**Supplementary Material**

**Metabolomics Approach to Understand the Hepatitis C Virus Induced Hepatocellular Carcinoma using LC-ESI-MS/MS**

**Sindhia Kumaria, Arslan Alib, Talat Roomec, Anam Razzakc, Ayesha Iqbalb, Amna Jabbar Siddiquib, Syed Muhammad Zahid Azamd, Hafeezullah Shaikhd, Hesham R. El‑Seedie,f,\* and Syed Ghulam Musharrafa,b,g,\***

*a H.E.J. Research Institute of Chemistry, International Center for Chemical and Biological Sciences, University of Karachi, Karachi-75270, Pakistan*

*b Dr. Panjwani Center for Molecular Medicine and Drug Research, International Center for Chemical and Biological Sciences, University of Karachi, Karachi-75270, Pakistan*

*c Section of Molecular Pathology, Department of Pathology, Dow International Medical College, Dow Diagnostic Reference and Research Laboratory, Dow University of Health Sciences, Ojha Campus, Karachi-74200, Pakistan*

*d National Institute of Liver & GI Diseases (NILGID), Dow University of Health Sciences, Ojha Campus, Karachi-74200, Pakistan*

*e Pharmacognosy Group, Department of Pharmaceutical Biosciences, Biomedical Centre, Uppsala University, SE- 751 23, Uppsala, Sweden*

*fInternational Research Center for Food Nutrition and Safety, Jiangsu University, Zhenjiang, 212013, China*

*g The Affiliated T.C.M Hospital of Southwest Medical University, Luzhou, Sichuan, China*

\*Corresponding author

*Correspondence:* **Prof. Dr. Syed Ghulam Musharraf, Ph.D.**

*H. E. J. Research Institute of Chemistry, International Center for Chemical and Biological Sciences, University of Karachi.*

*E-mail address:* [musharraf1977@yahoo.com](mailto:musharraf1977@yahoo.com); musharraf@iccs.edu

**Prof. Dr. Hesham R. El-Seedi**

*Pharmacognosy Group, Department of Pharmaceutical Biosciences, Biomedical Centre, Uppsala University, SE- 751 23, Uppsala, Sweden*

Tel.: +46 73 5668234

*E-mail address:* hesham.el-seedi@farmbio.uu.se

**Supplementary Figure Legends**

**Supplementary Figure 1.** Representative base peak chromatogram (BPC) of HS (green), HCV (blue), CLD (yellow), and HCC (red).

**Supplementary Figure 2.** PCA score plots showing **(A)** studied groups, **(B)** gender, and **(C)** age of individuals.

**Supplementary Figure 3.** Hotelling’s T2 plot with 95% and 99% confidence limit for the estimation of outlier in all groups.

**Supplementary Figure 4.** PCA analysis after outlier removal for healthy and diseased samples along with QC samples.

**Supplementary Figure 5.** ROC (Receiver Operating Characteristics) curve for OPLS-DA model, sensitivity on y-axis and 1-specificity on x-axis.

**Supplementary Figure 6.** Permutation tests for OPLS-DA model showing R2 and Q2 values. **(A)** HS, **(B)** HCV, **(C)** CLD, **(D)** HCC.

**Supplementary Figure 7.** OPLS-DA score plot showing original model with 16 blind samples for external validation studies.

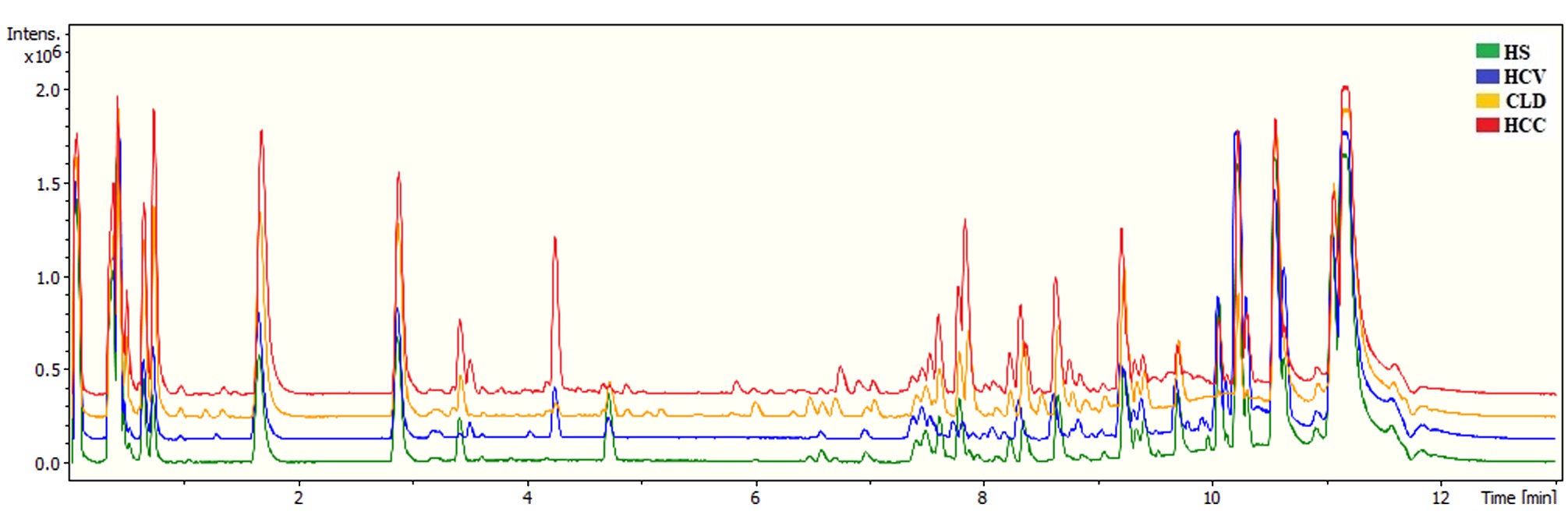
**Supplementary Figure 8.** Box-and-Whisker plot showing intensities of identified metabolites in HS (green), HCV (blue), CLD (yellow), and HCC (red).

**Supplementary Figure 9.** Glycerophospholipid metabolism, biological pathway showing glycerol-3-phosphocholine, LPC 18:2, phosphatidylcholine metabolites.

**Supplementary Figure 10.** Formation of L-tryptophan and L-phenylalanine in the phenylalanine, tyrosine, and tryptophan metabolism.

**Supplementary Figure 11.** Phenylalanine metabolism showing formation of phenylalanine.

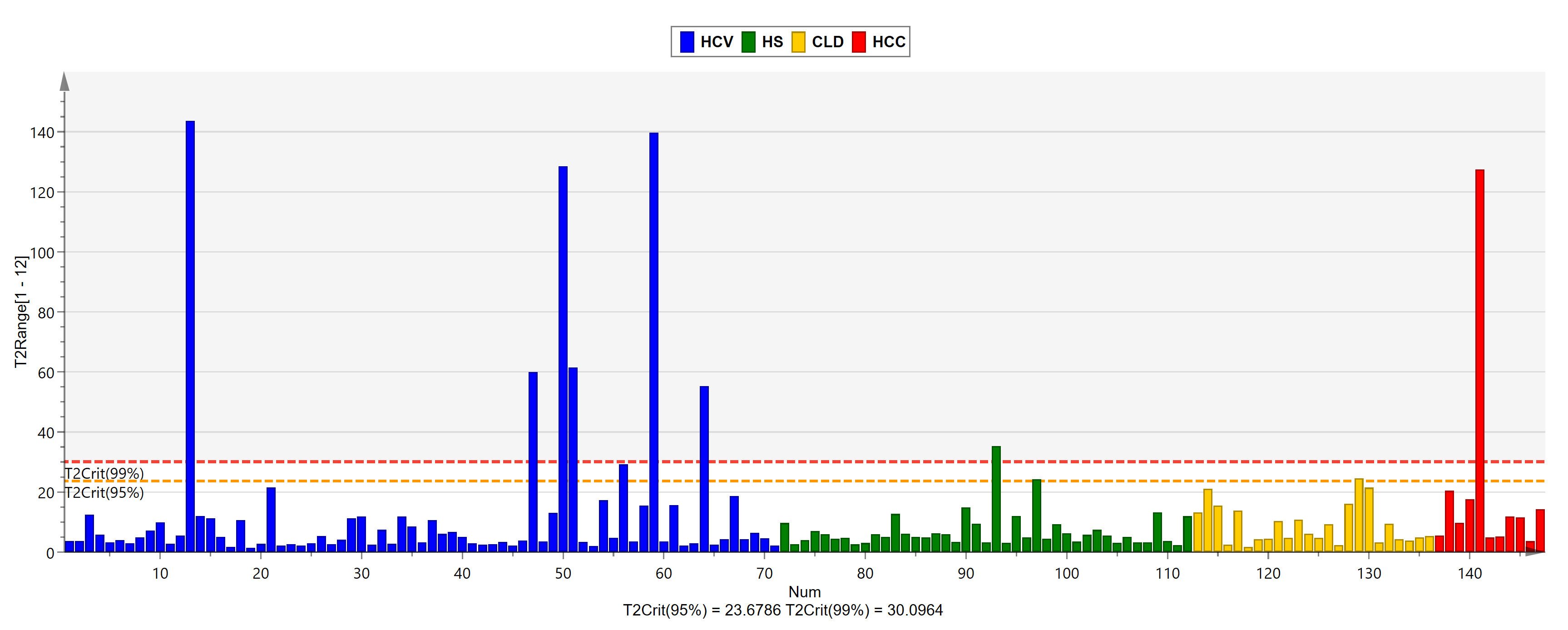
**Supplementary Figure 12.** Formation of hydroxyindoleacetic acid and L-tryptophan in the tryptophan metabolism.

****

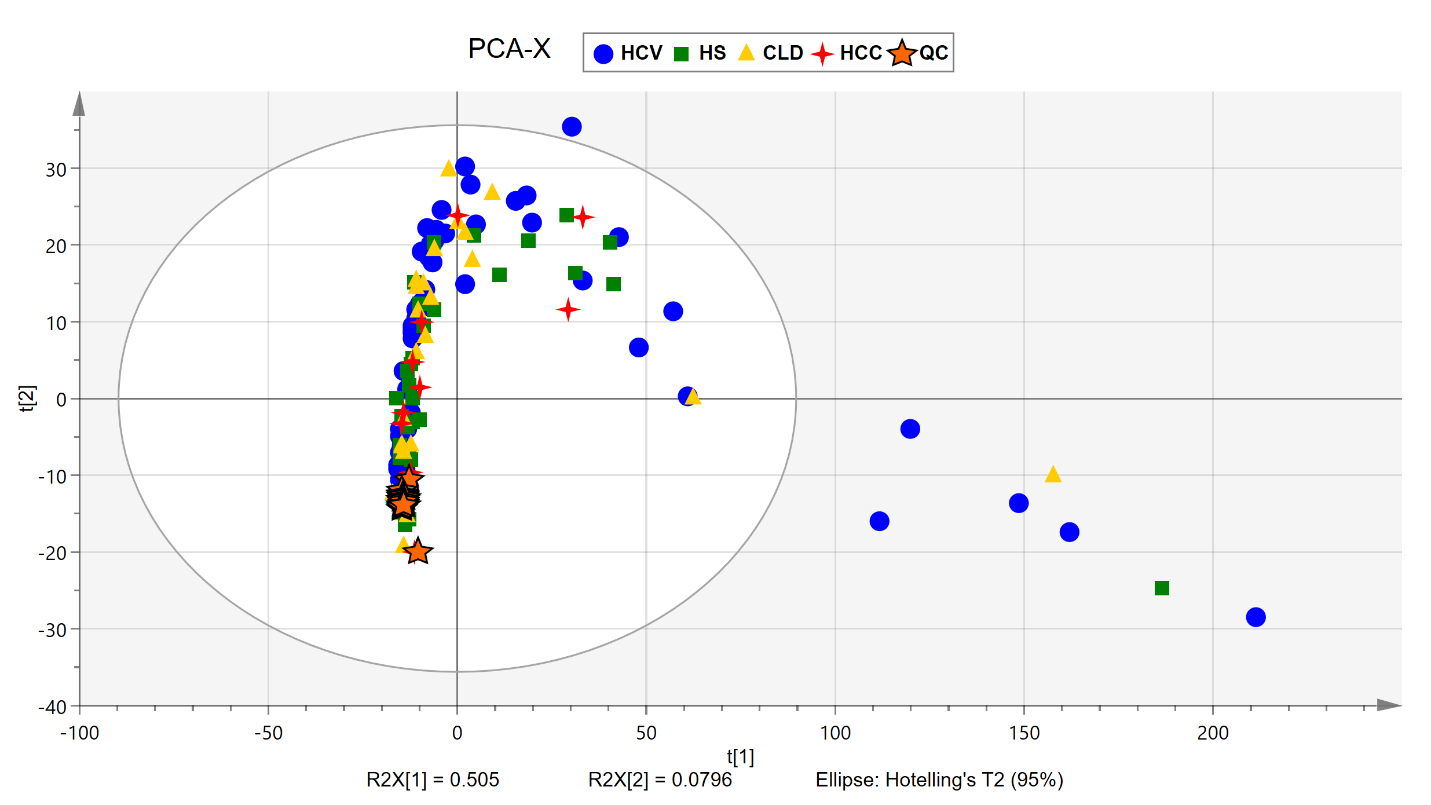
**Supplementary Figure 1.** Representative base peak chromatogram (BPC) of HS (green), HCV (blue), CLD (yellow), and HCC (red).

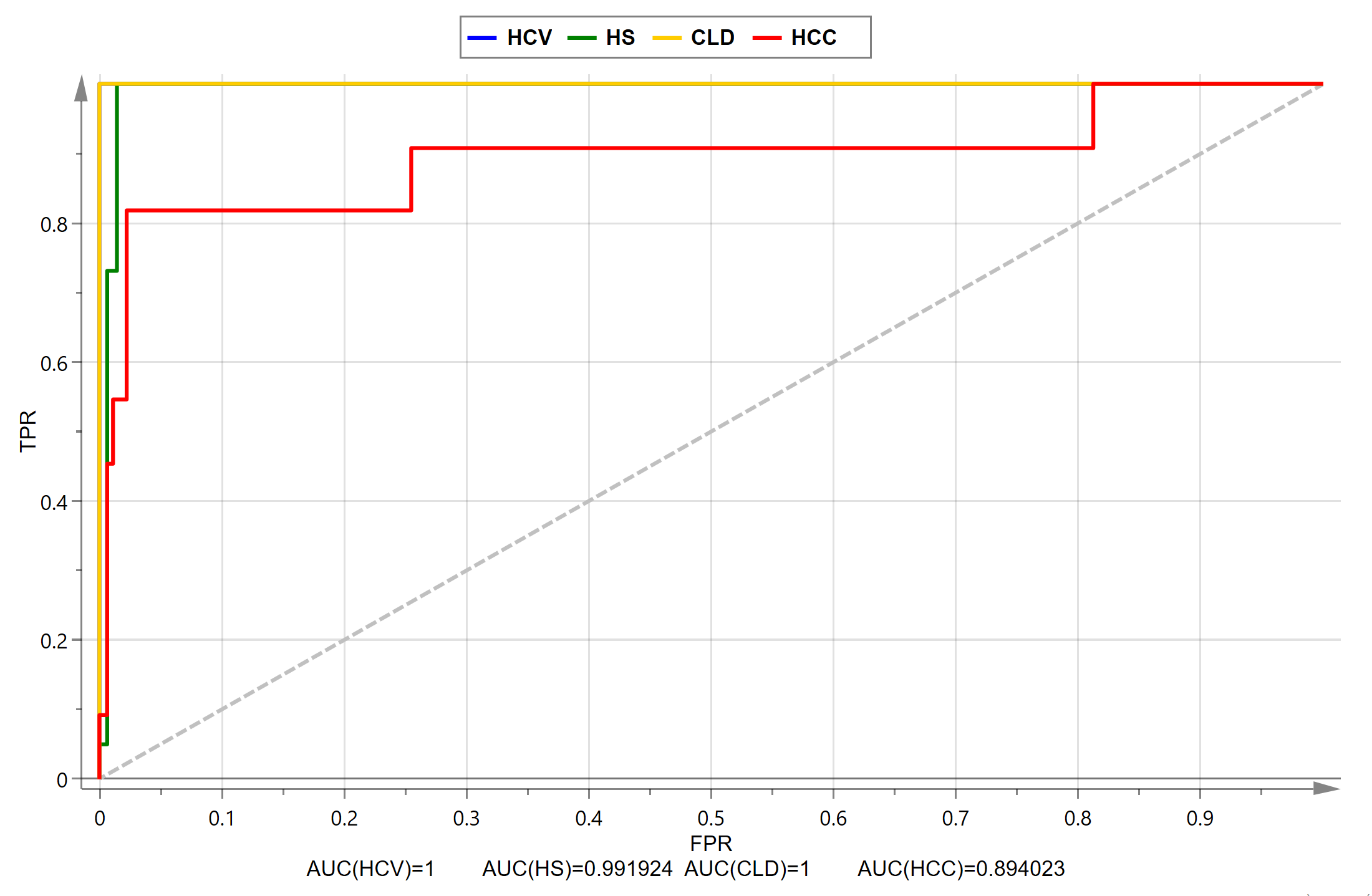


**Supplementary Figure 2.** PCA score plots showing **(A)** studied groups, **(B)** gender, and **(C)** age of individuals.

****

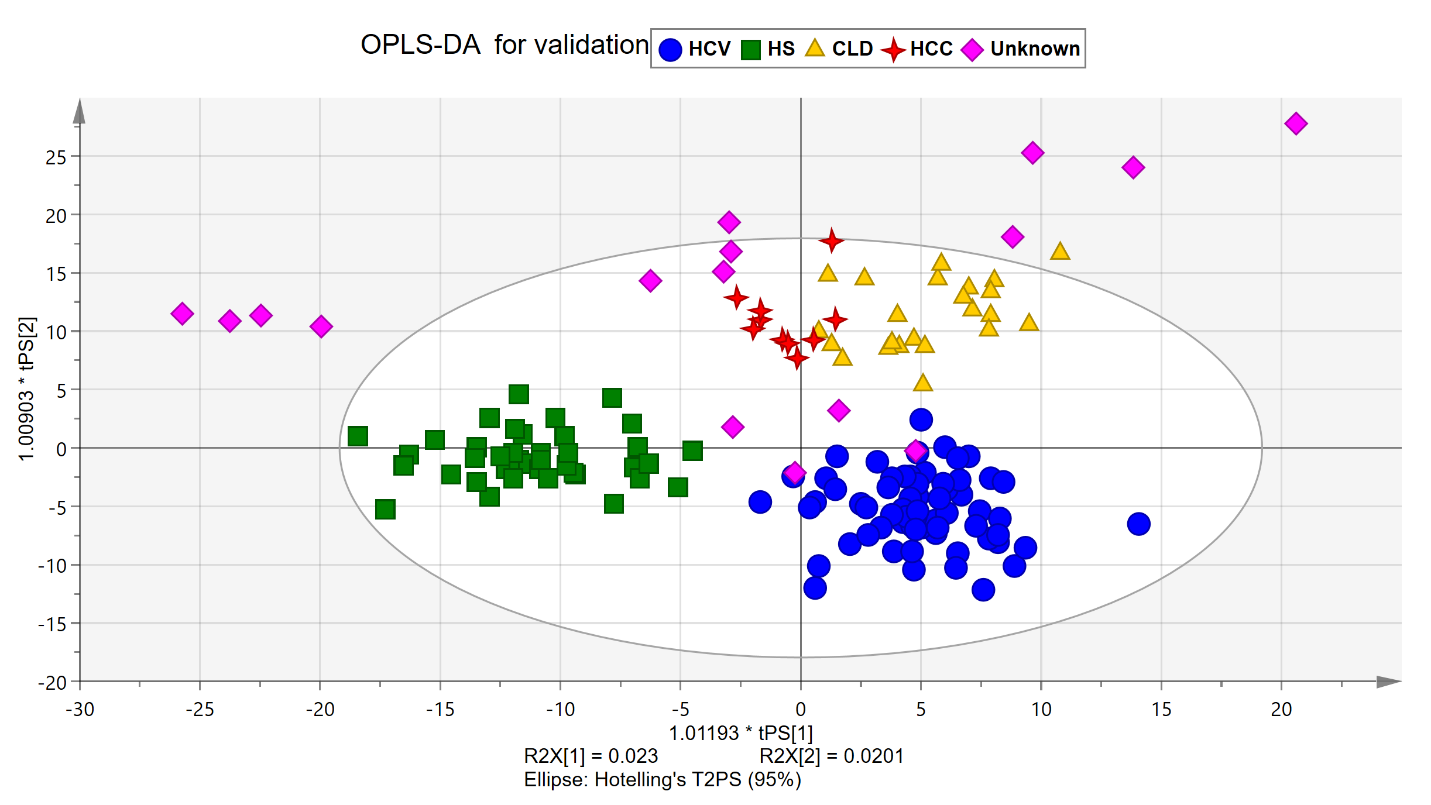
**Supplementary Figure 3.** Hotelling’s T2 plot with 95% and 99% confidence limit for the estimation of outlier in all groups.

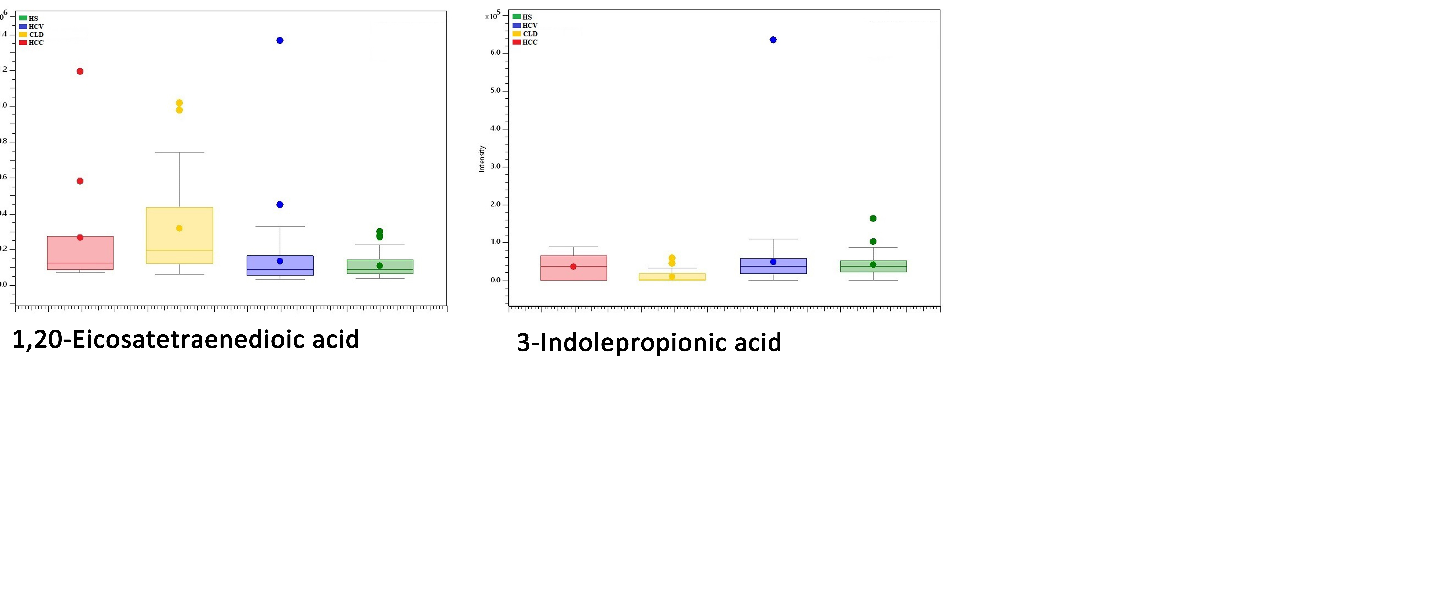
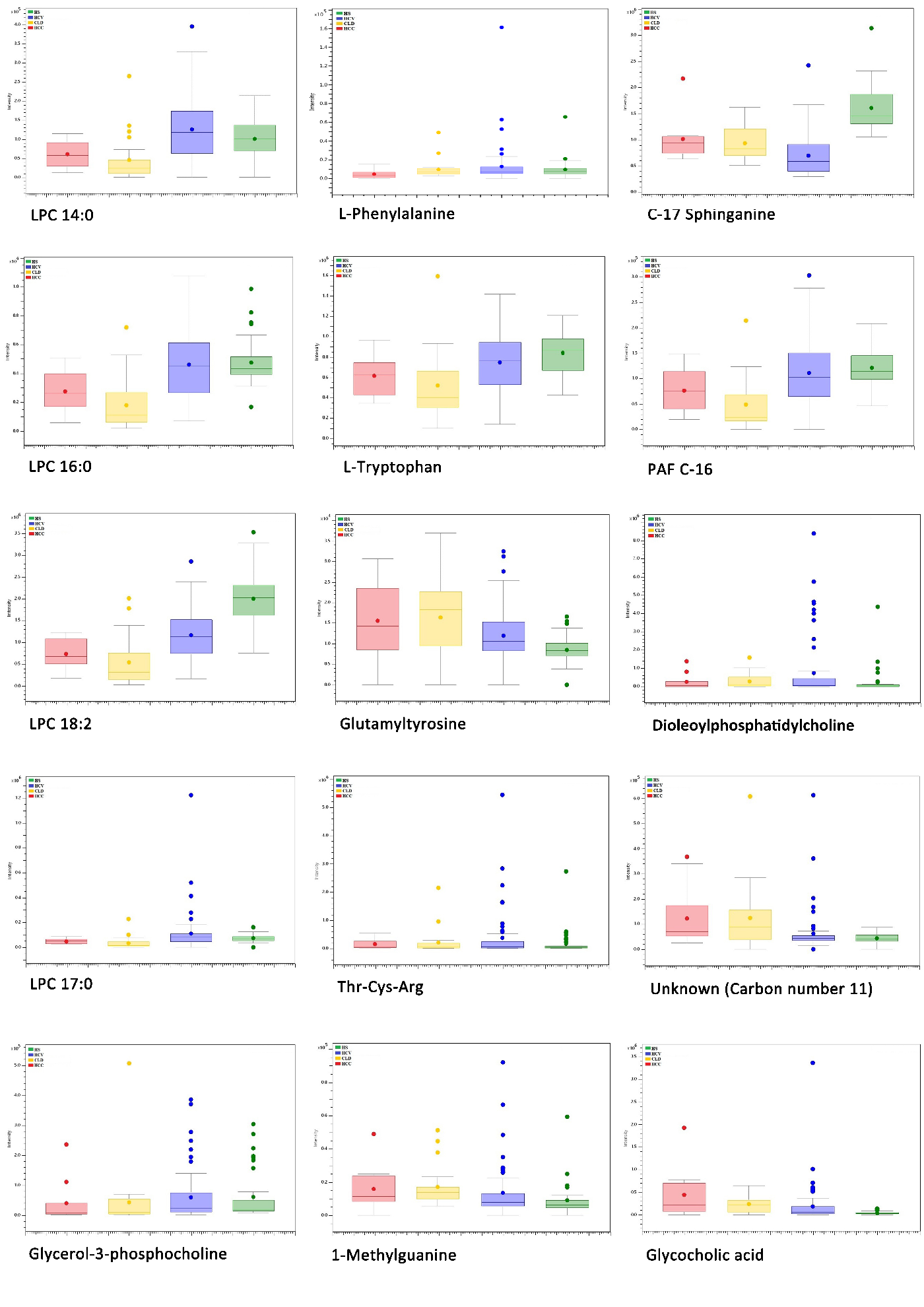
**** **Supplementary Figure 4.** PCA analysis after outlier removal for healthy and diseased samples along with QC samples.

****

**Supplementary Figure 5.** ROC (Receiver Operating Characteristics) curve for OPLS-DA model, sensitivity on y-axis and 1-specificity on x-axis.

**Supplementary Figure 6.** Permutation tests for OPLS-DA model showing R2 and Q2 values. **(A)** HS, **(B)** HCV, **(C)** CLD, **(D)** HCC.

** Supplementary Figure 7.** OPLS-DA score plot showing original model with 16 blind samples for external validation studies.

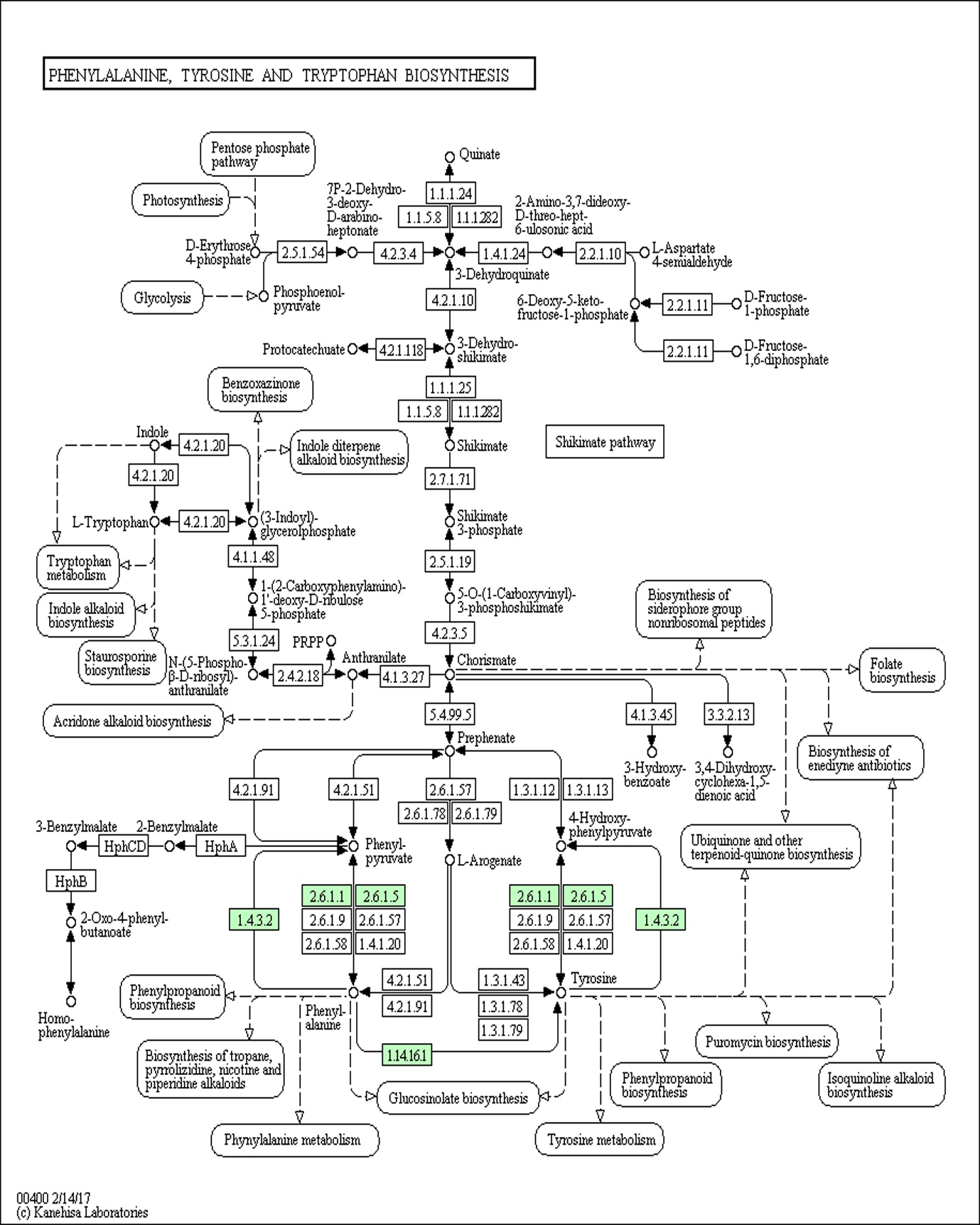
****

**Supplementary Figure 8.** Box-and-Whisker plot showing intensities of identified metabolites in HS (green), HCV (blue), CLD (yellow), and HCC (red).

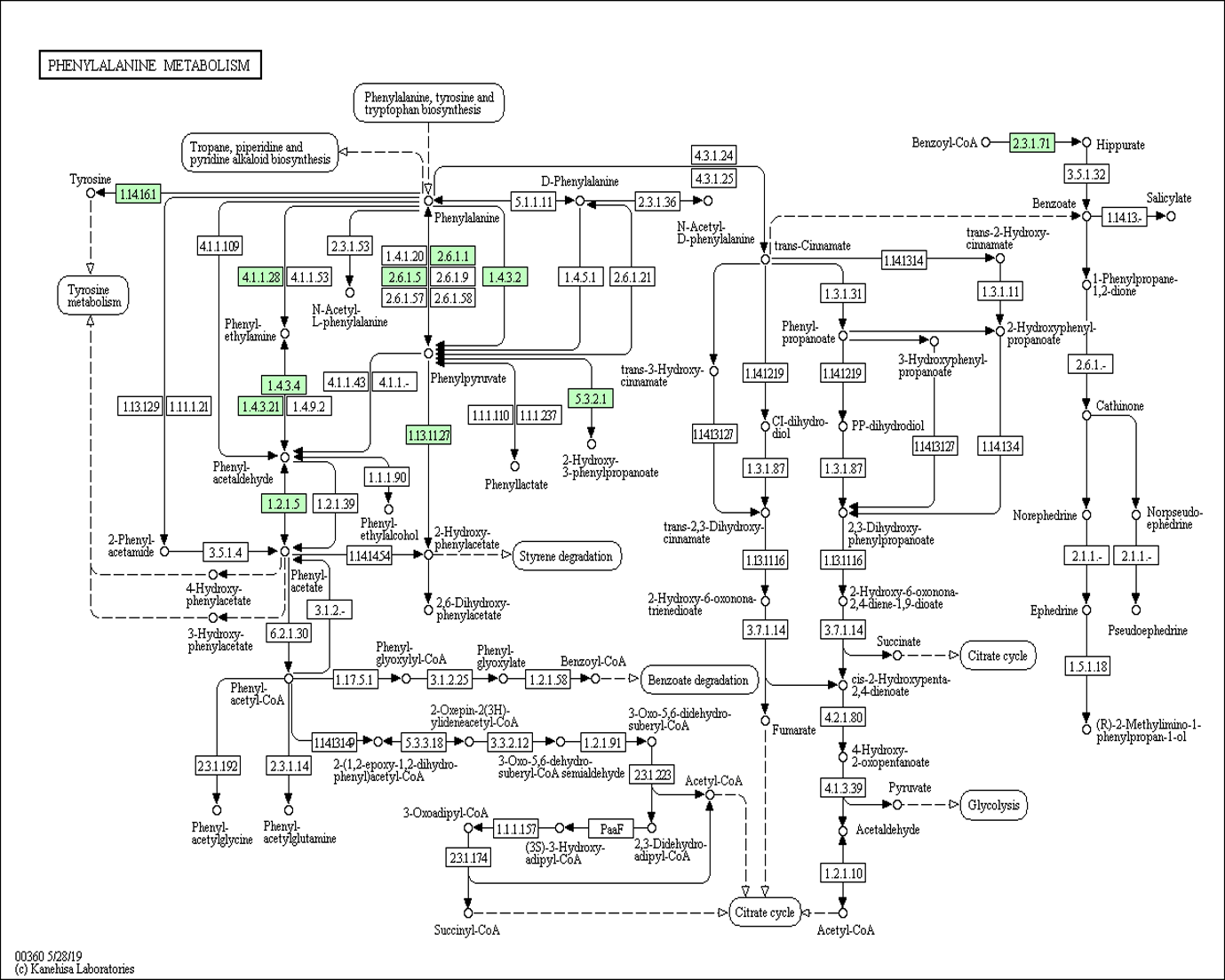
A picture containing text, map

Description generated with very high confidence

**Supplementary Figure 9.** Glycerophospholipid metabolism, biological pathway showing glycerol-3-phosphocholine, LPC 18:2, phosphatidylcholine metabolites.

****

**Supplementary Figure 10.** Formation of L-tryptophan and L-phenylalanine in the phenylalanine, tyrosine, and tryptophan metabolism.

****

**Supplementary Figure 11.** Phenylalanine metabolism showing formation of phenylalanine.

A picture containing text, map

Description generated with very high confidence

**Supplementary Figure 12.** Formation of hydroxyindoleacetic acid and L-tryptophan in the tryptophan metabolism.

**Supplementary Table Legends**

**Supplementary Table S1.** List of distinctive features among HS, HCV, CLD, and HCC at fold change >1.5 or *p-*value < 0.05.

**Supplementary Table S2.** Misclassification table for generated OPLS-DA model after removing outlier along with blind samples.

**Supplementary Table S1.** List of distinctive features among HS, HCV, CLD, and HCC at fold change >1.5 or *p-*value < 0.05.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Name | HS Vs HCV | | HS Vs CLD | | HS Vs HCC | |
|  | ***p*-value** | **FC** | ***p-*value** | **FC** | ***p*-value** | **FC** |
| 1,20-Eicosatetraenedioic acid | 0.39 | -1.22 | 0.02 | -2.89 | 0.85 | -2.45 |
| C17-Sphinganine | 0 | 2.31 | 0.000001 | 1.72 | 0.04 | 1.58 |
| Glutamyltyrosine | 0.001 | -1.47 | 0.0004 | -2.05 | 0.19 | -1.88 |
| Glycocholic acid | 0.07 | -5.49 | 0.001 | -6.99 | 0.45 | -13.54 |
| L-Aspartyl-L-phenylalanine | 0.42 | 1.15 | 0.81 | 1.31 | 0.002 | 4.27 |
| LPC 14:0 | 0.11 | -1.29 | 0.01 | 2.10 | 0.09 | 1.79 |
| LPC 16:0 | 0.77 | 1.02 | 0.000006 | 2.66 | 0.04 | 1.73 |
| LPC 17:0 | 0.24 | -1.59 | 0.06 | 1.91 | 0.07 | 1.54 |
| LPC 18:2 | 0 | 1.70 | 0 | 3.64 | 0.000001 | 2.70 |
| N-Fructosyl tyrosine | 0.02 | -2.02 | 0.024 | -2.22 | 0.15 | -5.01 |
| PAF C-16 | 0.84 | -1.02 | 0.0001 | 2.38 | 0.14 | 1.58 |
| Thr-Cys-Arg | 0.25 | -2.15 | 0.93 | -1.19 | 0.97 | -1.11 |
| Unknown (carbon number 11) | 0.25 | -1.43 | 0.04 | -2.87 | 0.55 | -2.69 |
| 1-Methylguanine | 0.23 | -1.56 | 0.07 | -1.85 | 0.62 | -1.92 |
| Dioleoylphosphatidylcholine | 0.25 | -2.70 | 0.97 | -1.03 | 0.98 | 1.08 |
| Hydroxyindoleacetic acid | 0.08 | -1.20 | 0.32 | -1.36 | 0.53 | -1.41 |
| L-Tryptophan | 0.25 | 1.12 | 0.005 | 1.60 | 0.12 | 1.36 |
| Thyroxine | 0.97 | 1.00 | 0.22 | 1.25 | 0.004 | 1.71 |
| 3-Indolepropionic acid | 0.1 | 1.22 | 0.000006 | 2.29 | 0.008 | 1.68 |
| Glycerol-3-phosphocholine | 0.0001 | 1.51 | 0.000009 | 2.65 | 0.000002 | 2.31 |
| L-Phenylalanine | 0.30 | 1.15 | 0.62 | 1.08 | 0.01 | 2.06 |

**Supplementary Table S2.** Misclassification table for generated OPLS-DA model after removing outlier along with blind samples.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Members | Correct | HCV | HS | CLD | HCC |
| HCV | 64 | 100% | 64 | 0 | 0 | 0 |
| HS | 39 | 100% | 0 | 39 | 0 | 0 |
| CLD | 23 | 100% | 0 | 0 | 23 | 0 |
| HCC | 10 | 100% | 0 | 0 | 0 | 10 |
| Blind samples | 16 |  | 3 | 4 | 4 | 5 |
| Total | 152 | 100% | 67 | 43 | 27 | 15 |