1-1-2020

Population-based case-control study revealed metabolomic biomarkers of suboptimal health status in Chinese population—potential utility for innovative approach by predictive, preventive, and personalized medicine

Hao Wang
Edith Cowan University

Qiuyue Tian

Jie Zhang

Hongqi Liu

Xiaoyu Zhang

See next page for additional authors

10.1007/s13167-020-00200-7

Comprehensive Matabolomic Features of Suboptimal Health Status: A Population-based Case-control Study in a Chinese Han Population

Hao Wang[1,2,*], Qiuyue Tian[1,*], Jie Zhang[1], Hongqi Liu[3], Xiaoyu Zhang[1], Weijie Cao[1], Jinxia Zhang[1], Xingang Li[2], Xueqing Wang[2], Di Liu[1], Yulu Zheng[2], Zheng Guo[2], Lijuan Wu[1], Manshu Song[1], Youxin Wang[1,#], and Wei Wang[1,2,4,#]

[1]Beijing Key Laboratory of Clinical Epidemiology, School of Public Health, Capital Medical University, Beijing, China.

[2]School of Medical and Health Sciences, Edith Cowan University, Perth, Australia.

[3]Student Health Center, Weifang University, Weifang, China

[4]School of Public Health, Shandong First Medical University, Taian, China.

[*]These authors contributed equally to this work.

Short title: *Matabolomic Features of Suboptimal Health Status*

Address correspondence to:
Prof. Youxin Wang, PhD
School of Public Health
Capital Medical University
10 Youanmenwai Xitoutiao Road
Fengtai District
Beijing 100069
China

E-mail: wangy@ccmu.edu.cn

AND

Prof. Wei Wang, MD, PhD, FFPH, FRSB, FRSM
School of Medical and Health Sciences
Edith Cowan University
270 Joondalup Drive
Perth WA 6027
Australia

Email: wei.wang@ecu.edu.au

Abstract

Background Metabolomics, the identification and quantification of all metabolites in a biological system, is considered as a promising technique for biomarker discovery and mechanism elucidation. Chronic diseases occur at an elevated frequency in people with suboptimal health status (SHS). The recognition of SHS plays a significant role in the preventive, predictive and personalisded medicine (PPPM) of chronic diseases. Although a subjective screening tool for SHS has been developed, the objective biomarkers that give a better understanding of the pathophysiology of SHS warrant further investigation.

Methods To report the original observations on the metabolomic feature of SHS, a liquid chromatography-mass spectrometry-based untargeted metabolomics analysis was conducted on plasma samples collected from 100 partciapnts (50 with SHS and 50 age- and sex-matched healthy controls).

Results After controlling for the confounding factors (smoking, drinking, low-density lipoprotein cholesterol level, sleep duration, insomnia, anxiety, depression, and physical activity level) 24 significantly differential metabolites were identified as the candidate biomarkers for SHS. Pathway analysis revealed that 1) sphingolipid metabolism, 2) taurine metabolism, and 3) steroid hormone biosynthesis are the disturbed metabolic pathways related to SHS. Protein-metabolite-disease network analysis showed that 148 proteins ? associated with 13 potential metabolic moelculaes for SHS, which could be potiential biomarkers for 17 typs of chornic diseases. A classification model was constructed based on the 24 candidate biomarkers, and the model yielded a sensitivity of 94.0%, a specificity of 90.0%, and an AUC of 0.977 (95% CI: 0.955 - 0.998, $P < 0.001$).

Conclusion These findings support that plasma metabolomics could be used as an objectively diagnostic tool for SHS. Metabolic biosignature of SHS warrants further research in a larger population and the applications of SHS in PPPM practice.

Abstract word count = 272

Keywords:

Metabolomics, suboptimal health status, biomarkers, sphingolipid metabolism, Chinese Han population

Introduction

Suboptimal health status (SHS) is a physical state between health and disease characterized by the perception of health complaints, chronic fatigue, and a constellation of physical symptoms lasting for at least three months (1, 2). It is recognized as a subclinical, reversible stage of chronic disease (3). Many health problems, such as psychosocial stress (4), cardiovascular risk factors (5, 6), and type 2 diabetes mellitus (T2DM) (7) occur at an elevated frequency in people with SHS. The recognition of SHS plays a significant role in the prevention of chronic disease from the perspective of predictive, preventive and personalized medicine (PPPM) (3).

Based on the perceived health complaints affected by SHS, a subjective screening tool for SHS, suboptimal health status questionnaire-25 (SHSQ-25), has been developed and validated in African, Asian and Caucasian ethnic groups (6-9). The SHSQ-25 includes 25 items encompassing five domains: fatigue, the cardiovascular system, the respiratory system, the immune system and mental status (2). A preliminary diagnostic criterion(score?) for SHS (SHSQ-25 score ≥ 35) has also been recommended (10, 11). In order to understand the pathophysiology of SHS, several objective biomarkers, including cortisol, adrenaline, and noradrenaline, have been investigated (4, 12). Although previous studies have indicated that chronic psychosocial stress is associated with SHS, the underlying mechanisms of SHS remain partially understood.

Metabolites are the small molecules of metabolism that drive essential cellular functions, such as energy production and storage, apoptosis and signal transduction (13). Metabolomics is an emerging “omics” tool involving the identification and quantification of all endogenous and exogenous metabolites in tissues, cells, and biofluids (14). It provides a snapshot of functional status of a biological system. Metabolomics has several advantages for the diagnosis of complex chronic disease. First, metabolites are the intermediate or end products of metabolism, which reflects what has been encoded by the genome and modified by environmental changes. Therefore, the levels of

metabolites could be regarded as the ultimate response of biological systems to both genetic and environmental factors (15). Second, metabolome, which is connected with genome, transcriptome and proteome, represents the most downstream stage in the dynamic system of cellular processes. Metabolites are therefore easier to correlate with phenotype (16). In addition, because metabolomics provides downstream information of underlying biological pathways, the results can provide directions to develop treatments for disease (17). Given these advantages, the metabolome has become widely accepted as the dynamic and sensitive measures of the phenotype at the molecular level, and the metabolomics is at the forefront of biomarker and mechanism discovery.

The main methodologies of metabolomics research are typically categorized as targeted metabolomics and untargeted metabolomics. Target metabolomics, which focuses on identifying and quantifying a predefined set of metabolites based on a priori information, is usually hypothesis-driven. In the targeted approach, the chemical properties of the investigated metabolites are known, and methods are optimized for the analysis of specific metabolites and metabolic pathways of interest. By contrast, untargeted metabolomics aims to measure as many metabolites as possible from biological samples without a priori knowledge. The major advantage of untargeted metabolomics is the discovery of novel metabolites in relation to the study context, and it is therefore considered as hypothesis-generating method. Although untargeted metabolomics can be performed using nuclear magnetic resonance (NMR), gas chromatography-mass spectrometry (GC-MS), or liquid chromatography-mass spectrometry (LC-MS) technologies, LC-MSenables to detect the broadest range of metabolites in a single analysis and therefore has been technique for metabolite profiling. LC-MS-based metabolomics has been applied in various research areas including biomarker discovery (18), precision medicine (15, 19), drug discovery (17), and systems biology (20).

The aims of this study were to describe comprehensive metabolomic biosignature for SHS, and

screen objectively metabolic biomarkers for SHS using LC-MS-based untargeted metabolomics. In addition, we investigated the metabolic pathways and the protein-metabolite-disease interaction network of the potential metabolite biomarkers to further understand the biological processes involved in SHS, which may prove to be useful for PPPM of chronic disease.

Materials and Methods

Study design and participants

A population-based case-control study was conducted in a Chinese Han population who received routine health check-up at the student health centre of *Weifang* University between September 2017 and November 2017. In order to minimize the influence of age on metabolomics, the source population was undergraduate students aged 18 to 20 who were living in the university dormitory.

All participants were required to meet the following inclusion criteria: (1) Chinese Han individuals; (2) aged 18 years or older; and (3) *Shandong* province native residents. Participants were excluded from the study if they met any of the following criteria: (1) history of somatic and psychiatric abnormalities; (2) intake of medication or supplements during the past three month; (3) history of any diseases involving the cardiovascular and cerebrovascular system, respiratory system, genitourinary system, digestive system, immune system and hematic system; (4) history of any surgeries or (5) pregnant or lactating women.

Figure 1 shows the flow of potential participants through the study. ??? participants were asked to complete SHSQ-25 questionnaires at two time points of baseline and three? Months later, and participants with SHSQ-25 score ≥ 35 in both two surveys were selected as cases. Then, age- and sex-matched healthy participants with SHSQ-25 score < 35 in both two surveys were selected as controls. In total, the case group comprised 50 participants with SHS and the control group

comprised 50 healthy participants without SHS were selected for blood biochemistry assays and LC-MS analysis .

This study was approved by the Ethics Committee of the *Weifang* University, Weifang, China. Written informed consent was obtained from each participant at the beginning of the study. The ethics approval was given in compliance with the Declaration of Helsinki.

Covariates

Demographic characteristics of participants, including age, sex, marital status, and ethnicity, were collected by questionnaires. Lifestyle information on smoking, drinking, and sleep duration were also collected. Participants were assigned to be non-smokers (never regularly smoked daily) or smokers (former smoker or current smoker) according to their smoking status. Drinking status was classified as non-drinker, light to moderate drinker, or heavy drinker according to the 2015-2020 Dietary Guidelines for Americans (21).

Anthropometric measurements were carried out by trained nurses and physicians. Height, weight, waist circumference (WC) and hip circumference (HC) were measured with the participants wearing only indoor clothing and without shoes. Body mass index (BMI) was calculated as weight in kilograms divided by height in metres squared (kg/m^2). Waist-to-hip ratio (WHR) was calculated as WC in centimetres divided by HC in centimetres. Systolic blood pressure (SBP) and diastolic blood pressure (DBP) were measured twice on the right arm using a standard mercury sphygmomanometer with the participants resting for at least 10 minutes in a sitting position.

The blood biochemistry assays included measurements of the levels of fasting plasma glucose (FPG), total cholesterol (TC), triglycerides (TG), high-density lipoprotein cholesterol (HDL), low-density lipoprotein cholesterol (LDL), uric acid (UA), creatinine, blood urea nitrogen (BUN), alanine transaminase (ALT), aspartate transaminase (AST), creatine kinase-MB (CK-MB), lactate

dehydrogenase (LDH), and alkaline phosphatase (ALP).

The levels of physical activity were measured using the long form of the International Physical Activity Questionnaire (IPAQ) and classified as low, moderate and high levels according to the protocol for IPAQ (22). The measurements of insomnia, anxiety, and depression were assessed using the Athens Insomnia Scale (23), the Hamilton Anxiety Rating Scale (24), and the Hamilton Depression Rating Scale (25), respectively.

Plasma collection and metabolomics analysis

Fasting blood samples were collected by venipuncture in the morning after an overnight fasting. The plasma samples were then separated in the laboratory after centrifugation at 3000 rpm for 10 minutes and immediately stored at -80°C until metabolomics analysis.

Plasma samples were thawed and proteins were precipitated by adding 300 µL of cold methanol to 100 µL of plasma. The mixture was then centrifuged at 12,000 rpm for 15 min at 4 °C and 200 µL of supernatant was transferred into polypropylene tubes for further analyses. To assess the stability and reproducibility of LC-MS analyses, a quality control (QC) sample was prepared by mixing equal volumes (10 µL) of the collected 100 plasma samples. The QC sample was also subjected to similar protein precipitation procedure and seven aliquots of QC sample were injected during LC-MS analyses.

LC-MS analysis was performed using an LC-MS system consisting of an UltiMate 3000 liquid chromatography system (Thermo Fisher Scientific Waltham, USA) coupled with an Orbitrap Elite mass spectrometer (Thermo Fisher Scientific Waltham, USA). A 4 µL aliquot of each sample was injected onto a C18 column (100 x 4.6 mm, 3 µm) with the column temperature maintained at 40 °C. The flow rate was 0.3 mL/min, and the mobile phase consisted of ultrapure water with 0.1% (v/v) formic acid (A) and acetonitrile with 0.1% (v/v) formic acid (B). The gradient of the mobile phase is

shown in Supplementary Table 1.

The column eluent was directed to the MS and full-scan profiling data were acquired in the Obitrap mass analyser with the scan range of 50-1000 mass-to-charge ratio (*m/z)* in both positive and negative ion mode. The source ion parameters applied were as follows: heater temperature of 300 °C, sheath gas flow rate of 45 arbitrary units, auxiliary gas flow rate of 15 arbitrary units, sweep gas flow rate of 1 arbitrary units, capillary temperature of 350 °C, spray voltage of 3 kV (positive ion mode) and 3.2 kV (negative ion mode).

Statistical analysis

The acquired mass spectrometry data were converted into .mzXML format by using ProteoWizard msConvert (26). The XCMS online was applied for pre-processing of raw metabolomics data, including feature detection, retention time correction and alignment (27). The 80% rule was used to treat the missing values, then a list including *m/z*, retention time and peak intensity was generated.

Normality distribution of all variables were tested by the *Shapiro-Wilk* test. Normally distributed continuous variables were reported as mean ± standard deviation (SD), and non-normally distributed continuous variables were represented as medians and interquartile ranges (IQR). Categorical variables were represented as frequencies and percentages. The differences in categorical variables between the two groups were tested by *Chi-square* test or *Fisher's* exact test. The differences in continuous variables between the two groups were tested by *Student t*-test or *Mann-Whitney U* test. The *Benjamini-Hochberg* method was used to perform the multiple testing corrections and false discovery rate (FDR) was set to 0.05. The principal component analysis (PCA) was used to provide an informative summary of the metabolomics dataset and relationships between groups. To identify the metabolites responsible for the discrimination, the orthogonal partial

least squares projection-discriminant analysis (OPLS-DA) model was performed by using SIMCA 14.1 software (Umetrics, Umea, Sweden). Open source databases, including Kyoto Encyclopedia of Genes and Genomes (KEGG) (28), Human Metabolome Database (HMDB) (29), METLIN (30), and PubChem (31) were used to identify the metabolites and metabolic pathways. Correlation analysis was performed to estimate the association between identified metabolites by *Spearman's* rank correlation method and correlation coefficients were shown in plot (R package "*corrplot*"). A protein-metabolite-disease interaction network was created by using MetaboAnalyst (32) based on information gathered from HMDB (29) and Search Tool for Interactions of Chemicals (STITCH) database (33), and the Cytoscape software 3.7.1 (National Institute of General Medical Sciences, Bethesda, USA) was used to plot the interaction networks. Multivariate binary logistic regression was used to construct classification models for SHS. Receiver operating characteristic (ROC) curves and the area under the curve (AUC) were used to assess the classification performance of the models.

Data analysis was performed using SPSS 25.0 (IBM Corporation, New York, USA) and R 3.4.3 (34). All reported *P* values were two-tailed, and $P < 0.05$ was considered statistically significant.

Results

Characteristics of participants

In total, 50 participants with SHS and 50 age- amnd sex-matched healthy controls without SHS were included in this study. The demographic, anthropometric and biochemical characteristics of participants are described in Table 1. The average ages of SHS and control groups were 19.00 ± 0.73 and 18.98 ± 0.65 years respectively, with the same age range of 18.00 - 20.00 years. The level of LDL and sleep duration in the SHS group was significantly lower than those in the control group, whereas higher SHS, insomnia, anxiety, and depression scores were observed in the SHS group (*P*

< 0.05). In terms of other characteristics, there were no statistically significant differences observed between the two groups (*P* > 0.05).

Discovery and identification of potential metabolic biomarkers

LC-MS technique was used to performed metabolomics analysis on 100 plasma samples and 7 QC samples, and the total ion chromatograms of these samples are presented in Supplementary Figure 1. PCA plots were performed to show the within-group and between-group variations (Supplementary Figure 2). After feature detection, retention time correction, alignment, and removal of missing values, 4399 features were detected in positive ion mode and 1976 features were detected in negative ion mode (Supplementary Table 2).

The OPLS-DA model was used to calculate the variable importance on projection (VIP) values of each feature, and features with VIP values > 1 were considered the potential differential metabolites. The OPLS-DA score plot showed that the SHS group was obviously separated from the control group, indicating that the levels of metabolites in SHS participants are different from the controls (Figure 2). The cumulative R^2Y and cumulative Q^2 values of OPLS-DA model were calculated to estimate the “goodness of fit” and the ability of prediction of the model. The OPLS-DA model yielded a cumulative R^2Y of 0.858 and a cumulative Q^2 of 0.646.

Among 6375 detected features, 225 detected features with VIP values > 1, *P* values < 0.05, FDR adjusted *P* values < 0.05, and fold change > 1.2 (or < 0.83) were selected as the candidate features for SHS (Supplementary Table 2). Then the *m/z* values and retention times of these features were used to identified potential differential metabolites, and 26 of the 255 differential features were identified successfully (Table 2). After adjusting for the confounding factors, including smoking, drinking, LDL, sleep duration, insomnia score, anxiety score, depression score, and physical activity level, 24 significantly differential metabolites between the SHS group and the control group were

finally selected as the candidate biomarkers for SHS (Table 2). Among these 24 metabolites, 12 metabolites in SHS participants were significantly higher than those in controls, whereas lower levers of 12 metabolites were found in SHS participants. A heat map was performed to visualize the relative quantities of 24 metabolites, which indicated significant differences in metabolites between SHS group and control group (Figure 3).

Pathway analysis of potential metabolic biomarkers

In order to reveal the disturbed metabolic pathways related to SHS, the pathway analysis of 24 candidate metabolites was performed using Metaboanalyst. As shown in Figure 4, these metabolites were involved in 12 metabolic pathways (Supplementary Table 3). Among these 12 metabolic pathways, three pathways with *P* values < 0.05 and impact values > 0.00 were identified as sphingolipid metabolism, taurine metabolism, and steroid hormone biosynthesis pathways, respectively. Sphinganine 1-phosphate, sphinganine, sphingosine, sphingomyelin are involved in the sphingolipid metabolism. Acetyl phosphate and glutaurine are metabolites participated in the taurine and hypotaurine metabolism. Androstenedione, progesterone, and pregnanolone are involved in the steroid hormone biosynthesis.

Correlation analysis of potential metabolic biomarkers

To investigate the potential relationships of the 24 differential metabolites, the *Spearman's* correlation coefficients between the metabolites were calculated based on the levels of metabolites (Supplementary Table 4). A correlation plot was performed to visualize the correlation coefficients of 24 metabolites in Figure 5. In terms of the metabolites in sphingolipid metabolism pathway, the level of sphinganine was positively correlated with the level of sphingosine (r = 0.975, P < 0.001), and negatively correlated with the levels of sphinganine 1-phosphate (r = -0.713, P < 0.001) and

sphingomyelin (r = -0.468, P < 0.001). No correlation was observed between the levels of acetyl phosphate and glutaurine (r = -0.026, P = 0.794) in the taurine and hypotaurine metabolism pathway. The level of androstenedione was positively correlated with the level of progesterone (r = 0.856, P < 0.001), whereas a negative correlation was found between the levels of androstenedione and pregnanolone (r = -0.323, P = 0.001) in the steroid hormone biosynthesis pathway.

Protein-metabolite-disease network of potential metabolic biomarkers

In order to provide a comprehensive view of potential functional relationships between potential metabolic biomarkers, proteins, and diseases, a protein-metabolite-disease interaction network was conducted based on the information searched from HMDB and STITCH databases. Finally, 13 candidate metabolites associated with 148 proteins which were correlatied ? with 17 diseases pheno types (Figure 6). Progesterone and androstenedione in the steroid hormone biosynthesis pathway are associated with three diseases, including adrenal hyperplasia, adrenal insufficiency, and schizophrenia. Pregnanolone and progesterone are associated with major depressive disorder. Androstenedione is associated with rheumatoid arthritis and polycystic ovary syndrome 1.

Metabolite profiling as an objective diagnostic test for SHS

To improve the classification accuracy of candidate metabolites for SHS, a logistic regression analysis was performed to construct a classification model based on 24 significantly differential candidate metabolites. Metabolomics dataset was standardized before logistic regression analysis and the step-wise method was used to select the best classification model. The classification model based on candidate metabolites was as follows:

Logit (p = SHS) = -2.536 - 6.638 × (Sphingosine) + 1.156 × (Pregnanolone) + 1.929 × (Taurolithocholate sulfate) + 1.432 × (Cervonyl carnitine)

ROC curve and the AUC were used to assess the diagnostic performance of the model. The model yielded a sensitivity of 94.0%, a specificity of 90.0%, and an AUC of 0.977 (95% Confidence Interval (CI): 0.955-0.998, $P < 0.001$) (Figure 7).

Discussion

Previous studies showed that untargeted metabolomics based on LC-MS is a promising high-throughput approach for biomarker discovery. In this study, we described a comprehensive matabolomic feature of SHS and the metabolic phenotypes revealed significant differences between individuals with SHS and individuals with ideal health status. With 24 significantly altered metabolites in plasma samples, the pathway analysis suggested that three pathways (?????? Speciciy ???) were disturbed in individuals with SHS. ROC curve analysis showed that combination of metabolic biomarkers can distinguish individuals with SHS from individuals with ideal health status with a sensitivity of 94.0%, a specificity of 90.0%, and an AUC of 0.977. To the best of our knowledge, this study is the first metabolomics study of SHS.

It is believed that SHS is an early stage of chronic disease >?, and reliable methods for the identification of the individuals with SHS are of great importance for PPPM of chronic disease (3). In the present study, SHSQ-25 was applied in a population of 2861 participants to assess their SHS scores at baseline. After follow-up for 3 months, we assessed their SHS scores for the second time. The aim of this design was to minimize the influence of subjective factors on SHS scores and to give each participant an SHS score at a longgidutial time frame. On the other hand, the 3-month follow-up also can define an individual who has been a state of prolonged, profound suboptimal health. The potential confounding factors, including age, sex, smoking, drinking, level of LDL, sleep duration, level of physical activity, insomnia, anxiety, and depression, were also investigated and adjusted in this study.

Sphingolipids are a class of bioactive lipids, which are key modulators of several physiologic and pathophysiologic processes, such as cell cycle, apoptosis, stress and inflammatory responses (35). Sphingolipid metabolism pathway is highly interconnected and branched, and the misregulation of one sphingolipid enzyme may lead to accumulation or depletion of one or more species of sphingolipids (36),. A umique (abnormal?) sphingolipid profile may result in a pathological condition triggered by accumulation of sphingolipids (37). In this study, the levels of sphingomyelin and sphinganine 1-phosphate in individuals with SHS were significantly higher than those in healthy controls, whereas lower levels of sphingosine and sphinganine were found in individuals with SHS. Our findings indicate that one or more sphingolipid enzymes may be misregulated in individuals with SHS. Sphingomyelin is one of the main structural components of biological membranes (38), and the overexpression of sphingomyelin synthase increases cholesterol accumulation (39). Sphingomyelin plays an essential role in the formation of atherosclerotic lesions, and it is a risk factor for subclinical atherosclerosis and coronary artery disease (40, 41). Furthermore, our previous studies showed that SHS is associated with cardiovascular disease (5, 6). In the present study, the level of sphingomyelin is significantly higher in individuals with SHS, which strongly supports our idea that SHS is the subclinical stage of cardiovascular disease.

Steroid hormones are the steroid that acts as hormones, and steroid hormones biosynthesis is the multi-step enzymatic conversion of cholesterol via intermediate steroid precursor into biological active steroid hormones (42). Steroid hormones are the key regulators of a diverse array of physiological processes, including the maintenance of carbohydrate metabolism, sodium and fluid homeostasis, development of secondary sex characteristics, and reproduction (43). This study observed lower levels of androstenedione and progesterone in individuals with SHS compared with individuals with ideal health status, whereas higher level of pregnanolone has been found in

individuals with SHS. Our previous study also found that the level of cortisol, a steroid hormone, is elevated in plasma of individuals with SHS (4). These findings also indicate that hormonal disorders might play an important role in the pathophysiology of SHS.

Taurine is one of the most abundant amino acids in the body and it can be synthesized from other amino acids such as methionine and cysteine (44). Taurine is involved in a wide range of biological processes, including antioxidant action, anti-inflammatory effect, glucose homeostasis and osmoregulation (45). In this study, acetyl phosphate and glutaurine, which involved in the taurine and hypotaurine metabolism, were identified as the metabolic biomarkers for SHS. The level of acetyl phosphate in participants with SHS is significantly lower than those in participants with ideal health status, whereas higher level of glutaurine was found in participants with SHS. These data proposed that taurine metabolism disorder is involved in the pathophysiology of SHS.

The metabolite-protein interaction network enables exploration and visualization of interactions between functionally related metabolites and proteins. Based on the reactions from pathway databases, literature associations, similar structures, and similar activities, the interactions between metabolites and proteins were extracted from STITCH database (33). This network visualization can be used to gain novel insights into pathophysiology of SHS or assist with the development of new hypotheses for SHS research. According to the association obtained from HMDB database, metabolite-disease interaction network was also conducted to explore the disease-related metabolites. Several digestive system diseases, including cholestasis (46), Crohn's disease (47) and bile acid synthesis defect (48), were found to be associated with bilirubin metabolism disorder. In this study, increased level of bilirubin was observed in the individuals with SHS, which indicates that bilirubin metabolism disorder might associate with the SHS phonetype of the digestive system symptoms, such as nausea and poor appetite. Previous studies reported that several mental disorders, including major depressive disorder (49) and schizophrenia (50), were associated with

progesterone metabolism disorder. The higher level of progesterone in individuals with SHS indicated that progesterone metabolism disorder is associated with that mental symptom of SHS. The metabolite-disease interaction network might assist with the discovery of functional connections between SHS and diseases in this network.

Several limitations in this study are noteworthy. First, because our study is a case-control study with a relatively small sample size, the universality of these findings could be questioned. However, the longitudial design ……. In addition, considering the semi-quantitative nature of the untargeted metabolomics method, a targeted metabolomics study is underway against the same cohort to validate the putative biomarkers and pathways based on the findings in this study. Finally, although we attempted to control as many potential confounders as possible, the dietary information ? was not collected due to the complexity of dietary survey. Despite the limitations, this study has provided a new idea that plasma metabolomics might offer a novel alternative for the recognition of SHS. Building on the present findings, further studies of larger cohorts from diverse geographical areas and populations with different age ranges are warranted.

Conclusions

Individuals with SHS may involve a universal metabolic disturbance. SHS has a biosignature that can be identified using untargeted metabolomics technique. Three metabolic pathway disturbances are related to SHS and 24 significantly differential metabolites could be used as the candidate biomarkers for SHS. The combination of metabolic biomarkers offered excellent diagnostic performance for distinguishing individuals with SHS from individuals with ideal health status. The finding of an objective biosignature in SHS will give us better understanding of the etiology and pathophysiology of SHS, and help clinicians recognize individuals with higher risk of chronic disease from the perspective of predictive, preventive and personalized medicine.

Author contributions

HW, YW, and WW participated in the design of the study. HW, QT, HL, DL, XZ, WC, JZ, LW and MS performed participant enrollment and collected the samples. HW, QT, XL, XW, YZ, and ZG performed the metabolomics analysis. HW and QT performed the statistical analysis and drafted the manuscript. YW and WW revised the manuscript.

Acknowledgments

The authors acknowledge the participants and their families who donated their time and effort in helping to make this study possible. This work was partially supported by National Natural Science Foundation of China (Grant Numbers: 81673247 & 81773527) and China- Australia International Collaborative Grant (NSFC 81561128020, NHMRC APP1112767). HW and XW were supported by the China Scholarship Council (CSC 201708110200 and CSC 201608230108).

Conflicts of interest

The authors declare that they have no conflicts of interest.

Abbreviations Used

SHS = suboptimal health status

T2DM = type 2 diabetes mellitus

SHSQ-25 = suboptimal health status questionnaire-25

NMR = nuclear magnetic resonance

GC-MS = gas chromatography-mass spectrometry

LC-MS = liquid chromatography-mass spectrometry

WC = waist circumference

HC = hip circumference

BMI = body mass index

WHR = Waist-to-hip ratio

SBP = systolic blood pressure

DBP = diastolic blood pressure

FPG = fasting plasma glucose

TC = total cholesterol

TG = triglycerides

HDL = high-density lipoprotein cholesterol

LDL = low-density lipoprotein cholesterol

UA = uric acid

BUN = blood urea nitrogen

ALT = alanine transaminase

AST = aspartate transaminase

CK-MB = creatine kinase-MB

LDH = lactate dehydrogenase

ALP = alkaline phosphatase

IPAQ = International Physical Activity Questionnaire

QC = quality control

m/z = mass-to-change ratio

SD = standard deviation

IQR = interquartile ranges

FDR = false discovery rate

PCA = principal component analysis

OPLS-DA = orthogonal partial least squares projection-discriminant analysis

KEGG = Kyoto Encyclopedia of Genes and Genomes

HMDB = Human Metabolome Database

ROC = Receiver operating characteristic

AUC = area under the curve

VIP = variable importance on projection

CI = confidence interval

References

1. Wang W, Yan Y. Suboptimal health: a new health dimension for translational medicine. Clin Transl Med. 2012;1(1):28.

2. Yan YX, Liu YQ, Li M, Hu PF, Guo AM, Yang XH, et al. Development and evaluation of a questionnaire for measuring suboptimal health status in urban Chinese. J Epidemiol. 2009;19(6):333-41.

3. Wang W, Russell A, Yan Y, Global Health Epidemiology Reference G. Traditional Chinese medicine and new concepts of predictive, preventive and personalized medicine in diagnosis and treatment of suboptimal health. EPMA J. 2014;5(1):4.

4. Yan YX, Dong J, Liu YQ, Zhang J, Song MS, He Y, et al. Association of suboptimal health status with psychosocial stress, plasma cortisol and mRNA expression of glucocorticoid receptor alpha/beta in lymphocyte. Stress. 2015;18(1):29-34.

5. Yan YX, Dong J, Liu YQ, Yang XH, Li M, Shia G, et al. Association of suboptimal health status and cardiovascular risk factors in urban Chinese workers. J Urban Health. 2012;89(2):329-38.

6. Wang Y, Liu X, Qiu J, Wang H, Liu D, Zhao Z, et al. Association between Ideal Cardiovascular Health Metrics and Suboptimal Health Status in Chinese Population. Sci Rep. 2017;7(1):14975.

7. Ge S, Xu X, Zhang J, Hou H, Wang H, Liu D, et al. Suboptimal health status as an independent risk factor for type 2 diabetes mellitus in a community-based cohort: the China suboptimal health cohort study. EPMA Journal. 2019:1-8.

8. Adua E, Roberts P, Wang W. Incorporation of suboptimal health status as a potential risk assessment for type II diabetes mellitus: a case-control study in a Ghanaian population. EPMA J. 2017;8(4):345-55.

9. Kupaev V, Borisov O, Marutina E, Yan YX, Wang W. Integration of suboptimal health status and endothelial dysfunction as a new aspect for risk evaluation of cardiovascular disease. EPMA J. 2016;7(1):19.

10. Wang Y, Ge S, Yan Y, Wang A, Zhao Z, Yu X, et al. China suboptimal health cohort study: rationale, design and baseline characteristics. J Transl Med. 2016;14(1):291.

11. Yan YX, Dong J, Li M, Yang S, Wang W. Establish the Cut off Point for Suboptimal Health Status Using SHSQ-25. Chinese Journal of Health Statistics. 2011;28(3):256-8. (In Chinese).

12. Yan YX, Wu LJ, Xiao HB, Wang S, Dong J, Wang W. Latent class analysis to evaluate performance of plasma cortisol, plasma catecholamines, and SHSQ-25 for early recognition of suboptimal health status. EPMA J. 2018;9(3):299-305.

13. Hatzimanikatis V, Li C, Ionita JA, Broadbelt LJ. Metabolic networks: enzyme function and metabolite structure. Curr Opin Struct Biol. 2004;14(3):300-6.

14. Patti GJ, Yanes O, Siuzdak G. Innovation: Metabolomics: the apogee of the omics trilogy. Nat Rev Mol Cell Biol. 2012;13(4):263-9.

15. Beger RD, Dunn W, Schmidt MA, Gross SS, Kirwan JA, Cascante M, et al. Metabolomics enables precision medicine: "A White Paper, Community Perspective". Metabolomics. 2016;12(10):149.

16. Fiehn O. Metabolomics - the link between genotypes and phenotypes. Plant MolBiol. 2002;48(1-2):155-71.

17. Wishart DS. Emerging applications of metabolomics in drug discovery and precision medicine. Nat Rev Drug Discov.

2016;15(7):473-84.

18. Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. Nat Rev Mol Cell Biol. 2016;17(7):451-9.

19. Jacob M, Lopata AL, Dasouki M, Abdel Rahman AM. Metabolomics toward personalized medicine. Mass Spectrom Rev. 2019;38(3):221-38.

20. Pinu FR, Beale DJ, Paten AM, Kouremenos K, Swarup S, Schirra HJ, et al. Systems Biology and Multi-Omics Integration: Viewpoints from the Metabolomics Research Community. Metabolites. 2019;9(4).

21. 2015–2020 Dietary Guidelines for Americans. 8 ed: U.S. Department of Health and Human Services; 2015.

22. Craig CL, Marshall AL, Sjostrom M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc. 2003;35(8):1381-95.

23. Soldatos CR, Dikeos DG, Paparrigopoulos TJ. Athens Insomnia Scale: validation of an instrument based on ICD-10 criteria. J Psychosom Res. 2000;48(6):555-60.

24. Shear MK, Vander Bilt J, Rucci P, Endicott J, Lydiard B, Otto MW, et al. Reliability and validity of a structured interview guide for the Hamilton Anxiety Rating Scale (SIGH-A). Depress Anxiety. 2001;13(4):166-78.

25. Bagby RM, Ryder AG, Schuller DR, Marshall MB. The Hamilton Depression Rating Scale: has the gold standard become a lead weight? Am J Psychiatry. 2004;161(12):2163-77.

26. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, et al. A cross-platform toolkit for mass spectrometry and proteomics. Nat Biotechnol. 2012;30(10):918-20.

27. Forsberg EM, Huan T, Rinehart D, Benton HP, Warth B, Hilmers B, et al. Data processing, multi-omic pathway mapping, and metabolite activity analysis using XCMS Online. Nat Protoc. 2018;13(4):633-51.

28. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017;45(D1):D353-D61.

29. Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, Vazquez-Fresno R, et al. HMDB 4.0: the human metabolome database for 2018. Nucleic Acids Res. 2018;46(D1):D608-D17.

30. Guijas C, Montenegro-Burke JR, Domingo-Almenara X, Palermo A, Warth B, Hermann G, et al. METLIN: A Technology Platform for Identifying Knowns and Unknowns. Anal Chem. 2018;90(5):3156-64.

31. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, et al. PubChem Substance and Compound databases. Nucleic Acids Res. 2016;44(D1):D1202-13.

32. Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, et al. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. Nucleic Acids Res. 2018;46(W1):W486-W94.

33. Szklarczyk D, Santos A, von Mering C, Jensen LJ, Bork P, Kuhn M. STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data. Nucleic Acids Res. 2016;44(D1):D380-4.

34. R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna,

Austria. URL https://www.R-project.org/.

35. Huwiler A, Kolter T, Pfeilschifter J, Sandhoff K. Physiology and pathophysiology of sphingolipid metabolism and signaling. Biochim Biophys Acta. 2000;1485(2-3):63-99.

36. Hannun YA, Luberto C, Argraves KM. Enzymes of sphingolipid metabolism: from modular to integrative signaling. Biochemistry. 2001;40(16):4893-903.

37. Maceyka M, Spiegel S. Sphingolipid metabolites in inflammatory disease. Nature. 2014;510(7503):58-67.

38. Slotte JP, Ramstedt B. The functional role of sphingomyelin in cell membranes. European journal of lipid science and technology. 2007;109(10):977-81.

39. Yan N, Ding T, Dong J, Li Y, Wu M. Sphingomyelin synthase overexpression increases cholesterol accumulation and decreases cholesterol secretion in liver cells. Lipids Health Dis. 2011;10:46.

40. Nelson JC, Jiang XC, Tabas I, Tall A, Shea S. Plasma sphingomyelin and subclinical atherosclerosis: findings from the multi-ethnic study of atherosclerosis. Am J Epidemiol. 2006;163(10):903-12.

41. Schlitt A, Blankenberg S, Yan D, von Gizycki H, Buerke M, Werdan K, et al. Further evaluation of plasma sphingomyelin levels as a risk factor for coronary artery disease. Nutr Metab (Lond). 2006;3:5.

42. Ghayee HK, Auchus RJ. Basic concepts and recent developments in human steroid hormone biosynthesis. Rev Endocr Metab Disord. 2007;8(4):289-300.

43. Sewer MB, Li D. Regulation of steroid hormone biosynthesis by the cytoskeleton. Lipids. 2008;43(12):1109-15.

44. Vitvitsky V, Garg SK, Banerjee R. Taurine biosynthesis by neurons and astrocytes. J Biol Chem. 2011;286(37):32002-10.

45. De la Puerta C, Arrieta F, Balsa J, Botella-Carretero J, Zamarrón I, Vázquez C. Taurine and glucose metabolism: a review. Nutricion hospitalaria. 2010;25(6):910-9.

46. Schukfeh N, Metzelder ML, Petersen C, Reismann M, Pfister ED, Ure BM, et al. Normalization of serum bile acids after partial external biliary diversion indicates an excellent long-term outcome in children with progressive familial intrahepatic cholestasis. J Pediatr Surg. 2012;47(3):501-5.

47. Lapidus A, Akerlund JE, Einarsson C. Gallbladder bile composition in patients with Crohn 's disease. World J Gastroenterol. 2006;12(1):70-4.

48. Huang H-Y, Zhou H, Wang H, Chen Y-X, Fang F. Novel Mutations in the 3β-hydroxy-Δ5-C27-steroid Dehydrogenase Gene (HSD3B7) in a Patient with Neonatal Cholestasis. Chinese medical journal. 2016;129(1):98.

49. Nieminen LR, Makino KK, Mehta N, Virkkunen M, Kim HY, Hibbeln JR. Relationship between omega-3 fatty acids and plasma neuroactive steroids in alcoholism, depression and controls. Prostaglandins Leukot Essent Fatty Acids. 2006;75(4-5):309-14.

50. Bicikova M, Hill M, Ripova D, Mohr P, Hampl R. Determination of steroid metabolome as a possible tool for laboratory diagnosis of schizophrenia. J Steroid Biochem Mol Biol. 2013;133:77-83.

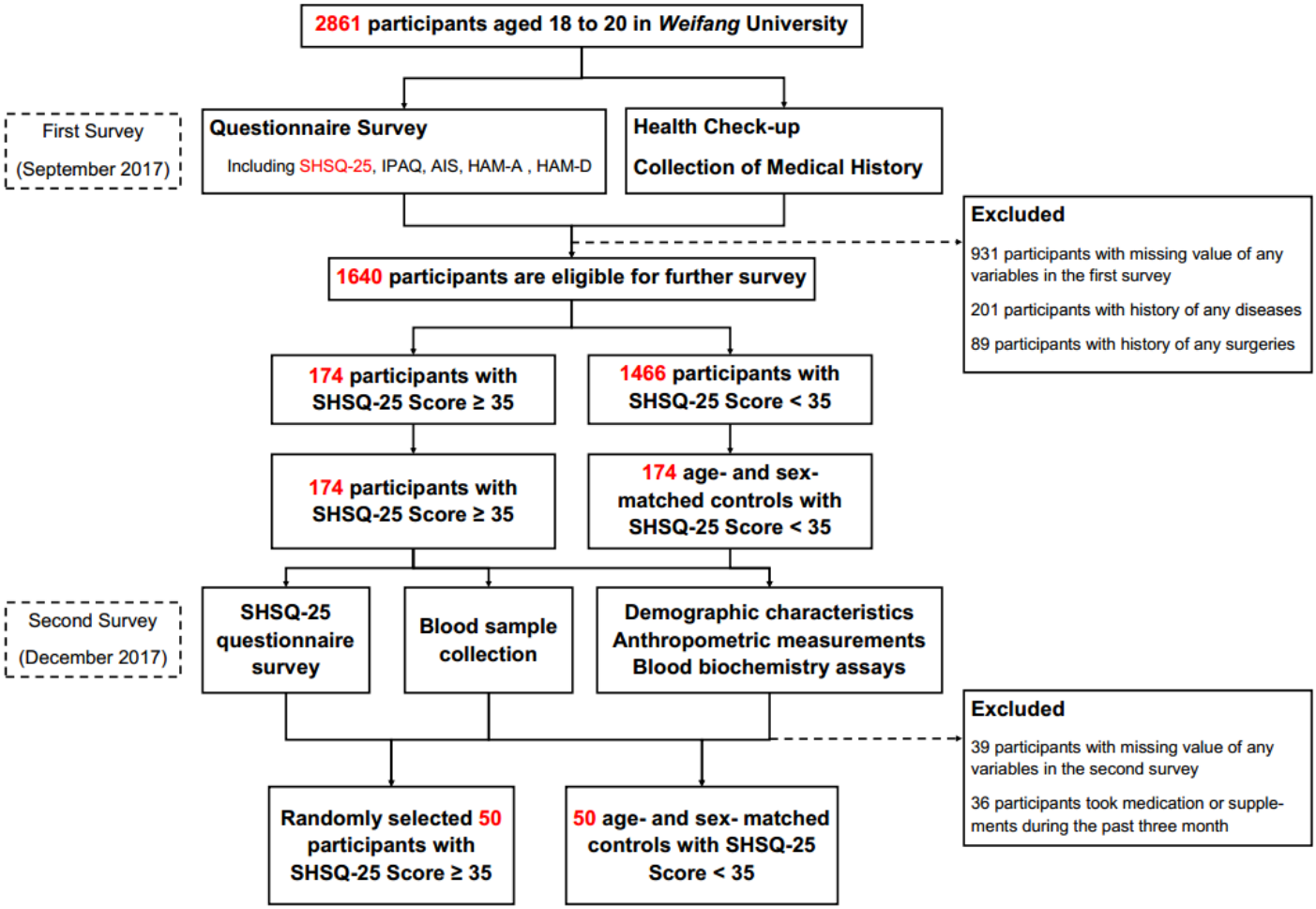

Figure 1. Study population flow chart

Table 1. Characteristics of the study participants

Characteristics	Control Group	SHS Group	*P* value
N	50.00	50.00	
Sex			1.000
Male	10.00 (20.00%)	10.00 (20.00%)	
Female	40.00 (80.00%)	40.00 (80.00%)	
Age (years)	19.00 (19.00 - 19.00)	19.00 (18.00 - 20.00)	0.888
Smoking			1.000
Non-smoker	47.00 (94.00%)	47.00 (94.00%)	
Smokers	3.00 (6.00%)	3.00 (6.00%)	
Drinking			0.695
Non-drinker	46.00 (92.00%)	44.00 (88.00%)	
Light to moderate drinker	1.00 (2.00%)	3.00 (6.00%)	
Heavy drinker	3.00 (6.00%)	3.00 (6.00%)	
Dietary			1.000
Omnivore	50.00 (100.00%)	50.00 (100.00%)	
Vegetarian	0.00 (0.00%)	0.00 (0.00%)	
SBP (mmHg)	115.50 (109.00 - 124.50)	118.00 (111.75 - 126.25)	0.235
DBP (mmHg)	74.86 ± 8.82	76.70 ± 9.95	0.330
BMI (kg/m^2)	21.41 (19.92 - 23.40)	21.26 (19.43 - 23.07)	0.442
WC (cm)	71.01 ± 8.13	70.91 ± 8.09	0.951
HC (cm)	92.03 ± 7.80	92.45 ± 6.76	0.774
WHR	0.77 (0.74 - 0.80)	0.75 (0.72 - 0.80)	0.427
FPG (mmol/L)	4.94 (4.78 - 5.14)	4.90 (4.75 - 5.09)	0.326
TC (mmol/L)	4.22 ± 0.66	3.98 ± 0.63	0.066
TG (mmol/L)	0.90 (0.80 - 1.10)	0.90 (0.70 - 1.00)	0.403
HDL (mmol/L)	1.44 ± 0.27	1.48 ± 0.28	0.547
LDL (mmol/L)	2.40 (2.00 - 2.90)	2.20 (1.90 - 2.50)	**0.044**
UA (μmol/L)	317.50 (261.05 - 406.25)	308.85 (269.65 - 377.08)	0.788
Cre (μmol/L)	59.55 (55.50 - 69.63)	58.95 (55.88 - 65.45)	0.674
BUN (mmol/L)	3.95 (3.30 - 4.50)	4.15 (3.30 - 4.93)	0.446
ALT (U/L)	12.65 (9.23 - 19.15)	13.30 (9.18 - 18.75)	0.852
AST (U/L)	20.90 (17.90 - 24.18)	21.50 (17.78 - 25.10)	0.627
CK-MB (U/L)	6.70 (5.28 - 8.55)	6.05 (5.18 - 7.33)	0.181
LDH (U/L)	147.44 ± 24.09	143.09 ± 21.41	0.343
ALP (U/L)	59.54 (50.12 - 75.72)	58.48 (51.71 - 72.44)	0.825
Sleep duration (hours)	8.00 (7.00 - 8.00)	7.00 (7.00 - 8.00)	**0.030**
SHS score (First survey)	12.00 (8.00 - 21.25)	40.00 (37.00 - 44.00)	**<0.001**
SHS score (Second survey)	16.00 (12.50 - 25.25)	35.00 (35.00 - 40.00)	**<0.001**
Insomnia score	3.00 (1.75 - 5.25)	6.50 (6.00 - 9.00)	**<0.001**
Anxiety score	2.00 (0.00 - 4.25)	9.00 (3.00 - 13.25)	**<0.001**
Depression score	2.00 (0.00 - 3.00)	5.00 (1.75 - 10.25)	**<0.001**
Physical activity			0.165
Moderate level	9.00 (18.00%)	16.00 (32.00%)	

High level	41.00 (82.00%)	34.00 (68.00%)

Note: Data are presented as means ± SD, medians (interquartile ranges) or frequencies (percentages). $P < 0.05$ is considered statistically significant. SHS, suboptimal health status; *N,* number of participants; SD, standard deviation; SBP, systolic blood pressure; DBP, diastolic blood pressure; BMI, body mass index; WC, waist circumference; HC, hip circumference; WHR, waist-to-hip ratio; FPG, fasting plasma glucose; TC, total cholesterol; TG, total triglycerides; HDL, high-density lipoprotein; LDL, low-density lipoprotein; UA, uric acid, Cre, creatinine; BUN, blood urea nitrogen; TBIL, total bilirubin; ALT, alanine transaminase; AST, aspartate transaminase; CK-MB, creatine kinase-MB; LDH, lactate dehydrogenase; ALP, alkaline phosphatase.

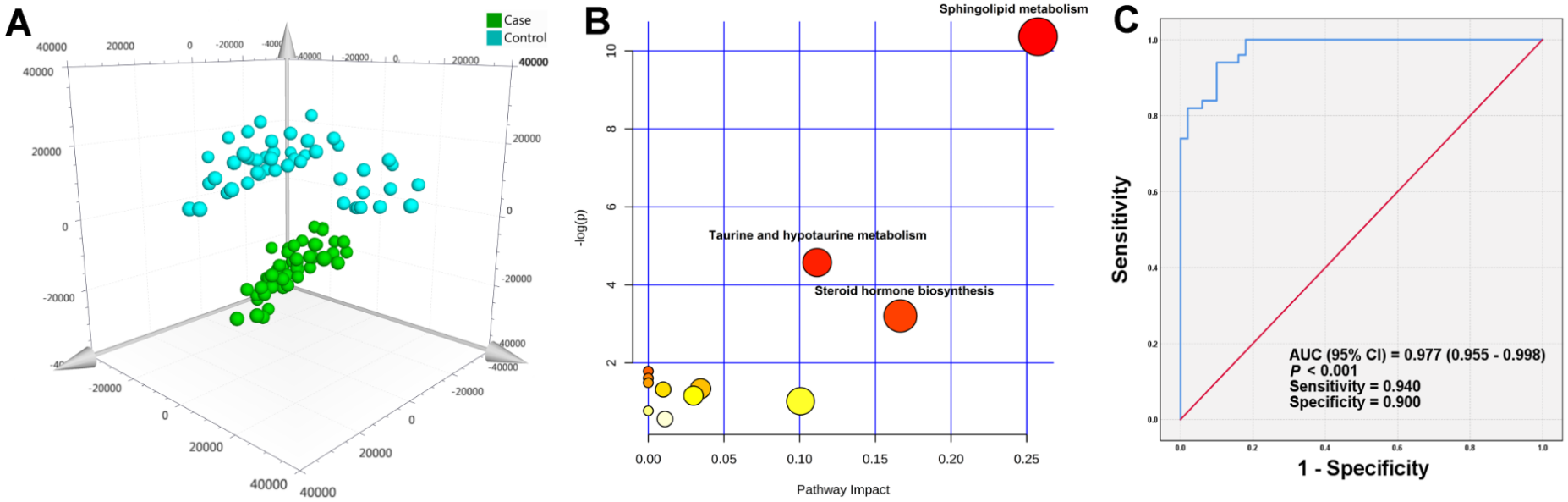

Figure 2. Metabolic biomarkers for SHS

Note: (A) The orthogonal projection to latent structure-discriminant analysis (OPLS-DS) score plots. (B) The disturbed metabolic pathways in suboptimal health status. (C) The diagnostic performance of potential metabolic metabolites for SHS.

Table 2. Differential metabolites identified between SHS participants and controls

Metabolites	Mode	Observed *m/z*	RT (min)	Trend	FC	VIP value	*P* Value	**P* Value	*#P* value
Acetyl phosphate	Positive	139.9881	15.8044	Down	0.815	3.744	0.004	0.006	0.016
Spermine	Positive	202.2167	5.0300	Down	0.528	1.282	<0.001	<0.001	<0.001
Androstenedione	Positive	219.2154	4.2883	Down	0.472	1.034	<0.001	<0.001	<0.001
Progesterone	Positive	247.2463	4.9932	Down	0.113	2.707	<0.001	<0.001	0.004
Glutaurine	Positive	255.0659	4.3251	Up	1.774	1.038	<0.001	<0.001	0.001
Glycerophosphocholine	Positive	258.1109	0.9428	Up	1.451	1.959	<0.001	<0.001	<0.001
Hexanoylcarnitine	Positive	260.1864	4.0222	Down	0.735	1.395	<0.001	0.017	0.476
L-Octanoylcarnitine	Positive	288.2178	4.6330	Down	0.739	2.524	0.005	0.008	0.397
Sphinganine 1-phosphate	Positive	298.3109	12.1612	Up	2.524	1.783	<0.001	<0.001	<0.001
Sphinganine	Positive	302.3061	6.8676	Down	0.085	15.831	<0.001	<0.001	0.042
Glyerophosphoinositol	Positive	317.0612	4.9279	Up	1.220	1.616	<0.001	<0.001	0.001
Sphingosine	Positive	318.3008	5.9321	Down	0.139	8.451	<0.001	<0.001	0.003
Pregnanolone	Positive	319.2636	11.3965	Up	1.386	3.833	<0.001	<0.001	0.001
2-Arachidonylglycerol	Positive	343.2636	11.0523	Up	1.401	3.040	<0.001	<0.001	0.001
4alpha-Methylzymosterol-4-carboxylate	Positive	375.3669	7.8852	Down	0.119	2.330	<0.001	<0.001	0.015
S-Adenosylhomocysteine	Positive	403.1396	5.6208	Down	0.065	1.119	<0.001	<0.001	0.002
Reduced Vitamin K	Positive	403.3981	8.7993	Down	0.088	1.909	<0.001	<0.001	0.046
7'-Carboxy-alpha-chromanol	Positive	416.2160	6.4400	Down	0.516	3.595	<0.001	<0.001	0.001
Taurolithocholate sulfate	Positive	480.3092	8.0890	Up	1.241	2.064	0.005	0.008	0.042
Cervonyl carnitine	Positive	494.3251	7.2723	Up	1.202	4.716	<0.001	<0.001	0.001
Hexacosanoyl carnitine	Positive	512.5042	12.2543	Down	0.399	1.353	<0.001	<0.001	0.001
3alpha,7alpha,12alpha-Trihydroxy-5beta-cholestanoate	Positive	519.3287	7.0182	Up	1.340	1.576	<0.001	<0.001	0.005
18-CoA-18-oxo-dinorleukotriene B4	Positive	542.1220	15.3080	Up	1.304	1.181	<0.001	<0.001	0.001
Sphingomyelin	Positive	703.5764	7.8690	Up	1.370	6.219	<0.001	<0.001	<0.001

9'-Carboxy-gamma-chromanol	Negative	374.2449	6.1306	Up	2.144	2.176	<0.001	<0.001	<0.001
Bilirubin	Negative	583.2558	4.9730	Down	0.723	1.769	0.001	0.001	0.032

Note: *m/z*, mass-to charge ratio; RT, retention time; FC, fold chnage; VIP, variable importance on projection. Down trend, relatively lower levels of metabolites present in SHS group. Up trend, relatively higher levels of metabolites present in SHS group.

P Value, *P* value from Student *t*-test without adjustment; **P* Value, *P* value adjusted for false discovery rate using Benjamini-Hochberg method. $^{\#}P$ Value, *P* value adjusted for smoking, drinking, low-density lipoprotein cholesterol, sleep duration, insomnia score, anxiety score, depression scores, physical activity and false discovery rate; $P < 0.05$ is considered statistically significant.

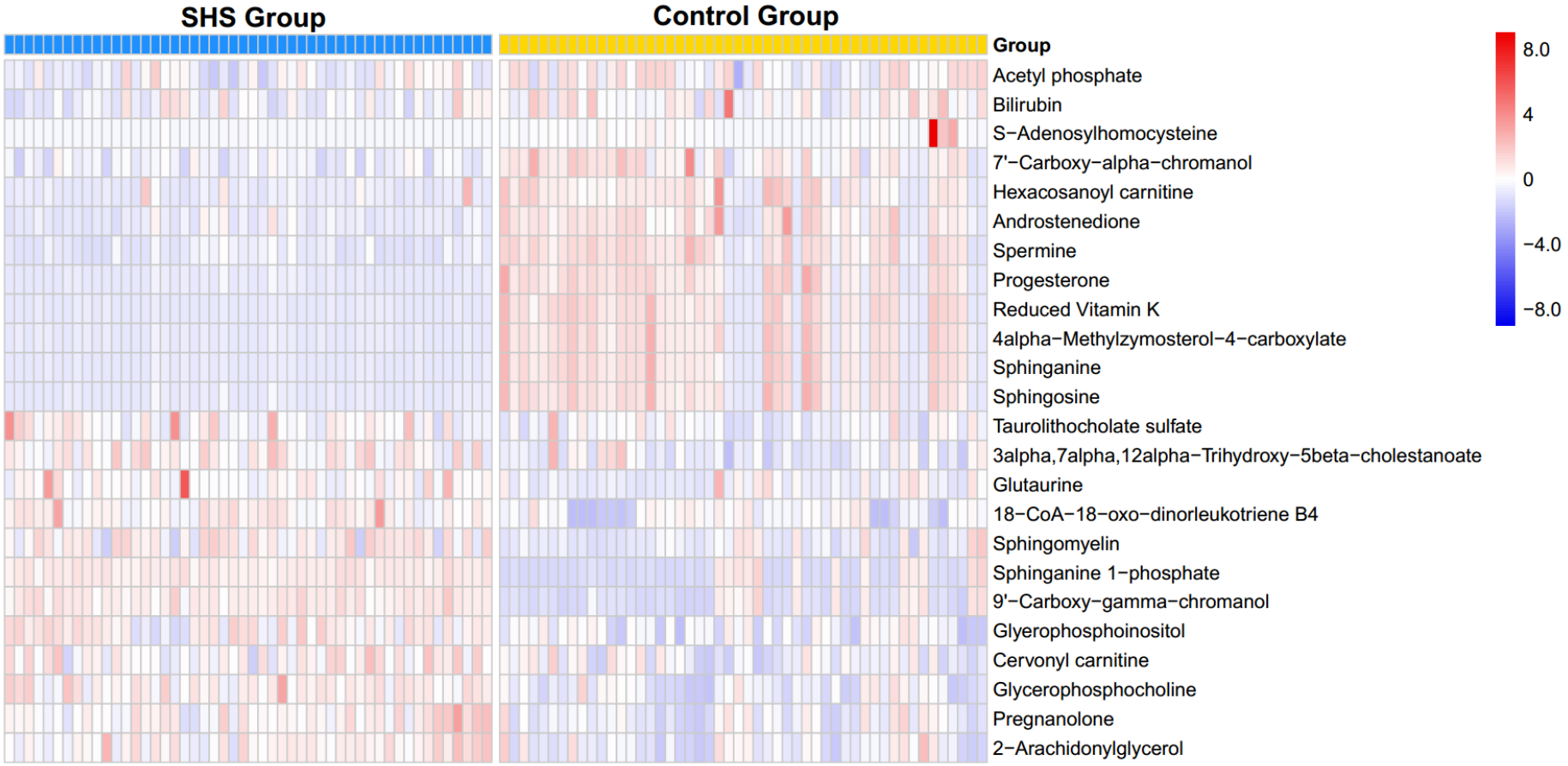

Figure 3. The Heat map of 24 potential metabolic biomarkers for SHS

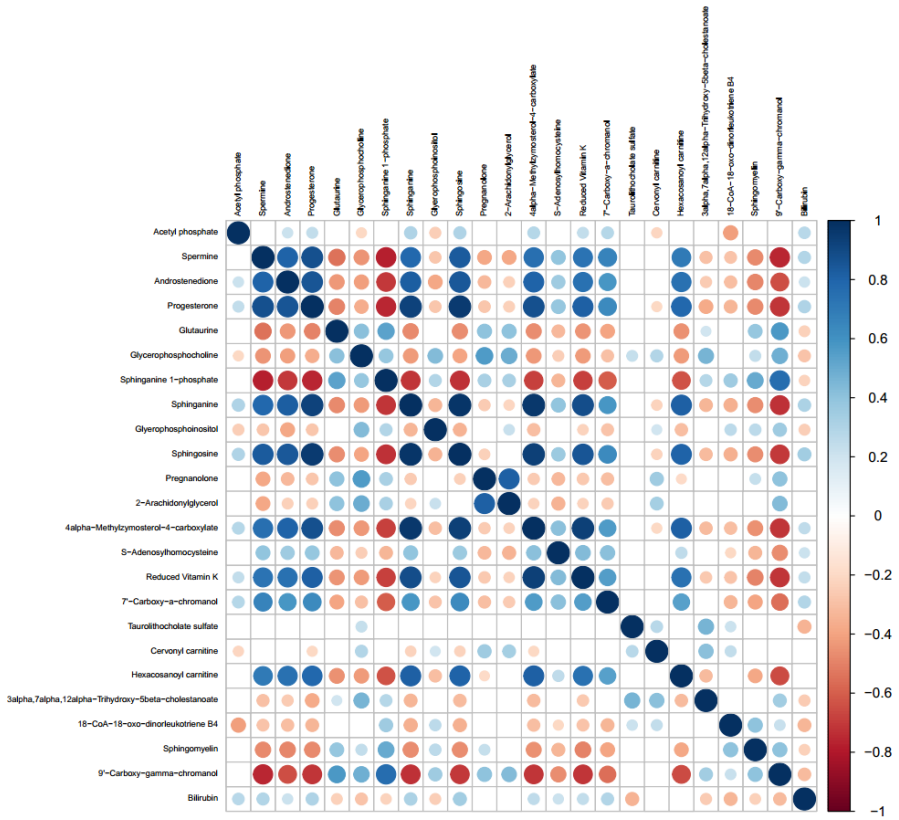

Figure 4. Correlation coefficients of 24 potential metabolic biomarkers for SHS

Note: Statistically significant correlations between two metabolites are shown, while the insignificant correlation coefficients are blank in the boxes. The positive correlations are represented by blue color, while negative correlations are represented by red color. $P < 0.05$ is considered statistically significant. The detailed correlation coefficients and *P* values were shown in Supplementary Table 4.

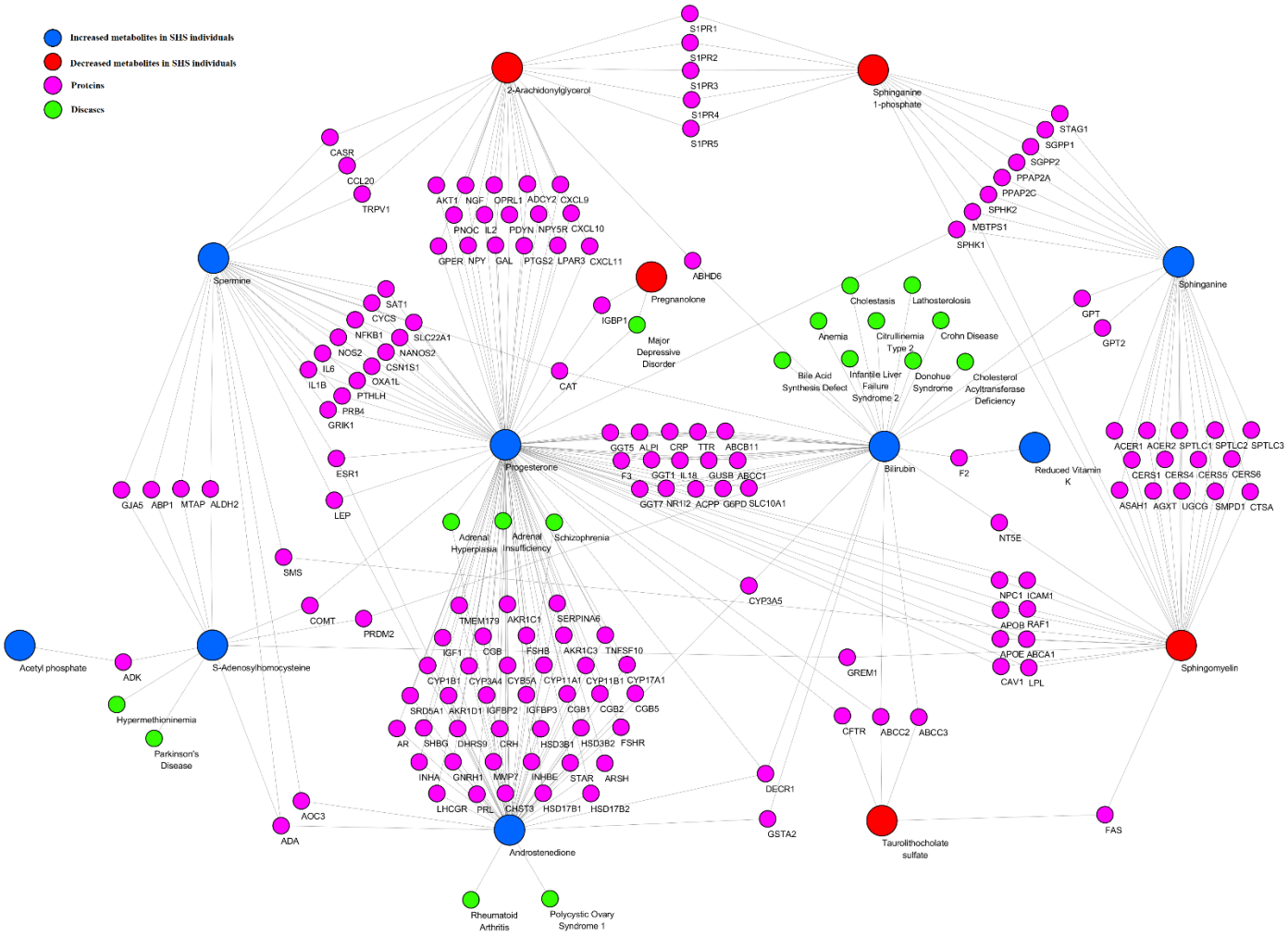

Figure 5. The metabolite-gene-disease network analysis of potential metabolites for SHS