Supplementary Information

**Zwitterionic mesoporous engineering aids peptide-dependent patterns profiling for determination of different liver diseases**

Zixing Xu,[a] Wantong Zhang,[a] Chunhui Deng[a,b]*, and Nianrong Sun[a,c]*

[a] *Department of Gastroenterology and Hepatology, Zhongshan Hospital, and Department of Chemistry, Fudan University, Shanghai, 200433, China.*

[b] *School of Chemistry and Chemical Engineering, Nanchang University, Nanchang 330031, China.*

[c] *Shanghai Institute of Liver Diseases, Shanghai 200032, China*

*Corresponding author. Email: sunnianrong@fudan.edu.cn; chdeng@fudan.edu.cn.

This file includes:

Materials and Methods

(Chemical and materials, Synthesis of ZMM, Characterization, Sample preparation, MALDI-TOF MS analysis, Statistical analysis, HPLC-MS/MS analysis and database searching, Matching criteria of peptide markers)

Figures S1 to S5

Tables S1 to S4

References

## Materials and Methods

### Chemical and materials

Tetraethyl orthosilicate (TEOS), cetrimonium bromide (CTAB), trifluoroacetic acid (TFA), sodium hydroxide (NaOH), ethanol and ethylene glycol were purchased from Shanghai Chemical Co., Ltd. Iminodiacetic acid (IDA), acetonitrile (ACN) and ammonium bicarbonate ($NH_4HCO_3$) were obtained from Sinopharm Chemical Reagent Co., Ltd. Trypsin, and 2,5-Dihydroxy-benzoic acid (DHB) were bought from Merck (Darmstadt, Germany). Ultrapure water (18.2 M$\Omega$ cm) purified by Milli-Q system (Millipore, Bedford, MA) was employed for all aqueous solutions. Serum of liver cancer patients, liver cirrhosis patients and healthy people were offered by Shanghai Zhongshan Hospital. All other chemical reagents were at least analytical grade and purchased from commercial sources.

### Synthesis of ZMM

As for the synthesis of ZMM, GLYMO-IDA solution was first prepared, 2.5 g IDA was dissolved in 50 mL of 2 M $Na_2CO_3$, and pH was tuned to 10 with 10 M NaOH. The solution was stirred in the condition of ice bath for 10 min, and then 1.6 mL of GLYMO was added into the solution drop by drop. After stirring at 0 °C for 30 min, the temperature was raised to 65 °C with stirring for 6 h. A total of 1.6 mL of GLYMO was added into the mixture again after being cooled down to 0 °C. Repeat the above steps and the obtained product was adjusted to pH 6 with concentrated HCl after naturally cooling, and the terminate product was denoted as GLYMO-IDA solution.

Then, $Fe_3O_4$ was obtained according to previously reported procedure, 50 mg $Fe_3O_4$ was dispersed in the mixed solution containing 500 mg CTAB and 50 mL of deionized water for sonication with 30 min. Afterwards, 50 mL of 10 mM NaOH and 400 mL deionized water were added to the above mixture, then the mixture was sonicated until getting uniform. The resultant mixture was stirred under 60 °C for 30 min, followed by 2.5 mL of mixed solvent (TEOS/ethanol=1:4, v/v) being introduced under continuous stirring. The reaction was proceeded under constant stirring at 60 °C overnight. The nanocomposites were separated by external magnetic field, followed by washing with distilled water and ethanol several times, respectively. Finally, the obtained nanocomposites were immersed into 100 mL acetone and mildly stirred under 60 °C overnight for reflux to remove CTAB. After drying in vacuum oven, the above magnetic silica was added into the GLYMO-IDA solution, and then the solution was treated ultrasonically for 10 min. The suspension was then heated to 95 °C and stirred for 2 h to obtain the final product (ZMM).

**Characterization**

Transmission electron microscope (TEM) images were recorded on a JEOL 2011 microscope (Japan) operated at 200 kV. Field emission scanning electron microscope (FeSEM) image was taken with a Hitachi S-4800 microscope (Japan). Nitrogen adsorption-desorption isotherms were conducted on a Micromeritcs Tristar 3000 analyzer (USA). The specific surface area was measured by the BET (Brunauer-Emmett-Teller) method, the pore size distribution curves were determined by BJH (Barrett-Joyner-Halenda) method.

**Sample preparation**

For 209 human serum samples, 2 μL serum of each sample was mixed with 100 μg ZMM in 200 μL loading buffer (ACN/$H_2O$/TFA=90/9/1, v/v/v), and incubated at 37 °C for 30 min. The supernatant was discarded with the aid of a magnet, the remaining nanocomposites were washed with loading buffer for 3 times. Afterwards, the nanocomposites were eluted with 10 μL eluting buffer (ACN/$H_2O$/TFA=30/69.9/0.1, v/v/v), the eluent containing serum peptides was collected after vibrating at 37 °C for 30 min. 1 μL eluent was mixed with 1 μL matrix solution (20 mg mL$^{-1}$ DHB, ACN/$H_2O$/TFA=50/49.9/0.1, v/v/v), the mixed solution was deposited onto the stainless-steel target plate for further MS analysis. Each sample underwent three biological and technical repetitions. For biological repetitions, three independent samples were collected from each serum and subjected to MS analysis. For technical repetitions, each independent sample was performed using MS analysis three times.

**MALDI-TOF MS analysis**

MALDI-TOF MS analysis was carried out using an UltrafleXtreme MALDI-TOF/TOF MS (Bruker Daltonic, Germany) equipped with a pulsed 355 nm smartbeam Nd:YAG laser in the frequency of 2000 Hz, under reflector positive mode with 90% laser intensity, accumulation of 1000 laser shots and 170 ns delayed extraction time. MALDI MS data were obtained from Flexcontrol 3.4 and processed in Flexanalysis 3.4.

**Statistical analysis**

The MALDI-TOF raw data underwent a series of processing steps, including peak extraction, alignment and average of the three technical repetition MS data using the R package MALDIquant[1] to build a peak intensity table, the table contains all the peaks and corresponding intensities of each sample. Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA) and filtration of the key *m/z* features were implemented through SIMCA-P 14.1 software (MKS Umetrics, Sweden). ROC analysis was powered with the software Orange.[2] The key peptide features were searched using Uniprot-SwissProt database (Taxonomy: Homo Sapiens, 20386 entries). volcano plots were plotted using Origin 2023.

**HPLC-MS/MS analysis and database searching**

The serum samples from two liver cirrhosis patients, two liver cancer patients and two healthy controls were prepared following the aforementioned procedure, then lyophilized and resuspended with 10 μL of buffer A (0.1% formic acid aqueous solution). The experiments were conducted using an EASY-nLC 1000 system (Thermo Fisher Scientific, Waltham, MA) coupled to an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific, San Jose, CA) equipped with an online nano-electrospray ion source. 3 μL peptide sample was injected into analytical column (Acclaim PepMap C18, 75 μm x 25 cm) and separated using a linear gradient, starting from 5% solvent A (0.1% formic acid aqueous solution) and reaching 30% buffer B (ACN with 0.1% formic acid) over 50 minutes. The column was then re-equilibrated at initial conditions for 10 minutes, while maintaining a constant flow rate of 300 nL/min. An electrospray voltage of 2.3 kV versus the mass spectrometer inlet was used.

The Orbitrap Fusion mass spectrometer operated in the data-dependent mode, automatically switching between MS and MS/MS acquisition. Survey full-scan MS spectra (*m/z* 350-1600) was collected in the Orbitrap at a resolution of 60,000 and *m/z* 200, the automatic gain control (AGC) target was set to 1,000,000 with a maximum injection time of 50 ms. MS/MS acquisition was performed in the Orbitrap with a resolution of 15,000 at *m/z* 200. The intensity threshold was set at 50,000, and the maximum injection time was 100 ms. The AGC target for MS/MS was 100,000, and the isolation window was 1.6 *m/z*. Ions with charge states 2+, 3+, and 4+ were sequentially fragmented by high-energy collisional dissociation (HCD) with a normalized collision energy (NCE) of 28%. dynamic exclusion was set to 21 seconds.

Tandem mass spectra were extracted using Proteome Discoverer software (Thermo Fisher Scientific, version 2.4.0.305). All MS/MS spectra were analyzed using Sequest HT, searching the Uniprot-SwissProt database (Taxonomy: Homo Sapiens, 20,386 entries). Spectra were searched with a fragment ion mass tolerance of 0.020 Da and a parent ion tolerance of 10.0 ppm, assuming non-enzyme digestion.

**Matching criteria of peptide markers**

The molecular weight of an identified peptide should be within a tolerance of 2000 ppm of the *m/z* value of the MALDI-TOF peak. Only the charge state of 1+ was taken into account for the MALDI-TOF analysis. In cases where multiple peptides matched the feature peak based on the given criteria, the identification result was determined by choosing the one with the smallest mass difference.
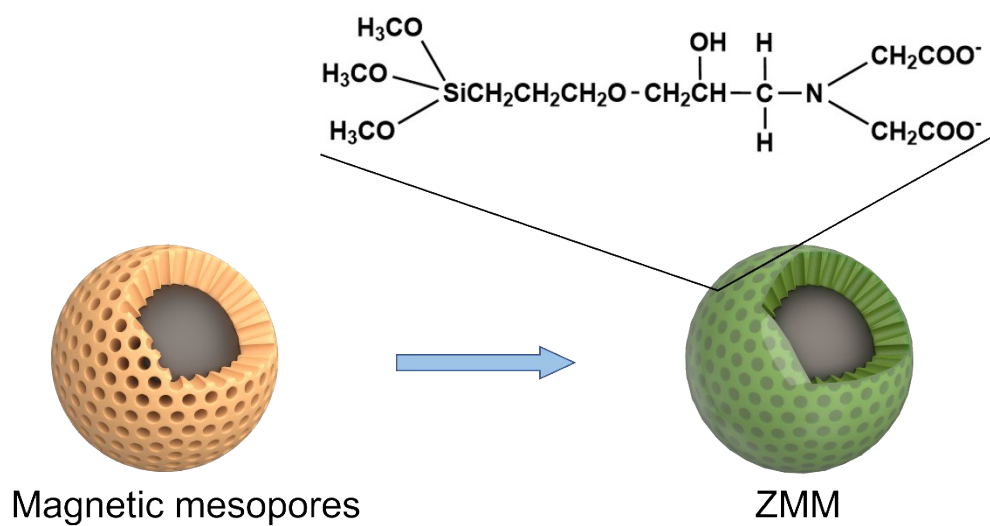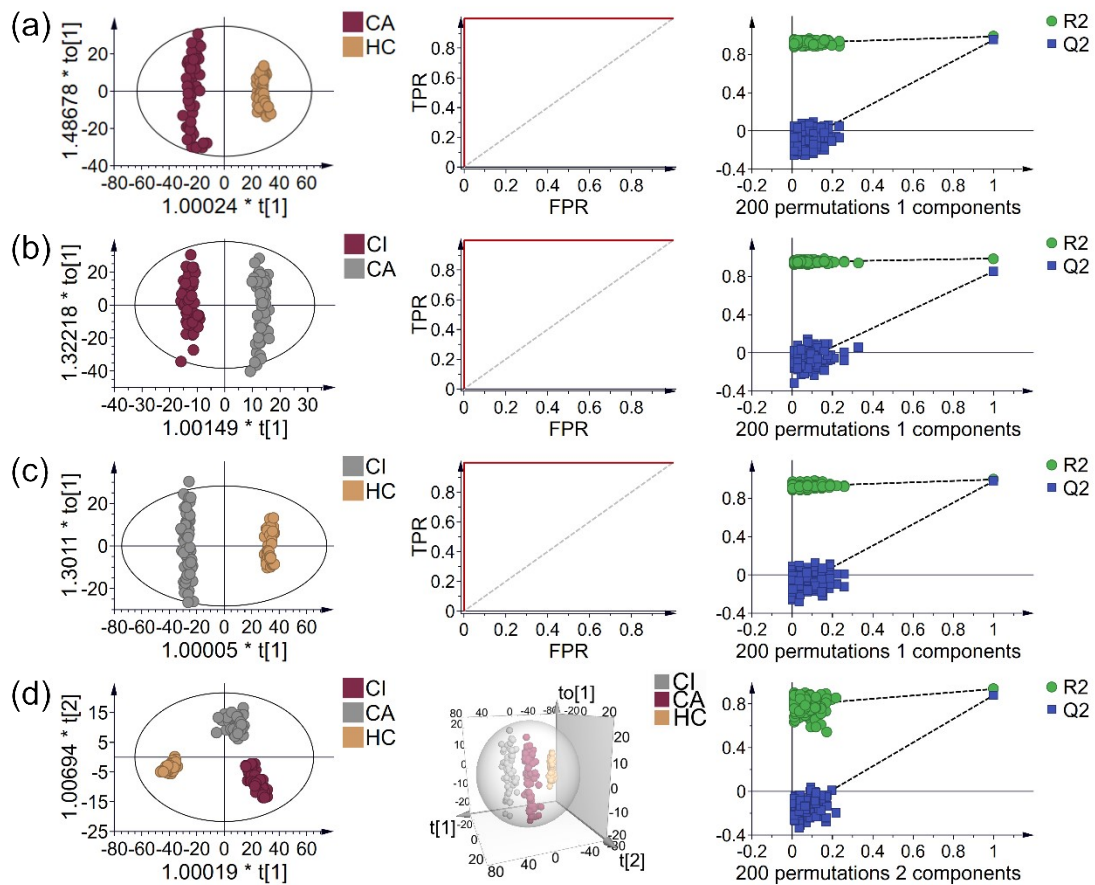
**Fig. S1.** The synthesis scheme of ZMM.
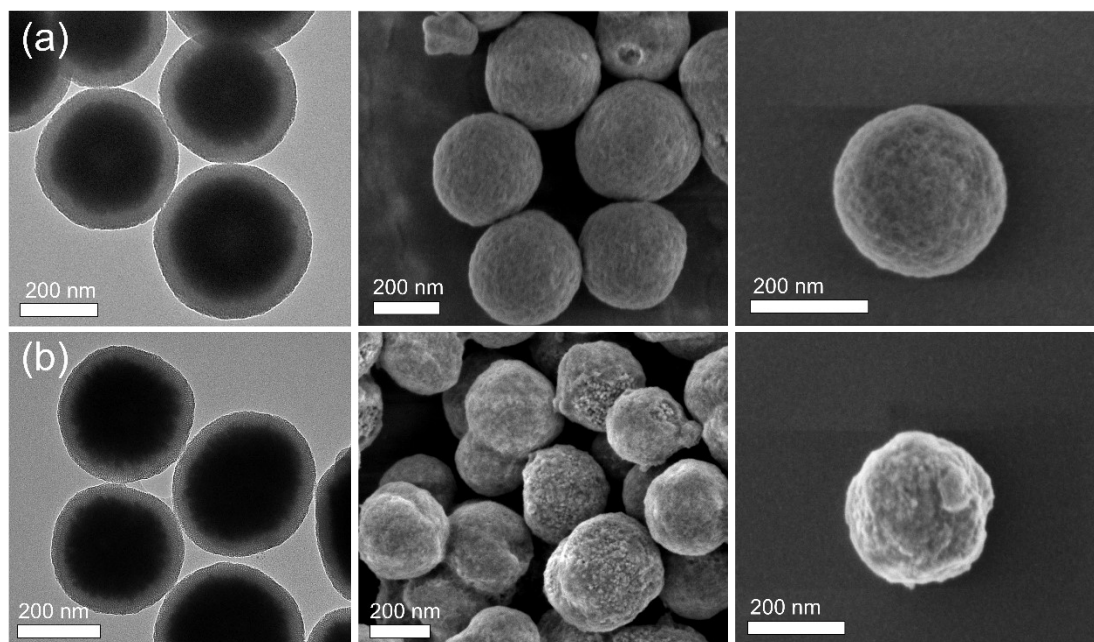


**Fig. S2.** TEM and FESEM images of magnetic silica (a) and ZMM (b).
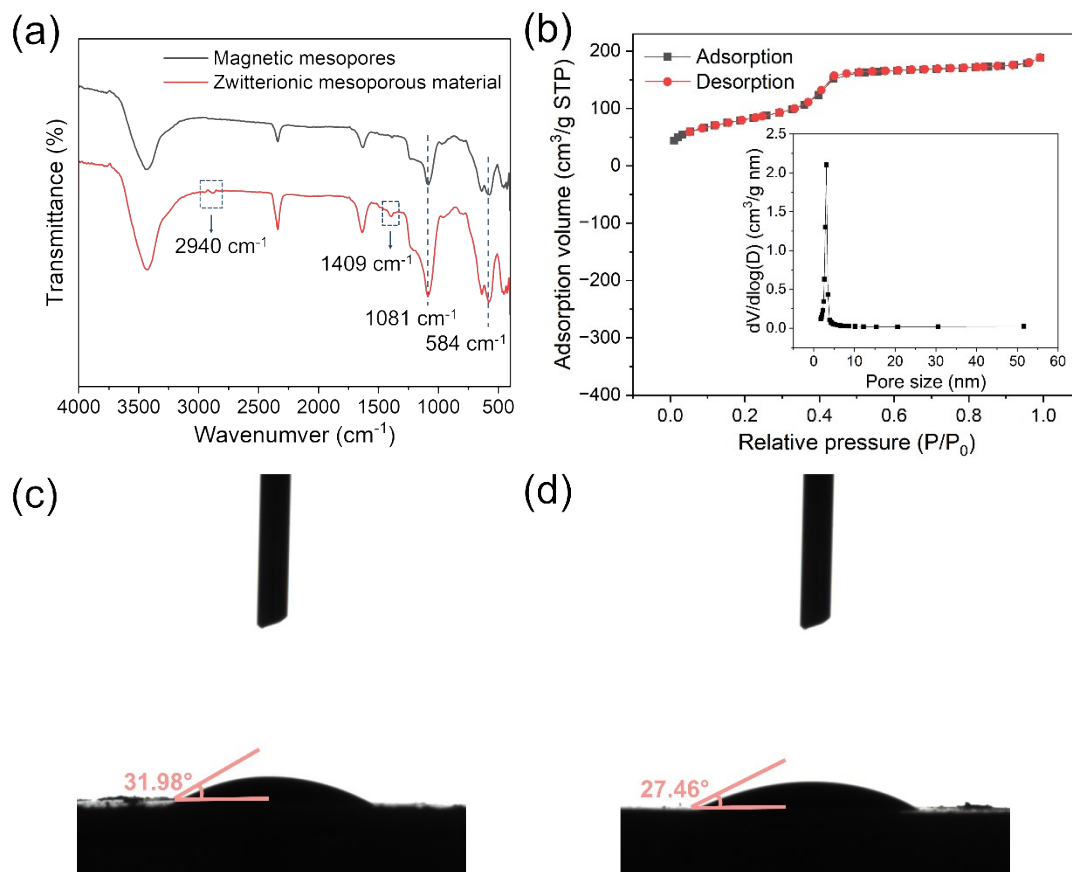
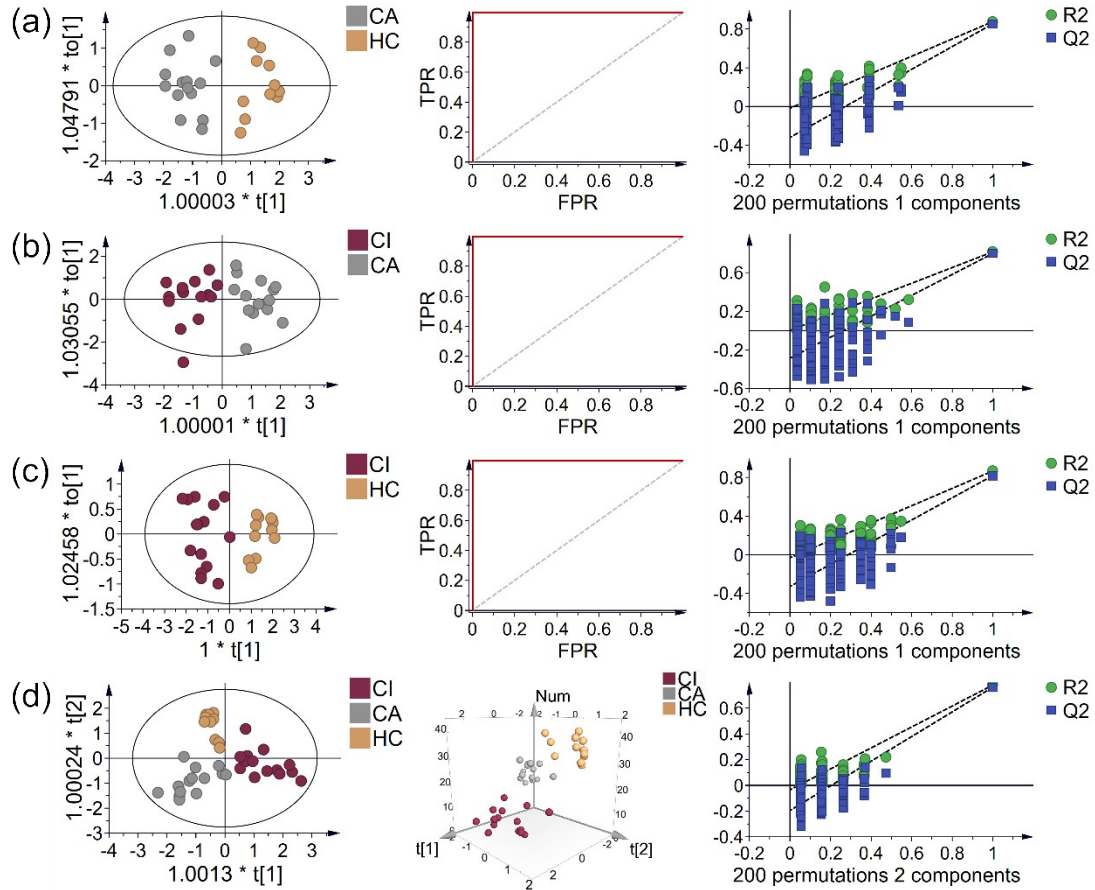**Fig. S3.** (a) FT-IR spectra of magnetic mesopores and ZMM. (b) The nitrogen adsorption-desorption isotherms of ZMM. Water contact angles of (c) magnetic mesopores and (d) ZMM.

**Fig. S4.** Evaluation of all *m/z* signals in the detection of CI and CA in discovery cohort. OPLS-DA score plots, ROC curves and corresponding permutations between (a) CA and HC, (b) CI and CA, (c) CI and HC, (d) CI, CA and HC.

**Fig. S5.** Evaluation of 3 markers in the detection of CI and CA in validation cohort. OPLS-DA score plots, ROC curves and corresponding permutations between (a) CA and HC, (b) CI and CA, (c) CI and HC, (d) CI, CA and HC.

**Table S1.** Mann-Whitney U test of gender distribution between liver cancer patients and healthy control groups.

| Sample | Total number | Male number (percentage) | Female number (percentage) | Z | P |
|--------|-------------|--------------------------|----------------------------|-----|-----|
| CA | 73 | 56(76.7%) | 17(23.3%) | -0.605 | 0.545 |
| HC | 61 | 44(72.1%) | 17(27.9%) | | |

**Table S2.** Mann-Whitney U test of gender distribution between liver cirrhosis patients and healthy control groups.

| Sample | Total number | Male number (percentage) | Female number (percentage) | Z | P |
|--------|-------------|--------------------------|----------------------------|-----|-----|

| | | | | | |
|---|---|---|---|---|---|
| CI | 75 | 43(57.3%) | 32(42.7%) | -1.781 | 0.075 |
| HC | 61 | 44(72.1%) | 17(27.9%) | | |

**Table S3.** Parameters of five different machine learning models using all *m/z* features of discovery cohort.

| Model | Model | AUC | Accuracy | Precision | F1 |
|---|---|---|---|---|---|
| | OPLS-DA | 1.000 | 0.992 | 0.992 | 0.992 |
| | KNN | 0.997 | 0.966 | 0.967 | 0.966 |
| CA vs HC | AdaBoost | 0.992 | 0.992 | 0.992 | 0.992 |
| | Tree | 0.992 | 0.992 | 0.992 | 0.992 |
| | Logistic Regression | 0.819 | 0.538 | 0.611 | 0.441 |
| | OPLS-DA | 1.000 | 1.000 | 1.000 | 1.000 |
| | KNN | 0.999 | 0.982 | 0.982 | 0.982 |
| CI vs HC | AdaBoost | 0.953 | 0.954 | 0.954 | 0.954 |
| | Tree | 0.980 | 0.982 | 0.982 | 0.982 |
| | Logistic Regression | 0.821 | 0.550 | 0.303 | 0.391 |
| | OPLS-DA | 1.000 | 0.991 | 0.991 | 0.991 |
| | KNN | 0.997 | 0.972 | 0.972 | 0.972 |
| CA vs CI | AdaBoost | 0.934 | 0.935 | 0.935 | 0.935 |
| | Tree | 0.956 | 0.963 | 0.963 | 0.963 |
| | Logistic Regression | 0.696 | 0.546 | 0.298 | 0.386 |

**Table S4.** Annotation of biomarkers.

| Feature peak (*m/z*) | Peptide | Protein | Gene name | Gravy | -lgP | Mass | Length | m/z | z | RT | Area | Intensity | Scan |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1335.67 | VDSGNDVTDIADD | Haptoglobin | ALB | -1.219 | 36.42 | 1334.547 | 13 | 668.2806 | 2 | 24.9055 | 5.04E+05 | 5.16E+06 | 10369 |
| 3378.12 | SSSYSKQFTSSTSY NRGDSTFESKSYK MAD | Fibrinogen alpha chain | HP | -0.631 | 26.48 | 3375.484 | 30 | 844.8773 | 4 | 31.9828 | 6.72E+05 | 1.29E+06 | 13451 |
| 1838.52 | DAHKSEVAHRFKD LGE | Albumin | FGA | -1.293 | 32.33 | 1837.907 | 16 | 613.6428 | 3 | 32.3106 | 2.16E+06 | 3.54E+07 | 13724 |

## References

1   S. Gibb, K. Strimmer, *Bioinformatics*, 2012, **28**, 2270-2271.
2   J. Demšar, T. Curk, A. Erjavec, Č. Gorup, T. Hočevar, M. Milutinovič, M. Možina, M. Polajnar, M. Toplak, A. Starič, *the Journal of machine Learning research*, 2013, **14**, 2349-2353.