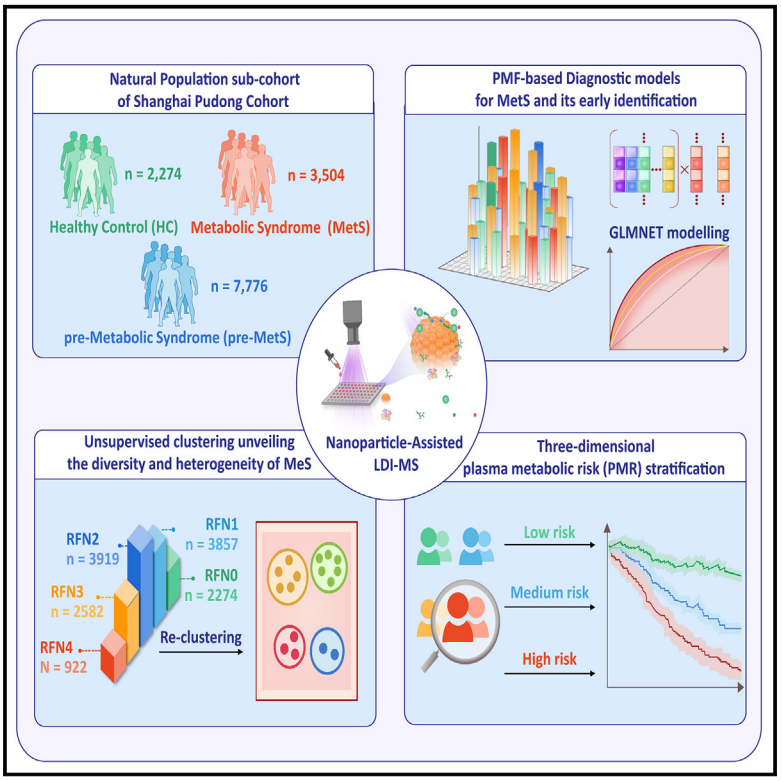
Cell Reports
Medicine

Article

Plasma metabolic fingerprints for large-scale screening and personalized risk stratification of metabolic syndrome

Graphical abstract



Highlights

- Advanced ferric particle-assisted LDI-MS enables plasma metabolomic analysis of MetS
- Plasma metabolic fingerprints (PMFs) highlight the diagnosis of MetS and pre-MetS
- PMFs prompt the assessment of the four significant risk factors contributing to MetS
- Personalized three-dimensional stratification reveals three metabolic risk patterns

Authors

Yifan Chen, Wei Xu, Wei Zhang, ..., Alex F. Chen, Kun Qian, Jun Pu

Correspondence

chenfengyuan@xinhuamed.com.cn (A.F.C.), k.qian@sjtu.edu.cn (K.Q.), pujun310@hotmail.com (J.P.)

In brief

Chen et al. develop diagnostic models that are robust and can be used for the personalized three-dimensional plasma metabolic risk (PMR) stratification of metabolic syndrome (MetS) as well as its early identification (pre-MetS) via plasma metabolic fingerprints. The results show that there are heterogeneous risk patterns in metabolic phenotypes.

Chen et al., 2023, Cell Reports Medicine *4*, 101109
July 18, 2023
https://doi.org/10.1016/j.xcrm.2023.101109

CellPress

Cell Reports Medicine

CellPress
OPEN ACCESS

Article

Plasma metabolic fingerprints for large-scale screening and personalized risk stratification of metabolic syndrome

Yifan Chen,[1,6] Wei Xu,[1,6] Wei Zhang,[1,6] Renyang Tong,[1] Ancai Yuan,[1] Zheng Li,[1] Huiru Jiang,[1] Liuhua Hu,[1] Lin Huang,[3] Yudian Xu,[2] Ziyue Zhang,[2] Mingze Sun,[1] Xiaoxiang Yan,[4] Alex F. Chen,[5,*] Kun Qian,[1,2,*] and Jun Pu[1,7,*]

[1]Division of Cardiology, State Key Laboratory of Systems Medicine for Cancer, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, 160 Pujian Road, Shanghai 200127, China

[2]School of Biomedical Engineering, Institute of Medical Robotics and Institute of Translational Medicine, Shanghai Jiao Tong University, Shanghai 200030, China

[3]Country Department of Clinical Laboratory Medicine, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai 200030, China

[4]Department of Cardiovascular Medicine, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China

[5]Institute for Developmental and Regenerative Cardiovascular Medicine, Xinhua Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200092, China

[6]These authors contributed equally

[7]Lead contact

*Correspondence: chenfengyuan@xinhuamed.com.cn (A.F.C.), k.qian@sjtu.edu.cn (K.Q.), pujun310@hotmail.com (J.P.)

https://doi.org/10.1016/j.xcrm.2023.101109

SUMMARY

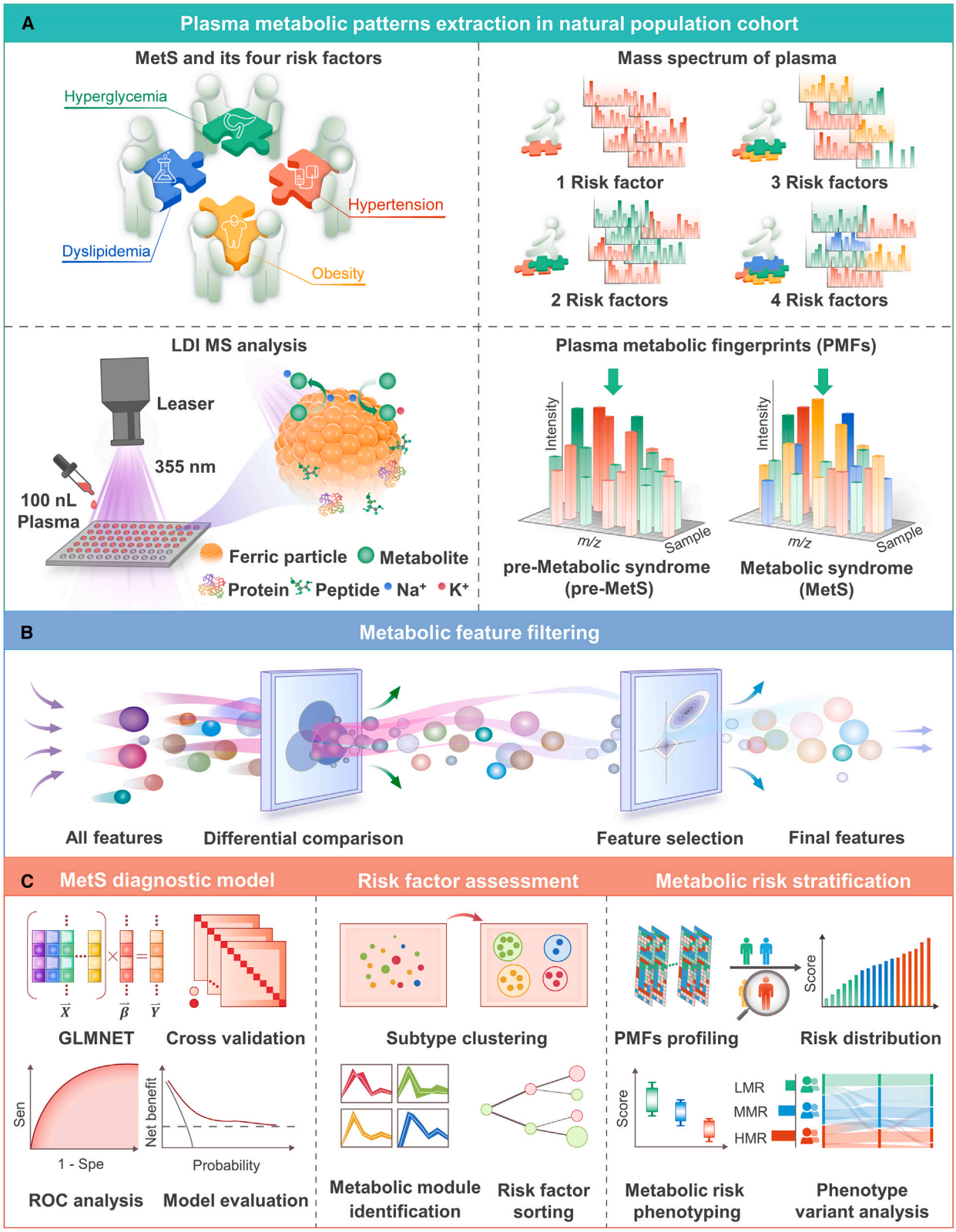
Direct diagnosis and accurate assessment of metabolic syndrome (MetS) allow for prompt clinical interventions. However, traditional diagnostic strategies overlook the complex heterogeneity of MetS. Here, we perform metabolomic analysis in 13,554 participants from the natural cohort and identify 26 hub plasma metabolic fingerprints (PMFs) associated with MetS and its early identification (pre-MetS). By leveraging machine-learning algorithms, we develop robust diagnostic models for pre-MetS and MetS with convincing performance through independent validation. We utilize these PMFs to assess the relative contributions of the four major MetS risk factors in the general population, ranked as follows: hyperglycemia, hypertension, dyslipidemia, and obesity. Furthermore, we devise a personalized three-dimensional plasma metabolic risk (PMR) stratification, revealing three distinct risk patterns. In summary, our study offers effective screening tools for identifying pre-MetS and MetS patients in the general community, while defining the heterogeneous risk stratification of metabolic phenotypes in real-world settings.

INTRODUCTION

The Centers for Disease Control and Prevention estimates that the prevalence of metabolic syndrome (MetS) is 34.7% in the United States[1] and 33.9% in China,[2] with increasing severity in recent decades and an annual growth rate of >8% among subjects >60 years old due to increasing rates of diabetes and obesity.[1,3] MetS is recognized as a progressive chronic pathophysiological state that significantly increases the risks of cardiovascular (odds ratio [OR], 1.80) and cerebrovascular events (OR, 2.05),[4] as well as liver (relative risk [RR], 1.43)[5] and colorectal cancers (RR, 1.25).[6] MetS is diagnosed primarily by the coincident identification of three or more of the following known risk factors, including obesity, hypertension, hyperglycemia, and dyslipidemia.[7–9] As an early sign of MetS, one or two risk factors for MetS are required for it to qualify as a pre-MetS (pre-MS). Although there are at least five different international MetS guidelines,[10–14] none of these definitions have a decisive advantage in predicting cardiovascular events,[15] and the different guidelines showed a discordance of 20%–41% in the same population.[16,17] Thus, there is a need to develop an effective and useful screening platform to promote individualized strategies for the management of MetS.[18,19]

A blood test serves as a common diagnostic method for detecting molecules inside the body. Notably, the comprehensive profiling of small-molecule metabolites, termed the plasma metabolome,[20] provides predictive biomarkers for MetS and a valuable source for understanding the underlying pathophysiologic mechanisms.[21,22] Given the increasing concern about the burden of MetS and the healthcare costs in an aging population, the development of rapid and portable community-based diagnostic screening will provide a powerful window for therapeutic intervention.

Recent studies have promoted the role of metabolomics research in the revolutionary progress of discovering biomarkers for MetS prediction.[23] For example, a study of 58 donors confirmed that 36 metabolites were closely associated with MetS using traditional ^{1}H nuclear magnetic resonance



(legend on next page)

(NMR).[24] Another study of 311 subjects found that plasma 5-hydroxyindole-3-acetic acid concentrations were higher in MetS donors than in those without MetS using liquid chromatography-mass spectrometry (LC-MS).[25] In addition, a study of 163 participants found that 2-hydroxybutyric acid, inositol, and D-glucose served as potential biomarkers for MetS using gas chromatography-mass spectrometry (GC-MS).[26] However, these traditional spectrometry methods require strict pretreatment procedures to separate and enrich the metabolites, limiting their use for large-scale screening (over 10,000 samples) in general population-based cohorts.[27] Despite the importance of these findings, these methods may not precisely reflect the real-world state of MetS, with several potentially important signaling molecules being missed. Recent technological advances have introduced the use of nanoparticle-assisted laser desorption/ionization mass spectrometry (LDI-MS) as a high-throughput screening (≈300 samples/h) tool for accurate metabolic analysis (mass error <50 ppm) with easier sample preparation, thus allowing the identification of specific biomarkers of MetS to reflect the status of the patients. This approach detects MetS metabolites in ultralow volumes without enrichment or purification and produces results after a short processing time.

Figure 1 illustrates how the overall experiment was conducted. We obtained plasma metabolic fingerprints (PMFs), including low-molecular-weight metabolites (100–300 Da), from 13,554 participants, including healthy volunteers, pre-MetS patients (having one or two MetS risk factors), and MetS patients (having at least three MetS risk factors). Through combined LDI-MS detection, we filtered out 26 hub PMFs to develop precise and efficient machine-learning (ML)-based diagnostic platforms for large-scale screening of pre-MetS and MetS, with areas under curves (AUCs) of 0.867 (95% confidence interval [CI], 0.854–0.881) and 0.864 (95% CI, 0.843–0.885) for healthy control (HC) vs. pre-MetS, 0.849 (95% CI, 0.838–0.860) and 0.835 (95% CI, 0.817–0.853) for pre-MetS vs. MetS, and 0.891 (95% CI, 0.880–0.902) and 0.886 (95% CI, 0.868–0.903) for HC vs. MetS in the discovery and validation sets, respectively. We reclassified the 16 subclusters (RF0000–RF1111) into four metabolic phenotypes (MPs) with 26 hub PMFs. By evaluating the contribution of each risk factor to MetS, we highlighted the importance of glycemic control for prevention and intervention of MetS. Finally, we constructed a three-dimensional plasma metabolic risk (PMR) stratification adjusted for age and gender to classify metabolic heterogeneity into three risk patterns. Consistent with the change in conventional clinical parameters and the cumulative all-cause death events by 4-year follow-up adjusted by gender and age (for medium risk, hazard ratio [HR] 1.54, 95% CI 1.05–2.28, p for log-rank test = 0.029. For high risk, HR 1.85, 95% CI 1.22–2.79, p for log-rank test = 0.004), this PMR assessment provided an alternative approach to reflect individual actual metabolic status.

RESULTS

Comprehensive clinical and metabolic characterization of MetS

There were 17,841 participants enrolled in the prospective Shanghai Community Cohort Establishment and Follow-up (NCT04517513) from January to March 2019. After filtering out participants (n = 4,287) according to the flow chart (Figure 2 and STAR Methods), a total of 13,554 (76%) participants (mean age ± standard deviation 67.84 ± 5.90 years; 53.0% females) were finally eligible for inclusion in the cohort study. In the study, we also adopted the statement of the Chinese Diabetes Society 2004 (Table S1)[28,29] to better reflect the Chinese population. To identify potential biomarkers for MetS and enable more granular analysis of the relationship between cardiovascular risk factors and the severity of the syndrome, the current study divided the participants into three groups: a MetS group (with at least three MetS risk factors), a pre-MetS group (with one or two MetS risk factors), and an HC group, including 3,504 (25.9%; age 68.29 ± 5.66 years; 52.9% females), 7,776 (57.4%; age 67.88 ± 5.99 years; 53.0% females), and 2,274 participants (16.8%; age 67.01 ± 5.85 years; 53.2% females), respectively (Figure 2). By this classification, we could better understand the nuanced differences among HC, pre-MetS, and MetS. Table 1 summarizes the demographic and clinical characteristics of the participants. The gender distribution was not significantly different among the three groups. Demographically, there was a stepwise increase across the groups in the proportions of obesity, hypertension, hyperglycemia, and dyslipidemia ($p < 0.001$) (Table 1). Based on the number of MetS risk factors, we further investigated the heterogeneity of MetS development by dividing the participants into five subgroups (RFN0–RFN4), including RFN0 (no risk factors), RFN1(one risk factor), RFN2 (two risk factors), RFN3 (three risk factors), and RFN4 (four risk factors) (Figure 2). Moreover, we classified the cohort into 16 subclusters (RF0000–RF1111) according to the permutations and combinations of the four MetS risk factors, each binary coded by "I" or "0" to indicate their presence or absence in the conditions (Figure 2 and Table 2).

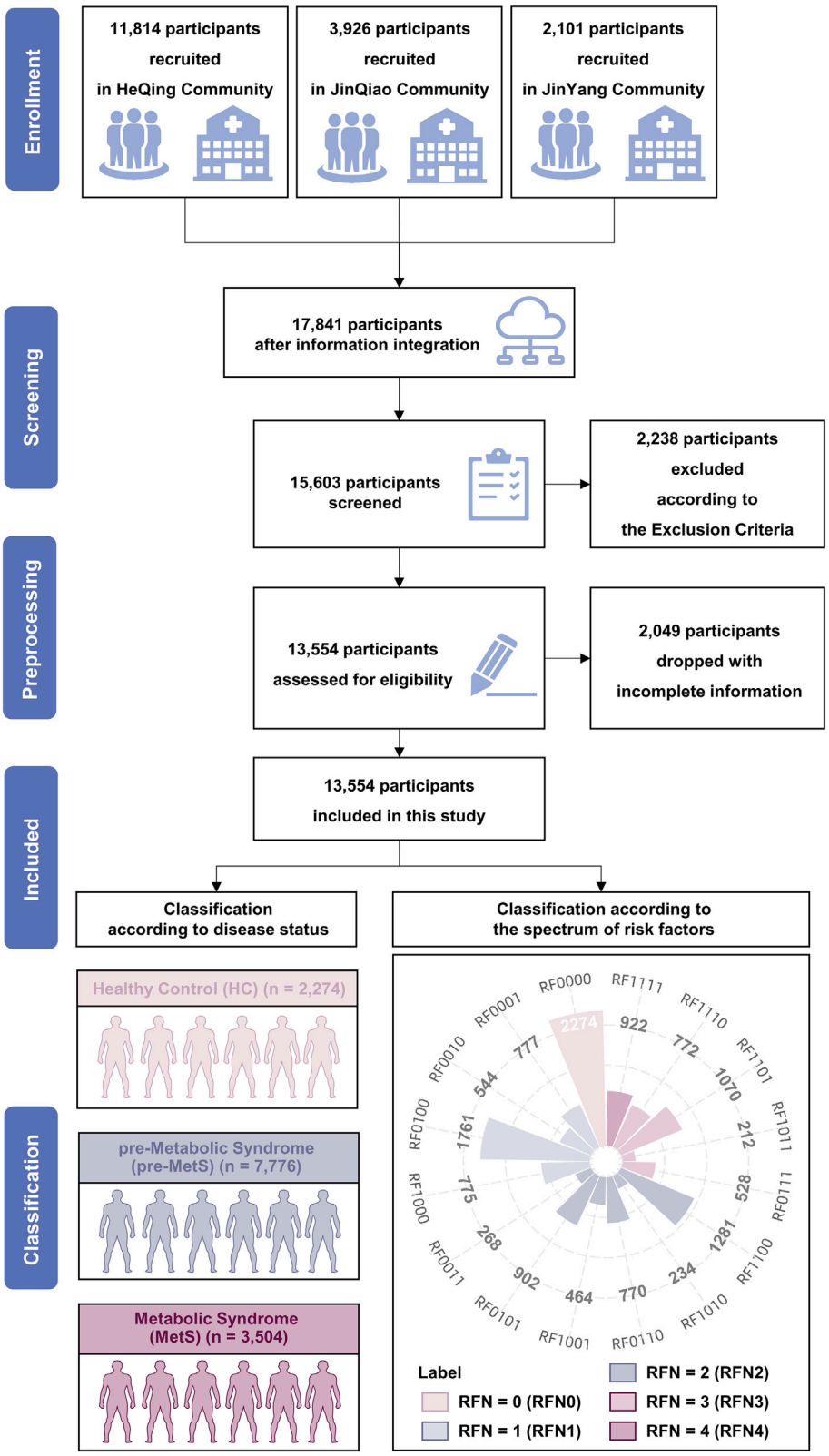
We applied a recently developed ferric particle-assisted LDI-MS technique, followed by an in-house computational pipeline (STAR Methods).[30] This procedure provided an effective detection platform with superior-quality and large-scale biospecimens for subsequent analysis. First, we prepared tailored ferric particles and observed rough-surfaced nanoparticles of uniform size 300 ± 20 nm in diameter (polydispersity index [PDI] of 0.358) by transmission electron microscopy (TEM) (Figure S1A) and

Figure 1. Overall schematics for evaluating MetS among the general aging population based on the combination of PMFs and ML

(A) Schematic workflow for the extraction of PMFs using ferric particle-assisted laser desorption/ionization mass spectrometry (LDI-MS). One hundred nanoliters of native plasma was used for direct analysis without pretreatment procedures. Only Na^+- and K^+-adducted metabolites can be selectively detected with the coexistence of high concentrations of peptides and proteins.

(B) Feature filtering for hub PMFs was carried out according to difference comparisons among the healthy control (HC), pre-MetS, and MetS groups.

(C) PMF-based diagnostic models were constructed for HC vs. pre-MetS, pre-MetS vs. MetS, and HC vs. MetS. Traditional risk factors were assessed by unsupervised clustering of various subtypes of MetS. Finally, PMF-based metabolic risk stratification was computed to define three metabolic risk patterns: low metabolic risk (LMR), moderate metabolic risk (MMR), and high metabolic risk (HMR).



(legend on next page)

dynamic light scattering (DLS) (Figure S1Bi). The ferric particles with elements (including Fe and O) were uniformly distributed in the elemental mapping and scanning electron microscopy (SEM) images shown in Figure S1A. The zeta potential of -15.13 ± 0.15 mV indicated that these ferric particles had a negative charge (Figure S1Bii). Accordingly, ferric particles had a unique extinction spectrum at 350–450 nm (Figure S1Biii) compatible with the LDI-MS laser. Subsequently, we tested different contents of standards, including lysine, D-glucose, and sucrose, and showed reliable results, affording R^2 values of 0.98, 0.92, and 0.92, respectively (Figure S1C). To further confirm the detection performance, we also compared the results of six standards using ferric particles and organic matrices by LDI-MS, including 2,5-dihydroxybenzoic acid (DHB) and α-cyano-4-hydroxycinnamic acid (CHCA). The results showed that the detection limit of standard metabolites for standards obtained by ferric particles was significantly lower than those obtained by DHB and CHCA (Figures S2 and S3 and Table 2). Furthermore, we compared the LDI-MS results of the plasma sample before and after preprocessing among the ferric particles, DHB, and CHCA. The results showed that ferric particle-assisted LDI-MS, with or without pretreatment, could detect more peaks of plasma samples than DHB and CHCA (Figure S4). Regarding detection speed, we achieved rapid LDI-MS detection of 13,554 samples at a speed of $\sim$2 s per sample (2,000 shots at a laser frequency of 1,000 Hz) using an automatic scanning model with a 1 s interval between samples. Overall, the LDI-MS of detection achieved a leap in performance with ferric nanoparticles' help in analytical reproducibility, speed, and throughput.

Top discriminating PMFs associated with pre-MetS and MetS

Exploiting the high-throughput metabolic profiling provided by LDI-MS, we developed a metabolic screening platform for the one-step diagnosis of MetS using small plasma samples. This study recorded the typical mass spectra results for 100 nL of plasma from participants in the HC, pre-MetS, and MetS groups (Figure 3A). For further analysis, we extracted 303 *m/z* signals as PMFs with stable uniformity from the raw mass spectrum (Figure S5).

To balance the sample categories, we randomly selected 4,548, 4,548, and 7,008 participants with 1:1 case-control matching from the general population to build diagnostic models for HC vs. MetS, HC vs. pre-MetS, and pre-MetS vs. MetS, respectively. We randomly split this cohort at a ratio of 0.7 to 0.3 into the discovery (for HC vs. pre-MetS n = 3,184; for pre-MetS vs. MetS n = 4,906; for HC vs. MetS n = 3,184) and validation (for HC vs. pre-MetS n = 1,364; for pre-MetS vs. MetS n = 2,102; for HC vs. MetS n = 1,364) sets, respectively. There were no statistical differences between the gender of controls and cases (Table S3).

Ninety-eight PMFs significantly differed between the HC and MetS groups in the discovery set (two-sided Kruskal-Wallis rank-sum test, $p < 0.05$; Figure 3B). Similarly, 76 differential PMFs were screened between the HC and pre-MetS groups in the discovery set, and 54 differential PMFs were screened between the MetS and pre-MetS groups in the discovery set (two-sided Kruskal-Wallis rank-sum test, $p < 0.05$; Figure 3B). Following Bonferroni correction and Dunnett's test for multiple comparisons among these three groups, 26 differential PMFs were further identified as the top discriminating metabolic patterns (adjusted $p < 0.05$; Figure 3B) and might be potential plasma metabolic biomarkers for diagnosing pre-MetS and MetS.

Hub PMFs optimizing the metabolic classification of MetS through ML

The current large sample sizes (for HC vs. MetS n = 4,548; for HC vs. pre-MetS n = 4,548; for pre-MetS vs. MetS n = 7,008) achieved 99.99% diagnostic power to detect statistical significance and thus produce reliable, valid, and generalizable results (Figure 3C).

We compared the model performance for HC vs. MetS through 5-fold cross-validation among five ML algorithms, including generalized linear models via least absolute shrinkage and selection operator and elastic-net regularization (GLMNET) (mean AUC 0.72), support vector machine (SVM) (mean AUC 0.71), multivariate adaptive regression splines (MARS) (mean AUC 0.71), random forest (RF) (mean AUC 0.69), and adaptive boosting (ADABOOST) (mean AUC 0.67) (Figure 3D). Although the SVM and MARS models showed a high sensitivity (0.70 and 0.64), the SVM model had the lowest specificity (0.60) compared to the MARS and glmnet models (both 0.69) (Figure S6 and Table S4). Due to its combined AUC, sensitivity, and specificity performance, the glmnet algorithm was identified as an appropriate ML algorithm to build diagnostic models for pre-MetS and MetS. Then, we sequentially optimized parameters for HC vs. pre-MetS, pre-MetS vs. MetS, and HC vs. MetS using all 26 hub PMFs adjusted for age and gender under 5-fold cross-validation. The AUCs for HC vs. pre-MetS were 0.867 (95% CI, 0.854–0.881) and 0.864 (95% CI, 0.843–0.885) (one-sided DeLong test; p = 0.813) in the discovery (n = 3,184) and validation sets (n = 1,364), respectively (Figures 3E and S7A). The AUCs for pre-MetS vs. MetS were 0.849 (95% CI, 0.838–0.860) and 0.835 (95% CI, 0.817–0.853) (one-sided DeLong

Figure 2. Overview flow chart of the study design and enrollment

The complete study design consisted of five steps, including enrollment, screening, preprocessing, inclusion, and classification. For enrollment, 11,814, 3,926, and 2,101 individuals were initially recruited in the HeQing, JinQiao, and JinYang communities based on the prospective Shanghai Community Cohort Establishment and Follow-up (NCT04517513). For screening, 2,238 participants were excluded according to the exclusion criteria (see STAR Methods). For preprocessing, 2,049 participants with incomplete information were excluded. Finally, 13,554 participants were included in this study. Participants were classified into the HC (n = 2,274; no use of lipid-lowering, antidiabetic, or antihypertensive drugs), pre-MetS (n = 7,776; clustering of <3 traditional risk factors, including obesity, hypertension, dyslipidemia, and dyslipidemia), and MetS (n = 3,504; clustering of $\geq$3 traditional risk factors) groups according to the statement of the Chinese Diabetes Society 2004. Participants were also classified into five subgroups (RFN1–5) according to the present number of four MetS risk factors or into 16 subclusters (RF0000–RF111) considering a combination of risk factors (0/1 = absence/presence of each risk factor). For more details, see Table 2. RFN indicates the number of traditional MetS risk factors.

Table 1. Baseline characteristics of the analyzed participants in the general population

	HC	pre-MetS	MetS	Overall	p value[a]
Number (%)	2,274 (16.8)	7,776 (57.4)	3,504 (25.9)	13,554 (100.0)	<0.001
Obesity (%)	0 (0.0)	2,754 (35.4)	2,976 (84.9)	5,730 (42.3)	<0.001
Hypertension (%)	0 (0.0)	4,714 (60.6)	3,292 (93.9)	8,006 (59.1)	<0.001
Hyperglycemia (%)	0 (0.0)	1,816 (23.4)	2,434 (69.5)	4,250 (31.4)	<0.001
Dyslipidemia (%)	0 (0.0)	2,411 (31.0)	2,732 (78.0)	5,143 (37.9)	<0.001
Male (%)	1,064 (46.8)	3,651 (47.0)	1,649 (47.1)	6,364 (47.0)	0.98
Age, mean (SD)	67.01 (5.85)	67.88 (5.99)	68.29 (5.66)	67.84 (5.90)	<0.001
BMI, mean (SD)	21.69 (2.11)	24.26 (2.98)	27.15 (2.86)	24.58 (3.34)	<0.001
TC, mean (SD)	4.77 (0.72)	4.92 (0.95)	4.96 (1.05)	4.91 (0.94)	<0.001
GLU, mean (SD)	5.10 (0.52)	5.78 (1.59)	7.16 (2.28)	6.02 (1.83)	<0.001
HDLC, mean (SD)	1.56 (0.38)	1.40 (0.36)	1.27 (0.38)	1.39 (0.38)	<0.001
LDLC, mean (SD)	2.75 (0.63)	2.93 (0.82)	2.97 (0.87)	2.91 (0.81)	<0.001
TGs, mean (SD)	1.02 (0.31)	1.58 (1.04)	2.44 (1.62)	1.71 (1.24)	<0.001
UA, mean (SD)	302.44 (70.21)	333.46 (83.11)	362.62 (86.86)	335.80 (84.35)	<0.001

HC, healthy control; pre-MetS, pre-metabolic syndrome; MetS, metabolic syndrome; SD, standard deviation; BMI, body mass index; TC, serum total cholesterol; GLU, glucose; HDLC, high-density lipoprotein cholesterol; LDLC, low-density lipoprotein cholesterol; TGs, triglycerides; UA, uric acid.

[a]p values calculated by χ^2 tests for categorical data and one-way analysis of variance for parametric data.

test; p = 0.195) in the discovery (n = 4,906) and validation sets (n = 2,012), respectively (Figures 3E and S7C). Other evaluation metrics, including accuracy, F1 score, negative predictive rate, positive predictive rate, specificity, and sensitivity, are shown in Figures S7B and S7D. The specificity of the HC vs. pre-MetS model was 0.99 in both the discovery and validation sets, and the F1 score was 0.88 in the discovery set and 0.87 in the validation set. The specificity of the pre-MetS vs. MetS model was 0.76 and 0.75 in the discovery and validation sets, while its F1 score was 0.78 and 0.77, respectively.

The model for HC vs. MetS (Figures S8A and S8B) showed good diagnostic performance in both the discovery (n = 3,184) and validation (n = 1,364) sets, with AUCs of 0.891 (95% CI, 0.880–0.902) and 0.886 (95% CI, 0.868–0.903) (one-sided DeLong test; p = 0.280), respectively (Figure 3F). In the HC vs. MetS model, the specificity was 0.79 in both the discovery and validation sets, while the F1 scores were 0.79 and 0.80, respectively (Figure S8C). The confusion matrix for the validation set is shown in Figure S8D. Furthermore, we compared the role of 26 hub PMFs with 303 features for HC vs. MetS to verify the representativeness of these hub features. The results showed that the AUCs were 0.885 (95% CI, 0.867–0.903) and 0.886 (95% CI, 0.868–0.903) (one-sided DeLong test; p = 0.967) in the validation sets of the model with 303 hub PMFs and 26 features, respectively (Figure S8E). Comparisons of other model evaluation metrics are shown in Figure S8F. The p values did not differ significantly between the two groups, demonstrating that hub PMFs were representative.

To effectively utilize the entire dataset to evaluate the performance of the PMFs-based diagnostic model, we conducted 100 independent and randomized 1:1 case-control matching sample selections for model evaluation. In the 100 repeated trials, each sample from the MetS group was selected at least 47 times (Figure S9A); 95% of the discovery set had AUC values in the range of [0.842, 0.869], and 95% of the validation set had AUC values in the range of [0.828, 0.867] (Figure S9B). Moreover, we randomly and independently split the discovery and validation sets 100 times using the same dataset as in Figure 3F. The results showed that 95% of the discovery set had AUC values in the range of [0.884, 0.896], and 95% of the validation set had AUC values in the range of [0.871, 0.899] (Figure S9C). These results indicate that the PMFs-based MetS diagnostic model is robust through the evaluation using the entire dataset.

Moreover, we compared the performance of GLMNET and categorical boosting (CatBoost) models in distinguishing between HC and MetS. The CatBoost model showed a significant difference between discovery and validation AUC ($p < 0.001$) (Figure S9D), suggesting overfitting. However, the advantage of CatBoost models is their ability to analyze multi-class labels. We also compared the AUC of the GLMNET model and the CatBoost model on the discovery set and found no statistically significant difference between the two models (p = 0.756) in diagnosing MetS (Figure S9E). Furthermore, we separately analyzed the importance of each feature using the coefficients of the GLMNET model and the Shapley Additive Explanation (SHAP) values of the CatBoost model (Figures S10A and S10B). We also compared the coefficients and SHAP values, finding that five variables showed positive effects in both evaluation methods, while four showed negative effects (Figure S10C).

Given the confounding effects of drugs, we conducted a sensitivity analysis to eliminate the potential impact of medication on our results. We collected medication information on participants in both the discovery and validation sets. In the discovery set, medication information was available for 58.4%, and it was available for 56.4% in the validation set (Figure S11A). Conventional principal-component analysis (PCA) was unable to differentiate between MetS patients who were using medication and those who were not (Figure S11B). We also evaluated the performance of our diagnostic model separately for the medication and non-medication groups. The AUC for distinguishing MetS from HC in the medication group was 0.920 (95% CI, 0.903–0.936) in the discovery set and 0.922 (95% CI, 0.898–0.946) in the validation set, with a p value of 0.890 (Figure S11C). The AUC for the non-medication group was 0.861 (95% CI, 0.843–0.879) in the discovery set and 0.863 (95% CI,

Table 2. Distribution and prevalence of 16 subcluster conditions in this study

	Obesity	Hypertension	Hyperglycemia	Dyslipidemia	Number (%)[a]
RF0000	0	0	0	0	2,274 (16.78)
RF0001	0	0	0	1	777 (5.73)
RF0010	0	0	1	0	544 (4.01)
RF0100	0	1	0	0	1,761 (12.99)
RF1000	1	0	0	0	775 (5.72)
RF0011	0	0	1	1	268 (1.97)
RF0101	0	1	0	1	902 (6.65)
RF1001	1	0	0	1	464 (3.42)
RF0110	0	1	1	0	770 (5.68)
RF1010	1	0	1	0	234 (1.73)
RF1100	1	1	0	0	1,281 (9.45)
RF0111	0	1	1	1	528 (3.90)
RF1011	1	0	1	1	212 (1.56)
RF1101	1	1	0	1	1,070 (7.89)
RF1110	1	1	1	0	772 (5.69)
RF1111	1	1	1	1	922 (6.80)

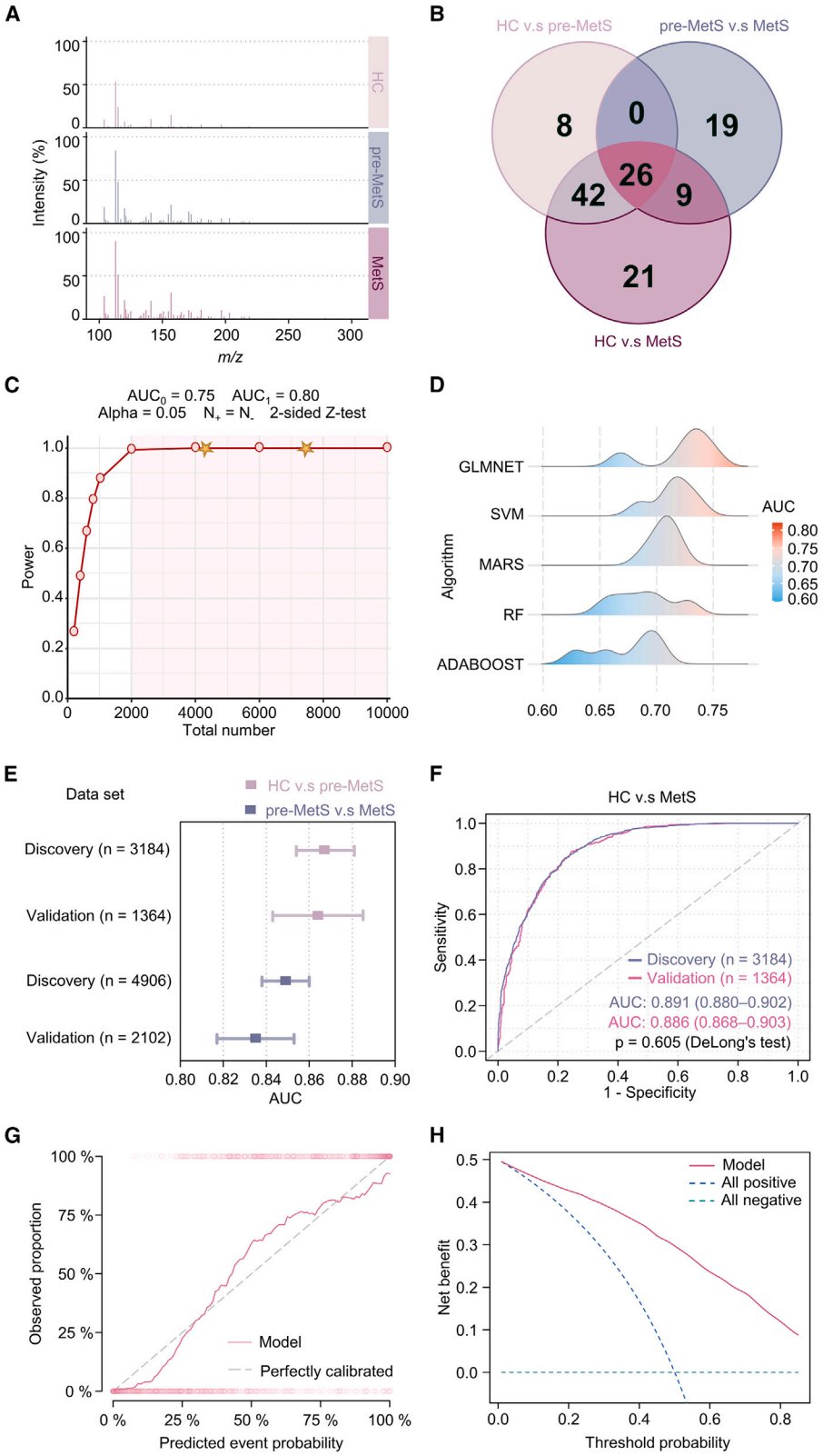
0/1 = absence/presence of each risk factor; RF0000–RF1111, different combinations of four risk factors.
[a]Data are presented as the number/total number (percentage) of subjects unless otherwise indicated.

0.836–0.890) in the validation set, with a p value of 0.923 (Figure S11C). Other model evaluation indicators also show good performance (Figure S11D). Previous literature has suggested that patients adhering to medication regimens exhibit alterations in metabolic profiles.[31–33] In this study, we observed a significant difference in the AUC within the validation set between the non-medication and medication groups, yielding a p value of 0.002, which implies a potential influence of medication on metabolism. Nevertheless, this effect does not facilitate a distinct PCA-based separation between MetS patients in the medication and non-medication groups (Figure S11B). These results underscore the robustness of the PMF-based diagnostic model for MetS, accounting for diverse real-world scenarios and including both medication and non-medication groups, thereby facilitating clinical diagnosis in practical healthcare settings.

Next, we examined whether the diagnostic model based on PMFs could serve as a clinical tool to define MetS. To that end, we introduced calibration curves to summarize model calibration over the observed range of predicted probabilities and assess the accuracy of the predictions.[34,35] MetS cases predicted by this PMF-based model were in high agreement with the actual observations in the calibration plot (Figure 3G). Additionally, this model increased the net benefit rate through decision-curve analysis (DCA) (Figure 3H). Overall, by comparing the plasma profiles of donors with distinct risk factors, we constructed reliable ML-based classifiers for pre-MetS and MetS using 26 hub PMFs. These PMF-based diagnostic models exhibited excellent discrimination, calibration, and clinical utility independent of traditional biochemical examinations.

Re-clustering MetS subtypes optimized the assessment of different MetS risk factors

To investigate the heterogeneity of four risk factors and their combinations in MetS, we calculated the average intensity of 26 hub PMFs in five subgroups (RFN0–RFN4) sorted by the number of risk factors (Figure 4A). However, using only the mean intensity did not distinguish the heterogeneity of the subgroups well. Therefore, we divided five subgroups into 16 subclusters (RF0000–RF1111) according to the composition of MetS risk factors. The population numbers for each subcluster are summarized in Figure 4B. We used two unsupervised ML strategies, K-means and hierarchical clustering, to reclassify the metabolic heterogeneity of 16 subclusters after dimensionality reduction by PCA. The first two principal components (PC1 and 2) explained over 96.4% of the total variation in the entire cohort (Figure 4C). The results showed that both analytical strategies classified these 16 subclusters into the same four MPs (Figures 4C and 4D). The optimal number of clusters was calculated according to the gap statistic (Figure S12 and Table S5). The subclusters of unitary dyslipidemia (RF0001) and dyslipidemia combined with hypertension (RF0101) clustered together (MP1 in Figures 4C and 4D). All subclusters containing obesity, except those with the complication of hyperglycemia (RF1000, RF1001, RF1101, RF1100), were highly consistent with the HC subcluster (RF0000) and the subcluster of unitary hypertension (RF0100) (MP2 in Figures 4C and 4D). MP3 contained the subclusters of obesity combined with hyperglycemia (RF1010), obesity combined with hypertension and hyperglycemia (RF1110), and hyperglycemia combined with dyslipidemia (RF0011) (MP3 in Figures 4C and 4D). The remaining five subclusters, including unitary hyperglycemia (RF0010), hyperglycemia combined with hypertension (RF0110), hyperglycemia combined with obesity and dyslipidemia (RF1011), and a combination of all four risk factors (RF111), were classified as MP4 (MP4 in Figures 4C and 4D). Based on these results validated with two different algorithms, our PMF-based metabolic reclassification provided the first real-world evidence of the metabolic phenotypic diversity of the traditional composition of four MetS



(legend on next page)

risk factors based on a large community cohort study of over 10,000 people.

Moreover, to precisely interpret the proportion of each risk factor involved in MetS, we re-clustered the five MetS subtypes (RF1110, RF1011, RF1101, RF0111, and RF1111) as well as the HC group (RF0000) using the K-means algorithm to clarify the contributions of the different risk factors to the pathogenesis of MetS. The MetS subtype with risk factors for obesity, hypertension, and dyslipidemia (RF1101) showed the highest similarity to the HC group (RF0000), suggesting that the missing risk factor, hyperglycemia, had the greatest impact on chronic metabolic disorders in the general population (Figure 4E). Similarly, hypertension had a greater effect on MetS than dyslipidemia (RF1011 vs. RF1110) or obesity (RF1011 vs. RF0111), as demonstrated by comparing groups in turn according to the degree of clustering similarity (Figure 4E). The contribution of dyslipidemia to MetS was similar to that of obesity (RF1110 vs. RF0111) (Figure 4E). The clustering results also divided the 26 hub PMFs into four intensity-specific metabolic modules (modules 1–4), which contained 2, 2, 16, and 6 hub PMFs, respectively (Figure 4E). The relative intensity of these modules among the five MetS subtypes and the HC group is depicted in Figure S13. Based on these results, the contributions of all the MetS risk factors could be sorted in descending order: hyperglycemia, hypertension, dyslipidemia, and obesity.

Individualized PMR stratification mirroring actual metabolic dysfunction

Compared to clinical parameters, plasma metabolites could better reflect an individual's disease progression and metabolic dysfunction.[36,37] This study constructed a personalized three-dimensional PMR stratification using 26 hub PMFs and CatBoost (an algorithm for gradient boosting on decision trees) in the general population (Figure 5). The composite PMR scores comprised three-dimensional indexes (indexes 1–3). To calculate indexes 1–3, we first determined each participant's mean predicted probability scores for HC, pre-MetS, and MetS. Then, we used these mean probability scores as cutoff values to categorize each participant into one of the three metabolic patterns.

The score of index 1 was closely related to the subgroup with no risk factors (RFN0) (Figure 5A) and the HC group (Figure 5B). The score of index 2 was extremely high in the subgroup with one or two risk factors (RFN1 and 2) (Figure 5A) and the pre-MetS group (Figure 5B), while, in other groups, it was close to zero. The score of index 3 was only sensitive to the subgroup with three or four risk factors (RFN3 and 4) (Figure 5A) and the MetS group (Figure 5B). These results suggested that the three indexes characterized different aspects of MetS. By calculating the priority of the three index scores, we reclassified the cohort (n = 13,554) into three metabolic risk patterns: low (LMR), medium (MMR), and high metabolic risk (HMR) patterns (Figure 5C). The LMR (n = 2,371), MMR (n = 7,674), and HMR (n = 3,509) patterns had the highest index 1, index 2, and index 3 scores, respectively.

Compared to individuals with the MMR pattern, those with the HMR pattern showed greater relative increases in serum creatinine (SCr), serum total cholesterol (TC), low-density lipoprotein cholesterol (LDLC), uric acid (UA), body mass index (BMI), glucose (GLU), and triglycerides (TGs) when the expression in the LMR pattern was taken as the baseline (Figure 5D and Table S6), which is consistent with clinical practice. Additionally, the HMR group exhibited a more pronounced relative decrease in high-density lipoprotein cholesterol (HDLC) (Figure 5D and Table S6), suggesting that HDLC is a protective factor for metabolic disorders. We also calculated scores for index 1, which reflects health status, and index 3, which reflects disease status, in the five MetS subtypes (RF1110, RF1011, RF1101, RF0111, and RF1111) as well as in the HC group (RF0000) to reassess the metabolic heterogeneity of MetS. The results showed that index 1 scores increased sequentially in the RF0000, RF1101, RF1011, RF1110, RF0111, and RF1111 subtypes (Figures 4C and 5E), while index 3 scores decreased sequentially in the same subtypes (Figure 5E). This ordering of MetS subtypes was consistent with the results of the K-means cluster analysis described above (Figure 5E), which further supports reliability and robustness of our PMR-based stratification model.

To investigate the clinical relevance of PMR-based stratification, we collected all-cause death events in the entire cohort within 4 years, including 291 (2.1%) all-cause death events as

Figure 3. Diagnostic models for pre-MetS and MetS by 26 hub PMFs adjusted for age and gender using ML

(A) Typical mass spectrometry spectra within an *m/z* range of 100–300 obtained by ferric particle-assisted LDI-MS of plasma samples from the HC, pre-MetS, and MetS groups.

(B) The final 26 hub PMFs were filtered out by comparison of differential PMFs among HC vs. pre-MetS, pre-MetS vs. MetS, and HC vs. MetS groups using Kruskal-Wallis rank-sum test and Bonferroni/Dunnett correction in the discovery cohorts.

(C) Power analysis of the diagnosis of pre-MetS and MetS using a two-sided Z test. AUC_0 and AUC_1 are the areas under the receiver operating characteristic (ROC) curves (AUCs) for the null and alternative hypotheses, respectively. N_+ and N_- are the numbers of items sampled from cases and controls, respectively. The stars indicate the numbers in our datasets (n = 4,548, 4,548, and 7,008) for classification among HC vs. pre-MetS, HC vs. MetS, and pre-MetS vs. MetS, respectively.

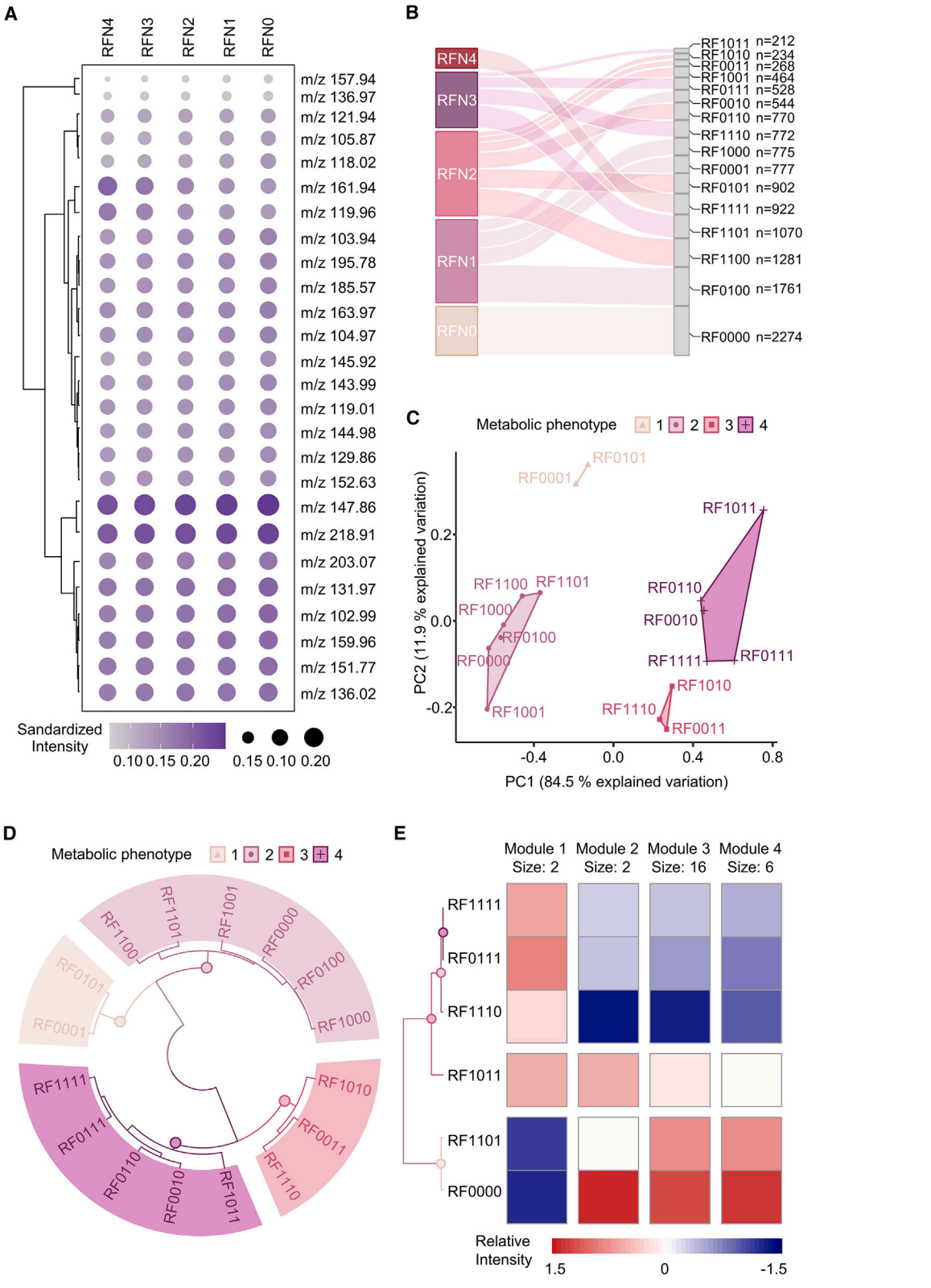
(D) Distribution of the AUC using generalized linear models via least absolute shrinkage and selection operator and elastic-net regularization (GLMNET), support vector machine (SVM), multivariate adaptive regression splines (MARS), random forest (RF), and adaptive boosting (ADABOOST) to distinguish between HC and MetS groups in the validation cohort (n = 1,364).

(E) Distribution of the AUCs of HC vs. pre-MetS and pre-MetS vs. MetS in both the discovery and validation sets (HC vs. pre-MetS in pink and pre-MetS vs. MetS in purple).

(F) ROC curves for the PMF-based MetS diagnostic model using the GLMNET algorithm to distinguish between MetS and HC in the discovery (n = 3,184) and validation sets (n = 1,364).

(G) Calibration curves for our model showed good correlation between predicted and observed outcomes. The calibration curve was close to the 45° perfectly calibrated line.

(H) DCA plot depicting the standardized net benefit of our model.



(legend on next page)

major points and 28 (0.2%) accidental death events as censors. Based on the traditional disease classification, the status was found to be significantly associated with a higher incidence of adverse events without adjusting for gender and age (for pre-MetS, HR 1.58, 95% CI 1.08–2.30, p for log-rank test = 0.018; for MetS, HR 1.78, 95% CI 1.18–2.66, p for log-rank test = 0.006) (Figure S14B). However, statistical significance disappeared for pre-MetS (adjust p = 0.063) (Figure 5F) after excluding age and gender as confounders. This finding implies that our conventional disease staging, which heavily relies on age and gender, may not adequately reflect the metabolic status of the patient. Notably, PMR-based stratification was also significantly related to worse event accumulation without gender and age correction (for medium risk, HR 1.76, 95% CI 1.19–2.59, p for log-rank test = 0.004; for high risk, HR 2.06, 95% CI 1.36–3.11, p for log-rank test = 0.001) (Figure S14A). Using the above PMR-based stratification model, we also revealed that patients with medium risk had a significantly worse event accumulation than those with low risk adjusted for gender and age (HR 1.54, 95% CI 1.05–2.28, p for log-rank test = 0.029). The patients with high risk had a significantly worse event accumulation than those with low risk adjusted for gender and age (HR 1.85, 95% CI 1.22–2.79, p for log-rank test = 0.004) (Figure 5F), suggesting the independent relationship between this PMR-based stratification and all-cause death events.

We further investigated the association between the traditional risk grouping and PMR-based metabolic patterns using variant analysis to illustrate the changes (Table S7). The RFN0 subgroup (n = 2,274) could be reclassified into LMR (n = 2,188, 96.2%), MMR (n = 74 3.3%), and HMR (n = 12 0.5%) patterns (Table S7). Similarly, the variant analysis of novel risk patterns and existing traditional classifications based on disease states is summarized in Table S7. The PMF-based metabolic risk patterns were at least 95.58% consistent with the traditional classification following the number of risk factors. These results indicated that the conventional determination of MetS status was too ambiguous to precisely reflect an individual's metabolic dysfunction, especially in the general population. Even a homogeneous subgroup of MetS patients (concerning traditional risk factors) had heterogeneous risk patterns of metabolic dysfunction. Therefore, individualized risk stratification and preventive measures are necessary for MetS patients. Our PMR stratification with personalized metabolic patterns and three-dimensional PMR scores provides a robust and feasible solution for accurate metabolic risk stratification.

To identify the specific metabolic pathways driving these risk factors in MetS, we matched 26 hub PMFs to the Human Metabolome Database (HMDB) and classified them into organic acids (n = 12), organoheterocyclic compounds (n = 4), benzenoids (n = 3), fatty acyls (n = 2), carbohydrates (n = 2), and nucleic acids (n = 1) (Figure S15A and Table S8). Recent studies have highlighted the role of lipid metabolites with large molecular weights in the pathological process of MetS.[38,39] However, our advanced LDI-MS revealed the importance of small-molecule metabolites in the diagnosis and risk stratification of MetS, thereby promoting the potential clinical application in the early identification of pre-MetS.

Using pathway topology analysis, we further computed the metabolic pathway enrichment of the 26 hub PMFs to regulate the process of MetS. Four differential metabolic pathways (enrichment ratio >5 and $p < 0.05$), including taurine and hypotaurine metabolism, phenylacetate metabolism, homocysteine degradation, and phosphatidylethanolamine biosynthesis, were identified to be involved in the development of MetS (Figure S15B and Table S9). These findings suggest that small-molecule metabolites play potential roles in MetS development and could serve as new targets for attenuating MetS risk.

DISCUSSION

Metabolic risk factors present significant global challenges, necessitating effective early intervention strategies.[40] In this study, we employed ferric particle-assisted LDI-MS to identify potential biomarkers for pre-MetS and MetS in a community-based setting. While prior research primarily focused on discovering new metabolites associated with MetS, our approach emphasized the heterogeneity of metabolic dysfunctions through ML-based integrated PMRs scoring rather than relying solely on traditional risk factors. Using unsupervised clustering, we successfully characterized MetS subtypes based on four distinct metabolic modules, revealing the varying contributions of hyperglycemia, hypertension, dyslipidemia, and obesity (Figure 4). This finding was further supported by the correlation between clinical parameters and the reclassified metabolic risk patterns (Figure 5). These results provide a novel and scalable screening approach for identifying MetS patients, mainly targeting elderly individuals at risk of metabolic dysfunction and facilitating interventions that effectively address modifiable risk factors.

Metabolomics is a powerful tool for precision medicine in biomarkers. However, metabolomics-based biomarkers for MetS are still in their infancy. Previous studies have primarily focused on analyzing MetS patients in bulk using small sample sizes[23,41] and employing univariate analysis methods.[42] These studies aimed to identify distinct population-level risk factors rather than investigating the phenotypic consequences from the perspective of integrated PMFs. Our study represents the largest investigation of PMFs in MetS patients, utilizing a general

Figure 4. PMF-based unsupervised ML revealed the heterogeneity of MetS

(A) The standardized intensity of all hub PMFs among five subgroups (RFN0–RFN4) according to the number of risk factors.

(B) The distribution of five subgroups (RFN0–RFN4) and 16 subclusters (RF0000–RF1111) classified according to traditional risk factors in the general population.

(C) K-means clustering analysis scatter diagram regrouping 16 subclusters into four metabolic phenotypes (MPs) based on the results of PCA. Each point represents a subcluster condition containing (or not) the risk factor according to Table 2. MPs are indicated by colored shading.

(D) Circular hierarchical cluster analysis dendrogram grouping 16 subclusters into the same four MPs based on phenotypic similarity. Colors of subcluster names based on the MPs.

(E) Relative risk assessment among the five MetS subgroups (RF0111, RF1011, RF1101, RF1110, RF1111) and HC group through K-means clustering analysis based on the relative intensity of all 26 hub PMFs. These PMFs were divided into four metabolic feature modules (modules 1–4).

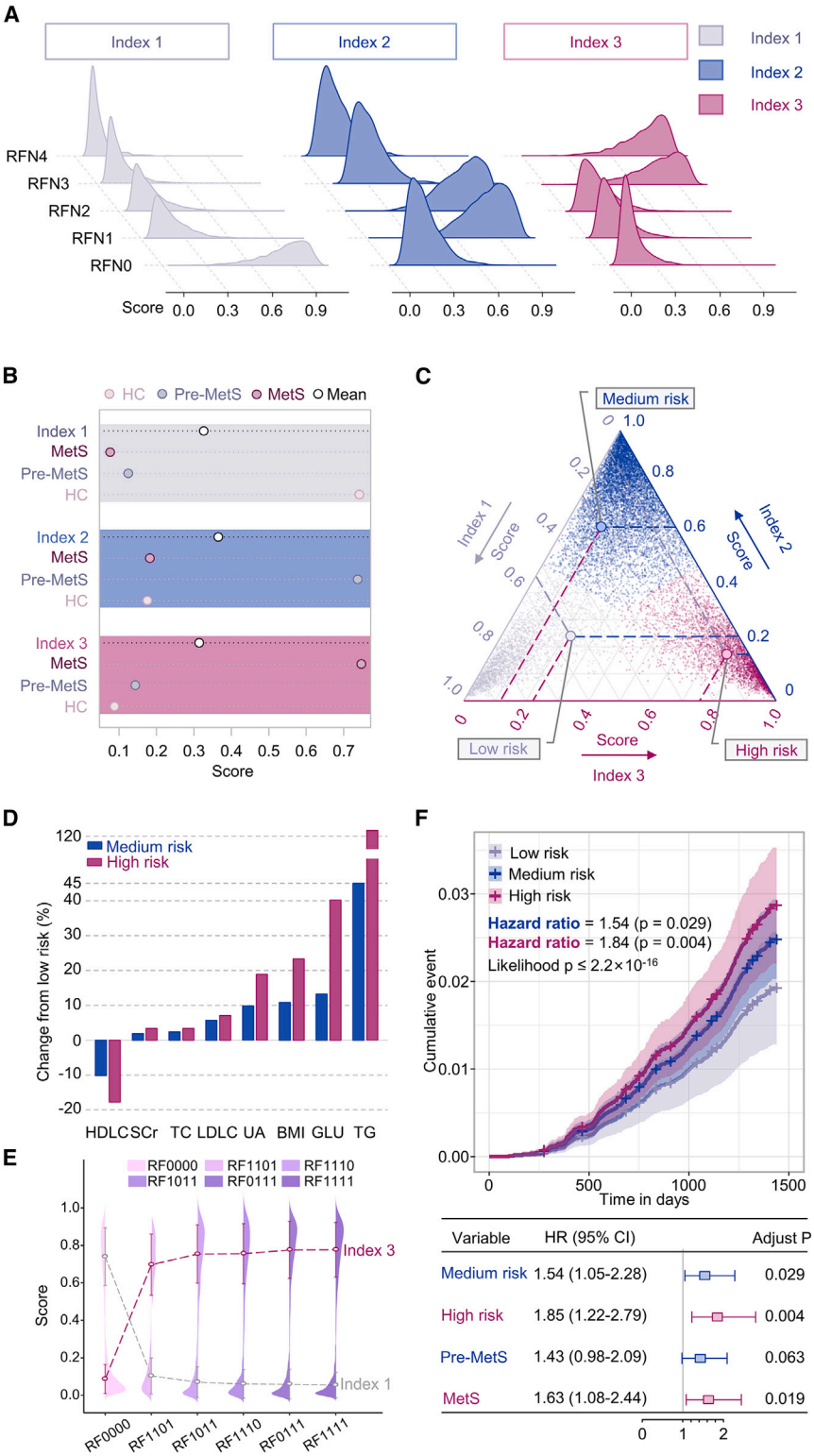


Figure 5. Construction of the three-dimensional PMR stratification for evaluating individual metabolic heterogeneity

(A) Distribution of the three PMR indexes (index 1–3) among the five subgroups (RFN0–RFN4) classified according to the present number of traditional risk factors.

(B) Mean scores of these three indexes (index 1–3) in the HC, pre-MetS, and MetS groups.

(C) Specific PMR patterns for all participants in the general population cohort. Gray, blue, and purple dots indicate individual status ranked by LMR, MMR, and HMR patterns, respectively.

(D) The relative changes in high-density lipoprotein cholesterol (HDLC), serum creatinine (SCr), serum total cholesterol (TC), low-density lipoprotein cholesterol (LDLC), uric acid (UA), body mass index (BMI), glucose (GLU), and triglycerides (TGs) in the MMR and HMR groups compared with those in the LMR group.

(E) Mean scores for index 1 (in gray) and index 3 (in purple) in the five MetS subtypes (RF1110, RF1011, RF1101, RF0111, and RF1111) and the HC group (RF0000).

(F) Cumulative curves and forest plots of the 4-year mortality events for 13,554 patients with three PMR statuses stratified by gender and age. The p value for multivariate Cox regression analysis models was calculated by the likelihood test. The p value for variables was obtained by a log-rank test and hazard ratio (HR).

community cohort of 13,554 individuals and achieving a remarkable diagnostic power of 99.99% (Table 1 and Figure 3C). A previous cross-sectional serum metabolomic study on MetS in 2021 included only 1,800 participants.[43] To ensure generalizability, we employed independent discovery and validation sets (Table S3).

Most MetS analyses have utilized NMR,[24] LC–MS,[25] or GC-MS[18] methods. However, conventional spectrometric techniques have been reported[27] to be challenging to achieve the required diagnostic performance using high-throughput quantitative metabolomic analysis of ultralow-volume samples when using these individual platforms separately to detect large numbers of samples (Table S10). NMR requires a pretreatment step and large sample volume (~microliters) to enhance sensitivity (~picomoles) by expanding nuclear-level rotation. On the other hand, GC-MS/LC-MS methods necessitate longer sophisticated sample handling (~hours) to reduce sample complexity.[44–46] In contrast, LDI-MS, as an alternative metabolite recognition method, offers a high-performance technique that enables large-scale MetS screening.

The traditional multi-process diagnosis of MetS is challenging to promote in community screening due to its need for cumbersome physical examination and clinical indicators. In addition, complex examinations of four risk factors using different instruments can result in significant measurement errors and low efficiency, particularly in large-scale screening programs.[16] Plasma metabolites provide a fundamental view into the dynamics of MetS.[47] Recent studies[41,48,49] focused on comparing MetS and non-MetS and therefore did not reveal the heterogeneity of metabolic disorders. Unlike the traditional classification, the current study divided the participants into three groups, a MetS group (n = 3,504), a pre-MetS group (n = 7,776), and an HC group (n = 2,274), to achieve widespread screening of MetS in the community. We systematically validated PMF-based classifiers for pre-MetS and MetS with high performance. The AUC for pre-MetS vs. HC was 0.867 (95% CI, 0.854–0.881) and 0.864 (95% CI, 0.843–0.885) in the discovery (n = 3,184) and validation (n = 1,364) set (Figures 3E and S7A). The AUC for MetS vs. HC was 0.891 (95% CI, 0.880–0.902) and 0.886 (95% CI, 0.868–0.903) in the discovery (n = 3,184) and validation (n = 1,364) set, respectively (Figure 3F). The AUC for pre-MetS vs. MetS was 0.849 (95% CI, 0.838–0.860) and 0.835 (95% CI, 0.817–0.853) in the discovery (n = 4,906) and validation (n = 2,012) sets (Figures 3E and S7C). These results were superior to prior research of 1,800 participants in 2021 (AUC 0.750, 95% CI 0.698–0.802).[43] Moreover, we further confirmed the favorable consistency through the calibration plot (Figure 3G), and the DCA suggested that our model had a reasonable clinical net benefit (Figure 3H). Therefore, these advanced classifiers based on PMFs demonstrated excellent discrimination, calibration, and clinical utility, which can pave the way for a future community screening of pre-MetS and MetS.

Obesity, hypertension, hyperglycemia, and dyslipidemia represent dominant risk factors for MetS development.[50] However, the specific contributions of each risk factor to metabolic impairment progression remain largely unknown.[51] Here, we proposed a metabolomics-based approach to evaluate the contribution of various risk factors using different unsupervised clustering methods. Through unsupervised analysis, we found that hyperglycemia emerged as the dominant risk factor in MetS. Metabolic disturbance phenotypes observed in MetS patients with hyperglycemia were similar to those with four risk factors (Figure 4E). We selected the 26 PMFs to explain more than 96.4% of total variation (Figure 4C). These selected PMFs effectively captured the underlying patterns and aligned with the theory of MetS development, where hyperglycemia plays a crucial role as a direct indicator of insulin resistance.[52] Additionally, the result shows obesity with smaller contributions to MetS (Figure 4E), suggesting that obesity is not synonymous, and metabolically benign obesity in MetS may be induced by insulin resistance.[53,54] These results provide systematic evidence from a metabolomics perspective and offer a solid theoretical foundation for optimizing cost-effective interventions tailored to patients with different risk factors.

The lack of consistent cutoff points among different diagnostic criteria[55] and the significant heterogeneity in coronary heart disease (CHD) risk among individuals with MetS[56,57] indicate that current diagnostic criteria do not correlate with an individual's metabolic status. Here, we aimed to focus on individual-level metabolic diversity and identify personalized metabolic heterogeneity. We established an integrated three-dimensional PMR stratification that considered age and gender (Figure 5). Thus, significant increases in SCr, TC, LDLC, UA, BMI, GLU, and TGs were observed in the HMR pattern compared to the MMR pattern (Figure 5D and Table S6), which aligned with clinical practices. Additionally, HDLC alteration declined in the HMR pattern (Figure 5D and Table S6), indicating that HDLC play a protective role to some extent against metabolic disorders.

To verify PMR stratification, we conducted a 4-year death event follow-up in large prospective cohort. Compared with the traditional disease classification, PMR stratification was independently related to all-cause death events in this large retrospective cohort of 13,554 participants (for medium risk, HR 1.54, 95% CI 1.05–2.28, p for log-rank test = 0.029; for high risk, HR 1.85, 95% CI 1.22–2.79, p for log-rank test = 0.004). Thus, our study highlights that risk stratification and preventive measures are person specific in the development of MetS.

We conducted a comparative analysis between PMF-based molecular typing and traditional typing of MetS using flow analysis (Table S7). The results indicated that the conventional determination of MetS status based on different risk factors was too ambiguous to reflect an individual's metabolic dysfunction. Even a homogeneous subgroup of MetS patients (concerning traditional risk factors) had heterogeneous risk patterns of metabolic dysfunction. The integrated PMR stratification with personalized metabolic patterns and three-dimensional PMR scores provides a robust and feasible solution for accurate metabolic risk stratification, which could not only mirror the actual individual MPs but also replace the traditional classification method based on the number of risks.

Limitations of the study

Several limitations should also be considered in our study. First, potential bias might limit the results of our study from being replicated in other settings. Therefore, we involved a relatively large number of participants among four communities controlling for sampling error to minimize drawbacks. Moreover, the data were randomly partitioned into discovery and validation sets in

a 7:3 ratio. Another limitation is that the study lacks potential experimental validation, which is important in establishing the reliability and robustness of the findings. Future efforts will focus on validating the models in different centers and conducting experimental verification through animal experiments.

Conclusions

Using PMF of MetS with LDI-MS, we established portable and efficient diagnostic models for pre-MetS and MetS with improved performance for clinical promotion in the general population. Furthermore, we assessed the contributions of different MetS risk factors based on specific metabolic modules using unsupervised clustering. We demonstrated the importance of glycemic control for the prevention and treatment of MetS. Finally, we devised a personalized three-dimensional PMR stratification to decode the individual metabolic heterogeneity into three metabolic patterns related to all-cause death in our large prospective cohort. Overall, our results introduce novel paradigms and interventional directions to accelerate the application of precision medicine for pre-MetS and MetS.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Participants and ethics
  - Exclusion criteria
  - Definition of metabolic syndrome (MetS)
  - Definition of pre-metabolic syndrome (pre-MS)
  - Baseline examination
  - Follow-up
  - Sample collection, preparation, and storage
- METHOD DETAILS
  - Matrix synthesis
  - Material characterization methods
  - Nano-assisted laser desorption/ionization mass spectrometry (LDI MS) profiling
  - Hub differential plasma metabolic fingerprints (PMFs) screening
  - Power analysis
  - Data sets splitting
  - Model selection
  - PMFs-based diagnostic models
  - Metabolomic clustering and assessment of MetS risk factors
  - Construction and validation of plasma metabolic risk (PMR) stratification
  - Pathway analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrm.2023.101109.

ACKNOWLEDGMENTS

The authors gratefully thank the National Natural Science Foundation of China (82230014, 81930007, 81930007, 81800307, 81971771), Medical-Engineering Joint Funds of Shanghai Jiao Tong University (YG2021ZD09, YG2022QN107, YG2023ZD08), National Science Fund for Distinguished Young Scholars (81625002), Joint Project of Health Commission of Shanghai Pudong District (PW2019D-11), Science and Technology Commission of Shanghai Municipality (22JC1402100, 22QA1405400, 22DZ2292400, 201409005200), Shanghai Municipal Health Commission (2022JC013), Shanghai Cancer Institute (ZZ-20-22SYL), Shanghai Hospital Development Center (SHDC12022102), Shanghai Institutions of Higher Learning project (2021-01-07-00-02-E00083), National Research Center for Translational Medicine Shanghai (TMSK-2021-124, NRCTM(SH)-2021-06), and National High Level Hospital Clinical Research Funding (2022-PUMCH-C-023) for financial support. This work was also financially supported by grants 2022YFE0103500, 2021YFF0703500, 2021YFA0910100, and 2022YFC2502801 from the National Ministry of Science and Technology (MOST). We thank all the participants, the physicians, nurses, and research staff for their contributions to the Shanghai Community Cohort Establishment and Follow-up (NCT04517513). The authors extend their heartfelt appreciation to the staff members, including Xiaonan Kang, Zhuqing Wang, Kai Luo, Yumeng Liu, and others, from the Biobank of Renji Hospital.

AUTHOR CONTRIBUTIONS

Conceptualization, J.P., K.Q., and A.C.; cohort construction and investigation, W.Z., R.T., A.Y., Z.L., H.J., and L.H.; data analysis, Y.C., W.X., L.H., Y.X., Z.Z., and M.S.; resources, Y.C. and W.X.; data curation, B.M. and X.Z.; writing – original draft, Y.C., X.W., and W.Z.; writing – review & editing, J.P., K.Q., and A.C.; visualization, Y.C. and W.X.; supervision, J.P., K.Q., and A.C.; project administration, J.P.; funding acquisition, J.P.

DECLARATION OF INTERESTS

The authors have filed patents for both the technology and the use of the technology to analyze biofluid samples.

INCLUSION AND DIVERSITY

We worked to ensure gender balance in the recruitment of human subjects. We worked to ensure ethnic or other types of diversity in the recruitment of human subjects. We worked to ensure that the study questionnaires were prepared in an inclusive way. While citing references scientifically relevant for this work, we also actively worked to promote gender balance in our reference list. We avoided "helicopter science" practices by including the participating local contributors from the region where we conducted the research as authors on the paper.

Received: December 22, 2022
Revised: April 1, 2023
Accepted: June 16, 2023
Published: July 18, 2023

REFERENCES

1. Hirode, G., and Wong, R.J. (2020). Trends in the Prevalence of Metabolic Syndrome in the United States, 2011-2016. JAMA *323*, 2526–2528. https://doi.org/10.1001/jama.2020.4501.
2. Lu, J., Wang, L., Li, M., Xu, Y., Jiang, Y., Wang, W., Li, J., Mi, S., Zhang, M., Li, Y., et al. (2017). Metabolic Syndrome Among Adults in China: The 2010

China Noncommunicable Disease Surveillance. J. Clin. Endocrinol. Metab. *102*, 507–515. https://doi.org/10.1210/jc.2016-2477.

3. Gurka, M.J., Golden, S.H., Musani, S.K., Sims, M., Vishnu, A., Guo, Y., Cardel, M., Pearson, T.A., and DeBoer, M.D. (2017). Independent associations between a metabolic syndrome severity score and future diabetes by sex and race: the Atherosclerosis Risk In Communities Study and Jackson Heart Study. Diabetologia *60*, 1261–1270. https://doi.org/10.1007/s00125-017-4267-6.

4. Wang, Y., Li, J., Zheng, X., Jiang, Z., Hu, S., Wadhera, R.K., Bai, X., Lu, J., Wang, Q., Li, Y., et al. (2018). Risk Factors Associated With Major Cardiovascular Events 1 Year After Acute Myocardial Infarction. JAMA Netw. Open *1*, e181079. https://doi.org/10.1001/jamanetworkopen.2018.1079.

5. Esposito, K., Chiodini, P., Colao, A., Lenzi, A., and Giugliano, D. (2012). Metabolic syndrome and risk of cancer: a systematic review and meta-analysis. Diabetes Care *35*, 2402–2411. https://doi.org/10.2337/dc12-0336.

6. Chen, H., Zheng, X., Zong, X., Li, Z., Li, N., Hur, J., Fritz, C.D., Chapman, W., Jr., Nickel, K.B., Tipping, A., et al. (2021). Metabolic syndrome, metabolic comorbid conditions and risk of early-onset colorectal cancer. Gut *70*, 1147–1154. https://doi.org/10.1136/gutjnl-2020-321661.

7. Grundy, S.M. (2016). Metabolic syndrome update. Trends Cardiovasc. Med. *26*, 364–373. https://doi.org/10.1016/j.tcm.2015.10.004.

8. Huang, P.L. (2009). A comprehensive definition for metabolic syndrome. Dis. Model. Mech. *2*, 231–237. https://doi.org/10.1242/dmm.001180.

9. Saklayen, M.G. (2018). The Global Epidemic of the Metabolic Syndrome. Curr. Hypertens. Rep. *20*, 12. https://doi.org/10.1007/s11906-018-0812-z.

10. Grundy, S.M., Cleeman, J.I., Daniels, S.R., Donato, K.A., Eckel, R.H., Franklin, B.A., Gordon, D.J., Krauss, R.M., Savage, P.J., Smith, S.C., Jr., et al. (2005). Diagnosis and management of the metabolic syndrome: an American Heart Association/National Heart, Lung, and Blood Institute Scientific Statement. Circulation *112*, 2735–2752. https://doi.org/10.1161/CIRCULATIONAHA.105.169404.

11. Alberti, K.G.M.M., Eckel, R.H., Grundy, S.M., Zimmet, P.Z., Cleeman, J.I., Donato, K.A., Fruchart, J.C., James, W.P.T., Loria, C.M., Smith, S.C., Jr., et al. (2009). Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. Circulation *120*, 1640–1645. https://doi.org/10.1161/CIRCULATIONAHA.109.192644.

12. Alberti, K.G.M.M., Zimmet, P., and Shaw, J.; IDF Epidemiology Task Force Consensus Group (2005). The metabolic syndrome–a new worldwide definition. Lancet *366*, 1059–1062. https://doi.org/10.1016/S0140-6736(05)67402-8.

13. Expert Panel on Detection, E., and Treatment of High Blood Cholesterol in, Expert Panel on Detection Evaluation and Treatment of High Blood Cholesterol in Adults (2001). Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, And Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III). JAMA 285, 2486-2497. https://doi.org/10.1001/jama.285.19.2486.

14. Alberti, K.G., and Zimmet, P.Z. (1998). Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. Diabet. Med. *15*, 539–553. https://doi.org/10.1002/(SICI)1096-9136(199807)15:7<539::AID-DIA668>3.0.CO;2-S.

15. Hosseinpanah, F., Asghari, G., Barzin, M., Golkashani, H.A., and Azizi, F. (2013). Prognostic impact of different definitions of metabolic syndrome in predicting cardiovascular events in a cohort of non-diabetic Tehranian adults. Int. J. Cardiol. *168*, 369–374. https://doi.org/10.1016/j.ijcard.2012.09.037.

16. Athyros, V.G., Ganotakis, E.S., Tziomalos, K., Papageorgiou, A.A., Anagnostis, P., Griva, T., Kargiotis, K., Mitsiou, E.K., Karagiannis, A., and Mikhailidis, D.P. (2010). Comparison of four definitions of the metabolic syndrome in a Greek (Mediterranean) population. Curr. Med. Res. Opin. *26*, 713–719. https://doi.org/10.1185/03007991003590597.

17. Herath, H.M.M., Weerasinghe, N.P., Weerarathna, T.P., and Amarathunga, A. (2018). A Comparison of the Prevalence of the Metabolic Syndrome among Sri Lankan Patients with Type 2 Diabetes Mellitus Using WHO, NCEP-ATP III, and IDF Definitions. Int. J. Chronic Dis. *2018*, 7813537. https://doi.org/10.1155/2018/7813537.

18. Kassi, E., Pervanidou, P., Kaltsas, G., and Chrousos, G. (2011). Metabolic syndrome: definitions and controversies. BMC Med. *9*, 48. https://doi.org/10.1186/1741-7015-9-48.

19. Sperling, L.S., Mechanick, J.I., Neeland, I.J., Herrick, C.J., Després, J.P., Ndumele, C.E., Vijayaraghavan, K., Handelsman, Y., Puckrein, G.A., Araneta, M.R.G., et al. (2015). The CardioMetabolic Health Alliance: Working Toward a New Care Model for the Metabolic Syndrome. J. Am. Coll. Cardiol. *66*, 1050–1067. https://doi.org/10.1016/j.jacc.2015.06.1328.

20. Holmes, E., Loo, R.L., Stamler, J., Bictash, M., Yap, I.K.S., Chan, Q., Ebbels, T., De Iorio, M., Brown, I.J., Veselkov, K.A., et al. (2008). Human metabolic phenotype diversity and its association with diet and blood pressure. Nature *453*, 396–400. https://doi.org/10.1038/nature06882.

21. Alshehry, Z.H., Mundra, P.A., Barlow, C.K., Mellett, N.A., Wong, G., McConville, M.J., Simes, J., Tonkin, A.M., Sullivan, D.R., Barnes, E.H., et al. (2016). Plasma Lipidomic Profiles Improve on Traditional Risk Factors for the Prediction of Cardiovascular Events in Type 2 Diabetes Mellitus. Circulation *134*, 1637–1650. https://doi.org/10.1161/CIRCULATIONAHA.116.023233.

22. Ussher, J.R., Elmariah, S., Gerszten, R.E., and Dyck, J.R.B. (2016). The Emerging Role of Metabolomics in the Diagnosis and Prognosis of Cardiovascular Disease. J. Am. Coll. Cardiol. *68*, 2850–2870. https://doi.org/10.1016/j.jacc.2016.09.972.

23. Monnerie, S., Comte, B., Ziegler, D., Morais, J.A., Pujos-Guillot, E., and Gaudreau, P. (2020). Metabolomic and Lipidomic Signatures of Metabolic Syndrome and its Physiological Components in Adults: A Systematic Review. Sci. Rep. *10*, 669. https://doi.org/10.1038/s41598-019-56909-7.

24. Li, Y., Wang, Y., Zhuang, Y., Zhang, P., Chen, S., Asakawa, T., and Gao, B. (2020). Serum Metabolomic Profiles Associated With Untreated Metabolic Syndrome Patients in the Chinese Population. Clin. Transl. Sci. *13*, 1271–1278. https://doi.org/10.1111/cts.12817.

25. Fukui, M., Tanaka, M., Toda, H., Asano, M., Yamazaki, M., Hasegawa, G., Imai, S., and Nakamura, N. (2012). High plasma 5-hydroxyindole-3-acetic acid concentrations in subjects with metabolic syndrome. Diabetes Care *35*, 163–167. https://doi.org/10.2337/dc11-1619.

26. Lin, Z., Vicente Gonçalves, C.M., Dai, L., Lu, H.M., Huang, J.H., Ji, H., Wang, D.S., Yi, L.Z., and Liang, Y.Z. (2014). Exploring metabolic syndrome serum profiling based on gas chromatography mass spectrometry and random forest models. Anal. Chim. Acta *827*, 22–27. https://doi.org/10.1016/j.aca.2014.04.008.

27. Wishart, D.S. (2016). Emerging applications of metabolomics in drug discovery and precision medicine. Nat. Rev. Drug Discov. *15*, 473–484. https://doi.org/10.1038/nrd.2016.32.

28. Xiang, K., Xiang, H., Yang, W., Jia, W., Qian, R., and Weng, J. (2004). Diagnosis and management of the metabolic syndrome: a Chinese Diabetes Society Scientific Statement. Chin J Diabetes, 156–161. (in Chinese).

29. Xu, H., Song, Y., You, N.C., Zhang, Z.F., Greenland, S., Ford, E.S., He, L., and Liu, S. (2010). Prevalence and clustering of metabolic risk factors for type 2 diabetes among Chinese adults in Shanghai, China. BMC Publ. Health *10*, 683. https://doi.org/10.1186/1471-2458-10-683.

30. Huang, L., Wang, L., Hu, X., Chen, S., Tao, Y., Su, H., Yang, J., Xu, W., Vedarethinam, V., Wu, S., et al. (2020). Machine learning of serum metabolic patterns encodes early-stage lung adenocarcinoma. Nat. Commun. *11*, 3556. https://doi.org/10.1038/s41467-020-17347-6.

31. Forslund, S.K., Chakaroun, R., Zimmermann-Kogadeeva, M., Markó, L., Aron-Wisnewsky, J., Nielsen, T., Moitinho-Silva, L., Schmidt, T.S.B., Falony, G., Vieira-Silva, S., et al. (2021). Combinatorial, additive and dose-dependent drug-microbiome associations. Nature *600*, 500–505. https://doi.org/10.1038/s41586-021-04177-9.

32. Guo, L., Milburn, M.V., Ryals, J.A., Lonergan, S.C., Mitchell, M.W., Wulff, J.E., Alexander, D.C., Evans, A.M., Bridgewater, B., Miller, L., et al. (2015). Plasma metabolomic profiles enhance precision medicine for volunteers of normal health. Proc. Natl. Acad. Sci. USA *112*, E4901–E4910. https://doi.org/10.1073/pnas.1508425112.

33. Matsumoto, M., Harada, S., Iida, M., Kato, S., Sata, M., Hirata, A., Kuwabara, K., Takeuchi, A., Sugiyama, D., Okamura, T., and Takebayashi, T. (2021). Validity Assessment of Self-reported Medication Use for Hypertension, Diabetes, and Dyslipidemia in a Pharmacoepidemiologic Study by Comparison With Health Insurance Claims. J. Epidemiol. *31*, 495–502. https://doi.org/10.2188/jea.JE20200089.

34. Amini, B., Bassett, R.L., Jr., Haygood, T.M., McEnery, K.W., and Richardson, M.L. (2020). Confidence Calibration: An Introduction With Application to Quality Improvement. J. Am. Coll. Radiol. *17*, 620–628. https://doi.org/10.1016/j.jacr.2019.12.009.

35. Alba, A.C., Agoritsas, T., Walsh, M., Hanna, S., Iorio, A., Devereaux, P.J., McGinn, T., and Guyatt, G. (2017). Discrimination and Calibration of Clinical Prediction Models: Users' Guides to the Medical Literature. JAMA *318*, 1377–1384. https://doi.org/10.1001/jama.2017.12126.

36. Fan, Y., Li, Y., Chen, Y., Zhao, Y.J., Liu, L.W., Li, J., Wang, S.L., Alolga, R.N., Yin, Y., Wang, X.M., et al. (2016). Comprehensive Metabolomic Characterization of Coronary Artery Diseases. J. Am. Coll. Cardiol. *68*, 1281–1293. https://doi.org/10.1016/j.jacc.2016.06.044.

37. Chen, X.W., Ding, G., Xu, L., and Li, P. (2021). A glimpse at the metabolic research in China. Cell Metabol. *33*, 2122–2125. https://doi.org/10.1016/j.cmet.2021.09.014.

38. Wang-Sattler, R., Yu, Z., Herder, C., Messias, A.C., Floegel, A., He, Y., Heim, K., Campillos, M., Holzapfel, C., Thorand, B., et al. (2012). Novel biomarkers for pre-diabetes identified by metabolomics. Mol. Syst. Biol. *8*, 615. https://doi.org/10.1038/msb.2012.43.

39. Suvitaival, T., Bondia-Pons, I., Yetukuri, L., Pöhö, P., Nolan, J.J., Hyötyläinen, T., Kuusisto, J., and Orešič, M. (2018). Lipidome as a predictive tool in progression to type 2 diabetes in Finnish men. Metabolism *78*, 1–12. https://doi.org/10.1016/j.metabol.2017.08.014.

40. Yifan, C., and Jun, P. (2020). Understanding the Clinical Features of Coronavirus Disease 2019 From the Perspective of Aging: A Systematic Review and Meta-Analysis. Front. Endocrinol. *11*, 557333. https://doi.org/10.3389/fendo.2020.557333.

41. Shim, K., Gulhar, R., and Jialal, I. (2019). Exploratory metabolomics of nascent metabolic syndrome. J. Diabet. Complicat. *33*, 212–216. https://doi.org/10.1016/j.jdiacomp.2018.12.002.

42. Antonio, L., Wu, F.C.W., O'Neill, T.W., Pye, S.R., Carter, E.L., Finn, J.D., Rutter, M.K., Laurent, M.R., Huhtaniemi, I.T., Han, T.S., et al. (2015). Associations between sex steroids and the development of metabolic syndrome: a longitudinal study in European men. J. Clin. Endocrinol. Metab. *100*, 1396–1404. https://doi.org/10.1210/jc.2014-4184.

43. Wu, Q., Li, J., Sun, X., He, D., Cheng, Z., Li, J., Zhang, X., Xie, Y., Zhu, Y., and Lai, M. (2021). Multi-stage metabolomics and genetic analyses identified metabolite biomarkers of metabolic syndrome and their genetic determinants. EBioMedicine *74*, 103707. https://doi.org/10.1016/j.ebiom.2021.103707.

44. Garcia-Perez, I., Posma, J.M., Serrano-Contreras, J.I., Boulangé, C.L., Chan, Q., Frost, G., Stamler, J., Elliott, P., Lindon, J.C., Holmes, E., and Nicholson, J.K. (2020). Identifying unknown metabolites using NMR-based metabolic profiling techniques. Nat. Protoc. *15*, 2538–2567. https://doi.org/10.1038/s41596-020-0343-3.

45. Bian, Y., Zheng, R., Bayer, F.P., Wong, C., Chang, Y.C., Meng, C., Zolg, D.P., Reinecke, M., Zecha, J., Wiechmann, S., et al. (2020). Robust, reproducible and quantitative analysis of thousands of proteomes by micro-flow LC-MS/MS. Nat. Commun. *11*, 157. https://doi.org/10.1038/s41467-019-13973-x.

46. Alseekh, S., Aharoni, A., Brotman, Y., Contrepois, K., D'Auria, J., Ewald, J., C Ewald, J., Fraser, P.D., Giavalisco, P., Hall, R.D., et al. (2021). Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. Nat. Methods *18*, 747–756. https://doi.org/10.1038/s41592-021-01197-1.

47. Liu, B., Chen, G., Zhao, R., Huang, D., and Tao, L. (2021). Temporal trends in the prevalence of metabolic syndrome among middle-aged and elderly adults from 2011 to 2015 in China: the China health and retirement longitudinal study (CHARLS). BMC Publ. Health *21*, 1045. https://doi.org/10.1186/s12889-021-11042-x.

48. Surowiec, I., Noordam, R., Bennett, K., Beekman, M., Slagboom, P.E., Lundstedt, T., and van Heemst, D. (2019). Metabolomic and lipidomic assessment of the metabolic syndrome in Dutch middle-aged individuals reveals novel biological signatures separating health and disease. Metabolomics *15*, 23. https://doi.org/10.1007/s11306-019-1484-7.

49. Olszanecka, A., Kawecka-Jaszcz, K., and Czarnecka, D. (2016). Association of free testosterone and sex hormone binding globulin with metabolic syndrome and subclinical atherosclerosis but not blood pressure in hypertensive perimenopausal women. Arch. Med. Sci. *12*, 521–528. https://doi.org/10.5114/aoms.2016.59925.

50. Asgharnezhad, M., Joukar, F., Naghipour, M., Nikbakht, H.A., Hassanipour, S., Arab-Zozani, M., and Mansour-Ghanaei, F. (2020). Exploratory factor analysis of gender-based metabolic syndrome components: Results from the PERSIAN Guilan cohort study (PGCS). Clin. Nutr. ESPEN *40*, 252–256. https://doi.org/10.1016/j.clnesp.2020.09.011.

51. Nichols, G.A., and Moler, E.J. (2011). Metabolic syndrome components are associated with future medical costs independent of cardiovascular hospitalization and incident diabetes. Metab. Syndr. Relat. Disord. *9*, 127–133. https://doi.org/10.1089/met.2010.0105.

52. Diabetes Canada Clinical Practice Guidelines Expert Committee; Punthakee, Z., Goldenberg, R., and Katz, P. (2018). Definition, Classification and Diagnosis of Diabetes, Prediabetes and Metabolic Syndrome. Can. J. Diabetes *42* (*Suppl 1*), S10–S15. https://doi.org/10.1016/j.jcjd.2017.10.003.

53. Engin, A. (2017). The Definition and Prevalence of Obesity and Metabolic Syndrome. Adv. Exp. Med. Biol. *960*, 1–17. https://doi.org/10.1007/978-3-319-48382-5_1.

54. Ortega, F.B., Lee, D.C., Katzmarzyk, P.T., Ruiz, J.R., Sui, X., Church, T.S., and Blair, S.N. (2013). The intriguing metabolically healthy but obese phenotype: cardiovascular prognosis and role of fitness. Eur. Heart J. *34*, 389–397. https://doi.org/10.1093/eurheartj/ehs174.

55. Suzuki, T., Zeng, Z., Zhao, B., Wei, Z., Tanabe, M., Shimbo, T., Kajio, H., Kato, N., and Naruse, M. (2016). Comparison of coronary heart disease risk among four diagnostic definitions of metabolic syndrome. J. Endocrinol. Invest. *39*, 1337–1346. https://doi.org/10.1007/s40618-016-0538-1.

56. Hoang, K.C., Ghandehari, H., Lopez, V.A., Barboza, M.G., and Wong, N.D. (2008). Global coronary heart disease risk assessment of individuals with the metabolic syndrome in the. Diabetes Care *31*, 1405–1409. https://doi.org/10.2337/dc07-2087.

57. Zhou, J., Gao, Q., Wang, J., Zhang, M., Ma, J., Wang, C., Chen, H., Peng, X., and Hao, L. (2018). Comparison of coronary heart disease risk assessments among individuals with metabolic syndrome using three diagnostic definitions: a cross-sectional study from China. BMJ Open *8*, e022974. https://doi.org/10.1136/bmjopen-2018-022974.

58. Sud, M., Fahy, E., Cotter, D., Azam, K., Vadivelu, I., Burant, C., Edison, A., Fiehn, O., Higashi, R., Nair, K.S., et al. (2016). Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. Nucleic Acids Res. *44*, D463–D470. https://doi.org/10.1093/nar/gkv1042.

59. World Medical Association (2013). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA *310*, 2191–2194. https://doi.org/10.1001/jama.2013.281053.

60. de las Fuentes, L., Brown, A.L., Mathews, S.J., Waggoner, A.D., Soto, P.F., Gropler, R.J., and Dávila-Román, V.G. (2007). Metabolic syndrome is associated with abnormal left ventricular diastolic function independent of left ventricular mass. Eur. Heart J. *28*, 553–559. https://doi.org/10.1093/eurheartj/ehl526.

61. Vidigal, F.d.C., Ribeiro, A.Q., Babio, N., Salas-Salvadó, J., and Bressan, J. (2015). Prevalence of metabolic syndrome and pre-metabolic syndrome in health professionals: LATINMETS Brazil study. Diabetol. Metab. Syndrome *7*, 6. https://doi.org/10.1186/s13098-015-0003-x.

62. Gesteiro, E., Megía, A., Guadalupe-Grau, A., Fernandez-Veledo, S., Vendrell, J., and González-Gross, M. (2021). Early identification of metabolic syndrome risk: A review of reviews and proposal for defining pre-metabolic syndrome status. Nutr. Metabol. Cardiovasc. Dis. *31*, 2557–2574. https://doi.org/10.1016/j.numecd.2021.05.022.

63. Zhu, Y., Qiao, L., Prudent, M., Bondarenko, A., Gasilova, N., Möller, S.B., Lion, N., Pick, H., Gong, T., Chen, Z., et al. (2016). Sensitive and fast identification of bacteria in blood samples by immunoaffinity mass spectrometry for quick BSI diagnosis. Chem. Sci. *7*, 2987–2995. https://doi.org/10.1039/c5sc04919a.

64. Obuchowski, N.A., and McClish, D.K. (1997). Sample size determination for diagnostic accuracy studies involving binormal ROC curve indices. Stat. Med. *16*, 1529–1542. https://doi.org/10.1002/(sici)1097-0258(19970715)16:13<1529::aid-sim565>3.0.co;2-h.

65. Vickers, A.J., van Calster, B., and Steyerberg, E.W. (2019). A simple, step-by-step guide to interpreting decision curve analysis. Diagn. Progn. Res. *3*, 18. https://doi.org/10.1186/s41512-019-0064-7.

66. Hancock, J.T., and Khoshgoftaar, T.M. (2020). CatBoost for big data: an interdisciplinary review. J. Big Data *7*, 94. https://doi.org/10.1186/s40537-020-00369-8.

67. Wishart, D.S., Guo, A., Oler, E., Wang, F., Anjum, A., Peters, H., Dizon, R., Sayeeda, Z., Tian, S., Lee, B.L., et al. (2022). HMDB 5.0: the Human Metabolome Database for 2022. Nucleic Acids Res. *50*, D622–D631. https://doi.org/10.1093/nar/gkab1062.

68. Pang, Z., Chong, J., Zhou, G., de Lima Morais, D.A., Chang, L., Barrette, M., Gauthier, C., Jacques, P.É., Li, S., and Xia, J. (2021). MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. Nucleic Acids Res. *49*, W388–W396. https://doi.org/10.1093/nar/gkab382.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Human plasma samples	Shanghai Renji Hospital, China	N/A
Chemicals, peptides, and recombinant proteins		
Bovine serum albumin (BSA, purity>98%)	Sigma, USA	Cat #V900933
2,5-dihydroxybenzoic acid (DHB, purity>98%)	Sigma, USA	Cat # 149357
α-cyano-4-hydroxycinnamic acid (CHCA, purity>99%)	Sigma, USA	Cat # 476870
acetonitrile (ACN, purity>99%)	Sigma, USA	Cat # 110086
cetyltrimethylammonium chloride (CTAC, purity>99%)	Sigma, USA	Cat # 52366
dithiothreitol (DTT, purity>99%)	Sigma, USA	Cat #D9163
L-lysine (purity>98%)	Sigma, USA	Cat #L5501
D-glucose (purity>99.5%)	Sigma, USA	Cat #G8270
glycine (purity>99%)	Sigma, USA	Cat #W328707
Sucrose	Sigma, USA	Cat # 1.07687
L-tryptophan (purity>99%)	Sigma, USA	Cat # 574597
L-glutamine (purity>99%)	Sigma, USA	Cat #G3126
Ethanol absolute (99.7%)	Sinopharm Chemical Reagent Beijing Co. Ltd. (Beijing, China)	Cat # 10009218
sodium acetate (purity>99%)	Sinopharm Chemical Reagent Beijing Co. Ltd. (Beijing, China)	Cat # 10018718
trisodium citrate (purity>99%)	Sinopharm Chemical Reagent Beijing Co. Ltd. (Beijing, China)	Cat # 10019408
ethylene glycol	Sinopharm Chemical Reagent Beijing Co. Ltd. (Beijing, China)	Cat # 10009818
ferric chloride (purity>97%)	Sinopharm Chemical Reagent Beijing Co. Ltd. (Beijing, China)	Cat # 81015228
trifluoroacetic acid (TFA, purity>99%)	Sinopharm Chemical Reagent Beijing Co. Ltd. (Beijing, China)	Cat # 80134716
Resorcinol (purity>99%)	Sinopharm Chemical Reagent Beijing Co. Ltd. (Beijing, China)	Cat # 81012016
Thiol-aptamers	Sangon Biotech Co., Ltd. (Shanghai, China)	Cat # A100958
Formaldehyde solution (CH2O, purity >36.0%)	Shanghai Aladdin Reagent Co. Ltd. (Shanghai, China)	Cat #F140859
Deposited data		
LDI-MS based metabolomics	This paper	NMDR: ST002733, Project DOI: https://doi.org/10.21228/M8BT6J
Software and algorithms		
R version 4.0.5	R project	https://www.r-project.org/
PASS 15.0.5	NCSS, LLC. Kaysville, Utah, USA	https://www.ncss.com/download/pass/updates/pass15/
MATLAB (R2016a)	The Mathworks, Natick, MA, USA	https://ww2.mathworks.cn/products/matlab.html
MetaboAnalyst 5.0	MetaboAnalyst	http://www.metaboanalyst.ca

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
mlbench 2.1	N/A	https://cran.r-project.org/web/packages/mlbench/index.html
pROC 1.18.2	N/A	https://cran.r-project.org/web/packages/pROC/index.html
ModelGood 1.0.9	N/A	https://cran.r-project.org/web/packages/ModelGood/index.html
glmnet 4.1	N/A	https://cran.r-project.org/web/packages/glmnet/index.html
caret 6.0	N/A	https://cran.r-project.org/web/packages/caret/index.html
cluster 2.1.4	N/A	https://cran.r-project.org/web/packages/cluster/index.html
catboost 1.2	N/A	https://pypi.org/project/catboost/
ggalluvial 0.12.5	N/A	https://cran.r-project.org/web/packages/ggalluvial/index.html
ggpubr 0.6.0	N/A	https://cran.r-project.org/web/packages/ggpubr/index.html
survival 3.5	N/A	https://cran.r-project.org/web/packages/survival/index.html
ggplot2 3.4.2	N/A	https://cran.r-project.org/web/packages/ggplot2/index.html
pheatmap 1.0.12	N/A	https://cran.r-project.org/web/packages/pheatmap/index.html
circlize 0.4.15	N/A	https://cran.r-project.org/web/packages/circlize/index.html

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Jun Pu (pujun310@hotmail.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The standardized metabolomics data have been deposited to the NIH Common Fund's National Metabolomics Data Repository (NMDR) Website, the Metabolomics Workbench, https://www.metabolomicsworkbench.org[58] where it has been assigned Study ID: ST002733 (as listed in the key resources table). This paper does not report original code. Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Participants and ethics

The Shanghai Community Cohort Establishment and Follow-up (NCT04517513) is registered with ClinicalTrials.gov as a large general population cohort in Pudong New District, Shanghai, China. As a part of the series of studies on this cohort, all subjects (n = 17,841) with different degrees of metabolic syndrome (MetS) were enrolled from three different communities including HeQing, JinQiao, and JinYang communities between January 2019 and March 2019, with approval by the Ethics Committee of Renji Hospital (KY2019-136), School of Medicine, Shanghai Jiao Tong University. To avoid potential interference from other confounding factors, 2,238 participants were excluded using the following exclusion criteria. After excluding subjects with missing clinical information, 13,554 volunteers were included in the current study. To better reflect the Chinese population, pre-MetS and MetS were diagnosed by the

following standard criteria according to the statement of the Chinese Diabetes Society 2004[28] (Table S1). According to the guidelines of the Declaration of Helsinki (2013),[59] written informed consent was obtained from all participants and their information was anonymized to protect their privacy.

Exclusion criteria

Patients with any acute and infectious clinical symptoms, including but not limited to fever, headache, cough, malaise, sore throat, loss of smell, runny nose, abdominal pain, and diarrhea, within 3 weeks before sampling. Patients with chronic inflammatory diseases, such as rheumatoid arthritis, systemic lupus erythematosus, and inflammatory bowel disease.

Definition of metabolic syndrome (MetS)

MetS is defined as the presence of three or more of the four risk factors established by the Chinese Diabetes Society 2004[28]: 1) obesity (BMI $\geq$ 25 kg/m^2); 2) hypertension (blood pressure $\geq$ 140/90mmHg and/or have been confirmed and treated as hypertension); 3) hyperglycemia (fasting plasma glucose $\geq$ 6.1 mmol/L [110 mg/dl] and/or 2-h postprandial blood glucose $\geq$ 7.8 mmol/L [140 mg/dl], and/or have been diagnosed and treated as diabetes); 4) dyslipidemia (high triglyceride $\geq$ 1.7 mmol/L [150 mg/dl], and/or low high-density lipoprotein cholesterol< 0.9 mmol/L [35 mg/dl] in men or <1.0 mmol/L [39 mg/dl] in women) (Table S1).

Definition of pre-metabolic syndrome (pre-MS)

pre-MS is defined as having one or two risk factors of MetS but does not meet the criteria for the diagnosis of MetS.[60–62]

Baseline examination

All community subjects underwent physical examinations and biochemical tests after overnight fasting without medication. Current regular medication, previous medical history, sleep quality, mental condition, and smoking habits were recorded based on questionnaires. Waist/hip circumference, height, and weight were measured by tape, calibrated stadiometer, and scale, respectively. Body mass index (BMI) was calculated as weight (kg)/height (m^2). The average of three systolic and diastolic blood pressure recordings was noted by qualified staff to the nearest 1 mmHg using cuff-based blood pressure measurement after at least 15 min of rest in the sitting position. Triglycerides, high-density lipoprotein cholesterol, and fasting blood glucose were detected using a standard spectrophotometer.

Follow-up

The follow-up period is 4 years and its deadline was December 2022. Our predefined primary outcome measure is the full cause of death within 4 years. For each participant, the death event would be verified through the official death-certificates by Shanghai Pudong New Area Public Security Bureau on 9th December 2022. To ensure complete and accurate registration of death information, highly trained clinicians coded the causes of death and each record was cross-checked by a local community physician and further validated by the Center for Disease Control and Prevention (CDC) at the municipal and district levels. If there were any discrepancies, we would re-check medical records, family reports or official death-certificates. All the causes of death were coded and classified according to the International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10) in 2010.

Sample collection, preparation, and storage

Samples were provided by Renji Hospital Biobank, Shanghai Jiaotong University School of Medicine. All human peripheral venous blood samples were obtained following the protocols approved by the Shanghai Jiao Tong University Institutional Review Board. In order to prevent any interference from dietary factors, peripheral vein blood samples were obtained from individuals following an overnight fast of at least 8 h. The collected blood was gently mixed in EDTA anticoagulant tubes. The plasma was separated after centrifugation for 10min (4000 rpm; 4°C) within 2 h of collection. All plasma samples were stored at −80°C for future assays.

METHOD DETAILS

Matrix synthesis

A modified solvo-thermal method was adopted to prepare the ferric particles (FPs), which provided a cost-effective approach for large-scale production. Initially, ethylene glycol solution was used to dissolve ferric chloride. Then, we added trisodium citrate (weights from 0 to 0.8g) to tune the surface charge of the mixture. After adding 1.8g of sodium acetate, we sonicated the products at room temperature for 30min. To form the FPs, we held the products at 200°C for 10h after transferring them into a Teflon-lined stainless-steel autoclave (capacity 50mL). After being washed with ethanol and deionized water, the final particles were dried at 60°C before use.

Material characterization methods

A JEM-2100F instrument (JEOL, Japan) was used to collect transmission electron microscopy (TEM) images by depositing 10μL of material suspension onto a copper grid. An S-4800 instrument (Hitachi, Japan) was used to record scanning electron microscopy (SEM) images by placing a drop of material suspension on aluminum foil. A UV1900 spectrophotometer (AuCy, China) was used

to obtain the optical absorption spectrum of the matrixes at room temperature. An ASAP 2020M (Micromeritics, USA) was used to test the nitrogen adsorption isotherm of the samples degassed in a vacuum. A Nano ZS90 instrument (Malvern, Worcestershire, UK) was used to perform zeta potential and dynamic light scattering (DLS) size measurements in water at 25°C.

Nano-assisted laser desorption/ionization mass spectrometry (LDI MS) profiling

LDI MS experiments were conducted as reported previously.[30] All MS measurements were performed on an Autoflex speed TOF/TOF mass spectrometer (Bruker Daltonics, Bremen, Germany). In a typical LDI MS experiment, ferric particles were dispersed in deionized water at a concentration of 1 mg mL^{-1} for use as a matrix. The CHCA and DHB organic matrices were dissolved in TA30 solution (acetonitrile/0.1% TFA in water, 7/3, v/v) at a concentration of 10 mg/mL. For the detection of the standards, 100nL of analyte solution (each standard listed in the part of chemicals and reagents) with different densities (100 ng mL^{-1}, 1 μg mL^{-1}, 10 μg mL^{-1}, 100 μg mL^{-1}, 1 mg mL^{-1}) was mixed with 100nL of matrix slurry on the plate and dried for LDI MS analysis. For plasma sample detection, the samples were first prepared through protein precipitation, centrifugation, and supernatant filtration according to a commonly applied procedure. Then, a volume of 100nL of plasma solution was spotted on the plate and dried in air at room temperature, followed by adding 100nL of matrix slurry and drying for LDI MS analysis. Then, mass spectra were performed on a 5800 Proteomics Analyzer (Applied Biosystems, Framingham, MA, USA) equipped with a Nd:YAG laser (2 kHz, 355 nm). The acquisitions were extracted in positive reflector ion mode employing delayed extraction with a repetition rate of 1,000Hz and an acceleration voltage of 20 kV. The delay time for this experiment was optimized to 250 ns. The 2,000 laser shots per analysis were for all LDI MS experiments. All the original mass spectra data were visualized in DataExplorer (Version 4.5). Only the m/z signals within 100–300Da and with a signal-to-noise ratio over 3 were then acquired without smoothing processes. For pre-processing, we applied a "home-developed" program using MATLAB (R2016a, The Mathworks, Natick, MA, USA) to normalize and standardize the mass spectra data after peak extraction and alignment.[63] And standard molecules for the accurate mass measurement (±0.05Da) of both Na^+-adducted and K^+-adducted signals were used to perform the mass calibration. The detection limit of standard metabolites (listed in chemicals and reagents' part) obtained by ferric particle, DHB, and CHCA-assisted LDI MS were calculated as previously reported.[30]

Hub differential plasma metabolic fingerprints (PMFs) screening

After testing for a normal distribution, significant differential *PMFs* among the three states of HC, pre-MetS, and MetS in the discovery cohorts were assessed using the Kruskal-Wallis rank-sum test through the primary function in R. The significance level was set as $p < 0.05$. For multiple comparisons, the final MetS-specific hub PMFs were filtered according to the Bonferroni correction and Dunnett's test in R version 4.0.5.

Power analysis

Power analysis for MetS diagnosis was calculated using PASS 15.0.5 (NCSS, LLC. Kaysville, Utah, USA) to determine the appropriate sample size required to detect an effect of a given size with a given degree of confidence.[64] For HC vs. pre-MetS and HC vs. MetS, samples of 2,274 from the positive group and 2,274 from the negative group achieved 99.99% power to detect a difference of 0.05 between the area under the receiver operating characteristic (ROC) curve (AUC) under the null hypothesis of 0.75 and an AUC under the alternative hypothesis of 0.80 using a two-sided Z-test at a significance level of 0.05. For pre-MetS vs. MetS, samples of 3,504 from the positive group and 3,504 from the negative group achieved 99.99% power to detect a difference of 0.05 between AUC under the null hypothesis of 0.75 and an AUC under the alternative hypothesis of 0.80 using a two-sided Z-test at a significance level of 0.05.

Data sets splitting

To improve the generalization of the diagnostic models, we randomly selected 4,548, 7,008, and 4,548 participants (1:1 case-control matched) for HC vs. pre-MetS, pre-MetS vs. MetS and HC vs. MetS, respectively. And all these data sets were split at a ratio of 0.7 to 0.3 into the discovery (n = 3,184, 4,906, 3,184 70%) and validation (n = 1,364, 2,102, 1,364 30%) sets for HC vs. pre-MetS, pre-MetS vs. MetS and HC vs. MetS, respectively.

Model selection

All 26 PMFs were included in the model construction as biomarker panels with age and gender adjustment. To choose the suitable machine learning (ML) algorithm for model construction in this study, five machine-learning algorithms including generalized linear models via least absolute shrinkage and selection operator and elastic-net regularization (GLMNET), support vector machine (SVM), multivariate adaptive regression splines (MARS), random forest (RF), and adaptive boosting (ADABOOST) were adopted in to build diagnostic model for HC vs. MetS by 5-fold cross-validation through the "caret" package and randomized searching through the "mlbench" package in R version 4.0.5. After comprehensively comparing the sensitivity, specificity, and AUCs in the validation cohort (n = 1,364), the GLMNET algorithm was used to construct the final classifiers for pre-MetS and MetS with powerful performance.

PMFs-based diagnostic models

The final PMF-based diagnostic models were constructed using all 60 hub PMFs, with age and gender correction, by 5-fold cross validation through the "glmnet" package. And the parameter optimization was performed by randomized searching and nested

cross-validation through the "mlbench" package in R version 4.0.5. To further evaluate the classification performance of our models, ROC curves and AUCs were generated using the "pROC" package in R version 4.0.5. We also compared the model performance of 26 hub features with all the features of 303 PMFs for HC vs. MetS. Similarly, calibration curves[35] were drafted according to the GLMNET regression results to assess the model reliability using the "ModelGood" package in R version 4.0.5. The individual benefits of our model were evaluated by decision curve analyses[65] in R version 4.0.5.

Metabolomic clustering and assessment of MetS risk factors

Four were identified according to the gap-statistic calculated using the "cluster" packages in R version 4.0.5 to optimize the number of clusters. Unsupervised learning of 16 sub-clusters (RF0000–RF1111) containing different MetS risk factors with standardized 26 hub PMFs was then conducted using the K-means algorithm and hierarchical clustering though the "cluster" and "circlize" packages in R version 4.0.5 after dimensionality deduction with principal component (PC) analysis. The five MetS subgroups (RF0111, RF1011, RF1101, RF1110, and RF1111) and the HC group (RF0000) were clustered and visualized with the K-means algorithm with the "pheatmap" packages in R version 4.0.5. And the 26 hub PMFs were clustered into four metabolic modules (MP).

Construction and validation of plasma metabolic risk (PMR) stratification

The average intensity of 26 hub PMFs among all the five subgroups (RFN0-RFN4) according to the numbers of MetS risk factors were shown using the "ggplot2" package in R version 4.0.5. The variant analysis was performed using the "ggalluvial" package in R version 4.0.5 to track differences between the five subgroups (RFN0-RFN4) and sixteen sub-clusters (RF0000-RF1111) on behalf of the Sankey diagram. In order to ascertain the true metabolic status of individuals within a general population cohort, a PMR model was constructed using candidate factors such as age, gender, and 26 hub PMFs. The model was developed using the categorical boosting (CatBoost) algorithm, a widely used method for gradient boosting on decision trees.[66] Three indexes (Index1-3) were calculated for each participant. By comparing the three index scores of everyone, all participants in the general population cohort were reclassified into three metabolic patterns: low-(LMR), medium-(MMR), and high (HMR) metabolic risk patterns, respectively. The relative change in different clinical parameters from the LMR pattern was calculated according to the following formula:

$$Change\ from\ LMR = \frac{(Exp_i - Exp_{LMR})}{Exp_{LMR}} \times 100\% \quad \text{(Equation 1)}$$

where Exp_i represented the mean expression of each clinical parameter in the MMR or HMR pattern. Variant analysis was performed using the "ggalluvial" package in R version 4.0.5 to track differences between individual metabolic risk subgroups and traditional MetS classifications on behalf of the Sankey diagram. To validate the predictive significance of this PMR stratification, the scores of these three indexes were compared among the five subgroups (RFN0-RFN4) and three groups (HC, pre-MetS, and MetS) in the general population cohort using the Kruskal-Wallis test in the "ggpubr" package in R version 4.0.5. For the survival analysis of the 4-year mortality events, we conducted univariate Cox regression, multivariate Cox regression and Kaplan-Meier analysis using the "survival" package in R version 4.0.5. The univariate and multivariate Cox proportional hazards models were used to determine the Hazard Ratio (HR) of the variables. Models were tested statistically using the Likelihood test. Cumulative curves were calculated using the Kaplan-Meier method and log-rank test was used to examine the intergroup differences. $p < 0.05$ was considered to indicate a statistically significant difference.

Pathway analysis

The 26 hub PMFs were validated as a metabolite feature panel by accurate mass measurement through the human metabolome database Version 5.0 (http://www.hmdb.ca).[67] Pathway topology analysis was conducted to explore the respective biochemical pathways among these differential metabolomic profiles using MetaboAnalyst 5.0 (http://www.metaboanalyst.ca).[68]

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis was performed using R version 4.0.5. Descriptive data were presented as mean and standard deviation. The Shapiro-Wilk test was applied to test for a normal distribution and Levene's test was used to test for homogeneity of variance. Differences among groups were tested by one-way analysis of variance (ANOVA) for parametric data such as age, and the Kruskal-Wallis rank-sum test was applied for non-parametric data, such as m/z signals. The Bonferroni test and Dunnett's test were applied for multiple comparisons. Categorical data, such as gender, were presented as numbers and percentages and analyzed by χ^2 tests. AUC was compared using a one-sided DeLong test. All significance levels were set at 5%.

ADDITIONAL RESOURCES

This study is based the prospective Shanghai Community Cohort Establishment and Follow-up (NCT04517513): https://classic.clinicaltrials.gov/ct2/show/NCT04517513.

Cell Reports Medicine, Volume *4*

Supplemental information

Plasma metabolic fingerprints for large-scale screening and personalized risk stratification of metabolic syndrome

Yifan Chen, Wei Xu, Wei Zhang, Renyang Tong, Ancai Yuan, Zheng Li, Huiru Jiang, Liuhua Hu, Lin Huang, Yudian Xu, Ziyue Zhang, Mingze Sun, Xiaoxiang Yan, Alex F. Chen, Kun Qian, and Jun Pu

Supplemental information

Plasma metabolic fingerprints for large-scale screening and personalized risk stratification of metabolic syndrome

Yifan Chen, Wei Xu, Wei Zhang, Renyang Tong, Ancai Yuan, Zheng Li, Huiru Jiang, Liuhua Hu, Lin Huang, Yudian Xu, Ziyue Zhang, Mingze Sun, Xiaoxiang Yan, Alex F. Chen, Kun Qian, Jun Pu

This section includes the following:

1. Figures S1–15
2. Tables S1–10

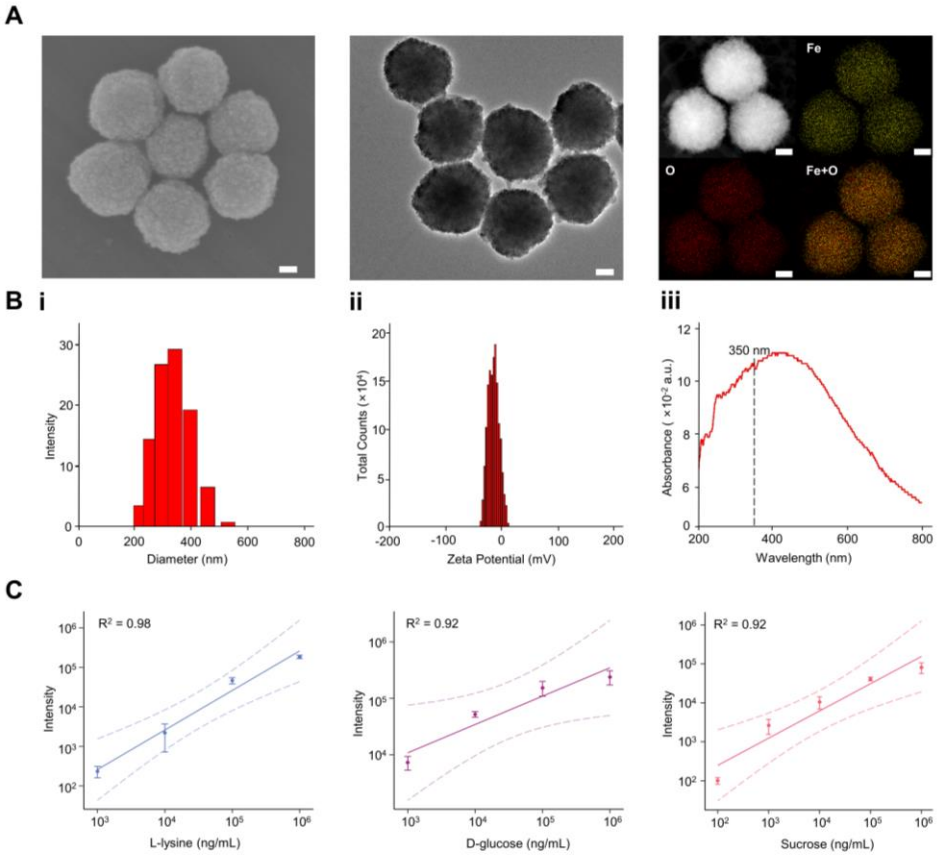


Figure S1. Material characterization of the ferric particles used in the LDI-MS process. Related to Figure 1. A) Electron micrograph images of the ferric particles. Scanning electron microscopy (SEM) images showed nanoscale surface roughness of ferric particles (n≥3 randomly selected). Transmission electron microscopy (TEM) images showed the polycrystalline structures of ferric particles (n≥3 randomly selected). Elemental mapping images of the ferric particles with Fe, O, and Fe+O (Fe in yellow and O in red). Scale bars=50 nm. **B) i)** Size distribution of ferric particles at the room temperature (25°C) in water by dynamic light scattering (DLS). **ii)** Zeta potential of ferric particles. **iii)**Absorption spectrum of ferric particles. **C)** Linear correlation between standard concentration and LDI-MS intensity ($M+[Na]^+$). Quantification results for samples consisted of different contents of lysine, D-glucose, and sucrose, affording R^2 values of 0.92-0.98 (n=3 independent mixed samples tested 5 times each).

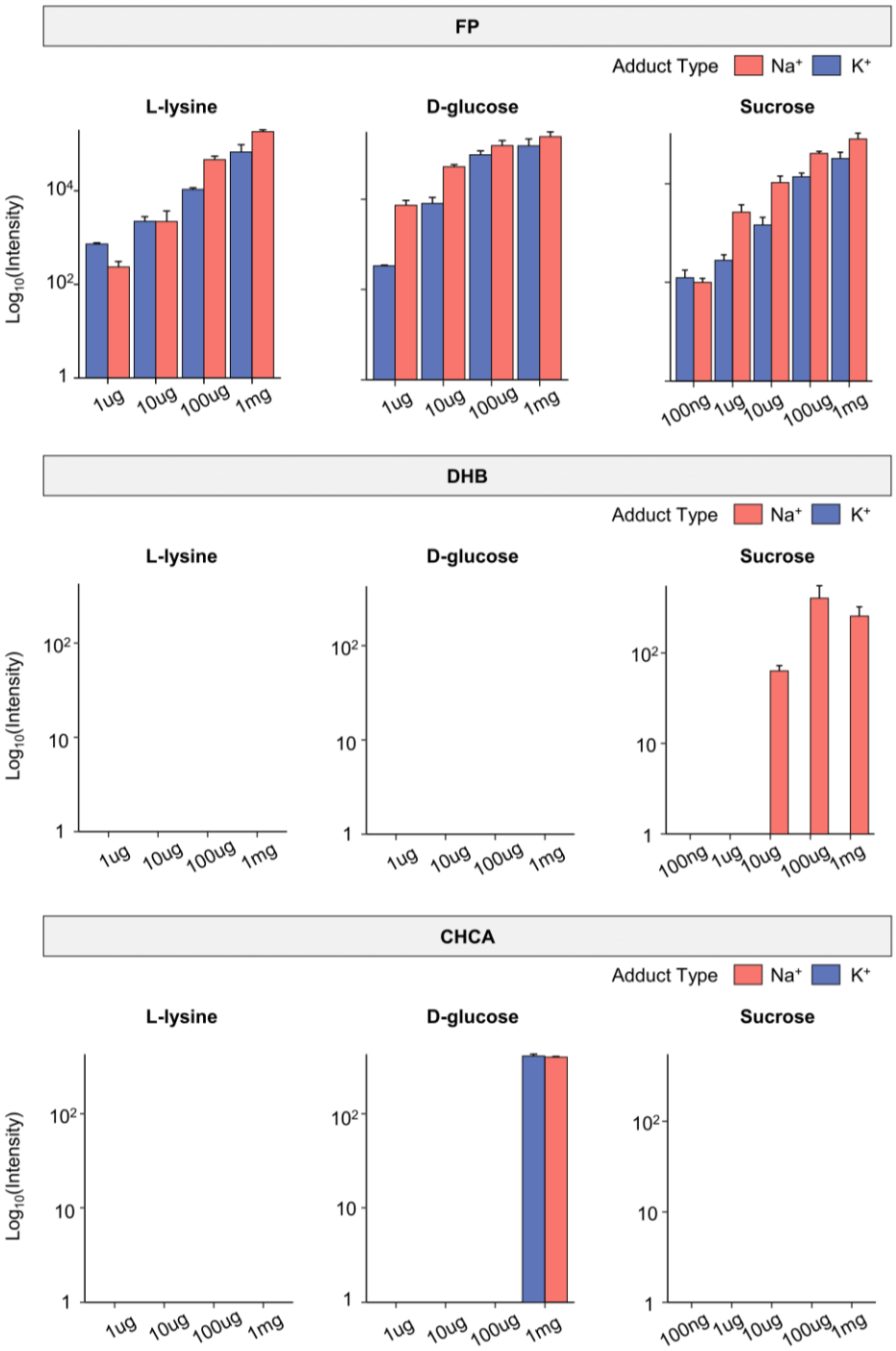


Figure S2. Quantification results for standards including L-lysine, D-glucose, and sucrose at different concentrations obtained by ferric particle, CHCA, and DHB-assisted LDI-MS. Related to Figure 1. n=3 independent mixed samples tested 5 times each. FP, ferric particles; CHCA, α-cyano-4-hydroxycinnamic acid; DHB, 2,5-dihydroxybenzoic acid.

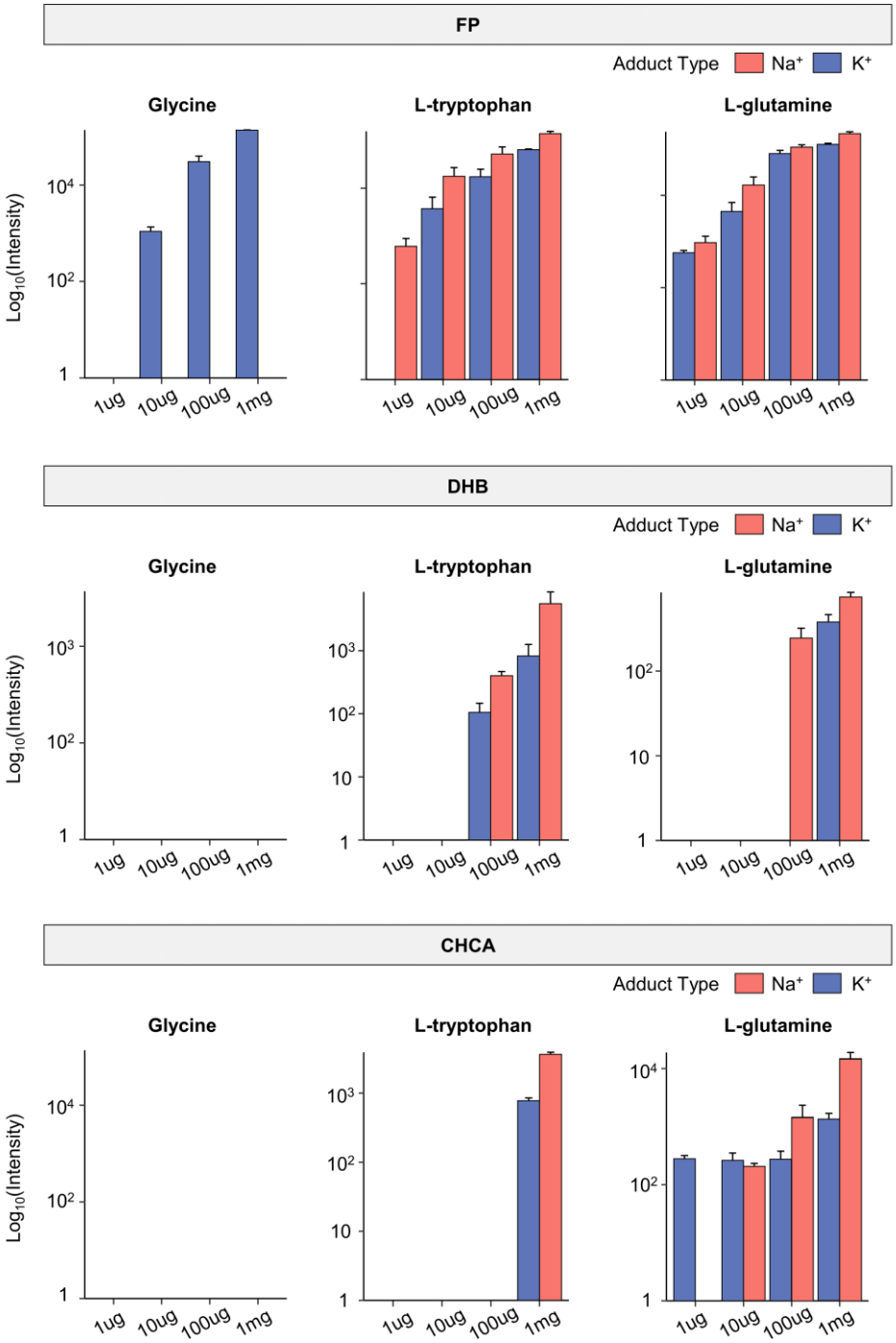


Figure S3. Quantification results for standards including glycine, L-tryptophan, and L-glutamine at different concentrations obtained by ferric particle, CHCA, and DHB-assisted LDI-MS. Related to Figure 1. n=3 independent mixed samples tested 5 times each. FP, ferric particles; CHCA, α-cyano-4-hydroxycinnamic acid; DHB, 2,5-dihydroxybenzoic acid.

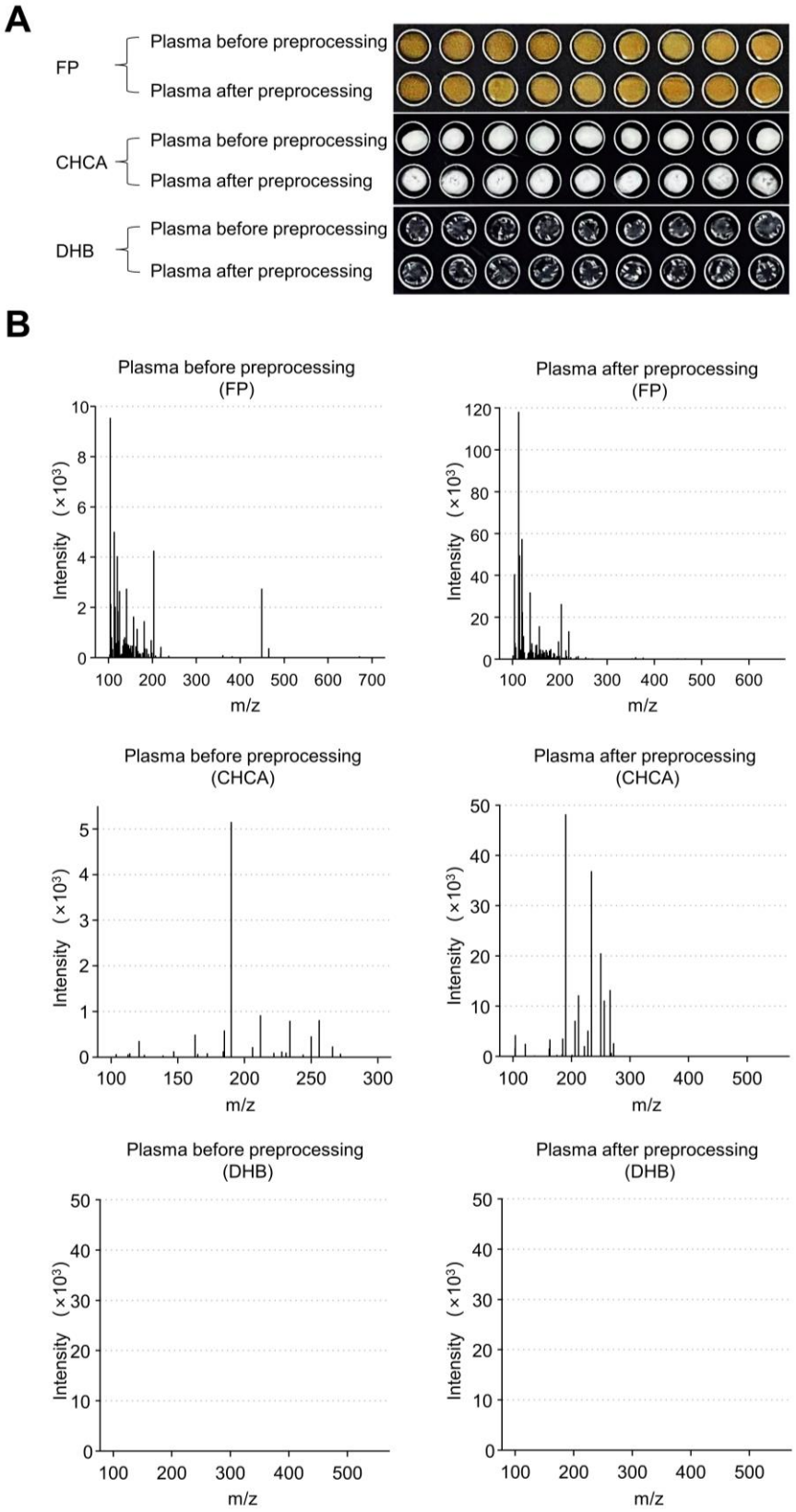


Figure S4. Plasma samples with and without pretreatment in three different matrices for LDI-MS. Related to Figure 1. A) Dried drops of the mixture of plasma samples and three different matrices including FP, CHCA, and DHB, on the plate. **B)** Typical mass spectrometry spectra of plasma samples

with and without pretreatment obtained by FP, CHCA, and DHB-assisted LDI-MS. FP, ferric particles; CHCA, α-cyano-4-hydroxycinnamic acid; DHB, 2,5-dihydroxybenzoic acid.

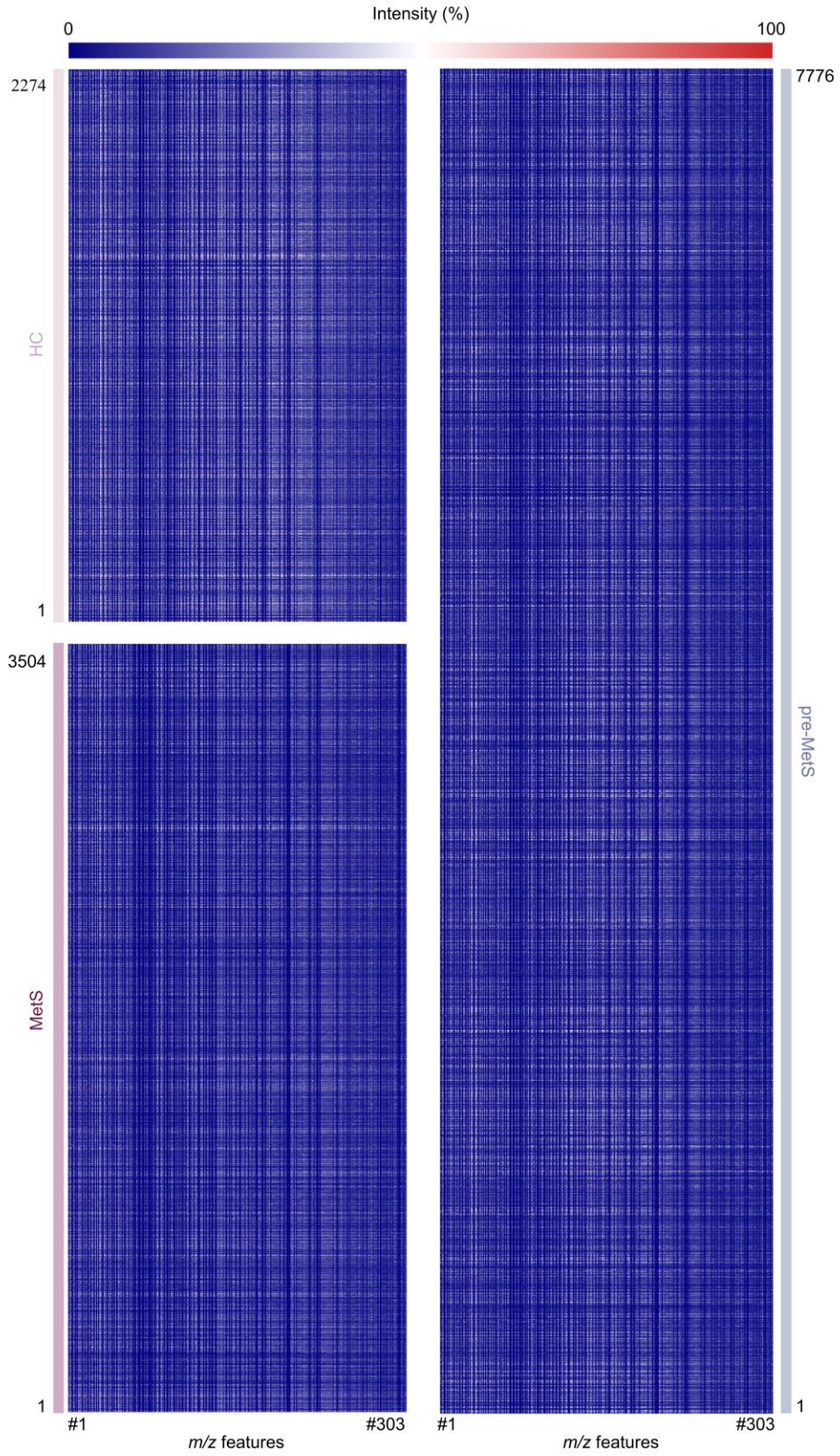

Intensity (%)
0
100
2274
HC
1
3504
MetS
1
#1
#303
m/z features
7776
pre-MetS
1
#1
#303
m/z features

Figure S5. Plasma metabolic fingerprints were extracted from raw mass spectra for all participants (n=13,554). Related to Figures 1 and 2. Three groups: healthy control (HC, n=2,274) (in pink); pre metabolic syndrome (pre-MetS, n=7,776) (in gray); metabolic syndrome (MetS, n=3,504) (in purple).

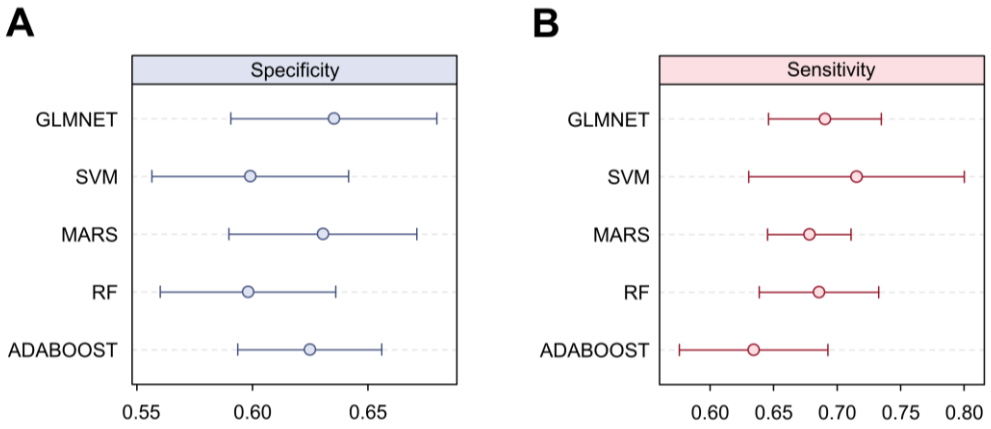


Figure S6. Distribution of performances of different machine-learning (ML)-based models for HC vs. MetS in the validation cohort (n=1,364). Related to Figure 3. A) Specificity of different ML-based models in the validation cohort. **B)** Sensitivity of different ML-based models in the validation cohort. GLMNET, generalized linear models via least absolute shrinkage and selection operator and elastic-net regularization; SVM, support vector machine; MARS, multivariate adaptive regression splines; RF, random forest; Adaboost, adaptive boosting. Error bars represent a confidence level of 0.95.

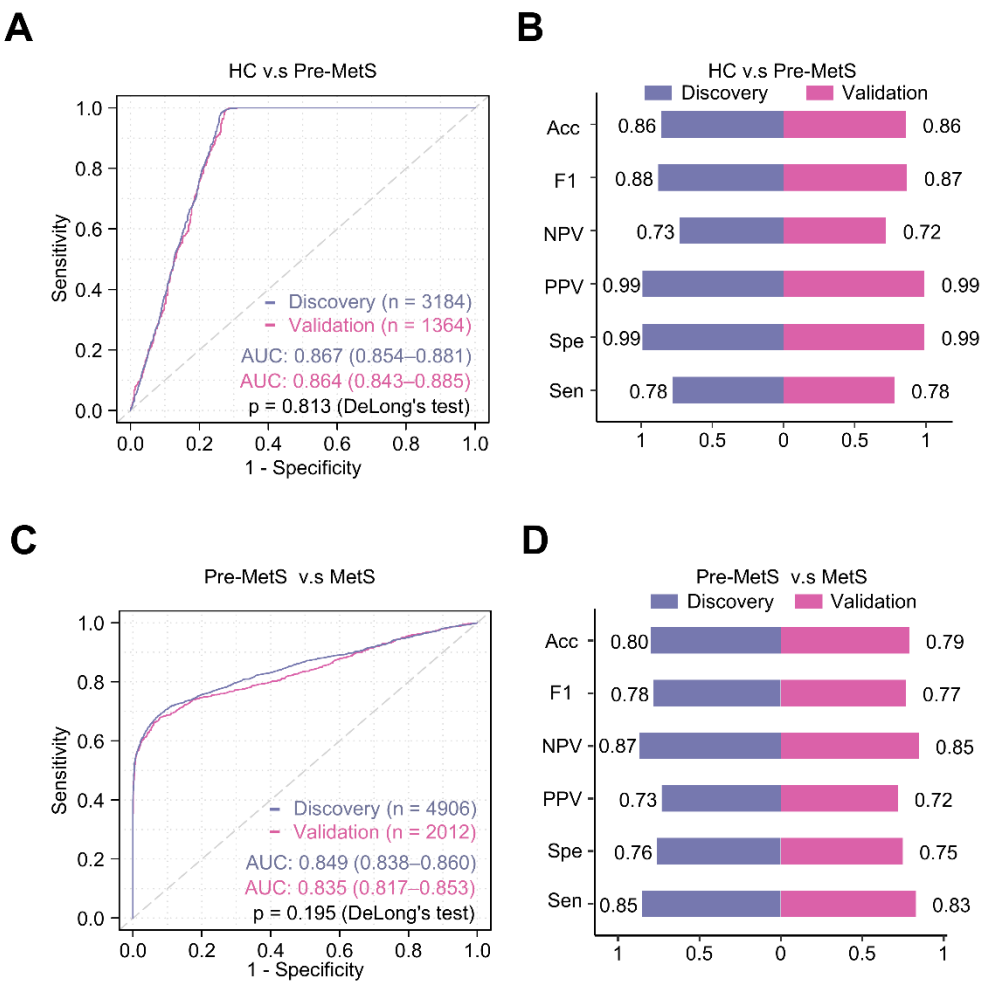


Figure S7. Construction of PMFs-based diagnostic model for HC vs. pre-MetS and pre-MetS vs. MetS. Related to Figure 3. A) Receiver operating characteristic (ROC) curves for HC vs. pre-MetS between the discovery (n=3,184) and validation (n=1,364) sets. **B)** Comparison of performances of the diagnostic model for HC vs. pre-MetS between the discovery and validation sets with different evaluation metrics. **C)** ROC curves for pre-MetS vs. MetS between the discovery (n=4,906) and validation (n=2,012) sets. **B)** Comparison of performances of the diagnostic model for pre-MetS vs. MetS between the discovery and the validation sets with different evaluation metrics. Acc, accuracy; F1, F1 score; NPV, negative predictive value; PPV, positive predictive value; Spe, specificity; Sen, sensitivity.

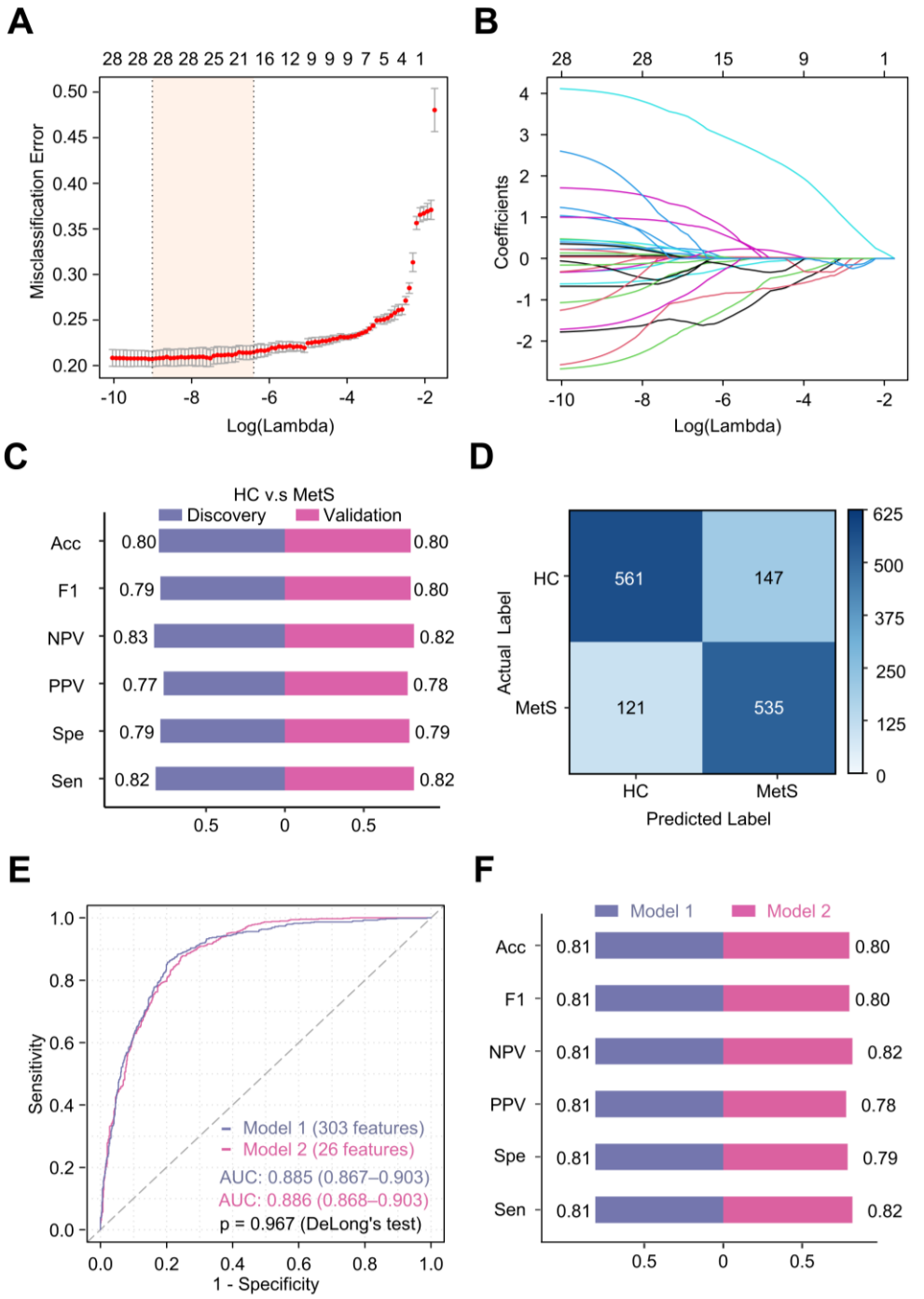


Figure S8. Construction of PMF-based diagnostic model for HC vs. MetS. Related to Figure 3. A) Generalized linear models via least absolute shrinkage and selection operator and elastic-net regularization (GLMNET) regression analysis results. The tuning parameter (lambda) was calculated based on the misclassification error by fivefold cross validation. Dotted vertical lines drawn at optimal values of lambda by minimum criteria and 1-standard error criteria. **B)** GLMNET variable trace profiles of hub metabolic features by eight-fold cross validation. Each curve represents the dynamic variation of the independent variable. The y-axis shows the coefficient level, the lower x-axis represents log(lambda), and the upper x-axis is the number of selected PMFs under each lambda. **C)** Comparison of performances between discovery (n=3,184) and validation (n=1,364) sets in our model with different evaluation metrics. **D)** Confusion matrix of the validation set (n=1,364) in our model. **E)** ROC curves for the validation

(n=1,364) set under the diagnostic models based on all 303 features (Model 1) and 26 hub features (Model 2) adjusted for age and gender. **F)** Comparison of performances between Model 1 and Model 2 in the validation cohort (n=1,364) with different evaluation metrics. Acc, accuracy; F1, F1 score; NPV, negative predictive value; PPV, positive predictive value; Spe, specificity; Sen, sensitivity.

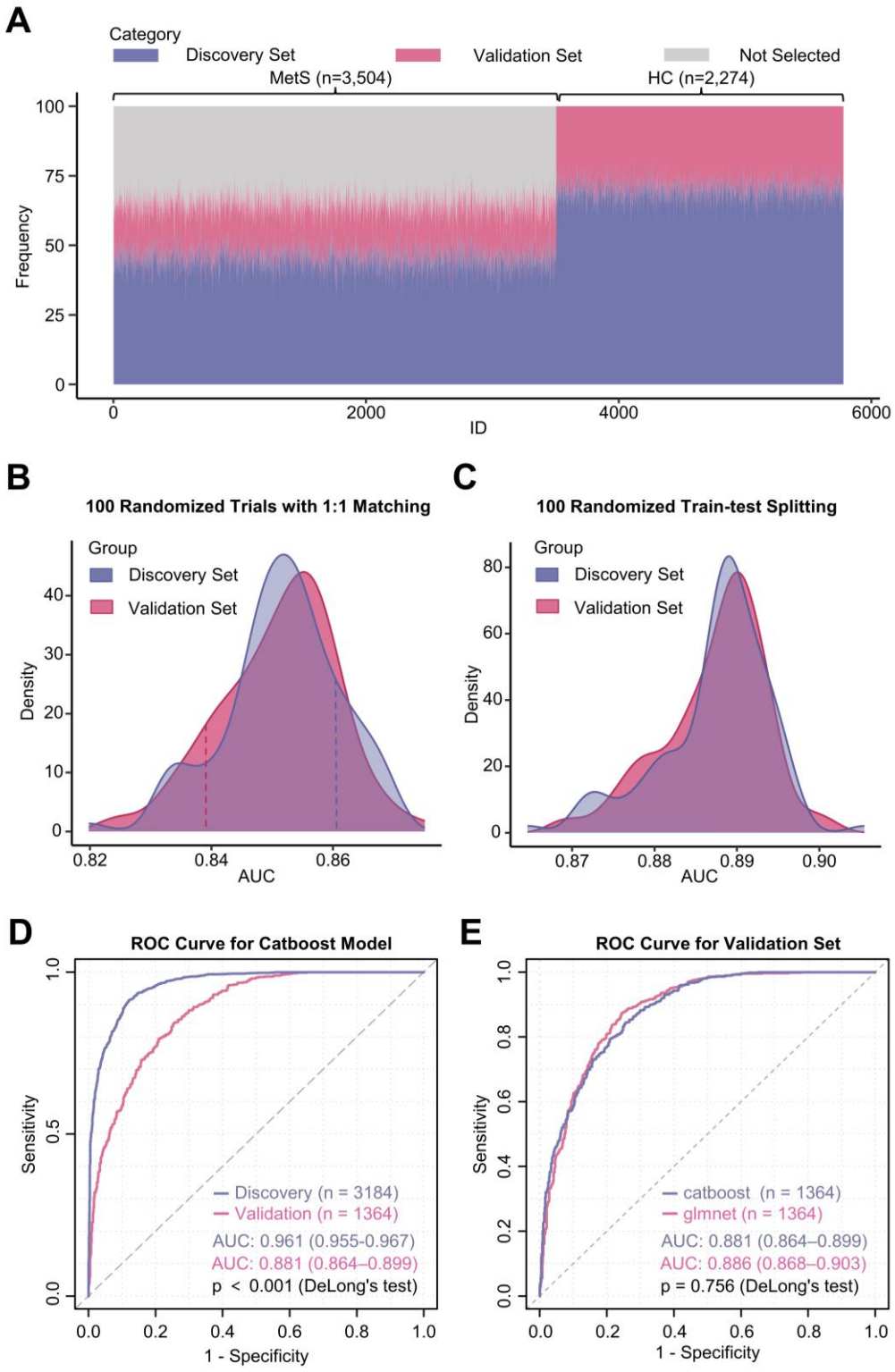
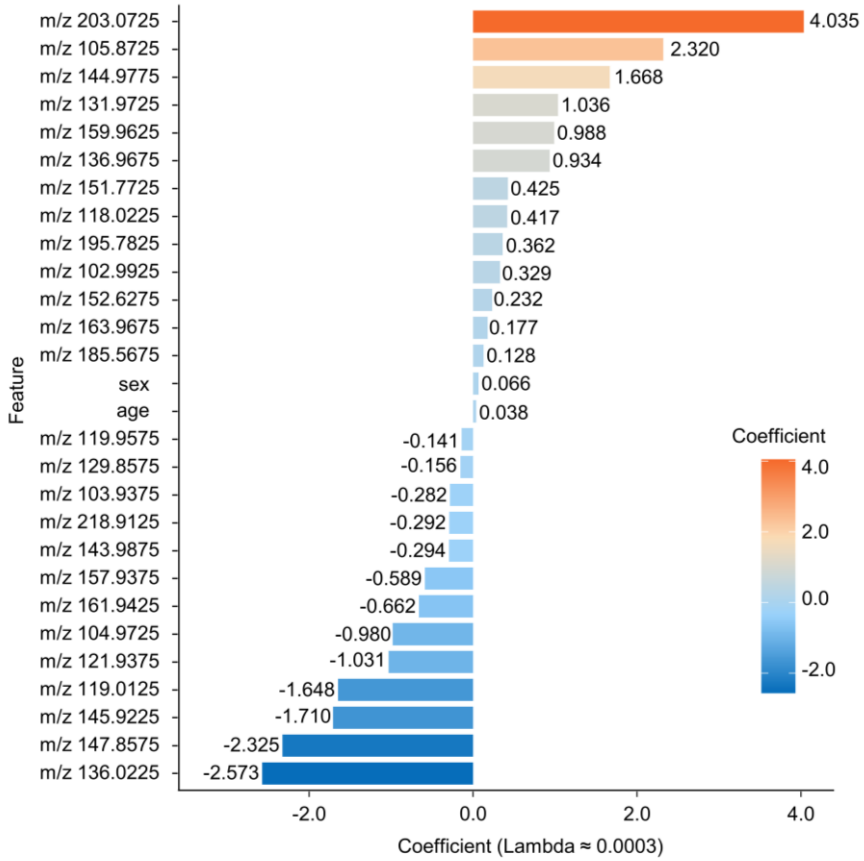


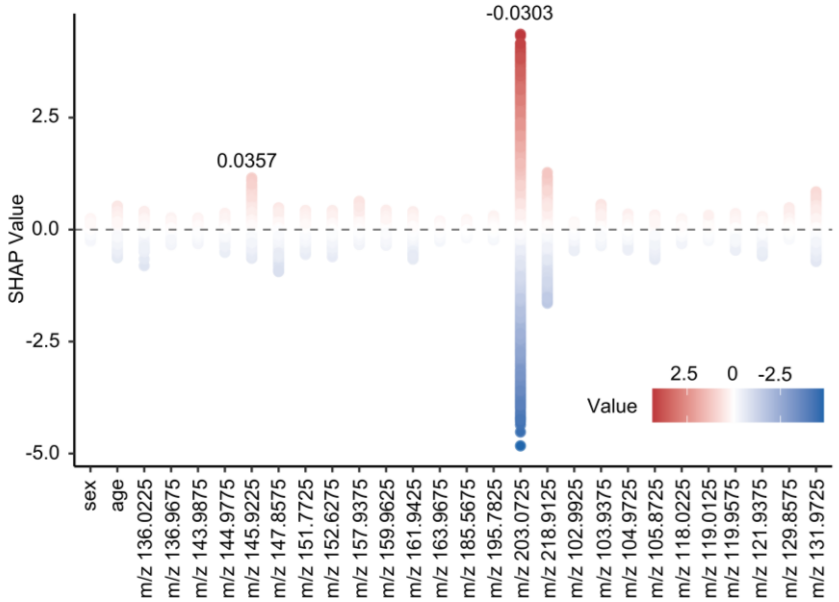
Figure S9. Validation of PMF-based diagnostic model for HC vs. MetS using 100 independent trials and Catboost-based models. Related to Figure 3. A) 100 independent randomized trials were conducted to generate discovery and validation sets with 7:3 split ratios from HC and MetS groups. **B)** Density distribution of AUC values in the 100 randomized training and testing sets. The dashed lines represent the sampled dataset used in Figure 3F. **C)** Density distribution of AUC values in 100 independent randomized splitting using the selected dataset shown by the dashed lines in B. **D)** ROC curve for the Catboost model trained on the dataset in Figure 4C. **E)** Comparison of ROC curves between

Catboost and GLMNET models for the same validation dataset. HC, healthy controls, MetS, metabolic syndrome, AUC, area under the receiver operating characteristic curve, ROC, receiver operating characteristic curve, CatBoost, categorical boosting, Glmnet, generalized linear models via least absolute shrinkage and selection operator and elastic-net regularization.

A


m/z 203.0725
m/z 105.8725
m/z 144.9775
m/z 131.9725
m/z 159.9625
m/z 136.9675
m/z 151.7725
m/z 118.0225
m/z 195.7825
m/z 102.9925
m/z 152.6275
m/z 163.9675
m/z 185.5675
sex
age
m/z 119.9575
m/z 129.8575
m/z 103.9375
m/z 218.9125
m/z 143.9875
m/z 157.9375
m/z 161.9425
m/z 104.9725
m/z 121.9375
m/z 119.0125
m/z 145.9225
m/z 147.8575
m/z 136.0225
4.035
2.320
1.668
1.036
0.988
0.934
0.425
0.417
0.362
0.329
0.232
0.177
0.128
0.066
0.038
-0.141
-0.156
-0.282
-0.292
-0.294
-0.589
-0.662
-0.980
-1.031
-1.648
-1.710
-2.325
-2.573
Feature
Coefficient
4.0
2.0
0.0
-2.0
-2.0
0.0
2.0
4.0
Coefficient (Lambda ≈ 0.0003)

B


-0.0303
0.0357
2.5
0.0
-2.5
-5.0
SHAP Value
2.5
0
-2.5
Value
sex
age
m/z 136.0225
m/z 136.9675
m/z 143.9875
m/z 144.9775
m/z 145.9225
m/z 147.8575
m/z 151.7725
m/z 152.6275
m/z 157.9375
m/z 159.9625
m/z 161.9425
m/z 163.9675
m/z 185.5675
m/z 195.7825
m/z 203.0725
m/z 218.9125
m/z 102.9925
m/z 103.9375
m/z 104.9725
m/z 105.8725
m/z 118.0225
m/z 119.0125
m/z 119.9575
m/z 121.9375
m/z 129.8575
m/z 131.9725

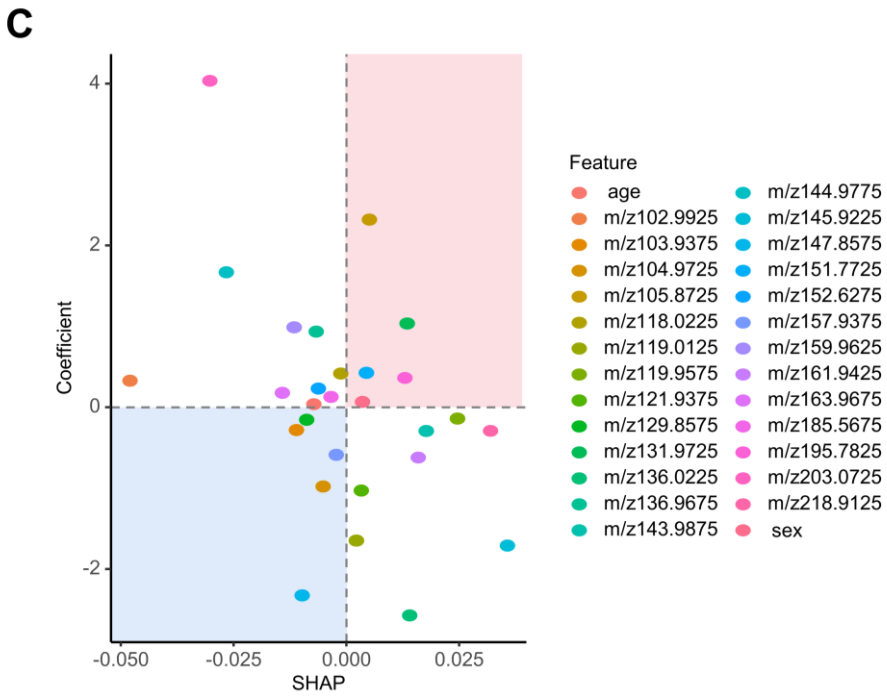


Figure S10. Feature importance of these 26 hub PMFs. Related to Figure 3. A) Feature coefficient of the generalized linear models via least absolute shrinkage and selection operator and elastic-net regularization regression analysis-based diagnostic model for HC vs. MetS with best lambda value 0.0003402896. Features with positive and negative coefficients are colored red and blue, respectively. **B)** SHAP values based on the Catboost-based model for HC vs. MetS in Figure S9D. **C)** Comparative analysis of GLMNET-based coefficients and Catboost-based SHAP values. SHAP, SHapley Additive exPlanations.

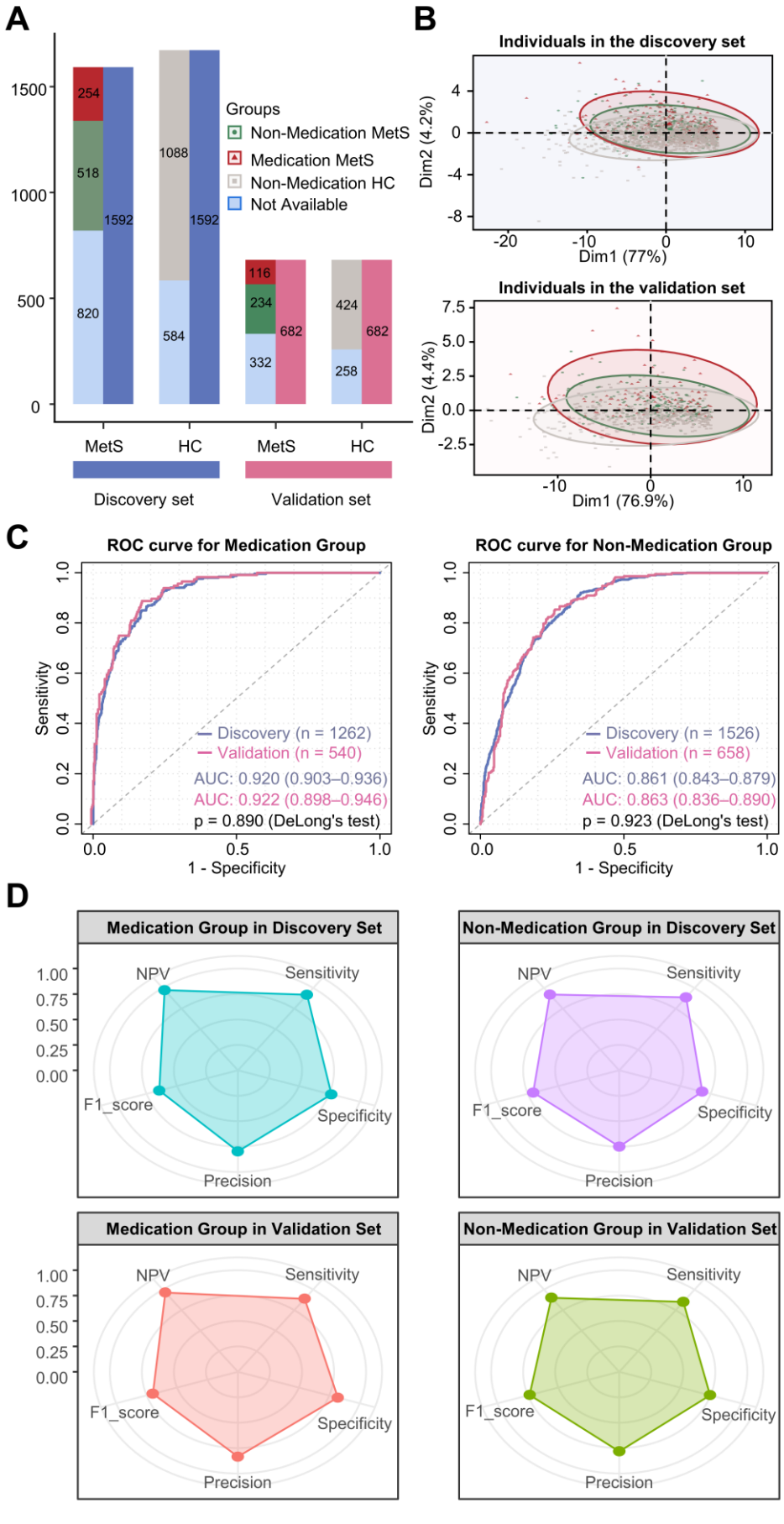

A
1500
1000
500
0
254
518
1592
820
1088
1592
584
116
234
682
332
424
682
258
Groups
Non-Medication MetS
Medication MetS
Non-Medication HC
Not Available
MetS
HC
MetS
HC
Discovery set
Validation set
B
Individuals in the discovery set
Dim2 (4.2%)
Dim1 (77%)
Individuals in the validation set
Dim2 (4.4%)
Dim1 (76.9%)
C
ROC curve for Medication Group
Sensitivity
1 - Specificity
Discovery (n = 1262)
Validation (n = 540)
AUC: 0.920 (0.903–0.936)
AUC: 0.922 (0.898–0.946)
p = 0.890 (DeLong's test)
ROC curve for Non-Medication Group
Discovery (n = 1526)
Validation (n = 658)
AUC: 0.861 (0.843–0.879)
AUC: 0.863 (0.836–0.890)
p = 0.923 (DeLong's test)
D
Medication Group in Discovery Set
Non-Medication Group in Discovery Set
Medication Group in Validation Set
Non-Medication Group in Validation Set
NPV
Sensitivity
Specificity
Precision
F1_score

Figure S11. Sensitivity analysis to exclude the effect of medication on the PMF-based diagnostic model for MetS. Related to Figure 3. A) Medication status distribution in the discovery and validation cohorts. Medication group: subjects taking two or more medications for different MetS risk factors. Non-medication group: subjects taking less than two medications for MetS risk factors. **B)** PCA analysis based on 26 hub PMFs. **C)** ROC curves for MetS diagnosis in the medication and non-medication groups. **D)** Other model performance evaluation metrics in the medication and non-medication groups for the discovery and validation cohorts. NPV, negative predictive value.

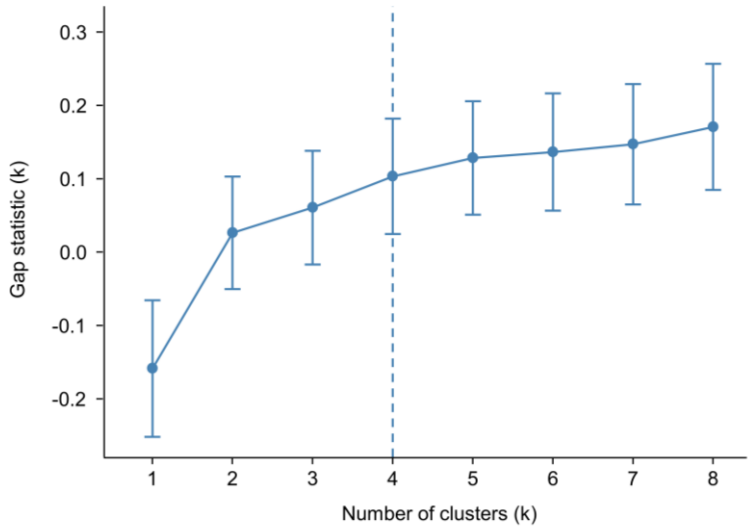


Figure S12. Gap statistic curve for choosing the optimal number of clusters. Related to Figure 4. The dotted vertical line suggests an optimal parameter value of the number of clusters (k) = 4 in our dataset.

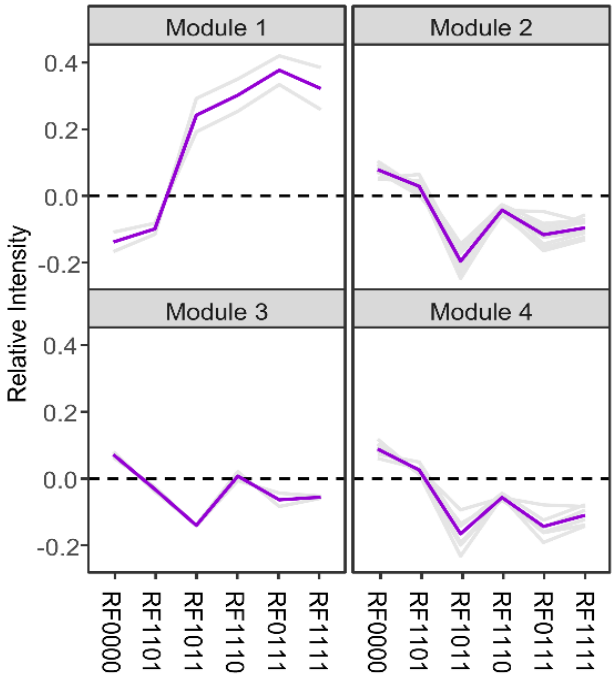


Figure S13. The relative intensity of four metabolic feature modules clustered through K-means algorithm in the five MetS subgroups and HC group. Related to Figure 4.

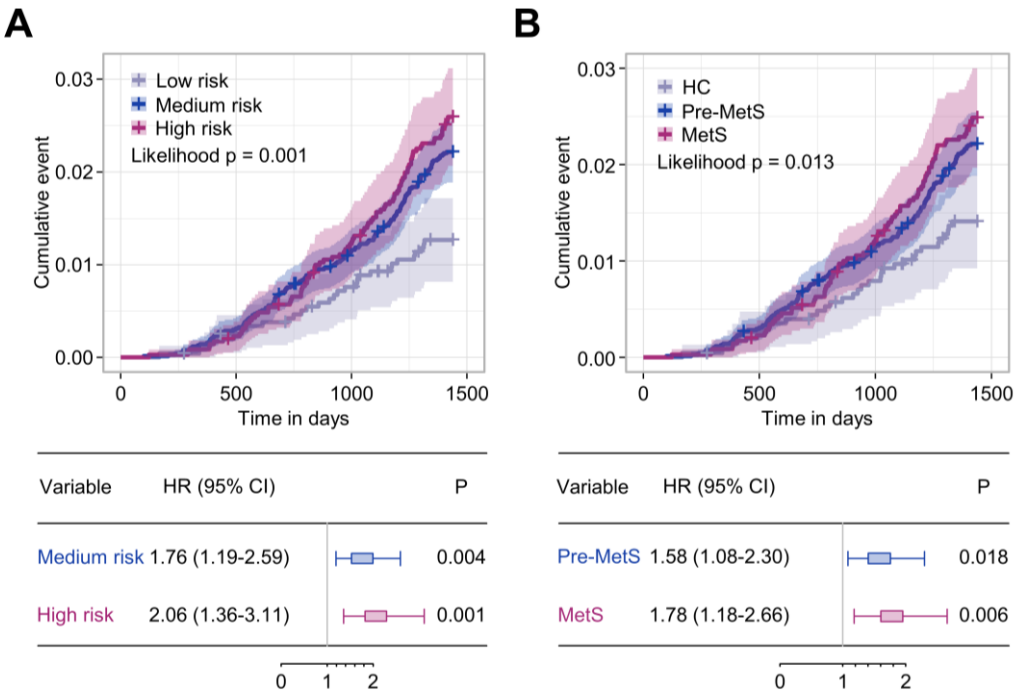


Figure S14. Plasma metabolic risk (PMR) stratification was related to 4-year mortality events in the longitudinal follow-up cohort of 13,554 patients among four communities without gender and age correction. Related to Figure 5. A) Cumulative curves and forest plots for 13,554 patients with three different plasma metabolic risk (PMR) statuses. **B)** Cumulative curves and forest plots for 13,554 patients with three different MetS statuses. The p value for univariate Cox regression analysis models was calculated by the likelihood test. The p value for variables was obtained by the log rank test. HR, Hazard ratio.

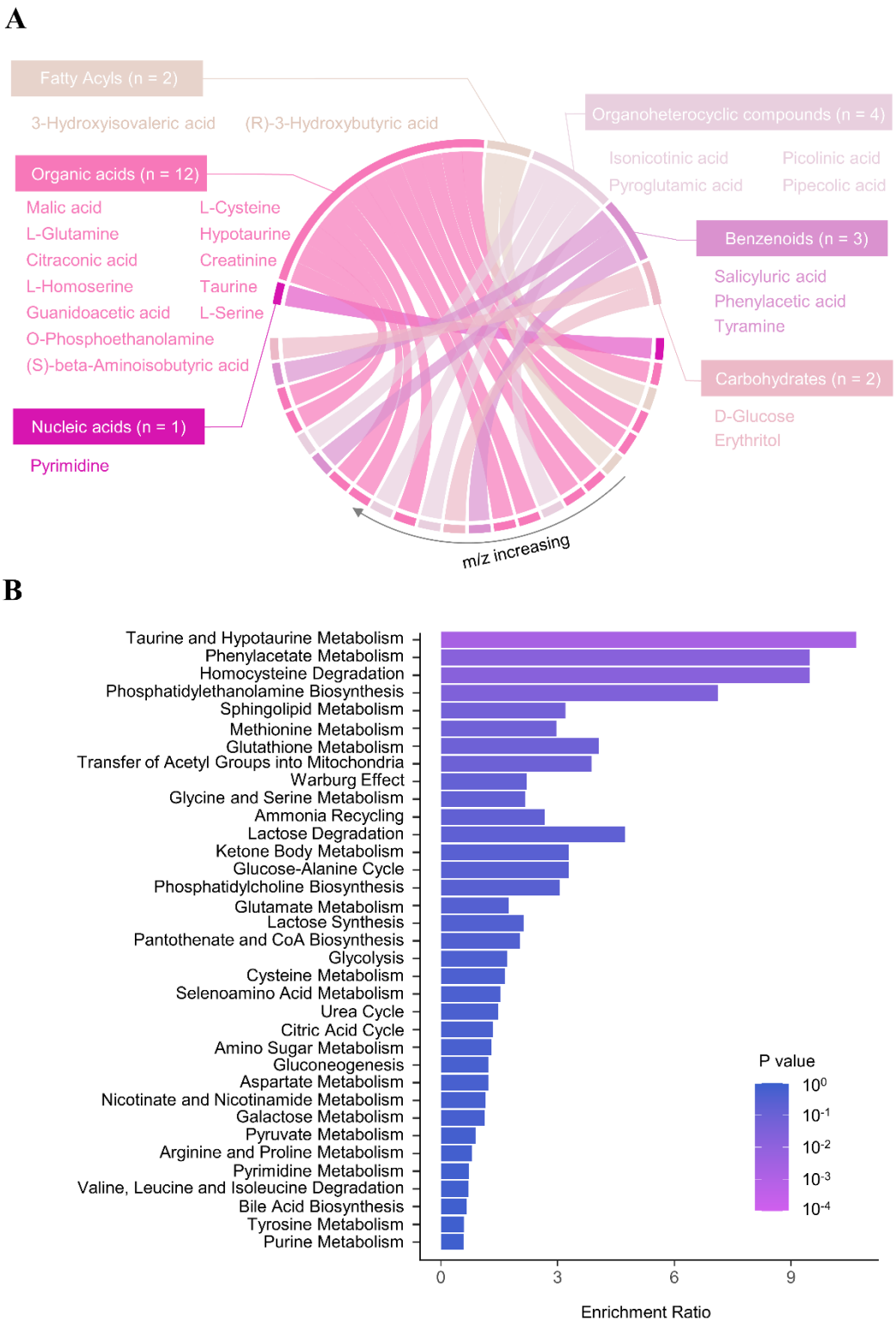


Figure S15. Potential biomarker identification and differential metabolic pathways in the pathological process of MetS. Related to Figure 5. A) Metabolite classification of the 26 hub PMFs by matching through the Human Metabolome Database (HMDB) and MetaboAnalyst 5.0. **B)** Potential metabolic pathways differentially regulating the pathological process of MetS. The color and size of each circle indicate the p value and pathway impact value. A total of four pathways (enrichment ratio>5 and $p<0.05$) were differentially regulated: (1) taurine and hypotaurine

metabolism, (2) phenylacetate metabolism, (3) homocysteine degradation, and (4) phosphatidylethanolamine biosynthesis.

Table S1. Four metabolic syndrome risk factors in a general population in this study. Related to Figure 1.

Risk factor (RF)	Definition [a, b)]
Obesity	BMI ⩾ 25 kg/m^2
Hypertension	BP ⩾ 140/90 mmHg and/or have been confirmed and treated as hypertension
Hyperglycemia	FPG ⩾ 6.1mmol/L (110 mg/dl) and/or 2hPG ⩾ 7.8 mmol/L (140 mg/dl), and/ or have been diagnosed and treated as diabetes
Dyslipidemia	high TG ⩾ 1.7mmol/L (150 mg/dl), and/or low HDLC< 0.9 mmol/L (35 mg/dl) in men or <1.0 mmol/L (39 mg/dl) in women

BMI, body mass index; BP, blood pressure; FPG, fasting plasma glucose; 2hPG, 2-hour postprandial blood glucose; TG, triglycerides; TC, serum total cholesterol; HDLC, high-density lipoprotein cholesterol.

[a)] Concentrating on racial differences, metabolic syndrome was determined by the presence of at least three of the above metabolic risk factors according to the statement of the Chinese Diabetes Society; [b)] pre metabolic syndrome (pre-MetS) was defined as the presence of one or two metabolic risk factors in this study.

Table S2. Detection limit of standard metabolites for standards obtained by ferric particle, DHB, and CHCA-assisted LDI-MS. Related to Figure 1.

Analytes	Detection limit(pmol)		
	FP	DHB	CHCA
L-lysine	6.84	＞6840.53	＞6840.53
D-glucose	5.55	＞5550.75	5550.75
Sucrose	0.29	29.21	＞2921.44
Glycine	133.22	＞13321.61	＞13321.61
L-tryptophan	4.90	489.66	4896.56
L-glutamine	6.84	684.25	68.42

FP, ferric particles; DHB, 2,5-dihydroxybenzoic acid; CHCA, α-cyano-4-hydroxycinnamic acid.

Table S3. Baseline characteristics of discovery and validation sets. Related to Figure 3.

	Control	Case	P value[a)]
A) Diagnostic model for HC vs. MetS			
Train cohort	**HC**	**MetS**	
Number (%)	1592 (50.0)	1592 (50.0)	/
Male (%)	748 (47.0)	726 (45.6)	0.455
Age (mean (SD))	67.09 (5.90)	68.46 (5.59)	<0.001
Validation cohort	**HC**	**MetS**	
Number (%)	682 (50.0)	682 (50.0)	/
Male (%)	316 (46.3)	330 (48.4)	0.481
Age (mean (SD))	66.82 (5.72)	68.36 (5.88)	<0.001
B) Diagnostic model for HC vs. pre-MetS			
Train cohort	**HC**	**pre-MetS**	
Number (%)	1592 (50.0)	1592 (50.0)	/
Male (%)	734 (46.1)	773 (48.6)	0.177
Age (mean (SD))	67.03 (5.87)	67.86 (6.00)	<0.001
Validation cohort	**HC**	**pre-MetS**	
Number (%)	682 (50.0)	682 (50.0)	/
Male (%)	330 (48.4)	337 (49.4)	0.745
Age (mean (SD))	66.95 (5.80)	67.76 (5.86)	0.01
C) Diagnostic model for pre-MetS vs. MetS			
Train cohort	**pre-MetS**	**MetS**	
Number (%)	2453 (50.0)	2453 (50.0)	/
Male (%)	1178 (48.0)	1169 (47.7)	0.819
Age (mean (SD))	67.87 (6.07)	68.38 (5.71)	0.002
Validation cohort	**pre-MetS**	**MetS**	
Number (%)	1051 (50.0)	1051 (50.0)	/
Male (%)	502 (47.8)	480 (45.7)	0.359
Age (mean (SD))	67.95 (5.95)	68.08 (5.54)	0.616

HC, healthy control; pre-MetS, pre metabolic syndrome; MetS, metabolic syndrome; SD, standard deviation.

[a)] p value calculated by χ^2 test for gender data and one-way analysis of variance for age data.

Table S4. Distribution of performances of different machine-learning-based models for HC vs. MetS in the validation cohort (n=1,364) using different evaluation metrics. Related to Figure 3.

	Min	1stQ	Median	Mean	3rdQ	Max
Area under the curve (AUC)						
ADABOOST	0.63	0.66	0.69	0.67	0.7	0.7
RF	0.65	0.67	0.69	0.69	0.7	0.73
GLMNET	0.67	0.72	0.74	0.72	0.74	0.75
MARS	0.69	0.70	0.71	0.71	0.71	0.72
SVM	0.69	0.71	0.71	0.71	0.72	0.74
Sensitivity (Sen)						
ADABOOST	0.59	0.6	0.63	0.63	0.64	0.71
RF	0.65	0.67	0.68	0.69	0.69	0.75
GLMNET	0.65	0.66	0.69	0.69	0.71	0.74
MARS	0.64	0.66	0.69	0.68	0.7	0.7
SVM	0.64	0.66	0.70	0.72	0.78	0.79
Specificity (Spe)						
ADABOOST	0.59	0.62	0.63	0.62	0.64	0.65
RF	0.56	0.59	0.60	0.60	0.60	0.64
GLMNET	0.60	0.60	0.63	0.64	0.67	0.67
MARS	0.59	0.61	0.64	0.63	0.65	0.67
SVM	0.55	0.6	0.6	0.6	0.61	0.64

1stQ, first quartile; 3rdQ, third quartile; GLMNET, generalized linear models via least absolute shrinkage and selection operator and elastic-net regularization; SVM, support vector machine; MARS, multivariate adaptive regression splines; RF, random forest; ADABOOST, adaptive boosting.

Table S5. Gap statistic for different numbers of clusters (k). Related to Figure 4.

K[a)]	logW[b)]	E.logW[c)]	gap[d)]	SE.sim[e)]
1	0.91918374	0.77143744	-0.14774630	0.08255203
2	0.21885086	0.25637335	0.03752249	0.08036719
3	-0.06642547	0.00154248	0.06796795	0.08097058
4	-0.31916315	-0.21126113	0.023563034	0.08061717
5	-0.52894093	-0.40092645	0.12801448	0.08022803
6	-0.72091521	-0.58165729	0.13925792	0.08249429
7	-0.90809419	-0.75470368	0.15339051	0.08436209
8	-1.07750200	-0.93779799	0.13970401	0.08662547

[a)] K represents the number of clusters; [b)] W is the within-cluster sum of squared distances from the cluster means; [c)] E.logW represents the expected value of logW of an appropriate null reference; [d)] gap represents the gap statistic; [e)] SE.sim corresponds to the standard error of the gap statistic.

Table S6. Relative changes in eight clinical parameters compared with the low-risk pattern (%). Related to Figure 5.

	TC	SCr	GLU	HDLC	LDLC	TG	UA	BMI
Medium -risk	2.36	1.89	13.2	-10.11	5.68	45.01	9.82	10.81
High -risk	3.35	3.34	40.19	-17.79	7.04	122.00	18.89	23.33

TC, serum total cholesterol; GLU, glucose; TG, triglycerides; UA, uric acid; HDLC, high-density lipoprotein cholesterol; LDLC, low-density lipoprotein cholesterol; BMI, body mass index.

Table S7. Relative population flow analysis of the three PMR patterns. Related to Figure 5.

	LMR (%)		MMR (%)		HMR (%)		Sum
A) Classification according to the present number of MetS risk factors							
RFN0	2188	(96.2)	74	(3.3)	12	(0.5)	2274
RFN1	76	(2)	3731	(96.7)	50	(1.3)	3857
RFN2	74	(1.9)	3766	(96.1)	79	(2)	3919
RFN3	30	(1.2)	84	(3.3)	2468	(95.6)	2582
RFN4	3	(0.3)	19	(2.1)	900	(97.6)	922
Sum	2371	(17.5)	7674	(56.6)	3509	(25.9)	13554
B) Classification according to disease status							
HC	2188	(96.2)	74	(3.3)	12	(0.5)	2274
pre-MetS	150	(1.9)	7497	(96.4)	129	(1.7)	7776
MetS	33	(0.9)	103	(2.9)	3368	(96.1)	3504
Sum	2371	(17.5)	7674	(56.6)	3509	(25.9)	13554

RFN, number of traditional MetS risk factors; HC, healthy control; pre-MetS, pre metabolic syndrome; MetS, metabolic syndrome.

Table S8. m/z signals selected as hub metabolic features for pre-MetS and MetS screening and staging. Related to Figures 4 and 5.

ID	m/z	Accession[a)]	Potential biomarkers	Adduct Type
1	102.9925	HMDB0003361	Pyrimidine	$[M+Na]^+$
2	103.9375	HMDB0002166	(S)-beta-Aminoisobutyric acid	$[M+H]^+$
3	104.9725	HMDB0000011	(R)-3-Hydroxybutyric acid	$[M+H]^+$
4	105.8725	HMDB0000187	L-Serine	$[M+H]^+$
5	118.0225	HMDB0000128	Guanidoacetic acid	$[M+H]^+$
6	119.0125	HMDB0000754	3-Hydroxyisovaleric acid	$[M+H]^+$
7	119.9575	HMDB0000719	L-Homoserine	$[M+H]^+$
8	121.9375	HMDB0000574	L-Cysteine	$[M+H]^+$
9	129.8575	HMDB0000070	Pipecolic acid	$[M+H]^+$
10	131.9725	HMDB0000965	Hypotaurine	$[M+Na]^+$
11	136.0225	HMDB0000562	Creatinine	$[M+Na]^+$
12	136.9675	HMDB0000209	Phenylacetic acid	$[M+H]^+$
13	143.9875	HMDB0000574	L-Cysteine	$[M+Na]^+$
14	144.9775	HMDB0002994	Erythritol	$[M+Na]^+$
15	145.9225	HMDB0002243	Picolinic acid	$[M+Na]^+$
16	147.8575	HMDB0000251	Taurine	$[M+Na]^+$
17	151.7725	HMDB0000267	Pyroglutamic acid	$[M+Na]^+$
18	152.6275	HMDB0000634	Citraconic acid	$[M+Na]^+$
19	157.9375	HMDB0000156	Malic acid	$[M+Na]^+$
20	159.9625	HMDB0000306	Tyramine	$[M+Na]^+$
21	161.9425	HMDB0060665	Isonicotinic acid	$[M+K]^+$
22	163.9675	HMDB0000224	O-Phosphoethanolamine	$[M+Na]^+$
23	185.5675	HMDB0000641	L-Glutamine	$[M+K]^+$
24	195.7825	HMDB0000840	Salicyluric acid	$[M+H]^+$
25	203.0725	HMDB0000122	D-Glucose	$[M+Na]^+$
26	218.9125	HMDB0000122	D-Glucose	$[M+K]^+$

[a)] Compound ID from the Human Metabolome Database (https://hmdb.ca/).

Table S9. Differential metabolic pathways regulated among the HC, pre-MetS and MetS groups. Related to Figure 5.

Pathway Hit[a)]	***P* value**[b)]	**-Log (p)**	**Enrichment Ratio**[c)]
Taurine and Hypotaurine Metabolism	0.00217	2.664	10.676
Phenylacetate Metabolism	0.0172	1.764	9.479
Homocysteine Degradation	0.0172	1.764	9.479
Phosphatidylethanolamine Biosynthesis	0.0301	1.521	7.117
Sphingolipid Metabolism	0.0633	1.199	3.198
Methionine Metabolism	0.0755	1.122	2.97
Glutathione Metabolism	0.0843	1.074	4.065
Transfer of Acetyl Groups into Mitochondria	0.0915	1.039	3.876
Warburg Effect	0.15	0.824	2.206
Glycine and Serine Metabolism	0.156	0.807	2.174
Ammonia Recycling	0.171	0.767	2.667
Lactose Degradation	0.193	0.714	4.739
Ketone Body Metabolism	0.267	0.573	3.279
Glucose-Alanine Cycle	0.267	0.573	3.279
Phosphatidylcholine Biosynthesis	0.284	0.547	3.049
Glutamate Metabolism	0.32	0.495	1.739
Lactose Synthesis	0.38	0.42	2.132
Pantothenate and CoA Biosynthesis	0.395	0.403	2.033
Glycolysis	0.451	0.346	1.706
Cysteine Metabolism	0.464	0.333	1.642
Selenoamino Acid Metabolism	0.49	0.31	1.524
Urea Cycle	0.502	0.299	1.471
Citric Acid Cycle	0.537	0.27	1.333
Amino Sugar Metabolism	0.549	0.26	1.294
Aspartate Metabolism	0.57	0.244	1.22
Gluconeogenesis	0.57	0.244	1.22
Nicotinate and Nicotinamide Metabolism	0.591	0.228	1.153
Galactose Metabolism	0.601	0.221	1.122

Pyruvate Metabolism	0.688	0.162	0.893
Arginine and Proline Metabolism	0.725	0.14	0.806
Pyrimidine Metabolism	0.763	0.117	0.725
Valine, Leucine and Isoleucine Degradation	0.769	0.114	0.709
Bile Acid Biosynthesis	0.797	0.099	0.658
Tyrosine Metabolism	0.83	0.081	0.592
Purine Metabolism	0.838	0.077	0.578

[a)] Differential metabolic pathway analysis performed using MetaboAnalyst (5.0) using the website analysis module; [b)] p value calculated from pathway enrichment analysis; [c)] enrichment ratio generated from pathway topology analysis.

Table S10. Comparison between different NMR and mass spectrometry platforms. Related to Figure 1.

Methods	Sample volume	Pre-treatment	NMR/MS analysis
NMR	20 - 500 μ L	1 - 1.5 hour for 96 samples	12 ~ 30 min per sample
LC-MS	10 - 60 μ L	1 - 2 hour for 96 samples	12 ~ 24 min per sample
GC-MS	30 - 400 μ L	45 - 60 min per sample	~ 30 min per sample
LDI MS	100 nL	~ 1 min per sample	~ 30 second per sample

MS, mass spectrometry; NMR, nuclear magnetic resonance; LC-MS, liquid chromatography-mass spectrometry; GC-MS, gas chromatography-mass spectrometry; LDI-MS, laser desorption/ionization mass spectrometry.