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

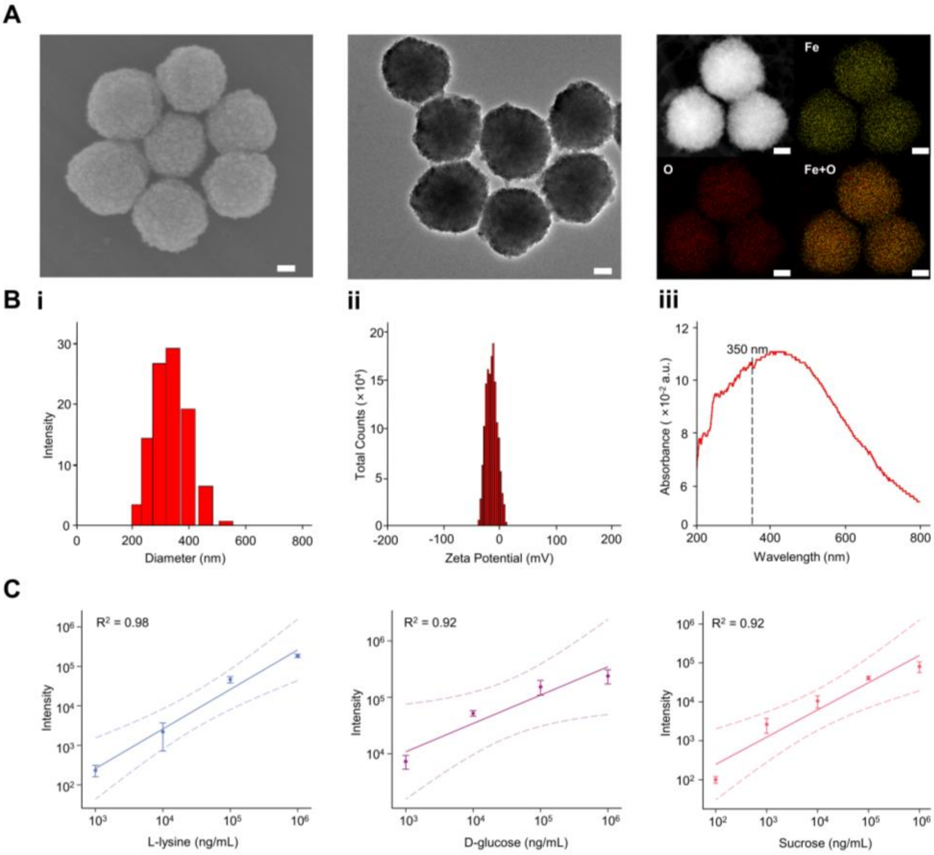


Figure S1. Material characterization of the ferric particles used in the LDI-MS process. Related to Figure 1. A) Electron micrograph images of the ferric particles. Scanning electron microscopy (SEM) images showed nanoscale surface roughness of ferric particles (n≥3 randomly selected). Transmission electron 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).

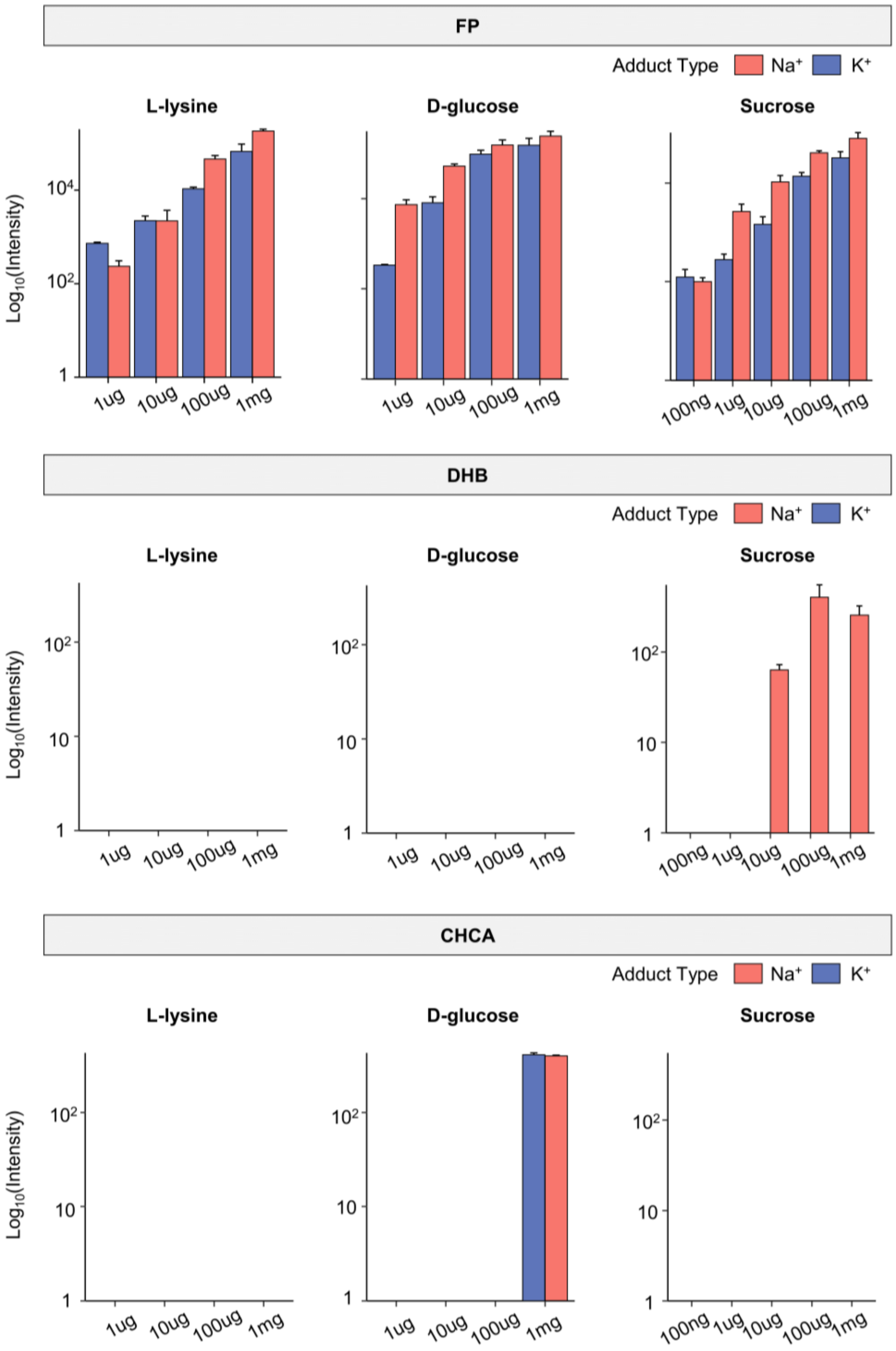


Figure S2. Quantification results for standards including L-lysine, D-glucose, and sucrose at different concentrations obtained by ferric particle, CHCA, and DHB-assisted LDI-MS. Related to Figure 1. n=3 independent mixed samples tested 5 times each. FP, ferric particles; CHCA, α-cyano-4-hydroxycinnamic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

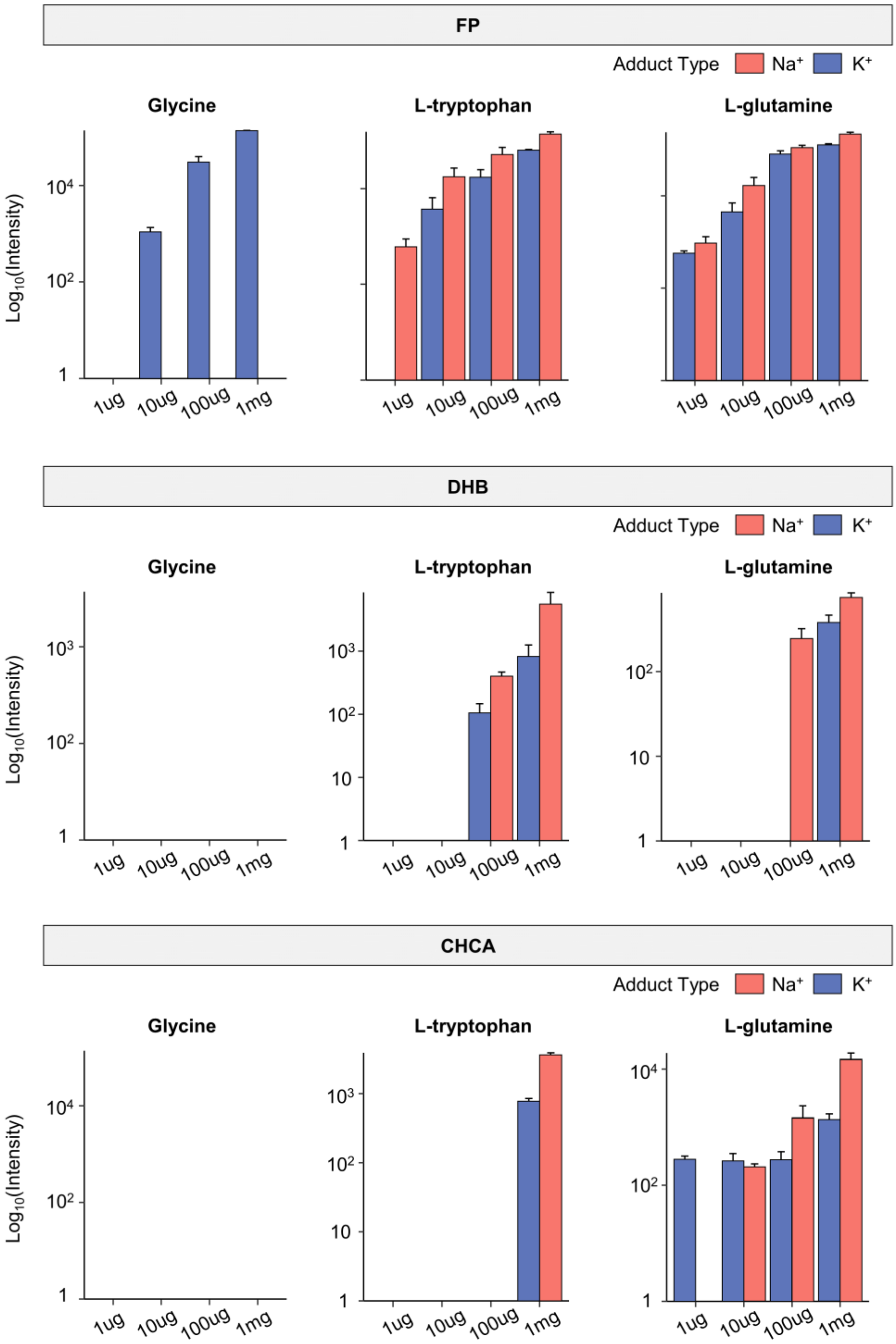


Figure S3. Quantification results for standards including glycine, L-tryptophan, and L-glutamine at different concentrations obtained by ferric particle, CHCA, and DHB-assisted LDI-MS. Related to Figure 1. n=3 independent mixed samples tested 5 times each. FP, ferric particles; CHCA, α-cyano-4-hydroxycinnamic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

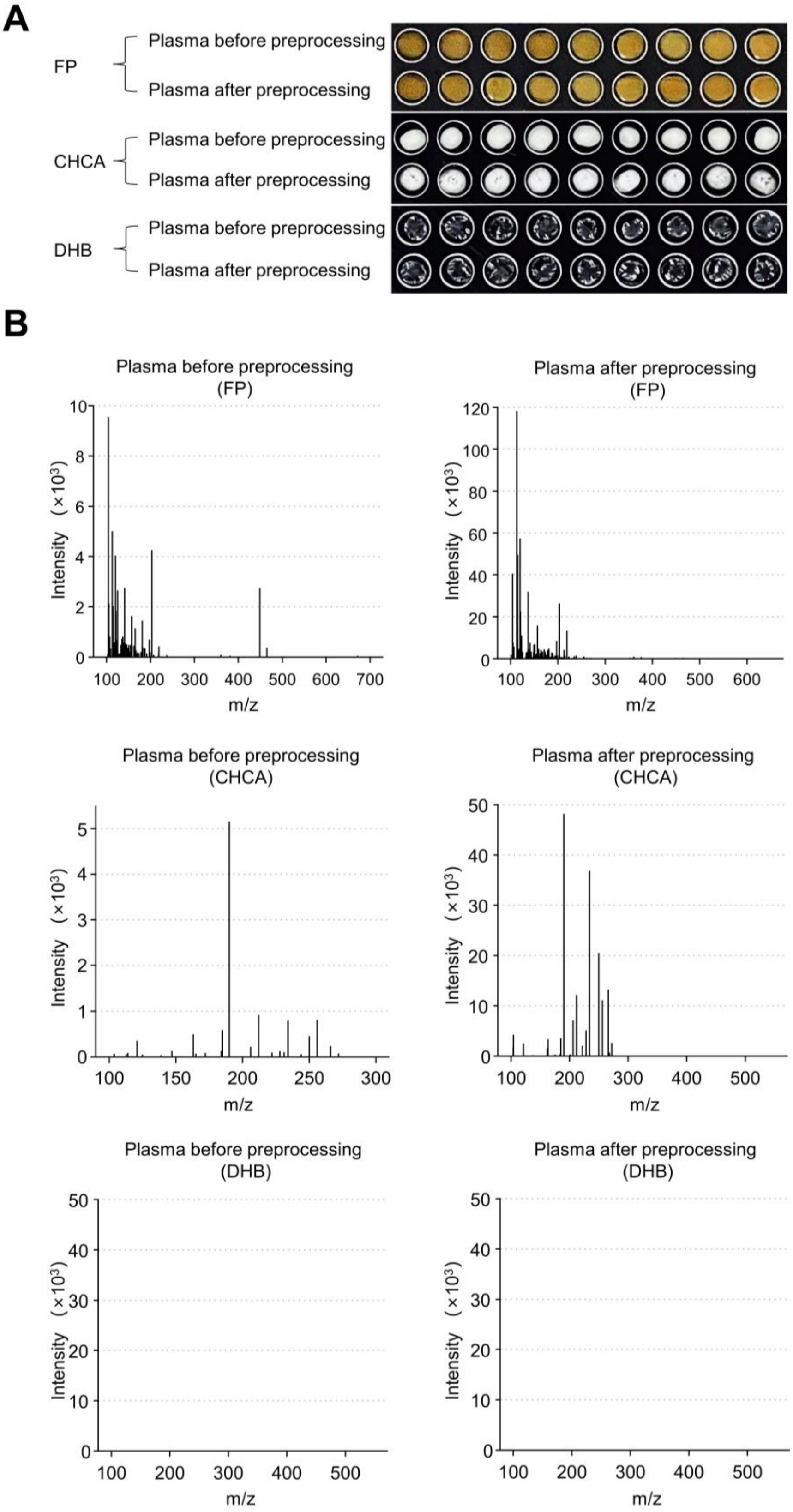


Figure S4. Plasma samples with and without pretreatment in three different matrices for LDI-MS. Related to Figure 1. A) Dried drops of the mixture of plasma samples and three different matrices including FP, CHCA, and DHB, on the plate. **B)** Typical mass spectrometry spectra of plasma samples

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.

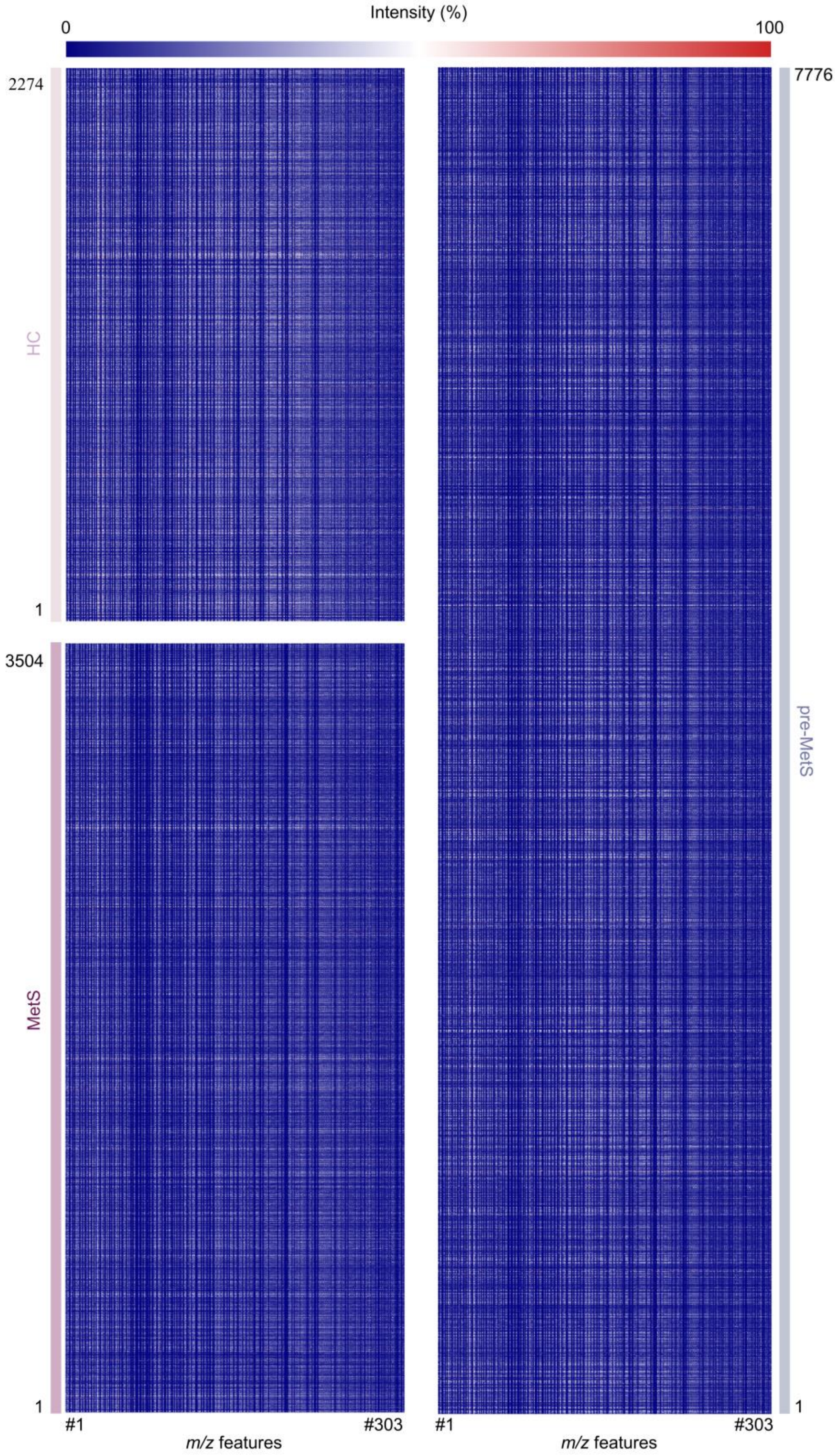

Intensity (%)
0
100
2274
HC
1
3504
MetS
1
7776
pre-MetS
1
#1
#303
m/z features
#1
#303
m/z features

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).

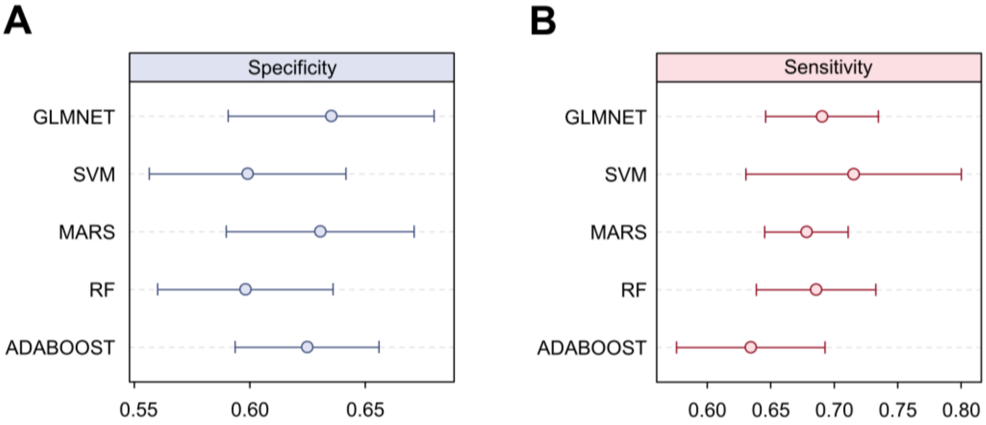


Figure S6. Distribution of performances of different machine-learning (ML)-based models for HC vs. MetS in the validation cohort (n=1,364). Related to Figure 3. A) Specificity of different ML-based models in the validation cohort. **B)** Sensitivity of different ML-based models in the validation cohort. 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.

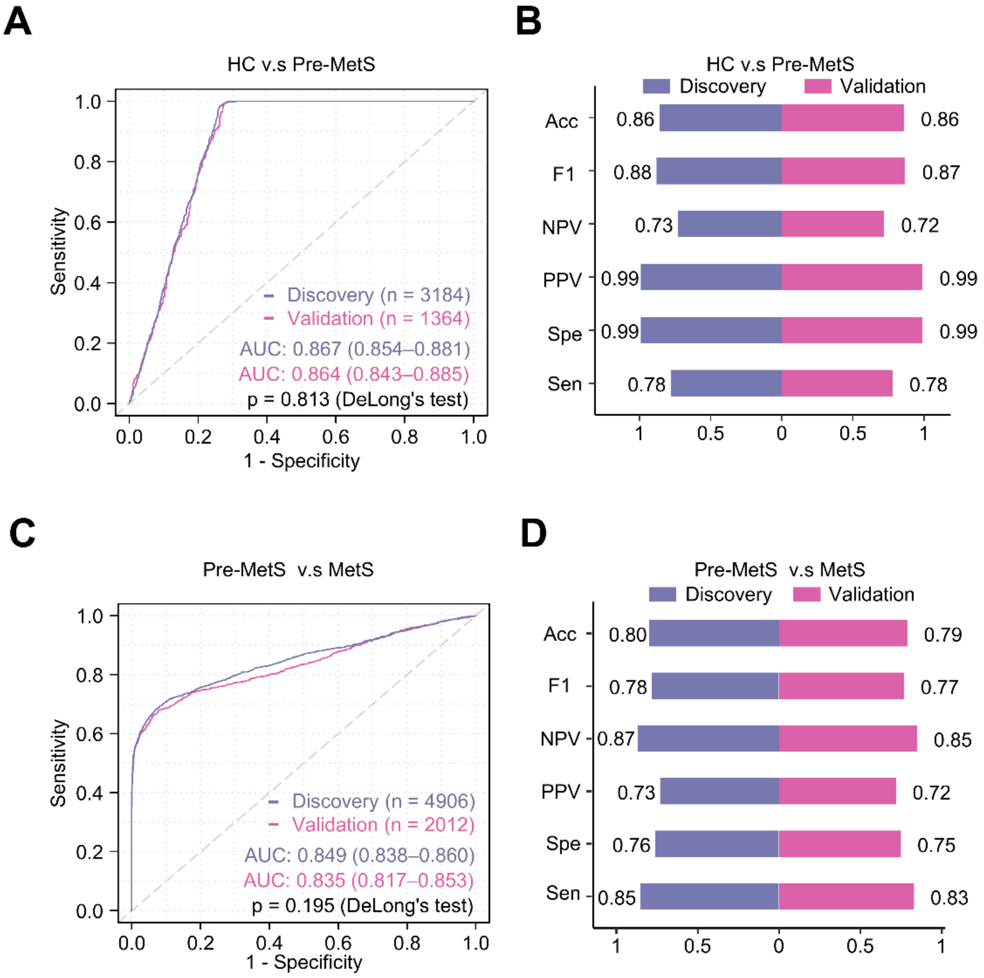


Figure S7. Construction of PMFs-based diagnostic model for HC vs. pre-MetS and pre-MetS vs. MetS. Related to Figure 3. A) Receiver operating characteristic (ROC) curves for HC vs. pre-MetS between the discovery (n=3,184) and validation (n=1,364) sets. **B)** Comparison of performances of the diagnostic 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.

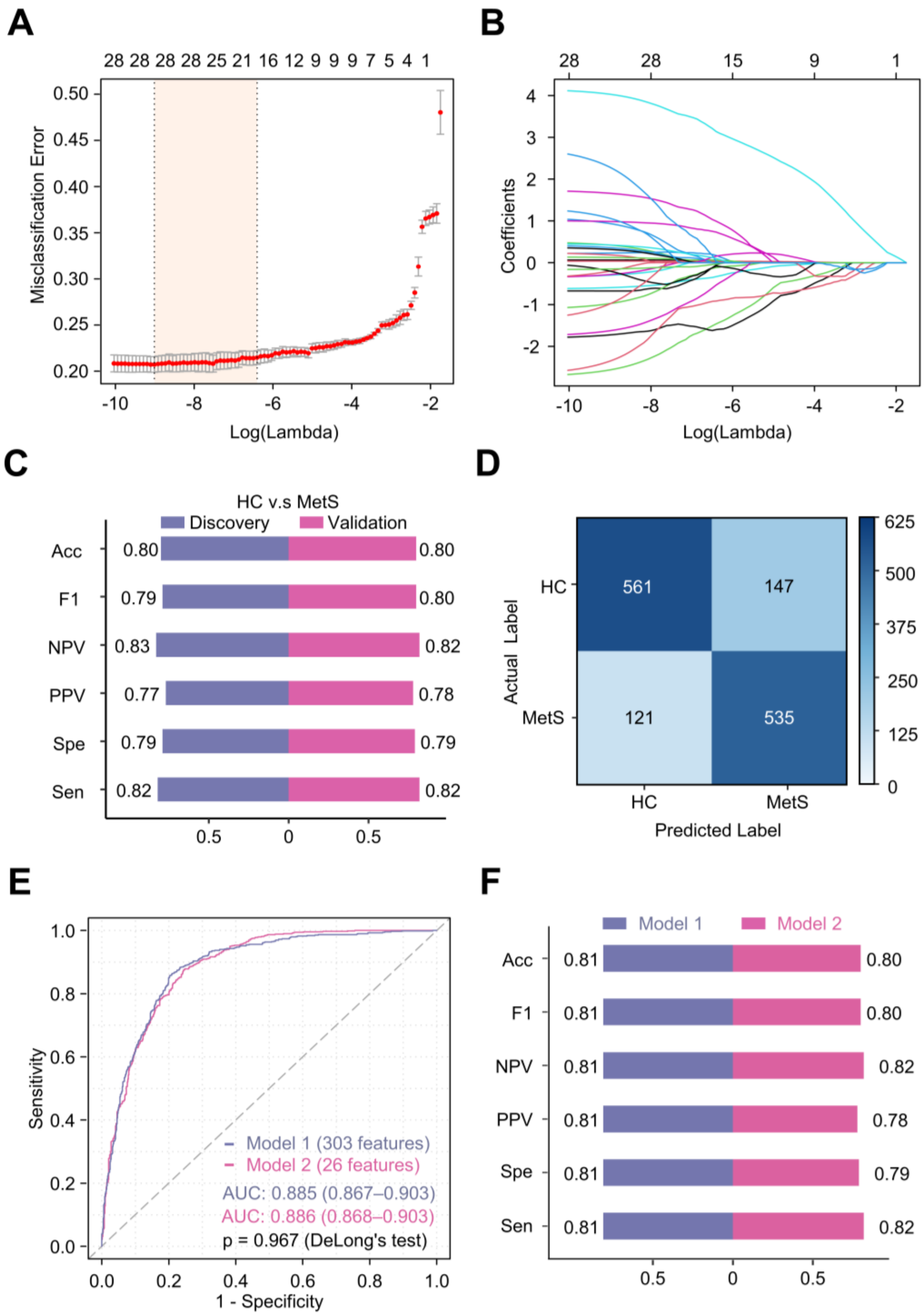


Figure S8. Construction of PMF-based diagnostic model for HC vs. MetS. Related to Figure 3. A) Generalized linear models via least absolute shrinkage and selection operator and elastic-net regularization (GLMNET) regression analysis results. The tuning parameter (lambda) was calculated based on the misclassification 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.

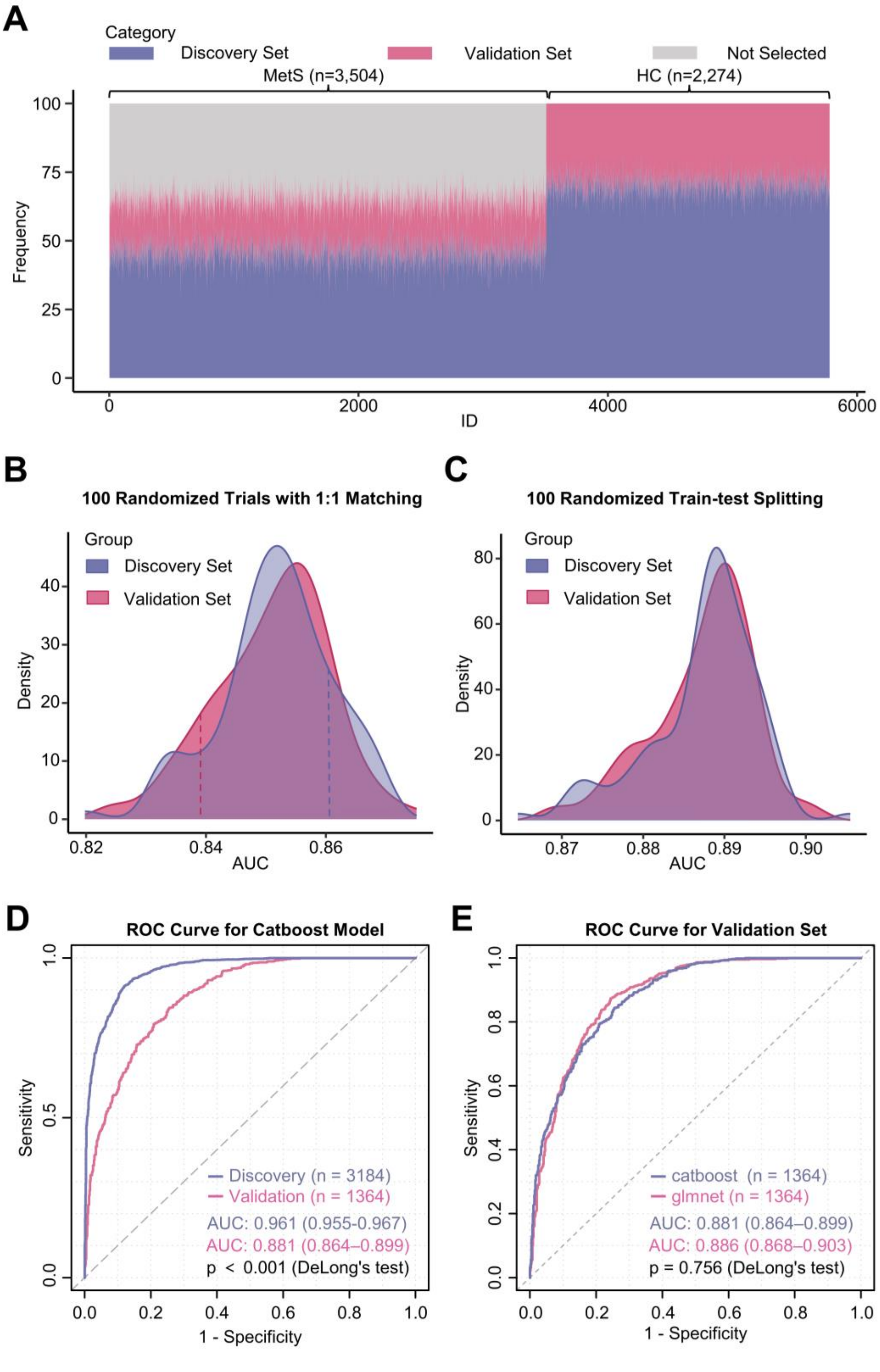
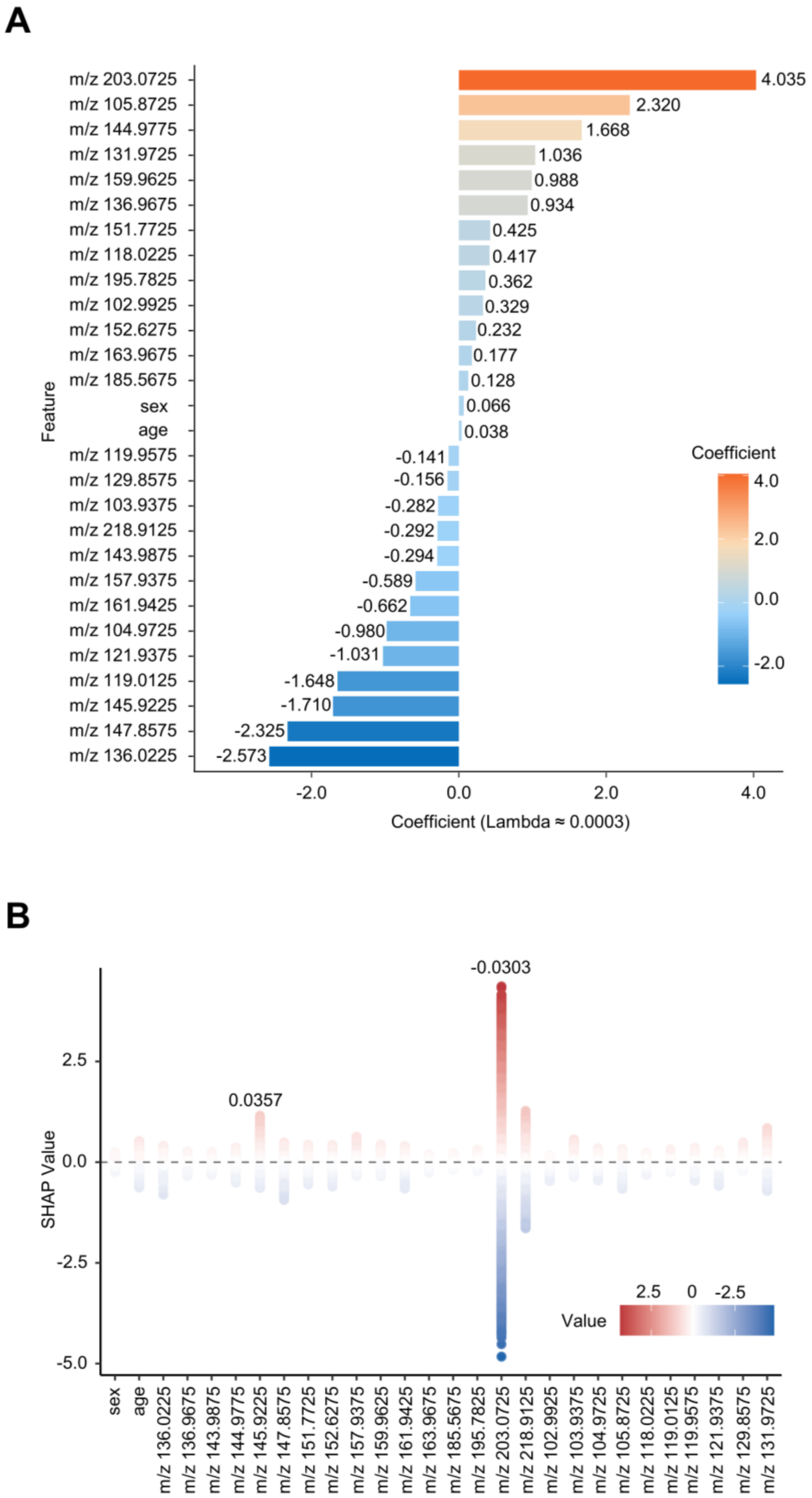


Figure S9. Validation of PMF-based diagnostic model for HC vs. MetS using 100 independent trials and Catboost-based models. Related to Figure 3. A) 100 independent randomized trials were conducted to generate discovery and validation sets with 7:3 split ratios from HC and MetS groups. **B)** Density 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
Feature
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
Coefficient
4.0
2.0
0.0
-2.0
-2.0
0.0
2.0
4.0
Coefficient (Lambda ≈ 0.0003)
B
SHAP Value
2.5
0.0
-2.5
-5.0
0.0357
-0.0303
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

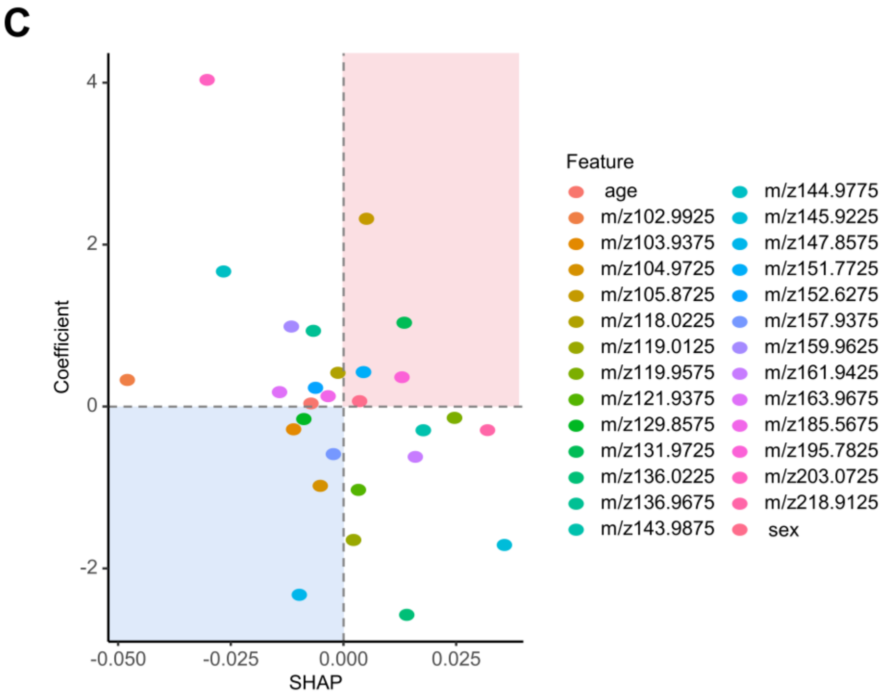


Figure S10. Feature importance of these 26 hub PMFs. Related to Figure 3. A) Feature coefficient of the generalized linear models via least absolute shrinkage and selection operator and elastic-net regularization regression analysis-based diagnostic model for HC vs. MetS with best lambda value 0.0003402896. 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.

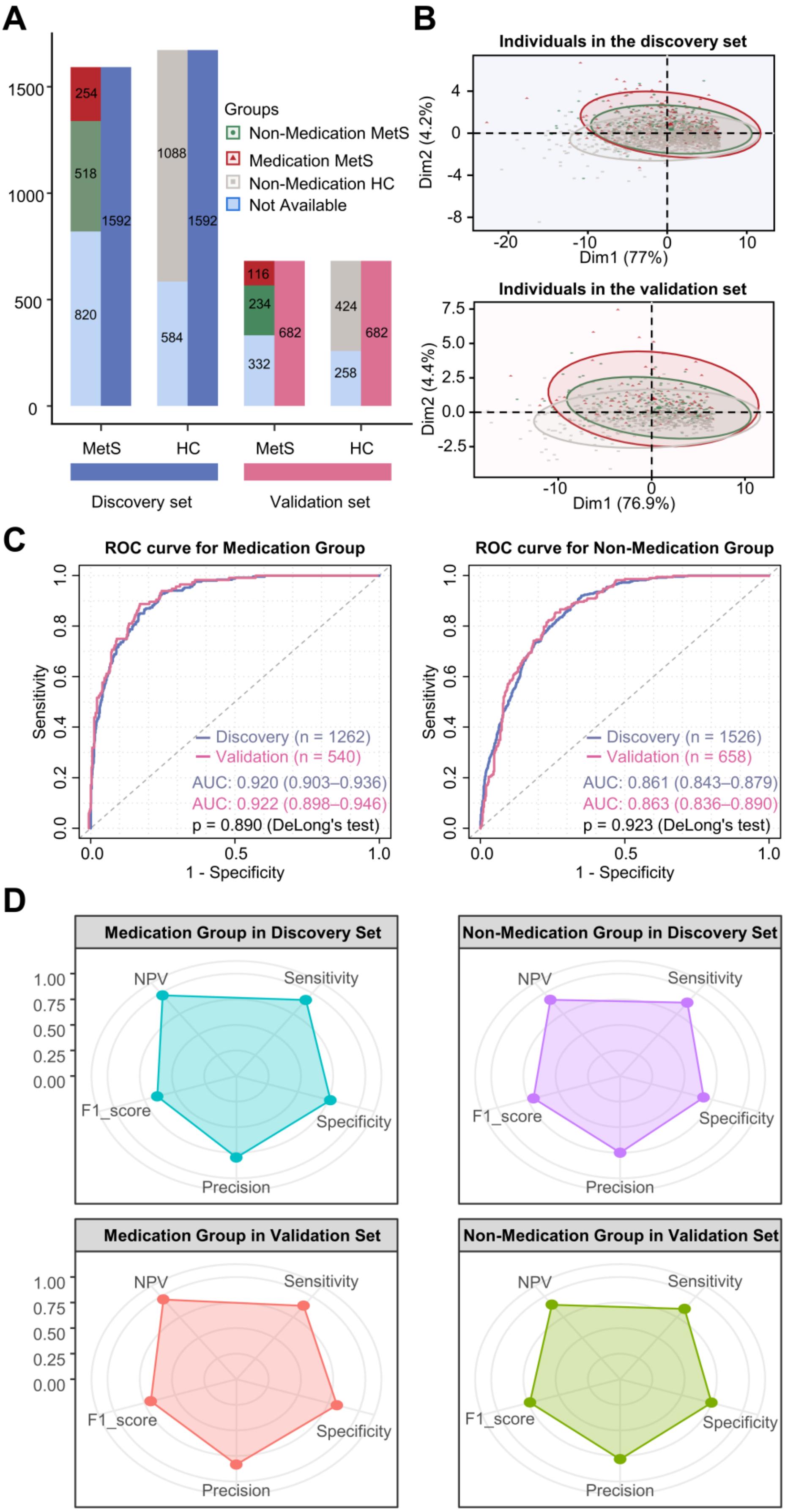

A
1500
1000
500
0
254
518
820
1592
1088
584
1592
116
234
332
682
424
258
682
MetS
HC
MetS
HC
Discovery set
Validation set
Groups
Non-Medication MetS
Medication MetS
Non-Medication HC
Not Available
B
Individuals in the discovery set
Dim2 (4.2%)
4
0
-4
-8
-20
-10
0
10
Dim1 (77%)
Individuals in the validation set
Dim2 (4.4%)
7.5
5.0
2.5
0.0
-2.5
-10
0
10
Dim1 (76.9%)
C
ROC curve for Medication Group
Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.5
1.0
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
Sensitivity
1 - Specificity
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
1.00
0.75
0.50
0.25
0.00
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.

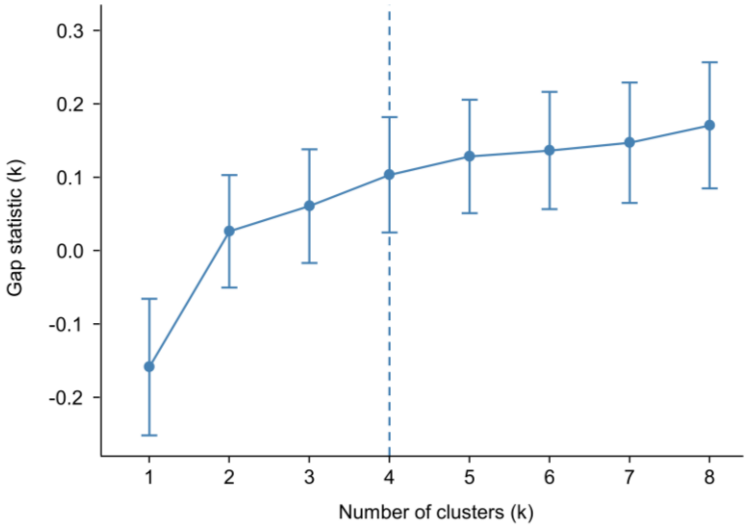


Figure S12. Gap statistic curve for choosing the optimal number of clusters. Related to Figure 4. The dotted vertical line suggests an optimal parameter value of the number of clusters (k) = 4 in our dataset.

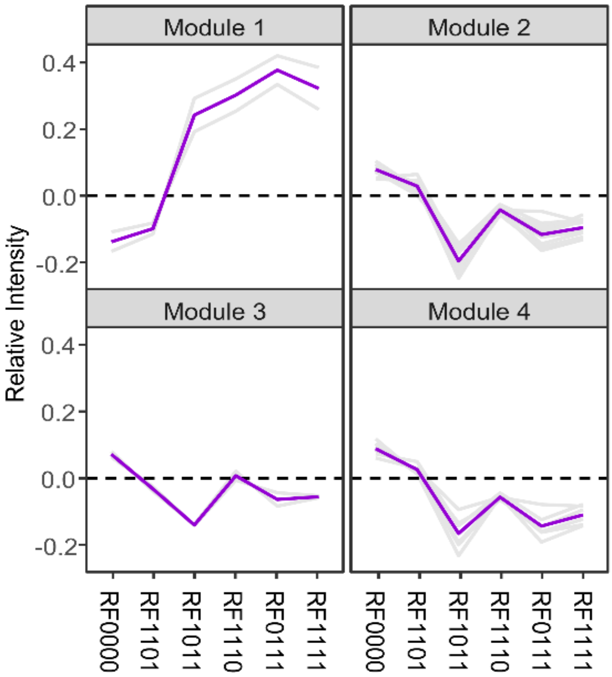
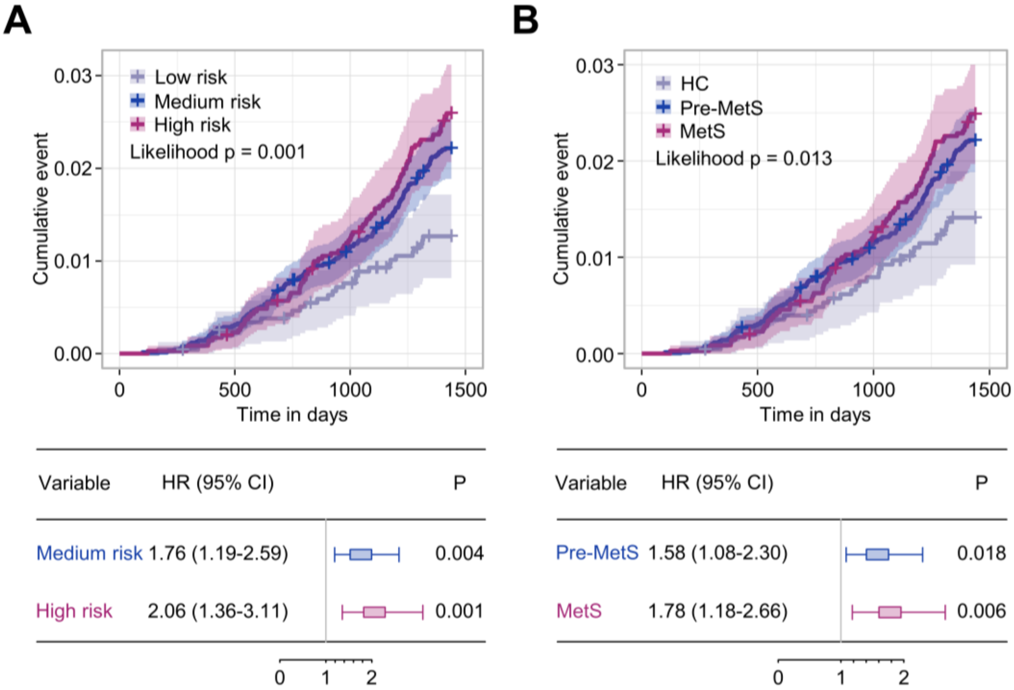


Figure S13. The relative intensity of four metabolic feature modules clustered through K-means algorithm in the five MetS subgroups and HC group. Related to Figure 4.



Variable	HR (95% CI)	P
Medium risk	1.76 (1.19-2.59)	0.004
High risk	2.06 (1.36-3.11)	0.001

Variable	HR (95% CI)	P
Pre-MetS	1.58 (1.08-2.30)	0.018
MetS	1.78 (1.18-2.66)	0.006

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.

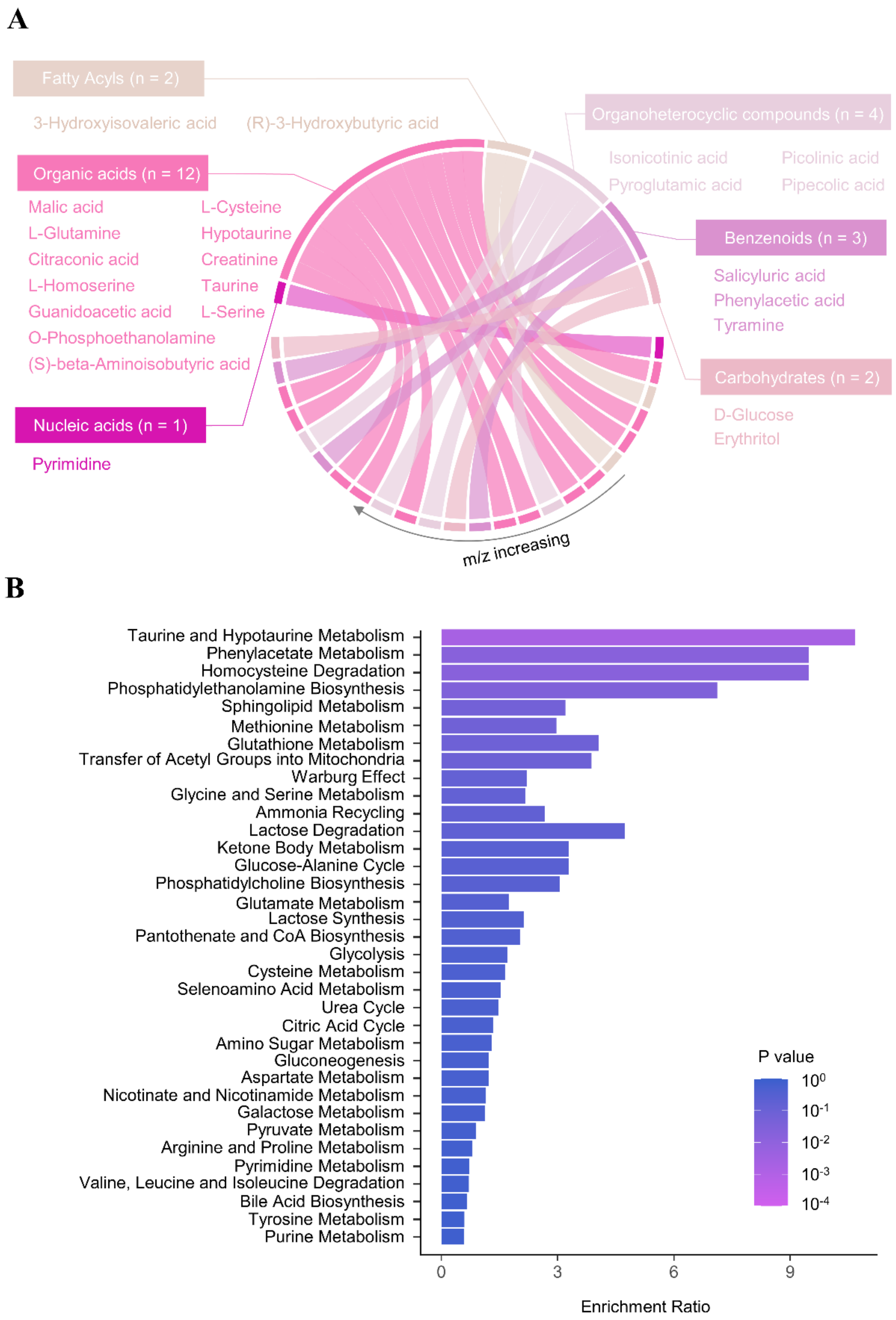


Figure S15. Potential biomarker identification and differential metabolic pathways in the pathological process of MetS. Related to Figure 5. A) Metabolite classification of the 26 hub PMFs by matching through the Human Metabolome Database (HMDB) and MetaboAnalyst 5.0. **B)** Potential metabolic pathways 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 μL	1 - 1.5 hour for 96 samples	12 ~ 30 min per sample
LC-MS	10 - 60 μL	1 - 2 hour for 96 samples	12 ~ 24 min per sample
GC-MS	30 - 400 μL	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.