Acupuncture and Herbal Medicine

Integrated causal inference modeling uncovers novel causal factors and potential therapeutic targets of Qingjin Yiqi granules for chronic fatigue syndrome

Junrong Li[1], Xiaobing Zhai[2], Jixing Liu[2], Chi Kin Lam[1], Weiyu Meng[2], Yuefei Wang[3,4], Shu Li[2], Yapeng Wang[1,*], Kefeng Li[2,*]

[1]Faculty of Applied Sciences, Macao Polytechnic University, Macau SAR, China; [2]Center for Artificial Intelligence Driven Drug Discovery, Faculty of Applied Sciences, Macao Polytechnic University, Macau SAR, China; [3]National Key Laboratory of Chinese Medicine Modernization, Tianjin Key Laboratory of TCM Chemistry and Analysis, Tianjin University of Traditional Chinese Medicine, Tianjin, China; [4]Haihe Laboratory of Modern Chinese Medicine, Tianjin, China

Abstract

Objective: Chronic fatigue syndrome (CFS) is a prevalent symptom of post-coronavirus disease 2019 (COVID-19) and is associated with unclear disease mechanisms. The herbal medicine Qingjin Yiqi granules (QJYQ) constitute a clinically approved formula for treating post-COVID-19; however, its potential as a drug target for treating CFS remains largely unknown. This study aimed to identify novel causal factors for CFS and elucidate the potential targets and pharmacological mechanisms of action of QJYQ in treating CFS.

Methods: This prospective cohort analysis included 4,212 adults aged ≥65 years who were followed up for 7 years with 435 incident CFS cases. Causal modeling and multivariate logistic regression analysis were performed to identify the potential causal determinants of CFS. A proteome-wide, two-sample Mendelian randomization (MR) analysis was employed to explore the proteins associated with the identified causal factors of CFS, which may serve as potential drug targets. Furthermore, we performed a virtual screening analysis to assess the binding affinity between the bioactive compounds in QJYQ and CFS-associated proteins.

Results: Among 4,212 participants (47.5% men) with a median age of 69 years (interquartile range: 69–70 years) enrolled in 2004, 435 developed CFS by 2011. Causal graph analysis with multivariate logistic regression identified frequent cough (odds ratio: 1.74, 95% confidence interval [CI]: 1.15–2.63) and insomnia (odds ratio: 2.59, 95% CI: 1.77–3.79) as novel causal factors of CFS. Proteome-wide MR analysis revealed that the upregulation of endothelial cell-selective adhesion molecule (ESAM) was causally linked to both chronic cough (odds ratio: 1.019, 95% CI: 1.012–1.026, $P = 2.75\ e^{-05}$) and insomnia (odds ratio: 1.015, 95% CI: 1.008–1.022, $P = 4.40\ e^{-08}$) in CFS. The major bioactive compounds of QJYQ, ginsenoside Rb2 (docking score: –6.03) and RG4 (docking score: –6.15), bound to ESAM with high affinity based on virtual screening.

Conclusions: Our integrated analytical framework combining epidemiological, genetic, and *in silico* data provides a novel strategy for elucidating complex disease mechanisms, such as CFS, and informing models of action of traditional Chinese medicines, such as QJYQ. Further validation in animal models is warranted to confirm the potential pharmacological effects of QJYQ on ESAM and as a treatment for CFS.

Keywords: Causal factors, Causal graph analysis, Chronic fatigue syndrome, Drug targets, Mendelian randomization, Qingjin Yiqi

Graphical abstract: http://links.lww.com/AHM/A95.

Introduction

Chronic fatigue syndrome (CFS), also known as myalgic encephalomyelitis (ME), is a severe illness characterized by persistent and debilitating fatigue that lasts for at least 6 months and is accompanied by various somatic manifestations, such as muscle pain, tender lymph nodes, joint pain, chills, and feverishness, which are not alleviated by rest[1]. According to a literature review by the Institute of Medicine, the prevalence of ME/CFS in the U.S. population is between 836,000 and 2.5 million, with an annual economic burden of $17 to $240 billion[2]. Globally, this illness affects 17 to 24 million people worldwide,

Junrong Li and Xiaobing Zhai contributed equally to this work.

**Corresponding author. Kefeng Li, E-mail: kefengl@mpu.edu.mo; Yapeng Wang, E-mail: yapengwang@mpu.edu.mo.*

Received 7 July 2023 / Accepted 31 January 2024

How to cite this article: *Li JR, Zhai XB, Liu JX, Lam CK, Meng WY, Wang YF, Li S, Wang YP, Li KF. Integrated causal inference modeling uncovers novel causal factors and potential therapeutic targets of Qingjin Yiqi granules for chronic fatigue syndrome. Acupunct Herb Med 2024;4(1):122–133. DOI: 10.1097/HM9.0000000000000098*

which is approximately equal to 1% of the population, with a greater proportion of female patients than male patients[3]. In addition, CFS is an important aspect of post-coronavirus disease 2019 (COVID-19) (PCC) and is associated with long-term fatigue, physical weakness, and cognitive impairment after COVID-19 infection[4]. A previous study has demonstrated that although most COVID-19 symptoms resolve after discharge from the hospital, approximately 69% of patients still report experiencing persistent fatigue[5]. Therefore, understanding the etiology and causes of CFS is critical for developing effective treatment strategies and improving the quality of life for individuals with this disease.

Previous studies have indicated an increased risk of CFS in patients with viral infections[6]. In addition, immune system dysfunctions, genetic causes, and environmental factors were reported to be significantly associated with CFS[7]. Although several potential risk factors have been identified, the pathogenesis of CFS remains unclear. This is largely because of the reliance on evidence from cross-sectional studies, which can establish only associations, not causality.

Causal inference employs statistical methods and artificial intelligence to infer causal relationships between variables based on observed data. This approach can elucidate disease mechanisms. In causal graphical models, graphs are used to describe conditional independence relationships between multiple random variables. Each vertex represents a variable, and each edge represents a causal link between variables. These graphs elucidate the important effects of variables on disease occurrence, including variable interactions. Two common algorithms in causal graph modeling are fast causal inference (FCI) and fast greedy equivalence search (FGES). FCI uses a constraint-based algorithm that constructs causal structures based on conditional independence constraints which includes two key steps[8]. First, the FCI creates a fully connected graph of the random set V. It uses a significance threshold α to test the conditional independence of adjacent statistical variables. Any edges between conditionally independent variables are deleted. Second, the FCI determines the directionality of the remaining edges in the graph. Furthermore, the FCI algorithm has the capability to detect potential confounding factors in the causal network. By contrast, FGES utilizes a score-based search[9], beginning with an empty graph and adding or deleting edges to maximize the Bayesian information criterion (BIC)[10]. The first search stage uses a greedy forward search to sequentially add edges to the empty graph until the BIC stops increasing. The second search stage employs a greedy backward search to continuously delete edges until the BIC can no longer increase[11]. Currently, causal graph learning has been applied in several disease studies. For example, causal determinants of length of hospital stay after heart surgery have been explored[12], the risk of coronary artery disease has been estimated[13], and the problem of genetic admixture has been studied[14]. However, causal graphs have not yet been used in CFS studies.

Mendelian randomization (MR) is another approach for inferring causal inference, leveraging natural genetic variations arising from meiosis to investigate potential causal relationships between exposures and outcomes[15]. A key strength of MR is that, unlike observational studies, it is less susceptible to bias from reverse causation and confounding since genetic variants are randomly assigned at conception[16]. Case–control studies have demonstrated alterations in circulating proteins and metabolites such as purines and phospholipids in patients with CFS versus control participants[17]. However, information on which proteins are causally linked to the determinants of CFS is limited. Thus, identifying these genes could reveal potential drug targets.

Qingjin Yiqi granules (QJYQ) is a formula containing the following 16 herbs: *Ginseng Radix* et Rhizoma (Renshen), *Ophiopogonis Radix* (Maidong), *Schisandrae chinensis* Fructus (Wuweizi), *Poria cocos* (Fuling), *Pinelliae Rhizoma* (Banxia), *Scrophulariae Radix* (Xuanshen), *Atractylodis Rhizoma* (Gangzhu), *Citri Reticulatae* Pericarpium (Chenpi), *Glycyrrhizae Radix* et Rhizoma (Gancao), *Bupleuri Radix* (Chaihu), *Cimicifugae Rhizoma* (Shengma), *Coicis Semen* (Yiyiren), *Scutellariae Radix* (Huangqin), *Verbenae Herba* (Mabiancao), *Phragmitis Rhizoma* (Lugen), and *Lophatheri Herba* (Danzhuye). Chemical fingerprint analysis revealed 163 phytochemicals in QJYQ, the main active ingredients of which were ginsenoside Rb1, ginsenoside Rb2, and glycyrrhizic acid[18]. Clinically, QJYQ is recommended for use in rehabilitation guidelines for patients with PCC in the Hebei and Tianjin Provinces of China. A recent randomized controlled trial has demonstrated that QJYQ significantly improved breathlessness and fatigue in patients with PCC after 14 d of treatment[19]. However, the specific drug targets and pharmacological mechanisms of action of QJYQ in treating CFS have not been identified.

To address these gaps, this study aimed to identify novel causal factors involved in CFS pathogenesis through causal graph analysis of data from the Wisconsin Longitudinal Study (WLS) prospective cohort. We subsequently performed a proteome-wide MR analysis to explore the proteins causally associated with the identified CFS risk factors. Furthermore, virtual screening analysis was conducted to elucidate the potential pharmacological mechanisms of QJYQ for treating CFS. Our study demonstrated a new approach for elucidating the mechanisms of complex diseases as well as the pharmacological actions of traditional Chinese medicine (TCM).

Materials and methods

Population

The WLS is a long-term study of 10,317 graduates from Wisconsin high schools conducted in 1957[20]. This study provides information on the life course of the respondents and their parents, covering social background, youthful aspirations, education, military service, family characteristics, social engagement, and psychological qualities. Since 1977, data have also been collected on respondents' siblings, who share a family identifier (familypub) that enables the investigation of genetic factors.

We examined WLS data from 20,470 participants from 2004 to 2011. We excluded those who died after 2004 (n = 3,735), had missing values for CFS (n = 8,130)

or follow-up time (n = 1,595), or had CFS in 2004 (n = 760), leaving a final sample of 4,212 participants. Altogether, 4,212 individuals without a history of CFS at baseline in 2004 were included to determine the causal factors associated with the 435 individuals who developed CFS by 2011 (Figure 1). The study protocol was approved by the Ethics Committee of Macao Polytechnic University (no. CI237/DEI/2022). Written or electronic informed consent was obtained from all the participants during the survey.

Outcome ascertainment

According to the US Centers for Disease Control and Prevention criteria, CFS is defined as extreme exhaustion lasting >6 months and accompanied by at least one type of pain[1]. Participants responded "approximately once a week" or "daily or more often" to the question "How often have you had fatigue/exhaustion in the past 6 months?" were considered potential CFS cases. Confirmation required reporting both regular fatigue and at least one form of pain (hand/wrist pain, ankle/knee pain, neck/shoulder pain, headache, joint swelling, chest discomfort, foot pain, back pain/strain, or bone pain).

Exposure variables for causal graph modeling

Categorical data extracted from the 2004 WLS study included 35 potential exposure variables across five domains: basic information, lifestyle, pre-existing diseases, pre-existing pain, and viral infection symptoms. Among them, marital status was classified as currently married, single/separated, divorced, or widowed[21]. Body mass index (BMI) was categorized as <25, 25 to 30, or ≥30. Education level was classified as attended/never attended college. Pre-existing diseases were reported as yes/no. Pre-existing pain and viral infection symptoms were defined as follows: no pain, monthly pain or less, approximately once a week, daily, or more. The continuous variables included income, age, light exercise time, vigorous exercise time, BMI, number of restaurant visits, number of diagnoses, medications taken, and alcohol intake [Supplementary Table S1, http://links.lww.com/AHM/A96].

Causal graph modeling

To enhance the accuracy of identifying CFS-associated factors, causal graphical models were constructed separately for the six variable groups [Supplementary Table S1, http://links.lww.com/AHM/A96]. The FCI and FGES algorithms were used to construct causal graphs using the Tetrad software package (https://github.com/cmu-phil/tetrad). For the FCI algorithm, the default settings were applied for the depth parameter (maximum size of the conditioning set) and maxPathLength parameter (maximum length for any discriminating path), imposing no restrictions. Similarly, in the FGES algorithm, the maximum degree of the graph was left unrestricted. The application of these two complimentary algorithms allowed us to leverage their respective strengths to increase confidence in the determination of CFS-associated factors.

Family cohort validation

Individuals from the same family (including full siblings, twins, and those with a common father or mother) were screened from among the 4,212 participants to further investigate whether genetic background could be a potential confounding variable. A logistic regression analysis was performed on 186 individuals from 93 families (two individuals per family, one with and one without CFS).

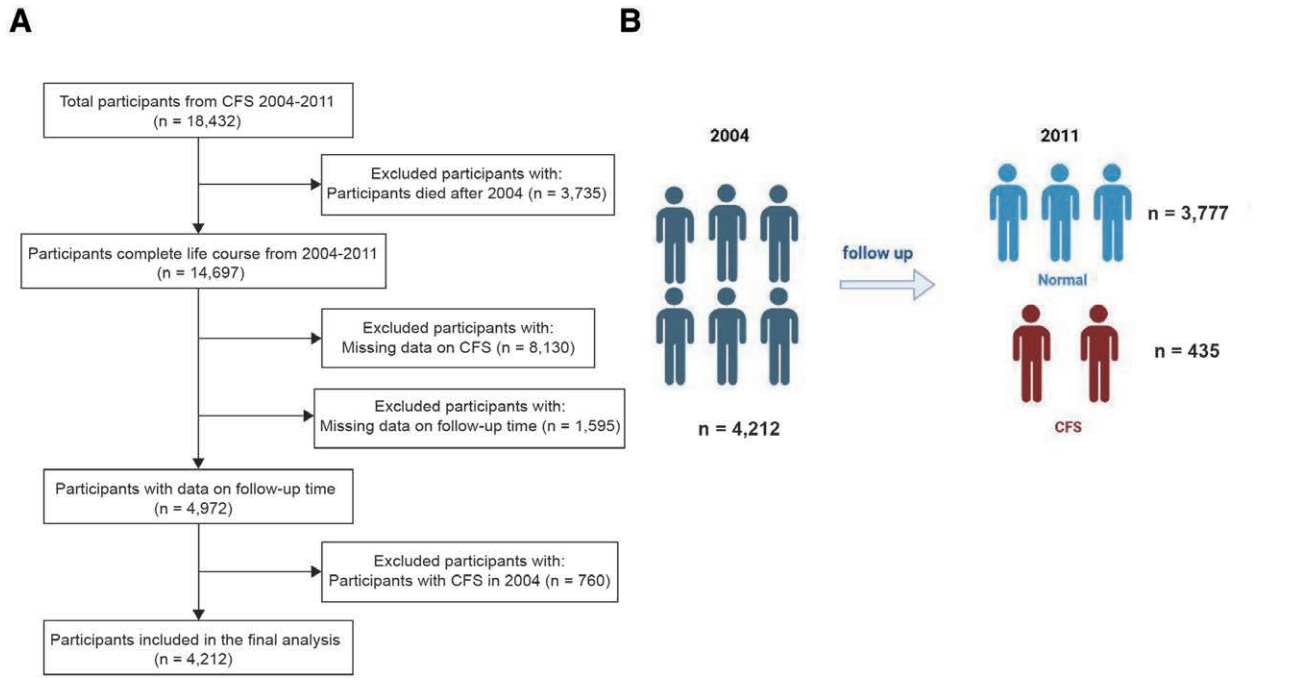

Figure 1. Flowchart illustrating participant inclusion in this study (A) and a schematic representation of the prospective cohort study design (B). CFS: Chronic fatigue syndrome.

MR analysis

In the MR analysis, genetically predicted circulating proteins from nine proteomic genome-wide association studies (GWASs) were treated as the exposures, and two CFS risk factors (cough and insomnia) were treated as the outcomes. The detailed data of the nine studies are presented in Supplementary Table S2, http://links.lww.com/AHM/A96, and public databases for outcomes were accessible from the IEU GWAS database (https://gwas.mrcieu.ac.uk/), with two databases used for each outcome. The cis-pQTLs in single-nucleotide polymorphisms (SNPs) were used as candidate instrumental variables because they have a more direct and specific biological effect on protein levels than trans-pQTLs[22]. SNPs were selected according to the following requirements. First, SNPs were associated with any protein at the genome-wide significance level in the corresponding study [Supplementary Table S2, http://links.lww.com/AHM/A96]. Second, SNPs within the human major histocompatibility complex region (chr6: from 26 to 34 Mb) were excluded due to complex linkage disequilibrium (LD) patterns in this region. Lastly, SNPs were clumped to avoid LD using a strict window threshold ($r^2 > 0.01$ within a clumping window of 5,000 kb).

In the main analysis, the inverse-variance weighted method was used for primary analysis to estimate MR effects between plasma proteins and CFS risk factors. In addition, the Wald ratio method was applied when only a single pQTL was available.

To avoid potential heterogeneity and horizontal pleiotropy among instruments, we conducted Cochran's *Q* test and MR–Egger intercept analysis as further sensitivity analyses. If heterogeneity existed, the weighted median method was employed; similarly, if horizontal pleiotropy was identified, MR–Egger was performed to correct the pleiotropy among the genetic instruments[23].

Protein structure prediction, preparation, and structural-based virtual screening

The structure of the endothelial cell-selective adhesion molecule (ESAM; UniProt ID: Q96AP7) identified by MR analysis was predicted using AlphaFold2 accessed through the EBI repository (https://AlphaFold.ebi.ac.uk/entry/Q96AP7). The predicted structure was then imported into the Schrödinger Maestro software (New York, USA), and any missing side chains were reconstructed with Prime in the Protein Preparation Wizard. The protonation states of the titratable residues were assigned using PROPKA at pH 7.0. Furthermore, hydrogen bonding optimization and restrained minimization were conducted considering a root-mean-square deviation cutoff value of 0.5 Å using the OPLS3 (Optimized Potentials for Liquid Simulations) force field. All the ligands were structurally prepared, and up to eight tautomers and four stereoisomers were generated per tautomer at pH 7.0 ± 0.4 using Schrödinger LigPrep and OPLS3 energy minimization.

Protein–ligand docking was performed using the Schrödinger's Glide software (New York, USA). The standard-precision Glide mode was selected to provide more accurate scoring of ligand binding affinities. The 163 previously identified bioactive compounds from QJYQ were used as ligands[18]. We utilized the HDOCK server (http://hdock.phys.hust.edu.cn/) to analyze protein–protein interactions. To ensure accuracy, the top poses were selected based on a combination of scoring metrics and known biological data.

Statistical analysis

Descriptive statistics were used to analyze the characteristics and distribution of the populations. Categorical variables are presented as numbers and proportions. Continuous variables are presented as medians and interquartile ranges (IQRs) as they were non-normal distributions. All continuous variables were normalized using the *z* score before analysis. A two-proportional *z* test was used to compare the CFS and non-CFS groups across variables.

The potential causal factors identified from the FCI and FGES analyses were evaluated using forward and backward stepwise multivariate logistic regression, and the 95% confidence intervals (CIs) were calculated.

All the analyses were conducted using the Tetrad software package[24], SAS software version 9.4, GraphPad Prism 9.5 (San Diego, CA, USA), "TwoSampleMR" package (http://github.com/MRCIEU/TwoSampleMR) in R (version 4.1.2), and PyMOL (PyMOL Molecular Graphics System, Version 2.5.2; Schrödinger, LLC, New York, USA). *P* values were Bonferroni-adjusted to account for multiple testing, and a two-sided $P < 0.05$ was considered to indicate statistical significance.

Results

Population characteristics

The baseline characteristics of the 4,212 participants are presented in Table 1. The median age was 69 (IQR: 69–70) years, and data for 435 patients with CFS were recorded. Participants resided in Wisconsin (65.5%) or elsewhere (34.5%). The cohort included 1,999 (47.5%) men and 2,213 (52.5%) women. The median age and BMI were 69 (69, 70) years and 27 (24, 30), respectively. Most were currently married (81.7%), and almost half (50.2%) had attended college. High blood pressure and high cholesterol were the most prevalent conditions, accounting for 42.7% and 44.9%, respectively. Allergic reactions were reported by 29.6% of the individuals, which is a relatively high proportion. Only 1.4% of individuals had experienced a stroke. The prevalence of other diseases ranged from 7% to 10%. The incidence of each type of pain symptom was very similar, with 36% to 46% of patients experiencing some form of pain symptom within 6 months. The distributions of sociodemographic characteristics in 2011 are presented in the supplemental materials [Supplementary Table S3, http://links.lww.com/AHM/A96].

Causal graph modeling reveals the potential causal factors linked to CFS

The FCI algorithm identified 15 variables that may ultimately lead to CFS (Figure 2) as follows: BMI and sex in

Table 1

Baseline characteristics of the 4,212 participants enrolled in the study in 2004

Variables	Baseline
Age, median (IQR: 25%, 75%)	69 (69, 70)
Male, n [%]	1,999 [47.5]
Marital status, n [%]	
Currently married	3,441 [81.7]
Single/separated	146 [3.5]
Divorced	350 [8.3]
Widowed	275 [6.5]
Race, n [%]	
White	3,927 [93.2]
Black/African American	75 [1.8]
Asian	7 [0.2]
Black/African American and American Indian	40 [1.0]
Missing	163 [3.9]
Education, n [%]	
Attended college	2,116 [50.2]
Never attended college	2,096 [49.8]
Child, median (IQR: 25%, 75%)	3 (2,4)
BMI, median (IQR: 25%, 75%)	27 (24,30)
Residence, n [%]	
Foreign Country	1,453 [34.5]
Wisconsin	2,759 [65.5]
Trouble sleeping, n [%]	
Have not had	2,026 [48.1]
Monthly or less often	1,154 [27.4]
About once a week	806 [19.1]
Daily or more often	226 [5.4]
Do light exercise, n [%]	
Never	881 [20.9]
Rarely	1,286 [30.5]
Often	1,584 [37.6]
Do vigorous exercise, n [%]	
Never	1,344 [31.9]
Rarely	1,159 [27.5]
Often	1,248 [29.6]
Go to restaurant/bar frequency, n [%]	
Never	2,335 [55.4]
Rarely	1,668 [39.6]
Often	41 [1.0]
Smoke, n [%]	
Never smoke	1,967 [46.7]
Ever smoke	1,827 [43.4]
Currently smoke	418 [9.9]
Drink alcohol, median (IQR: 25%, 75%)	6 (1, 20)
Income, median (IQR: 25%, 75%), thousand (dollar)	52.4 (28.8, 87.2)
Light exercise time, median (IQR: 25%, 75%), hours	20 (8, 35)
Vigorous exercise time, median (IQR: 25%, 75%), hours	0 (0, 8)
Going to restaurant/bar time, median (IQR: 25%, 75%), hours	6 (3, 10)
Diagnosed illnesses number, median (IQR: 25%, 75%)	1 (1, 2)
Prescriptions taken number, median (IQR: 25%, 75%)	2 (0, 3)
Have diabetes, n [%]	498 [11.8]
Have heart attack, n [%]	436 [10.4]
Have high blood pressure, n [%]	1,797 [42.7]
Have stroke, n [%]	59 [1.4]
Have circulation problem, n [%]	242 [5.8]
Have asthma, n [%]	302 [7.2]
Have irritable bowel syndrome, n [%]	312 [7.4]
Have allergy, n [%]	1,245 [29.6]
Have high cholesterol, n [%]	1,889 [44.9]
Hands/wrists pain, n [%]	
Have not had	2,681 [63.7]
Monthly or less often	813 [19.3]
About once a week	368 [8.7]
Daily or more often	350 [8.3]
Neck/shoulder pain, n [%]	
Have not had	2,268 [53.9]
Monthly or less often	1,140 [27.1]
About once a week	456 [10.8]
Daily or more often	348 [8.3]
Ankles/knees pain, n [%]	
Have not had	2,362 [56.1]
Monthly or less often	894 [21.2]
About once a week	474 [11.3]
Daily or more often	482 [11.4]
Headache, n [%]	
Have not had	2,386 [56.7]
Monthly or less often	1,488 [35.3]
About once a week	286 [6.8]
Daily or more often	52 [1.2]
Stiff/swollen joints, n [%]	
Have not had	2,644 [62.8]
Monthly or less often	801 [19.0]
About once a week	351 [8.3]
Daily or more often	416 [9.9]
Palpitations, n [%]	
Have not had	3,536 [84.0]
Monthly or less often	544 [12.9]
About once a week	105 [2.5]
Daily or more often	27 [0.6]

(*Continued*)

Table 1

(Continued)

Variables	Baseline
Excessive sweating, n [%]	
Have not had	3,538 [84.0]
Monthly or less often	377 [9.0]
About once a week	150 [3.6]
Daily or more often	147 [3.5]
Ringing in the ears, n [%]	
Have not had	3,048 [72.4]
Monthly or less often	466 [11.1]
About once a week	176 [4.2]
Daily or more often	521 [12.4]
Shortness of breath, n [%]	
Have not had	3,421 [81.2]
Monthly or less often	577 [13.7]
About once a week	151 [3.6]
Daily or more often	63 [1.5]
Coughing/wheezing, n [%]	
Have not had	3,067 [72.8]
Monthly or less often	787 [18.7]
About once a week	175 [4.2]
Daily or more often	183 [4.3]
Numbness, n [%]	
Have not had	3,453 [82.0]
Monthly or less often	420 [10.0]
About once a week	166 [3.9]
Daily or more often	164 [3.9]
Upset stomach, n [%]	
Have not had	3,204 [76.1]
Monthly or less often	821 [19.5]
About once a week	158 [3.8]
Daily or more often	29 [0.7]
Dizziness/faintness, n [%]	
Have not had	3,659 [86.9]
Monthly or less often	493 [11.7]
About once a week	48 [1.1]
Daily or more often	12 [0.3]

IQR: Interquartile range. Continuous variables are either the mean ± standard deviation or median and the (IQR: 25%, 75%) depending on the data distribution. Categorical data are expressed as *n* [%].

Group A; trouble sleeping and light exercise in Group B; heart attack, circulation issues, and irritable bowel syndrome in Group C; ankles/knees pain and neck/shoulder pain in Group D; shortness of breath, excessive sweating, numbness, and cough in Group E; and number of medications taken, diagnosed illnesses, and BMI in Group F. Edges with round circle ends in Figure 2 indicate the presence of potential unmeasured confounders associated with these variables [definitions of causal graphs are provided in Supplementary Table S4, http://links.lww.com/AHM/A96].

Figure 3 presents the models from the FGES algorithm, revealing seven CFS-associated variables as follows: BMI in Group A, trouble sleeping in Group B, circulation problems in Group C, ankles/knees pain in Group D, shortness of breath in Group E, and number of medications taken, diagnosed illnesses, and BMI in Group F. The exposure variables that have arrows pointing toward CFS in Figure 3 represent potential causal factors for CFS, either directly or indirectly through the definition of causal graphs in Supplementary Table S4, http://links.lww.com/AHM/A96.

Stepwise multivariate logistic regression

To explore the robustness of the results from causal graphical models, forward and backward stepwise multivariate logistic regression analyses were conducted. According to the unadjusted model (Figure 4A), the number of medications, coughing/wheezing, shortness of breath, ankle/knee pain, neck/shoulder pain, circulation problems, trouble sleeping, light exercise, BMI, and sex were significantly associated with an increased risk of CFS.

In the adjusted model (Figure 4B), the number of medications, coughing/wheezing, shortness of breath, ankle/knee pain, circulation problems, trouble sleeping, light exercise, and BMI remained statistically significant. Daily sleep troubles (odds ratio [OR]: 2.59, 95% CI: 1.77–3.78), cough (OR: 1.74, 95% CI: 1.15–2.63), and ankle/knee pain (OR: 1.77, 95% CI: 1.30–2.41) increased the risk of CFS, whereas weekly shortness of breath (OR: 2.07, 95% CI: 1.34–3.20) decreased such risk. Frequent light exercise (OR: 0.59, 95% CI: 0.45–0.79) also lowered the risk of CFS. However, circulation problems (OR: 1.68, 95% CI: 1.17–2.41), increased number of medications taken (OR: 1.11, 95% CI: 1.06–1.17), higher BMI (OR: 1.03, 95% CI: 1.00–1.05), and the female sex (OR: 1.49, 95% CI: 1.19–1.87) significantly increased the risk of CFS.

Influence of genetic background on potential causal factors

To investigate the potential genetic confounding factors for CFS, a family cohort validation was performed. According to the unadjusted model (Figure 5A), the number of medications taken and the presence of cough were significantly associated with an increased risk of CFS in the family cohort with similar genetic backgrounds.

According to the adjusted model (Figure 5B), these two variables maintained statistical significance. Daily or more frequent coughing (OR: 4.05, 95% CI: 1.14–14.38) and more medications taken (OR: 1.21, 95% CI: 1.03–1.42) increased the likelihood of CFS.

Identification of causal factor-associated proteins as drug targets of CFS

After processing nine plasma databases, 8,285 pQTLs for 4,421 proteins (2,518 unique proteins) remained; among them were 3,811 cis-pQTLs for 2,958 proteins (1,558 unique proteins) as instruments for MR analysis [Supplementary Table S5, http://links.lww.com/AHM/

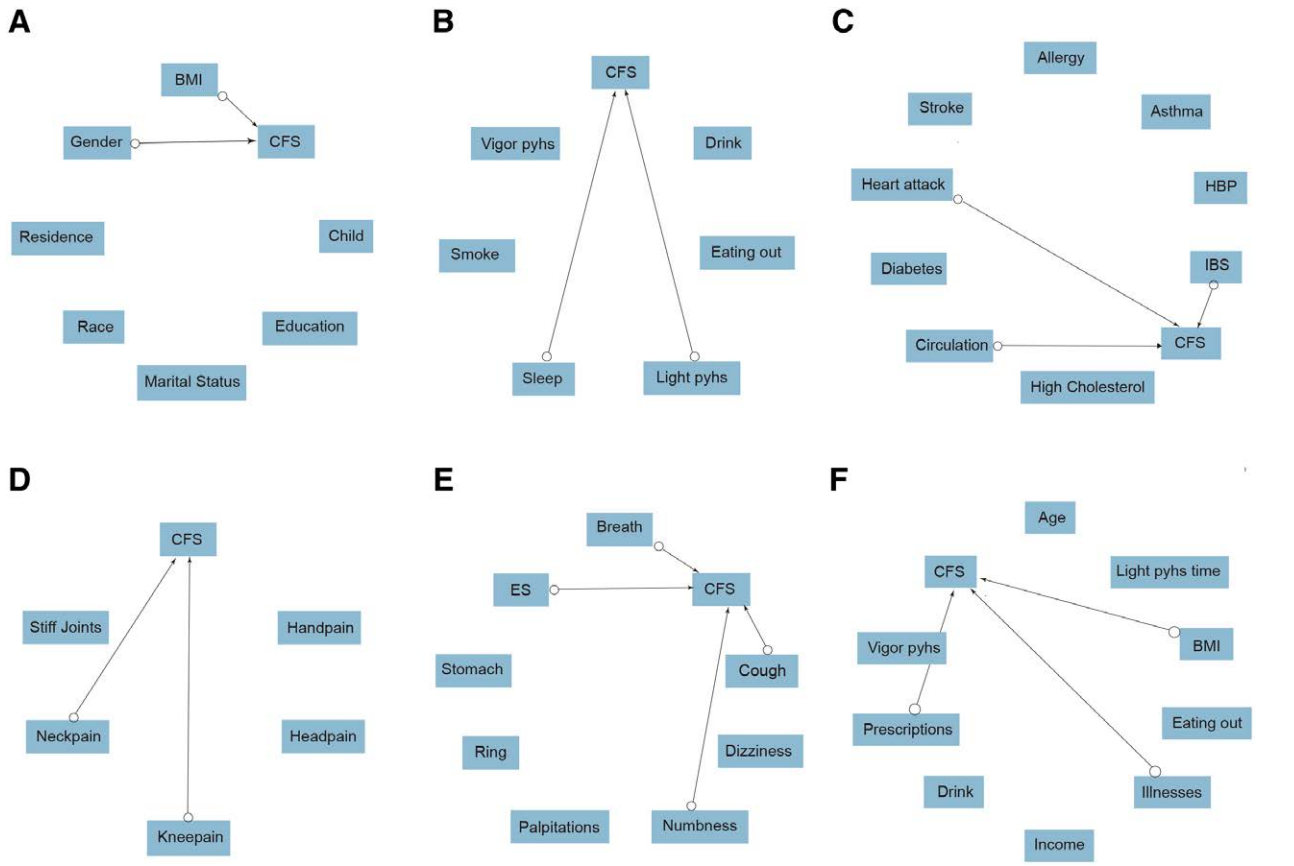

Figure 2. Causal graph models constructed by the FCI algorithm. The arrows in the figure indicate the possible direct or indirect causal relationships between the variables and CFS. An arrow pointing to the CFS from a variable suggests that the variable may contribute to the development or exacerbation of CFS. Additionally, the circles at the end of the arrows represent potential unmeasured confounding factors that may be associated with the causal relationships. BMI, Body mass index; BP, High blood pressure; Breath: Shortness of breath; CFS, Chronic fatigue syndrome; ES, Excessive sweating; FCI: Fast causal inference; IBS, Irritable bowel syndrome; Illnesses: Diagnosed illnesses; Prescriptions: Number of prescriptions; Ringing: Ringing in ears; Sleep, Trouble sleeping; Stomach: Upset stomach.

A96]. Based on the causal diagram analysis, we considered the CFS-associated factors of cough and insomnia as outcomes for MR. Overall, 19 protein-cough associations reached the Bonferroni-corrected threshold for 12 proteins, and 28 proteins were significantly associated with insomnia. The Bonferroni-corrected thresholds of the MR analyses were set separately for each outcome [Supplementary Table S6, http://links.lww.com/AHM/A96].

The primary results of the main MR analysis for chronic cough and insomnia are presented in Figure 6 and Supplementary Tables S7–S10, http://links.lww.com/AHM/A96. Genetically determined higher levels of ESAM (UniProt: Q96AP7) (OR: 1.020, 95% CI: 1.012–1.028) and lactase/phlorizin hydrolase (UniProt: P09848) (OR: 1.007, 95% CI: 1.004–1.009) increased chronic cough risk. Similarly, higher levels of interleukin-1 receptor type 2 (UniProt: P27930) (OR: 1.001, 95% CI: 1.001–1.002), HEXIM2 (UniProt: Q96MH2) (OR: 1.154, 95% CI: 1.114–1.196), interleukin-9 (UniProt: P15248) (OR: 1.063, 95% CI: 1.044–1.081), and inactive tyrosine-protein kinase 7 (UniProt: Q13308) (OR: 1.095, 95% CI: 1.056–1.136) were significantly associated with an increased risk of insomnia. Moreover, we determined that the protein ESAM was associated with both cough and insomnia.

The sensitivity analyses validated all significant associations originally identified in the main analyses, whereas two associations with MR evidence were newly identified [Supplementary Table S11, http://links.lww.com/AHM/A96].

Virtual screening analysis revealed potential pharmacological mechanisms of action of QJYQ for treating CFS

Of the 163 phytochemical compounds identified in QJYQ, ginsenoside RG4 and ginsenoside Rb2 exhibited good binding affinity with the ESAM protein. Interaction analysis and visualization of two protein–ligand complexes revealed intricate binding modes. The interaction details for each complex are provided in the following paragraphs.

Ginsenoside RG4 (docking score: –6.15)

The ligand forms hydrophobic interactions with Pro36, Val43, Glu47, Pro51, and Trp53 (3.48–3.95 Å). Gln41 and Glu47 form multiple hydrogen bonds, indicating strong, dynamic binding. Additional hydrogen bonds with Val43 and Arg112 confer specificity (Figure 7A).

Ginsenoside Rb2 (docking score: –6.03)

This ligand makes hydrophobic contacts with Arg178 and Lys180. Multiple hydrogen bonds with Arg39,

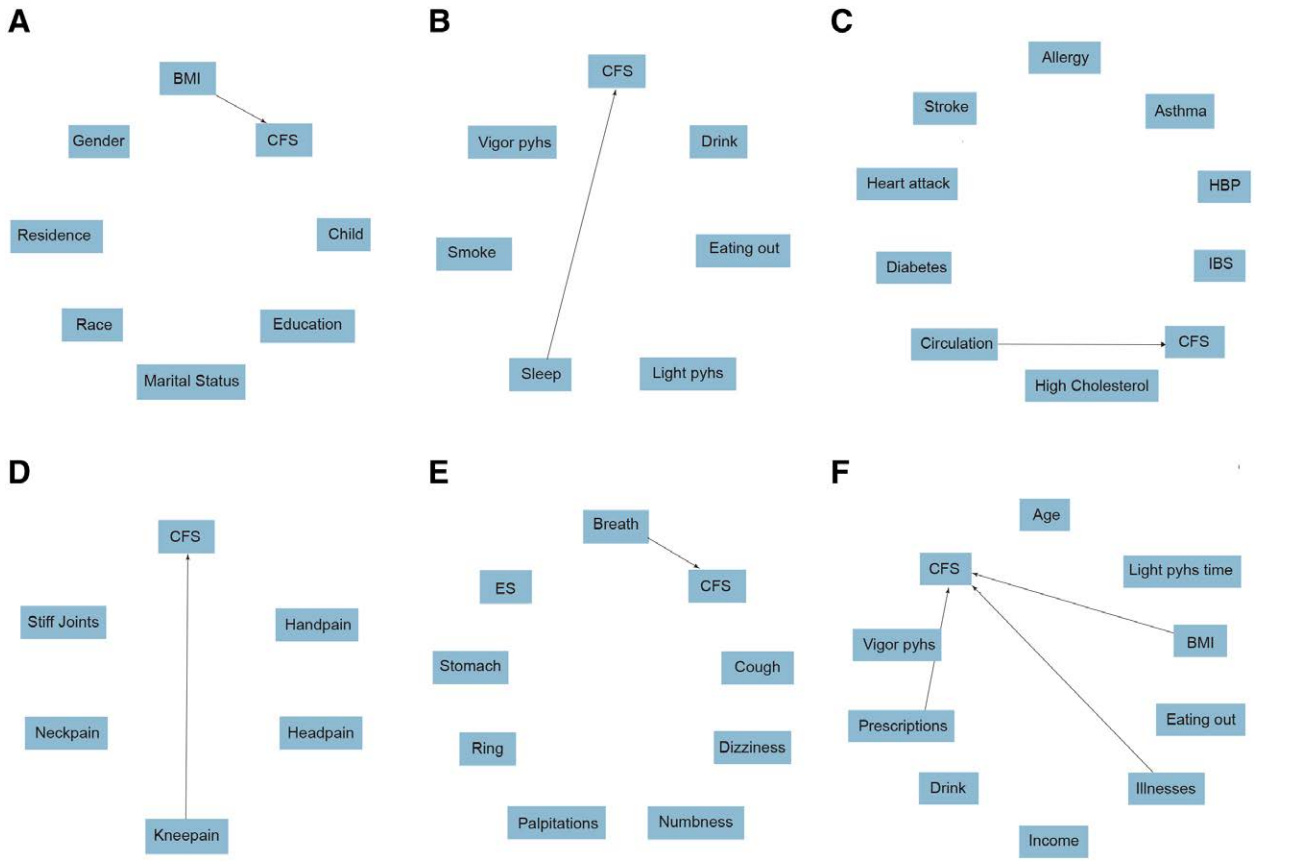

Figure 3. Causal graph models constructed by the FGES algorithm. The arrows in the figure indicate the presence of either direct or indirect causes of CFS from the variables. Specifically, an arrow pointing to the CFS from a variable suggests that the variable may directly or indirectly contribute to the development or exacerbation of CFS. BMI, Body mass index; Breath: Shortness of breath; CFS: Chronic fatigue syndrome; ES: Excessive sweating; FGES: Fast greedy equivalence search; HBP, High blood pressure; IBS: Irritable bowel syndrome; Illnesses: Diagnosed Illnesses Number; Prescriptions: Number of prescriptions; Ringing: Ringing in ears; Sleep, trouble sleeping; Stomach: Upset stomach.

Glu146, Asn148, Gln175, Ser176, Ser179, and Tyr185 (1.92–4.01 Å) demonstrate stable, targeted ESAM binding, with Asn148 forming critical interactions (Figure 7B).

Discussion

In this prospective cohort study, we identified that chronic cough and insomnia were causally associated with CFS. Additionally, the MR analysis revealed that the upregulation of the endothelial protein ESAM was causally linked with chronic cough and insomnia pathogenesis. Furthermore, we identified potential mechanisms of action of ginsenoside on target proteins, highlighting the potential biological therapeutic pathway of QJYQ for treating CFS.

In our study, chronic cough was identified as a causal factor contributing significantly to the development of CFS. The international consensus defines cough based on duration, with acute cough lasting up to 8 weeks and chronic cough exceeding 8 weeks[25]. Prolonged cough may be related to persistent lung inflammation from various etiologies, such as chronic obstructive pulmonary disease, bronchial asthma, and cough variant asthma[26]. This sustained inflammation is hypothesized to deplete energy and cause fatigue. Moreover, enduring coughing has the potential to increase mental stress by disrupting daily activities and exacerbating fatigue. Recent evidence indicates that TCMs such as QJYQ are promising in treating PCC conditions marked by chronic cough, thereby potentially preventing the progression to CFS[27].

Other significant risk factors include sleep disturbances and shortness of breath. Difficulty falling asleep, interrupted sleep, or light sleep prevents adequate restorative rest and recovery, potentially resulting in chronic fatigue; shortness of breath, an intrinsic symptom of breathing difficulty; often causes discomfort, restricts activity, and diminishes quality of life[28]; regular practices such as progressive muscle relaxation and deep breathing exercises effectively reduce breathing issues and fatigue symptoms[29]; ankle and knee pain commonly accompany myalgia sufferers, as research has demonstrated that physical fatigue is positively associated with augmented pain severity[30]; and circulation issues may also impact fatigue. Both systemic and pulmonary circulations appropriately respond to localized oxygen deficiency, the former through vasodilation to provide more oxygen and the latter by constriction, diverting circulation[31]. However, disruptions to circulation can induce insufficient or abnormal red blood cells, limiting the oxygen supply and potentially causing fatigue. In addition, previous research has demonstrated that women have a greater incidence of CFS than men[32].

The MR analysis revealed several notable findings related to chronic cough and insomnia. We observed that 19 proteins were significantly associated with chronic

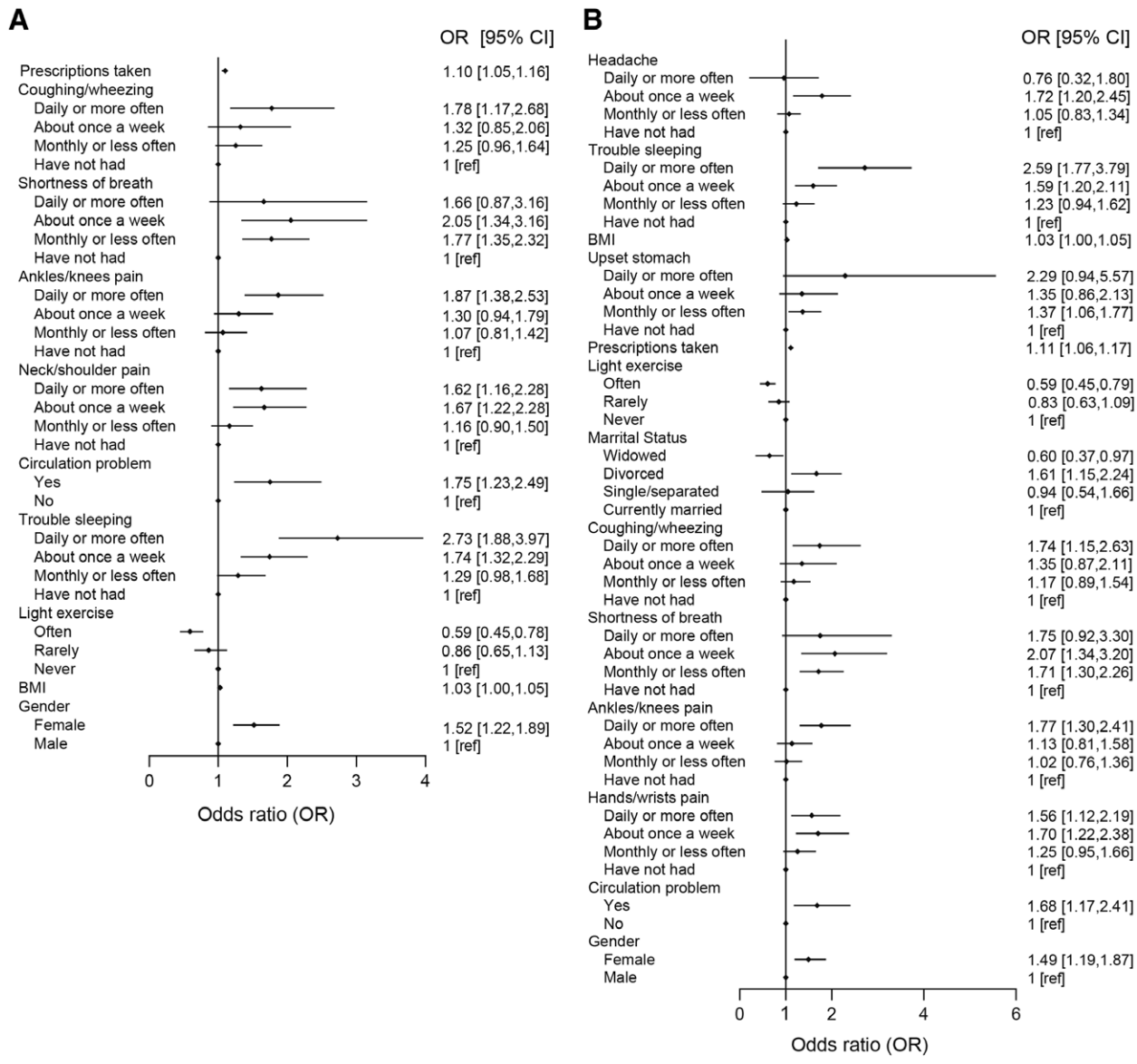

Figure 4. Stepwise multivariate logistic regression validating the associations between CFS and the potential causal factors identified by the FCI algorithm (A), as well as all 43 variables included in this study (B). An odds ratio was considered to be statistically significant if the null value of 1 was excluded from the 95% CI range. Only significant variables are shown in the forest plots. CFS: Chronic fatigue syndrome; CI: Confidence interval; FCI: Fast causal inference.

cough, and 28 proteins were significantly associated with insomnia, which suggested that specific proteins may play a role in the development or exacerbation of these symptoms in patients with CFS. ESAM scores demonstrated causal associations with both chronic cough and insomnia. Previous studies have implicated ESAM in schizophrenia[33] and cardiovascular diseases[34]. The current findings indicate that ESAM may also represent a shared biological mechanism impacting multiple symptoms of CFS.

Furthermore, exploring the molecular interactions between the main bioactive compounds ginsenosides and ESAM provides insights into the potential therapeutic mechanisms of QJYQ for preventing PCC CFS. The complexes exhibit various hydrophobic contacts, hydrogen bonds, and salt bridges that support flexible and adaptable ligand binding modes. This enables the formation of stable ligand–protein complexes, the key mechanism underlying the therapeutic function of ginsenosides in QJYQ.

However, this study has some limitations. First, although potential CFS factors were identified, their precise impacts warrant validation. Second, although rigorous statistical techniques were employed, unrecorded uncertainties in longitudinal observational data spanning 2004 to 2011 could influence the causal inferences drawn. Third, further validation using cell culture and animal models is necessary to confirm the potential pharmacological effects of QJYQ on ESAM and its efficacy in treating CFS. Finally, the data analysis was limited to data in the WLS database and only to 2011. In future studies, we aim to elucidate these findings by replicating our analyses using more recent longitudinal data from other databases.

The important strengths of our study were as follows. To the best of our knowledge, this research represents the first application of causal graph learning to a well-powered prospective cohort for identifying risk factors for CFS. The temporal sequence afforded by

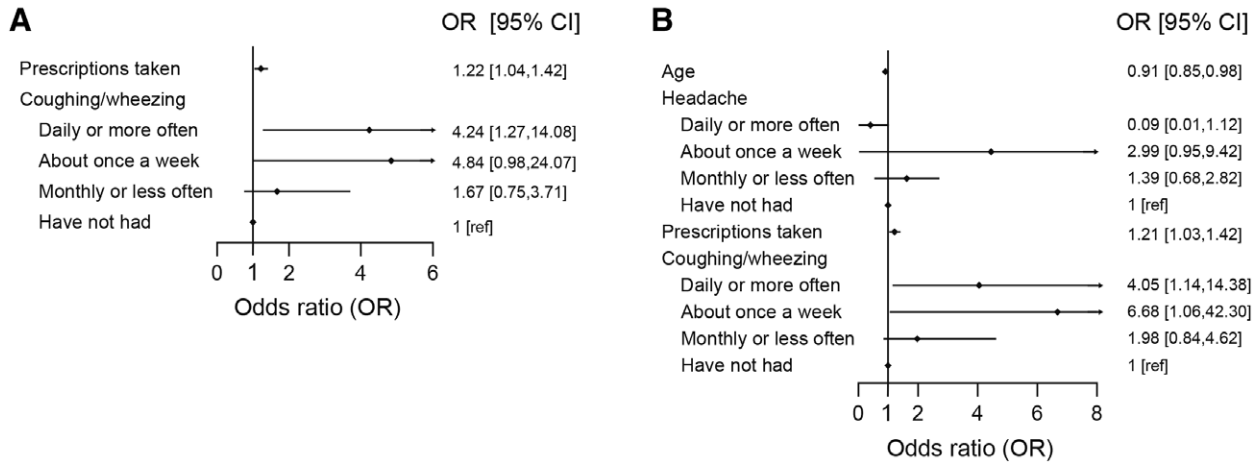

Figure 5. Stepwise multivariate logistic regression analysis of the family cohort with similar genetic backgrounds was performed to exclude genetic influences on the causal relationships between CFS and the identified potential causal factors. (A) The associations between the CFS score and potential causal variables identified by the FCI algorithm in the family cohort. (B) Associations between the CFS score and all 43 variables in the family cohort. The forest plot displays the ORs and 95% CIs for each variable. Only significant variables for which the null value of 1 was excluded from the 95% CI range are shown. CFS: Chronic fatigue syndrome; CIs: Confidence intervals; FCI: Fast causal inference; ORs: Odds ratios.

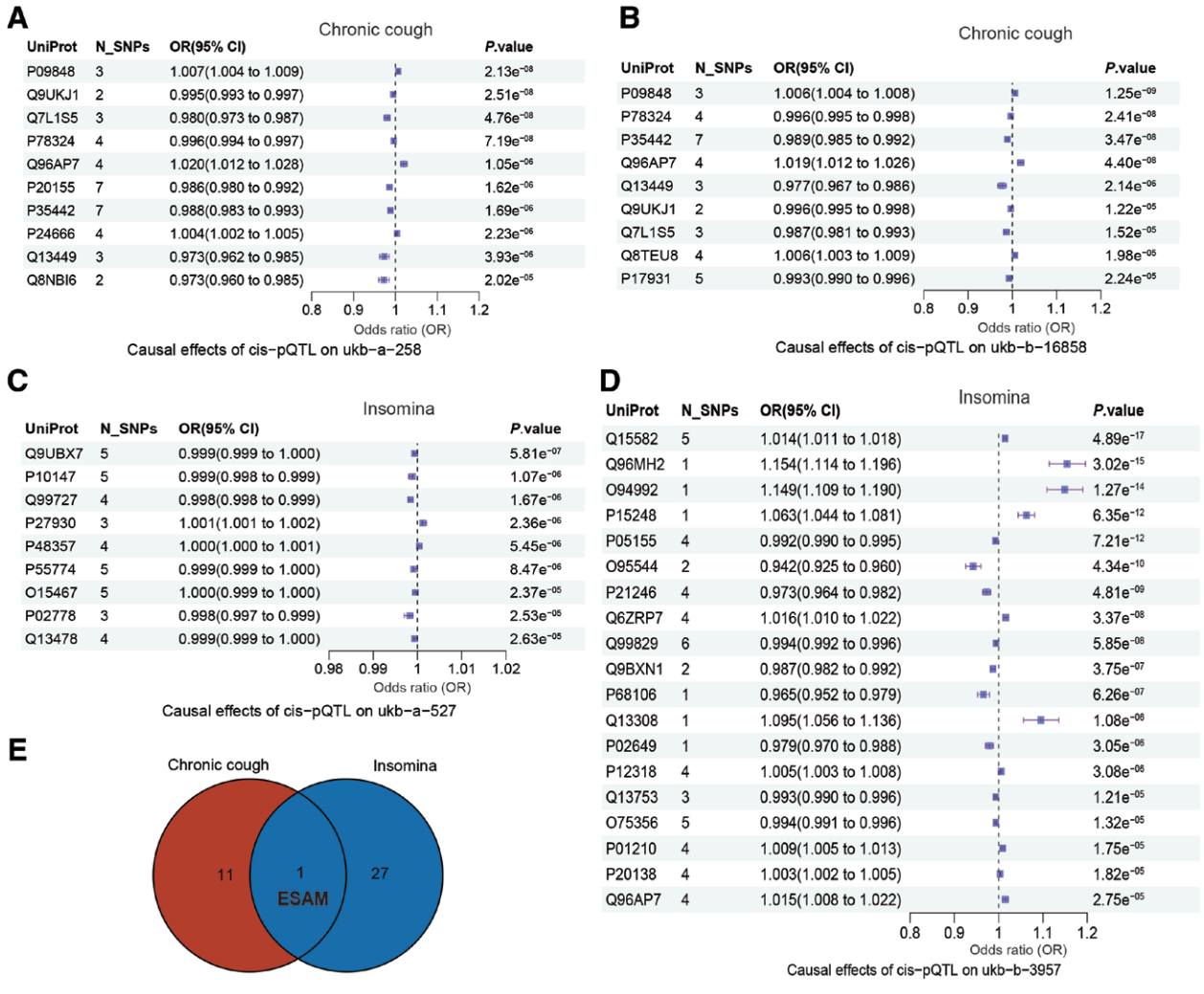

Figure 6. Proteome-wide MR analysis identifying circulating proteins causally associated with chronic cough and insomnia, the symptoms linked to CFS, by causal graphical modeling. (A–B) Proteins causally associated with chronic cough based on the GWAS IDs ukb-a-258 (A) and ukb-b-16858 (B). (C–D) Proteins causally linked to the outcome of insomnia based on GWAS IDs ukb-a-527 (C) and ukb-b-3957 (D). (E) ESAM was identified as the shared causal protein for both chronic cough and insomnia, two potential symptoms contributing to CFS. N_SNPs: number of SNPs used for the estimation of the causal effects in this plot. *P* values were determined using the IVW or Wald ratio two-sample MR method. CFS: Chronic fatigue syndrome; CIs: Confidence intervals; ESAM: Endothelial cell-selective adhesion molecule; IVW: Inverse-variance weighting; MR: Mendelian randomization; ORs: Odds ratios; SNP: Single-nucleotide polymorphism.

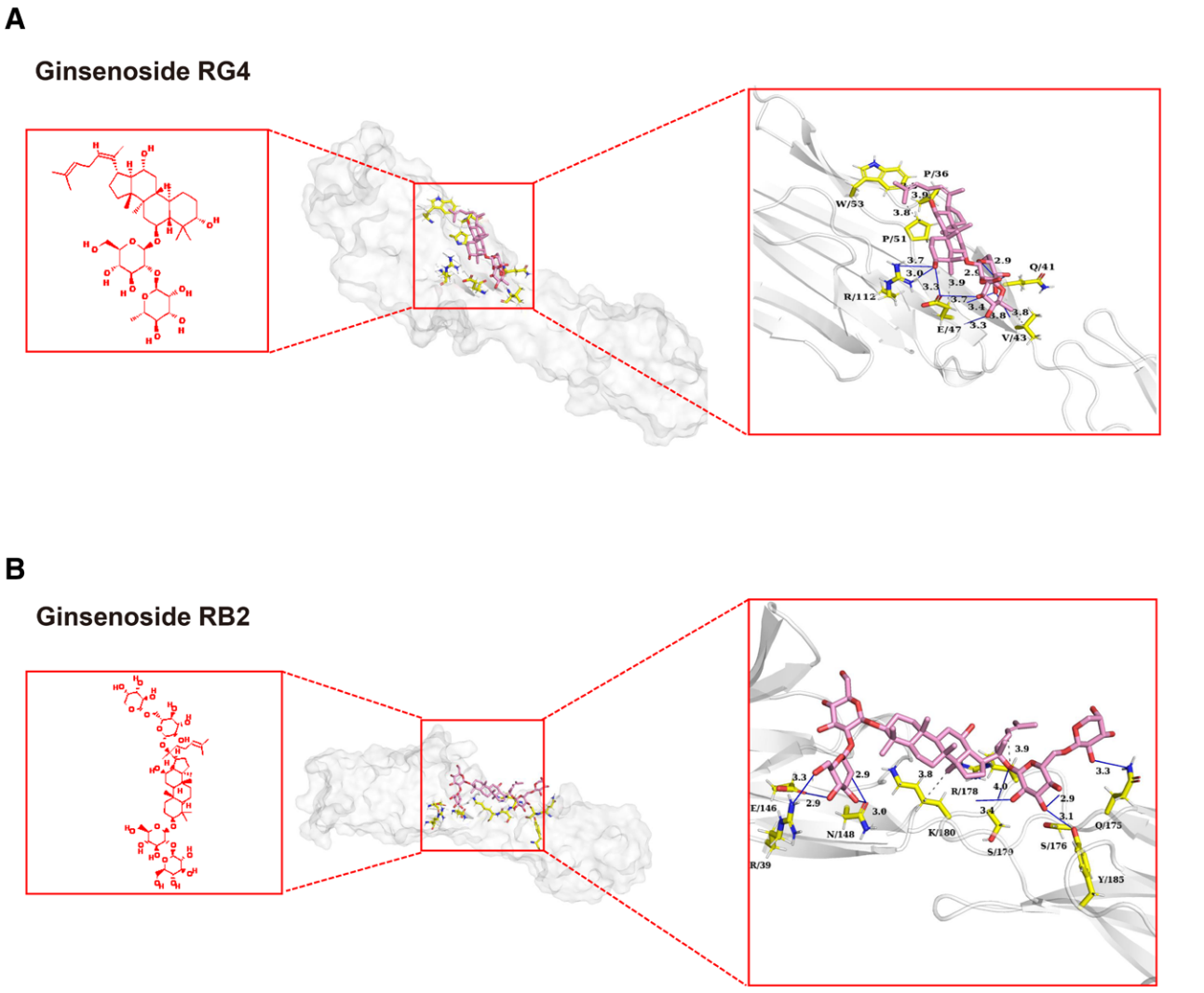

Figure 7. The molecular interactions between the bioactive compounds in QJYQ granules and ESAM, a protein causally linked to CFS development. (A) Binding mode between the QJYQ ingredient ginsenoside RG4 and ESAM (docking score: –6.15). (B) Binding mode between the main bioactive phytochemical ginsenoside Rb2 in QJYQ and ESAM (docking score: –6.03). CFS: Chronic fatigue syndrome; ESAM: Endothelial cell-selective adhesion molecule; QJYQ: Qingjin Yiqi.

the longitudinal design provides more robust evidence than could be attained through cross-sectional analyses alone. Additionally, complementary statistical modeling approaches were integrated to enhance interpretation and corroborate findings from causal graphical modeling. Leveraging epidemiological, genetic, and computational modeling within an integrative framework presented a novel strategy for elucidating the complex etiology of CFS and illuminating the pharmacology of TCM such as QJYQ.

Conclusion

This integrative epidemiological and computational study provides novel insights into the pathogenesis and management of CFS. Causal graph modeling identified chronic cough and insomnia as modifiable causal risk factors for CFS. The endothelial protein ESAM was identified as a potential shared drug target causally linked to the CFS risk factors for chronic cough and insomnia. QJYQ may prevent CFS development through the inhibition of ESAM through the bioactive ingredients ginsenoside RG4 and ginsenoside Rb2. Our study has important clinical and public health implications for preventing and treating CFS using TCM. Additionally, this study provides a new analytical framework for elucidating the pharmacological mechanisms of action of TCMs through the integration of epidemiological, genetic, and *in silico* data.

Conflict of interest statement

The authors declare no conflict of interest.

Funding

This study was supported by an internal fund from Macao Polytechnic University (RP/FCSD-02/2022).

Author contributions

Junrong Li, Xiaobing Zhai, Shu Li, and Kefeng Li conceptualized and designed the study. Junrong Li, Xiaobing Zhai, Chi Kin Lam, and Weiyu Meng collected and analyzed data, as well as prepared the figures and tables.

Junrong Li, Jixing Liu, and Xiaobing Zhai drafted the manuscript. Chi Kin Lam, Weiyu Meng, Yuefei Wang, Shu Li, Yapeng Wang, and Kefeng Li provided critical review and editing of the manuscript. Kefeng Li and Yapeng Wang secured funding for the research. All authors reviewed and approved the final version of the manuscript.

Ethical approval of studies and informed consent

The study protocol was approved by the Ethics Committee of Macao Polytechnic University (no. CI237/DEI/2022). Written or electronic informed consent was obtained from all the participants during the survey.

Acknowledgments

We thank the participants and staff of the Wisconsin Longitudinal Study (WLS), whose contributions have made this analysis feasible.

Data availability statement

Data from the WLS are available on the application at https://wls.wisc.edu/. Codes of WLS survey for extracting the variables used in this study are listed in Table S1, http://links.lww.com/AHM/A96. For the MR analysis, details of the plasma protein studies are provided in Table S2, http://links.lww.com/AHM/A96. GWAS data for chronic cough and insomnia were accessed from the IEU GWAS database (https://gwas.mrcieu.ac.uk/). The listed GWAS IDs ("ukb-b-16858" and "ukb-a-258" for cough; "ukb-b-3957" and "ukb-a-527" for insomnia) can be used to retrieve summary statistics for these traits from MR-Base on the MRC IEU site.

References

[1] Yancey JR, Thomas SM. Chronic fatigue syndrome: diagnosis and treatment. *Am Fam Physician* 2012;86(8):741–746.
[2] Bested AC, Marshall LM. Review of myalgic encephalomyelitis/chronic fatigue syndrome: an evidence-based approach to diagnosis and management by clinicians. *Rev Environ Health* 2015;30(4):223–249.
[3] Reid S, Chalder T, Cleare A, et al. Chronic fatigue syndrome. *BMJ Clin Evid* 2011;2011:1101.
[4] Davis HE, Assaf GS, McCorkell L, et al. Characterizing long COVID in an international cohort: 7 months of symptoms and their impact. *EClinicalMedicine* 2021;38:101019.
[5] Mandal S, Barnett J, Brill SE, et al. 'Long-COVID': a cross-sectional study of persisting symptoms, biomarker and imaging abnormalities following hospitalisation for COVID-19. *Thorax* 2021;76(4):396–398.
[6] Lacerda EM, Geraghty K, Kingdon CC, et al. A logistic regression analysis of risk factors in ME/CFS pathogenesis. *BMC Neurol* 2019;19(1):275.
[7] Rasa S, Nora-Krukle Z, Henning N, et al. Chronic viral infections in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). *J Transl Med* 2018;16(1):268.
[8] Shen X, Ma S, Vemuri P, et al. Alzheimer's disease neuroimaging I. Challenges and opportunities with causal discovery algorithms: application to Alzheimer's pathophysiology. *Sci Rep* 2020;10(1):2975.
[9] Kitson NK, Constantinou AC. Learning Bayesian networks from demographic and health survey data. *J Biomed Inform* 2021;113:103588.
[10] Mulder J, Raftery AE. BIC extensions for order-constrained model selection. *Sociol Methods Res* 2022;51(2):471–498.
[11] Ramsey J, Glymour M, Sanchez-Romero R, et al. A million variables and more: the Fast Greedy Equivalence Search algorithm for learning high-dimensional graphical causal models, with an application to functional magnetic resonance images. *Int J Data Sci Anal* 2017;3(2):121–129.
[12] Lee JJR, Srinivasan R, Ong CS, et al. Causal determinants of postoperative length of stay in cardiac surgery using causal graphical learning. *J Thorac Cardiovasc Surg* 2023;166(5):e446–e462.
[13] Gupta A, Slater JJ, Boyne D, et al. Probabilistic graphical modeling for estimating risk of coronary artery disease: applications of a flexible machine-learning method. *Med Decis Making* 2019;39(8):1032–1044.
[14] Hines O, Diaz-Ordaz K, Vansteelandt S, et al. Causal graphs for the analysis of genetic cohort data. *Physiol Genomics* 2020;52(9):369–378.
[15] Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014;23(R1):R89–R98.
[16] Palmer TM, Sterne JA, Harbord RM, et al. Instrumental variable estimation of causal risk ratios and causal odds ratios in Mendelian randomization analyses. *Am J Epidemiol* 2011;173(12):1392–1403.
[17] Naviaux RK, Naviaux JC, Li K, et al. Metabolic features of chronic fatigue syndrome. *Proc Natl Acad Sci U S A* 2016;113(37):E5472–E5480.
[18] Yang X, Wang S, Qi L, et al. An efficient method for qualitation and quantitation of multi-components of the herbal medicine Qingjin Yiqi Granules. *J Pharm Biomed Anal* 2023;227:115288.
[19] Pang W, Yang F, Zhao Y, et al. Qingjin Yiqi granules for post-COVID-19 condition: a randomized clinical trial. *J Evid Based Med* 2022;15(1):30–38.
[20] Herd P, Carr D, Roan C. Cohort profile: Wisconsin longitudinal study (WLS). *Int J Epidemiol* 2014;43(1):34–41.
[21] Rong S, Snetselaar LG, Xu G, et al. Association of skipping breakfast with cardiovascular and all-cause mortality. *J Am Coll Cardiol* 2019;73(16):2025–2032.
[22] Zheng J, Haberland V, Baird D, et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat Genet* 2020;52(10):1122–1131.
[23] Schmidt AF, Dudbridge F. Mendelian randomization with Egger pleiotropy correction and weakly informative Bayesian priors. *Int J Epidemiol* 2018;47(4):1217–1228.
[24] Scheines R, Spirtes P, Glymour C, et al. The TETRAD project: constraint based aids to causal model specification. *Multivariate Behav Res* 1998;33(1):65–117.
[25] Kruger K, Holzinger F, Trauth J, et al. Chronic cough. *Dtsch Arztebl Int* 2022;119(5):59–65.
[26] Morice AH, Fontana GA, Sovijarvi AR, et al. The diagnosis and management of chronic cough. *Eur Respir J* 2004;24(3):481–492.
[27] Meng N, Lyu Y, Zhang X, et al. The exciting and magical journey of components from compound formulae to where they fight. *Acupunct Herb Med* 2022;2(4):240–252.
[28] Yilmaz CK, Kapucu S. The effect of progressive relaxation exercises on fatigue and sleep quality in individuals with COPD. *Holist Nurs Pract* 2017;31(6):369–377.
[29] Nese A, Samancioglu Baglama S. The effect of progressive muscle relaxation and deep breathing exercises on dyspnea and fatigue symptoms of COPD patients: a randomized controlled study. *Holist Nurs Pract* 2022;36(4):E18–E26.
[30] Tuzun S, Keles A, Okutan D, et al. Assessment of musculoskeletal pain, fatigue and grip strength in hospitalized patients with COVID-19. *Eur J Phys Rehabil Med* 2021;57(4):653–662.
[31] Farhi LE, Sheehan DW. Pulmonary circulation and systemic circulation: similar problems, different solutions. *Adv Exp Med Biol* 1990;277:579–586.
[32] Bierl C, Nisenbaum R, Hoaglin DC, et al. Regional distribution of fatiguing illnesses in the United States: a pilot study. *Popul Health Metr* 2004;2(1):1.
[33] Lu T, Forgetta V, Greenwood CMT, et al. Circulating proteins influencing psychiatric disease: a Mendelian Randomization study. *Biol Psychiatry* 2023;93(1):82–91.
[34] Kacso IM, Potra AR, Bondor CI, et al. ESAM predicts cardiovascular mortality in diabetic hemodialysis patients. *Biomarkers* 2015;20(5):323–327.