

## Supplementary information

Yanyan Li<sup>1+</sup>, Hua Zhang<sup>1+</sup>, Jingjing Jiang<sup>2</sup>, Lin Zhao<sup>2\*</sup>, Yunbing Wang<sup>1\*</sup>

<sup>1</sup>National Engineering Research Center for Biomaterials, Sichuan University,  
Chengdu, Sichuan, 610065, China

<sup>2</sup>Department of Endocrinology and Metabolism, Fudan Institute of Metabolic  
Diseases, Zhongshan Hospital, Fudan University, Shanghai, 200032, China.

\* To whom correspondence should be addressed.

29 Wang Jiang Road, Chengdu, China, 610065

E-mail: [zhao.lin@zs-hospital.sh.cn](mailto:zhao.lin@zs-hospital.sh.cn), [yunbing.wang@scu.edu.cn](mailto:yunbing.wang@scu.edu.cn)

## **Table of Contents**

Table S1. Reaction conditions for adjusting SiO<sub>2</sub>@Au shell thickness.

Figure S1. Ultraviolet-visible absorption spectra of Au colloids.

Figure S2. Construction of SiO<sub>2</sub> nanoparticles.

Figure S3. Construction of Au nanoparticles.

Figure S4. The necessity of elemental composition of SiO<sub>2</sub>@Au nano shells in LDI MS.

Figure S5. LDI MS spectra of small molecules.

Overall workflow of data analysis.

Figure S6. The detection performance of serum small molecular metabolites by Au NPs and SiO<sub>2</sub>@Au-2.

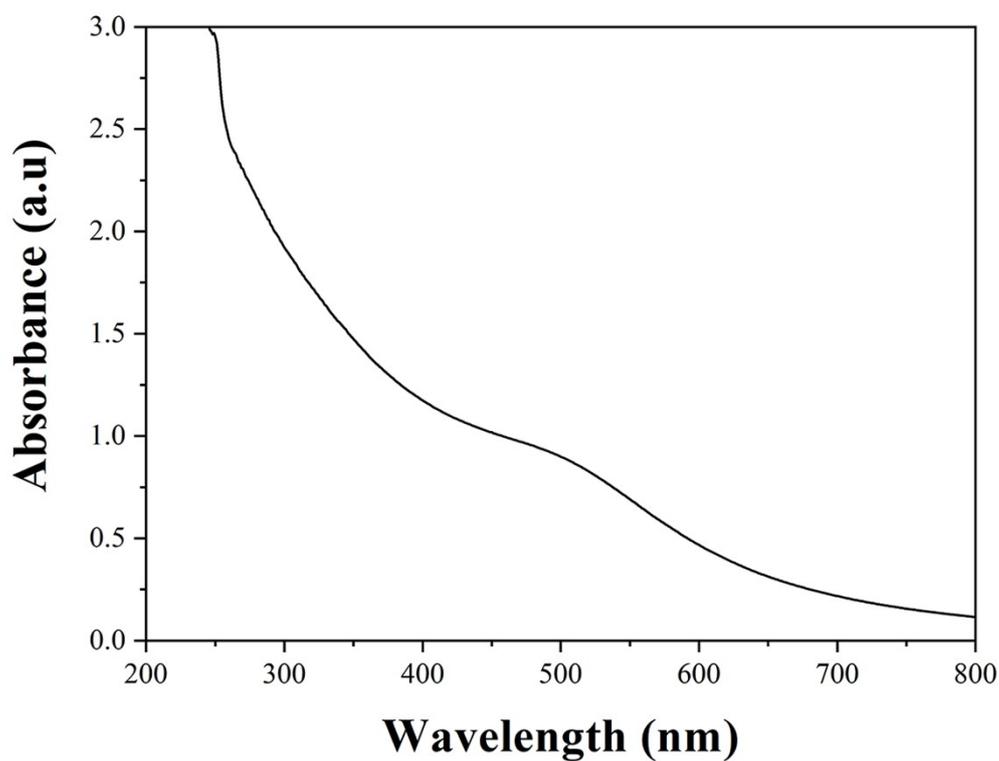
Figure S7. Overall workflow of data analysis.

Machine learning classifier.

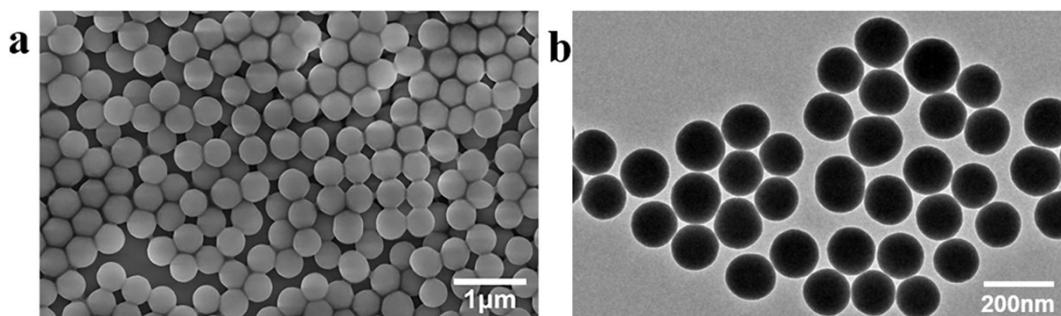
Table S2. 15 m/z from the lasso regression.

|                        | K <sub>2</sub> CO <sub>3</sub><br>(mM) | HAuCl <sub>4</sub> 1%<br>(mL) | Formaldehyde (μL) | SiO <sub>2</sub> @Au-seed<br>(μL) |
|------------------------|--|-------------------------------|-------------------|-----------------------------------|
| SiO <sub>2</sub> @Au-1 | 1.8                                    | 0.16                          | 50                | 20                                |
| SiO <sub>2</sub> @Au-2 | 3.6                                    | 0.32                          | 100               | 20                                |
| SiO <sub>2</sub> @Au-3 | 7.2                                    | 0.64                          | 200               | 20                                |
| SiO <sub>2</sub> @Au-4 | 18                                     | 1.6                           | 500               | 20                                |

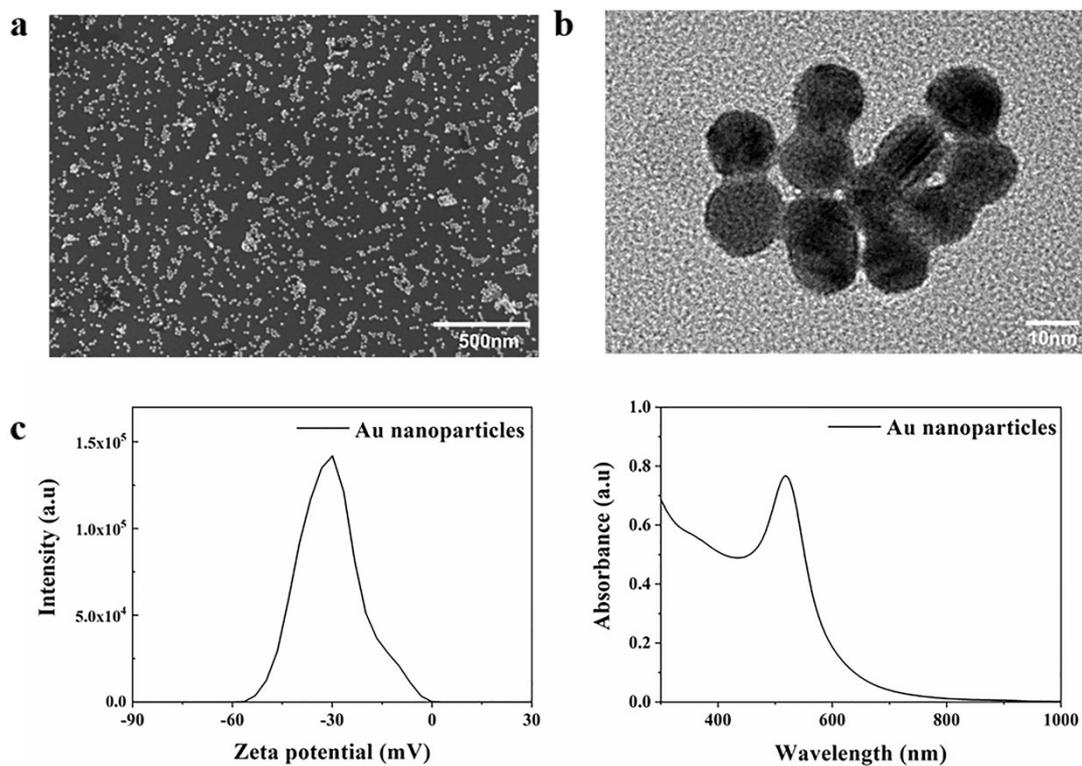
**Table S1.** Reaction conditions for adjusting SiO<sub>2</sub>@Au shell thickness.



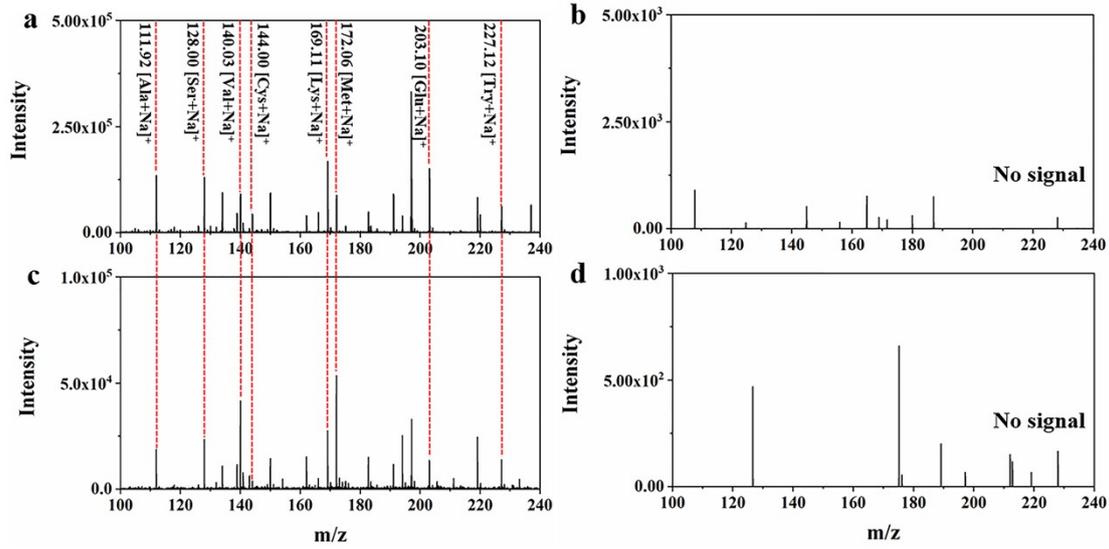
**Figure S1.** Ultraviolet-visible absorption spectra of Au colloids.



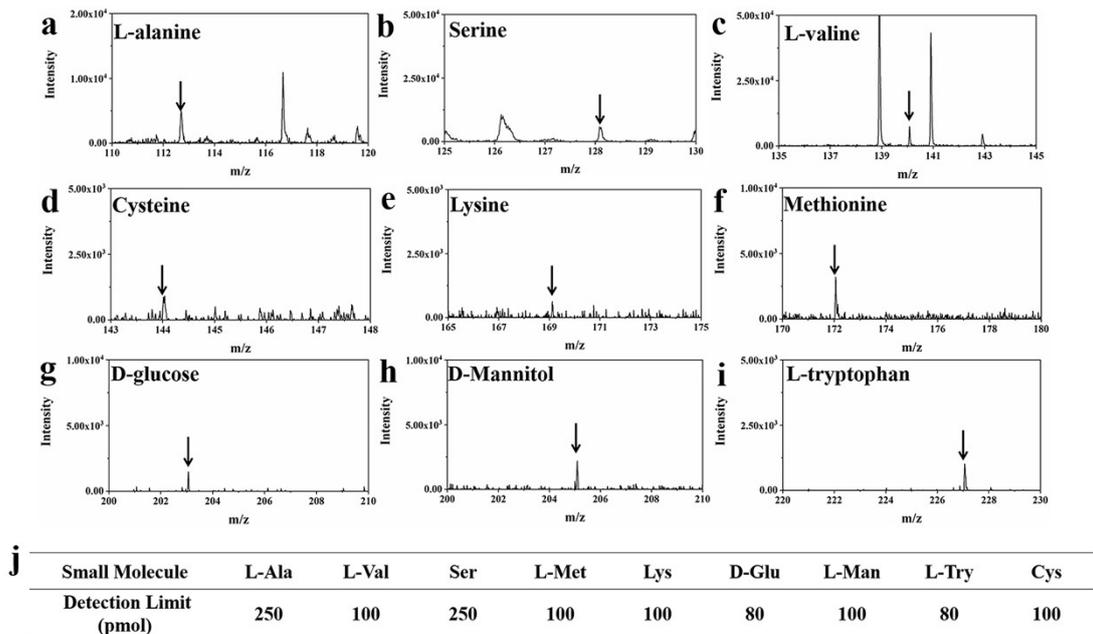
**Figure S2.** Construction of SiO<sub>2</sub> nanoparticles. a) Scanning electron microscopy (SEM) images. b) transmission electron microscopy (TEM) images.



**Figure S3.** Construction of Au nanoparticles. a) Scanning electron microscopy (SEM) images. b) transmission electron microscopy (TEM) images. c) ultraviolet-visible absorption spectra and d) Zeta potential of Au nanoparticles.

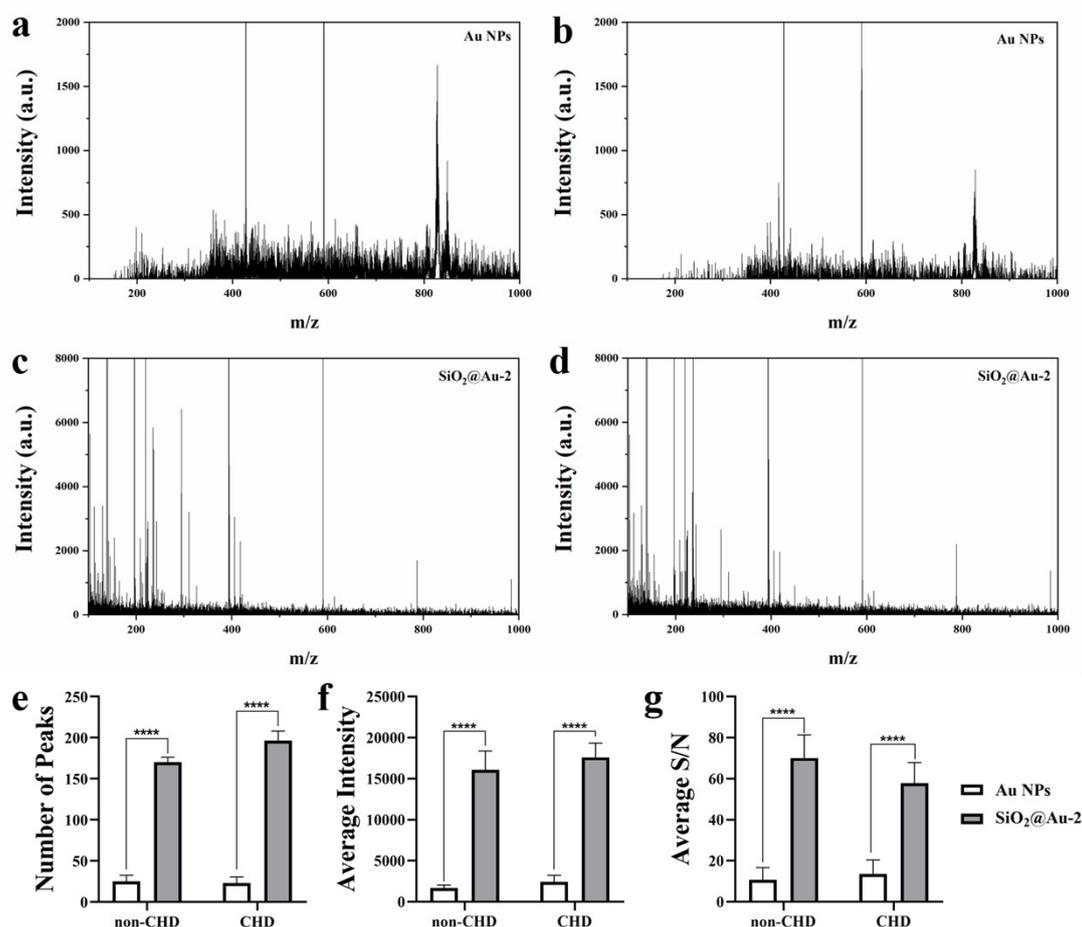


**Figure S4.** The necessity of elemental composition of SiO<sub>2</sub>@Au nano shells in LDI MS. Typical mass spectra of metabolites in the positive ion mode using a) SiO<sub>2</sub>@Au, b) bare silica core, c) Au nanoparticles and d)  $\alpha$ -cyano-4-hydroxycinnamic acid ( $\alpha$ -CHCA) as matrices of 10mM mixture (including L-alanine, serine, L-valine, cysteine, lysine, methionine, D-glucose and L-tryptophan in 0.9% NaCl).



**Figure S5.** LDI MS spectra of small molecules. Mass spectrum of Na<sup>+</sup> adducted a)

L-alanine at  $m/z$  of 112.11 for  $[M+Na]^+$ , b) serine at  $m/z$  of 128.10 for  $[M+Na]^+$ , c) L-valine at  $m/z$  of 140.17 for  $[M+Na]^+$ , d) cysteine at  $m/z$  of 144.19 for  $[M+Na]^+$ , e) lysine at  $m/z$  of 169.14 for  $[M+Na]^+$ , f) methionine at  $m/z$  of 172.05 for  $[M+Na]^+$ , g) D-glucose at  $m/z$  of 203.12 for  $[M+Na]^+$ , h) D-Mannitol at  $m/z$  of 205.08 for  $[M+Na]^+$  and i) L-tryptophan at  $m/z$  of 227.15 for  $[M+Na]^+$  in 0.9% NaCl and j) the detect limits using Au nanoparticles as matrix in positive ion mode.



**Figure S6.** The detection performance of serum small molecular metabolites by Au NPs and SiO<sub>2</sub>@Au-2. a) Typical mass spectra of non-CHD serum by AuNPs. b) Typical mass spectra of CHD serum by AuNPs. c) Typical mass spectra of non-CHD serum by SiO<sub>2</sub>@Au-2. d) Typical mass spectra of CHD serum by SiO<sub>2</sub>@Au-2. e) The quantitative

analysis of the number of peaks. e) The quantitative analysis of the average intensity.  
f) The quantitative analysis of the signal-to-noise ratio (S/N).

### **Overall workflow of data analysis**

The overall workflow was composed of data preprocessing, model training and model validation. A validation cohort from CHD mass spectrometry data were used to test gold-nanoshell assisted laser desorption/ionization mass spectrometry to diagnosing CHD. After the threshold value ( $\theta$ ) was set for different specificity, SVM model was built to compute the average validation error.

#### *Data Preprocessing*

In order to obtain reliable results, the Mass Spectrometry dataset T was randomly divided into two partitions. A part of 80% was used as a training cohort; the other 20% was left as validation cohort to verify the supervised machine learning model. The training cohort and the validation cohort were prepared separately but using the same strategy. First, like other analytical platforms, the raw data were preprocessed with several data preprocessing steps such as baseline correction and noise reduction. This was followed by normalization step, which could ensure reproducible comparisons. Finally, a calibration procedure was employed.

#### *Training Model*

The supervised decision-making model proposed for CHD discrimination had following steps:

1. Five times five-fold cross-validation. We divided the training cohort into two subsets, one was T-training cohort with 80% of training cohort, the remaining 20% was a T-

validation data set. This process would be repeated for 5 times to get the average training error.

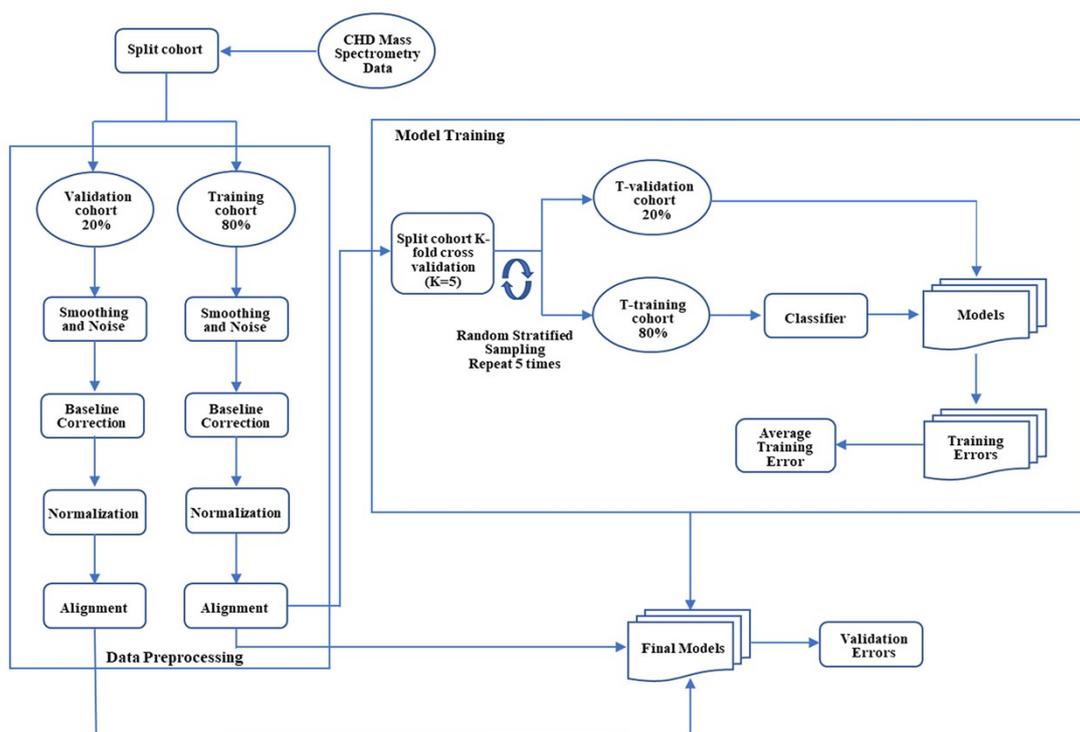
2. Model training. Support Vector Machine (SVM) was applied as the classifier, data from T-training cohort was considered as input, then the trained SVM model was tested by the T-validation cohort and we could get

the training error. After we developed all the 5 times 5-fold cross-validation, the average train error was obtained to train the hyper-parameters. In the end, the training model was built by the whole training cohort with the hyper-parameters.

### Validation Model

1. Internal validation. When a threshold value was selected, the internal validation dataset was used to compute the internal validation error.

2. External validation. Another single-blinded external validation cohort was tested to obtain the final external validation error.



**Figure S7.** Overall workflow of data analysis

**Machine learning classifier**

*Support-vector machines (SVM)*

SVM are supervised learning models using learning algorithms that examine data for classification and regression analysis in machine learning. An SVM training algorithm creates a model that assigns new examples to one of two categories, making it a non-probabilistic binary linear classifier, given a collection of training examples, each marked as belonging to one of two categories. SVM translates training examples to points in space in order to widen the distance between the two categories as much as possible. New instances are then mapped into the same space and classified according to which side of the gap they land on.

*Random forest (RF)*

RF is a machine learning technique that's used to solve regression and classification problems. It utilizes ensemble learning, which is a technique that combines many classifiers to provide solutions to complex problems. A RF algorithm consists of many decision trees. The 'forest' generated by the RF algorithm is trained through bagging or bootstrap aggregating. Bagging is an ensemble meta-algorithm that improves the accuracy of machine learning algorithms. The RF algorithm establishes the outcome based on the predictions of the decision trees. It predicts by taking the average or mean of the output from various trees. Increasing the number of trees increases the precision of the outcome. A RF eradicates the limitations of a decision

tree algorithm. It reduces the overfitting of datasets and increases precision. It generates predictions without requiring many configurations in packages (like scikit-learn).

### *Lasso regression (LR)*

LR is a regularization technique. It is used over regression methods for a more accurate prediction. This model uses shrinkage. Shrinkage is where data values are shrunk towards a central point as the mean. The lasso procedure encourages simple, sparse models. This particular type of regression is well-suited for models showing high levels of multicollinearity or when you want to automate certain parts of model selection, like variable selection/parameter elimination. LR uses L1 regularization technique. It is used when we have more number of features because it automatically performs feature selection. If a regression model uses the L1 Regularization technique, then it is called LR. If it used the L2 regularization technique, it's called Ridge Regression. L1 regularization adds a penalty that is equal to the absolute value of the magnitude of the coefficient. This regularization type can result in sparse models with few coefficients. Some coefficients might become zero and get eliminated from the model. Larger penalties result in coefficient values that are closer to zero (ideal for producing simpler models). On the other hand, L2 regularization does not result in any elimination of sparse models or coefficients. Thus, Lasso Regression is easier to interpret as compared to the Ridge.

**Table S2.** 15 m/z from the lasso regression.

| Number | m/z      | coef      | Metabolites   | Origin of the signal                                       |
|--------|----------|-----------|---|--|
| 1      | 466.3429 | 42.439316 | Glycocholic acid <sup>[1]</sup> /Ethylchenodeoxycholic acid <sup>[1]</sup>                      | [M+H] <sup>+</sup> /[M+Na] <sup>+</sup>                    |
| 2      | 125.4685 | 37.045935 | 3-Methyl-2-butene-1-thiol <sup>[1]</sup>  | [M+Na] <sup>+</sup>  |
| 3      | 161.6978 | 25.165203 | Carnitine <sup>[1]</sup> /Indole-3-ethanol <sup>[1]</sup>                                       | [M+H] <sup>+</sup> /[M+H] <sup>+</sup>                     |
| 4      | 145.0451 | 21.378481 | octanone <sup>[1]</sup> /Nicotinamide <sup>[2]</sup>  | [M+H] <sup>+</sup> /[M+Na] <sup>+</sup>                    |
| 5      | 113.9622 | 18.254777 | Uracil <sup>[2]</sup> /Trimethylamine N-oxide <sup>[1]</sup> /Potassium chloride <sup>[3]</sup> | [M+H] <sup>+</sup> /[M+K] <sup>+</sup> /[M+K] <sup>+</sup> |
| 6      | 132.1074 | 13.866250 | Creatine <sup>[1]</sup> /leucine <sup>[3]</sup>   | [M+H] <sup>+</sup> /[M+H] <sup>+</sup>                     |
| 7      | 146.4392 | 13.283424 | N-Methylisoleucine <sup>[1]</sup>   | [M+H] <sup>+</sup>   |
| 8      | 620.4997 | 9.828630  | N-(hexadecanoyl)-sphinganine-1-phosphate <sup>[1]</sup>   | [M+H] <sup>+</sup>   |
| 9      | 103.7001 | 7.558457  | $\gamma$ -Aminobutyric acid <sup>[1]</sup>  | [M+H] <sup>+</sup>   |
| 10     | 615.8975 | 5.627639  | NA  | NA   |
| 11     | 118.4850 | 4.291872  | Valine <sup>[1]</sup>   | [M+H] <sup>+</sup>   |
| 12     | 201.5301 | 2.494084  | Huppuric acid <sup>[1]</sup>  | [M+K] <sup>+</sup>   |
| 13     | 161.0779 | 1.516102  | 2-Methyl-3-(methylthio)furan <sup>[3]</sup>   | [M+H] <sup>+</sup>   |
| 14     | 378.2284 | 1.260060  | 5-amino-1-(5-phospho-D-ribose)imidazole-4-carboxylate <sup>[3]</sup>                            | [M+K] <sup>+</sup>   |
| 15     | 193.0717 | 1.156917  | 2-Nonynoic acid <sup>[1]</sup> /Isocitrate <sup>[2]</sup>                                       | [M+H] <sup>+</sup> /[M+K] <sup>+</sup>                     |

**Reference:**

[1] Y. Fan, Y. Li, Y. Chen, Y.-J. Zhao, L.-W. Liu, J. Li, S.-L. Wang, R.N. Alolga, Y. Yin, X.-M. Wang, et al., Comprehensive Metabolomic Characterization of Coronary Artery Diseases, *J. Am. Coll. Cardiol.* 68(12) (2016) 1281-1293.

[2] H. Liu, X. Chen, X. Hu, H. Niu, R. Tian, H. Wang, H. Pang, L. Jiang, B. Qiu, X. Chen, et al., Alterations in the gut microbiome and metabolism with coronary artery disease severity, *Microbiome* 7(1) (2019) 68.

[3] N.P. Paynter, R. Balasubramanian, F. Giulianini, D.D. Wang, L.F. Tinker, S. Gopal, A.A. Deik, K. Bullock, K.A. Pierce, J. Scott, et al., Metabolic Predictors of Incident Coronary Heart Disease in Women, *Circulation* 137(8) (2018) 841-853.