Graphdiyne-Assisted LDI-MS for Rapid, Non-Invasive Urine Metabolomic Profiling in Tuberculosis Screening

Yile Yu†[1,2], Xiaoyan Zhong†[3], Xi Yu[1,2], Jinhang Li[1,2], Huihui Liu[1,2], Xin Liu [4*], Lixia Jiang[3*], Zongxiu Nie[1,2*]

1 Beijing National Laboratory for Molecular Sciences, Key Laboratory of Analytical Chemistry for Living Biosystems, Institute of Chemistry, Chinese Academy of Sciences, Beijing, 100190, China

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

3 Department of Laboratory Medicine, First Affiliated Hospital of Gannan Medical University, Nanchang, Jiangxi Province 341000, China

4 Department of Laboratory Medicine, Ganzhou Fifth People's Hospital, Ganzhou Key Laboratory of respiratory diseases, Ganzhou Institute of Respiratory Disease Prevention and control

† These authors contributed equally to this work.

Corresponding author:

***Huihui Liu** **Email: hhliu@iccas.ac.cn**

***Xin Lin** **Email: 15170732977@163.com**

***Lixia Jiang** **Email: jlx7310@gmu.edu.cn**

***Zongxiu Nie** **Email: znie@iccas.ac.cn**

Abstract

Tuberculosis (TB) remains a major global health burden. Here we report a rapid, noninvasive urine-based metabolomics approach using Graphdiyne (GD)-assisted laser desorption ionization mass spectrometry (LDI-MS) combined with machine learning. We applied GD-assisted LDI-TOF MS to urine samples from healthy controls (HC) and active TB patients, generating rich metabolite fingerprints. Supervised classifiers trained on the GD-assisted LDI-TOF MS spectral features achieved excellent discrimination, consistent with prior reports of Matrix-Assisted Laser Desorption Ionization-Time-of-flight-mass spectrometry (MALDI-TOF MS) with machine learning as a powerful screening tool. Key discriminatory urine metabolites and pathways were putatively annotated and included markers of altered energy, nucleotide, and amino-acid metabolism in TB patients, reflecting a shift in cellular energy handling and immune-related nucleotide turnover. These biochemical insights underscore TB-associated inflammatory and energetic perturbations. Overall, the GD-assisted LDI-TOF MS platform enables fast, high-throughput metabolite profiling and, when coupled with machine learning, offers a patient-friendly, noninvasive screening strategy for early TB detection and monitoring.

Methods

Machine learning method Machine learning analysis was conducted with Orange (Version 3.38.1). in Python 3.10. The build in classifier logistic regression (LR), gradient boosting (GB), neural network (NN), random forest (RF) and support vector machine (SVM) were applied. Model parameters were set as follows: LR in distinguishing RPL and HC: Regularization: Lasso (L1), C=1, class weights=False. GB in distinguishing RPL and HC: Method: Gradient Boosting (scikit-learn), Number of trees: 100, Learning rate: 0.100, Replicable training: Ture, Limit depth of individual trees: 3, Do not split subsets smaller than: 2, Fraction of training instances: 1.00. NN in distinguishing RPL and HC: Hidden layers: 100, Activation: ReLu, Solver: Adam, Alpha: 0.0001, Max iterations: 90, Replicable training: True. RF in distinguishing RPL and HC: Number of trees: 25, Number of attributes considered at each split: False, Limit depth of individual trees: 6, Do not split subsets smaller than: 12. SVM in distinguishing RPL and HC: SVM type: v-SVM, C=1.00, v=0.50, Kernel: RBF, $\exp(-\text{auto}|x-y|^2)$, Numerical tolerance: 0.0010, Iteration limit: 100.

Equations

In evaluating the performance of machine learning classifiers, key metrics including Area Under the Curve (AUC), Classification Accuracy (CA), Precision, Specificity and F1-score were employed. The effectiveness of these classifiers was assessed through the examination of true positive (TP), false positive (FP), true negative (TN), and false negative (FN) outcomes.

The calculation equation is as follows:

$$\text{Classifiction Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$F1 = \frac{2 \times \text{Precison} \times \text{Recall}}{\text{Precison} + \text{Recall}}.$$

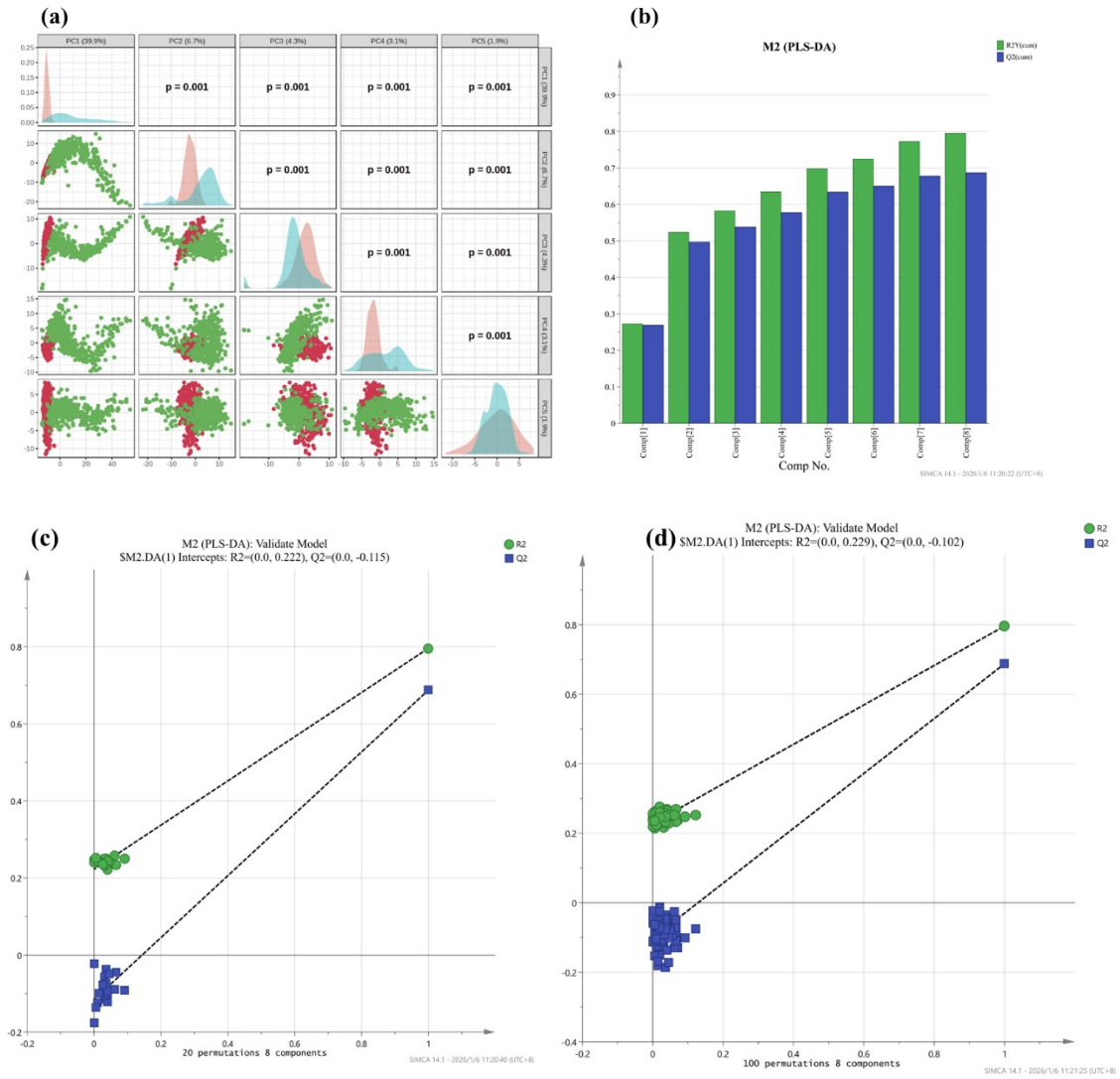


Figure S1 Overview of PCA results(a) and PLS-DA results after 5-fold cross-validation (b) in the classification of Tuberculosis (TB) patients and healthy control (HC). Permutation validation of the PLS-DA model with 8 components. Class labels were randomly permuted 20 (c) and 100 times (d), and the corresponding R^2 and Q^2 values were calculated for each permuted model. The original model (rightmost point) shows substantially higher R^2 and Q^2 values than those obtained from permuted models, indicating that the classification performance is unlikely to arise from random chance.

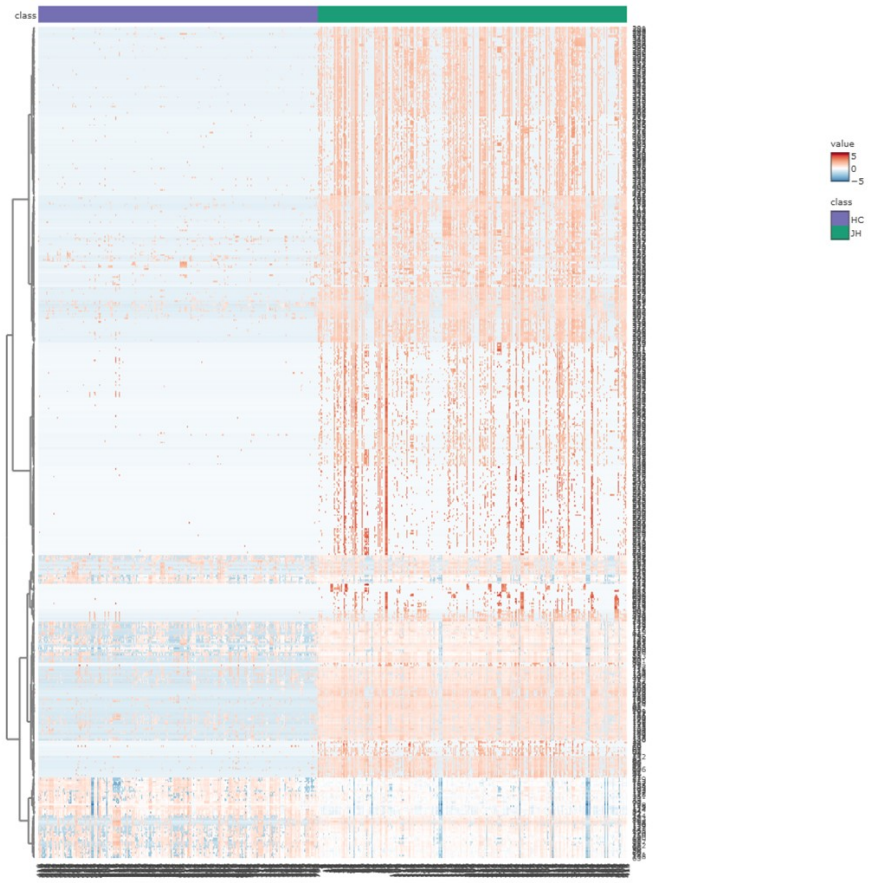


Figure S2 Heatmap of the full-spectrum metabolic features (rows: m/z features; columns: samples) showing z-scored intensities after mean-centering and unit-variance scaling. Samples are annotated by group (TB vs HC) and hierarchical clustering illustrates distinct metabolic patterns between the two groups.

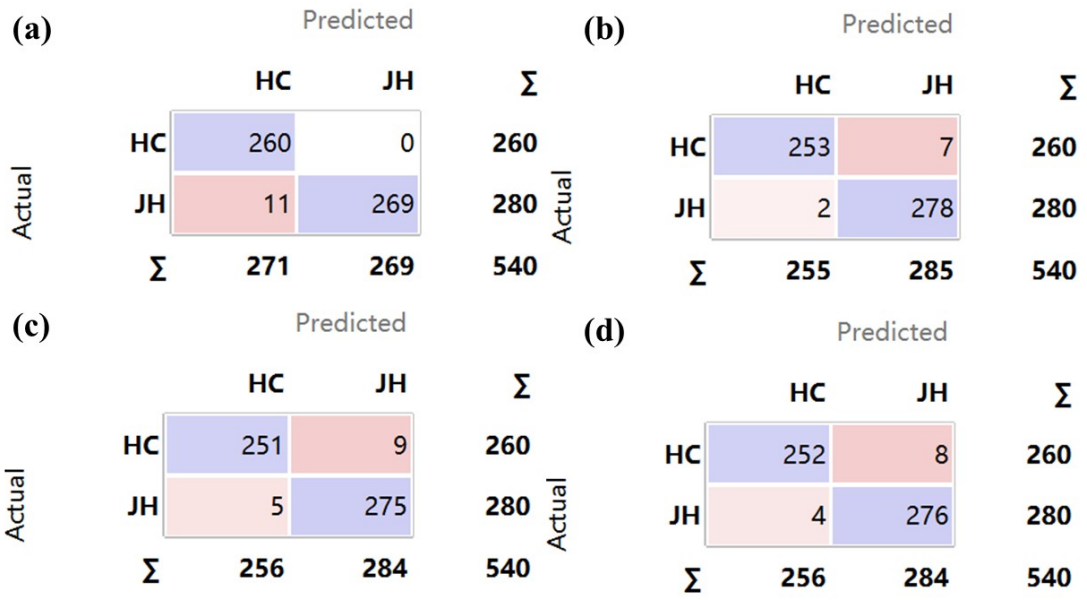


Figure S3 a) Confusion matrix of the training set with RF. (b) Confusion matrix of the training set with NN. (c) Confusion matrix of the training set with AB. (d) Confusion matrix of the training set with GB

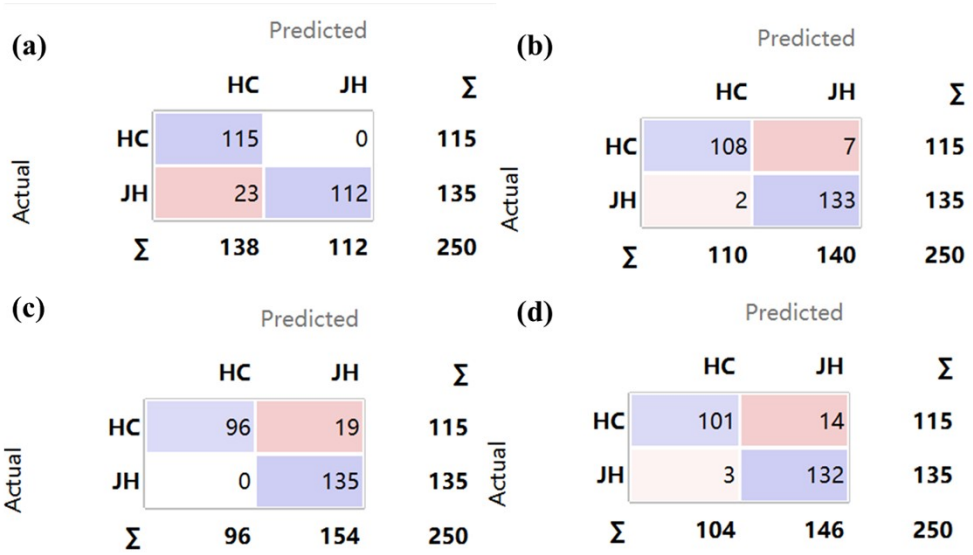


Figure S4 a) Confusion matrix of the test set with NN. (b) Confusion matrix of the test set with AB. (c) Confusion matrix of the test set with LR. (d) Confusion matrix of the test set with GB

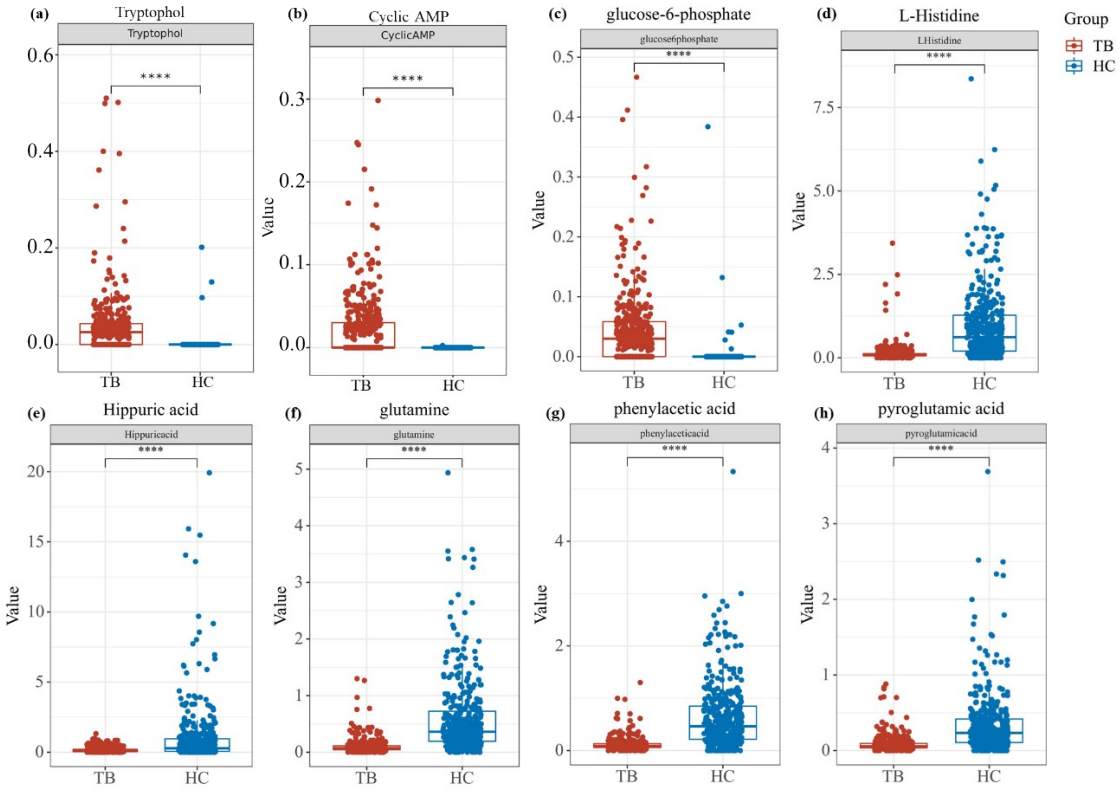


Figure S5 Box plots of upregulated metabolites (a) Tryptophol, (b) cyclic AMP, (c) glucose-6-phosphate and downregulated metabolites (d) L-histidine, (e) hippuric acid, (f) glutamine, (g) phenylacetic acid, (h) Pyroglutamic acid in TB patients.

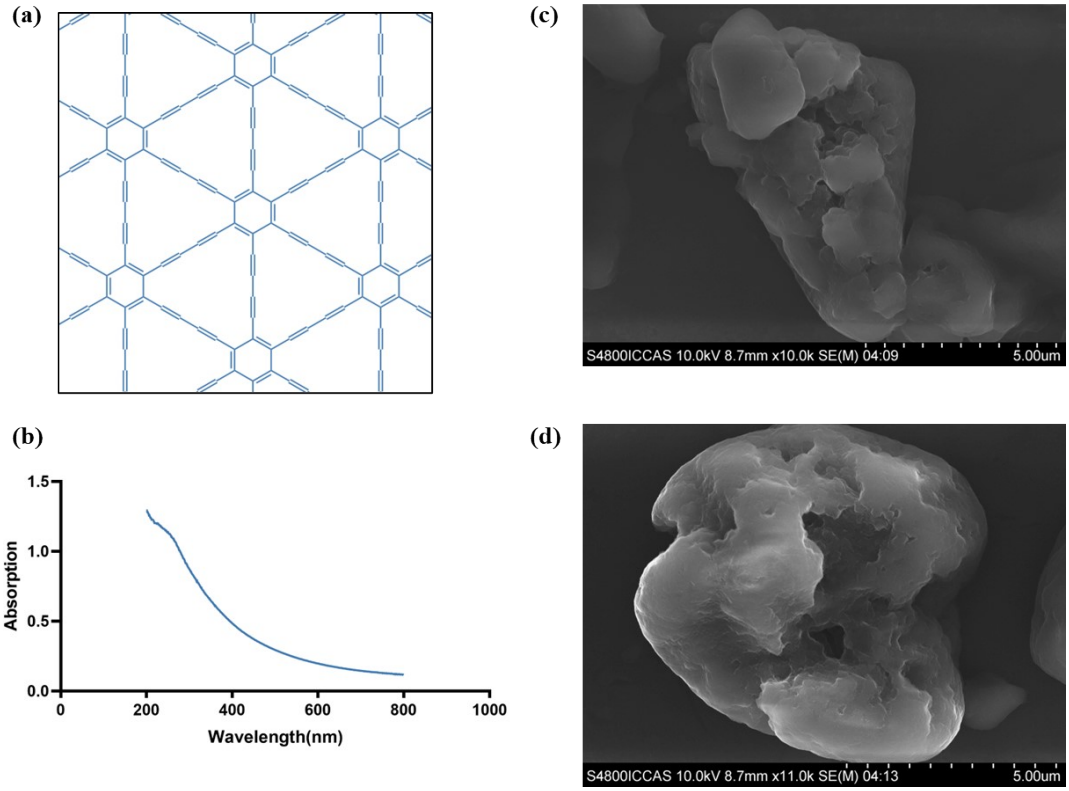


Fig. S6. Characterization of GD. (a) Molecular structure of GD. (b) UV–vis absorption spectrum of 0.1 mg/mL GD. (c) and (d) SEM image of GD (bar: 5.00 μm).

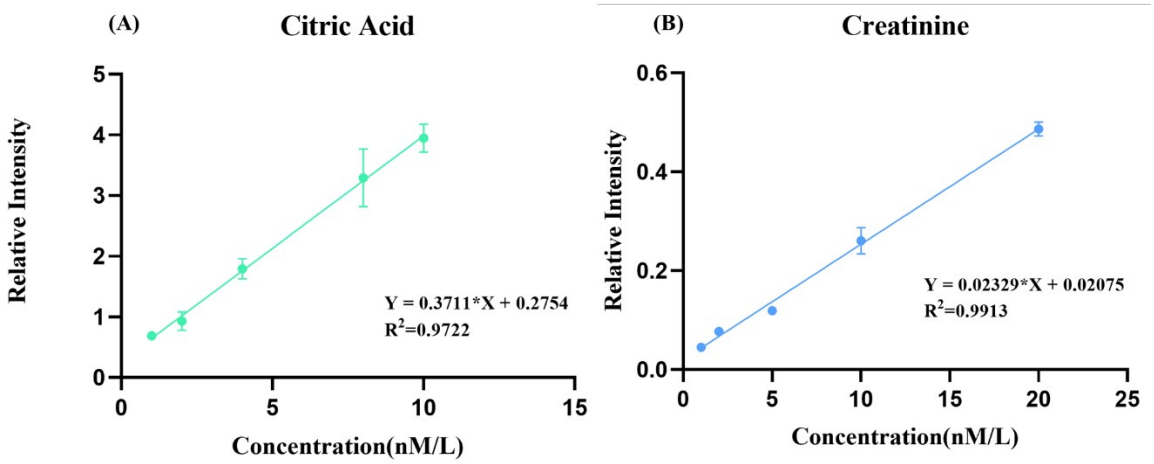
The molecular structure of graphdiyne (GD) consists of benzene rings interconnected by diacetylene linkages, forming an extended π-conjugated network (Fig. S6a). UV–visible absorption and electron microscopy were employed to evaluate the optical and morphological properties relevant to laser desorption/ionization performance.

UV–vis spectroscopy. UV–vis absorption spectra were recorded for GD dispersions (prepared by sonicating GD [amount; e.g., X mg] in [solvent; e.g., DMF or ethanol] at a concentration of [Y mg mL^{-1}]) using a UV–vis spectrophotometer (model: [replace with instrument]). The spectrum (Fig. S6b) displays a broad, intense absorption band spanning approximately 220–400 nm, thereby fully covering the 355 nm Nd:YAG laser line used in this study. The strong UV absorbance is consistent with the extended π-conjugation of GD and supports efficient coupling of laser energy during LDI.

Scanning electron microscopy (SEM). Morphological analysis was performed on samples prepared by drop-casting a dilute GD dispersion onto a silicon (or carbon) substrate and drying under ambient conditions. SEM imaging (Fig. 1c,d and Supporting Fig. S6c) shows few-layered, sheet-like structures with lateral dimensions of [typical range, e.g., hundreds of nm to μm — replace if measured]. These micrographs confirm the two-dimensional

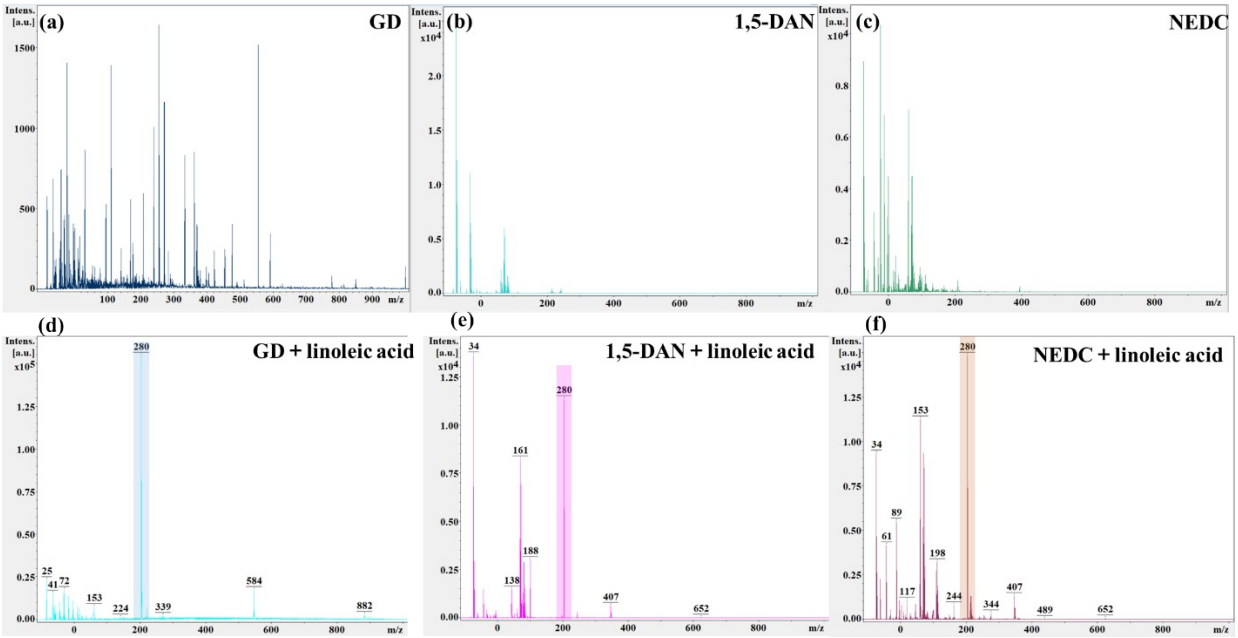
morphology expected for GD and are consistent with an exfoliated, layered architecture.

Summary. Together, the structural (Fig. S6a), optical (Fig. S6b) and morphological (Fig. 1c,d; Fig. S6c) data demonstrate that the GD material forms well-defined 2D layers and possesses strong ultraviolet absorbance — key attributes that underpin its performance as an LDI substrate.



Supplementary Fig. S7. Calibration curves for citric acid and creatinine.

Calibration curves for citric acid (a) and creatinine (b) obtained on the GD-LDI platform. Peak areas were normalized to the internal standard (3-chloro-L-phenylalanine). Linear regression equations and R^2 values are shown in each panel. Apparent LOD and LOQ were estimated as described in Methods.



Supplementary Fig. S8. Representative MALDI mass spectra acquired using different matrices. (a–c) Matrix-only spectra obtained with GD, 1,5-DAN, and NEDC, respectively. (d–f) Mass spectra of linoleic acid acquired using GD, 1,5-DAN, and NEDC, respectively, under identical experimental conditions.

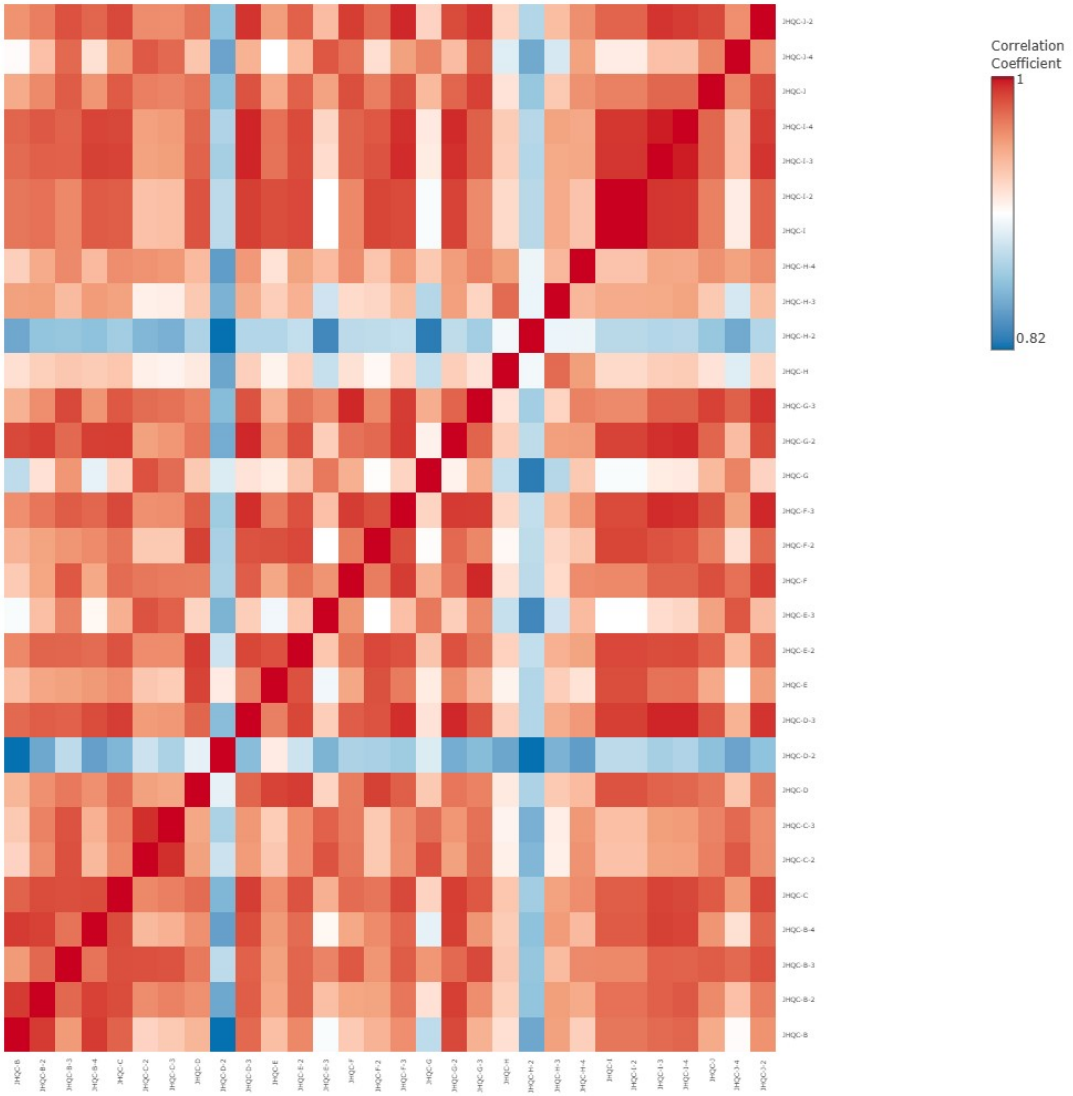


Figure S9. Correlation heatmap of quality control (QC) samples.

This heatmap depicts the Pearson correlation coefficients among the QC samples, which were injected at intervals throughout the analytical sequence, indicates high similarity across all QC replicates. The color scale represents the correlation coefficient, from strong positive correlation (red, 1) to strong negative correlation (blue, 0.82).

Table. S1 The urine metabolites identified by AP-SMALDI MS in negative ion mode

Name	formula	Adduct ions	Theoretical adduct m/z	Experimental	ΔPPM
Phosphate	H_3O_4P	M-H	96.9597	96.9597	0
Isovaleric acid	$C_5H_{10}O_2$	M-H	101.0597	101.0604	7
2-Furoic acid	$C_5H_4O_3$	M-H	111.0077	111.0085	7
Fumaric acid	$C_4H_4O_4$	M-H	115.0026	115.0034	7
Benzoic acid	$C_7H_6O_2$	M-H	121.0284	121.0293	7
Pyruvic acid	$C_3H_4O_3$	M+Cl	122.9851	122.9843	7
Pyroglutamic acid	$C_5H_7NO_3$	M-H	128.0342	128.0353	9
L-Aspartic acid	$C_4H_7NO_4$	M-H	132.0291	132.0303	9
Malic acid	$C_4H_6O_5$	M-H	133.0132	133.0143	8
Phenylacetic acid	$C_8H_8O_2$	M-H	135.0441	135.0452	8
Salicylic acid	$C_7H_6O_3$	M-H	137.0233	137.0245	9
Malonic acid	$C_3H_4O_4$	M+Cl	138.9804	138.9793	8
L-Serine	$C_3H_7NO_3$	M+Cl	140.0109	140.0118	6
L-Glyceric acid	$C_3H_6O_4$	M+Cl	140.996	140.9949	8
p-Cresol	C_7H_8O	M+Cl	143.0258	143.0253	3
Fumaric acid	$C_4H_4O_4$	M+Cl	150.9792	150.9805	9
Succinic acid	$C_4H_6O_4$	M+Cl	152.9949	152.996	7
3-Methylcrotonylglycine	$C_7H_{11}NO_3$	M-H	156.0655	156.065	3
Hypoxanthine	$C_5H_4N_4O$	M+Cl	171.0068	171.0066	1
Tartaric acid	$C_4H_6O_6$	M+Cl	184.9847	184.9859	6
2-Hydroxyadipic acid	$C_6H_{10}O_5$	M+Cl	197.0224	197.0211	7

Name	formula	Adduct ions	Theoretical adduct m/z	Experimental	ΔPPM
L-Phenylalanine	$C_9H_{11}NO_2$	M+Cl	200.0473	200.0469	2
N-Acetyltaurine	$C_4H_9NO_4S$	M+Cl	201.9945	201.9935	5
Ascorbic acid	$C_6H_8O_6$	M+Cl	211.0004	211.0016	6
Isoxanthopterin	$C_6H_5N_5O_2$	M+Cl	214.0126	214.0139	6
Carnosine	$C_9H_{14}N_4O_3$	M-H	225.0982	225.0999	8
D-Glucuronic acid	$C_6H_{10}O_7$	M+Cl	229.011	229.0123	6
(R)-Lipoic acid	$C_8H_{14}O_2S_2$	M+Cl	241.0118	241.0121	1
Deoxycytidine	$C_9H_{13}N_3O_4$	M+Cl	262.0589	262.057	7
Hydroxypropionylcarnitine	$C_{10}H_{19}NO_5$	M+Cl	268.0946	268.0966	7
Tiglylcarnitine	$C_{12}H_{21}NO_4$	M+Cl	278.1154	278.1153	0
Canavaninosuccinate	$C_9H_{16}N_4O_7$	M-H	291.0935	291.091	9
Xanthosine	$C_{10}H_{12}N_4O_6$	M+Cl	319.044	319.045	3
Orotidine	$C_{10}H_{12}N_2O_8$	M+Cl	323.0277	323.0286	3
1-Salicylate glucuronide	$C_{13}H_{14}O_9$	M+Cl	349.0321	349.0322	0
Cyclic AMP	$C_{10}H_{12}N_5O_6P$	M+Cl	364.0208	364.0225	5
6-Hydroxy-5-methoxyindole glucuronide	$C_{15}H_{17}NO_8$	M+Cl	374.0637	374.0613	6
Dihydrocaffeic acid 3-O-glucuronide	$C_{15}H_{18}O_{10}$	M+Cl	393.0583	393.0555	7
Epinephrine glucuronide	$C_{15}H_{21}NO_9$	M+Cl	394.0899	394.0869	8
Tryptophan 2-C-mannoside	$C_{17}H_{22}N_2O_7$	M+Cl	401.111	401.1124	3
beta-1,4-Mannosyl-N-acetylglucosamine	$C_{14}H_{25}NO_{11}$	M+Cl	418.1111	418.11	3
4-Hydroxy-5-(3',5'-dihydroxyphenyl)-valeric acid-O-glucuronide	$C_{17}H_{22}O_{11}$	M+Cl	437.0845	437.0872	6

Name	formula	Adduct ions	Theoretical adduct m/z	Experimental	ΔPPM
3-hydroxyhexadecanoyl carnitine	$C_{23}H_{45}NO_5$	M+Cl	450.2996	450.298	4
Glycocholic acid	$C_{26}H_{43}NO_6$	M+Cl	500.2773	500.2791	4
Cholesterol sulfate	$C_{27}H_{46}O_4S$	M+Cl	501.28	501.2821	4
Enterolactone 3'-glucuronide	$C_{24}H_{26}O_{10}$	M+Cl	509.1233	509.1209	5
3'-O-Methylepicatechin 7-O-glucuronide	$C_{22}H_{24}O_{12}$	M+Cl	515.096	515.095	2
Uridine diphosphate glucose	$C_{15}H_{24}N_2O_{17}P_2$	M-H	565.0466	565.0484	3
Cholic acid glucuronide	$C_{30}H_{48}O_{11}$	M-H	583.3112	583.3055	10
Deoxycholic acid 3-glucuronide	$C_{30}H_{48}O_{10}$	M+Cl	603.2901	603.2861	7

Table. S2 The urine metabolites identified by Bruker solarix FT-ICR mass spectrometer Bruker in negative ion mode

Name	formula	Adduct ions	Theoretical adduct m/z	Experimental	ΔPPM
Creatinine	$C_4H_7N_3O$	M-H	112.0505	112.0516	10
Erythronilic acid	$C_5H_{10}O_3$	M-H	117.0546	117.0557	9
Glutamine	$C_5H_{10}N_2O_3$	M-H	145.0608	145.0619	8
L-Histidine	$C_6H_9N_3O_2$	M-H	154.0622	154.06223	0
Ribonic acid	$C_5H_{10}O_6$	M-H	165.0394	165.0405	7
Uric acid	$C_5H_4N_4O_3$	M-H	167.0211	167.02109	0
Methylhistidine	$C_7H_{11}N_3O_2$	M-H	168.0768	168.0779	7
Aconitic acid	$C_6H_6O_6$	M-H	173.0092	173.00919	0
N-Acetyl-aspartic acid	$C_6H_9NO_5$	M-H	174.0408	174.04081	0
Ascorbic acid	$C_6H_8O_6$	M-H	175.0248	175.02484	0
Pyrophosphate	$H_4O_7P_2$	M-H	176.936	176.93598	0
Hippuric acid	$C_9H_9NO_3$	M-H	178.051	178.05099	0
Glucose	$C_6H_{12}O_6$	M-H	179.055	179.0561	6
Methyluric acid	$C_6H_6N_4O_3$	M-H	181.0367	181.03674	0
3-Dehydroquinic acid	$C_7H_{10}O_6$	M-H	189.0405	189.04049	0
Citric acid	$C_6H_8O_7$	M-H	191.0197	191.01975	0
Iduronic acid	$C_6H_{10}O_7$	M-H	193.0354	193.0354	0
Salicyluric acid	$C_9H_9NO_4$	M-H	194.0459	194.04587	0
Gluconic acid	$C_6H_{12}O_7$	M-H	195.051	195.05104	0
Tryptophol	$C_{10}H_{11}NO$	M+Cl	196.0524	196.0535	6
Xanthurenic acid	$C_{10}H_7NO_4$	M-H	204.0291	204.0302	5

Name	formula	Adduct ions	Theoretical adduct m/z	Experimental	ΔPPM
2-Methylcitric acid	$C_7H_{10}O_7$	M-H	205.0354	205.03541	0
3-(3,4-Dimethoxyphenyl)-2-propenoic acid	$C_{11}H_{12}O_4$	M-H	207.0652	207.0663	5
Glucaric acid	$C_6H_{10}O_8$	M-H	209.0303	209.03034	0
6-Carboxy-5,6,7,8-tetrahydropterin	$C_7H_9N_5O_3$	M-H	210.0633	210.06334	0
Gluconolactone	$C_6H_{10}O_6$	M+Cl	213.016	213.0171	5
Glucose	$C_6H_{12}O_6$	M+Cl	215.0317	215.0328	5
Ethyl glucuronide	$C_8H_{14}O_7$	M-H	221.0656	221.0667	5
Prolylhydroxyproline	$C_{10}H_{16}N_2O_4$	M-H	227.1037	227.10382	1
L-Cystine	$C_6H_{12}N_2O_4S_2$	M-H	239.0155	239.0166	5
2,4-Dihydroxy-7,8-dimethoxy-2H-1,4-benzoxazin-3(4H)-one	$C_{10}H_{11}NO_6$	M-H	240.0514	240.05138	0
Cytidine	$C_9H_{13}N_3O_5$	M-H	242.0782	242.07829	0
Uridine	$C_9H_{12}N_2O_6$	M-H	243.0623	243.06228	0
Methylphenobarbital	$C_{13}H_{14}N_2O_3$	M-H	245.0921	245.0932	4
Avenic acid B	$C_8H_{15}NO_6$	M+Cl	256.0582	256.0593	4
Ribothymidine	$C_{10}H_{14}N_2O_6$	M-H	257.0779	257.07794	0
Glucose-6-phosphate	$C_6H_{13}O_9P$	M-H	259.0213	259.0224	4
O-Sulfotyrosine	$C_9H_{11}NO_6S$	M-H	260.0223	260.0234	4
Homovanillic acid sulfate	$C_9H_{10}O_7S$	M-H	261.0074	261.00743	0
Phenylacetylglutamine	$C_{13}H_{16}N_2O_4$	M-H	263.1037	263.10376	0
5-Carboxy-2'-deoxyuridine	$C_{10}H_{12}N_2O_7$	M-H	271.0572	271.05714	0

Name	formula	Adduct ions	Theoretical adduct m/z	Experimental	ΔPPM
Diphenol glucuronide	$C_{12}H_{14}O_8$	M-H	285.0605	285.0616	4
N-Acetylglucosamine 6-sulfate	$C_8H_{15}NO_9S$	M-H	300.0395	300.03945	0
Pyrogallol-2-O-glucuronide	$C_{12}H_{14}O_9$	M-H	301.0554	301.0565	4
Indoleacetyl glutamine	$C_{15}H_{17}N_3O_4$	M-H	302.1135	302.115	5
Glutathione	$C_{10}H_{17}N_3O_6S$	M-H	307.0838	306.0754	4
N-Acetylneuraminic acid	$C_{11}H_{19}NO_9$	M-H	308.0987	308.09871	0
Acetaminophen glucuronide	$C_{14}H_{17}NO_8$	M-H	326.0851	326.087	6
Neomenthol-glucuronide	$C_{16}H_{28}O_7$	M-H	331.1751	331.1762	3
Tyrosyl-Tyrosine	$C_{18}H_{20}N_2O_5$	M-H	343.1299	343.12993	0
Nepetaside	$C_{16}H_{26}O_8$	M-H	345.1544	345.1555	3
Ipratropium bromide	$C_{20}H_{30}NO_3$	M+Cl	367.1909	367.1909	0
Testosterone sulfate	$C_{19}H_{28}O_5S$	M-H	367.1585	367.15846	0
Androsterone sulfate	$C_{19}H_{30}O_5S$	M-H	369.1741	369.17415	0
5-(3',4'-Dihydroxyphenyl)-gamma-valerolactone-4'-O-glucuronide	$C_{17}H_{20}O_{10}$	M-H	383.0984	383.09845	0
Chondroitin	$(C_{14}H_{21}NO_{11})_nH_2O$	M-H	396.1147	396.11474	0
Sinapinic acid-O-glucuronide isomer	$C_{17}H_{20}O_{11}$	M-H	399.0933	399.09337	0
Fenoprofen glucuronide	$C_{21}H_{22}O_9$	M-H	417.118	417.1191	3
Etiocholanolone glucuronide	$C_{25}H_{38}O_8$	M-H	465.2483	465.2494	2
Hesperetin 7-O-glucuronide	$C_{22}H_{22}O_{12}$	M-H	477.1039	477.10391	0

Name	formula	Adduct ions	Theoretical adduct m/z	Experimental	ΔPPM
11-beta-Hydroxyandrosterone-3-glucuronide	$C_{25}H_{38}O_9$	M-H	481.2432	481.2443	2
Tetrahydroaldosterone-3-glucuronide	$C_{27}H_{40}O_{11}$	M-H	539.2498	539.25001	0
Cortolone-3-glucuronide	$C_{27}H_{42}O_{11}$	M-H	541.2654	541.26552	0

Table. S3 Putatively annotated differential metabolites between TB and HC with statistical metrics (fold change, p-value, FDR.)

Name	t.stat	p.value	FDR
Tartaric acid	-31.348	1.11E-140	1.25E-138
N-Acetylglucosamine 6-sulfate	-31.187	1.05E-139	5.91E-138
Tiglylcarnitine	-28.625	4.17E-124	1.57E-122
Glucose	-27.381	1.61E-116	4.56E-115
Hypoxanthine	-26.655	4.35E-112	9.83E-111
Tryptophol	-25.079	1.68E-102	3.16E-101
Methylhistidine	-24.528	3.69E-99	5.96E-98
Glucose phosphate	-24.228	2.40E-97	3.39E-96
Creatinine	23.696	3.85E-94	4.84E-93
L-Glyceric acid	-23.346	4.87E-92	5.51E-91
2-Furoic acid	-23.298	9.49E-92	9.75E-91
O-Sulfotyrosine	-23.224	2.66E-91	2.50E-90
Ribothymidine	-22.32	6.68E-86	5.80E-85
Glucose.1	-21.973	7.74E-84	6.25E-83
Acetaminophen glucuronide	-21.705	2.97E-82	2.24E-81
Indoleacetyl glutamine	-20.74	1.39E-76	9.85E-76
Homovanillic acid sulfate	-20.332	3.30E-74	2.19E-73
Deoxycytidine	-20.172	2.79E-73	1.75E-72
Diphenol glucuronide	-19.367	1.21E-68	7.17E-68
N-Acetylneuraminic acid	-19.286	3.50E-68	1.98E-67
Isovaleric acid	-19.229	7.33E-68	3.94E-67
5-(3',4'-Dihydroxyphenyl)-gamma-valerolactone-4'-O-glucuronide	-19.153	2.00E-67	1.03E-66

Name	t.stat	p.value	FDR
N-Acetyl-aspartic acid	-19.029	1.02E-66	4.99E-66
Canavaninosuccinate	-18.837	1.24E-65	5.85E-65
Gluconolactone	-18.777	2.71E-65	1.23E-64
Ethyl glucuronide	-18.022	4.62E-61	2.01E-60
N-Acetyltaurine	-18.003	5.85E-61	2.45E-60
Cytidine	-17.861	3.60E-60	1.45E-59
Xanthosine	-17.735	1.79E-59	6.96E-59
6-Carboxy-5,6,7,8-tetrahydropterin	-17.145	3.10E-56	1.17E-55
1-Salicylate glucuronide	-17.049	1.02E-55	3.61E-55
Nepetaside	-17.049	1.02E-55	3.61E-55
Pyrogallol-2-O-glucuronide	-17.045	1.07E-55	3.67E-55
L-Cystine	-16.755	3.96E-54	1.32E-53
Aspartylphenylalanine	-16.623	2.04E-53	6.60E-53
Glutathione	-16.553	4.80E-53	1.51E-52
p-Cresol	-16.279	1.40E-51	4.27E-51
Pyruvic acid	-16.121	9.56E-51	2.84E-50
2,4-Dihydroxy-7,8-dimethoxy-2H-1,4-benzoxazin-3(4H)-one	-16.037	2.66E-50	7.71E-50
Fumaric acid	-15.876	1.87E-49	5.28E-49
Cyclic AMP	-15.613	4.41E-48	1.21E-47
L-Serine	-15.56	8.31E-48	2.24E-47
Ascorbic acid.1	-15.468	2.50E-47	6.56E-47
(R)-Lipoic acid	-15.367	8.31E-47	2.14E-46
Tyrosyl-Tyrosine	-15.189	6.80E-46	1.71E-45
Xanthurenic acid	-15.144	1.16E-45	2.85E-45

Name	t.stat	p.value	FDR
Glucaric acid	-15.09	2.20E-45	5.28E-45
Hydroxypropionylcarnitine	-15.056	3.26E-45	7.68E-45
Orotidine	-14.595	6.99E-43	1.61E-42
Ipratropium bromide	-14.497	2.17E-42	4.91E-42
2-Hydroxyadipic acid	-14.374	8.89E-42	1.97E-41
Pyrophosphate	-14.321	1.62E-41	3.52E-41
6-Hydroxy-5-methoxyindole glucuronide	-14.14	1.27E-40	2.72E-40
Glutamine	14.092	2.19E-40	4.58E-40
Epinephrine glucuronide	-14.053	3.41E-40	7.01E-40
Androsterone sulfate	-13.689	2.02E-38	4.08E-38
Chondroitin	-13.593	5.86E-38	1.16E-37
D-Glucuronic acid	-13.452	2.78E-37	5.41E-37
Dihydrocaffeic acid 3-O-glucuronide	-13.154	7.17E-36	1.37E-35
Sinapinic acid-O-glucuronide isomer	-13.116	1.09E-35	2.05E-35
5-Carboxy-2'-deoxyuridine	-13.085	1.52E-35	2.82E-35
Neomenthol-glucuronide	-12.845	2.01E-34	3.66E-34
Tryptophan 2-C-mannoside	-12.577	3.44E-33	6.17E-33
3-Methylcrotonylglycine	-12.276	8.05E-32	1.42E-31
2-Methylcitric acid	-12.217	1.47E-31	2.56E-31
Methylphenobarbital	-11.955	2.17E-30	3.71E-30
4-Hydroxy-5-(3',5'-dihydroxyphenyl)-valeric acid-O-glucuronide	-11.774	1.36E-29	2.29E-29
Uric acid	11.537	1.46E-28	2.43E-28
Ascorbic acid	-11.271	2.01E-27	3.28E-27
Fenoprofen glucuronide	-10.946	4.66E-26	7.53E-26

Name	t.stat	p.value	FDR
beta-1,4-Mannosyl-N-acetylglucosamine	-10.886	8.30E-26	1.32E-25
L-Histidine	10.766	2.60E-25	4.08E-25
Hesperetin 7-O-glucuronide	-10.632	9.18E-25	1.42E-24
Salicyluric acid	-10.342	1.36E-23	2.07E-23
Etiocholanolone glucuronide	-10.245	3.30E-23	4.97E-23
3-Dehydroquinic acid	-10.229	3.82E-23	5.68E-23
Malonic acid	-9.9775	3.69E-22	5.42E-22
Isoxanthopterin	-9.9517	4.65E-22	6.74E-22
Phenylacetic acid	9.8143	1.58E-21	2.25E-21
Erythronilic acid	-9.5697	1.34E-20	1.89E-20
3-hydroxyhexadecanoyl carnitine	-9.0276	1.32E-18	1.85E-18
Ribonic acid	-8.8301	6.70E-18	9.23E-18
L-Phenylalanine	-8.8061	8.14E-18	1.11E-17
Glycocholic acid	-8.5536	6.18E-17	8.31E-17
Avenic acid B	-8.4726	1.17E-16	1.56E-16
3-(3,4-Dimethoxyphenyl)-2-propenoic acid	-8.4425	1.48E-16	1.95E-16
11-beta-Hydroxyandrosterone-3-glucuronide	-8.2163	8.59E-16	1.12E-15
Phosphate	8.1939	1.02E-15	1.31E-15
Cortolone-3-glucuronide	-7.7869	2.16E-14	2.75E-14
Pyroglutamic acid	7.7415	3.01E-14	3.78E-14
3'-O-Methylepicatechin 7-O-glucuronide	-7.6624	5.36E-14	6.66E-14
Cholesterol sulfate	-7.5738	1.02E-13	1.25E-13
Methyluric acid	-7.1672	1.76E-12	2.14E-12
Tetrahydroaldosterone-3-glucuronide	-7.1086	2.63E-12	3.16E-12
Enterolactone 3'-glucuronide	-7.0339	4.36E-12	5.19E-12

Name	t.stat	p.value	FDR
Cholic acid glucuronide	-6.2928	5.17E-10	6.09E-10
Prolylhydroxyproline	-5.9756	3.47E-09	4.04E-09
Uridine diphosphate glucose	-5.8423	7.53E-09	8.68E-09
Deoxycholic acid 3-glucuronide	-5.7329	1.41E-08	1.60E-08
Aconitic acid	5.7122	1.58E-08	1.79E-08
Uridine	5.5374	4.18E-08	4.68E-08
Iduronic acid	5.1095	4.06E-07	4.49E-07
Hippuric acid	3.6021	0.00033546	0.00036803
Gluconic acid	3.5838	0.00035946	0.00039056
L-Aspartic acid	-2.5833	0.009965	0.010724
Malic acid	-2.2169	0.026912	0.028689

Table S4 Metrics of classifiers for MTB versus HC in training set. Performance metrics are averaged over 20 repeats of stratified random train/test splits (70% training set) in Orange.

Model	AUC	CA	F1	Prec	Recall	MCC
Logistic Regression	0.999	0.979	0.980	0.954	0.992	0.959
Random Forest	0.999	0.979	0.979	0.978	0.960	0.960
Neural Network	0.998	0.983	0.984	0.885	0.971	0.966
AdaBoost	0.996	0.974	0.975	1	0.992	0.948
Gradient Boosting	0.996	0.977	0.978	0.971	0.985	0.955

Table S5 Metrics of classifiers for MTB versus HC in testing set.

Model	AUC	CA	F1	Prec	Recall	MCC
Neural Network	0.997	0.984	0.985	0.985	0.985	0.968
AdaBoost	0.994	0.988	0.989	0.985	0.993	0.976
Random Forest	0.99	0.944	0.945	1	0.896	0.894
Logistic Regression	0.988	0.924	0.934	0.877	1	0.855
Gradient Boosting	0.986	0.932	0.939	0.915	0.963	0.864

Table S6. Evaluation results for target JH (comparison of classifiers). Performance metrics are averaged over 50 repeats of stratified random train/test splits (70% training set) in Orange.

Model	AUC	CA	F1	Prec	Recall	MCC
Random Forest	0.999	0.97	0.971	1	0.943	0.942
Neural Network	0.999	0.982	0.983	0.979	0.988	0.965
Logistic Regression	0.999	0.976	0.977	0.965	0.989	0.952
Gradient Boosting	0.996	0.972	0.973	0.972	0.975	0.944
AdaBoost	0.995	0.972	0.973	0.971	0.975	0.943

Table S7. Evaluation results for target JH (comparison of classifiers). Performance metrics are averaged over 100 repeats of stratified random train/test splits (70% training set) in Orange.

Model	AUC	CA	F1	Prec	Recall	MCC
Logistic Regression	0.999	0.976	0.978	0.965	0.99	0.953
Random Forest	0.999	0.972	0.972	0.997	0.948	0.945
Neural Network	0.998	0.979	0.979	0.976	0.983	0.957
AdaBoost	0.996	0.974	0.975	0.972	0.977	0.947
Gradient Boosting	0.995	0.97	0.971	0.968	0.974	0.939