4-20-2021

Blood transcriptome profiling as potential biomarkers of suboptimal health status: Potential utility of novel biomarkers for predictive, preventive, and personalized medicine strategy

Hao Wang
Edith Cowan University

Qiuyue Tian

Jie Zhang

Hongqi Liu

Jinxia Zhang

See next page for additional authors

10.1007/s13167-021-00238-1
This is a post-peer-review, pre-copyedit version of an article published in EPMA Journal. The final authenticated version is available online at: http://dx.doi.org/10.1007/s13167-021-00238-1
Wang, H., Tian, Q., Zhang, J., Liu, H., Zhang, J., Cao, W., ... Wang, Y. (2021). Blood transcriptome profiling as potential biomarkers of suboptimal health status: Potential utility of novel biomarkers for predictive, preventive, and personalized medicine strategy. *EPMA Journal, 12*(2), 103-115.
https://doi.org/10.1007/s13167-021-00238-1

Blood Transcriptome Profiling as Potential Biomarkers of Suboptimal Health Status: Potential Utility of Novel Biomarkers for Predictive, Preventive, and Personalized Medicine Strategy

Hao Wang[1,2,3,*], Qiuyue Tian[2,*], Jie Zhang[2], Hongqi Liu[4], Jinxia Zhang[2], Weijie Cao[2,3], Xiaoyu Zhang[2,5], Xingang Li[3], Lijuan Wu[2], Manshu Song[3], Yuanyuan Kong[1], Wei Wang[3], and Youxin Wang[2]

[1]Department of Clinical Epidemiology and Evidence-Based Medicine, National Clinical Research Center for Digestive Disease, Beijing Friendship Hospital, Capital Medical University, Beijing, China.

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

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

[4]Student Healthcare Center, Weifang University, Weifang, China

[5]Department of Anesthesiology, Sanbo Brain Hospital, Capital Medical University, Beijing, China.

*These authors contributed equally to this work.

Short title: Transcriptomic biomarkers for Suboptimal Health Status

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

Abstract

The early identification of Suboptimal Health Status (SHS) creates a window opportunity for the predictive, preventive, and personalized medicine (PPPM) in chronic diseases. Previous studies have observed the alterations in several mRNA levels in SHS individuals. As a promising “omics” technology offering comprehension of genome structure and function at RNA level, transcriptome profiling can provide innovative molecular biomarkers for the predictive identification and targeted prevention of SHS. To explore the potential biomarkers, biological functions, and signaling pathways involved in SHS, an RNA sequencing (RNA-Seq)-based transcriptome analysis was firstly conducted on buffy coat samples collected from 30 participants with SHS and 30 age- and sex-matched healthy controls. Transcriptome analysis identified a total of 46 differentially expressed genes (DEGs), in which 22 transcripts were significantly increased and 24 transcripts were decreased in the SHS group. A total of 23 transcripts was selected as candidate predictive biomarkers for SHS. Gene Ontology (GO) annotations and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis revealed that several biological processes were related to SHS, such as ATP-binding cassette (ABC) transporter and neurodegeneration. Protein-protein interaction (PPI) network analysis identified 10 hub gene related to SHS, including *GJA1*, *TWIST2*, *KRT1*, *TUBB3*, *AMHR2*, *BMP10*, *MT3*, *BMPER*, *NTM* and *TMEM98*. A transcriptome predictive model can distinguish SHS individuals from the healthy controls with a sensitivity of 83.3% (95% Confidence Interval (CI): 73.9%-92.7%), a specificity of 90.0% (95% CI: 82.4%-97.6%), and an

area under the receiver operating characteristic curve of 0.938 (95% CI: 0.882-0.994). In the present study, we have proved that blood (buffy coat) samples appear to be a very promising and easily accessible biological material for the transcriptomic analyses focused on the objective identification of SHS by using our transcriptome predictive model. The pattern of particularly determined DEGs can be used as predictive transcriptomic biomarkers for the identification of SHS in an individual who may, subjectively, feel healthy but at the level of subcellular mechanisms the changes can provide an early information about potential health problems in this person. Our findings also indicate the potential therapeutic targets in dealing with chronic diseases related to SHS, such as T2DM and CVD, and an early onset of neurodegenerative diseases, such as Alzheimer's and Parkinson's diseases, as well as the findings suggest the targets for personalized interventions as promoted in PPPM.

Keywords Predictive preventive personalized medicine, Transcriptome profiling, Suboptimal health status, Predictive biomarkers, Novel transcriptomic biomarkers, Transcriptome predictive model, ATP-binding cassette transporter, Neurodegeneration, Glucokinase regulator

Introduction

Suboptimal health status - a window opportunity for the PPPM in chronic diseases

Predictive, preventive, and personalized medicine (PPPM) is a holistic strategy in healthcare that aims to predict individual predisposition, to provide targeted prevention, and to create personalized treatment [1]. Chronic diseases are usually treated after disease onset, which is a very much delayed approach from the perspective of PPPM [2]. Suboptimal health status (SHS) is an intermediate physical state between ideal health and disease, which fits within the paradigm of PPPM [3]. Several studies have suggested that SHS might precede the occurrence of chronic diseases, including type 2 diabetes mellitus (T2DM) [4, 5] and cardiovascular diseases (CVD) [6, 7]. As a subclinical stage of chronic diseases, the early identification of SHS plays a significant role in the targeted prevention and personalized treatment of chronic diseases from the perspective of PPPM [8].

Suboptimal health status questionnaire-25 (SHSQ-25), the most widely used screening tool for SHS, has been developed based on the perceived health complaints and physical symptoms affected by SHS [9]. The SHSQ-25 includes 25 items constituting five dimensions: the immune system, the cardiovascular system, the digestive tract, fatigue, and mental status [9], and it has been used in Caucasian [10], African [4, 11], and Asian populations [6, 12]. Screening multi-omics biomarkers for the early identification of SHS have a great potential to implement the advanced paradigm of PPPM [13]. Several multi-omics biomarkers, including cortisol [14],

relative telomere length [15], intestinal microbiota [16], and metabolites [17], have been investigated. However, the transcriptomic biomarkers for the identification of SHS has not been investigated.

Transcriptome – a promising "omics" technology provides novel biomarkers for SHS identification

The transcriptome is a promising "omics" technology involving the identification and quantification of the complete set of transcripts in a specific tissue or cell type. It contains the full information about all RNA transcribed by the genome at a particular developmental stage, and under a certain physiological or pathological condition [18]. Transcriptome analysis not only provides us a better understanding of the human genome at the transcription level, but also gives a comprehensive perspective of genome structure and function [19]. Moreover, it may reveal the key alterations of biological processes that respond to pathogens, diseases, and environmental challenges, thus offering novel predictive biomarkers, useful not only for the comprehension of their underlying mechanisms but also for their predictive diagnosis and targeted prevention [20].

Changes in transcriptome might be involved in the underlying mechanism of SHS

Our previous study observed the association between chronic psychosocial stress and SHS, as well as it found that decreased mRNA expression of glucocorticoid

receptor α is associated with a high level of SHS [21]. The adrenaline and cortisol could impact glucocorticoid receptor expression and splicing [14, 21]. In addition, changes in the transcriptome of circulating immune cells were also observed in patients who suffered from chronic fatigue syndrome, a disease resembling SHS [22]. The biological pathways, including mitochondrial function, oxidative stress, and chronic inflammation are involved in the pathophysiologic mechanism of chronic fatigue syndrome [22]. Tomas-Roig et al. found that changes in mRNA expression, such as mRNA level of dopamine receptor, are associated with long-term psychosocial stress [23]. The association between chronic psychosocial stress and SHS, together with the alterations of mRNA expression in individuals with chronic psychosocial stress, leads to the hypothesis of this study that the alteration of transcriptome profiling might occur in SHS participants and changes in gene expression might be involved in the underlying mechanism of SHS.

Study Objectives

This study aimed to firstly describe comprehensive transcriptomic biosignature for SHS, and objectively screen predictive biomarkers for SHS, using RNA-Seq-based transcriptome profiling. In addition, the Gene Ontology (GO) annotation, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis and protein-protein interaction (PPI) network analysis of the potential differentially expressed genes (DEGs) were also conducted, so that further understanding of the biological processes involved in SHS and SHS-related chronic diseases was obtained,

which proves to be useful for the PPPM attitude towards chronic diseases.

Materials and methods

Study design and participants

From September 2017 to November 2017, a case-control study was conducted among a Chinese Han population who received annual health examination at the Student Healthcare Centre of *Weifang* University. In order to minimize the influence of confounding factors, all the participants were undergraduate students aged 18 to 20 years who were living in the university campus apartments. The inclusion and exclusion criteria were described previously in our study [17]. All the participants were required to complete SHSQ-25 in September 2017. After follow-up for 3 months, we assessed their SHS scores for the second time. 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, 30 SHS participants and 30 age- and sex-matched controls were included in the current study.

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.

Data collection

Demographic characteristics and lifestyle information of participants, including age, sex, ethnicity, smoking, drinking dietary, and sleep duration were collected by questionnaires. The anthropometric measurements, routine biochemical tests, and clinical characteristics were measured as described in our previous study [17]. The physical activity levels, insomnia, anxiety, and depression were measured using the International Physical Activity Questionnaire [24], the Athens insomnia scale [25] the Hamilton anxiety rating scale [26], and the Hamilton depression rating scale [27], respectively.

Blood sample collection and RNA extraction

Two tubes of blood samples (5 ml) were collected from each participant by venipuncture in the morning after an overnight fasting. One tube of blood sample was collected using a vacuum tube containing polymeric gel to acquire serum, which was used for routine biochemical tests. The other blood sample was collected for RNA extraction. Total RNA was extracted from buffy coat samples of 30 SHS individuals and 30 healthy controls using TRIzol reagent (ThermoFisher Scientific, Waltham, USA). The concentration and purity of extracted RNA were measured using Nanodrop 2000 (ThermoFisher Scientific, Waltham, USA). The quality of RNA was assessed using Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, USA). Then, RNA samples were stored at -80°C until RNA-Seq library preparation.

RNA-Seq library preparation and sequencing

Total RNA samples were quantified using Qubit 2.0 Fluorometer (ThermoFisher Scientific, Waltham, USA). Then, 100 ng of RNA from each sample was used for library preparation using NEBNext Ultra™ RNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, USA) according to the manufacturer's protocols. Briefly, the mRNA was purified using magnetic beads, and isolated mRNA was reversely transcribed to double-stranded complementary DNA (cDNA). Then, the cDNA libraries were denatured as single-stranded DNA molecules, captured on Illumina flow cells, amplified as clusters, and finally sequenced using an Illumina HiSeq Xten Sequencer (Illumina, San Diego, USA) for 150 bp paired-end reads.

Statistical analysis

Raw FASTQ files were filtered to remove the adaptors, polyN reads, and low-quality bases using FastQC software (Babraham Institute, Cambridge, UK). Clean reads were mapped and aligned to the human genome using HISAT2 program, and mapped reads were counted using featureCounts software [28]. Normality distribution of all variables was 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 the *Chi-square* test or *Fisher's* exact test. The differences in continuous variables between the two groups were tested by the *Student t*-test or *Mann-Whitney*

U test. To identify DEGs related to SHS, mRNA expression profile data in read counts were analysed using the "*DESeq*2" package in R [29]. The *Benjamini-Hochberg* method was used to adjust the false discovery rate (FDR). In order to identify the biological functions and pathways in which DEGs were enriched, GO and KEGG enrichment analyses were performed using the "*clusterProfiler*" package in R [30]. To further explore the interaction among the DEGs, a PPI network of DEGs was constructed using the Search Tool for the Retrieval of Interacting Genes (STRING) database [31]. The top 10 genes ranked by degree were selected as hub genes using the cytoHubba application [32], while the Cytoscape software 3.8.1 (National Institute of General Medical Sciences, Bethesda, USA) was used to visualize the PPI network. Multivariate binary logistic regression was used to construct transcriptome diagnostic models for SHS. Receiver operating characteristic (ROC) curves and the area under the curve (AUC) were used to assess the diagnostic performance of the models.

Data analysis was performed using SPSS 25.0 (IBM Corporation, New York, USA) and R 4.0.2 [33]. All reported *P* values were two-tailed, and $P < 0.05$ was considered statistically significant.

Results

Characteristics of participants

The average ages of SHS and control groups were 19.03 and 18.87 years, respectively. Higher SHS, insomnia, anxiety, and depression scores were observed in the SHS group (All $P < 0.05$). Aside from these, there were no statistically significant

differences observed between the SHS and control groups in terms of the other variables (All $P > 0.05$). The characteristics of participants were summarized in Supplementary Table 1.

Identification of potential transcriptomic biomarkers

A total of 46 DEGs were identified between SHS and control group, including 22 upregulated genes and 24 downregulated genes in SHS group, based on the criteria of P value < 0.05 and the fold change > 2 (Figure 1). The detailed list of 46 DEGs is presented in Supplementary Table 2. The AUC values were calculated to assess the diagnostic accuracy of 46 candidate transcripts (Supplementary Table 3). As shown in Table 1, 23 transcripts with statistically significant AUC values were selected as the candidate biomarkers for SHS.

GO enrichment analysis

To investigate the biological functions related to SHS, GO enrichment analysis of 46 DEGs was performed. Significantly enriched terms were those with a P value < 0.05 and gene count ≥ 2. The results revealed that the 22 upregulated DEGs were significantly enriched in intermediate filament, contractile fiber, intermediate filament cytoskeleton, and apical plasma membrane in terms of cellular component (CC) (Figure 2A). Regarding biological process (BP), the upregulated DEGs were enriched in adult heart development, organophosphate ester transport, male gonad development, development of primary male sexual characteristics, male sex

differentiation, gonad development, protein heterotetramerization, development of primary sexual characteristics, regulation of cardiac muscle cell proliferation, and cardiac muscle cell proliferation (Figure 2B). Under molecular function (MF), the upregulated DEGs were enriched in ATPase-coupled transmembrane transporter activity, primary active transmembrane transporter activity, lipid transporter activity, and growth factor activity (Figure 2C). For the downregulated DEGs, significant enrichment was observed in anchored component of membrane and vacuolar lumen regarding CC (Figure 2D). In terms of BP, the downregulated DEGs were enriched in negative regulation of bone morphogenetic proteins (BMPs) signalling pathway, positive regulation of extracellular signal-regulated kinase-1 (ERK1) and ERK2 cascade, negative regulation of ossification, gliogenesis, regulation of cellular response to growth factor stimulus, regulation of ERK1 and ERK2 cascade, regulation of BMP signalling pathway, ERK1 and ERK2 cascade, cellular transition metal ion homeostasis, and regulation of gliogenesis (Figure 2E). For MF, the downregulated DEGs were enriched in protein serine/threonine kinase activity (Figure 2F).

KEGG enrichment analysis

To explore the biological pathways related to SHS, KEGG enrichment analysis of DEGs was performed. As shown in Figure 3A, the upregulated DEGs were significantly enriched in ATP-binding cassette (ABC) transporters. In addition, KEGG enrichment analysis identified three significant pathways of the downregulated DEGs, including proteoglycans in cancer, Parkinson disease, and neurodegeneration (Figure 3B).

PPI network analysis and hub gene recognition

In order to provide a comprehensive view of potential functional relationships of DEGs, the 22 upregulated genes and 24 downregulated genes were mapped by the STRING database to establish a PPI network (Figure 4). Protein pairs with a combination score > 0.15 were selected in the PPI network. The top 10 genes with the highest degree of connectivity were defined as critical hub genes (Figure 5), including gap junction protein α1 (*GJA1*), twist family basic helix-loop-helix transcription factor 2 (*TWIST2*), keratin 1 (*KRT1*), tubulin β3 (*TUBB3*), anti-mullerian hormone receptor type 2 (*AMHR2*), bone morphogenetic protein 10 (*BMP10*), metallothionein 3 (*MT3*), bone morphogenetic protein binding endothelial regulator (*BMPER*), neurotrimin (*NTM*), and transmembrane protein 98 (*TMEM98*).

Transcriptome predictive model for SHS

Multivariate binary logistic regression analysis was conducted to construct a predictive model based on the transcripts of 10 hub genes. Transcriptome dataset was normalized before logistic regression analysis and stepwise method (entry criteria: $P < 0.05$ and exclusion criteria: $P > 0.10$) was used to select the best combination of transcripts. Consequently, two transcripts were selected in the model and the model based on candidate transcripts was as follows:

Logit (*p* = SHS) = 0.137 - 66.878 × (*BMPER* mRNA) + 29.328 × (*AMHR2* mRNA)

The model yielded a sensitivity of 83.3% (95% CI: 73.9%-92.7%), a specificity of

90.0% (95% CI: 82.4%-97.6%), and an AUC of 0.938 (95% Confidence Interval (CI): 0.882-0.994, $P < 0.001$) (Table 1 & Figure 6).

Discussion

Predictive diagnosis and targeted prevention is the cost-effective strategy for personalized medicine [34]. As a reversible stage in advance of chronic diseases, SHS proposes a new effective conception for preselection of healthy but pathology-predisposed individuals under the perspective of PPPM. Identifying key biological pathways relevant to SHS and its progression towards chronic diseases are considered as viable strategies for predictive diagnosis, targeted prevention and personalized therapy of chronic diseases. With the emergence of RNA-Seq-based technologies, transcriptome profiling could play a significant role in deciphering gene expressions on RNA level and identifying molecular biomarkers for SHS. The findings of present study have found that blood transcriptomic biomarkers have a promising prospect of clinical application in the prediction and prevention of chronic diseases, as we discuss below.

- **Glucokinase regulator (GCKR)**

In the present study, significantly lower level of *GCKR* mRNA was found in individuals with SHS. Glucokinase regulator protein, encoded by *GCRK* gene, is a hepatocyte-specific regulatory protein that inhibits glucokinase in liver cells [35]. Glucokinase, a hexokinase isozyme, is a key regulator of glucose disposal and storage,

and responds to increases in circulating glucose concentration by initiating a signalling cascade that results in insulin secretion from pancreatic islets β cell [36]. Alterations in glucokinase expression and activity are associated with poorly controlled T2DM [37] and nonalcoholic fatty liver disease (NAFLD) [38]. It has been reported that common variants in the *GCKR* gene are associated with increased blood triglycerides [39, 40], lower fasting glucose [40], and NAFLD [41]. The glucokinase regulator protein, which binds with glucokinase and inactivates it from carbohydrate metabolism, is able to serve as a new treatment target for T2DM [42]. Significantly lower level of *GCKR* mRNA in the SHS individuals in this study indicates that disorders of glucose metabolism might play an important role in the pathophysiology of SHS. Given the findings, *GCKR* mRNA might be a potential predictive diagnostic biomarker for the progression of SHS towards T2DM, and glucokinase regulator protein could be applied as potential therapeutic or preventive targets for SHS and T2DM.

- **Gap junction protein α1 (GJA1)**

The present study observed that *GJA1* is upregulated in individuals with SHS. Gap junction protein α1, also known as connexin 43 protein, is protein subunit that constitute gap junction channels [43]. The intercellular channels of gap junction facilitate the transfer of ions and small molecular from cell to cell, and are thought to modulate several processes, including embryogenesis, differentiation, and electrotonic coupling [44]. *GJA1* expression is affected by several pathophysiological conditions, such as hypertension, hypercholesterolemia, and diabetes [45]. In the

present study, the higher level of *GJA1* mRNA in SHS individuals indicates that *GJA1* mRNA could be associated with the molecular mechanisms of SHS related to chronic diseases, such as hypertension, CVD and T2DM. Building on these findings, *GJA1* mRNA could serve as a potentially predictive biomarker for the early identification of SHS. Gap junction protein α1 is proposed to be a therapeutic or preventive target for the progression of SHS toward chronic diseases.

Furthermore, Squecco et al. has reported that the bioactive sphingolipid, sphingosine 1-phosphate, can enhance GJA1 protein expression [46]. Our previous study has found sphingolipids metabolism is the disturbed metabolic pathway related to SHS, and significantly higher levels of sphinganine 1-phosphate and sphingomyelin are observed in SHS individuals [17]. Taken in consideration these findings, we suggest that the upregulated *GJA1* mRNA could be affected by the disturbed sphingolipids metabolism in SHS individuals. However, further studies are needed to validate the interactions between gap junction protein α1 and sphingolipids in SHS individuals.

- **BMPs signalling pathway**

Functional annotation and pathway enrichment analysis of DEGs provided an intuitive overview of the molecular mechanisms of SHS. The significant GO terms related with SHS were BMPs signalling pathway and regulation of ERK1/2 cascade. BMPs, a group of signalling molecules, are part of the transforming growth factor-β superfamily of proteins. It is initially discovered for their ability to induce bone formation

[47], BMPs are now known to play important roles in the adult vascular endothelium, promoting angiogenesis and mediating oxidative stress [48]. Due to the critical roles of BMPs in maintenance of adult tissue homeostasis, it is found that dysregulation in BMPs signalling pathway contribute to various diseases, including cancer, skeletal disorders and CVD [49]. Our previous studies suggested that SHS might precede the occurrence of CVD [6, 7]. In the present study, significantly lower levels of *BMPER* and hemojuvelin BMP co-receptor (*HJV*) mRNA, which involved in the BMPs signalling pathway, were found in individuals with SHS. Our findings indicate that BMPs signalling pathway may play an important role in the pathophysiology of SHS. Transcripts of *BMPER* and *HJV* could be potential biomarkers for the predictive identification, targeted prevention, and personalized management of the progression of SHS towards chronic diseases.

- **ERK1 and ERK2 cascade**

ERK1 and ERK2 cascade is key signalling pathway that regulates a large variety of cellular processes, including adhesion, migration, differentiation, metabolism, and proliferation [50]. This signalling cascade is dysregulated in a variety of diseases or conditions, including CVD [51], insulin resistance [52], and inflammation [53]. In the current study, significantly lower levels of *metallothionein 3* (*MT3*) and *C-C motif chemokine ligand 3* (*CCL3*) mRNA, which involved in the ERK1 and ERK2 cascade, were observed in individuals with SHS. Our findings demonstrate that the dysregulation of ERK1 and ERK2 cascade are associated with the molecular

mechanisms of SHS. The association between the ERK1 and ERK2 cascade and several pathologies indicate that insulin resistance and inflammation may be involved in the pathophysiology of SHS. Effectively inhibits or blocks the occurrence, development, and progression of the chronic inflammation process playing important roles in the prevention of inflammation-related diseases [54]. The transcripts of *MT3* and *CCL3* could be used as potential biomarkers for population risk stratification, as well as therapeutic or preventive targets for inflammation-related diseases.

- **ABC transporters**

The KEGG enrichment analysis revealed that ABC transporter and neurodegeneration are the biological pathways related to SHS. ABC transporters are a large family of transmembrane proteins. These proteins bind ATP and use the energy to drive the transport of various molecules across cell membranes [55]. In human, the 48 ABC proteins are divided into seven subfamilies, from A to G, based on sequence and organization of their ATP-binding domain [55]. The ABCA4 protein transports vitamin A derivatives and perform a crucial role in the visual cycle [56]. In the present study, the downregulation of *ABCA4* was observed in the individuals with SHS, which indicates that the decreased level of *ABCA4* mRNA is associated with the SHS phenotype of eyes, such as eye ache and fatigue.

The ABCG8 protein functions to facilitate the transport of sterols in the intestine and liver [57]. Our previous study found that steroid hormone biosynthesis pathway is disturbed in SHS individuals [17], which indicates that the upregulation of *ABCG8*

might be associated with the disorder of steroid hormone biosynthesis in SHS individuals. Chen and colleagues have observed that trimethylamine-N-oxide, a metabolite produced by gut microbiota, is associated with increased ABCG8 expression [58]. In addition, Zhu et al. has proved that intestinal microbiota, *Enterococcus faecalis,* increase the expression of ABCG8 [59]. Our previous study has found that alterations of intestinal microbiota occur in SHS individuals [16]. In the present study, the higher level of *ABCG8* mRNA might be associated with the altered diversity of intestinal microbiota in SHS individuals.

- **Neurodegeneration**

Neurodegeneration is an age-related progressive deterioration of neuronal structures and functions ultimately leading to cognitive disability. It is the main pathological feature of neurodegenerative diseases, including Parkinson's disease and Alzheimer's disease. Timely and prompt diagnosis of the neurodegeneration disease is critical as it offers a chance for an early intervention that may be helpful in preventing further progression of the diseases. Predictive transcriptome patterns in blood can serve as biomarkers for the early identification of neurodegeneration [60]. In the present study, significantly lower levels of *TUBB3* and calcium/calmodulin dependent protein kinase II β (*CAMK2B*) mRNA, involved in the neurodegeneration diseases, were observed in SHS individuals. These findings indicate that neurodegeneration might be involved in the pathophysiology of SHS. Our previous study also found the association between SHS and cognitive impairment [61]. Given

these findings, transcripts of *TUBB3* and *CAMK2B* could be potential biomarkers for the predictive identification, targeted prevention, and personalized management of the progression of SHS towards neurodegeneration diseases, such as Parkinson's disease and Alzheimer's disease.

Strength and Limitations

To the best of our knowledge, this is the first study to provide the comprehensive transcriptomic characterization for SHS using RNA-Seq-based transcriptome profiling. Our findings exhibit strong predictive abilities of transcriptomic biomarkers for the early identification of SHS. These transcriptomic biomarkers also provide better insight into the critical genes associated with SHS and a deeper understanding of its biological progression towards chronic diseases. Thus, the SHS-related transcripts and the particular pattern of transcripts can be used as potential biomarkers for the predictive diagnosis, population risk stratification, personalized management, and targeted intervention of SHS towards chronic diseases and neurodegenerative diseases.

Several limitations in the present study are noteworthy. First, our study is a case-control study with a relatively small sample size, hence the generalisation of these findings could be questioned. Considering the fact that our study provides the original observations on the transcriptomic characteristics of SHS population, the present study has provided a new idea that blood transcripts might offer a novel alternative for the predictive diagnosis, targeted prevention and personalized treatment of chronic diseases. Therefore, further validations for these transcriptomic biomarkers in multi-

centered clinical studies with large sample sizes and in multiple ethnic populations are also needed. In addition, considering the quantitative accuracy of RNA-Seq technology, a RT-qPCR study is underway against the same cohort to validate the putative transcript biomarkers and selected hub genes based on the findings in this study. Building on the present findings, it is essential to deeply investigate the functions of those SHS-related DEGs and the biological pathways in the progression of SHS toward chronic diseases.

Conclusions and expert recommendations

The early identification of SHS has the potential of predicting chronic diseases at early stage, and evidence-based intervention on SHS may be a cost-effective way for the targeted prevention, and personalized therapy of chronic diseases. A total of 23 transcripts was selected as candidate predictive biomarkers for the identification of SHS. The present study clearly revealed the potential value of transcriptomic biomarkers for the predictive identification of SHS from the perspective of PPPM. The pattern of the differentially expressed genes can be used as biomarkers for patient stratification. We suggest that SHS-related transcriptomic biomarkers and other SHS-related biomarkers at multi-omics levels are the key promise for the practice of PPPM in chronic diseases.

The downregulation of *GCKR* in SHS individuals indicated that glucose metabolism disorder plays a significant role in the pathophysiologic mechanism of SHS. The GCKR protein could serve as a potentially preventive/therapeutic target for

the progression from SHS towards T2DM. We recommend that further studies of large prospective cohorts should be conducted to investigate the crucial role of *GCKR* in the SHS progression, which could be useful for the predictive diagnosis and targeted prevention of T2DM. BMPs signalling pathway, ERK 1/2 cascade and ABC transporters play potential roles in the pathophysiologic mechanism of SHS. These findings determined the potential utility of SHS-related signalling pathways for targeted prevention and personalized therapy of chronic diseases. We recommend strengthening the studies of signaling pathways in SHS with different omics strategies. Ten genes, including *GJA1*, *TWIST2*, *KRT1*, *TUBB3*, *AMHR2*, *BMP10*, *MT3*, *BMPER*, *NTM* and *TMEM98,* play hub roles in the underlying mechanisms of SHS. We also suggest that it is crucial to deeply study the functions and activities of these differentially expressed genes in the underlying mechanism of SHS. SHS may precede the actual onset of neurodegeneration diseases, such as Alzheimer's disease and Parkinson's disease. The integration of subjective health measure (SHSQ-25) and objective biomarkers (transcripts of *TUBB3* and *CAMK2B*) enables clinician and public health workers to predict individual's high risk of developing neurodegeneration diseases.

SHSQ-25 can be used as an alternative health screening tool in the population-based health survey, particularly when there is a lack of laboratory-based resources. Primary healthcare providers must be able to detect and manage SHS to fight delayed diagnosis, untargeted prevention, and ineffective intervention of chronic disease.

Acknowledgments

The authors acknowledge the participants and their families who donated their time and effort in helping to make this study possible.

Authors' contribution

HW and YW participated in the design of the study. HW, QT, JZ, HL, WC, XZ, XL, LW, MS and YK performed participant enrollment and collected the samples. HW, QT and JXZ performed the Transcriptome analysis. HW and QT performed the statistical analysis and drafted the manuscript. YK, WW and YW revised the manuscript.

Funding information

This work was partially supported by the "Beijing Talents Project (2020A17)" and the National Natural Science Foundation of China (Grant Numbers: 81872682 & 81773527). HW was supported by the China Scholarship Council (CSC 201708110200).

Compliance and Ethical Standards

Conflicts of interest

The authors confirm that there are no conflicts of interest.

Consent for publication

Not applicable

Ethical approval and consent to participate

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.

Abbreviations

ABC	ATP-binding cassette
AUC	Area under the curve
BMPs	Bone morphogenetic proteins
BP	Biological process
CC	Cellular component
cDNA	Complementary DNA
CI	Confidence interval
CVD	Cardiovascular disease
DEGs	Differentially expressed genes
ERK1	Extracellular signal-regulated kinase-1
FDR	False discovery rate
GO	Gene ontology
IQR	Interquartile ranges
KEGG	Kyoto encyclopedia of genes and genomes
MF	Molecular function
NAFLD	Nonalcoholic fatty liver disease
PPI	Protein-protein Interaction
PPPM	Preventive, predictive and personalised medicine
RNA-Seq	RNA sequencing
ROC	Receiver operating characteristic
SD	Standard deviation

SHS	Suboptimal health status
SHSQ-25	Suboptimal health status questionnaire-25
STRING	Search tool for the retrieval of interacting genes
T2DM	Type 2 diabetes mellitus

Abbreviations for all particular genes can be found at the following link: www.ncbi.nlm.nih.gov/gene/

1

2

References

1. Golubnitschaja O, Watson ID, Topic E, Sandberg S, Ferrari M, Costigliola V. Position paper of the EPMA and EFLM: a global vision of the consolidated promotion of an integrative medical approach to advance health care. EPMA J. 2013;4(1):12.

2. Golubnitschaja O, Baban B, Boniolo G, Wang W, Bubnov R, Kapalla M et al. Medicine in the early twenty-first century: paradigm and anticipation - EPMA position paper 2016. EPMA J. 2016;7:23.

3. Golubnitschaja O, Costigliola V, Epma. General report & recommendations in predictive, preventive and personalised medicine 2012: white paper of the European Association for Predictive, Preventive and Personalised Medicine. EPMA J. 2012;3(1):14.

4. 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.

5. 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 J. 2019;10(1):1-8.

6. Yan Y, Dong J, Liu Y, Yang X, 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.

7. 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.

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

9. Yan Y, Liu Y, Li M, Hu P, Guo A, Yang X et al. Development and evaluation of a questionnaire for measuring suboptimal health status in urban Chinese. J Epidemiol. 2009;19(6):333-41.

10. 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.

11. Anto EO, Roberts P, Coall D, Turpin CA, Adua E, Wang Y et al. Integration of suboptimal health status evaluation as a criterion for prediction of preeclampsia is strongly recommended for healthcare management in pregnancy: a prospective cohort study in a Ghanaian population. EPMA J. 2019;10(3):211-26.

12. 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.

13. Gerner C, Costigliola V, Golubnitschaja O. Multiomic Patterns in Body Fluids: Technological Challenge with a Great Potential to Implement the Advanced Paradigm of 3P Medicine. Mass Spectrom Rev. 2020;39(5-6):442-51.

14. Yan Y, Wu LJ Xiao H, 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.

15. Alzain MA, Asweto CO, Zhang J, Fang H, Zhao Z, Guo X et al. Telomere Length

and Accelerated Biological Aging in the China Suboptimal Health Cohort: A Case-Control Study. OMICS. 2017;21(6):333-9.

16. Sun Q, Xu X, Zhang J, Sun M, Tian Q, Li Q et al. Association of suboptimal health status with intestinal microbiota in Chinese youths. J Cell Mol Med. 2019;24(2):1837-47.

17. Wang H, Tian QY, Zhang J, Liu HQ, Zhang XY, Cao WJ et al. 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. EPMA J. 2020;11(2):147-60.

18. Byron SA, Van Keuren-Jensen KR, Engelthaler DM, Carpten JD, Craig DW. Translating RNA sequencing into clinical diagnostics: opportunities and challenges. Nat Rev Genet. 2016;17(5):257-71.

19. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10(1):57-63.

20. Casamassimi A, Federico A, Rienzo M, Esposito S, Ciccodicola A. Transcriptome Profiling in Human Diseases: New Advances and Perspectives. Int J Mol Sci. 2017;18(8):1652.

21. Yan Y, Dong J, Liu Y, Zhang J, Song M, 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.

22. Sweetman E, Ryan M, Edgar C, MacKay A, Vallings R, Tate W. Changes in the transcriptome of circulating immune cells of a New Zealand cohort with myalgic

encephalomyelitis/chronic fatigue syndrome. Int J Immunopathol Pharmacol. 2019;33:2058738418820402.

23. Tomas-Roig J, Havemann-Reinecke U. Gene expression signature in brain regions exposed to long-term psychosocial stress following acute challenge with cannabinoid drugs. Psychoneuroendocrinology. 2019;102:1-8.

24. 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.

25. 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.

26. 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.

27. 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.

28. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923-30.

29. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

30. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284-7.

31. von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: a database of predicted functional associations between proteins. Nucleic Acids Res. 2003;31(1):258-61.

32. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol. 2014;8 Suppl 4:S11.

33. R Foundation for Statistical Computing. R: A language and environment for statistical computing. 2017. https://www.R-project. org/. Accessed 22 Feb 2021.

34. Golubnitschaja O, Costigliola V. Common origin but individual outcomes: time for new guidelines in personalized healthcare. Per Med. 2010;7(5):561-8.

35. Van Schaftingen E. A protein from rat liver confers to glucokinase the property of being antagonistically regulated by fructose 6-phosphate and fructose 1-phosphate. Eur J Biochem. 1989;179(1):179-84.

36. Agius L. Glucokinase and molecular aspects of liver glycogen metabolism. Biochem J. 2008;414(1):1-18.

37. Haeusler RA, Camastra S, Astiarraga B, Nannipieri M, Anselmino M, Ferrannini E. Decreased expression of hepatic glucokinase in type 2 diabetes. Mol Metab. 2015;4(3):222-6.

38. Peter A, Stefan N, Cegan A, Walenta M, Wagner S, Konigsrainer A et al. Hepatic glucokinase expression is associated with lipogenesis and fatty liver in humans. J Clin Endocrinol Metab. 2011;96(7):E1126-30.

39. Orho-Melander M, Melander O, Guiducci C, Perez-Martinez P, Corella D, Roos C

et al. Common missense variant in the glucokinase regulatory protein gene is associated with increased plasma triglyceride and C-reactive protein but lower fasting glucose concentrations. Diabetes. 2008;57(11):3112-21.

40. Diabetes Genetics Initiative of Broad Institute of H, Mit LU, Novartis Institutes of BioMedical R, Saxena R, Voight BF, Lyssenko V et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science. 2007;316(5829):1331-6.

41. Zain SM, Mohamed Z, Mohamed R. Common variant in the glucokinase regulatory gene rs780094 and risk of nonalcoholic fatty liver disease: a meta-analysis. J Gastroenterol Hepatol. 2015;30(1):21-7.

42. Kaushik A, Kaushik M. Recent Updates on Glucokinase Activators and Glucokinase Regulatory Protein Disrupters for the Treatment of Type 2 Diabetes Mellitus. Curr Diabetes Rev. 2019;15(3):205-12.

43. Fishman GI, Eddy RL, Shows TB, Rosenthal L, Leinwand LA. The human connexin gene family of gap junction proteins: distinct chromosomal locations but similar structures. Genomics. 1991;10(1):250-6.

44. Beyer EC, Paul DL, Goodenough DA. Connexin family of gap junction proteins. J Membr Biol. 1990;116(3):187-94.

45. Schulz R, Gorge PM, Gorbe A, Ferdinandy P, Lampe PD, Leybaert L. Connexin 43 is an emerging therapeutic target in ischemia/reperfusion injury, cardioprotection and neuroprotection. Pharmacol Ther. 2015;153:90-106.

46. Squecco R, Sassoli C, Nuti F, Martinesi M, Chellini F, Nosi D et al. Sphingosine 1-

phosphate induces myoblast differentiation through Cx43 protein expression: a role for a gap junction-dependent and -independent function. Mol Biol Cell. 2006;17(11):4896-910.

47. Urist MR. Bone: formation by autoinduction. Science. 1965;150(3698):893-9.

48. Dyer LA, Pi X, Patterson C. The role of BMPs in endothelial cell function and dysfunction. Trends Endocrinol Metab. 2014;25(9):472-80.

49. Wang RN, Green J, Wang Z, Deng Y, Qiao M, Peabody M et al. Bone Morphogenetic Protein (BMP) signaling in development and human diseases. Genes Dis. 2014;1(1):87-105.

50. Roskoski R, Jr. ERK1/2 MAP kinases: structure, function, and regulation. Pharmacol Res. 2012;66(2):105-43.

51. Muslin AJ. MAPK signalling in cardiovascular health and disease: molecular mechanisms and therapeutic targets. Clin Sci (Lond). 2008;115(7):203-18.

52. Tanti JF, Jager J. Cellular mechanisms of insulin resistance: role of stress-regulated serine kinases and insulin receptor substrates (IRS) serine phosphorylation. Curr Opin Pharmacol. 2009;9(6):753-62.

53. Kim EK, Choi EJ. Pathological roles of MAPK signaling pathways in human diseases. Biochim Biophys Acta. 2010;1802(4):396-405.

54. Qian S, Golubnitschaja O, Zhan X. Chronic inflammation: key player and biomarker-set to predict and prevent cancer development and progression based on individualized patient profiles. EPMA J. 2019;10(4):365-81.

55. Higgins CF. ABC transporters: from microorganisms to man. Annu Rev Cell Biol.

1992;8:67-113.

56. Sun H, Molday RS, Nathans J. Retinal stimulates ATP hydrolysis by purified and reconstituted ABCR, the photoreceptor-specific ATP-binding cassette transporter responsible for Stargardt disease. J Biol Chem. 1999;274(12):8269-81.

57. Berge KE, Tian H, Graf GA, Yu L, Grishin NV, Schultz J et al. Accumulation of dietary cholesterol in sitosterolemia caused by mutations in adjacent ABC transporters. Science. 2000;290(5497):1771-5.

58. Chen Y, Weng Z, Liu Q, Shao W, Guo W, Chen C et al. FMO3 and its metabolite TMAO contribute to the formation of gallstones. Biochim Biophys Acta Mol Basis Dis. 2019;1865(10):2576-85.

59. Zhu Y, Li T, Din AU, Hassan A, Wang Y, Wang G. Beneficial effects of Enterococcus faecalis in hypercholesterolemic mice on cholesterol transportation and gut microbiota. Appl Microbiol Biotechnol. 2019;103(7):3181-91.

60. Yeghiazaryan K, Flammer J, Golubnitschaja O. Predictive molecular profiling in blood of healthy vasospastic individuals: clue to targeted prevention as personalised medicine to effective costs. EPMA J. 2010;1(2):263-72.

61. Ding G, Zhao X, Wang Y, Song D, Chen D, Deng Y et al. Evaluation of the relationship between cognitive impairment and suboptimal health status in a northern Chinese population: a cross-sectional study. J Glob Health. 2020;10(1):010804.

Table 1. Diagnostic accuracy of candidate transcripts

Gene Symbol	Sensitivity	Specificity	AUC	95% CI	*P* value
Proposed Predictive Model	83.3%	90.0%	0.938	(0.882 – 0.994)	<0.001
ABCA4	50.0%	83.3%	0.667	(0.528 - 0.806)	0.026
AMHR2	60.0%	83.3%	0.703	(0.569 - 0.838)	0.007
BMPER	50.0%	90.0%	0.705	(0.572 - 0.838)	0.006
CAMK2B	56.7%	76.7%	0.682	(0.545 - 0.819)	0.016
CCL3	73.3%	56.7%	0.660	(0.520 - 0.800)	0.033
CYP2W1	63.3%	73.3%	0.691	(0.555 - 0.828)	0.011
DNAJC12	73.3%	70.0%	0.740	(0.612 - 0.868)	0.001
EMX1	56.7%	70.0%	0.649	(0.510 - 0.789)	0.047
EPGN	33.3%	96.7%	0.662	(0.523 - 0.801)	0.031
GCKR	50.0%	86.7%	0.693	(0.558 - 0.828)	0.010
HJV	50.0%	80.0%	0.657	(0.518 - 0.796)	0.036
HRNR	73.3%	33.3%	0.711	(0.580 - 0.842)	0.005
LCN8	83.3%	50.0%	0.683	(0.549 - 0.818)	0.015
NKAIN4	43.3%	90.0%	0.650	(0.510 - 0.790)	0.046
NPW	50.0%	86.7%	0.674	(0.537 - 0.812)	0.020
OR10AD1	73.3%	60.0%	0.687	(0.551 - 0.822)	0.013
PCDHGB4	73.3%	53.3%	0.678	(0.542 - 0.814)	0.018
RANBP17	66.7%	73.3%	0.709	(0.575 - 0.842)	0.005
SLC12A1	60.0%	63.3%	0.652	(0.513 - 0.792)	0.043
TBC1D3	70.0%	70.0%	0.697	(0.562 - 0.831)	0.009
TMEM211	53.3%	83.3%	0.653	(0.513 - 0.793)	0.041
TUBB3	60.0%	70.0%	0.679	(0.543 - 0.815)	0.017
TWIST2	76.7%	56.7%	0.695	(0.561 - 0.829)	0.009

AUC, area under the receiver operating characteristic curves; CI, confidence interval. The full names of all particular genes are presented in Supplementary Table 2. Abbreviations for all particular genes can be found at the following link: www.ncbi.nlm.nih.gov/gene. $P < 0.05$ is considered statistically significant.

Figure legends

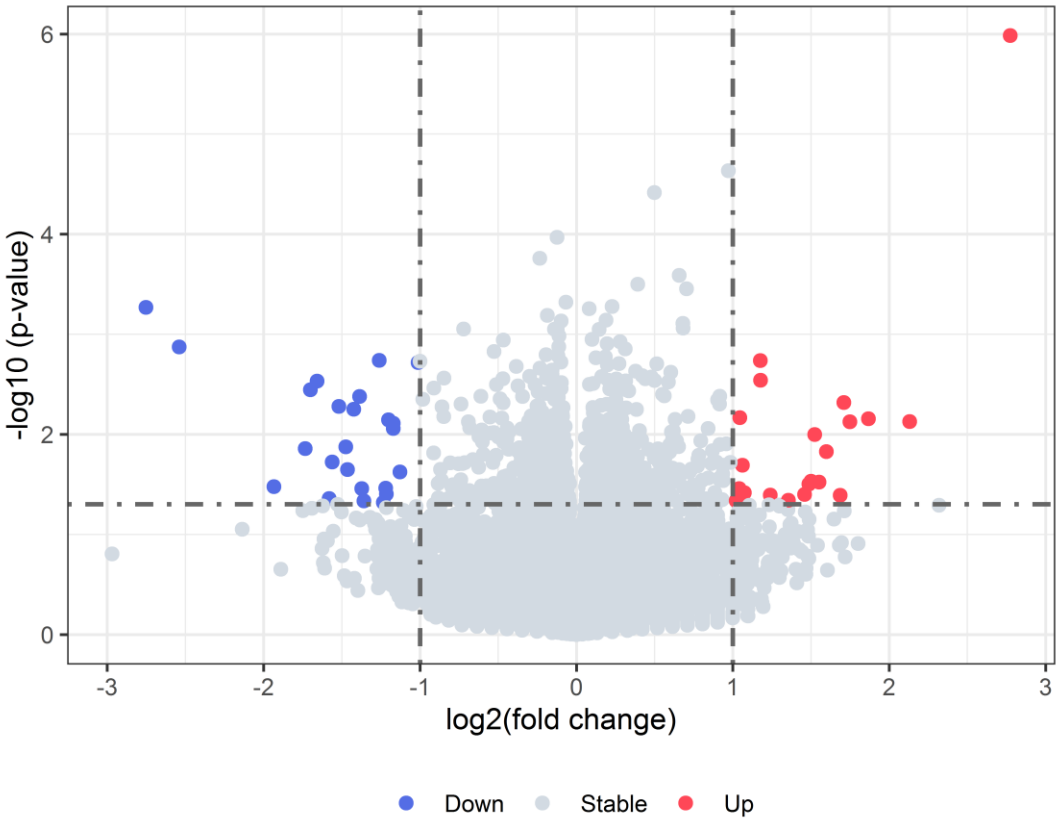

Fig. 1 Volcano plot of DEGs

DEGs (*P* value < 0.05 and Log2 (fold change) > 1.0) between SHS and control groups are presented using a volcano plot. The upregulated genes are presented by red color, while the downregulated genes are presented by blue color; The detailed information of DEGs are presented in Supplementary Table 2. DEGs. differential expressed genes.

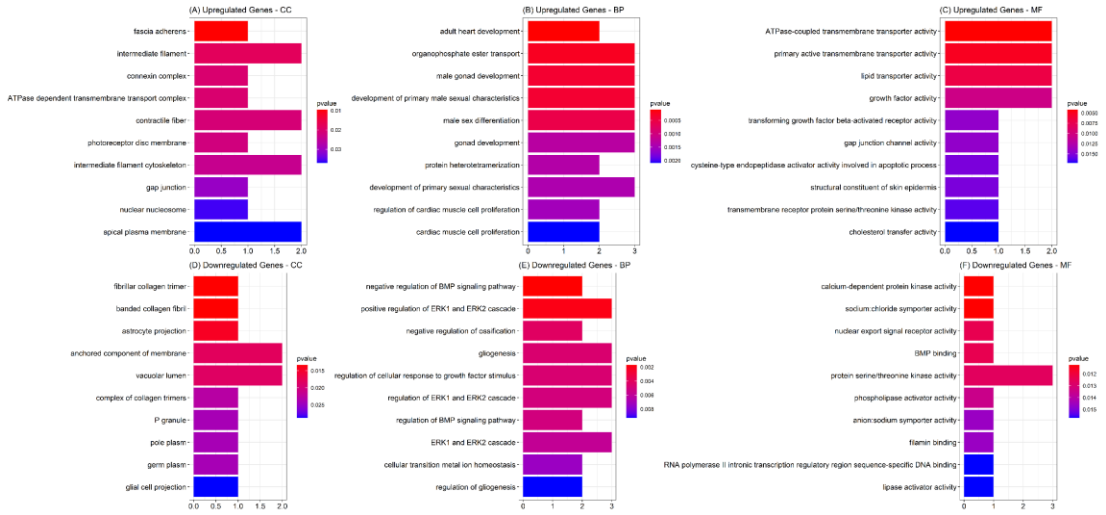

Fig. 2 GO enrichment analysis of DEGs.

GO enrichment analysis was performed to investigate the biological functions related to SHS based on 46 DEGs. The y-axis represents GO enriched terms. The x-axis represents the gene ratio under specific GO terms. The color of the bar represents the *P* value. BP, biological process; CC, cellular component; MF molecular function; GO, Gene ontology; DEGs, differentially expressed genes.

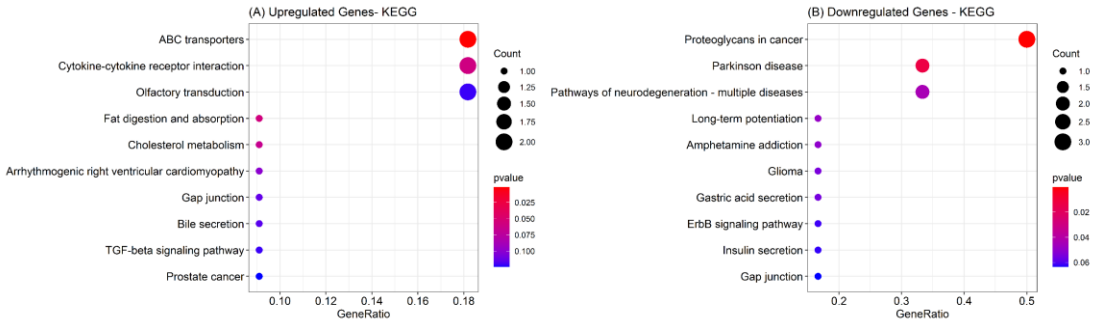

Fig. 3 KEGG enrichment analysis of DGEs

KEGG enrichment analysis was performed to investigate the biological pathways related to SHS based on 46 DEGs. The y-axis represents KEGG enriched terms. The x-axis represents the gene ratio under a specific KEGG terms. The color of the bar represents the *P* value. The size of dot represents the count of genes under specific KEGG terms. KEGG, Kyoto Encyclopedia of Genes and Genomes; DEGs, differentially expressed genes.

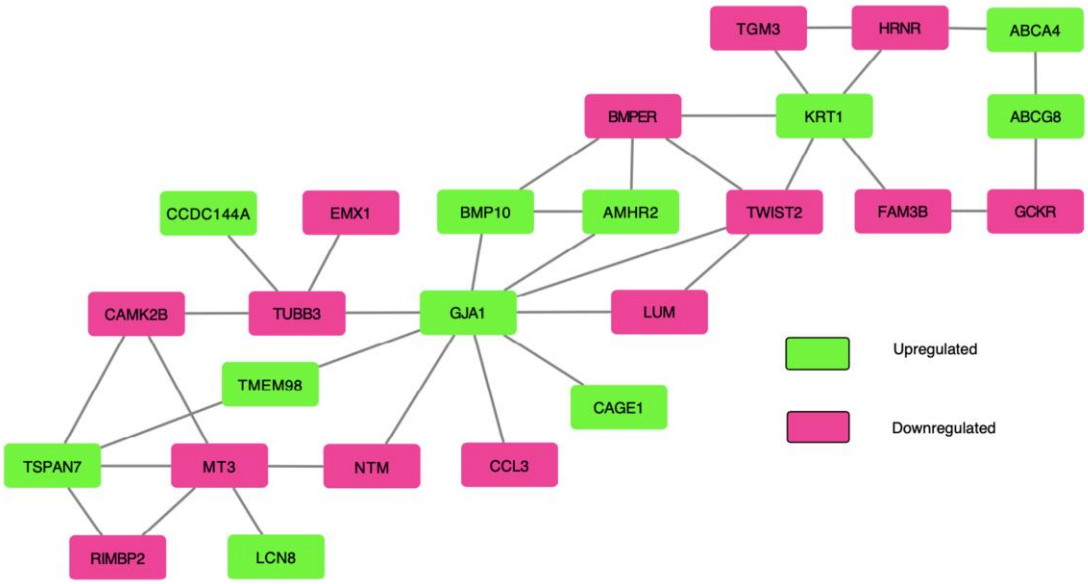

Fig. 4 PPI network of DEGs

The PPI network was mapped to show the potential functional relationships of DEGs. The nodes represent the significantly upregulated/downregulated genes. The edges represent the interaction of proteins coded by DEGs. The resource for the abbreviations of the protein names is from the STRING database. DEGs, differentially expressed genes. PPI, protein-protein interaction; STRING, Search Tool for the Retrieval of Interacting Genes.

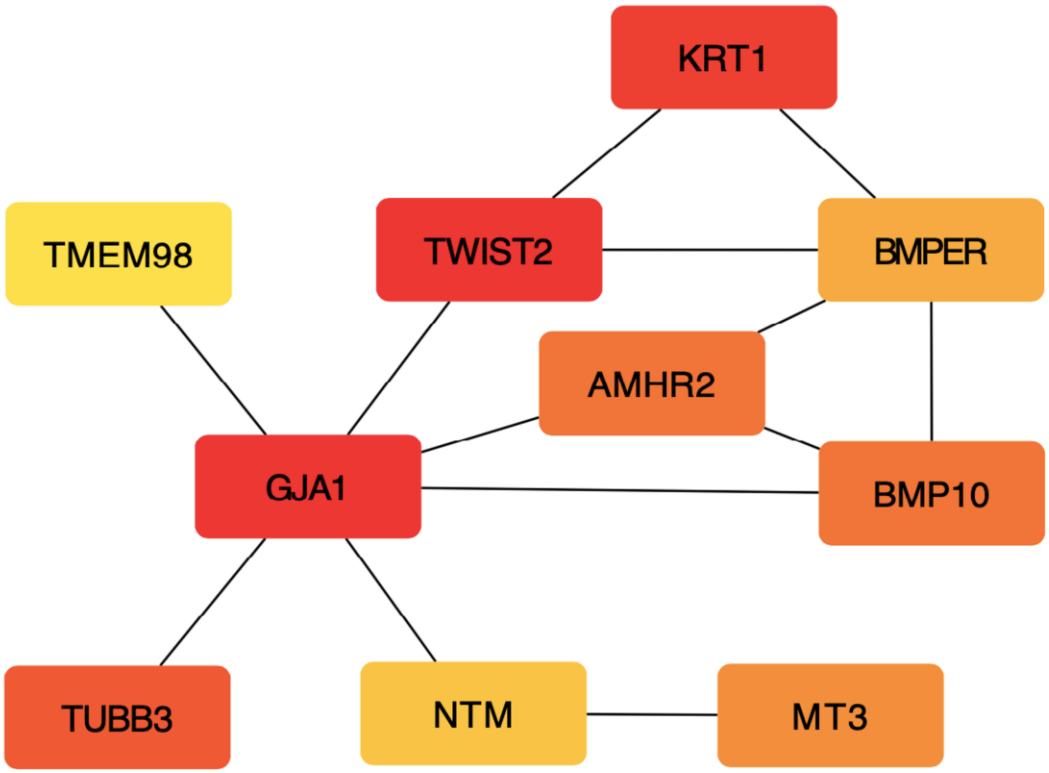

Fig. 5 Hub genes identified in PPI network

The top 10 genes of 46 DEGs with the highest degree of connectivity (combination score > 0.15) were defined as critical hub genes. The nodes represent the significantly upregulated/downregulated genes. The edges represent the interaction of proteins coded by DEGs. The node color changes gradually from yellow to red in ascending order according to the degree of the genes. The resource for the abbreviations of the protein names is form the STRING database. PPI, protein-protein interaction; STRING, Search Tool for the Retrieval of Interacting Genes.

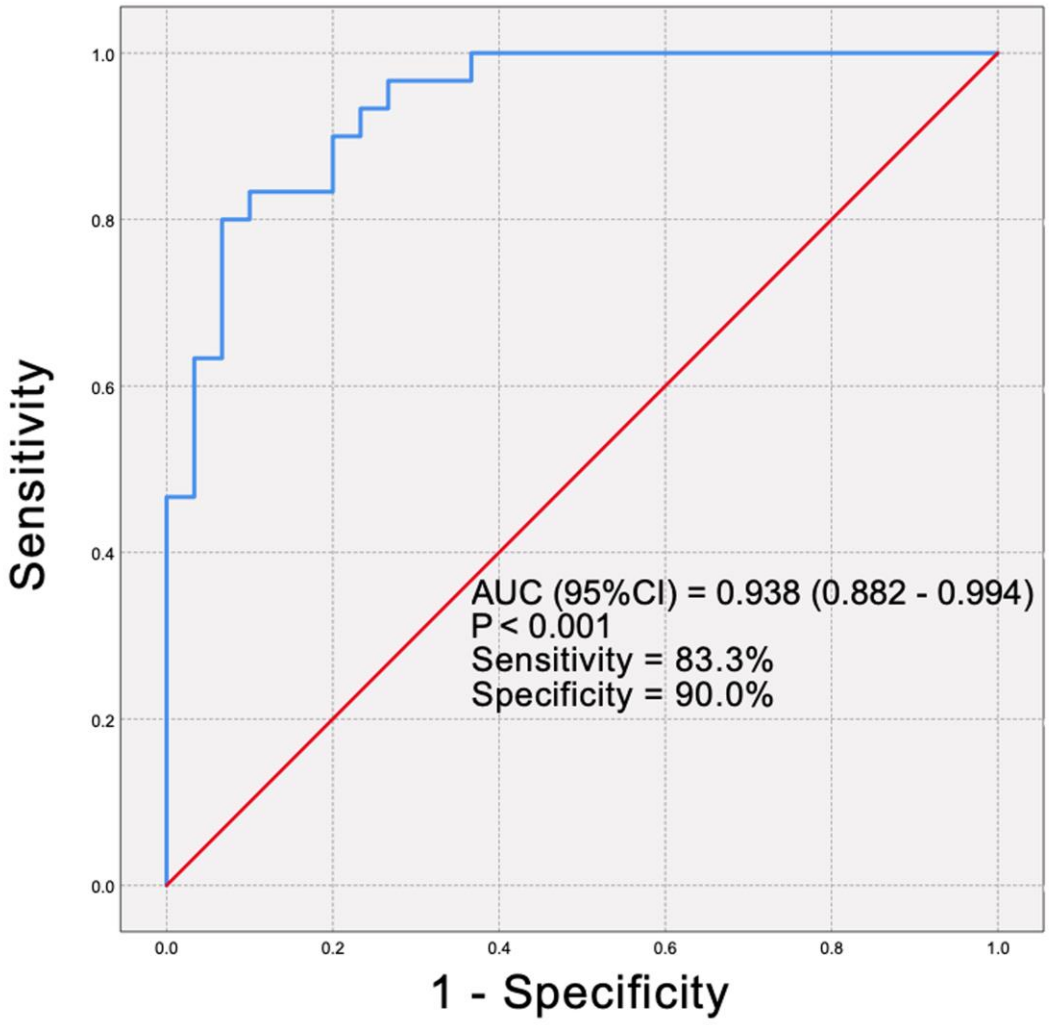

Fig. 6 ROC curve for transcriptome predictive model

Proposed predictive model：Logit (p = SHS) = 0.137 - 66.878 × (*BMPER* mRNA) + 29.328 × (*AMHR2* mRNA). AUC, area under the curve; CI, confidence interval; ROC, receiver operating characteristic.