resulting in final sets of 1506, 77, and 107 CpG sites for BMI, DBP, and SBP, respectively (details provided in File S1).

To construct the MRSs, we used three different methods: linear regression 1, which included all CpG sites from the

studies without filtering; linear regression 2, which selected only CpG sites with a P value $<$ 0.05; and LASSO regression, which applied penalized linear regression to optimize the model. Different CpG sets with corresponding beta coefficients were generated in the tuning set, then evaluated and validated in the testing and validation sets.

The best-performing MRS was selected and applied to the validation set for multi-omics prediction analysis. In all MRS analyses, phenotypes were regressed on age, sex, and cell composition, with the residuals serving as the dependent variable for MRS modeling. This approach enabled us to report adjusted R^2 values and 95% CI, calculated through bootstrap resampling ($k = 10,000$), that account for the variance in MRS explained beyond the confounding effects of age, sex, and cell composition.

Grouping analysis of MRS-predicted values

A grouping analysis of MRS-predicted values was performed in the validation set to determine whether discrepancies between MRS-predicted and observed values for BMI, DBP, and SBP were associated with metabolic and lifestyle factors. Participants were grouped into three groups based on the quantiles of the prediction error (predicted minus observed): the lowest 10% quantile, where MRS significantly underestimated the trait; the middle 80% quantiles, where MRS predictions closely matched observed values; and the highest 10% quantile, where MRS significantly overestimated the trait. Metabolic factors (including HDL, LDL, TC, TG, FBG) and healthy lifestyle scores were then compared across these groups using independent t -tests to identify significant differences between the underestimated, accurately predicted, and overestimated groups.

Multi-omics risk prediction

The best-performing PRS and MRS were combined into multi-omics scores, and their prediction performance was assessed in the validation set, using linear regression for continuous traits (BMI, SBP, and DBP) and logistic regression for binary traits (obesity, HTN, and OrH). These models were adjusted for age and sex as covariates. Performance was evaluated using a five-fold cross-validation design. In brief, models were trained on four folds and tested on the remaining one, and this process was repeated five times, each time using a different part of the data for testing. The R^2 values for continuous traits and AUC values for binary traits were averaged across the five iterations to provide a more generalized estimate of the model's prediction power.

Predictors were specified as follows. For continuous traits, separate models were developed using the corresponding PRS, MRS, or both (multi-omics) as predictors. For binary traits, obesity models used PRS_BMI, MRS_BMI, or both as predictors; HTN models incorporated the average of PRS_DBP and PRS_SBP, the average of MRS_DBP and MRS_SBP, or both averages as predictors; OrH models included PRS_BMI plus the average of PRS_DBP and PRS_SBP, MRS_BMI plus the average of MRS_DBP and MRS_SBP, or the combination of all four predictors.

This modeling strategy ensured a consistent and robust comparison of baseline (sex and age only), PRS, MRS, and multi-omics models. All analyses were performed using R (v4.0.3) and Python (v3.6.4).

Ethical statement

This study was approved by the Ethics Committee of the Beijing Institute of Genomics, CAS (Approval Nos. 2015H023 and 2021H001), the Ethics Committee of Beijing Zhongguancun Hospital (Approval No. 20201229), the Ethics Committee of Fudan University (Approval No. 14117), and the Ethics Committee of Shanghai Institutes for Biological Sciences (Approval No. ER-SIBS-261410). All participants provided written informed consent for the use of their samples and data for medical research.

Code availability

C+T and SCT are available at <https://github.com/privefl/bigsnp/tree/cef0482c3c87ff51b63f5f2b0c896c75717ab92d/vignettes>. PRS-CS is available at <https://github.com/getian107/PRScs>. PRS-CSx is available at <https://github.com/getian107/PRScsx>. CT-SLEB is available at <https://andrewhaoyu.github.io/CTSLEB/>. JointPRS is available at <https://github.com/LeqiXu/JointPRS>. PROSPER is available at <https://github.com/Jingning-Zhang/PROSPER>. LDpred and lassosum are available at <https://privefl.github.io/bigsnp/articles/LDpred2.html>. PolyPred-P+ is available at <https://github.com/omerwe/polyfun>. The code has also been submitted to BioCode at the National Genomics Data Center (NGDC), China National Center for Bioinformation (CNCB) (BioCode: BT007949), which is publicly accessible at <https://ngdc.cncb.ac.cn/biocode/tools/BT007949>.

Data availability

The raw sequencing data from the CAS cohort have been deposited in the Genome Sequence Archive [90] at the National Genomics Data Center (NGDC), China National Center for Bioinformation (CNCB) (GSA: CRA000631), and are publicly accessible at <https://ngdc.cncb.ac.cn/gsa>. The genome sequence has been deposited in the Genome Warehouse [91] at the NGDC, CNCB (GWH: GWHAAAS00000000), and is publicly accessible at <http://bigd.big.ac.cn/gwh>. Additionally, the methylation data from the CAS and NSPT cohorts have been deposited in the Open Archive for Miscellaneous Data [90] at the NGDC, CNCB (OMIX: OMIX004333 for CAS, OMIX00436 for NSPT), and are publicly accessible at <https://ngdc.cncb.ac.cn/omix>.

CRedit author statement

Yaning Zhang: Conceptualization, Methodology, Formal analysis, Writing – original draft. **Qiwen Zheng:** Conceptualization, Visualization, Writing – review & editing. **Qili Qian:** Data curation, Formal analysis. **Na Yuan:** Data curation. **Tianzi Liu:** Data curation. **Xingjian Gao:** Data curation. **Xiu Fan:** Data curation. **Youkun Bi:** Data curation. **Guangju Ji:** Data curation. **Peilin Jia:** Data curation. **Sijia Wang:** Conceptualization, Supervision, Writing – review & editing. **Fan Liu:** Conceptualization, Supervision, Writing – review & editing. **Changqing Zeng:** Conceptualization, Supervision, Writing – review & editing. All authors have read and approved the final manuscript.

Competing interests

The authors have declared no competing interests.

Supplementary material

Supplementary material is available at *Genomics, Proteomics & Bioinformatics* online (<https://doi.org/10.1093/gpbjnl/qzaf048>).

Acknowledgments

We thank the participants who contributed their data in the CAS cohort and the NSPT cohort study. This work was supported by the Science and Technology Service Network Initiative of the CAS (Grant No. KFJ-STS-ZDTP-079), the Strategic Priority Research Program of the CAS (Grant No. XDB38010400), the Science and Technology Research Project of Henan Province (Grant No. 232102310066), the Basic Research Fund of Henan Academy of Sciences (Grant No. 230618032), the Startup Research Fund of Henan Academy of Sciences (Grant No. 231816040 and 232016009), the National Natural Science Foundation of China (Grant No. 32325013), the CAS Project for Young Scientists in Basic Research (Grant No. YSBR-077), the Strategic Priority Research Program of the CAS (Grant No. XDB38020400 to Sijia Wang), and Shanghai Excellent Academic Leaders Program (Grant No. 22XD1424700), China.

ORCID

0000-0002-8073-3137 (Yaning Zhang)
 0000-0001-6921-6254 (Qiwen Zheng)
 0009-0005-8797-5242 (Qili Qian)
 0000-0003-3614-2441 (Na Yuan)
 0000-0002-7401-3571 (Tianzi Liu)
 0000-0002-6142-8102 (Xingjian Gao)
 0000-0002-8590-9859 (Xiu Fan)
 0000-0002-0195-9294 (Youkun Bi)
 0000-0001-8626-3490 (Guangju Ji)
 0000-0003-4523-4153 (Peilin Jia)
 0000-0001-6961-7867 (Sijia Wang)
 0000-0001-9241-8161 (Fan Liu)
 0000-0002-0037-1771 (Changqing Zeng)

References

- [1] Wang J, Feng B, Xiong X. Chinese herbal medicine for the treatment of obesity-related hypertension. *Evid Based Complement Alternat Med* 2013;2013:757540.
- [2] Julius S, Valentini M, Palatini P. Overweight and hypertension: a 2-way street? *Hypertension* 2000;35:807–13.
- [3] Wilson PW, D'Agostino RB, Sullivan L, Parise H, Kannel WB. Overweight and obesity as determinants of cardiovascular risk: the Framingham experience. *Arch Intern Med* 2002; 162:1867–72.
- [4] Landsberg L, Aronne LJ, Beilin LJ, Burke V, Igel LI, Lloyd-Jones D, et al. Obesity-related hypertension: pathogenesis, cardiovascular risk, and treatment: a position paper of The Obesity Society and the American Society of Hypertension. *J Clin Hypertens (Greenwich)* 2013;15:14–33.
- [5] Zidek W, Naditch-Brulé L, Perlini S, Farsang C, Kjeldsen SE. Blood pressure control and components of the metabolic syndrome: the GOOD survey. *Cardiovasc Diabetol* 2009;8:51.

- [6] Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 years of GWAS discovery: biology, function, and translation. *Am J Hum Genet* 2017;101:5–22.
- [7] Wei S, Tao J, Xu J, Chen X, Wang Z, Zhang N, et al. Ten years of EWAS. *Adv Sci (Weinh)* 2021;8:e2100727.
- [8] Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* 2019;47:D1005–12.
- [9] Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res* 2023;51:D977–85.
- [10] McCartney DL, Hillary RF, Stevenson AJ, Ritchie SJ, Walker RM, Zhang Q, et al. Epigenetic prediction of complex traits and death. *Genome Biol* 2018;19:136.
- [11] Do WL, Sun D, Meeks K, Dugué PA, Demerath E, Guan W, et al. Epigenome-wide meta-analysis of BMI in nine cohorts: examining the utility of epigenetically predicted BMI. *Am J Hum Genet* 2023;110:273–83.
- [12] Li W, Xia M, Zeng H, Lin H, Teschendorff AE, Gao X, et al. Longitudinal analysis of epigenome-wide DNA methylation reveals novel loci associated with BMI change in East Asians. *Clin Epigenetics* 2024;16:70.
- [13] Dick KJ, Nelson CP, Tsaprouni L, Sandling JK, Aïssi D, Wahl S, et al. DNA methylation and body-mass index: a genome-wide analysis. *Lancet* 2014;383:1990–8.
- [14] Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature* 2017;541:81–6.
- [15] Chen Y, Kassam I, Lau SH, Kooner JS, Wilson R, Peters A, et al. Impact of BMI and waist circumference on epigenome-wide DNA methylation and identification of epigenetic biomarkers in blood: an EWAS in multi-ethnic Asian individuals. *Clin Epigenetics* 2021;13:195.
- [16] Richard MA, Huan T, Ligthart S, Gondalia R, Jhun MA, Brody JA, et al. DNA methylation analysis identifies loci for blood pressure regulation. *Am J Hum Genet* 2017;101:888–902.
- [17] Kato N, Loh M, Takeuchi F, Verweij N, Wang X, Zhang W, et al. Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation. *Nat Genet* 2015;47:1282–93.
- [18] Hong X, Miao K, Cao W, Lv J, Yu C, Huang T, et al. Association between DNA methylation and blood pressure: a 5-year longitudinal twin study. *Hypertension* 2023;80:169–81.
- [19] Hoyt MF. Stepping into retirement: a postcard from the threshold. *J Clin Psychol* 2015;71:1121–7.
- [20] Huan T, Joehanes R, Song C, Peng F, Guo Y, Mendelson M, et al. Genome-wide identification of DNA methylation QTLs in whole blood highlights pathways for cardiovascular disease. *Nat Commun* 2019;10:4267.
- [21] Si J, Yang S, Sun D, Yu C, Guo Y, Lin Y, et al. Epigenome-wide analysis of DNA methylation and coronary heart disease: a nested case-control study. *Elife* 2021;10:e68671.
- [22] Kou M, Li X, Shao X, Grundberg E, Wang X, Ma H, et al. DNA methylation of birthweight-blood pressure genes and changes of blood pressure in response to weight-loss diets in the POUNDS Lost trial. *Hypertension* 2023;80:1223–30.
- [23] Wand H, Lambert SA, Tamburro C, Iacocca MA, O'Sullivan JW, Sillari C, et al. Improving reporting standards for polygenic scores in risk prediction studies. *Nature* 2021;591:211–9.
- [24] Choi SW, Mak TS, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc* 2020;15:2759–72.
- [25] Nabais MF, Gadd DA, Hannon E, Mill J, McRae AF, Wray NR. An overview of DNA methylation-derived trait score methods and applications. *Genome Biol* 2023;24:28.
- [26] Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum Mol Genet* 2018;27:3641–9.
- [27] Yoon N, Cho YS. Development of a polygenic risk score for BMI to assess the genetic susceptibility to obesity and related diseases in the Korean population. *Int J Mol Sci* 2023;24:11560.
- [28] Huang Y, Ollikainen M, Muniandy M, Zhang T, van Dongen J, Hao G, et al. Identification, heritability, and relation with gene expression of novel DNA methylation loci for blood pressure. *Hypertension* 2020;76:195–205.
- [29] Hamilton OKL, Zhang Q, McRae AF, Walker RM, Morris SW, Redmond P, et al. An epigenetic score for BMI based on DNA methylation correlates with poor physical health and major disease in the Lothian Birth Cohort. *Int J Obes (Lond)* 2019;43:1795–802.
- [30] Shah S, Bonder MJ, Marioni RE, Zhu Z, McRae AF, Zhernakova A, et al. Improving phenotypic prediction by combining genetic and epigenetic associations. *Am J Hum Genet* 2015;97:75–85.
- [31] Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet* 2019;51:584–91.
- [32] Thomas M, Su YR, Rosenthal EA, Sakoda LC, Schmit SL, Timofeeva MN, et al. Combining Asian and European genome-wide association studies of colorectal cancer improves risk prediction across racial and ethnic populations. *Nat Commun* 2023;14:6147.
- [33] Whelton PK, He J, Appel LJ, Cutler JA, Havas S, Kotchen TA, et al. Primary prevention of hypertension: clinical and public health advisory from The National High Blood Pressure Education Program. *JAMA* 2002;288:1882–8.
- [34] Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res* 2007;17:1520–8.
- [35] Privé F, Vilhjálmsson BJ, Aschard H, Blum MGB. Making the most of clumping and thresholding for polygenic scores. *Am J Hum Genet* 2019;105:1213–21.
- [36] Ge T, Chen CY, Ni Y, Feng YA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun* 2019;10:1776.
- [37] Vilhjálmsson BJ, Yang J, Finucane HK, Gusev A, Lindström S, Ripke S, et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am J Hum Genet* 2015;97:576–92.
- [38] Privé F, Arbel J, Vilhjálmsson BJ. LDpred2: better, faster, stronger. *Bioinformatics* 2021;36:5424–31.
- [39] Mak TSH, Porsch RM, Choi SW, Zhou X, Sham PC. Polygenic scores via penalized regression on summary statistics. *Genet Epidemiol* 2017;41:469–80.
- [40] Ruan Y, Lin YF, Feng YA, Chen CY, Lam M, Guo Z, et al. Improving polygenic prediction in ancestrally diverse populations. *Nat Genet* 2022;54:573–80.
- [41] Zhang H, Zhan J, Jin J, Zhang J, Lu W, Zhao R, et al. A new method for multi-ancestry polygenic prediction improves performance across diverse populations. *Nat Genet* 2023;55:1757–68.
- [42] Weissbrod O, Kanai M, Shi H, Gazal S, Peyrot WJ, Khera AV, et al. Leveraging fine-mapping and multipopulation training data to improve cross-population polygenic risk scores. *Nat Genet* 2022;54:450–8.
- [43] Xu L, Zhou G, Jiang W, Zhang H, Dong Y, Guan L, et al. JointPRS: a data-adaptive framework for multi-population genetic risk prediction incorporating genetic correlation. *Nat Commun* 2025;16:3841.
- [44] Zhang J, Zhan J, Jin J, Ma C, Zhao R, O'Connell J, et al. An ensemble penalized regression method for multi-ancestry polygenic risk prediction. *Nat Commun* 2024;15:3238.
- [45] Ma S, Xi B, Yang L, Sun J, Zhao M, Bovet P. Trends in the prevalence of overweight, obesity, and abdominal obesity among Chinese adults between 1993 and 2015. *Int J Obes (Lond)* 2021;45:427–37.
- [46] Pan XF, Wang L, Pan A. Epidemiology and determinants of obesity in China. *Lancet Diabetes Endocrinol* 2021;9:373–92.

- [47] Chen K, Shen Z, Gu W, Lyu Z, Qi X, Mu Y, et al. Prevalence of obesity and associated complications in China: a cross-sectional, real-world study in 15.8 million adults. *Diabetes Obes Metab* 2023;25:3390–9.
- [48] Qi SF, Zhang B, Wang HJ, Yan J, Mi YJ, Liu DW, et al. Prevalence of hypertension subtypes in 2011 and the trends from 1991 to 2011 among Chinese adults. *J Epidemiol Community Health* 2016;70:444–51.
- [49] Wang Z, Chen Z, Zhang L, Wang X, Hao G, Zhang Z, et al. Status of hypertension in China: results from the China Hypertension Survey, 2012–2015. *Circulation* 2018;137:2344–56.
- [50] Zhao Z, Zhang M, Wu J, Xu X, Yin P, Huang Z, et al. E-cigarette use among adults in China: findings from repeated cross-sectional surveys in 2015–16 and 2018–19. *Lancet Public Health* 2020;5:e639–49.
- [51] Zhou L, Cao D, Si Y, Zhu X, Du L, Zhang Y, et al. Income-related inequities of adult obesity and central obesity in China: evidence from the China Health and Nutrition Survey 1997–2011. *BMJ Open* 2020;10:e034288.
- [52] Jones-Smith JC, Gordon-Larsen P, Siddiqi A, Popkin BM. Cross-national comparisons of time trends in overweight inequality by socioeconomic status among women using repeated cross-sectional surveys from 37 developing countries, 1989–2007. *Am J Epidemiol* 2011;173:667–75.
- [53] Doak C, Adair L, Bentley M, Fengying Z, Popkin B. The underweight/overweight household: an exploration of household socio-demographic and dietary factors in China. *Public Health Nutr* 2002;5:215–21.
- [54] Dearth-Wesley T, Wang H, Popkin BM. Under- and overnutrition dynamics in Chinese children and adults (1991–2004). *Eur J Clin Nutr* 2008;62:1302–7.
- [55] Wang H, Du S, Zhai F, Popkin BM. Trends in the distribution of body mass index among Chinese adults, aged 20–45 years (1989–2000). *Int J Obes (Lond)* 2007;31:272–8.
- [56] Monda KL, Adair LS, Zhai F, Popkin BM. Longitudinal relationships between occupational and domestic physical activity patterns and body weight in China. *Eur J Clin Nutr* 2008;62:1318–25.
- [57] Jones-Smith JC, Gordon-Larsen P, Siddiqi A, Popkin BM. Emerging disparities in overweight by educational attainment in Chinese adults (1989–2006). *Int J Obes (Lond)* 2012;36:866–75.
- [58] Wang L, Zhou B, Zhao Z, Yang L, Zhang M, Jiang Y, et al. Body-mass index and obesity in urban and rural China: findings from consecutive nationally representative surveys during 2004–18. *Lancet* 2021;398:53–63.
- [59] Sart G, Bayar Y, Danilina M. Impact of educational attainment and economic globalization on obesity in adult females and males: empirical evidence from BRICS economies. *Front Public Health* 2023;11:1102359.
- [60] Witkam R, Gwinnutt JM, Humphreys J, Gandrup J, Cooper R, Verstappen SMM. Do associations between education and obesity vary depending on the measure of obesity used? A systematic literature review and meta-analysis. *SSM Popul Health* 2021;15:100884.
- [61] Monteiro CA, Conde WL, Popkin BM. Independent effects of income and education on the risk of obesity in the Brazilian adult population. *J Nutr* 2001;131:881S–6S.
- [62] Wang Y, Ye C, Kong L, Zheng J, Xu M, Xu Y, et al. Independent associations of education, intelligence, and cognition with hypertension and the mediating effects of cardiometabolic risk factors: a Mendelian randomization study. *Hypertension* 2023;80:192–203.
- [63] Lu J, Wu C, Zhang X, Yang Y, Cui J, Xu W, et al. Educational inequalities in mortality and their mediators among generations across four decades: nationwide, population based, prospective cohort study based on the ChinaHEART project. *BMJ* 2023;382:e073749.
- [64] Bell CG. *The epigenomic analysis of human obesity*. Obesity (Silver Spring) 2017;25:1471–81.
- [65] Ng SW, Norton EC, Popkin BM. Why have physical activity levels declined among Chinese adults? Findings from the 1991–2006 China Health and Nutrition Surveys. *Soc Sci Med* 2009;68:1305–14.
- [66] Aslani A, Faraji A, Allahverdzadeh B, Fathnezhad-Kazemi A. Prevalence of obesity and association between body mass index and different aspects of lifestyle in medical sciences students: a cross-sectional study. *Nurs Open* 2021;8:372–9.
- [67] Sauerberg M, Klüsener S, Mühlichen M, Grigoriev P. Sex differences in cause-specific mortality: regional trends in seven European countries, 1996–2019. *Eur J Public Health* 2023;33:1052–9.
- [68] Qing H, Desrouleaux R, Israni-Winger K, Mineur YS, Fogelman N, Zhang C, et al. Origin and function of stress-induced IL-6 in murine models. *Cell* 2020;182:372–87.e14.
- [69] Zheng Y, Joyce BT, Hwang SJ, Ma J, Liu L, Allen NB, et al. Association of cardiovascular health through young adulthood with genome-wide DNA methylation patterns in midlife: the CARDIA Study. *Circulation* 2022;146:94–109.
- [70] Yousefi PD, Suderman M, Langdon R, Whitehurst O, Davey Smith G, Relton CL. DNA methylation-based predictors of health: applications and statistical considerations. *Nat Rev Genet* 2022;23:369–83.
- [71] Xu Y, Li H, Wang A, Su Z, Yang G, Luo Y, et al. Association between the metabolically healthy obese phenotype and the risk of myocardial infarction: results from the Kailuan study. *Eur J Endocrinol* 2018;179:343–52.
- [72] Bramlage P, Pittrow D, Wittchen HU, Kirch W, Boehler S, Lehnert H, et al. Hypertension in overweight and obese primary care patients is highly prevalent and poorly controlled. *Am J Hypertens* 2004;17:904–10.
- [73] Li F, Yang CP, Wu CY, Ho CA, Yeh HC, Chan YS, et al. Contribution of body mass index stratification for the prediction of maximal oxygen uptake. *Int J Med Sci* 2022;19:1929–41.
- [74] Williams HC, Burden-Teh E. On the definition of dermatological disease. Part 2: approaches for defining dermatological diseases. *Clin Exp Dermatol* 2022;47:1812–9.
- [75] Fan J, Yu C, Pang Y, Guo Y, Pei P, Sun Z, et al. Adherence to healthy lifestyle and attenuation of biological aging in middle-aged and older Chinese adults. *J Gerontol A Biol Sci Med Sci* 2021;76:2232–41.
- [76] Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 2009;25:1754–60.
- [77] DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;43:491–8.
- [78] Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38:e164.
- [79] Du Z, Ma L, Qu H, Chen W, Zhang B, Lu X, et al. Whole genome analyses of Chinese population and *de novo* assembly of a northern Han genome. *Genomics Proteomics Bioinformatics* 2019;17:229–47.
- [80] Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009;5:e1000529.
- [81] 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 2015;526:68–74.
- [82] Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, et al. An integrated map of structural variation in 2504 human genomes. *Nature* 2015;526:75–81.
- [83] Tian Y, Morris TJ, Webster AP, Yang Z, Beck S, Feber A, et al. ChAMP: updated methylation analysis pipeline for Illumina BeadChips. *Bioinformatics* 2017;33:3982–4.
- [84] Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007;8:118–27.

- [85] Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 2012;28:882–3.
- [86] Zheng SC, Webster AP, Dong D, Feber A, Graham DG, Sullivan R, et al. A novel cell-type deconvolution algorithm reveals substantial contamination by immune cells in saliva, buccal and cervix. *Epigenomics* 2018;10:925–40.
- [87] Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. *Nat Genet* 2018;50:1593–9.
- [88] Kanai M, Akiyama M, Takahashi A, Matoba N, Momozawa Y, Ikeda M, et al. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat Genet* 2018;50:390–400.
- [89] Akiyama M, Okada Y, Kanai M, Takahashi A, Momozawa Y, Ikeda M, et al. Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nat Genet* 2017;49:1458–67.
- [90] Chen T, Chen X, Zhang S, Zhu J, Tang B, Wang A, et al. The Genome Sequence Archive Family: toward explosive data growth and diverse data types. *Genomics, Proteomics Bioinformatics* 2021;19:578–83.
- [91] Ma Y, Zhao X, Jia Y, Han Z, Yu C, Fan Z, et al. The updated Genome Warehouse: enhancing data value, security, and usability to address data expansion. *Genomics Proteomics Bioinformatics* 2025;23:qzaf010.