Original Article

Modular characteristics and the mechanism of Chinese medicine's treatment of gastric cancer: a data mining and pharmacology-based identification

Xintian Xu[1,2#], Yaling Chen[3#], Xingxing Zhang[4#], Ruijuan Zhang[2], Xu Chen[2], Shenlin Liu[1], Qingmin Sun[5]

[1]Oncology Department, Jiangsu Province Hospital of Chinese Medicine, Affiliated Hospital of Nanjing University of Chinese Medicine, Nanjing, China; [2]No. 1 Clinical Medical College, Nanjing University of Chinese Medicine, Nanjing, China; [3]College of Pharmacy, Nanjing University of Chinese Medicine, Nanjing, China; [4]Gastroenterology Department, Jiangsu Province Hospital of Chinese Medicine, Affiliated Hospital of Nanjing University of Chinese Medicine, Nanjing, China; [5]Science and technology Department, Jiangsu Province Hospital of Chinese Medicine, Affiliated Hospital of Nanjing University of Chinese Medicine, Nanjing, China

Contributions: (I) Conception and design: Q Sun, S Liu; (II) Administrative support: None; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: Y Chen; (V) Data analysis and interpretation: X Xu, X Zhang, R Zhang, X Chen; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Qingmin Sun. Jiangsu Province Hospital of Chinese Medicine, Affiliated Hospital of Nanjing University of Chinese Medicine, 155 Hanzhong Road, Nanjing 210029, Jiangsu, China. Email: qingminsun@njucm.edu.cn; Shenlin Liu. Jiangsu Province Hospital of Chinese Medicine, Affiliated Hospital of Nanjing University of Chinese Medicine, 155 Hanzhong Road, Nanjing 210029, Jiangsu, China. Email: lsljsszyy@126.com.

Background: Traditional Chinese medicine (TCM) is increasingly extensively being applied as a complementary and alternative therapy for gastric cancer (GC); however, there is a lack of large-scale evidence-based deep learning for the guidance of its clinical prescription.

Methods: The combinational search terms of "Gastric cancer and/or gastric malignancy" and "Traditional Chinese Medicine" were used to retrieve clinical study-based herbal prescriptions from public database over the past 3 decades [1990–2020]. Association rules mining (ARM) was used to analyze the prescription patterns of the herbs extracted from the eligible studies. Deep machine learning and computational prediction were conducted to explore candidate prescriptions with general applicability for GC. The action mechanism of the preferred prescription was investigated through network pharmacology, and further validated via *in vivo* and *in vitro* experiments.

Results: A total of 194 clinical study-based herbal prescriptions with good efficacy for GC were collected. TCM with focus on invigorating the Spleen and tonifying the vital-*Qi* is a promising adjuvant therapy for GC. The preferred prescription is composed of *Atractylodis Macrocephalae Rhizoma*, *Astragali Radix*, *Pinelliae Rhizoma*, *Citri Reticulatae Pericarpium*, *Herba Hedyotidis Diffusae*, *Crataegi Fructus*, and so on. We screened 74 bioactive compounds and 2,128 predictive targets of the preferred prescription from public databases. Eventually, 135 GC-related genes were identified as the targets of the preferred prescription. The compound-target network revealed that the crucial substances in the preferred prescription are quercetin, kaempferol, baicalein, and nobiletin. Experimentally, the preferred prescription was validated to modulate GC cell survival and inhibit tumor progression mainly via the hTERT/MDM2-p53 signaling pathway *in vivo* and *in vitro*.

Conclusions: TCM aimed at invigorating the Spleen and tonifying the vital-*Qi* is a promising adjuvant therapy for GC, which offers a guidance for worldwide use of TCM in the treatment of GC.

Keywords: Traditional Chinese medicine (TCM); data mining; machine learning; gastric cancer (GC); network pharmacology

Submitted Oct 27, 2021. Accepted for publication Dec 17, 2021.
doi: 10.21037/atm-21-6301
View this article at: https://dx.doi.org/10.21037/atm-21-6301

Introduction

Globally, gastric cancer (GC) is the fourth most common cancer and the second leading cause of cancer deaths (1). Current approaches to GC management largely consist of endoscopic detection followed by gastrectomy and chemotherapy (CT) or chemo-radiotherapy (CRT); however, the available treatments have adverse side effects and are associated with high recurrence rates (2). Therefore, there is a need to address the current limitations of the various therapeutic strategies to facilitate possible clinical applications.

With the development of personalized and complementary medicine, multi-compound and multi-targeting traditional Chinese medicine (TCM) has been shown to be clinically effective in treating GC (3,4). However, due to the lack of large-scale evidence-based medicine, the extensive application of TCM remains inhibited. The use of TCM as an adjuvant therapy is greatly subjective because understanding on GC treatment varies among physicians in terms of etiology, syndrome differentiation, and medicinal prescriptions. Generally, the principal theory of Chinese traditional medicine for GC is invigorating the Spleen and tonifying the vital-*Qi*, and eliminating blood stasis and removing toxins. The largely unknown mechanism of these empirical prescriptions is another limiting factor for the use of TCM. Therefore, it is important and innovative to screen clinical prescriptions with good efficacy, based on which the obtainment of a basic prescription with general applicability for treating GC could be achieved by machine learning. What's more, the elucidation of prescription patterns by data mining may promote both clinical application and basic researches on herbal pairs. To the best of our knowledge, another study with such an aim has not been previously reported.

In this study, we proposed a method of combining data mining and network pharmacology to systematically elucidate the prescription patterns of TCM, and unravel the modular functions and potential action mechanisms of TCM for treating GC. Additionally, the effects of the machine learning-based preferred prescription were validated *in vivo* and *in vitro*. We present the following article in accordance with the ARRIVE reporting checklist (available at https://dx.doi.org/10.21037/atm-21-6301).

Methods

Big data mining and machine learning

Source of literature, inclusion and exclusion criteria

All literature was obtained from the China National Knowledge Infrastructure (CNKI) database, which is the world's largest Chinese knowledge portal website. The sources of the literature included the Academic Journals Full-text Database, Doctoral Dissertations Full-text Database, and Masters'. These Full-text Database (01/1990–12/2020). The combinational search terms were "Gastric cancer and/or gastric malignancy" and "Traditional Chinese Medicine". Literature with the following criteria were included: (I) relevant to clinical research on using TCM in treating GC; (II) containing randomized controlled trial (RCT) as the study design; (III) containing prescriptions with complete and specific names of Chinese herbs; (IV) studies must have reported one or more of the following efficacy endpoints: progression-free survival (PFS), overall survival (OS), objective response rate (ORR), or adverse events (AEs) (5). The exclusion criteria were as follows: (I) duplicate publications reporting the same group of participants; (II) non-clinical studies including experimental research on cell lines, xenografts or animal models, or theoretical studies; (III) where TCM and western medicines were integrated as a therapeutic regimen; (IV) non-oral administrations including injection and nasogastric tube nutrition; (V) non-decoction dosage types including Chinese patent medicines and TCM for external use; (VI) use of prescriptions composed of an unspecified or single herb.

Data extraction

Firstly, the names of the prescriptions (ancient prescriptions or recombinant personalized prescriptions) and their constituent herbs were extracted from the eligible literature. Secondly, we referred to the *Chinese Pharmacopoeia* (2020 Edition) *Volume I* to standardize the names of each herb (6). Furthermore, the basic information on each herb was extracted from the *Chinese Pharmacopoeia*, including its Latin name, property, taste, and meridian tropism. The 5 properties of TCM herbs include cold, hot, warm, cool, and neutral. The 5 tastes of TCM herbs include sour, bitter, sweet, pungent, and salty. The various combinations of property and taste determine the herbs' specific attributes, which can influence the Yin and Yang of the body. For example, herbs with warm and hot properties are used to invigorate the Yang in patients with heat-deficiency disorders. Likewise, sour, bitter, and salty tastes are related to Yin, whereas pungent and sweet pertain to Yang. The meridian serves as the pathway for the transportation of Qi and Blood throughout the body, and its tropism represents the selective therapeutic effects of a medicinal herb on

a certain region of the human body (7). Moreover, the principal function of each herb was classified according to the *Chinese Pharmacy* (8).

Association rule mining (ARM)

To investigate the rules of herbal combinations in the prescriptions used in various studies, ARM, an in-silico screening process, was applied. In this scheme, the dataset and the association rules are defined as follows: an association rule has the form left hand side (LHS) ⇒ right hand side (RHS), where LHS and RHS are sets of items, with the likely occurrence of the RHS whenever the LHS set occurs (9). The Apriori algorithm was used to extract the significant associations from all possible combinations of the items from the main dataset (10). There are 3 evaluation metrics which are critical in describing the power and significance of the rules generated by ARM (11). Support is the frequency of the rule occurrence in the total dataset, measuring whether an association between the LHS and the RHS happens by chance. Confidence is the frequency of rule occurrence in the cases of the dataset fulfilling the LHS of the rule, thus, representing the reliability of the association. Lift is the ratio of observed support to the expected support when the LHS and the RHS are independent, indicating the dependency of the occurrences of the 2 items when its value is larger than 1 (12). To establish a proper threshold, we detected the central tendency of the association rules to be more obvious at the *s*upport of 0.1 and confidence of 0.6 in the correlation analysis of herbal combination patterns. Then, the herbs were categorized in Microsoft Excel 2010 (Microsoft Corp., Redmond, WA, USA) according to their properties, tastes, meridian tropisms, and functions. The software platform IBM SPSS Modeler 18.1 (IBM Corp., Armonk, NY, USA) was used to analyze the categorization-based frequency and the correlations of the prescription patterns and to generate a visual network diagram.

Cluster analysis

Clustering is central to many data-driven bioinformatics research and serves a powerful computational method. Deep learning can be effective means to transform mappings from a high-dimensional data space into a lower-dimensional feature space, leading to improved clustering results (13). In this study, we used IBM SPSS Modeler software platform to perform deep learning-based cluster analysis to identify the preferred regroups of the most frequently used herbs based on their attributes (14). In our study, k-means cluster analysis was considered since the variables were quantitative at the interval or ratio level rather than being binary or counts. To avoid unreliable results through omitted variable bias, we included all the attributes, including the 5 properties, 5 tastes, and meridian tropism, and investigated the therapeutic preferences of the candidate clusters. To assess the reliability of a given solution, we compared the results from analyses with different permutations of the initial center values to ensure an appropriate number of clusters.

Mechanism investigation of the candidate formulae by network pharmacology

Compounds library construction and active components screening of the candidate formulae

To build a compound library of the core herbs for GC, we extracted all the compounds of the candidate formulae from the Traditional Chinese Medicines for Systems Pharmacology Database and Analysis Platform (TCMSP; http://lsp.nwu.edu.cn/index.php), Traditional Chinese Medicines Integrated Database (TCMID; http://bionet.ncpsb.org/batman-tcm/), Bioinformatics Analysis Tool for Molecular Mechanism of Traditional Chinese Medicine (BATMAN-TCM; http://bionet.ncpsb.org/batman-tcm/), and wide-scale literature mining (15,16). To optimize the use of the high cost and time-consuming biological experiments and clinical research, absorption, distribution, metabolism, and excretion (ADME) evaluations are critical procedures for active components screening (17). In this study, oral bioavailability (OB) ≥30% and drug-likeness (DL) ≥0.18 were set as the threshold; however, compounds that did not meet these inclusion criteria but were supported by the literature were retained.

Therapeutic targets prediction of the candidate formulae

Computational predictions of bioactive molecule targets based on similarity with known ligands are powerful in narrowing down the number of potential targets and the rationalization of possible side effects of the known molecules (18). The prediction algorithms of the ligand-based strategies include systematic drug targeting (SysDT) (19) and weighted ensemble similarity (WES) models (20). The SysDT model was developed based on random forest (RF) and support vector machine (SVM), which performed impressively on systematic predictions for drug-target associations and interactions involving enzymes,

ion channels, nuclear receptors, and G-protein coupled receptors (19). In the WES model, the standardized ensemble similarities (Z score) by Bayesian network are utilized and the targets are predicted using the multivariate kernel approach (21). In our study, the predictive therapeutic targets of the candidate formulae were obtained from web tools including Search Tool for Interacting Chemicals (STITCH; http://stitch.embl.de/), similarity ensemble approach (SEA; http://sea.bkslab.org/) and SwissTargetPrediction (www.swisstargetprediction.ch) (22-24). Targets with RF ≥0.7, SVM ≥0.8, or Z score ≥7 were considered for further analysis and standardized to corresponding genes for homo sapiens through the UniProt database (https://www.uniprot.org/uploadlists/).

To evaluate the performances of the candidate formulae in treating GC, we mapped the predictive therapeutic targets to the GC-related genes/proteins, which were comprehensively collected from online databases including MalaCards (https://www.malacards.org/), Online Mendelian Inheritance in Man (OMIM; https://omim.org/), and DisGeNET v7.0 (https://www.disgenet.org/home/) (25,26). We visualized the results and generated an additional protein-protein interaction (PPI) network using Metascape (https://metascape.org/gp/index.html).

Construction and topological analysis of the compound-target network of the preferred prescription

Cytoscape v3.7.2 (https://cytoscape.org/) was used to construct a compound-target (C-T) network of the preferred prescription, and to analyze its degree, a key topological parameter for evaluation (15). In the C-T network, compounds sharing interactions with GC-related genes were determined as components that were beneficial for GC. Moreover, we considered the targets (compounds) with degree values equal to or above the mean value to be the predominant therapeutic targets (crucial substances).

Gene Ontology (GO) and pathway enrichment analysis of the preferred prescription

GO analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG), and Reactome pathway enrichment of the preferred prescription were carried out using the Database for Annotation, Visualization, and Integrated Discovery system v6.8 (DAVID; https://david.ncifcrf.gov/) (27). We also used ClueGO, a Cytoscape v3.7.2 plug-in to identify the interactions among the various signaling pathways by generating a functionally grouped network (28,29). Based on the mechanism of GC, we further constructed a multi-regulation map of KEGG pathways of the crucial components in the core herbs.

Moreover, the modular functional characteristics of TCM in GC treatment were demonstrated in PPI networks, which visualized the interactions among significant targets that could be regulated by the crucial components in the preferred prescription. The PPI networks were generated by the GeneMANIA web site (http://genemania.org/) which offers a fast prediction on the functions of the given gene sets through the application of a guilt-by-association approach (30).

Molecular docking

The three-dimensional (3D) structures of the predominant targets of the preferred prescription were collected from protein data bank (PDB; http://www.rcsb.org). AutoDock Tools 1.5.6 software (https://autodock.scripps.edu/) was used to remove the water molecules, isolate proteins, add nonpolar hydrogen, and calculate Gasteiger charges for the structure (31). The preprocessed structures were saved as PBD with partial changes and AutoDock 4 atom types (PDBQT) files. The PubChem database (https://pubchem.ncbi.nlm.nih.gov/) was applied to download the two-dimensional (2D) structures of the crucial substances of the preferred prescription. The 2D structure was processed and transformed into PDB format via Open Babel (32), and then saved in PDBQT format as docking ligands in AutoDock Tools 1.5.6 software. The target proteins were used as receptors while the substances were used as ligands. The active site of molecular docking was determined by the complex of ligand and target protein. Autodock Vina 1.1.2 (https://vina.scripps.edu/) was used to dock small molecules with their target proteins. The conformation with the best affinity was selected as the final docking conformation and visualized in Pymol 2.5 (https://pymol.org/2/).

Experimental validation

Preparation of the preferred prescription and components identification

Crude TCM herbs [dried roots of *Atractylodes macrocephala* Koidz. 12 g, dried roots of *Astragalus membranaceus* (Fisch.) Bge. 30 g, dried mature pericarp of *Citrus reticulata* Blanco 10 g, dried tuber of *Pineilia ternate* (Thunb.) Breit. 9 g, dried root of *Aucklandia lappa* Decne. 6 g, dried mature fruits of *Amomum villosum* Lour. 3 g, dried immature fruits of *Citrus aurantium* L. 10 g, dried gizzard lining of *Gallus gallus domesticus* Brisson 10 g, dried mature fruits

of *Crataegus pinnatifida* Bge. 12 g, dried mature fruits of *Hordeum vulgare* L. 15 g, *Radix Actinidiae Chinensis* 15 g, and *Herba Hedyotidis Diffusae* 15 g] were provided by Sanyue Chinese Traditional Medicine Co. (Nantong, China). All the herbs were soaked for 30 min in 1,800 mL double-distilled water and then boiled at minimum temperature for 30 min before being refluxed and extracted. The boiling process was repeated with 1,800 mL double-distilled water for 30 min. Then, 2 parts of the extracted solutions were mixed and vaporized to 60 mL. The decoction was finally concentrated to 1 g/mL and stored at −20 ℃ after being sterilized and filtered through a 0.22 μm filter. The extracts of the preferred prescription were detected and analyzed using high-performance liquid chromatography diode array detection (HPLC-DAD) (detailed information shown in the Supplementary materials).

In vitro, to determine the decoction dose, the half-maximal inhibitory concentration (IC50) of different GC cell lines were assessed by 3-(4,5-Dimethyl-2-thizolyl)-2,5-diphenyltetrazolium bromide (MTT) assay (detailed information shown in the Supplementary materials), and the dose range of 2, 4, 8 mg/mL was selected.

Cell apoptosis and cell cycle analyses

Human GC cell lines AGS, HGC27, MKN28, and SGC7901 were purchased from the Cell Bank of the Chinese Academy of Sciences (Shanghai, China). All cell lines were kept in a humidified atmosphere of 5% CO_2 at 37 ℃. For apoptosis analysis, the cells were measured using Annexin V-FITC/PI apoptosis detection kit (Keygen Biotech Co., Nanjing, China) by flow cytometry [Becton, Dickinson, and Co. (BD) Biosciences, Franklin Lakes, NJ, USA] according to the manufacturer's instructions. Cell cycle distributions were determined using a cell cycle and apoptosis analysis kit (Beyotime Biotech Co., Shanghai, China) by flow cytometry (BD Biosciences).

Wound-healing assay

Cells ($1{,}000\times10^3$ cells/well) were seeded into 6-well plates for 24 h, and scraped with a sterile pipette tip when 80% of the cells were adherent to the walls. Cells were treated with various concentrations of the preferred prescription after removing debris by phosphate-buffered saline (PBS). The scratch area was observed by microscopy at 0, 12, 24, and 48 h, respectively.

Invasion assay

The upper surface of the Transwell inserts (8 μm pore size, Merck & Millipore, Darmstadt, Germany) were coated with Matrigel (100 μL, diluted 1:29 with PBS) (Corning, Corning, NY, USA) before serum-free medium containing 2×10^5 cells were loaded. The lower chamber included 500 μL media containing 10% fasting blood sugar (FBS) and various concentrations of the preferred prescription. After 48 h, the chambers were removed, and nonpenetrative cells were washed from the top chamber with PBS. The invaded cells were fixed with 95% ethanol and stained with crystal violet. Image J (https://imagej.nih.gov/ij/) was applied to count the number of cells in images randomly taken under a microscope.

Western blot assay

Protein lysates were separated using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to polyvinylidene difluoride (PVDF) membranes. Membranes were blocked with 5% bovine serum albumin (BSA) for 1 h and incubated with primary antibodies at 4 ℃ overnight. The primary antibodies included β-actin, Bax, Bcl2, N-cadherin, Snail, Slug, hTERT, MDM2, p53, p21, cyclinE, and CDK2 [all antibodies were purchased from Cell Signaling Technology (CST) Danvers, MA, USA]. The secondary goat anti-rabbit horseradish peroxidase-conjugated antibody (ZSGB-BIO, Beijing, China) was incubated at room temperature for 1 h. Signals were examined using the Image Lab system, version 5.1 (Bio-Rad, Hercules, CA, USA).

In vivo study

Male BALB/c athymic nude mice (4–6 weeks old, 18–20 g) obtained from Charles River Co. (Beijing, China) were housed in a specific pathogen-free (SPF) environment. An appropriate amount of the preferred prescription extracts was collected and prepared into the 0.735 g/mL solution with distilled water, and used for the intragastric administration of the experimental animals. The MKN28 cells were collected and cultured in the logarithmic growth phase, and the density was adjusted to 5×10^7/mL. Each mouse was inoculated with 0.2 mL of cell suspension in the right armpit after disinfection. After 10 days, the diameter of the induration reached 3–7 mm, suggesting the establishment of a successful model. The 20 nude mice were divided into 4 groups (n=5 each) as follows: (I) model group with transplanted tumors given the gavage of distilled water; (II) 5-fluorouracil (5-FU) group with transplanted tumors given the intraperitoneal injection at a dose of 20 mg/kg body weight (BW) every 3 days; (III) preferred

prescription group with transplanted tumors given the gavage of decoction at a dose of 14.7 g/kg BW every day for 14 days; (IV) the 5-FU+ preferred prescription group with transplanted tumors given the intraperitoneal injection at a dose of 20 mg/kg BW every 3 days and gavage of decoction at a dose of 14.7 g/kg BW every day for 14 days. To calculate the volume of the tumors, the dimension was measured by length (L) and width (W) using a caliper every 3 days. Mice were sacrificed by cervical dislocation, and the tumors were excised and weighed.

Ethical statement

This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). Animal experiment was performed under a project license (No. 2021DW-35-01) granted by the Animal Ethics Committee of Affiliated Hospital of Nanjing University of Chinese Medicine (Nanjing, China), in compliance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. A protocol was prepared before the study without registration.

Statistical analysis

The data were described as means ± standard error of the mean (SEM). Statistical significance was determined using one-way analysis of variance (ANOVA, comparison between multiple groups) and Tukey multiple comparison processing (comparison between the two groups), with a P value <0.05 indicating statistical significance. All experiments were repeated at least three times under the same conditions. The statistical analyses were performed using GraphPad Prism software (GraphPad Software, La Jolla, CA, USA).

Results

Screening of eligible literature, clinical study-based prescriptions, core herbs, and frequency distributions according to herbal attributes and principal functional categorizations

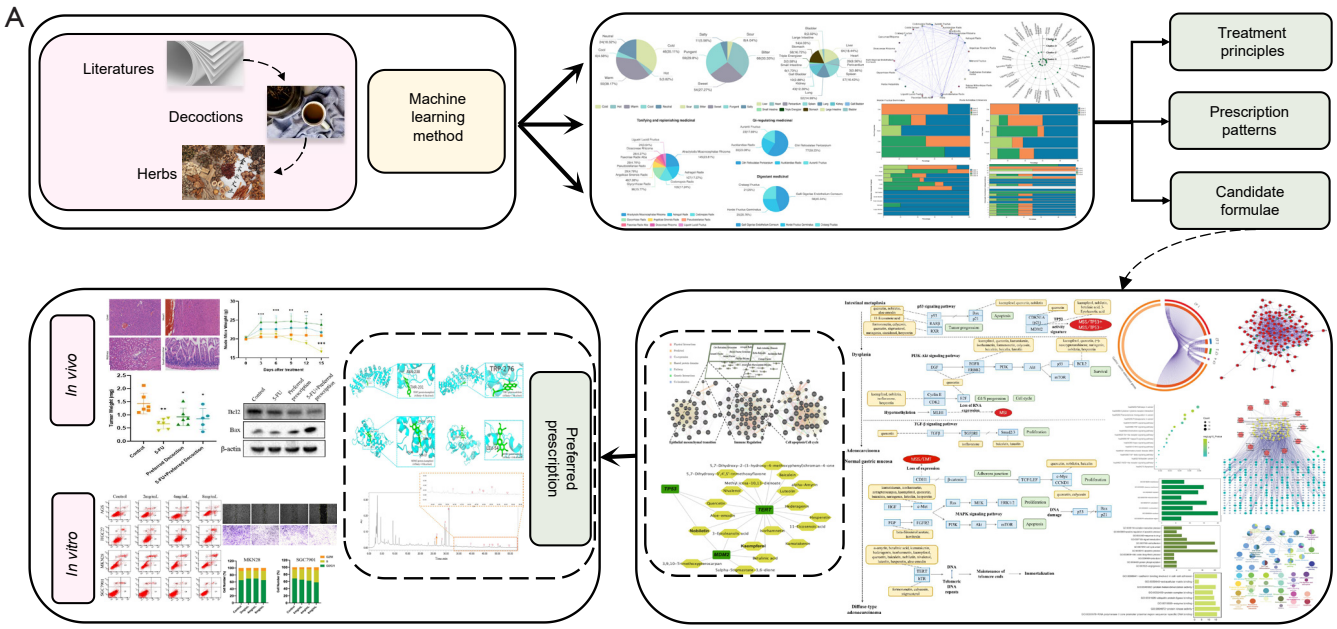
The framework of this study can be summarized as follows: (I) screening of clinical study based TCM prescriptions for GC treatment; (II) data mining of the treatment principles, prescription patterns, and generation of candidate formulae by deep machine learning; (III) prediction of the action mechanism of the preferred prescription by network pharmacology; (IV) validation of the antitumor effects of the preferred prescription by experiments *in vivo* and *in vitro* (*Figure 1A*).

A total of 194 eligible prescriptions and 148 herbs with standardized names were screened from clinical studies spanning from January 1990 to December 2020. The screening process is summarized as a PRISMA flow diagram (33) (*Figure 1B*). The total cumulative occurrences of the 148 herbs in 194 prescriptions were 2,103 times. Herbs with over 20 times frequency of occurrence were selected as predominant ones used in clinic. The top 24 core herbs and their functional categorizations are listed in Table S1. Descriptive statistics of herbal attributes are shown in *Figure 2A*. In terms of the 5 properties; herbs with warm property were the most frequently prescribed. With regards to the 5 tastes, herbs with bitter (44.59%), pungent (39.86%), and sweet (36.49%) tastes ranked the top 3 in clinical application. In terms of meridian tropism, herbs with a propensity for the Liver (LR) (43.24%), Stomach (ST) (39.19%), and Spleen (SP) meridians were the most frequently used. The top 3 principal functions of the core herbs are demonstrated in *Figure 2B*. In summary, the treatment principle of TCM in GC is mainly invigorating the Spleen and tonifying the vital-*Qi*.

To facilitate better application of the core herbs in clinic, we summarized their clinical indications (Table S2). Particularly, herbs with the 3 major functions that embody TCM treatment principles for GC are listed in *Table 1*.

Frequently prescribed herbal combination patterns by ARM and novel candidate formula prediction by cluster analysis

The ARM method was applied to analyze the combination patterns of the 194 prescriptions. Guided by the theory of synergy and attenuation in TCM, couplet herbs are 2 herbs administered together to enhance therapeutic effects or reduce toxicity. With a threshold of minimum *support* of 0.1 and con*fidence* of 0.6, the prescribed pairs of couplet herbs with the top 3 *confidence* included; *Atractylodis Macrocephalae Rhizoma* (Bai Zhu) paired with *Dioscoreae Rhizoma* (Shan Yao; 92.31%), *Atractylodis Macrocephalae Rhizoma* paired with *Aucklandiae Radix* (Mu Xiang; 90%), and *Atractylodis Macrocephalae Rhizoma* paired with *Codonopsis Radix* (Dang Shen; 88.57%) (*Table 2*). Triplet herbs are a combination of 3 herbs, which interact with each other and are usually contained in a decoction or used as an independent decoction. Based on the established threshold above, the triplet combinations of herbs with the top 3 *confidence* included *Poria* (Fu Ling)-*Aucklandiae Radix*-*Glycyrrhizae*



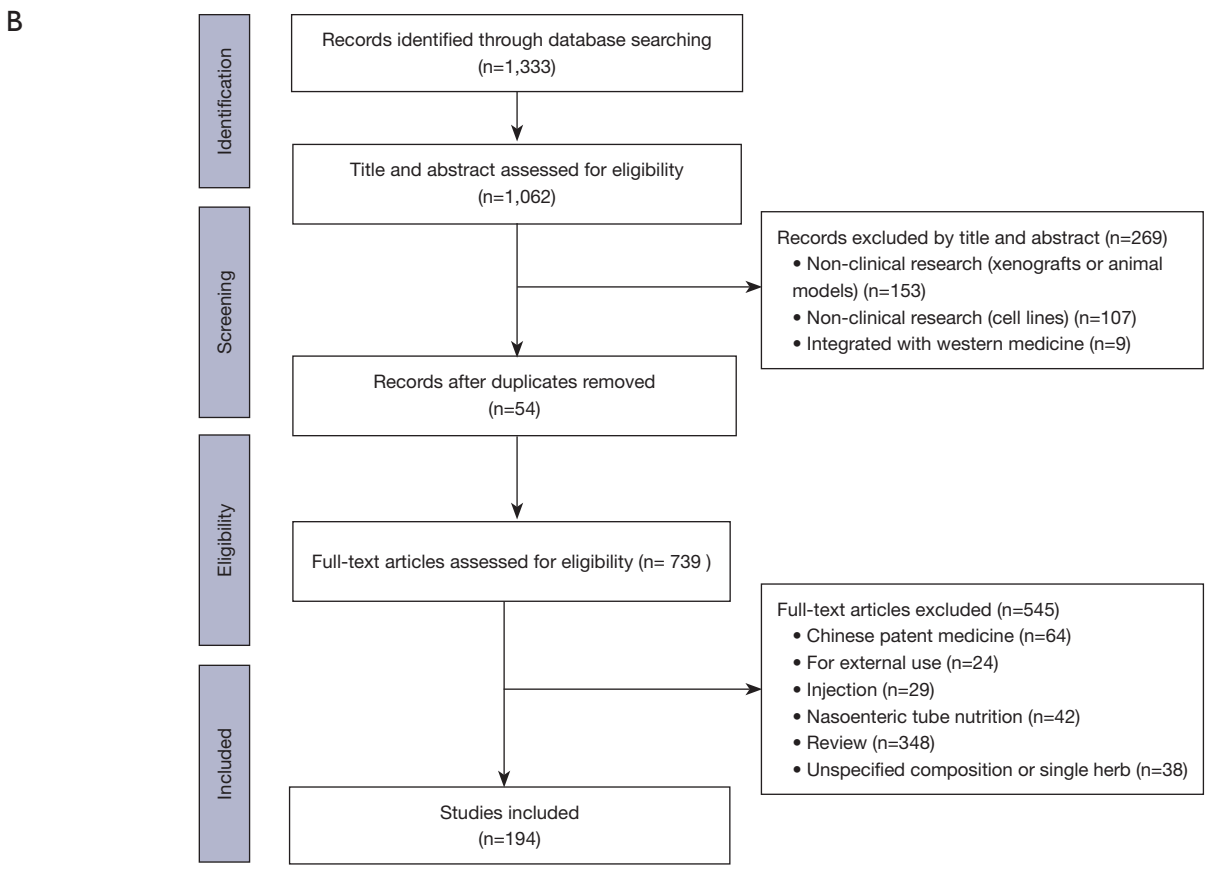
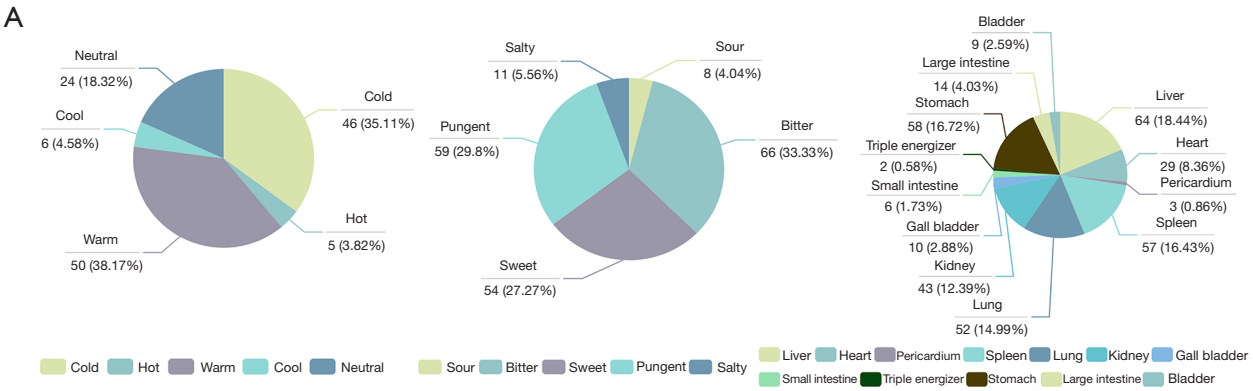
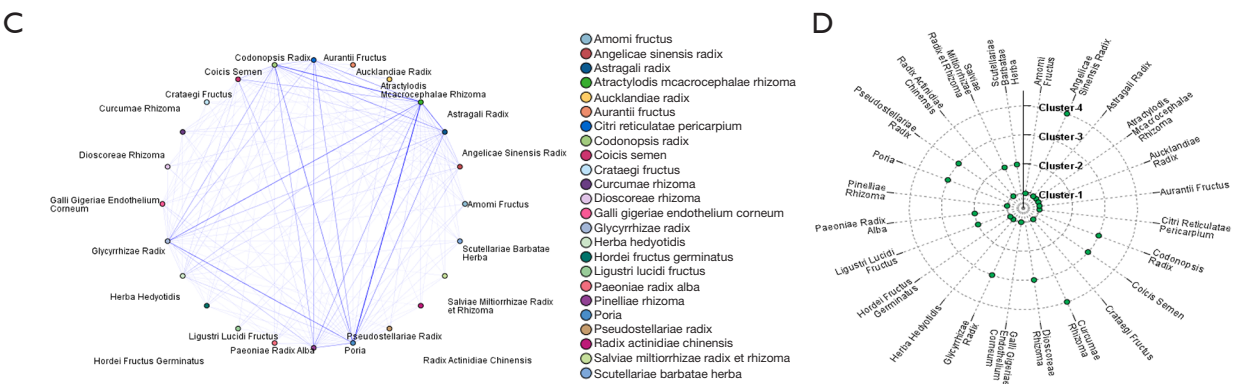


Figure 1 The technical roadmap of the current study. (A) The framework of the current study is summarized as data mining and machine learning combined with network pharmacology and experimental validation. (B) Flow chart of literature mining. A total of 1,333 records were retrieved, and 194 prescriptions were extracted.





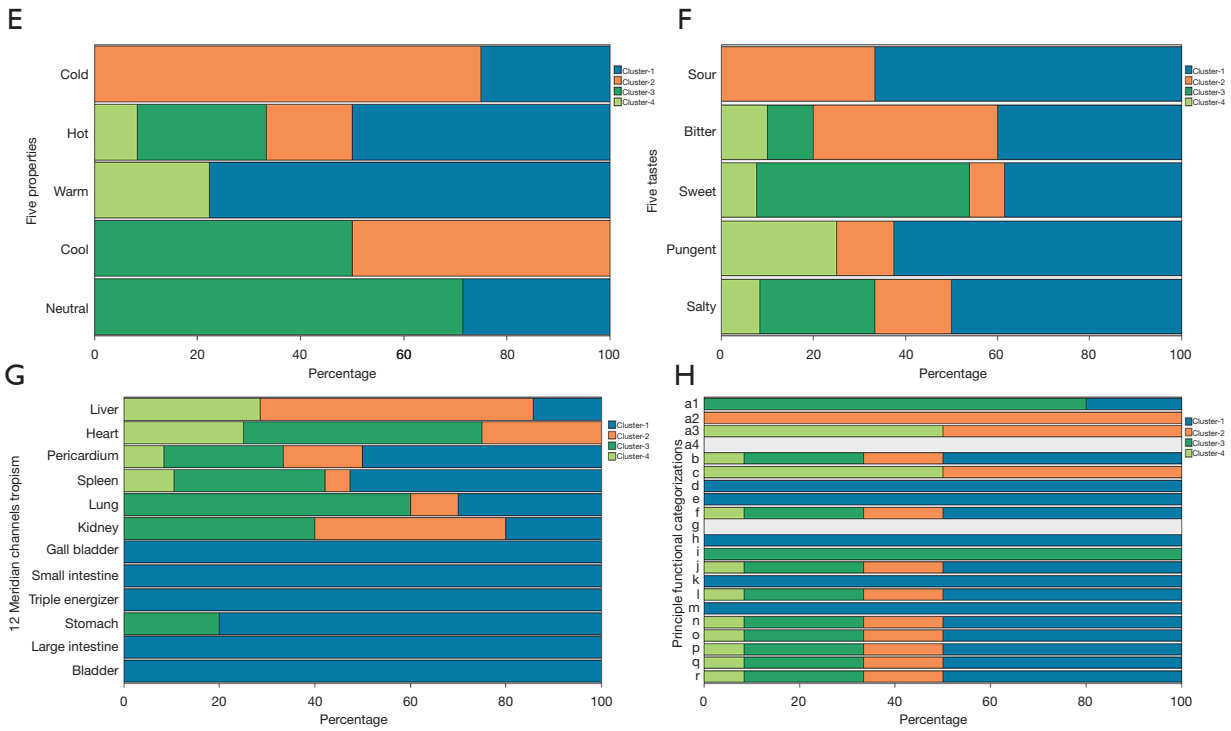


Figure 2 Frequently prescribed herbal combination patterns and novel candidate formulae prediction. (A) Descriptive statistics of herbal attributes including the 5 properties, 5 tastes, and the meridian tropism. (B) The top 3 principal functions of the core herbs for treating GC. (C) Network diagram of herbal combination patterns (support ≥10%, confidence ≥60%). (D) Novel candidate formulae prediction. Proportion distributions of the four clusters (candidate formulae) according to 5 properties (E), 5 tastes (F), 12 meridians tropism (G), and principal functional categorization (H). a1: *Qi*-tonifying medicinal; a2: *Yin*-tonifying medicinal; a3: *Blood*-tonifying medicinal; a4: *Yang*-tonifying medicinal; b: heat-clearing medicinal; c: blood-activating and stasis-dispelling medicinal; d: *Qi*-regulating medicinal; e: cough-suppressing and panting-calming medicinal; f: interior-warming medicinal; g: Liver-pacifying medicinal; h: digestant medicinal; i: dampness-draining diuretic medicinal; j: exterior-releasing medicinal; k: dampness-resolving medicinal; l: hemostatic medicinal; m: wind-dampness dispelling medicinal; n: astringent medicinal; o: purgative medicinal; p: orifice-opening medicinal; q: repellent medicinal; r: attacking poison, insects and itch-relieving medicinal. GC, gastric cancer.

Radix (Gan Cao) (100%), *Atractylodis Macrocephalae Rhizoma-Aucklandiae Radix-Codonopsis Radix* (100%), and *Atractylodis Macrocephalae Rhizoma-Galli Gigeriae Endothelium Corneum* (Ji Nei Jin)-*Codonopsis Radix* (100%) (Table S3). A network diagram was generated to visualize the association rules among the core herbs (*Figure 2C*).

Additionally, the core herbs in the 194 prescriptions were regrouped into 4 clusters by machine learning. The cluster analysis result was presented in a 2D scatter diagram (*Figure 2D*). Cluster 1 (candidate formula 1, CF 1) included; *Atractylodis Macrocephalae Rhizoma*, *Astragali Radix* (Huang Qi), *Pinelliae Rhizoma* (Zhi Ban Xia), *Citri Reticulatae Pericarpium* (Chen Pi), *Herba Hedyotidis* (Bai Hua She She Cao), *Galli Gigeriae Endothelium Corneum*, *Aucklandiae Radix*, *Amomi Fructus* (Sha Ren), *Hordei Fructus Germinatus* (Mai Ya), *Aurantii Fructus* (Zhi Ke), *Radix Actinidiae Chinensis* (Mi Hou Tao Gen), *Crataegi Fructus* (Shan Zha); cluster 2 (candidate formula 2, CF 2) included; *Paeoniae Radix Alba* (Bai Shao), *Scutellariae Barbatae Herba* (Ban Zhi Lian), *Salviae Miltiorrhizae Radix et Rhizoma* (Dan Shen), *Ligustri Lucidi Fructus* (Nu Zhen Zi); cluster 3 (candidate formula 3, CF 3) included; *Poria*, *Codonopsis Radix*, *Glycyrrhizae Radix*, *Coicis Semen* (Yi Yi Ren), *Pseudostellariae Radix* (Tai Zi Shen), *Dioscoreae Rhizoma*; and cluster 4 (candidate formula 4, CF 4) included; *Angelicae Sinensis Radix* (Dang Gui), *Curcumae Rhizoma* (E Zhu). For a clearer understanding of the functions of the candidate formulae, the herbal attributes preferences of the 4 candidate formulae are shown in a distribution histogram (*Figure 2E-2H*). The CF 1 and CF 2 were composed of more herbs with warm and cold

Table 1 Top 3 functions of TCM for GC treatment and clinical indications of the representative herbs

Herbal nature	Principal functional categorizations	Number of prescriptions using the herbs	Frequency of use (%)	Syndromes	Key signs & symptoms	Treatment principles	*Representative herbs*
Sweet, warm	*Qi*-tonifying	182	93.81	*Middle-Jiao* Deficiency	Poor appetite, dislike to talk, lassitude, weak limbs, borborygmus, loose stools, heavy descending sensation in abdominal cavity, prolapse of rectum.	Tonify *Middle-Jiao Qi*	*Atractylodis Macrocephalae Rhizoma, Astragali Radix, Codonopsis Radix, Glycyrrhizae Radix, Pseudostellariae Radix, Dioscoreae Rhizoma*
Pungent, warm/bitter, warm	*Qi*-regulating	119	61.34	Stagnation of Liver *Qi*	Mental depression, restlessness, sighing, distension, wandering pain in the costal and hypochondriac region, distress in epigastrium, poor appetite or vomiting, irregular bowel movements, thin greasy tongue coating, wiry pulse.	Disperse Liver *Qi*	*Citri Reticulatae Pericarpium, Aucklandiae Radix, Aurantii Fructus, Fructus Evodiae*
				Qi stagnation transforming into *Fire*	Irritability, stuffiness in the chest, hypochondriac distension, acid regurgitation, dry & bitter mouth, constipation or headache, tinnitus, red tongue & yellow coating, wiry-rapid pulse.	Purge *Fire* from Liver	
Sweet, neutral	Food abating	55	28.35	Stomach excessive	Epigastric and abdominal distension and fullness or pain, which are aggravated by food intake, belching with foul smell, anorexia, constipation, acidic regurgitation, nausea, vomiting, diarrhea with foul smell or fermented contents or constipation	Dissolve the stagnation	*Galli Gigeriae Endothelium Corneum, Hordei Fructus Germinatus, Crataegi Fructus*

TCM, traditional Chinese medicine; GC, gastric cancer.

properties respectively, while the property of CF 3 appeared to be milder. With regards to the 5 tastes, the majority of the herbs in the CF 1 possessed sour and pungent tastes, while most herbs with sweet tastes were clustered in CF 3. Meridian tropism represents the selective therapeutic effects of a Chinese herb on a certain region of the human body (7). The CF 1, CF 2, and CF 3 prescribed more herbs belonging to SP (ST), LR, and KI meridians, respectively. Specifically, the CF 1 and CF 3 clusters were distinguished for tonifying Qi and invigorating the Spleen, and regulating Qi and resolving dampness, which meant both of them could increase appetite, alleviate lassitude, fullness sensation in the upper abdomen, and loose stools, as well as help GC patients feel less depressive. While the CF 2 and CF 4 seemed to play a significant role in tonifying the Blood and promoting blood circulation, which indicated they are more applicable for GC patients with symptoms like pale complexion, dizziness, insomnia, distending pain of the hypochondrium, and so on.

Active components library construction and therapeutic targets prediction of the candidate formulae

Complied with OB ≥30% and DL ≥0.18, 305 compounds of the 24 core herbs were screened out as bioactive components (Table S4). The numbers of active components

Table 2 Top 10 pairs of couplet herbs used in clinical prescriptions

Herb (LHS)	Number of prescriptions	Herb (RHS)		Number of occurrences	Support (LHS) (%)	Confidence (LHS ≥ RHS) (%)	LIFT
Atractylodis Macrocephalae Rhizoma	145	*Dioscoreae Rhizoma*	→	26	13.40	92.31	1.24
Atractylodis Macrocephalae Rhizoma	145	*Aucklandiae Radix*	→	30	15.46	90.00	1.20
Atractylodis Macrocephalae Rhizoma	145	*Codonopsis Radix*	→	105	54.12	88.57	1.19
Astragali Radix	107	*Ligustri Lucidi Fructus*	→	24	12.37	87.50	1.59
Atractylodis Macrocephalae Rhizoma	145	*Poria*	→	126	64.95	87.30	1.17
Atractylodis Macrocephalae Rhizoma	145	*Coicis Semen*	→	63	32.47	87.30	1.17
Poria	126	*Pseudostellariae Radix*	→	29	14.95	86.21	1.33
Poria	126	*Amomi Fructus*	→	28	14.43	85.71	1.32
Atractylodis Macrocephalae Rhizoma	145	*Amomi Fructus*	→	28	14.43	85.71	1.15
Codonopsis Radix	105	*Hordei Fructus Germinatus*	→	25	12.89	84.00	1.55

LHS, left hand side; RHS, right hand side.

in CF 1-4 were 74, 106, 154, and 5, respectively. With 35 compounds hitting no corresponding targets, a total of 2,128 predictive targets were retrieved and normalized via prediction databases and UniProt, with the potential to interact with 305 active components.

To investigate the relationship between the predictive targets and GC, 429 GC-related genes were screened, and 136 targets overlapped. As shown in *Figure 3A*, CF 1 targeted the most GC-related genes and was defined as the "preferred prescription" in our study. The CF 1-4 contributed to 135, 12, 13, and 4 genes/proteins, respectively. As shown in *Figure 3B*, the shared genes/proteins among the 4 formulae included human telomerase reverse transcriptase (*hTERT*), tyrosine-protein phosphatase non-receptor type 11 (*PTPN11*), estrogen receptor (*ESR1*), and sonic hedgehog protein (SHH), G2/mitotic-specific cyclin-B1 (CCNB1), fibroblast growth factor 2 (FGF2), and so on.

Construction and topological analysis of the compound-target network of the preferred prescription

Topological analysis of the C-T network was conducted to identify the crucial components and targets in the preferred prescription. As shown in *Figure 3C*, the network embodied 505 nodes (11 herbs, 61 active components, and 429 target genes/proteins), and 952 C-T interactions. The mean degree of the active components was 15.61. There were 23 compounds with a degree value higher than 15.61. In this network, crucial substances quercetin (*Astragali Radix*, *Herba Hedyotidis*, *Radix Actinidiae Chinensis*, *Crataegi Fructus*), kaempferol (*Astragali Radix*), baicalein (*Pinelliae Rhizoma*), and nobiletin (*Citri Reticulatae Pericarpium*, *Aurantii Fructus*) targeted 256, 191, 144, and 131 GC-related genes, respectively. There were 49 targets with degree values higher than 6.57, the mean degree of the predicted targets. The *TP53*, *hTERT*, vascular endothelial growth factor A (*VEGFA*), caspase-3 (*CASP3*), murine double minute 2 (*MDM2*), matrix metalloproteinase 2 (*MMP2*), and apoptosis regulator Bcl-2 (*BCL2*) genes were targeted by 26, 19, 16, 16, 14, 14, and 12 compounds respectively, which indicated they may be involved in the underlying mechanisms of the preferred prescription.

GO and pathway enrichment analysis of the preferred prescription

To explore the potential mechanism of the preferred prescription, we utilized the DAVID database to decipher the information related to gene ontology. The GO analysis on the targets of the preferred prescription is shown in *Figure 4A*. The significant biological processes (BP) ($P<0.05$) included apoptotic process (GO: 0006915), cell adhesion (GO: 0007155), cell cycle arrest (GO: 0007050), and signal transduction (GO: 0007165). The significant molecular functions (MF) ($P<0.05$) included protein kinase activity (GO: 0004672), cadherin binding involved in cell-cell adhesion (GO:0098641), enzyme binding (GO:0019899), and ubiquitin-protein ligase binding (GO:0031625). The

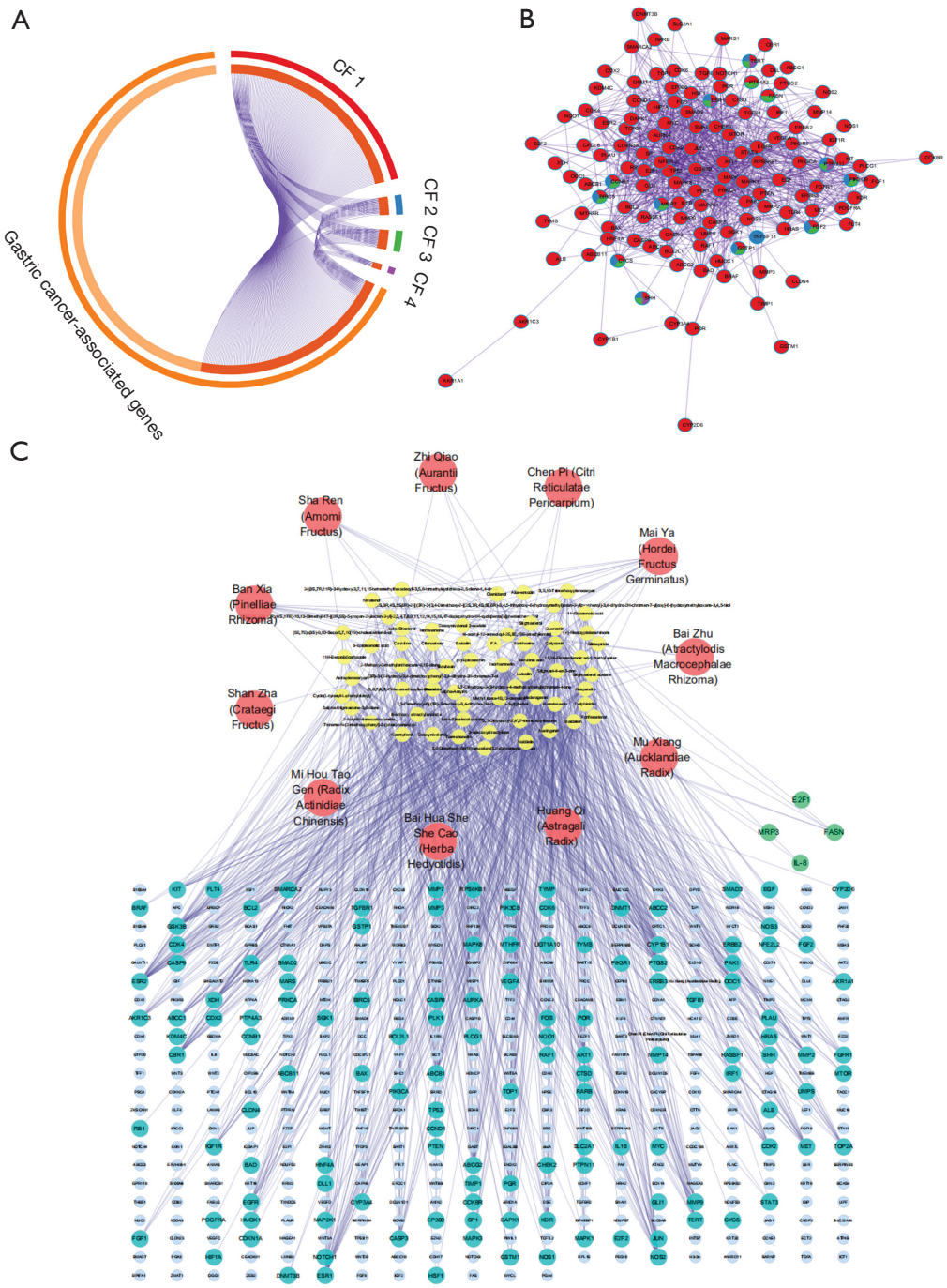


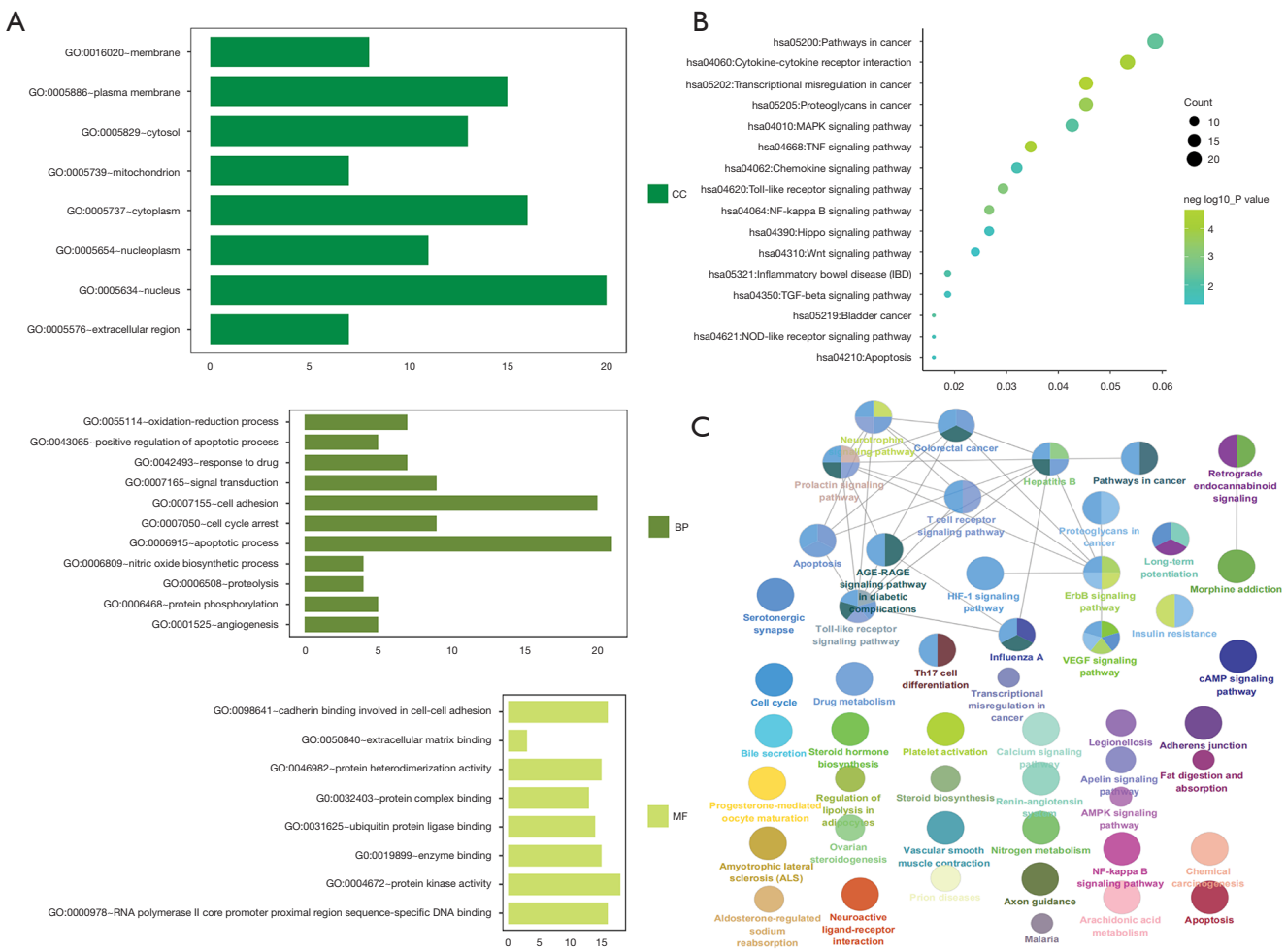
Figure 3 Target mapping of the candidate formulae to GC-related genes and compound-target network construction of the preferred prescription. (A) Overlapping diagram. The segments of the outside circle represent GC-associated genes (light orange), CF 1 targets (red), CF 2 targets (blue), CF 3 (green) and CF 4 targets (purple). The inside circle, specifically, the dark orange segments represent the overlapping parts. (B) PPI network of GC-related genes of the candidate formulae. The nodes represent GC-related genes/proteins from CF 1 targets (red), CF 2 targets (blue), CF 3 (green), and CF 4 targets (purple). Nodes with more than 1 color represent the shared genes/proteins among different formulae. (C) Compound-target network of the preferred prescription. The nodes represent Chinese herbs (red ellipse), active components (yellow ellipse), GC-associated genes (light blue ellipse), GC-related (turquoise ellipse), and GC-unrelated (green ellipse) predicted targets of the preferred prescription. GC, gastric cancer; CF, candidate formula; PPI, protein-protein interaction.

significant cellular components (CC) (P<0.05) included nucleus (GO:0005634), cytoplasm (GO:0005737), plasma membrane (GO:0005886), and cytosol (GO:0005829).

The significant KEGG pathways were mainly the pathways in cancer, T cell receptor signaling pathway, Toll-like receptor signaling pathway, apoptosis, and the VEGF signaling pathway (*Figure 4B*). Cross-talk pathways network of the preferred prescription is shown in *Figure 4C*. Moreover, a multi-regulation map of KEGG pathways was demonstrated, indicating the preferred prescription may exert inhibition on both tumorigenesis and progression of GC (*Figure 4D*).

Modular characteristics and molecular mechanism of the preferred prescription for GC treatment

To elucidate modular characteristics of the preferred prescription, we summarized the significantly enriched BP, KEGG signaling, and reactome pathway of the decoction, which was mainly distributed in the modules of immune regulation, epithelial-mesenchymal transition (EMT), and cell apoptosis/cell cycle (*Table 3*). Then, an herb-crucial compound-biological functional module-molecule network was constructed to determine the relationships among these elements (*Figure 5A*). For example, *Atractylodis Macrocephalae Rhizoma*, a core herb in different combinational patterns, is known to invigorate the Spleen and tonify the vital-*Qi*, which predominantly regulates the immune module. *Astragali Radix*, *Pinelliae Rhizoma*, *Citri Reticulatae Pericarpium*, *Amomi Fructus*, *Hordei Fructus Germinatus*, and *Aurantii Fructus* are known to regulate the movement of Qi, promote blood circulation, and disperse blood stasis, which mainly regulate the EMT module. In addition, GeneMANIA was used to analyze the interactions among the significant targets, which were enriched in the pathways of each functional biological module. The results



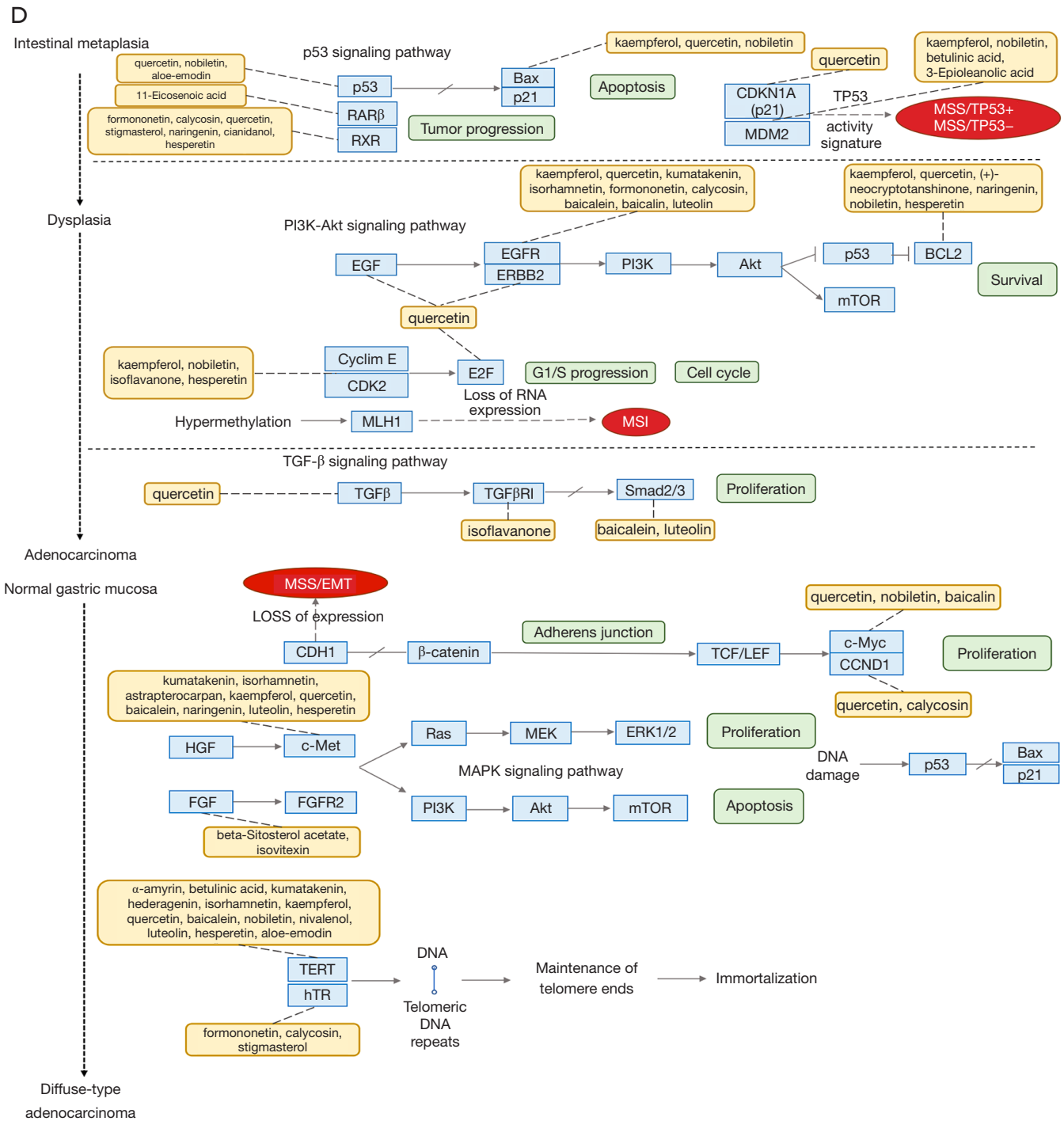


Figure 4 GO and pathway enrichment analysis of the preferred prescription. (A) GO analysis of significant BP, MF, and CC (P<0.05). (B) The significant KEGG pathways (P<0.05). (C) Cross-talk pathways network of the preferred prescription. The nodes represent KEGG pathway terms (P<0.05), and the closer colors they have, the more similar potential functions they possess. The size of nodes represents the enrichment significance of KEGG pathway terms. (D) Multiregulation map of KEGG pathways reflects interactions among crucial components and targets overlapped with GC-related genes. GO, Gene Ontology; BP, biological process; MF, molecular function; CC, cellular component; KEGG, Kyoto Encyclopedia of Genes and Genomes; GC, gastric cancer; MSS, microsatellite stability; MSI, microsatellite instability; EMT, epithelial-mesenchymal transition.

Table 3 Modular functional pathways of the targets in the preferred prescription

Module	Type	GO biological process/KEGG signaling pathway/reactome pathway	P value (Benjamini adjusted)
Immune	Reactome	Immune system	6.18E-05
	KEGG	T cell receptor signaling pathway	5.71E-09
	KEGG	Toll-like receptor signaling pathway	8.98E-08
	KEGG	NOD-like receptor signaling pathway	2.66E-04
Cell apoptosis/cell cycle	KEGG	Apoptosis	1.35E-06
	KEGG	Cell cycle	9.49121E-08
Epithelial mesenchymal transition	GO	Angiogenesis	0.000915735
	Reactome	Extracellular matrix organization	3.78222E-05
	KEGG	Focal adhesion	4.85619E-08
	KEGG	Wnt signaling pathway	0.005181823

GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; NOD, nucleotide-binding and oligomerization domain.

indicated that the preferred prescription has the exact substance basis to regulate the biological modules related to the pathophysiology of GC.

To confirm the molecular mechanism underlying the preferred prescription, molecular docking was performed. Based on the sub-network between the crucial substances and predominant targets of the preferred prescription (*Figure 5B*), we found that both nobiletin and kaempferol have strong affinity with TERT and MDM2 molecules, and p53 may function as a downstream target (*Figure 5C-5F*).

The preferred prescription suppressed GC proliferation and induced cell apoptosis

The typical HPLC-DAD chromatogram of all 12 major components in the preferred prescription is shown in *Figure 6*. The calycosin 7-O-glucoside, rutin, narirutin, naringin, hesperidin, neohesperidin, calycosin, naringenin, kaempferol, formononetin, nobiletin and atractylenolide II contents in the decoction were determined as 0.010, 0.014, 0.229, 0.214, 0.253, 0.239, 0.002, 0.002, 0.004, 0.001, 0.003, and 0.002 mg/g, respectively.

To determine the effect of the preferred prescription on GC cells, the AGS, HGC27, MKN28, and SGC7901 cell activities were assessed by MTT assay. As shown in *Figure 7A*, the cellular viabilities of the 4 GC cell lines were all significantly inhibited. In nude mouse xenograft models, we further validated that the preferred prescription-treated mice showed dramatically decreased tumor weights compared to the control (*Figure 7B*,*7C*). Notably, body mass did not change over the preferred prescription treatment time courses and mice appeared healthy over the duration of the experiments, suggesting that no significant adverse side-effects were experienced (*Figure 7D*, Figure S1). These results suggested that the preferred prescription treatment suppressed GC proliferation *in vivo and* in vitro.

Inducing the apoptosis of cancer cells is a vital way for anticancer drugs to take effect. Therefore, the GC cells were treated with different concentrations of the preferred prescription, and Annexin V-positive cells were detected by FITC analysis to evaluate whether the preferred prescription could induce apoptotic cell death. As shown in *Figure 7E*,*7F*, the preferred prescription significantly and dose-dependently increased the apoptosis rates of GC cells compared to the control. We also detected that the preferred prescription significantly increased the expression of pro-apoptotic protein Bax and decreased the expression of antiapoptotic Bcl2 protein in a concentration-dependent manner (*Figure 7G-7I*). Similar results were found *in vivo* (*Figure 7J*,*7K*). Taken together, our findings indicated that the preferred prescription inhibited the growth of GC by inducing apoptosis both *in vivo* and *in vitro*.

The preferred prescription induced GC cell cycle arrest via hTERT/MDM2-p53 signaling pathway

From the results of KEGG pathway enrichment and the sub-network among the crucial substances and

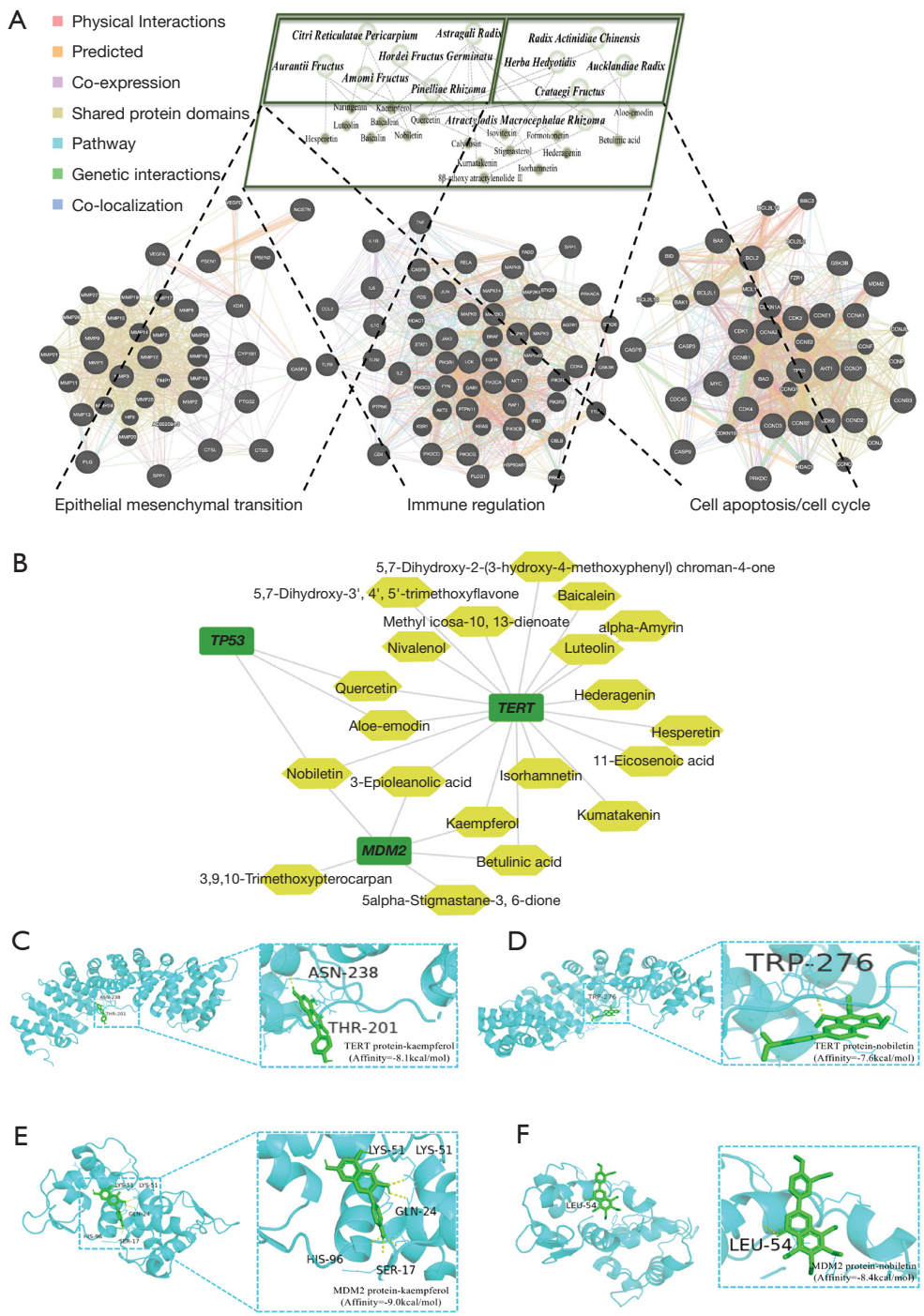


Figure 5 Modular characteristics and molecular docking of the preferred prescription. (A) Herb-key compound-biological functional module-molecule network. Gray dotted line stands for the predicted relationship between herb and crucial compound. Black dotted line stands for the predicted relationship between herb and functional module of the predictive targets. The prefuse force directed layout of the PPI network by GeneMANIA is based on edge betweenness score. The black nodes represent queried proteins. The network weighting of relationships between proteins are shown at the left top. (B) Sub-network among the crucial substances and predominant targets. Molecular docking scores of kaempferol and nobiletin with TERT and MDM2 protein targets were −8.1 (C), −7.6 (D), −9.0 (E), −8.4 (F) kcal/mol, respectively. PPI, protein-protein interaction; TERT, telomerase reverse transcriptase; MDM2, murine double minute 2.

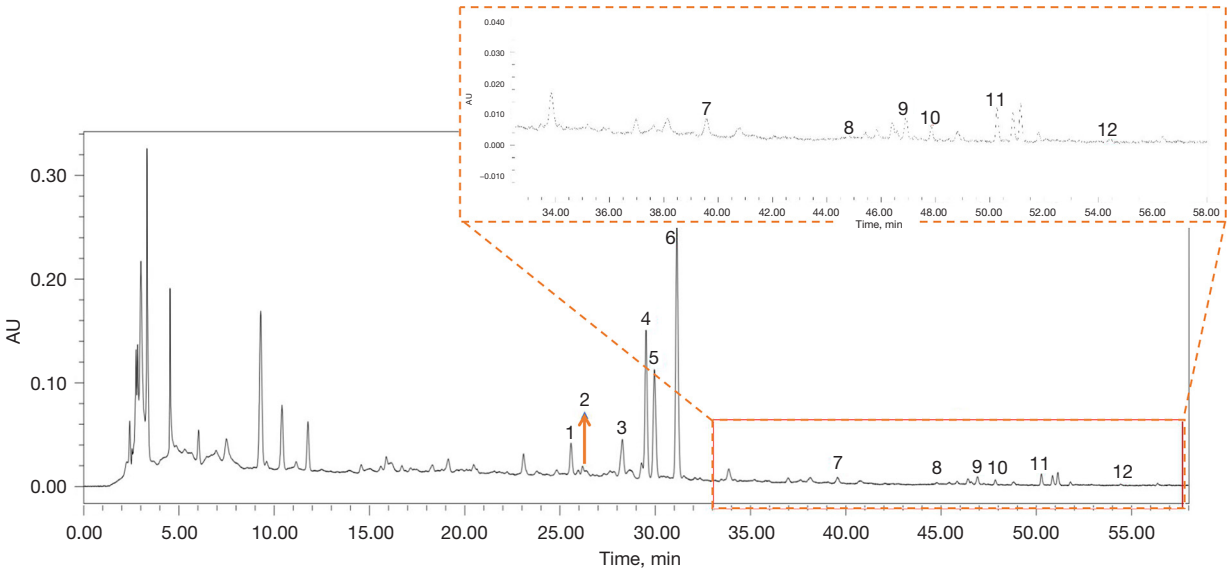


Figure 6 The HPLC-DAD chromatogram of the main components in the preferred prescription. 1: calycosin 7-O-glucoside; 2: rutin; 3: narirutin; 4: naringin; 5: hesperidin; 6: neohesperidin; 7: calycosin; 8: naringenin; 9: kaempferol; 10: formononetin; 11: nobiletin; 12: atractylenolide II. The contents of them in the preferred prescription were 0.010, 0.014, 0.229, 0.214, 0.253, 0.239, 0.002, 0.002, 0.004, 0.001, 0.003, and 0.002 mg/g respectively. HPLC-DAD, high-performance liquid chromatography diode array detection.

the predominant targets, the effects of the preferred prescription were evaluated on the hTERT/MDM2-p53 signaling pathway. The activation of the p53 protein initiates a program of cell cycle arrest, cellular senescence, or apoptosis (34). During different phases of cell cycle, p53 controls both the G1 and G2/M checkpoints (35). Therefore, we performed flow cytometry assay to evaluate whether the preferred prescription modulated the cell cycle of GC cells. As shown in *Figure 8A*,*8B*, the proportion of MKN28 cells in the G1 phase was increased and the proportion of cells in the S phase was decreased, while there was a significant increment in G2/M in AGS, HGC-27, SGC-7901 cells.

By binding to p53, MDM2 inactivates the suppressive function of the tumor in p53 and prevents it from intervening in the cell cycle (36). Cells lacking TERT possessed elevated p53 levels and transcriptional signatures were consistent with p53 up-regulation. Thus, we examined the lysates of the MKN28 cells treated with different concentrations of the preferred prescription using western blot assay. As shown in *Figure 8C*, the preferred prescription treatment significantly and dose-dependently decreased the expressions of hTERT and MDM2, and significantly increased the expression of p53. A major player in the p53-mediated G1 arrest is the p21 gene product that inhibits cyclin E-cdk2 (34). Therefore, we further detected the expressions of p21, cyclinE, and CDK2. It was found that the preferred prescription significantly increased the expression ratio of p21 and reduced the ratios of cyclinE and CDK2 in MKN28 cells (*Figure 8C*,*8D*). Similar results were verified in nude mouse xenograft models (*Figure 8E*,*8F*). The above findings implied that the preferred prescription induced cell cycle arrest in GC cells via hTERT/MDM2-p53 signaling pathway. It also supported the causal link between the elevated p53 by the preferred prescription and the induction of pro-apoptosis proteins of Bax, and the depletion of anti-apoptosis proteins Bcl-2.

The preferred prescription inhibited EMT of GC cells via the hTERT/MDM2-p53 signaling pathway

In cancer, EMT is associated with tumor initiation, invasion, metastasis, and resistance to therapy (37). The role of p53 in EMT has been well studied (38). Recently, it has been reported that p53, p21, and MDM2 bind to the EMT-inducing transcriptional factors Snail/Slug, and promote its ubiquitin-mediated proteasomal degradation (39,40). Based on these existing studies and our findings above, we hypothesized that the preferred prescription also exerted inhibition on EMT of GC via the hTERT/MDM2-p53 signaling pathway. We performed wound healing assay, which revealed that the preferred prescription

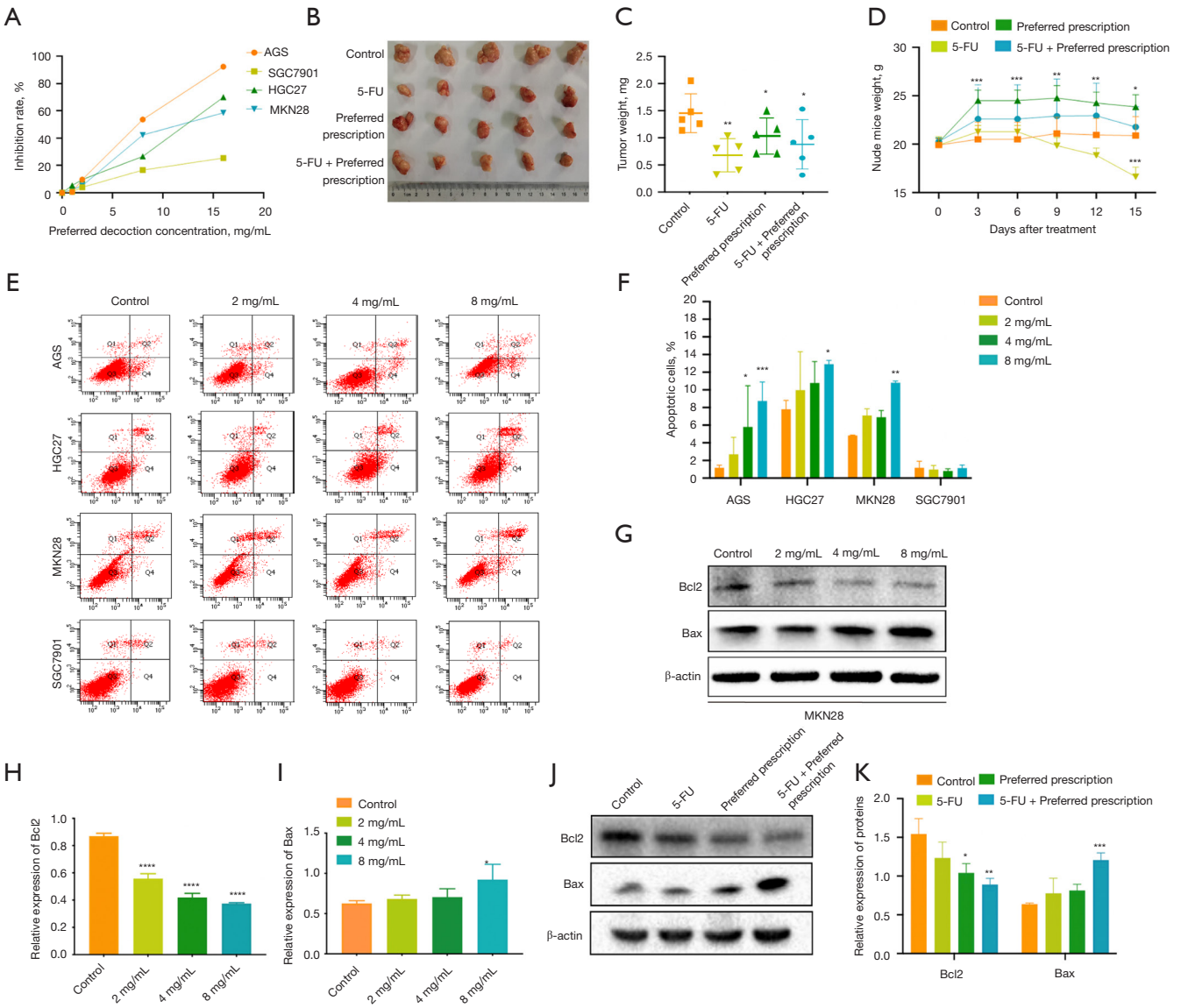


Figure 7 The preferred prescription suppressed GC cells proliferation and induced cell apoptosis. (A) MTT assay showing a concentration-dependent effect of the preferred prescription on the viability of AGS, HGC27, MKN28, and SGC7901 cells. (B) The inhibitory effect of the preferred prescription on the tumor growth of nude mouse xenograft models. The weights of the tumors (C), and the weights of the nude mice (D) were monitored (n=5). (E,F) Flow cytometry depicting cell apoptosis of GC cells treated with different concentrations of the preferred prescription. (G-I) The pro- and anti-apoptotic proteins were detected in MKN28 cells treated with different concentrations of the preferred prescription by western blotting. (J-K) The pro- and antiapoptotic proteins were detected in tumors of nude mouse xenograft models (n=5) by western blotting; *, P<0.05, **, P<0.01, ***, P<0.001, ****, P<0.0001 *vs.* Control. GC, gastric cancer; MTT, 3-(4,5-Dimethyl-2-thizolyl)-2,5- diphenyltetrazolium bromide.

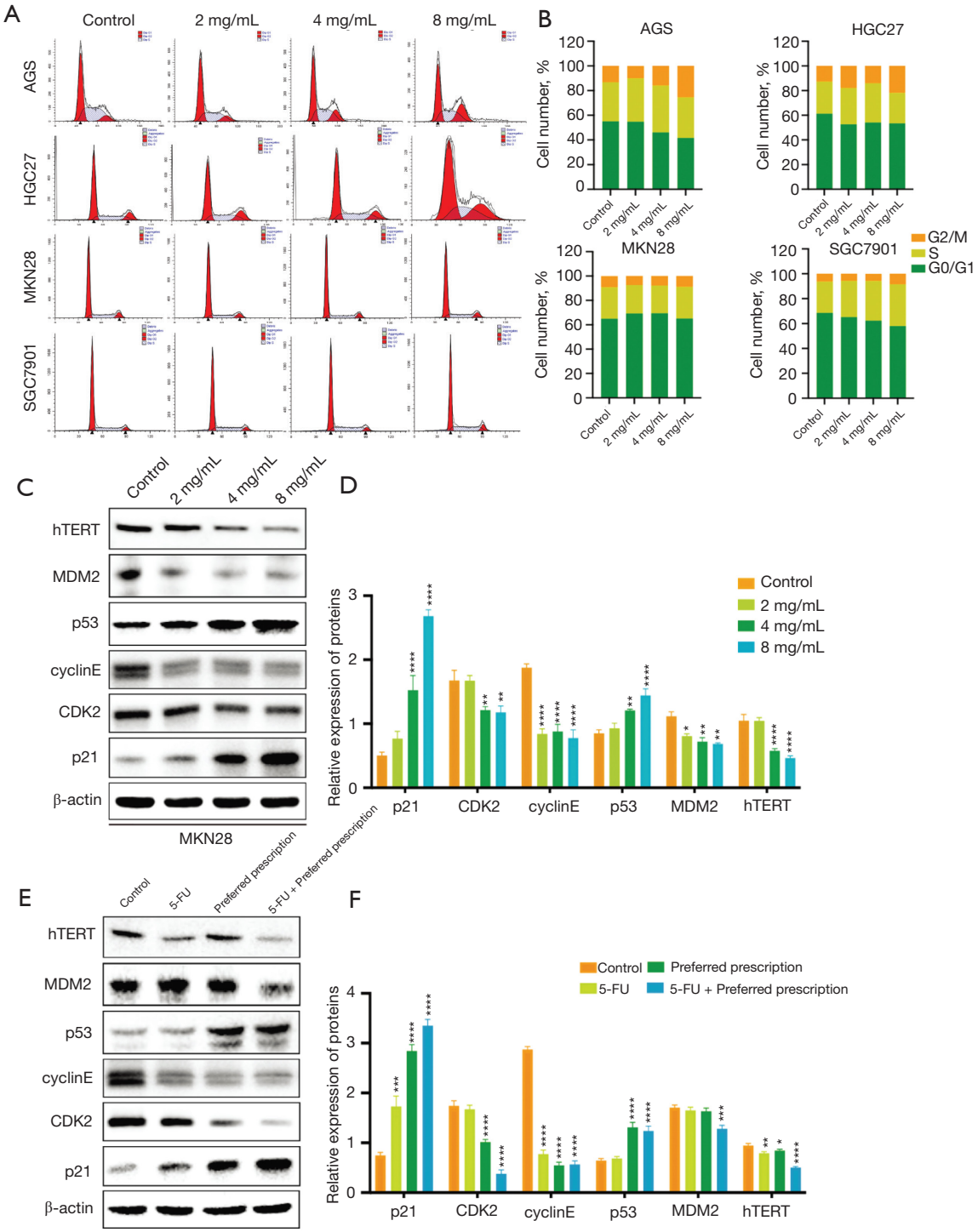


Figure 8 The preferred prescription induced GC cell cycle arrest via hTERT/MDM2-p53 signaling pathway. (A,B) Flow cytometry depicting the effect of the preferred prescription on cell cycle checkpoints in GC cells treated with different concentrations of the preferred prescription. The expressions of hTERT, MDM2, p53, p21, cyclinE, and CDK2 proteins were detected and quantified in MKN28 cells treated with different concentrations of the preferred prescription (C,D), and in extracts from transplanted tumors (n=5) (E,F) by western blotting; *, P<0.05, **, P<0.01, ***, P<0.001, ****, P<0.0001 *vs.* Control. GC, gastric cancer; hTERT, human telomerase reverse transcriptase; MDM2, murine double minute 2; CDK2, cyclin-dependent kinase 2.

remarkably suppressed the migration of MKN28 cells in a concentration-dependent manner, and notably the wound healing area of the preferred prescription (8 mg/mL) group was still large after culturing for 48 h (*Figure 9A,9B*). Transwell assay revealed that the preferred prescription dramatically inhibited GC invasion, even at the lowest concentration of 2 mg/mL (*Figure 9C,9D*). Further, we examined the EMT-related markers, and found that after the preferred prescription treatment, N-cadherin, Snail, and Slug expression ratios were significantly reduced in MKN28 GC cells (*Figure 9E,9F*) and *in vivo* (*Figure 9G,9H*) compared to the controls. Collectively, these data suggest that the preferred prescription may suppress EMT of GC via the hTERT/MDM2-p53 signaling pathway.

Discussion

Globally, GC is the most common cancer and is the leading cause of cancer deaths (1). Therefore, the discovery of novel therapeutic strategies is urgent to enhance the therapeutic effects of existing drugs. Recently, with the development of bioinformatics and network pharmacology, more researchers have applied these methods to unravel the therapeutic effects of TCM formulae (4,41). Elaborately prescribed herbal formulae are being increasingly beneficial for GC patients in relieving adverse events caused by CT or CRT, expediting postoperative recovery, and reducing recurrence or metastasis incidence. However, the modular functional characteristics and molecular mechanisms of TCM in ameliorating GC have remained unclear. In the current study, we conducted a comprehensive data mining of clinical prescriptions, based on which, we obtained a basic TCM prescription with general applicability for GC treatment via machine learning. Based on network pharmacology exploration, the pharmacological mechanism of this preferred prescription against GC was also clarified via experimental verification.

In this study, all the TCM prescriptions for GC treatment from eligible clinical studies over the past 3 decades were collected via CNKI, which is a predominant academic database containing the most comprehensive and authoritative information on TCM. A total of 194 prescriptions were retrieved, among which the most common used couplet herb pairs and triplet herbal combinations were analyzed through ARM. The top 3 recommended herb pairs were all led by *Atractylodis Macrocephalae Rhizoma*. As one of the most potent herbs to invigorate the Spleen and tonify the Qi, it is especially indicated for poor appetite, loose stools, and diarrhea, which are the common manifestations in GC patients, especially after postoperative CT or CRT. When paired with *Dioscoreae Rhizoma*, with the potential to nourish the Yin and tonify the Lung and Kidney, it improves appetite, and mitigates loose stools. When paired with *Aucklandiae Radix*, it promotes digestion and relieves pain. *Codonopsis Radix* can be used with *Atractylodis Mcacrocephalae Rhizoma* in GC patients with weak physique and lassitude, or those experiencing cold pain in the stomach and abdomen, vomiting, or diarrhea (8). In addition, we found that *Poria-Pseudostellariae Radix-Atractylodis Mcacrocephalae Rhizoma* was one of the most recommended triplet herbal combinations. Notably, they are the major compositions of Si Jun Zi decoction, a classic traditional Chinese herbal prescription, which is well known for treating digestive function disorders (42). A deep machine learning method-based cluster analysis was further conducted to obtain an optimized prescription. In this preferred prescription, *Atractylodis Macrocephalae Rhizoma* and *Astragali Radix* were found to be major herbs responsible for tonifying the *Middle-Jiao* and invigorating the Spleen, which could relieve fatigue, poor appetite, loose stools, and other symptoms in GC patients. *Pinelliae Rhizoma*, *Citri Reticulatae Pericarpium*, *Aucklandiae Radix*, *Amomi Fructus*, and *Aurantii Fructus* worked cooperatively to alleviate symptoms like fullness in the abdomen, belching, nausea, and vomiting in GC patients with dampness stagnation. Dyspepsia is very common among GC patients, especially after surgery. Thus, *Galli Gigeriae Endothelium Corneum*, *Hordei Fructus Germinatus*, and *Crataegi Fructus* in the preferred prescription were used to improve digestive dysfunction. For patients enduring a long course of the disease, *Herba Hedyotidis* and *Radix Actinidiae Chinensis* could help in the elimination of internal toxins. Generally, the treatment principles and clinical indications of the preferred prescription are summarized in Figure S2.

Next, we identified the bioactive components and the potential molecule targets of the preferred prescription by network pharmacology. A total of 74 bioactive components were acquired from literature and various public databases, 2,128 genes relevant to the preferred prescription were obtained via target prediction, and 429 GC-related targets were retrieved from disease databases. Eventually, 135 overlapping genes were identified as disease-associated targets. The disease of GC is heterogeneous, whereby the presence of molecular heterogeneity has been described based on anatomic histopathology, the anatomic site, gene expression, and so on (43). The multi-compound and

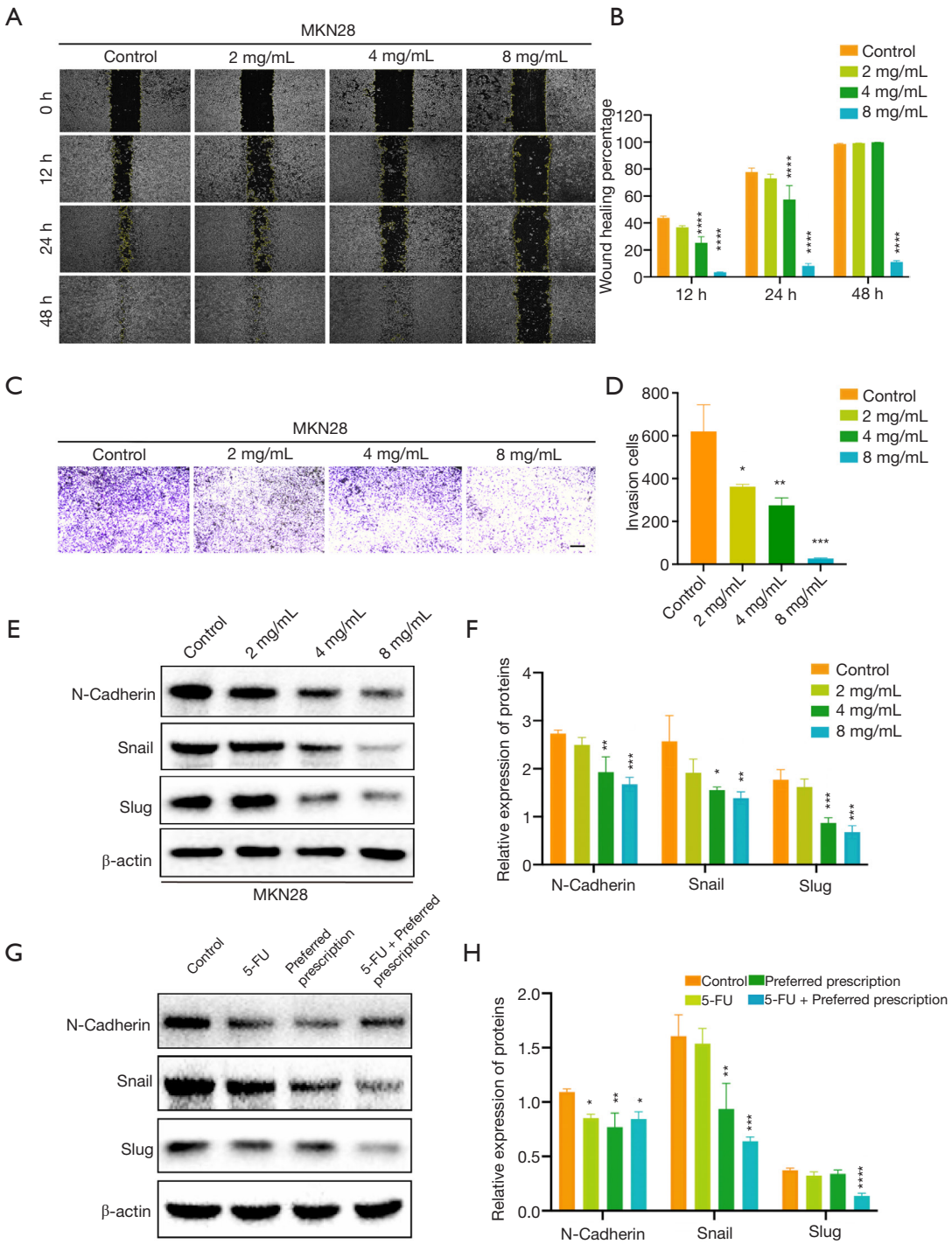


Figure 9 The preferred prescription inhibited EMT of GC cells via the hTERT/MDM2-p53 signaling pathway. Wound healing assay and Transwell assay detecting the migration (A,B) and invasion (C,D) abilities of MKN28 cells treated with different concentrations of the preferred prescription (scale bar =80 μm). The expressions of EMT-related markers were detected and quantified in MKN28 cells treated with different concentrations of the preferred prescription (E,F), and in extracts of the transplanted tumors (n=5) (G,H) by western blotting; *, P<0.05, **, P<0.01, ***, P<0.001, ****, P<0.0001 *vs.* Control. EMT, epithelial-mesenchymal transition; GC, gastric cancer; hTERT, human telomerase reverse transcriptase; MDM2, murine double minute 2.

multi-target characteristics of the preferred prescription potentiate its multiple biological functions in treating GC. To acquire an in-depth understanding of the overlapping targets, GO, KEGG, and reactome pathway enrichment were performed. The modular functional network revealed part of the combinational rules of herbs in the context of biological functional molecules (*Figure 5A*). In this study, the crucial compounds ranked by degree were quercetin, kaempferol, baicalein, nobiletin, and luteolin. Existing studies on these bioactive substances have shown diverse anti-GC mechanisms. For example, luteolin could shift the Bax/Bcl ratio in human GC cells by increasing the expressions of pro-apoptotic proteins (44-46). Treatment with luteolin was also observed to up-regulate p21/cip1 (CDKN1A), a TP53 activity signature (47). The mixture of *Radix Actinidiae Chinensis* could down-regulate the expressions of stromal cell-derived factor-1 (SDF-1), MMP-2, and MMP-9 in SGC-7901 cells (48). Quercetin was found to restrain transforming growth factor (TGF)-β1-induced EMT by inhibiting Twist1 and regulating E-cadherin expression (49). Also, quercetin-3-methyl ether (Q3ME) is a natural flavonoid compound capable of inhibiting esophageal carcinogenesis by targeting the receptor tyrosine kinases (RTKs) (50). Although there was no literature on some crucial components associated with GC, the efficacy was noteworthy in other cancers. All literature, together with the experimental studies, provided a valuable hint in identifying the action mechanism of the preferred prescription against GC.

From the *in vivo* results, it was revealed that treatment with the preferred prescription significantly suppressed tumor growth compared to the control. Notably, the preferred prescription did not compromise the mice's body weights compared to the 5-FU group, indicating that it has a better safety profile, or at least in part, is favorable for patients who are intolerant of 5-FU treatment. To explore the potential mechanism, multiple biological function assays were conducted *in vitro* in GC cell lines, including AGS, HGC27, MKN28, and SGC-7901. It was demonstrated that the preferred prescription promoted cellular apoptosis and attenuated the metastatic capability in GC cells.

From the results of target prediction and pathway analysis, the preferred prescription might suppress the survival and metastasis of GC cells via the hTERT/MDM2-p53 signaling pathway. The TERT protein is often overexpressed in tumor cells and mediates cellular immortalization (51). Recent research revealed that cells lacking TERT possessed elevated p53 levels and transcriptional signatures were consistent with p53 up-regulation. The up-regulation of the MDM2 oncogene plays a role in the diffuse type of GC (52). By binding to p53, MDM2 inactivates the anti-tumor function of p53 and prevents it from intervening in the cell cycle (36). The activation of p53 induces p53-dependent cell death and p53- and p21-dependent cell cycle arrests, which is characterized by depletion of the S-phase cells and accumulation at the G1/S and/or G2/M phase boundaries of the cell cycle (53). In the present study, the flow cytometry and western blot results supported the prediction. After treatment with the preferred prescription, the proportion of MKN28 cells in the G1 phase was increased and the proportion of cells in the S phase was decreased, while there was a significant increment in G2/M in AGS, HGC-27, and SGC-7901 cells. Additionally, we validated that the preferred prescription exerted negative modulation on the expressions of hTERT and MDM2, and positively modulated the expressions of p53 and p21. The activation of p53 stimulates the synthesis of the p21 protein, which inhibits cyclin E-cdk2 activity, and this in turn acts upon the retinoblastoma (Rb)-MDM2 complex that promotes p53 activity and apoptosis (34). In this study, we demonstrated that after the preferred prescription treatment, increased p53 activity induced the pro-apoptosis protein Bax and depleted the anti-apoptosis protein Bcl-2. We also detected decreased expressions of Slug and Snail under the preferred prescription treatment, which was possibly due to p53, p21, and MDM2 interacting with the EMT-inducing transcriptional factors, and leading to their ubiquitination (39). Taken together, the preferred prescription might play a role in inducing cell cycle arrest, cellular apoptosis, and inhibiting EMT process of GC via the hTERT/MDM2-p53 signaling pathway.

Meanwhile, our research had several limitations. Firstly, the eligible literature in our study was drawn only from Chinese databases. With the development of TCM, we will be able to include more information from other Asian countries like Japan and South Korea. Secondly, the bioactive substances of the Chinese herbs screened in the existing databases need further preclinical and clinical verification. Lastly, a clinical trial on the preferred prescription is required to reliably assess the roles of TCM in the recurrence and metastasis of GC.

Conclusions

To conclude, data mining and machine learning combined with network pharmacology analysis and experimental verification may elucidate the modular functions and

pharmacological mechanisms of TCM on GC from an innovative perspective. It was demonstrated that the preferred prescription may suppress the survival and metastasis of GC cells via modulating the hTERT/MDM2-p53 signaling pathway. Meanwhile, in-depth pharmacological mechanisms by which the preferred prescription ameliorates GC need to be further explored. Also, as the core concept of TCM, syndrome differentiation cannot be completely replaced by the results of machine learning. Hence, clinically effective combinations of herbs should also be encouraged as individualized strategies for GC patients. This study will facilitate the application of TCM in GC treatment with the purpose of improving therapeutic strategy in clinic.

Acknowledgments

Funding: This study was supported by the National Natural Science Foundation of China (81973609, 81704031, 81973782); Jiangsu Provincial Medical Youth Talent (QNRC2016641); Jiangsu Provincial Hospital of Traditional Chinese Medicine Academic Talent Program (Y2018RC33); the Open Projects of the Discipline of Chinese Medicine of Nanjing University of Chinese Medicine Supported by the Subject of Academic priority discipline of Jiangsu Higher Education Institutions (ZYX03KF019, ZYX03KF021, ZYX03KF029); Jiangsu Province Scientific Research and Practice Innovation Program (SJCX21_0738, SJCX20_0568).

Footnote

Reporting Checklist: The authors have completed the ARRIVE reporting checklist. Available at https://dx.doi.org/10.21037/atm-21-6301

Data Sharing Statement: Available at https://dx.doi.org/10.21037/atm-21-6301

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://dx.doi.org/10.21037/atm-21-6301). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). Animal experiment was performed under a project license (No. 2021DW-35-01) granted by the Animal Ethics Committee of Affiliated Hospital of Nanjing University of Chinese Medicine (Nanjing, China), in compliance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health.

References

1. Jemal A, Bray F, Center MM, et al. Global cancer statistics. CA Cancer J Clin 2011;61:69-90.
2. Guo S, Shang MY, Dong Z, et al. A nomogram for predicting cancer-specific survival in different age groups for operable gastric cancer: a population-based study. Transl Cancer Res 2020;9:2758-68.
3. Zhao Y, Wang XW, Lu Y, et al. Effect of Xiaotan Sanjie Decoction treatment based on syndrome differentiation on quality of life of patients with intermediate and advanced gastric cancer. Ti Erh Chun I Ta Hsueh Hsueh Pao 2016;37:1333-7.
4. Pan B, Wang Y, Wu C, et al. A Mechanism of Action Study on Danggui Sini Decoction to Discover Its Therapeutic Effect on Gastric Cancer. Front Pharmacol 2020;11:592903.
5. Catenacci DV, Chao J, Muro K, et al. Toward a Treatment Sequencing Strategy: A Systematic Review of Treatment Regimens in Advanced Gastric Cancer/Gastroesophageal Junction Adenocarcinoma. Oncologist 2021;26:e1704-29.
6. Pharmacopoeia Committee of the People's Republic of China (Beijing). Chinese Pharmacopoeia. 2020 ed. Beijing: Chinese Medical Science and Technology Press, 2020.
7. Liu P, Liu S, Chen G, et al. Understanding channel tropism in traditional Chinese medicine in the context of systems biology. Front Med 2013;7:277-9.
8. Gao XM. Chinese Pharmacy. 8th ed. Beijing: China Traditional Chinese Medicine Publishing Press, 2007.

9. Creighton C, Hanash S. Mining gene expression databases for association rules. Bioinformatics 2003;19:79-86.
10. Agrawal R, Tomasz, Arun S. Mining association rules between sets of items in large databases. Acm Sigmod Conference on Management of Data 1993;22:207-15.
11. Vougas K, Sakellaropoulos T, Kotsinas A, et al. Machine learning and data mining frameworks for predicting drug response in cancer: An overview and a novel in silico screening process based on association rule mining. Pharmacol Ther 2019;203:107395.
12. Leem J, Jung W, Kim Y, et al. Exploring the combination and modular characteristics of herbs for alopecia treatment in traditional Chinese medicine: an association rule mining and network analysis study. BMC Complement Altern Med 2018;18:204.
13. Karim MR, Beyan O, Zappa A, et al. Deep learning-based clustering approaches for bioinformatics. Brief Bioinform 2021;22:393-415.
14. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017;2:230-43.
15. Ru J, Li P, Wang J, et al. TCMSP: a database of systems pharmacology for drug discovery from herbal medicines. J Cheminform 2014;6:13.
16. Liu Z, Guo F, Wang Y, et al. BATMAN-TCM: a Bioinformatics Analysis Tool for Molecular mechANism of Traditional Chinese Medicine. Sci Rep 2016;6:21146.
17. Su X, Kong L, Lei X, et al. Biological fingerprinting analysis of traditional Chinese medicines with targeting ADME/Tox property for screening of bioactive compounds by chromatographic and MS methods. Mini Rev Med Chem 2007;7:87-98.
18. Gfeller D, Michielin O, Zoete V. Shaping the interaction landscape of bioactive molecules. Bioinformatics 2013;29:3073-9.
19. Yu H, Chen J, Xu X, et al. A systematic prediction of multiple drug-target interactions from chemical, genomic, and pharmacological data. PLoS One 2012;7:e37608.
20. Zheng C, Guo Z, Huang C, et al. Large-scale Direct Targeting for Drug Repositioning and Discovery. Sci Rep 2015;5:11970.
21. Li J, Zhao P, Li Y, et al. Systems pharmacology-based dissection of mechanisms of Chinese medicinal formula Bufei Yishen as an effective treatment for chronic obstructive pulmonary disease. Sci Rep 2015;5:15290.
22. Szklarczyk D, Santos A, von Mering C, et al. STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data. Nucleic Acids Res 2016;44:D380-4.
23. Keiser MJ, Roth BL, Armbruster BN, et al. Relating protein pharmacology by ligand chemistry. Nat Biotechnol 2007;25:197-206.
24. Daina A, Michielin O, Zoete V. SwissTargetPrediction: updated data and new features for efficient prediction of protein targets of small molecules. Nucleic Acids Res 2019;47:W357-64.
25. Rappaport N, Twik M, Plaschkes I, et al. MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. Nucleic Acids Res 2017;45:D877-87.
26. Piñero J, Bravo À, Queralt-Rosinach N, et al. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. Nucleic Acids Res 2017;45:D833-9.
27. Dennis G Jr, Sherman BT, Hosack DA, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol 2003;4:P3.
28. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28:27-30.
29. Bindea G, Mlecnik B, Hackl H, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics 2009;25:1091-3.
30. Warde-Farley D, Donaldson SL, Comes O, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res 2010;38:W214-20.
31. Morris GM, Huey R, Lindstrom W, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem 2009;30:2785-91.
32. O'Boyle NM, Banck M, James CA, et al. Open Babel: An open chemical toolbox. J Cheminform 2011;3:33.
33. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009;6:e1000097.
34. Wu D, Prives C. Relevance of the p53-MDM2 axis to aging. Cell Death Differ 2018;25:169-79.
35. Agarwal ML, Agarwal A, Taylor WR, et al. p53 controls both the G2/M and the G1 cell cycle checkpoints and mediates reversible growth arrest in human fibroblasts. Proc Natl Acad Sci U S A 1995;92:8493-7.
36. Momand J, Zambetti GP, Olson DC, et al. The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation. Cell 1992;69:1237-45.
37. Pastushenko I, Blanpain C. EMT Transition States during

Tumor Progression and Metastasis. Trends Cell Biol 2019;29:212-26.
38. Powell E, Piwnica-Worms D, Piwnica-Worms H. Contribution of p53 to metastasis. Cancer Discov 2014;4:405-14.
39. Wang SP, Wang WL, Chang YL, et al. p53 controls cancer cell invasion by inducing the MDM2-mediated degradation of Slug. Nat Cell Biol 2009;11:694-704.
40. Kim J, Bae S, An S, et al. Cooperative actions of p21WAF1 and p53 induce Slug protein degradation and suppress cell invasion. EMBO Rep 2014;15:1062-8.
41. Wang T, Feng Y, Wang H, et al. The Mechanisms of Sijunzi Decoction in the Treatment of Chronic Gastritis Revealed by Network Pharmacology. Evid Based Complement Alternat Med 2020;2020:8850259.
42. Liu L, Han L, Wong DY, et al. Effects of Si-Jun-Zi decoction polysaccharides on cell migration and gene expression in wounded rat intestinal epithelial cells. Br J Nutr 2005;93:21-9.
43. Cristescu R, Lee J, Nebozhyn M, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. Nat Med 2015;21:449-56.
44. Lu J, Li G, He K, et al. Luteolin exerts a marked antitumor effect in cMet-overexpressing patient-derived tumor xenograft models of gastric cancer. J Transl Med 2015;13:42.
45. Liu JF, Ma Y, Wang Y, et al. Reduction of lipid accumulation in HepG2 cells by luteolin is associated with activation of AMPK and mitigation of oxidative stress. Phytother Res 2011;25:588-96.
46. Lee WJ, Wu LF, Chen WK, et al. Inhibitory effect of luteolin on hepatocyte growth factor/scatter factor-induced HepG2 cell invasion involving both MAPK/ERKs and PI3K-Akt pathways. Chem Biol Interact 2006;160:123-33.
47. Imran M, Rauf A, Abu-Izneid T, et al. Luteolin, a flavonoid, as an anticancer agent: A review. Biomed Pharmacother 2019;112:108612.
48. Zhang F, Zhang N, Liang F, et al. Effects of Tengligen Mixture on Expression of MMP-2,MMP-9 and SDF-1 in Gastric Cancer SGC-7901 Cells. Acta Chinese Medicine 2018;33:175-80.
49. Feng J, Song D, Jiang S, et al. Quercetin restrains TGF-β1-induced epithelial-mesenchymal transition by inhibiting Twist1 and regulating E-cadherin expression. Biochem Biophys Res Commun 2018;498:132-8.
50. Zhao S, Jiang Y, Zhao J, et al. Quercetin-3-methyl ether inhibits esophageal carcinogenesis by targeting the AKT/mTOR/p70S6K and MAPK pathways. Mol Carcinog 2018;57:1540-52.
51. Konnikova L, Simeone MC, Kruger MM, et al. Signal transducer and activator of transcription 3 (STAT3) regulates human telomerase reverse transcriptase (hTERT) expression in human cancer and primary cells. Cancer Res 2005;65:6516-20.
52. Günther T, Schneider-Stock R, Häckel C, et al. Mdm2 gene amplification in gastric cancer correlation with expression of Mdm2 protein and p53 alterations. Mod Pathol 2000;13:621-6.
53. Shangary S, Wang S. Small-molecule inhibitors of the MDM2-p53 protein-protein interaction to reactivate p53 function: a novel approach for cancer therapy. Annu Rev Pharmacol Toxicol 2009;49:223-41.

Cite this article as: Xu X, Chen Y, Zhang X, Zhang R, Chen X, Liu S, Sun Q. Modular characteristics and the mechanism of Chinese medicine's treatment of gastric cancer: a data mining and pharmacology-based identification. Ann Transl Med 2021;9(24):1777. doi: 10.21037/atm-21-6301

Supporting Information

Supporting materials and methods

1. HPLC-DAD analysis of the preferred prescription

To facilitate the identification of the preferred prescription components, the preferred prescription aqueous extract was analyzed by the HPLC-DAD. The extracts of the preferred prescription were filtered with a 0.22 μm filter membrane, a Phenomenex Synergi 4μ Fusion-RP 80A (250 mm*4.6 mm, 5 μm) column was used for chromatographic separation. The mobile phase was composed of acetonitrile (A) and aqueous solution (B). The linear elution gradient was 0-40 min, 5–35% A; 40–50 min, 35–70% A; 50–55 min, 70–90% A; and 55–58 min, 90–5% A. The injection volume was 10 μL, and the flow rate was 1.0 mL/min. The reference compounds were precisely weighted and diluted in methanol with fixed volume. The detection wavelengths were as follows: 248 nm (calycosin 7-O-glucoside, calycosin, and formononetin), 356 nm (rutin), 282 nm (narirutin, naringin, hesperidin, and neohesperidin), 288 nm (naringenin), 266 nm (kaempferol), 330 nm (nobiletin), and 220 nm (atractylenolide II). The linear range, average recovery, precision, exclusivity, and stability data were analyzed.

HPLC-DAD, high-performance liquid chromatography diode array detection

2. MTT assay

AGS, HGC-27, MKN28, SGC-7901 cells (5 × 103 cells/well) were seeded into 96-well plates separately for 24 h to allow adherence to the walls. To determine the decoction dose, GC cells were treated with different concentrations (0, 1, 2, 8, 16 mg/mL) of the preferred prescription for 48 h. MTT (120 μL, 5 mg/mL) (Sigma, USA) was added after the medium was removed, and the cells were incubated for 4 h in the incubator. The supernatant was removed, and 150 mL of dimethyl sulfoxide (DMSO) was added for 10 min. Absorbance at 490 nm was detected on an ELX800 Automatic microplate reader (Bio-Tek, USA) to calculate the absorbance (OD490). The IC50 was calculated by Graphpad software.

MTT, 3-(4,5-Dimethyl-2-thizolyl)-2,5- diphenyltertazolium bromide; IC50, half-maximal inhibitory concentration

https://dx.doi.org/10.21037/atm-21-6301

Table S1 Top 24 principal function-categorized herbs

No.	Herb	Functional category	Occurrence frequency (%)	Number of occurrences	Frequency of use (%)	Classification category
1	*Atractylodis Macrocephalae Rhizoma*	a1	26.37	145	74.74	Tonifying and replenishing medicinal
2	*Poria*	i	10.07	126	64.95	Dampness-draining diuretic medicinal
3	*Astragali Radix*	a1	26.37	107	55.15	Tonifying and replenishing medicinal
4	*Codonopsis Radix*	a1	26.37	105	54.12	Tonifying and replenishing medicinal
5	*Glycyrrhizae Radix*	a1	26.37	96	49.48	Tonifying and replenishing medicinal
6	*Pinelliae Rhizoma*	e	6.51	85	43.81	Cough-suppressing and panting-calming medicinal
7	*Citri Reticulatae Pericarpium*	d	8.74	77	39.69	Qi-regulating medicinal
8	*Coicis Semen*	i	10.07	63	32.47	Dampness-draining diuretic medicinal
9	*Herba Hedyotidis Diffusae*	b1	8.08	54	27.84	Heat-clearing medicinal
10	*Angelicae Sinensis Radix*	a3	5.04	48	24.74	Tonifying and replenishing medicinal
11	*Galli Gigeriae Endothelium Corneum*	h	5.51	38	19.59	Digestant medicinal
12	*Curcumae Rhizoma*	c	6.89	34	17.53	Blood-activating and stasis-dispelling medicinal
13	*Aucklandiae Radix*	d	8.74	30	15.46	Qi-regulating medicinal
14	*Pseudostellariae Radix*	a1	26.37	29	14.95	Tonifying and replenishing medicinal
15	*Paeoniae Radix Alba*	a3	5.04	29	14.95	Tonifying and replenishing medicinal
16	*Scutellariae Barbatae Herba*	b1	8.08	29	14.95	Heat-clearing medicinal
17	*Amomi Fructus*	k	2.57	28	14.43	Dampness-resolving medicinal
18	*Dioscoreae Rhizoma*	a1	26.37	26	13.40	Tonifying and replenishing medicinal
19	*Hordei Fructus Germinatus*	h	5.51	25	12.89	Digestant medicinal
20	*Salviae Miltiorrhizae Radix et Rhizoma*	c	6.89	24	12.37	Blood-activating and stasis-dispelling medicinal
21	*Ligustri Lucidi Fructus*	a2	4.18	24	12.37	Tonifying and replenishing medicinal
22	*Aurantii Fructus*	d	8.74	23	11.86	Qi-regulating medicinal
23	*Radix Actinidiae Chinensis*	m	2.38	23	11.86	Wind-dampness dispelling medicinal
24	*Crataegi Fructus*	h	5.51	21	10.82	Digestant medicinal

Occurrence frequency = number of occurrences for the herbs appearing in 194 prescriptions / total cumulative occurrences for 148 herbs appearing in 194 prescriptions (i.e. 2,103); Frequency of use = number of prescriptions recording the herbs in use/ total number of the eligible prescriptions (i.e. 194). Abbreviations: a1: Qi-tonifying medicinal; a2: Yin-tonifying medicinal; a3: Blood-tonifying medicinal; a4: Yang-tonifying medicinal; b: heat-clearing medicinal; c: blood-activating and stasis-dispelling medicinal; d: Qi-regulating medicinal; e: cough-suppressing and panting-calming medicinal; f: interior-warming medicinal; g: Liver-pacifying medicinal; h: digestant medicinal; i: dampness-draining diuretic medicinal; j: exterior-releasing medicinal; k: dampness-resolving medicinal; l: hemostatic medicinal; m: wind-dampness dispelling medicinal; n: astringent medicinal; o: purgative medicinal; p: orifice-opening medicinal; q: repellent medicinal; r: attacking poison, insects and itch-relieving medicinal.

https://dx.doi.org/10.21037/atm-21-6301

Table S2 Principal functional categorizations and clinical application of the most recorded herbs

Herbal nature	Principal functional categorizations	Number of decoctions using the herbs	Frequency of use (%)	Syndromes	Key signs & symptoms	Treatment principles	Representative herbs
Sweet, warm	Qi-tonifying	182	93.81	*Middle-Jiao* Deficiency	Poor appetite, dislike to talk, lassitude, weak limbs, borborygmus, loose stools, heavy descending sensation in abdominal cavity, prolapse of rectum.	Tonify Middle-Jiao *Qi*	*Atractylodis Macrocephalae Rhizoma, Astragali Radix, Codonopsis Radix, Glycyrrhizae Radix, Pseudostellariae Radix, Dioscoreae Rhizoma*
Bitter, warm/ pungent, warm	Dampness-draining	149	76.80	Cold damp obstructs Spleen	Fullness sensation in upper abdomen, poor appetite, sticky sensation in mouth, heavy sensation on head, loose stools or diarrhea.	Tonify Spleen to transform damp	*Poria, Coicis Semen, Amomi Fructus*
				Damp-heat stagnates interiorly	Full abdomen & hypochondrium distension, no desire for food, bitter taste in mouth, thirsty, heavy sensation on body, yellow urine, loose stools, jaundice, itchy skin.	Remove damp and heat	
				Damp from Spleen affecting Lung	Cough/vomiting of phlegm, saliva, congestion in chest, shortness of breath (SOB), poor appetite.	Dry damp, remove phlegm	
Pungent, warm/ bitter, warm	Qi-regulating	119	61.34	Stagnation of Liver *Qi*	Mental depression, restlessness, sighing, distension, wandering pain in the costal and hypochondriac region, distress in epigastrium, poor appetite or vomiting, irregular bowel movements, thin greasy tongue coating, wiry pulse.	Disperse Liver *Qi*	*Citri Reticulatae Pericarpium, Aucklandiae Radix, Aurantii Fructus, Fructus Evodiae*
				Qi stagnation transforming into *Fire*	Irritability, stuffiness in the chest, hypochondriac distension, acid regurgitation, dry & bitter mouth, constipation or headache, tinnitus, red tongue & yellow coating, wiry-rapid pulse.	Purge *Fire* from Liver	
Pungent, warm/ bitter, warm	Blood-activating and stasis-dispelling	84	42.27	*Qi* stagnation, *Blood* stasis	Moving/fixed pain, distending pain on hypochondrium, masses in abdominal cavity, stabbing pain aggravated by pressure, purplish tongue body, purple spots; thready, string-taut pulse.	Invigorate blood circulation, eliminate blood stasis	*Curcumae Rhizoma, Salviae Miltiorrhizae Radix et Rhizoma*
Sweet, warm/sweet, cold	Blood-tonifying	69	35.57	*Blood* deficiency	Pale or yellowish complexion, pale lustreless lips & nails, dizziness, vertigo, palpitations, insomnia, numbness of limbs, pale tongue body, thready-weak pulse.	Replenish *Blood*	*Angelicae Sinensis Radix,Paeoniae Radix Alba*
Sweet, neutral	Food abating	55	28.35	Stomach excessive	Epigastric and abdominal distension and fullness or pain, which are aggravated by food intake, belching with foul smell, anorexia, constipation, acidic regurgitation, nausea, vomiting, diarrhea with foul smell or fermented contents or constipation.	Dissolve the stagnation	*Galli Gigeriae Endothelium Corneum, Hordei Fructus Germinatus, Crataegi Fructus*
Sweet, cold	Yin-tonifying	55	28.35	Stomach *Yin* deficiency	Dry mouth & lips, thirsty, hunger but no desire for food, retching, hiccups, constipation/ dry stools, red dry tongue with little coating or mirror red tongue, thready-rapid pulse.	Tonify Stomach *Yin* fluids	*Herba Dendrobii, Rhizoma Polygonati Odorati*
Sweet, warm/ pungent, warm	Yang-tonifying	27	13.92	Spleen *Yang* deficiency	Lustreless withered yellow complexion, cold sensation at epigastric region, vomiting clear water, poor appetite, distension, preference for hot drinks, etc.	Warm *Middle-Jiao Yang*	*Semen Cuscutae, Fructus Psoraleae, Semen Myristicae, Rhizoma Zingiberis*
				Yang deficiency of Spleen & Kidney	SOB, dislike to talk, cold & sore loins & knees, pre-dawn diarrhea.	Warm Spleen & Kidney *Yang*	
				Cold in Stomach	Cold pain in Stomach, aggravated by cold, relieved by warmth, no thirst, vomiting clear water, hiccup.	Warm Stomach to dispel cold	

Frequency of use = number of prescriptions recording the herbs in use/ total number of the eligible prescriptions (i.e. 194). SOB, shortness of breath.

https://dx.doi.org/10.21037/atm-21-6301

Table S3 Top 10 triplet herbal combinations

Herb (LHS)	Number of decoctions	Herbs (RHS)		Number of *occurrences*	*Support* (LHS) (%)	*Confidence* (LHS=>RHS) (%)	*LIFT*
Poria	126	*Aucklandiae Radix, Glycyrrhizae Radix*	→	21	10.82	100.00	1.54
Atractylodis Macrocephalae Rhizoma	145	*Aucklandiae Radix, Codonopsis Radix*	→	20	10.31	100.00	1.34
Atractylodis Macrocephalae Rhizoma	145	*Galli Gigeriae Endothelium Corneum, Codonopsis Radix*	→	20	10.31	100.00	1.34
Atractylodis Macrocephalae Rhizoma	145	*Coicis Semen, Glycyrrhizae Radix*	→	29	14.95	96.55	1.29
Atractylodis Macrocephalae Rhizoma	145	*Aucklandiae Radix, Poria*	→	25	12.89	96.00	1.28
Poria	126	*Pseudostellariae Radix, Atractylodis Macrocephalae Rhizoma*	→	24	12.37	95.83	1.48
Atractylodis Macrocephalae Rhizoma	145	*Hordei Fructus Germinatus, Codonopsis Radix*	→	21	10.82	95.24	1.27
Codonopsis Radix	105	*Hordei Fructus Germinatus, Atractylodis Macrocephalae Rhizoma*	→	21	10.82	95.24	1.76
Atractylodis Macrocephalae Rhizoma	145	*Aucklandiae Radix, Glycyrrhizae Radix*	→	21	10.82	95.24	1.27
Atractylodis Macrocephalae Rhizoma	145	*Paeoniae Radix Alba, Poria*	→	20	10.31	95.00	1.27

Occurrence= number of occurrences of the herbal pairs appearing in the eligible prescriptions (i.e. 194). LHS, left hand side; RHS, right hand side

https://dx.doi.org/10.21037/atm-21-6301

Table S4 Bioactive compounds of the 24 core herbs

Molecule ID	Compound	Abbr. of Compounds	MW	OB (%)	DL	Herb
MOL000028	alpha-Amyrin	01BZ	426.8	39.51	0.76	*Atractylodis Macrocephalae Rhizoma*
MOL000049	3β-acetoxyatractylone	02BZ	274.39	54.07	0.22	*Atractylodis Macrocephalae Rhizoma*
MOL000072	8β-ethoxy atractylenolide III	03BZ	276.41	35.95	0.21	*Atractylodis Macrocephalae Rhizoma*
MOL000211	Betulinic acid	01HQ	456.78	55.38	0.78	*Astragali Radix*
MOL000239	Kumatakenin	02HQ	314.31	50.83	0.29	*Astragali Radix*
MOL000296	Hederagenin	03HQ	414.79	36.91	0.75	*Astragali Radix*
MOL000354	Isorhamnetin	04HQ	316.28	49.6	0.31	*Astragali Radix*
MOL000387	Bifendate	05HQ	418.38	31.1	0.67	*Astragali Radix*
MOL000392	Formononetin	06HQ	268.28	69.67	0.21	*Astragali Radix*
MOL000398	Isoflavanone	07HQ	224.25	109.99	0.3	*Astragali Radix*
MOL000417	Calycosin	08HQ	284.28	47.75	0.24	*Astragali Radix*
MOL000422	Kaempferol	09HQ	286.25	41.88	0.24	*Astragali Radix*
MOL000098	Quercetin	10HQ	302.25	46.43	0.28	*Astragali Radix*
MOL002670	Cavidine	01BX	353.45	35.64	0.81	*Pinelliae Rhizoma*
MOL002714	Baicalein	02BX	270.25	33.52	0.21	*Pinelliae Rhizoma*
MOL002776	Baicalin	03BX	446.39	40.12	0.75	*Pinelliae Rhizoma*
MOL000358	beta-Sitosterol	04BX	414.79	36.91	0.75	*Pinelliae Rhizoma*
MOL000449	Stigmasterol	05BX	412.77	43.83	0.76	*Pinelliae Rhizoma*
MOL005030	11-Eicosenoic acid	06BX	310.58	30.7	0.2	*Pinelliae Rhizoma*
MOL000519	(+)-Neocryptotanshinone	07BX	314.41	31.11	0.32	*Pinelliae Rhizoma*
MOL003578	Cyclo(L-tyrosyl-L-phenylalanyl)	08BX	310.3	38.69	0.78	*Pinelliae Rhizoma*
MOL000359	beta-Sitosterol	01CP	414.79	36.91	0.75	*Citri Reticulatae Pericarpium*
MOL004328	Naringenin	02CP	272.27	59.29	0.21	*Citri Reticulatae Pericarpium*
MOL005828	Nobiletin	03CP	402.43	61.67	0.52	*Citri Reticulatae Pericarpium*
MOL001659	Poriferasterol	01BHSSC	412.77	43.83	0.76	*Herba Hedyotidis*
MOL000449	Stigmasterol	02BHSSC	412.77	43.83	0.76	*Herba Hedyotidis*
MOL000358	beta-Sitosterol	03BHSSC	414.79	36.91	0.75	*Herba Hedyotidis*
MOL000098	Quercetin	04BHSSC	302.25	46.43	0.28	*Herba Hedyotidis*
MOL010828	Cynaropicrin	01MX	346.41	67.5	0.38	*Aucklandiae Radix*
MOL010839	Lappadilactone	02MX	494.68	38.56	0.73	*Aucklandiae Radix*
MOL000211	Betulinic acid	03MX	456.78	55.38	0.78	*Aucklandiae Radix*
MOL000359	beta-Sitosterol	04MX	414.79	36.91	0.75	*Aucklandiae Radix*
MOL000449	Stigmasterol	05MX	412.77	43.83	0.76	*Aucklandiae Radix*
MOL001973	beta-Sitosterol acetate	01SR	456.83	40.39	0.85	*Amomi Fructus*
MOL000358	beta-Sitosterol	02SR	414.79	36.91	0.75	*Amomi Fructus*
MOL000449	Stigmasterol	03SR	412.77	43.83	0.76	*Amomi Fructus*
MOL007180	Glimepiride	04SR	490.69	32.29	0.7	*Amomi Fructus*
MOL007514	11,14-Eicosadienoic acid, methyl ester	05SR	322.59	39.67	0.23	*Amomi Fructus*
MOL010846	Deoxynivalenol	01MY	296.35	31.16	0.25	*Hordei Fructus Germinatus*
MOL005088	Nivalenol	02MY	312.35	35.68	0.28	*Hordei Fructus Germinatus*
MOL002322	Isovitexin	03MY	432.41	31.29	0.72	*Hordei Fructus Germinatus*
MOL000358	beta-Sitosterol	04MY	414.79	36.91	0.75	*Hordei Fructus Germinatus*
MOL004798	Delphinidin	05MY	338.69	40.63	0.28	*Hordei Fructus Germinatus*
MOL000492	Cianidanol	06MY	290.29	54.83	0.24	*Hordei Fructus Germinatus*
MOL000569	Digallic acid	07MY	322.24	61.85	0.26	*Hordei Fructus Germinatus*
MOL000006	Luteolin	08MY	286.25	36.16	0.25	*Hordei Fructus Germinatus*
MOL007180	Glimepiride	09MY	490.69	32.29	0.7	*Hordei Fructus Germinatus*
MOL000073	(+)-Epicatechin	10MY	290.29	48.96	0.24	*Hordei Fructus Germinatus*
MOL013381	Marmin	01ZQ	332.43	38.23	0.31	*Aurantii Fructus*
MOL002341	Hesperetin	02ZQ	302.3	70.31	0.27	*Aurantii Fructus*
MOL000358	beta-Sitosterol	03ZQ	414.79	36.91	0.75	*Aurantii Fructus*
MOL004328	Naringenin	04ZQ	272.27	59.29	0.21	*Aurantii Fructus*
MOL005828	Nobiletin	05ZQ	402.43	61.67	0.52	*Aurantii Fructus*
MOL000358	beta-Sitosterol	01TLG	414.79	36.91	0.75	*Radix Actinidiae Chinensis*
MOL000359	beta-Sitosterol	02TLG	414.79	36.91	0.75	*Radix Actinidiae Chinensis*
MOL000471	Aloe-emodin	03TLG	270.25	83.38	0.24	*Radix Actinidiae Chinensis*
MOL000492	Cianidanol	04TLG	290.29	54.83	0.24	*Radix Actinidiae Chinensis*
MOL000073	(+)-Epicatechin	05TLG	290.29	48.96	0.24	*Radix Actinidiae Chinensis*
MOL000098	Quercetin	06TLG	302.25	46.43	0.28	*Radix Actinidiae Chinensis*
MOL001918	paeoniflorgenone	01BS	318.35	87.59	0.37	*Paeoniae Radix Alba*
MOL001919	Palbinone	02BS	358.52	43.56	0.53	*Paeoniae Radix Alba*
MOL001921	(+)-Lactiflorin	03BS	462.49	49.12	0.8	*Paeoniae Radix Alba*
MOL001924	Paeoniflorin	04BS	480.51	53.87	0.79	*Paeoniae Radix Alba*
MOL001928	Albiflorin	05BS	480.5	66.64	0.33	*Paeoniae Radix Alba*
MOL001930	Benzoyloxypaeoniflorin	06BS	600.6	31.27	0.75	*Paeoniae Radix Alba*
MOL000211	Betulinic acid	07BS	456.78	55.38	0.78	*Paeoniae Radix Alba*
MOL000359	beta-Sitosterol	08BS	414.79	36.91	0.75	*Paeoniae Radix Alba*
MOL000422	Kaempferol	09BS	286.25	41.88	0.24	*Paeoniae Radix Alba*
MOL000492	Cianidanol	10BS	290.29	54.83	0.24	*Paeoniae Radix Alba*
MOL002776	Baicalin	01BZL	446.39	40.12	0.75	*Scutellariae Barbatae Herba*
MOL005043	Campesterol	02BZL	400.76	37.58	0.71	*Scutellariae Barbatae Herba*
MOL000953	Cholesterol	03BZL	386.73	37.87	0.68	*Scutellariae Barbatae Herba*
MOL000358	beta-Sitosterol	04BZL	414.79	36.91	0.75	*Scutellariae Barbatae Herba*
MOL012266	Rivularin (flavone)	05BZL	344.34	37.94	0.37	*Scutellariae Barbatae Herba*
MOL001973	beta-Sitosterol acetate	06BZL	456.83	40.39	0.85	*Scutellariae Barbatae Herba*
MOL000449	Stigmasterol	07BZL	412.77	43.83	0.76	*Scutellariae Barbatae Herba*
MOL000173	Wogonin	08BZL	284.28	30.68	0.23	*Scutellariae Barbatae Herba*
MOL001735	Hispidulin	09BZL	300.28	30.97	0.27	*Scutellariae Barbatae Herba*
MOL001755	Stigmast-4-en-3-one	10BZL	412.77	36.08	0.76	*Scutellariae Barbatae Herba*
MOL002714	Baicalein	11BZL	270.25	33.52	0.21	*Scutellariae Barbatae Herba*
MOL002719	Carthamidin	12BZL	288.27	33.23	0.24	*Scutellariae Barbatae Herba*
MOL002915	Salvigenin	13BZL	328.34	49.07	0.33	*Scutellariae Barbatae Herba*
MOL000351	Rhamnazin	14BZL	330.31	47.14	0.34	*Scutellariae Barbatae Herba*
MOL000359	beta-Sitosterol	15BZL	414.79	36.91	0.75	*Scutellariae Barbatae Herba*
MOL005190	Eriodictyol	16BZL	288.27	71.79	0.24	*Scutellariae Barbatae Herba*
MOL000006	Luteolin	17BZL	286.25	36.16	0.25	*Scutellariae Barbatae Herba*
MOL008206	5-Hydroxy-7,8-dimethoxyflavone	18BZL	298.31	44.09	0.25	*Scutellariae Barbatae Herba*
MOL000098	Quercetin	19BZL	302.25	46.43	0.28	*Scutellariae Barbatae Herba*
MOL001601	Trijuganone B	01DANS	280.34	38.75	0.36	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL001659	Poriferasterol	02DANS	412.77	43.83	0.76	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL001771	Clionasterol	03DANS	414.79	36.91	0.75	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL001942	Isoimperatorin	04DANS	270.3	45.46	0.23	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL002222	Sugiol	05DANS	300.48	36.11	0.28	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL002651	Dehydrotanshinone II A	06DANS	292.35	43.76	0.4	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL002776	Baicalin	07DANS	446.39	40.12	0.75	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL000569	Digallic acid	08DANS	322.24	61.85	0.26	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL000006	Luteolin	09DANS	286.25	36.16	0.25	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL006824	α-amyrin	10DANS	426.8	39.51	0.76	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007036	Arucadiol	11DANS	298.41	33.77	0.29	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007048	Tournefolic acid A	12DANS	312.29	48.24	0.31	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007061	Methylenetanshinquinone	13DANS	278.32	37.07	0.36	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007064	Przewalskin B	14DANS	330.46	110.32	0.44	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007068	Przewaquinone B	15DANS	292.3	62.24	0.41	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007069	Tanshinol B	16DANS	296.34	55.74	0.4	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007079	Tanshinaldehyde	17DANS	310.3	52.47	0.45	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007081	Danshenol B	18DANS	354.48	57.95	0.56	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007082	Danshenol A	19DANS	336.41	56.97	0.52	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007085	Salvilenone	20DANS	292.4	30.38	0.38	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007088	Cryptotanshinone	21DANS	296.39	52.34	0.4	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007098	Deoxyneocryptotanshinone	22DANS	298.41	49.4	0.29	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007101	1,2-Dihydrotanshinquinone	23DANS	278.32	45.04	0.36	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007105	Epidanshenspiroketallactone	24DANS	268.31	68.27	0.31	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007107	Ferruginol	25DANS	286.5	36.07	0.25	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007108	Isocryptotanshinone	26DANS	296.39	54.98	0.39	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007111	Isotanshinone IIA	27DANS	294.37	49.92	0.4	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007115	Manool	28DANS	290.5	45.04	0.2	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007118	Microstegiol	29DANS	298.46	39.61	0.28	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007121	Miltipolone	30DANS	300.43	36.56	0.37	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007122	Miltirone	31DANS	282.41	38.76	0.25	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007124	Deoxyneocryptotanshinone	32DANS	298.4	39.46	0.23	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007125	Neocryptotanshinone	33DANS	314.41	52.49	0.32	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007127	Nortanshinone	34DANS	280.29	34.72	0.37	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007141	Salvianolic acid G	35DANS	418.3	45.56	0.61	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007143	Salvilenone	36DANS	292.4	32.43	0.23	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007145	Salviolone	37DANS	268.38	31.72	0.24	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007149	Sugiol	38DANS	300.48	34.49	0.28	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007150	Tanshindiol A	39DANS	312.34	75.39	0.46	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007151	Tanshindiol B	40DANS	312.34	42.67	0.45	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007152	Tanshindiol C	41DANS	312.34	42.85	0.45	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007154	Tanshinone IIA	42DANS	294.37	49.89	0.4	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL007156	Danshenxinkun A	43DANS	296.34	45.64	0.3	*Salviae Miltiorrhizae Radix et Rhizoma*
MOL000358	beta-Sitosterol	01NZZ	414.79	36.91	0.75	*Ligustri Lucidi Fructus*
MOL000422	Kaempferol	02NZZ	286.25	41.88	0.24	*Ligustri Lucidi Fructus*
MOL004576	Taxifolin	03NZZ	304.27	57.84	0.27	*Ligustri Lucidi Fructus*
MOL005190	Eriodictyol	04NZZ	288.27	71.79	0.24	*Ligustri Lucidi Fructus*
MOL005209	Lucidusculine	05NZZ	401.6	30.11	0.75	*Ligustri Lucidi Fructus*
MOL000006	Luteolin	06NZZ	286.25	36.16	0.25	*Ligustri Lucidi Fructus*
MOL000098	Quercetin	07NZZ	302.25	46.43	0.28	*Ligustri Lucidi Fructus*
MOL000275	Trametenolic acid	01FL	456.78	38.71	0.8	*Poria*
MOL000276	Dehydropachymic acid	02FL	526.83	35.11	0.81	*Poria*
MOL000279	Cerevisterol	03FL	430.74	37.96	0.77	*Poria*
MOL000280	Dehydrotumulosic acid	04FL	484.79	31.07	0.82	*Poria*
MOL000282	Stellasterol	05FL	398.74	43.51	0.72	*Poria*
MOL000283	Ergosterol peroxide	06FL	428.6	40.36	0.81	*Poria*
MOL000285	Polyporenic acid C	07FL	482.77	38.26	0.82	*Poria*
MOL000287	Eburicoic acid	08FL	470.81	38.7	0.81	*Poria*
MOL000289	Pachymic acid	09FL	528.85	33.63	0.81	*Poria*
MOL000290	Poricoic acid A	10FL	498.77	30.61	0.76	*Poria*
MOL000291	Poricoic acid B	11FL	484.74	30.52	0.75	*Poria*
MOL000292	Poricoic acid C	12FL	482.77	38.15	0.75	*Poria*
MOL000296	Hederagenin	13FL	414.79	36.91	0.75	*Poria*
MOL000300	Dehydroeburicoic acid	14FL	468.7	44.17	0.83	*Poria*
MOL001006	Chondrillasterol	01DANGS	412.77	42.98	0.76	*Codonopsis Radix*
MOL002140	Perlolyrine	02DANGS	264.3	65.95	0.27	*Codonopsis Radix*
MOL002879	Diisooctyl phthalate	03DANGS	390.62	43.59	0.39	*Codonopsis Radix*
MOL000449	Stigmasterol	04DANGS	412.77	43.83	0.76	*Codonopsis Radix*
MOL003896	7-Methoxy-2-methyl-3-phenyl-4H-chromen-4-one	05DANGS	266.31	42.56	0.2	*Codonopsis Radix*
MOL004355	alpha-Spinasterol	06DANGS	412.77	42.98	0.76	*Codonopsis Radix*
MOL004492	Chrysanthemaxanthin	07DANGS	584.96	38.72	0.58	*Codonopsis Radix*
MOL005321	Frutinone A	08DANGS	264.24	65.9	0.34	*Codonopsis Radix*
MOL000006	Luteolin	09DANGS	286.25	36.16	0.25	*Codonopsis Radix*
MOL006554	Taraxerol	10DANGS	426.8	38.4	0.77	*Codonopsis Radix*
MOL008400	Glycitein	11DANGS	284.28	50.48	0.24	*Codonopsis Radix*
MOL008406	Spinoside A	12DANGS	716.95	39.97	0.4	*Codonopsis Radix*
MOL008407	Stigmasterone	13DANGS	410.75	45.4	0.76	*Codonopsis Radix*
MOL008411	11-Hydroxyrankinidine	14DANGS	356.46	40	0.66	*Codonopsis Radix*
MOL000211	Betulinic acid	01GC	456.78	55.38	0.78	*Glycyrrhizae Radix et Rhizoma*
MOL000359	beta-Sitosterol	02GC	414.79	36.91	0.75	*Glycyrrhizae Radix et Rhizoma*
MOL000422	Kaempferol	03GC	286.25	41.88	0.24	*Glycyrrhizae Radix et Rhizoma*
MOL004328	Naringenin	04GC	272.27	59.29	0.21	*Glycyrrhizae Radix et Rhizoma*
MOL004805	Shinflavanone	05GC	390.51	31.79	0.72	*Glycyrrhizae Radix et Rhizoma*
MOL004806	euchrenone	06GC	406.56	30.29	0.57	*Glycyrrhizae Radix et Rhizoma*
MOL004810	Glyasperin F	07GC	354.38	75.84	0.54	*Glycyrrhizae Radix et Rhizoma*
MOL004811	Glyasperin C	08GC	356.45	45.56	0.4	*Glycyrrhizae Radix et Rhizoma*
MOL004814	Isotrifoliol	09GC	298.26	31.94	0.42	*Glycyrrhizae Radix et Rhizoma*
MOL004815	Kanzonol B	10GC	322.38	39.62	0.35	*Glycyrrhizae Radix et Rhizoma*
MOL004820	Kanzonol W	11GC	336.36	50.48	0.52	*Glycyrrhizae Radix et Rhizoma*
MOL004827	Semilicoisoflavone B	12GC	352.36	48.78	0.55	*Glycyrrhizae Radix et Rhizoma*
MOL004828	Glepidotin A	13GC	338.38	44.72	0.35	*Glycyrrhizae Radix et Rhizoma*
MOL004829	Glepidotin B	14GC	340.4	64.46	0.34	*Glycyrrhizae Radix et Rhizoma*
MOL004833	Phaseolinisoflavan	15GC	324.4	32.01	0.45	*Glycyrrhizae Radix et Rhizoma*
MOL004838	Kanzonol U	16GC	308.35	58.44	0.38	*Glycyrrhizae Radix et Rhizoma*
MOL004841	Licochalcone B	17GC	286.3	76.76	0.19	*Glycyrrhizae Radix et Rhizoma*
MOL004849	Licoarylcoumarin	18GC	368.41	59.62	0.43	*Glycyrrhizae Radix et Rhizoma*
MOL004855	Licoricone	19GC	382.44	63.58	0.47	*Glycyrrhizae Radix et Rhizoma*
MOL004856	Gancaonin A	20GC	352.41	51.08	0.4	*Glycyrrhizae Radix et Rhizoma*
MOL004857	Gancaonin B	21GC	368.41	48.79	0.45	*Glycyrrhizae Radix et Rhizoma*
MOL004863	Gancaonin L	22GC	354.38	66.37	0.41	*Glycyrrhizae Radix et Rhizoma*
MOL004864	Gancaonin M	23GC	352.41	30.49	0.41	*Glycyrrhizae Radix et Rhizoma*
MOL004866	Gancaonin O	24GC	354.38	44.15	0.41	*Glycyrrhizae Radix et Rhizoma*
MOL004879	Glycyrin	25GC	382.44	52.61	0.47	*Glycyrrhizae Radix et Rhizoma*
MOL004882	Licocoumarone	26GC	340.4	33.21	0.36	*Glycyrrhizae Radix et Rhizoma*
MOL004883	Licoisoflavone A	27GC	354.38	41.61	0.42	*Glycyrrhizae Radix et Rhizoma*
MOL004884	Licoisoflavone B	28GC	352.36	38.93	0.55	*Glycyrrhizae Radix et Rhizoma*
MOL004885	Licoisoflavanone	29GC	354.38	52.47	0.54	*Glycyrrhizae Radix et Rhizoma*
MOL004891	Shinpterocarpin	30GC	322.38	80.3	0.73	*Glycyrrhizae Radix et Rhizoma*
MOL004903	Liquiritin	31GC	418.43	65.69	0.74	*Glycyrrhizae Radix et Rhizoma*
MOL004904	Licopyranocoumarin	32GC	384.41	80.36	0.65	*Glycyrrhizae Radix et Rhizoma*
MOL004907	Glyzaglabrin	33GC	298.26	61.07	0.35	*Glycyrrhizae Radix et Rhizoma*
MOL004908	Glabridin	34GC	324.4	53.25	0.47	*Glycyrrhizae Radix et Rhizoma*
MOL004910	Glabranin	35GC	324.4	52.9	0.31	*Glycyrrhizae Radix et Rhizoma*
MOL004911	Glabrene	36GC	322.38	46.27	0.44	*Glycyrrhizae Radix et Rhizoma*
MOL004912	Glabrone	37GC	336.36	52.51	0.5	*Glycyrrhizae Radix et Rhizoma*
MOL004915	Eurycarpin A	38GC	338.38	43.28	0.37	*Glycyrrhizae Radix et Rhizoma*
MOL004917	Glycyroside	39GC	562.57	37.25	0.79	*Glycyrrhizae Radix et Rhizoma*
MOL004924	(-)-Medicocarpin	40GC	432.46	40.99	0.95	*Glycyrrhizae Radix et Rhizoma*
MOL004935	Sigmoidin B	41GC	356.4	34.88	0.41	*Glycyrrhizae Radix et Rhizoma*
MOL004945	Isobavachin	42GC	324.4	36.57	0.32	*Glycyrrhizae Radix et Rhizoma*
MOL004948	Isoglycyrol	43GC	366.39	44.7	0.84	*Glycyrrhizae Radix et Rhizoma*
MOL004949	Isolicoflavonol	44GC	354.38	45.17	0.42	*Glycyrrhizae Radix et Rhizoma*
MOL004957	Isoformononetin	45GC	268.28	38.37	0.21	*Glycyrrhizae Radix et Rhizoma*
MOL004959	1-Methoxyphaseollidin	46GC	354.43	69.98	0.64	*Glycyrrhizae Radix et Rhizoma*
MOL000497	Licochalcone a	47GC	338.43	40.79	0.29	*Glycyrrhizae Radix et Rhizoma*
MOL004996	Gadelaidic acid	48GC	310.58	30.7	0.2	*Glycyrrhizae Radix et Rhizoma*
MOL000500	Vestitol	49GC	272.32	74.66	0.21	*Glycyrrhizae Radix et Rhizoma*
MOL005000	Gancaonin G	50GC	352.41	60.44	0.39	*Glycyrrhizae Radix et Rhizoma*
MOL005001	Gancaonin H	51GC	420.49	50.1	0.78	*Glycyrrhizae Radix et Rhizoma*
MOL005003	Licoagrocarpin	52GC	338.43	58.81	0.58	*Glycyrrhizae Radix et Rhizoma*
MOL005008	Glycyrrhiza flavonol A	53GC	370.38	41.28	0.6	*Glycyrrhizae Radix et Rhizoma*
MOL005012	Licoagroisoflavone	54GC	336.36	57.28	0.49	*Glycyrrhizae Radix et Rhizoma*
MOL005016	Odoratin	55GC	314.31	49.95	0.3	*Glycyrrhizae Radix et Rhizoma*
MOL005017	Phaseol	56GC	336.36	78.77	0.58	*Glycyrrhizae Radix et Rhizoma*
MOL005018	Xambioona	57GC	388.49	54.85	0.87	*Glycyrrhizae Radix et Rhizoma*
MOL005020	Dehydroglyasperin C	58GC	354.4	53.82	0.37	*Glycyrrhizae Radix et Rhizoma*
MOL000098	Quercetin	59GC	302.25	46.43	0.28	*Glycyrrhizae Radix et Rhizoma*
MOL001323	alpha1-Sitosterol	01YYR	426.8	43.28	0.78	*Coicis Semen*
MOL001494	Ethyl linoleate	02YYR	308.56	42	0.19	*Coicis Semen*
MOL000359	beta-Sitosterol	03YYR	414.79	36.91	0.75	*Coicis Semen*
MOL000449	Stigmasterol	04YYR	412.77	43.83	0.76	*Coicis Semen*
MOL008118	Coixenolide	05YYR	591.08	32.4	0.43	*Coicis Semen*
MOL008121	2-Monoolein	06YYR	356.61	34.23	0.29	*Coicis Semen*
MOL000953	Cholesterol	07YYR	386.73	37.87	0.68	*Coicis Semen*
MOL001506	Squalene	01TZS	410.8	33.55	0.42	*Pseudostellariae Radix*
MOL001689	Acacetin	02TZS	284.28	34.97	0.24	*Pseudostellariae Radix*
MOL001790	Linarin	03TZS	592.6	39.84	0.71	*Pseudostellariae Radix*
MOL000358	beta-Sitosterol	04TZS	414.79	36.91	0.75	*Pseudostellariae Radix*
MOL000006	Luteolin	05TZS	286.25	36.16	0.25	*Pseudostellariae Radix*
MOL006554	Taraxerol	06TZS	426.8	38.4	0.77	*Pseudostellariae Radix*
MOL006756	Schottenol	07TZS	414.79	37.42	0.75	*Pseudostellariae Radix*
MOL002464	1-Monolinolein	08TZS	354.59	37.18	0.3	*Pseudostellariae Radix*
MOL001559	Piperlonguminine	01SY	273.36	30.71	0.18	*Dioscoreae Rhizoma*
MOL001736	(-)-Taxifolin	02SY	304.27	60.51	0.27	*Dioscoreae Rhizoma*
MOL000310	Denudatin B	03SY	356.45	61.47	0.38	*Dioscoreae Rhizoma*
MOL000322	Kadsurenone	04SY	356.45	54.72	0.38	*Dioscoreae Rhizoma*
MOL005429	Hancinol	05SY	372.5	64.01	0.37	*Dioscoreae Rhizoma*
MOL005430	Hancinone C	06SY	400.51	59.05	0.39	*Dioscoreae Rhizoma*
MOL005043	Campesterol	07SY	400.76	37.58	0.71	*Dioscoreae Rhizoma*
MOL005440	Fucosterol	08SY	412.77	43.78	0.76	*Dioscoreae Rhizoma*
MOL000449	Stigmasterol	09SY	412.77	43.83	0.76	*Dioscoreae Rhizoma*
MOL000546	Diosgenin	10SY	414.69	80.88	0.81	*Dioscoreae Rhizoma*
MOL005461	Adonixanthin	11SY	582.9	38.16	0.54	*Dioscoreae Rhizoma*
MOL005465	Garcinone B	12SY	394.45	45.33	0.77	*Dioscoreae Rhizoma*
MOL000953	Cholesterol	13SY	386.73	37.87	0.68	*Dioscoreae Rhizoma*
MOL000358	beta-Sitosterol	01DG	414.79	36.91	0.75	*Angelicae Sinensis Radix*
MOL000449	Stigmasterol	02DG	412.77	43.83	0.76	*Angelicae Sinensis Radix*
MOL000296	Hederagenin	01EZ	414.79	36.91	0.75	*Curcumae Rhizoma*
MOL000906	Wenjine	02EZ	282.37	47.93	0.27	*Curcumae Rhizoma*
MOL000940	Bisdemethoxycurcumin	03EZ	308.35	77.38	0.26	*Curcumae Rhizoma*
MOL000098	Quercetin	01SZ	302.25	46.43	0.28	*Crataegi Fructus*

Abbr., abbreviation; MW, molecular weight; OB, oral bioavailability; DL, drug-likeness; BZ, bai zhu; HQ, huang qi; BX, ban xia; CP, chen pi; BHSSC, bai hua she she cao; MX, mu xiang; SR, sha ren; MY, mai ya; ZQ, zhi qiao; TLG, teng li gen; BS, bai shao; BZL, ban zhi lian; DANS, dan shen; NZZ, nvu zhen zi; FL, fu ling; DANGS, dang shen; GC, gan cao; YYR, yi yi ren; TZS, tai zi shen; SY, shan yao; DG, dang gui; EZ, e zhu.

 https://dx.doi.org/10.21037/atm-21-6301

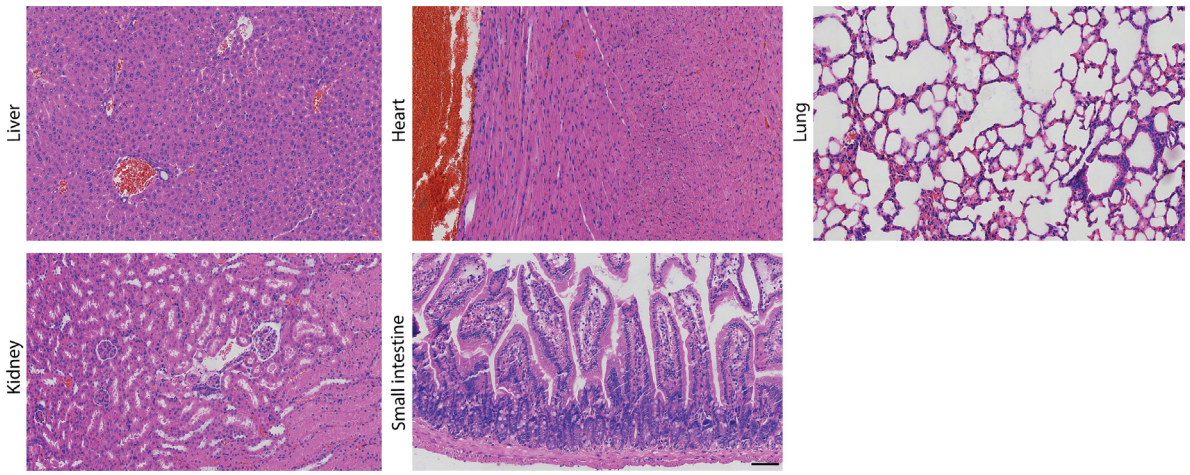


Figure S1 The preferred prescription had no toxic and side effect *in vivo*. Hematoxylin and eosin (H&E) staining of liver, heart, lung, kidney, and small intestine tissues of sacrificed nude mice (n = 5). The scale bar indicates 100 μm. H&E, hematoxylin and eosin.

Treatment principles	Herbs	Indications of GC
Tonify *Middle-Jiao Qi* & invigorate Spleen	*Atractylodis Macrocephalae Rhizoma*, *Astragali Radix*	Fatigue, poor appetite, loose stools or diarrhea, SOB, spontaneous sweating, prolapse of rectum, etc.
Disperse *Qi* stagnation & dry dampness	*Citri Reticulatae Pericarpium, Pinelliae Rhizoma, Aucklandiae Radix, Amomi Fructus, Aurantii Fructus*	Fullness and painful in the abdomen, belching, sour regurgitation, nausea, vomiting, constipation or diarrhoea, etc.
Abate food & promote digestion	*Galli Gigeriae Endothelium Corneum, Hordei Fructus Germinatus, Crataegi Fructus*	Epigastric and abdominal distension and fullness or pain, which are aggravated by food intake, belching with foul smell, anorexia, constipation, acidic regurgitation, nausea, vomiting, diarrhea with foul smell or fermented contents or constipation, etc.
Clear heat & remove toxin	*Radix Actinidiae Chinensis, Herba Hedyotidis*	Internal abscesses, jaundice, indigestion, vomiting, diarrhea, etc.

Figure S2 Working mode of the preferred prescription. Treatment principles of the preferred prescription include tonifying *Middle-Jiao* and invigorating the Spleen, dispersing *Qi* stagnation and drying dampness, abating food and promoting digestion, clearing heat, and removing toxins. SOB, short of breath.

https://dx.doi.org/10.21037/atm-21-6301