Supplementary information

Bacterial Identification by Metabolite-Level Interpretable Surface-Enhanced Raman Spectroscopy

Haoran Chen[1#], Ruike Zhao[2#], Xinyuan Bi[1], Nan Shen[2], Xi Mo[2], Yue Tao[2], Zhou Chen[1,3]* and Jian Ye[1,3,4]**

[1]Sixth People's Hospital, School of Medicine & School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200030, P.R. China

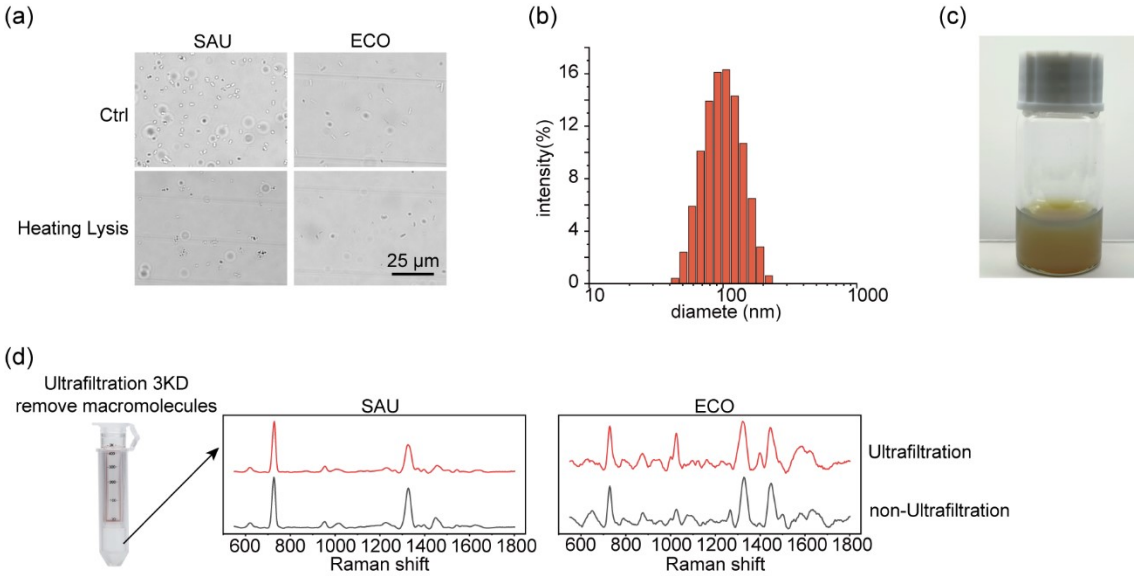
[2]Pediatric Translational Medicine Institute, Shanghai Children's Medical Center, School of Medicine, Shanghai Jiao Tong University, Shanghai 200127, P.R. China

[3]Institute of Medical Robotics, Shanghai Jiao Tong University, Shanghai 200240, P.R. China

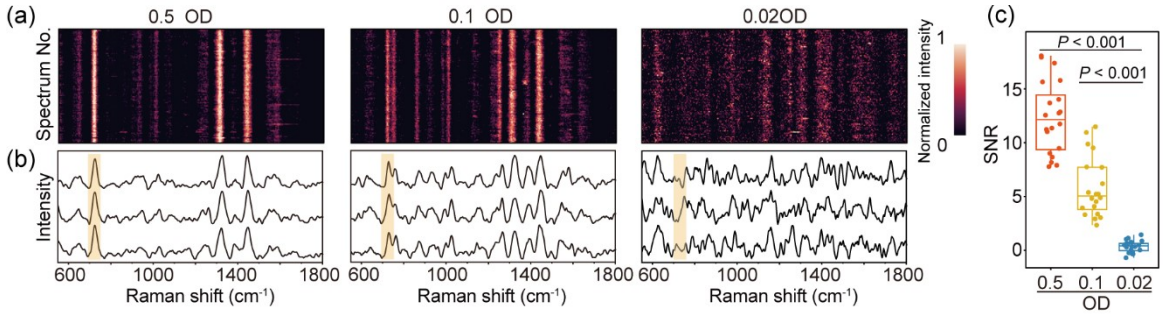
[4]Shanghai Key Laboratory of Gynecologic Oncology, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200127, P.R. China

[#]These authors contributed equally to this work

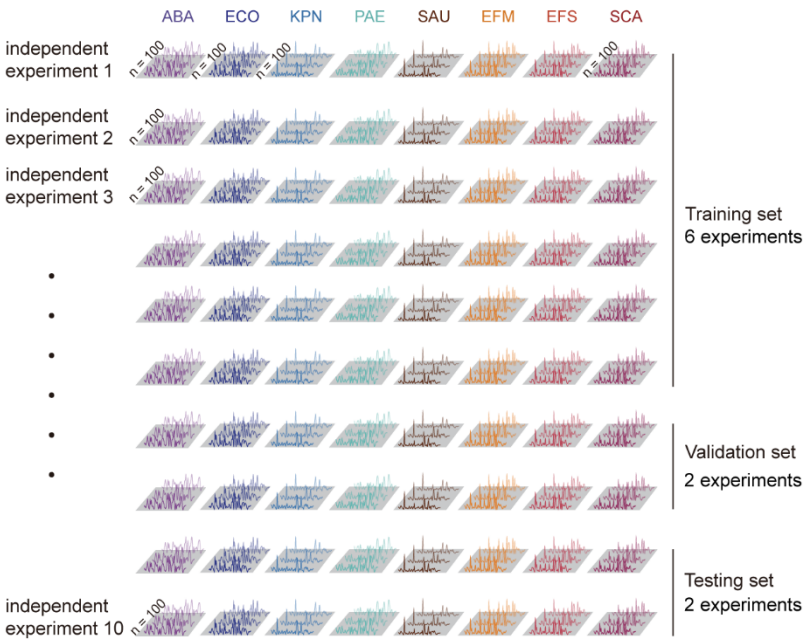
*To whom correspondence should be addressed. E-mail: yejian78@sjtu.edu.cn (J.Y.); chenzhou96@sjtu.edu.cn (Z.C.); taoyue@scmc.com.cn (Y.T.)



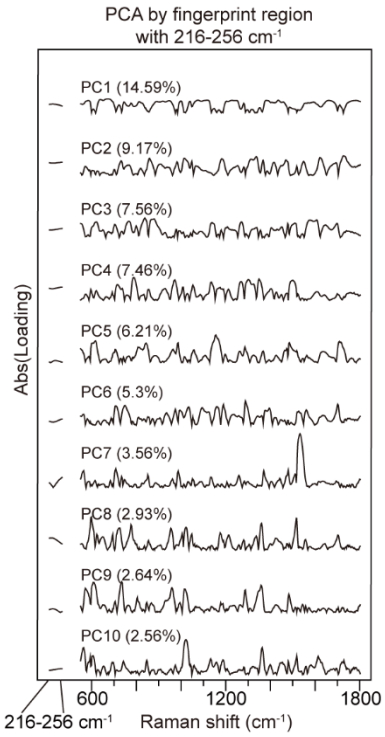
Supplementary Figure 1. (a) Images of bacteria lysed by heating. **(b)** Histogram of the hydrodynamic diameters via DLS. **(c)** The citrate-reduced Ag NPs. **(d)** Comparison of the ultrafiltration and non-ultrafiltration samples (*S. aureus* and *E. coli* at 0.5 OD). In this section, macromolecules, such as proteins, were removed through ultrafiltration.



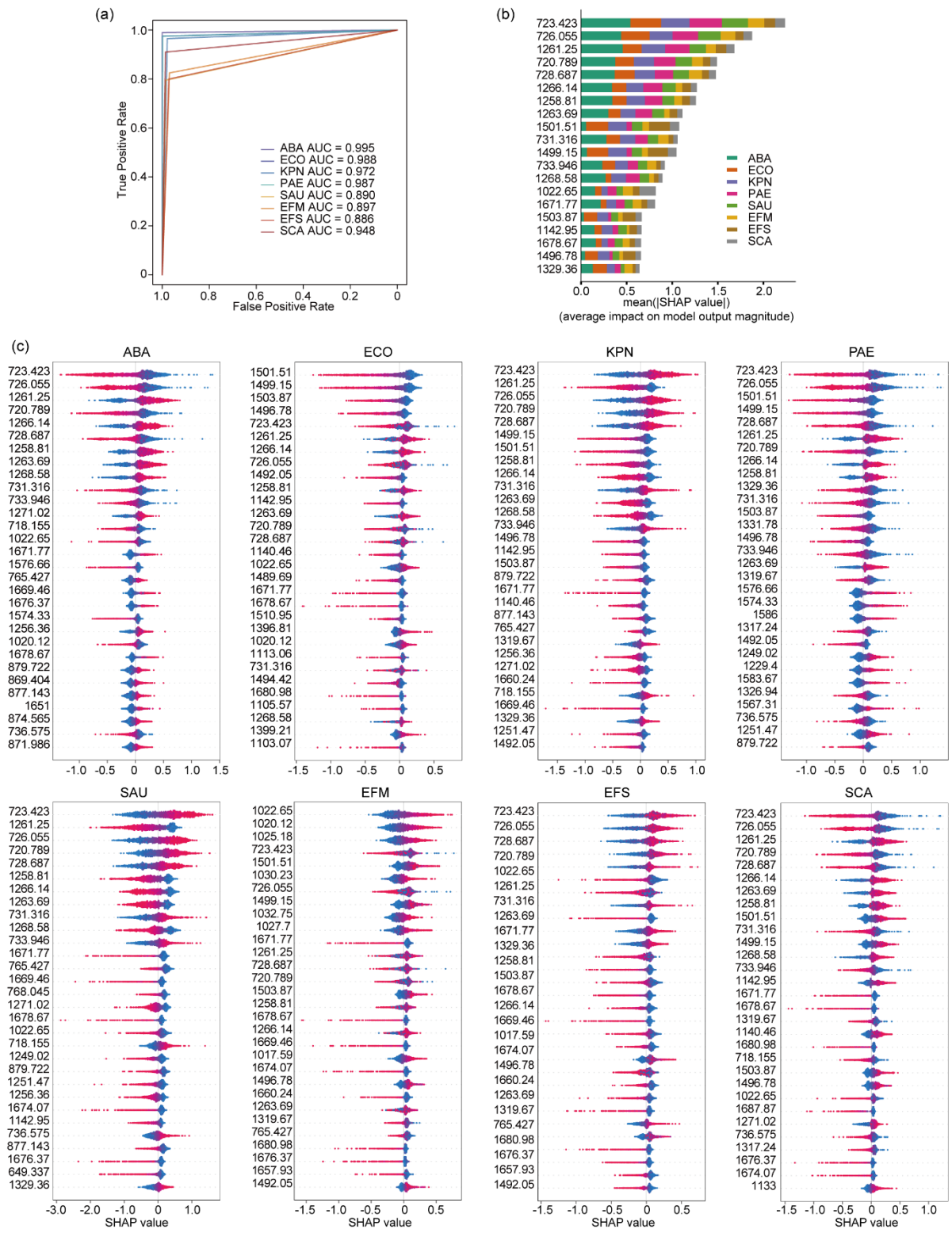
Supplementary Figure 2. (a) Heatmaps of SERSomes from *E. coli* at various concentrations. All spectra were normalized. **(b)** Typical single SERS spectra of *E. coli* at different concentrations. The yellow region encompasses the Raman shift range of 700-749 cm^{-1}. **(c)** The signal-to-noise ratio (SNR) of *E. coli* at different concentrations computed with the peak indicated by yellow stripes in panel B (n = 20). *P* values were calculated using Wilcoxon rank-sum test.



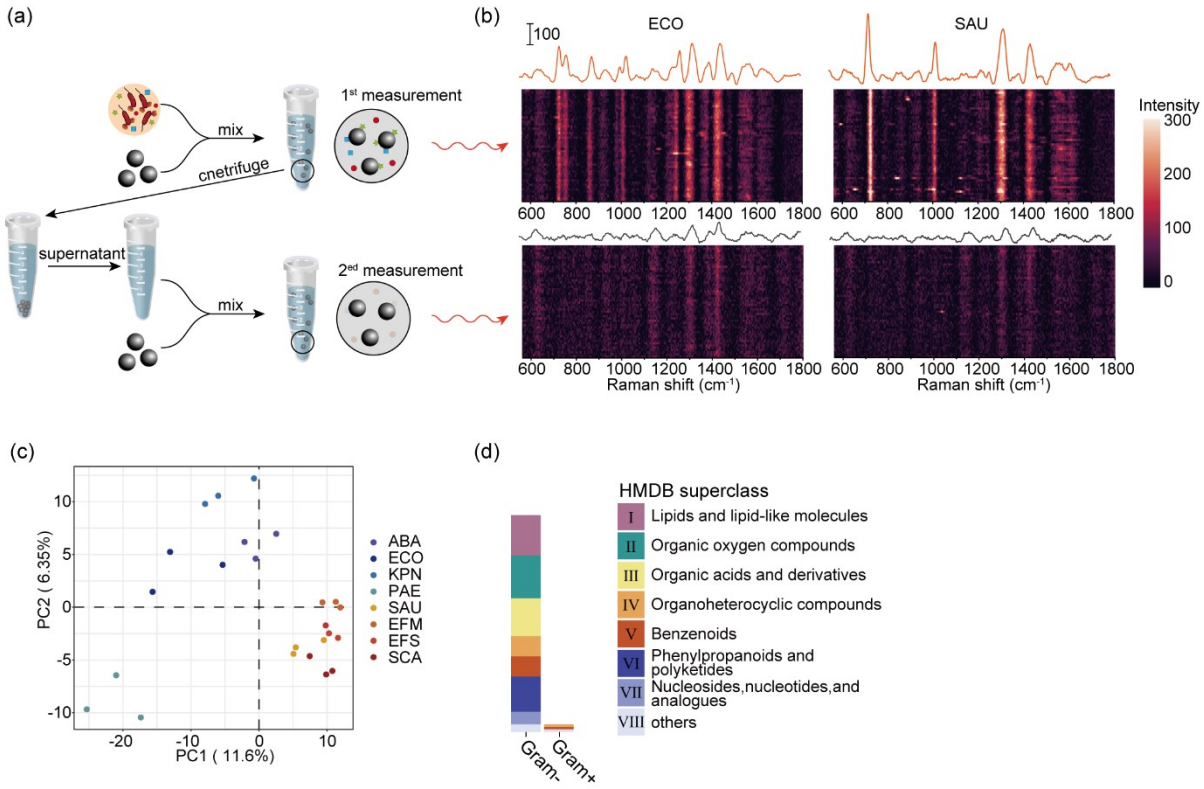
Supplementary Figure 3. Schematic diagram of the data acquisition and separation design. Through ten independent experiments at different time, we obtained the bacterial spectral dataset. In the stage of establishing the diagnostic model, we randomly divided the dataset by independent experiment, with the training set : validation set : testing set = 60% : 20% : 20%.



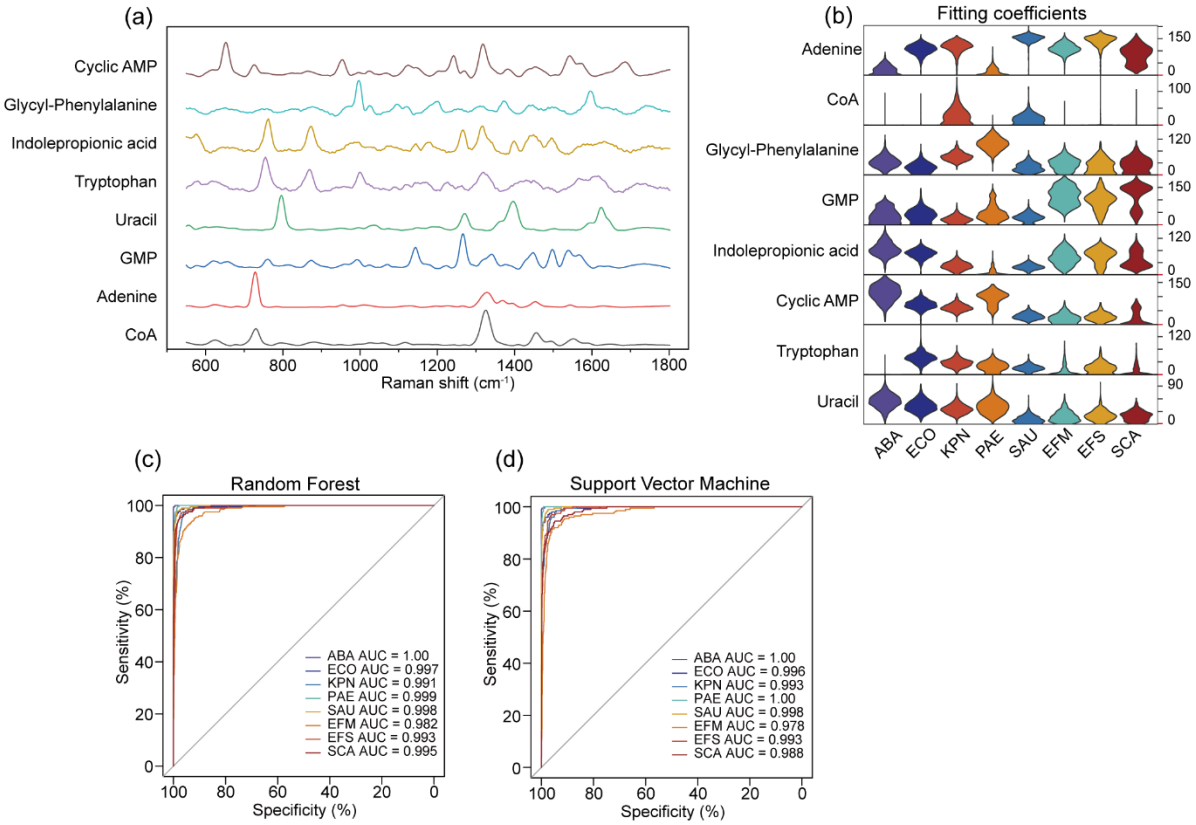
Supplementary Figure 4. The absolute value of PC loadings corresponding to the 216-256 cm^{-1} region and fingerprint region.



Supplementary Figure 5. (a) ROC curve for the eight bacterial species under the CNN diagnostic model. **(b), (c)** Top contributing features for classification according to SHAP values.



Supplementary Figure 6. (a) Schematic diagram of the experiment to verify the adsorption effect of nanoparticles for metabolites. The first adsorption: the lysate was mixed with Ag NPs and measured by SERS. The second adsorption: The mixture was centrifuged to obtain the supernatant which was mixed with Ag NPs again and measured by SERS. **(b)** The Raman spectra of ECO and SAU from the 1st and 2nd SERS measurements. **(c)** PCA plot based on the bacteria features from LDI-MS. **(d)** Superclasses of differential metabolites between Gram+ and Gram- bacteria.



Supplementary Figure 7. (a) Metabolite panel for fitting. **(b)** Coefficient distribution of each metabolite in the metabolite panel obtained by spectral decomposition (n = 1000 spectra per bacterial species). **(c)** ROC curve based on a random forest model. **(d)** ROC curve based on a SVM model.

Supplementary Table 1. Bacterial species and abbreviations.

	Abbreviate1	Abbreviate2	
Acinetobacter baumannii	*A. baumannii*	ABA	Gram-
Escherichia coli	*E. coli*	ECO	Gram-
Klebsiella pneumoniae	*K. pneumoniae*	KPN	Gram-
Pseudomonas aeruginosa	*P. aeruginosa*	PAE	Gram-
Staphylococcus aureus	*S. aureus*	SAU	Gram+
Enterococcus faecium	*E. faecium*	EFM	Gram+
Enterococcus faecalis	*E. faecalis*	EFS	Gram+
Staphylococcus capitis	*S. capitis*	SCA	Gram+

Supplementary Table 2. Metabolites identified from LDI-MS and screened for spectral fitting.

No.	molecule	HMDB ID	Adduct.	*m/z*	ABA	ECO	PAE	KPN	SAU	EFM	EFS	SCA	Test for SERS spectra
1	Adenine	HMDB0000034	$[M+K]^+$	174.016	+	+	+			+	+		Adenine
2	Cyclic AMP	HMDB0000058	$[M+H]^+$/ $[M+Na]^+$	330.060/ 352.043	+	+	+	+	+				Cyclic AMP
3	Uridine diphosphate-N-acetylglucosamine	HMDB0000290	$[M+H]^+$	608.089		+	+		+		+	+	Uridine
4	Glycyl-Phenylalanine	HMDB0028848	$[M+H]^+$/ $[M+Na]^+$	223.108/ 245.090	+	+	+	+	+				Glycyl-Phenylalanine
5	Alanyltryptophan	HMDB0013209	$[M+H]^+$	276.134	+	+	+	+		+	+	+	Tryptophan
6	Tryptophyl-Tryptophan	HMDB0029094	$[M+Na]^+$	413.155		+	+			+		+	
7	5-Hydroxy-L-tryptophan	HMDB0000472	$[M+H]^+$	221.092	+	+	+						
8	Palmityl-CoA	HMDB0001338	$[M+H]^+$	1006.360	+	+	+	+	+	+	+	+	CoA
9	Butyryl-CoA/ Isobutyryl-CoA	HMDB0001243	$[M+H]^+$	838.166	+		+	+	+	+	+	+	
10	Indoleacrylic acid	HMDB0000734	$[M+Na]^+$/ $[M+K]^+$	210.052/ 226.026	+	+	+	+	+				Indolepyruvate
11	Indolepyruvate	HMDB0060484	$[M+K]^+$	242.019	+	+	+	+	+				
12	5-Hydroxyindoleacetic acid	HMDB0000763	$[M+H]^+$	192.065	+	+	+		+				
13	Guanosine diphosphate mannose	HMDB0001163	$[M+H]^+$	606.088	+	+						+	GMP
14	Guanosine tetraphosphate adenosine	HMDB0001454	$[M+K]^+$	891.009			+	+		+			
15	Guanosine	HMDB0000133	$[M+H]^+$	284.101		+	+						
16	Diguanosine tetraphosphate	HMDB0001340	$[M+K]^+$	906.999		+	+						

"+": The corresponding metabolite was detected above the signal-to-noise ratio threshold in the mass spectrum of the bacterium.