Acta Pharmaceutica Sinica B 2023;13(9):3802−3816

Chinese Pharmaceutical Association
Institute of Materia Medica, Chinese Academy of Medical Sciences

Acta Pharmaceutica Sinica B

www.elsevier.com/locate/apsb
www.sciencedirect.com

ORIGINAL ARTICLE

Functional annotation map of natural compounds in traditional Chinese medicines library: TCMs with myocardial protection as a case

Xudong Xing, Mengru Sun, Zifan Guo, Yongjuan Zhao, Yuru Cai, Ping Zhou, Huiying Wang, Wen Gao*, Ping Li*, Hua Yang*

State Key Laboratory of Natural Medicines, School of Traditional Chinese Pharmacy, China Pharmaceutical University, Nanjing 211198, China

Received 14 February 2023; received in revised form 14 May 2023; accepted 31 May 2023

KEY WORDS

Knowledge discovery;
Metabolomics;
High content screening;
Cell phenotype;
Ginseng;
Ginsenosides

Abstract The chemical complexity of traditional Chinese medicines (TCMs) makes the active and functional annotation of natural compounds challenging. Herein, we developed the TCMs-Compounds Functional Annotation platform (TCMs-CFA) for large-scale predicting active compounds with potential mechanisms from TCM complex system, without isolating and activity testing every single compound one by one. The platform was established based on the integration of TCMs knowledge base, chemome profiling, and high-content imaging. It mainly included: (1) selection of herbal drugs of target based on TCMs knowledge base; (2) chemome profiling of TCMs extract library by LC−MS; (3) cytological profiling of TCMs extract library by high-content cell-based imaging; (4) active compounds discovery by combining each mass signal and multi-parametric cell phenotypes; (5) construction of functional annotation map for predicting the potential mechanisms of lead compounds. In this stud TCMs with myocardial protection were applied as a case study, and validated for the feasibility and utility of the platform. Seven frequently used herbal drugs (*Ginseng*, etc.) were screened from 100,000 TCMs formulas for myocardial protection and subsequently prepared as a library of 700 extracts. By using TCMs-CFA platform, 81 lead compounds, including 10 novel bioactive ones, were quickly identified by correlating 8089 mass signals with 170,100 cytological parameters from an extract library. The TCMs-CFA platform described a new evidence-led tool for the rapid discovery process by data mining strategies, which is valuable for novel lead compounds from TCMs. All computations are done through Python and are publicly available on GitHub.

*Corresponding authors.
E-mail addresses: gw_cpu@126.com (Wen Gao), liping2004@126.com (Ping Li), 104yang104@163.com (Hua Yang).

Peer review under the responsibility of Chinese Pharmaceutical Association and Institute of Materia Medica, Chinese Academy of Medical Sciences.

https://doi.org/10.1016/j.apsb.2023.06.002

1. Introduction

Natural compounds and their derivatives have significantly contributed to clinical therapeutics[1,2], which counted for 33.6% of all US Food and Drug Administration (FDA)-approved small-molecule drugs between 1981 to 2019[3,4]. Compared with conventional synthetic small molecules, natural compounds confer advantages for drug discovery due to their high structural diversity and unique biological activities[5,6]. Particularly, traditional Chinese medicines (TCMs)-derived molecules were a rich resource and valuable treasure for drug discovery. TCMs theory and clinical practices for thousands of years could be used as prior knowledge to greatly increase the probability of discovering functional compounds. For instance, artemisinin from the plant *Artemisia annua* which has anti-malarial efficacy is a miracle of TCMs, known around the world. However, drug or lead compounds discovery from TCMs, a complex chemical system, is always been a challenge. To date, the well-known TCMs-sourced new drugs have been discovered through individual activity testing after isolation and purification, and/or based on the iterative activity-oriented fractionation. This strategy is very labor-intensive and time-consuming, leading to relatively low rates of discovering bioactive compounds[7].

Efficient chemical annotation of TCMs was an essential precondition for the bioactive natural compound discovery from a complex mixture. Metabolomics based on modern analytical techniques such as high performance liquid chromatography coupled with mass spectrometry (HPLC−MS) enabled the rapid detection and identification of known or unknown compounds in TCMs. Each compound could be labeled as some specific LC or mass signals. Advances in metabolomics have provided a better understanding of the constituents, but using LC−MS alone cannot be used for activity evaluation[8,9].

Active screening techniques, such as affinity chromatography[10,11], were thus often combined with MS for the discovery of lead compounds of specific efficacy from complex matrix[12]. These methods are simple and effective, but not systematically investigate the activities of all fractions and their characteristics. With the rapid development of modern biotechnology, cell mapping-based high-content imaging technology has been widely used in the field of drug research and biomedicine[13−17]. The cytological profiling allowed high-throughput and automated multicolor imaging of complex cellular events[18,19]. For example, researchers quantified 74 features defining filamentous actin (F-actin) and cellular morphology in >25 million cells after treatment with a library of 114,400 structurally diverse compounds, each defined by distinct quantitative features that could be machine learned[13]. Cellular perturbations by microenvironments or chemicals could bring about a lot of phenotypic changes[13,14]. Significantly, high-content imaging provides untargeted and unbiased multiple phenotypic parameters from an intact cell perspective, which might be suitable for in-depth understanding of the "multi-fractions, multi-targets" integration mechanism of TCMs. However, a potential disadvantage is a challenge of how metabolomic and phenotypic profiling technologies integrate to directly discover the hits from TCMs complex system. Kurita et al.[16] reported Compound Activity Mapping to connect phenotypic signatures to small molecules to predict novel bioactive natural products. Moreover, Lamb et al.[20] reported the Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease.

In this work, TCMs-Compounds Functional Annotation Platform (TCMs-CFA) was constructed, which combines a TCM knowledge-oriented strategy and comprehensive functional annotation of fractions in TCMs (Fig. 1). The TCMs-CFA allows the prediction of the identities and modes of action of biologically active compounds directly from TCMs, providing a mechanism for rational lead selection based on target compound. To evaluate this platform for natural compounds discovery we take the TCMs with myocardial protective activity as a case study. Firstly, 7 natural medicine were screened out as the most high-scored cardioprotective TCMs based on the knowledge base from 100,000 prescriptions. Then, 100 Fractions of each herb were prepared, and their untargeted chemomics and high-content imaging data of cardioprotection were integrated by Python, which to establish the high-content imaging data for annotation of activity compounds and prediction of mechanism of action. The 81 cardioprotective lead compounds were discovered from which we derived 170,100 cytological parameters and 8089 mass spectral features. Moreover, the target of lead compounds was predicted by machine learning (ML) and 12 target compounds in the library. Clustering the cytological profiles of ginsenosides (vina-ginsenoside R_4, etc.) with those of the 12 target compounds training set suggests that these compounds protect cardiomyocyte by protecting mitochondrial function. Thus, an efficient annotation map was constructed to predict the target of the screening functional compound from TCMs.

2. Materials and methods

2.1. Chemicals and reagents

Acetonitrile and formic acid (HPLC grade) were purchased from Merck (Darmstadt, Germany) and ROE (Newark, USA), respectively. Other organic reagents and chemicals were all analytical grades. Mito-Tracker® Deep Red™ (M22426), Hoechst 33342 (R37605), Alexa Fluor™ 488 Phalloidin (A12379), Mito-SOX™ Red Mitochondrial Superoxide Indicator (M36008), Image-iT™ LIVE Mitochondrial Transition Pore Assay Kit (I35103), Nonyl Acridine Orange (NAO, A1372), Mito-Probe™ TMRM Assay Kit for Flow Cytometry (M20036), Pierce™ Microplate BCA Protein Assay Kit (23225) and Pierce™ Fast Western Kit (35050) were purchased from Thermo Fisher Scientific (USA). Anti-GAPDH (ab8245); anti-VDAC1 (ab15895); anti-Drp1 (ab184247) were obtained from Abcam (Cambridge, UK). Anti-Cox IV (bs-10257R) were obtained from Bioss (China). Goat anti-mouse IgG (H + L) HRP (SA00001-1) and Goat anti-rabbit IgG (H + L) HRP (SA00001-2) were purchased from Proteintech (Manchester, UK).

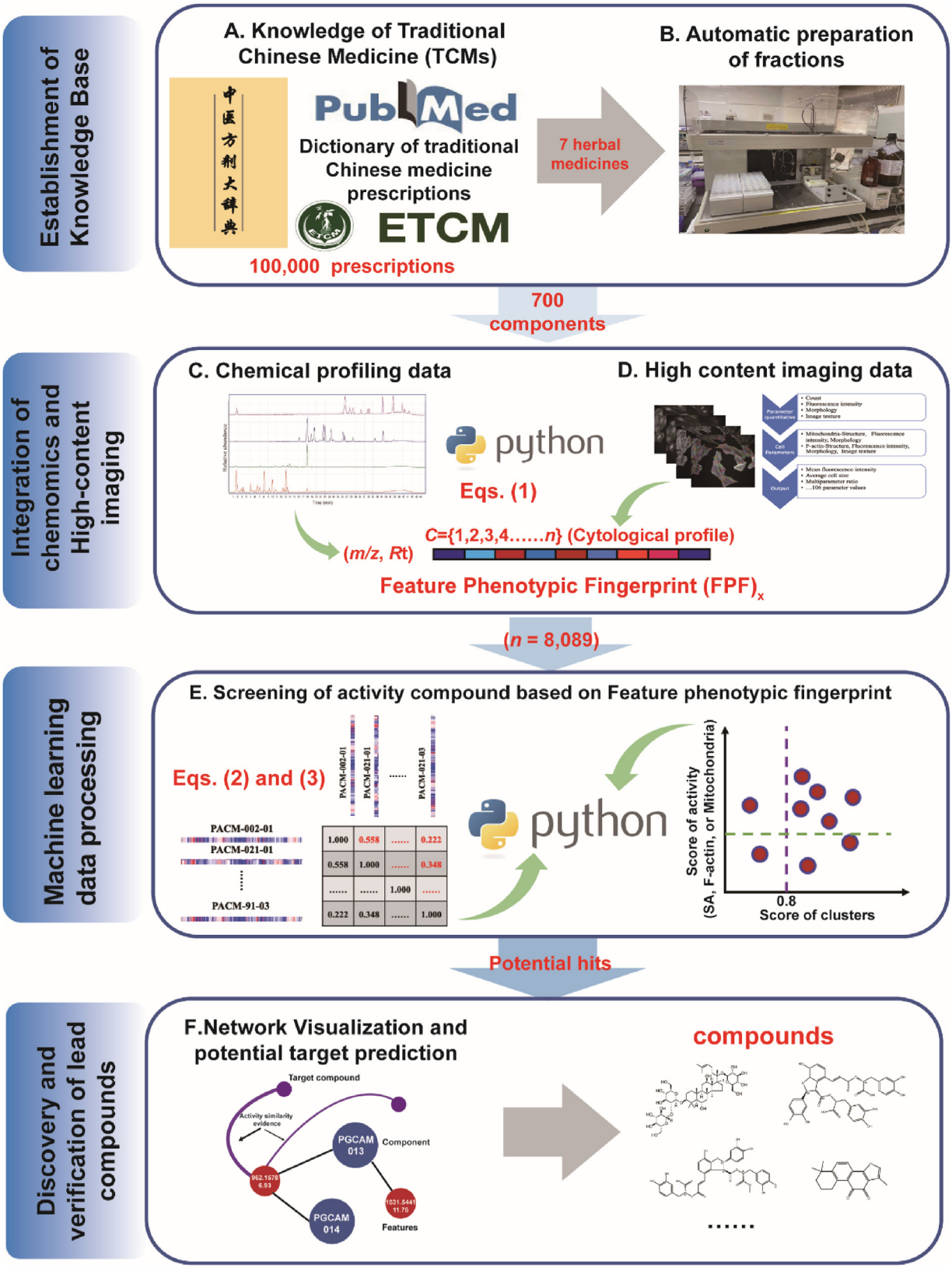

Figure 1 Workflow of traditional Chinese Medicines-Compounds Functional Annotation Platform (TCMs-CFA).

The target compounds used were derived from Med Chem Express (MCE) library, which information and chemical structure are given in Supporting Information Table S1. Reference standards, including ginsenoside Rf, vina-ginsenoside R_4, ginsenoside Rg_6, ginsenoside Rg_1, ginsenoside Rb_1, Danshensu, salvianolic acid E, salvianolic acid B, 9″-methyl lithospermate B,

and tanshinone IIA was obtained from Chengdu Munster Phytochemicals Ltd. (Chengdu, China), and the purity was determined using HPLC (> 98%).

2.2. TCMs knowledge analysis

There will also collect TCMs guided by TCMs knowledge. Data including formula name, dosage form and compositional formulation, TCMs indication and clinical application was collected from all formulae recorded in 'the dictionary of Chinese medicine prescription'. And according to the myocardial protection related Chinese herbal compound prescription or Chinese patent medicine was sorted into Excel for later screening (Supporting Information Table S2).

2.3. Fraction preparation of TCMs and chemical profiling

Each herbal medicine was extracted with 50% methanol/water (*v*/*v*), and the 2 mL of extracted solution was fractionated on an X-Bridge Prep C18 OBDTM column (19 mm × 250 mm, 5 μm, Waters) with an eluotropic series of water/methanol 10%−100% (*v*/*v*) wash in 95 min with 5 min for re-equilibration. The preparative-HPLC (waters Prep 150) was run on a Waters FractionLynx system with an Autosampler combined Automated Fraction Collector (Waters 2767). The flow rate was 5 mL/min. The automated collection platform was moved every minute by an automated robotic arm. Therefore, 100 extract samples were obtained from each herbal medicines.

These Fractions were dried by a personal solvent evaporator (GeneVac EZ-2 Plus, UK). Subsequently, the dried extracts were redissolved into 10 mg/mL stock solutions using DMSO. The DMSO stock solutions were 1:1000 for cytotoxicity assay or cell phenotype profile, and 1:20,000 in 50% methanol/water (*v*/*v*) for untargeted chemome profiling. Fractions resulting in reduced cell count, as well as toxicity, were submitted for continuous dilution and rescreening (Supporting Information Fig. S1).

Chemical profiling was carried out using an Agilent 1290 ultra-performance liquid chromatography coupled to an Agilent 6545 quadrupole time-of-flight mass spectrometer (UHPLC−QTOF-MS) with a Jetstream ESI source. Chromatographic separation was carried out on an Agilent ZORBAX Eclipse Plus C18 column (2.1 mm × 150 mm 1.8 μm) under 30 °C, and the mobile phase has consisted of 0.1% formic acid-water (A) and acetonitrile (B) at flow rate of 0.3 mL/min. The gradient procedure is as follows: 0−40 min, 10%−100% B. The acquisition parameters of TOF-MS were as follows: sheath gas temperature, 350 °C; drying gas temperature, 350 °C; sheath gas flow rate, 11.0 L/min; drying gas flow rate, 10.0 L/min; nebulizer gas pressure, 35 psi; capillary voltage, 3500 V; fragmentor voltage, 135 V. For MS/MS parameters, the collision energy was set at 15−45 eV. The sample was analyzed in both positive and negative ion mode and the mass was recorded from *m*/*z* 100−1500.

2.4. Cell culture and high-content imaging analysis of TCMs fraction

H9c2 cardiomyocytes cell lines were purchased from the National Collection of Authenticated Cell Cultures, Chinese Academy of Sciences (Shanghai, China). H9c2 cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM, Thermo Fisher Scientific, USA) containing 10% of Fetal Bovine Serum (FBS, Gibco, USA) and 1% of penicillin and streptomycin (1000 U/mL, Thermo, USA).

In the hypoxia model experiments, nitrogen (N_2) was flushed up to a partial oxygen pressure of about 1.0%. To simulate ischemia moreover, the cell culture medium for hypoxia was free of serum. By comparing the measured high-content imaging parameters in different modeling times, 2 h of hypoxia was finally selected. The molding time was optimized and 2 h of hypoxia was selected.

H9c2 cells were seeded at a density of 2000 cells/well into CellCarrier-96 ultra microplates (PerkinElmer, USA) and incubated at 37 °C for 24 h. When cells have reached the desired confluency and after the model, remove the media from the dish and add prewarmed (37 °C) staining solution containing MitoTracker® Deep RedTM and Hoechst 33342 (1:5000), incubation for 15 min. After cells were washed with PBS to remove the residual stain before immobilization with cold 4% paraformaldehyde. After fixing for 20 min, the cells were permeabilized with 0.1% Triton X-100 for 15 min and then incubated with a mixture of Alexa FluorTM 488 Phalloidin (1:1000) at room temperature for 15 min.

Cell phenotype profiles were collected on an high-content imaging system (Opera Phenix, PerkinElmer, USA) equipped with high-speed automatic spinning disk confocal optics, using a water immersion 40× water objective. A total of 81 fields of view were acquired for each well using three excitation sources, a Xenon UV lamp (80 ms exposure), a 200 mW 488 nm solid-state laser (120 ms exposure), and a 200 mW 561 nm solid-state laser (100 ms exposure) and. Each image was taken with a 12-bit high-sensitivity QE CCD camera applied to the images. Image data were then uploaded to both the Columbus data storage and Harmony 4.9 software analysis server (PerkinElmer, USA).

2.5. LC−MS method for chemical profiling

Chemical profiling was carried out using an Agilent 1290 ultra-performance liquid chromatography coupled to an Agilent 6545 quadrupole time-of-flight mass spectrometer (UHPLC−QTOF-MS) with a Jetstream ESI source. Chromatographic separation was carried out on an Agilent ZORBAX Eclipse Plus C18 column (2.1 mm × 150 mm 1.8 μm) under 30 °C, and the mobile phase has consisted of 0.1% formic acid/water (A) and acetonitrile (B) at flow rate of 0.3 mL/min. The gradient procedure is as follows: 0−40 min, 10%−100% B. The acquisition parameters of TOF-MS were as follows: sheath gas temperature, 350 °C; drying gas temperature, 350 °C; sheath gas flow rate, 11.0 L/min; drying gas flow rate, 10.0 L/min; nebulizer gas pressure, 35 psi; capillary voltage, 3500 V; fragmentor voltage, 135 V. For MS/MS parameters, the collision energy was set at 15−45 eV. The sample was analyzed in both positive and negative ion modes and the mass was recorded from *m*/*z* 100−1500.

2.6. Data integration of chemical and cytological data sets

The parameters of F-actin and mitochondrial phenotypes were analyzed by Harmony 4.9 software. Mover, the chemomics database by Molecular-Feature-Extraction (MFE) algorithm based on the Masshunter software (Agilent Technology). The data was converted into matrix mode and saved in Excel for subsequent data analysis. Integrating untargeted chemomics and high-content imaging data of cell phenotype to Feature Phenotypic Fingerprint (FPF) was described below.

2.6.1. High-content imaging data processing

The high-content image data were processed by Harmony 4.9 software (184 parameters, *i.e.*, cytological attribute values were extracted from each well) and then normalized and further analyzed using Python (Supporting Information). All of the 184 cytological attribute values of the model group (DMSO-treated) and the component-treated group were compared with the control group, and then conducted as a fingerprint, respectively. The value of each parameter in the fingerprint was between −1 and 1, which was shown as blue (negative) and red (positive), respectively.

2.6.2. Feature Phenotypic Fingerprint (FPF) processing

The Feature Phenotypic Fingerprint of a Feature is the average of each cytological attribute value for the set of Fractions in which the Feature is detected. This calculated "FPF" represents the predicted cytological profile for each Feature in the sample set.

$$C = \{1, 2, 3, \cdots i\}$$

$$F = \{f_1, f_2, f_3, \cdots f_i\}$$

$$f_{(1-i)} = \{a_1, a_2, a_3, \cdots a_n\}$$

$$FPF_x = \{\overline{a}_1, \overline{a}_2, \overline{a}_3, \cdots \overline{a}_n\} \quad (1)$$

where C was the Fraction set in which the Feature x was detected, numbered as 1 to i. F was the set of each Fraction's phenotype fingerprint f, which contained cytological attributes a_1 to a_n; Then the Feature phenotypic fingerprint of Feature x could be calculated using Eq. (1) which is the set containing the average of each cytological attribute value from Fraction 1 to i (Fig. 1).

2.6.3. Score of activity

From an FPF, the score of activity (SA) was defined as the sum of the square of each attribute to f (Eq. (2)). This defined value was used to evaluate the protective effect of each Feature on H9c2 cardiomyocytes cytoskeleton (SA of F-actin, SAF) or mitochondria (SA of mitochondria, SAM).

$$\mathrm{SA} = \sum_{x}^{i} a_x^2 \quad (2)$$

2.6.4. Score of clusters

The score of clusters (SC) of a Features is the mean value of Pearson correlation cube between all combinations of two different $f_{(1-i)}$ (Eq. (3)):

$$\mathrm{SC} = \frac{\sum_y^n \sum_z^n \mathrm{Paerson}\left(P_y, P_z\right)^3 - n}{n^2 - n} \quad (3)$$

The Pearson correlation values were cubed to reduce the importance of low values since the concentration of one or more extracts containing a given Feature, which is high enough to be detected by a UHPLC−QTOF MS (Fig. 1E). This value was used to assess each Feature associated with cytological data sets.

2.6.5. Screening of activity compound based on FPF

In the primary analysis selected SAF or SAM that met the threshold for modulating (>2.5 robust Z-scores) and score of clusters >0.8, which is a strong correlation[21].

The process was completed by Python, integrated with a program, and each step through the Python package (Supporting Information) output results and image analysis. Furthermore, the codes supporting this study are available on GitHub: https://github.com/CPU-XING/TCMs-High-content-data-and-chemicalome-integration-platform.git.

2.7. Network visualization and annotation of functional compounds

The heatmap summarizing hierarchical clustering of all phenotypic clusters and all Z-normalized quantitative features was generated using the 'Wald 'method function from the 'Pheatmap' package in the Python package. Agglomerative clustering proceeded *via* complete linkage as described for the default 'Wald' method' function.

In part, Excel and Gephi were used to create the networks, and the network data was analyzed and visualized (Fig. 1F). In the network, blue nodes represent Fractions and are connected to red nodes representing the Feature detected in those Fractions, green nodes represent the compound inferred by the second-order mass spectrum, and purple nodes represent the target compounds. Next, the coefficient was calculated by the K-means algorithm, and the visualized network was also used to quickly capture the relationship between Fractions, Features, inferential compounds, and target compounds. This section uses Gephi's built-in Fruchterman Reingold algorithm to distribute nodes with default parameters except the following: approximate repulsion of 0.4, the gravity of 5, scaling of 8, and setting the layout as "prevent overlap".

2.8. Primary neonatal rat ventricular myocytes (NRVMs) isolation and culture

NRVMs were isolated and cultured as described previously[22]. In short, NRVMs were isolated from 1 to 2 day-old Sprague−Dawley rats (B&K Universal Group Ltd., Shanghai, China). The heart tissue was cut into pieces and washed with precooled phosphate buffer saline (PBS) and then digested continuously with 0.1% type II collagenase (BS164) in a 37 °C water bath 5−7 times. After removing erythrocytes using red blood cell lysate, the harvested cells were pre-incubated with DMEM containing 10% FBS for 2 h. Cardiomyocytes were isolated and purified by discontinuous Percoll gradient (Cytiva, 17-0891-01), and then grown in DMEM supplemented with 100 mol/L 5-BrdU to inhibit fibroblast proliferation. The cells were used for experiments after 3−4 days of culture.

2.9. The assay of ROS, mitochondrial membrane potential ($\Delta\psi m$), mitochondrial number, ATP contents and calcium contents

For the observation of superoxide disputants of mitochondria, H9c2 cardiomyocytes were treated with 500 nmol/L Mito-SOX™ red mitochondrial superoxide indicator, for live-cell imaging working solution at 37 °C for 15 min, 50 nmol/L Mito-Tracker Green solution at 37 °C for 25 min and avoid light. Subsequently, images were acquired on a high-content imaging system, using a 63 × water objective to quantify fluorescence intensity.

For the detection of mitochondrial membrane potential ($\Delta\psi m$), H9c2 cells were incubated with 15 mol/L TMRM working solution dilution at 37 °C for 30 min in darkness. The data were

collected by Beckman cytometry (Beckman Coulter, Inc., USA) and by FlowJo software (version 10.8.1) analysis. Fluorescence intensity data were collected using the same method described above.

To quantify the number of mitochondria, H9c2 cells were treated with 0.1 mol/L nonyl acridine orange (NAO), synthesized at 37 °C, and cultured in darkness for 30 min. According to the manufacturer's instructions, cell fluorescence was analyzed by flow cytometry (CytoFLEX, Beckman Coulter, Inc., USA). And fluorescence intensity data were collected using the same method described above.

For the observation of mitochondria calcium ion change, the treated H9c2 cells were cultured with 1.0 mol/L calcein AM stock solution, 2 nmol/L Mito-Tracker Red CMXRos stock solution, 1.0 mol/L Hoechst 33342, and 1.0 mmol/L $CoCl_2$. The cells were washed twice in the modified HBSS buffer, the buffer was sucked out of the cells, and a sufficient labeling solution was applied to cover the cells attached to the covering. Incubate for 15 min at 37 °C, protected from light. Subsequently, images where fluorescence intensity data were collected using the same method described above.

The intracellular ATP contents were assayed with a commercial kit (Beyotime, S0027). According to the manufacturer's instructions, was using a microplate analyzer analysis.

2.10. Mitochondrial isolation and purification

Targeted analysis was performed to screen out the effects of Fractions on mitochondria. Therefore, mitochondria in H9c2 cells or NRVMs were isolated and purified by mitochondrial isolation and extraction kit (Beyotime, C3601). Pellet 4×10^8 cells by centrifuging harvested cell suspension in a 1.5 mL microcentrifuge tube at $600\times g$ for 10 min. Carefully remove and discard the supernatant. Add 1 mL of Mitochondria Isolation Reagent. The cells were then broken up using a cell-crushing apparatus. Centrifuge at $700\times g$ for 10 min at 4 °C. Transfer the supernatant to a new, 2.0 mL tube and centrifuge at $12{,}000\times g$ for 15 min at 4 °C. Carefully remove and discard the supernatant. 20 μL of mitochondrial lysate was added to each centrifuge tube. Subsequently, the protein concentration was determined by the Pierce™ Microplate BCA Protein Assay Kit.

2.11. Western blot analysis

The same amounts of protein were separated by SDS-PAGE and then transferred to the nitrocellulose (NC) membrane. After blocking with 5% non-fat milk at room temperature for 2 h, the membranes were immunoblotted with primary antibody (1:1000) overnight at 4 °C and then followed by incubation with the HRP-conjugated secondary antibody (1:10,000) at room temperature for 2 h. Protein bands were visualized using a Pierce™ Fast Western Kit and Image-J software was applied to quantify the band intensity values.

2.12. Interactions between small molecules and Drp1

Surface plasmon resonance (SPR) experiments were carried out at 25 °C by a Biacore T200 system (Cytiva). PBS-P buffer was filtered through a 0.45-μm membrane and degassed before use. After the surface of the CM5 sensor chip was prepared, Drp1 was immobilized on the surface according to the following procedure. Sodium acetate solution was used to prepare 30 μg/mL of Drp1 at a pH of 4.0.

Various hit compounds were diluted in PBS-P buffer according to a series of concentrations and flowed over immobilized Drp1 at a flow rate of 30 μL/min. The contact time and dissociation time were respectively set to, 60 and 150 s, respectively. A flow cell without a coupling protein was used as the blank reference. Blank solutions without analyte were adopted to correct system errors. The binding activity of small molecules was evaluated with the equilibrium dissociation constant (K_D). Experimental data were analyzed using a Biacore T200 instrument.

2.13. Statistical analysis

All data are expressed as mean ± standard deviation (SD) unless otherwise noted. Significance between the two groups was performed by Student's two-tailed *t*-test with GraphPad Prism 9.01 (Graph-Pad Software, La Jolla, CA, USA). In other cases, significance across more than two groups was done in Prism with one-way ANOVA. $P < 0.05$ was considered significant for all tests.

3. Results

3.1. Establishment of a knowledge base grid of TCMs for myocardial protection

To screen and select herbal drugs with potential myocardial protection, 'the dictionary of Chinese medicine prescription', which records 10,000 proprietaries of Chinese medicine including traditional Chinese prescription or Chinese patent medicines were used. A total of 472 prescriptions were filtered out based on the keywords of treating syndrome of heart blood stasis and obstruction, syndrome of Qi stagnation and blood stasis in the heart, and chest painful impediment, which were sorted by name, dosage form, and composition in Supporting Information Tables S2 and S3. Among them, 305 crude drugs including 258 medicinal plants (herbal medicines), 27 animal derived drugs, and 20 mineral drugs were involved (Supporting Information Table S2). According to the classification and statistics of TCMs properties, one-fourth of those medicinal materials (Supporting Information Fig. S2) were Qi-tonifying drugs (13.43%), and blood-activating and stasis-eliminating drugs (12.58%), which were often considered the best medicinal materials for the treatment of cardiovascular diseases[23,24].

The total appeared frequency of each crude drug in the prescription was analyzed. The top 21 with a frequency of more than 10% were defined as "highlighted medicines" (Table S4). Herb pairs was the smallest unit of compatibility of TCM compounds. It has the advantages of a simple composition structure, clear compatibility effect characteristics, and is suitable for scientific research[5]. Therefore, chord diagrams were used for pairing analysis (Fig. 2A). Among them, herbal drugs Renshen (Chinese name, 129 times), Danshen (Chinese name, 118 times), and Danggui (Chinese name, 106 times) ranked as the top 3 with a frequency of more than 100 times. Besides, a chord diagram (Fig. 2A) was generated to show the compatibility relationships between the 21 highlighted medicines in terms of frequency (column width) and herbal pairing (connections width) in the 472 prescriptions. The most frequent herbal pairs (co-occurrences) of

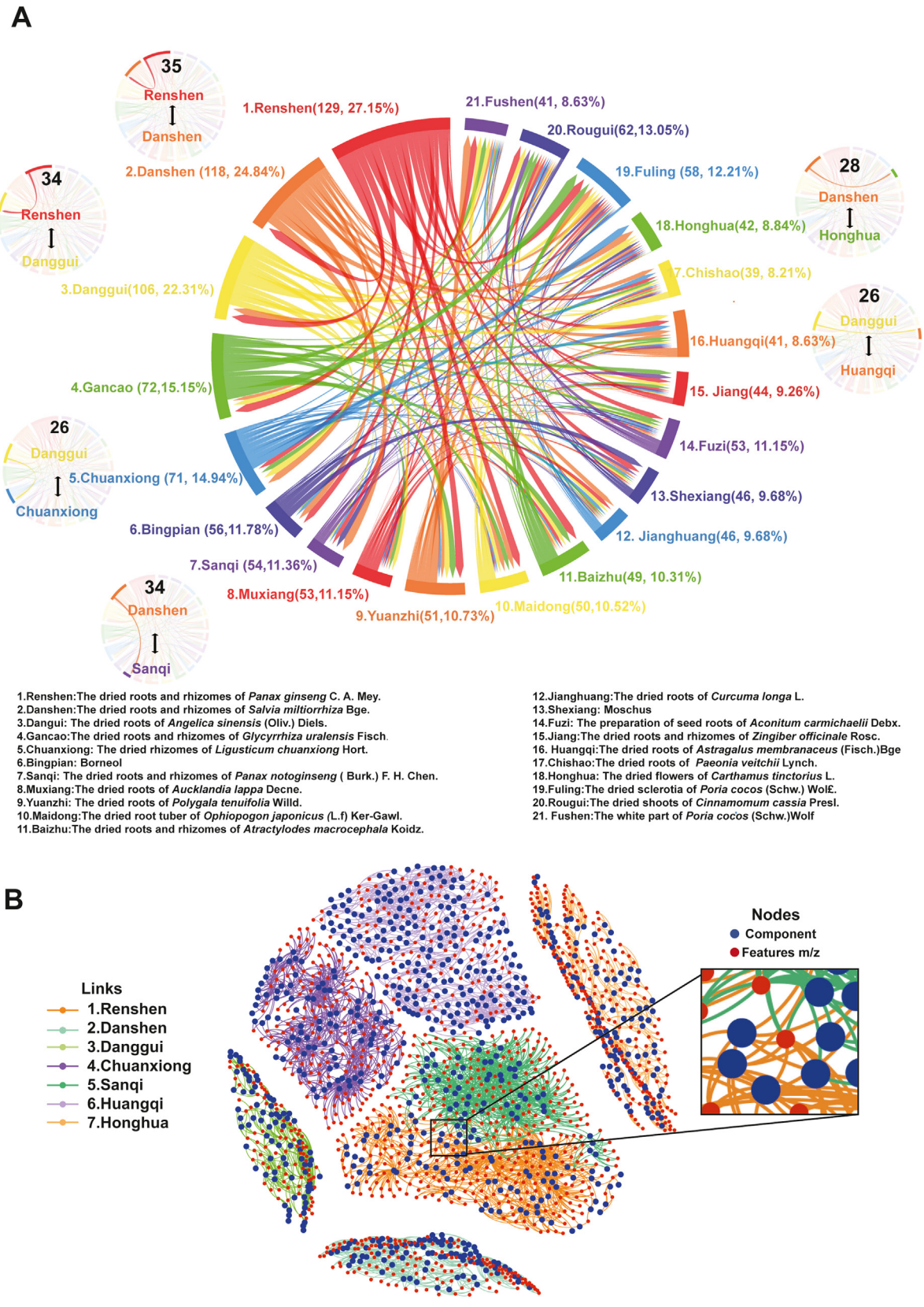

Figure 2 The cardioprotective TCMs and its chemomes. (A) The chord diagram of the 21 most high-frequency crude drugs' occurrence and coexisting in the prescription of cardioprotection recorded in the encyclopedia of TCMs (the middle circle), and the top 6 of appearing herb pairs (the surrounding small circles). (B) The "TCMs-Components-Features" graph of the selected 7 high-frequency herbal medicines; red nodes: Features (n = 8098); blue nodes: components (n = 700); colored links: the feature was detected in which component and its herb sources.

Renshen, Danshen, and Danggui were counted and selected, respectively. As an example, Renshen occurred 35 times with Danshen in myocardial protection, and the second herbal pair was Renshen and Danggui with 34 times (Fig. 2A, little chord diagram). Regarding Danshen, besides co-occurring with Renshen, the top 2 herbal pairs were Danshen-Sanqi (34 times) and Danshen-Honghua (28 times, Supporting Information Table S4), while the top 2 herbal pairs in terms of Danggui (besides Renshen-Dangui) were Danggui-Huangqi and Danggui-Chuanxiong (equally 26 times). Based on the results, 7 most high scored herbal drugs (Supporting Information Table S5), including Renshen [roots and rhizomes of *Panax ginseng*], Danshen [roots and rhizomes of *Salvia miltiorrhiza* Bge.), Danggui [roots of *Angelica sinensis* (Oliv.) Diels.], Sanqi [roots and rhizome of *Panax notoginseng* (Burk.) F. H. Chen.], Chuanxiong [rhizomes of *Ligusticum chuanxiong* Hort], Huangqi [roots of *Astragalus membranaceus* (Fisch.) Bge.], Honghua [Flowers of *Carthamus tinctorius* L.) (Table S5), were selected as candidates for development and validation of the TCMs-CFA platform.

3.2. *TCMs extract library and chemical profiling*

Each selected herb were extracted and prepared into 100 fractions, and a total of 700-member TCMs extract library was constructed. These fractions were subsequently subjected to UHPLC−QTOF MS for chemome profiling based on the untargeted metabolomics approaches. The molecular features detected in at least two Fractions were retained and then integrated into a database. The TCMs extract library are referred to as Feature lists, which were characterized by their accurate mass combined with their abundance and retention time (R_t), as well as the originated herbal fraction. Because the positive ion and negative ion data were two independent sets, their corresponding networks are constructed, respectively. To calculate the cluster value and the correlation between Fractions, the Feature ($n = 8089$) appearing from at least two Fractions was utilized, including 1240 from Renshen, 1145 from Danshen, 1046 from Danggui, 1226 from Sanqi, 1051 from Chuanxiong, 990 from Huangqi, and 1391 from Honghua. The 700 Fractions are broken up into different 8089 Features. Features (Fig. 2B). In the network diagram (right enlarged drawing in Fig. 2B), the red nodes represent the Features, which are linked to the Fractions (blue dots) in which the features were detected. And different colored edges represent different herb sources. This graph could help to understand the Features' information intuitively from Fractions to herbs. As shown in Fig. 2B, since Renshen and Sanqi have many identical saponins, there was some common point between the two herbs in the network diagram, resulting a close distance without obvious boundary. Simultaneously, clear boundaries between other medicinal materials in the network distinguish the compounds and Features from the different herbs.

3.3. *Multiparametric high content imaging for cytological profiling of TCMs extract library*

Cardiovascular disease was usually caused by acute and persistent coronary artery occlusion, and leads to cardiomyocyte ischemia/hypoxia injury. It has been reported that long-term hypoxia can induce various harmful effects in these cells through various signal cascades, including cell hypertrophy, apoptosis, and necrosis[24]. As cardiomyocytes were terminally differentiated cells that lack regenerative capabilities, preventing cardiomyocyte loss was an important challenge. Therefore, the H9c2 cardiomyocyte was used as a model under hypoxia. Hypoxic-induced myocardial apoptosis was a complex process, which includes disruption of ionic homeostasis. To characterize and annotate the Features from the selected 7 herbs on myocardial protection effects based on cell phenotypic screening, OGD-exposed H9c2 cells were applied for the compound activity mapping development. Image segmentation could extract and quantitative 100 morphological parameters (Supporting Information Table S6) according to cell morphology, texture, and properties of F-actin (Alexa Fluor™ 488 Phalloidin), Mitochondria (Mito-Tracker® Deep red), and nucleus (Hoechst 33342). First, H9c2 cells were exposed to OGD for 0.5−5.0 h to evaluate significant morphological changes (Fig. S3). Among 56 F-actin parameters, 17 such as ratio width to length and axial length decrease with the increase of OGD time (Supporting Information Figs. S2A and S3), indicated that OGD could significantly lead to the contraction of H9c2 cytoskeleton, and eventually lead to cell atrophy and apoptosis. With the increase of OGD time, the cytoskeleton showed shrinkage, and corresponding morphological parameters (such as Alexa 488 Ser 0 Px) and cytoskeleton texture parameters (such as Spot Alexa 488 Gabor Max 2 Px W2 and Spot Alexa 488 Gabor Min 2 Px W2) showed a decreasing trend (Fig. S3). In addition, the values of 9 cell symmetry parameters increased significantly with the extension of OGD time, indicating the cells lost the original spindle structure and became round (Fig. S3) While for 32 mitochondrial parameters and cytoskeletal parameters (Supporting Information Figs. S4 and S5), the morphological parameters were more in line with the requirements of mitochondrial kinetic energy for mitochondrial morphology analysis. The number of mitochondria increased with OGD time, indicating mitochondrial cleavage under hypoxia. All the parameters with morphological changes were statistically analyzed, and 80% of the parameters had significant changes (*vs.* control). Most of the cytological attribute values remained stable at 2 h ($P < 0.05$). Further, the robustness of the imaging assay was assessed with the Z′ factor values[25], which showed that 81 cell phenotypic parameters met the requirements (Z' factor $\geq$ 0.5, Supporting Information Fig. S6) in 2 h of hypoxia. In addition, we analyzed the correlation between morphological parameters and cell viability. The results showed that these parameters were correlated with the time of OGD and cell viability (Supporting Information Fig. S7).

Thus, the 81 morphological parameters (35 F-actin and 46 mitochondria) of H9c2 cells after 2 h induction of oxidative glucose were selected. Subsequently, cytological analysis was carried out to screen the active compounds of 7 herbs on H9c2 cells. In total, image data sets were collected. The data results ($81 \times 700 \times 3 = 170{,}100$) were output to Excel sheets in batches for later machine learning analysis.

3.4. *Screening of activity compound by correlating the chemome and cytological database*

In order to correlate chemome and cytological database, the Feature Phenotypic Fingerprint (FPF) was calculated for the correlation of the chemome and cytological database. Firstly, we normalized the high-content imaging's cytological profiling data of 700 Fractions. FPF of each Feature was the set containing the average of each cytological attribute value corresponding to Fractions in which the Feature was detected. For example, the Feature = 1007.5425 @6.938 was detected in 8 Fractions (Renshen-18-01, Renshen-18-02, Renshen-19-01, Renshen-19-02,

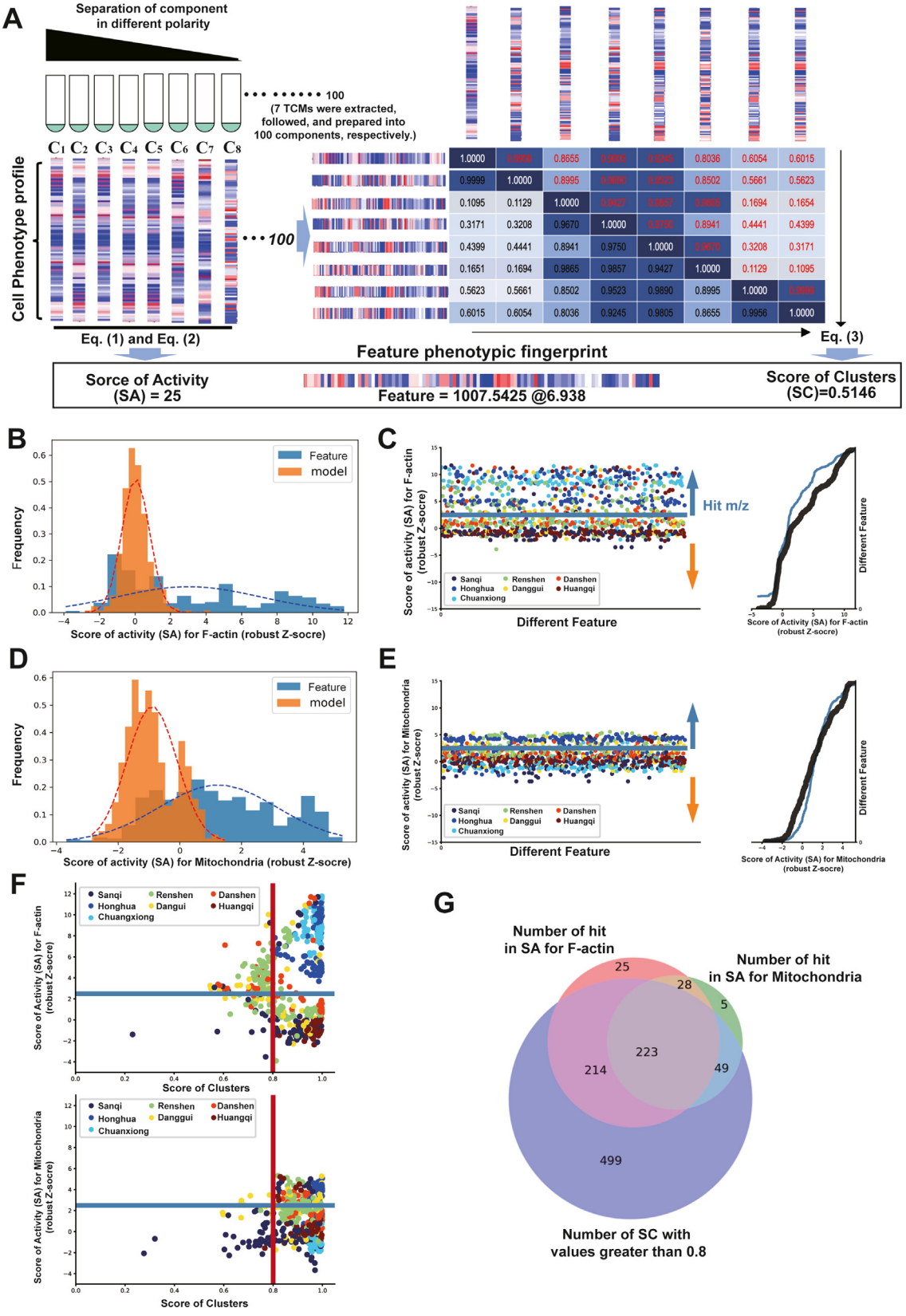

Figure 3 Integrating chemomics and cell phenotype profile database for activity compounds screening. (A) Schematic diagram of feature phenotypic fingerprint (FPF), a score of activity (SA), and the score of clusters (SC) calculation, using the feature 1007.5425@6.938 as an example. The 81-cell parameters with values between −1 and 1 of each component were displayed in blue (negative perturbation) or red (positive perturbation), respectively. (B) Relative frequency distributions of average robust Z-scores for SA of F-action. (C) Average robust Z-scores (left) and the cumulative distributions (right) for SA of F-action. (D) Relative frequency distributions of average robust Z-scores for SA of

Renshen-20-01, Renshen-20-02, Renshen-21-01, Renshen-21-02). Then, the FPF of Feature = 1007.5425 @6.938 was generated by the average 81 × 8 morphological parameters according to Eq. (1) in the 2.6 section. The score of activity of Feature = 1007.5425 @6.938 could be further calculated based on Eq. (2) in section 2.6 (Fig. 3A, left). Since the 81 morphological parameters include 35 related to F-actin and 46 to mitochondria, the score of activity in terms of F-actin (SAF) or mitochondria (SAM) were generated, respectively. Meanwhile, the score of clusters was calculated based on the Pearson correlations of the cytological profiling, according to Eq. (3) (Fig. 3A, right). Similarly, the scores of clusters of each Feature could be differentiated for F-actin and mitochondria respectively, which may be helpful in screening the refining mechanism of active compounds. Finally, 1093 Features' scores of clusters and SA were obtained for the integration of chemome profiling and cytological profiling database. Then the active compounds were screened based on the score of activity and scores of clusters. Firstly, the score of activity (SAF or SAM) and SC parameter values of 1093 features from 7 herbs were analyzed by the cumulative distribution, respectively. In the analysis, we selected 490 putative hits (44.95%) of SAF and 305 putative hits (27.98%) of SAM that met the threshold for modulating (>2.5 robust Z-scores) (Fig. 3B−E). Those selected Features were fitted with Gaussian distribution which is shown in the frequency histograms (Fig. 3B and D). Meanwhile, 985 features with scores of clusters >0.8, a strong correlation[21], were screened out (Fig. 3F). A three-way Venn diagram was applied for the features that meet the requirement of SAF, SAM, and SC (Fig. 3G). Of the confirmed hits, 437 (214 + 223, scores of clusters > 0.8) potential hits (40.09%) were correlated with the parameters of F-actin, and 272 (223 + 49, scores of clusters > 0.8) potential hits (24.95%) were correlated with the parameters of mitochondrial (Fig. 3G).

3.5. *Functional annotation map of functional feature (compounds)*

The network map of "Fraction-Feature (Compounds)-Prediction target" in the primary setting up (Supporting Information Fig. S8) could be simplified according to the screened potential hits according to F-actin and mitochondria. To discover the activity compound associated with the exact mechanism, the FPF of 12 compounds with known targets, derived from the MedChemExpress (MCE) library (Table S1) was used as a reference. Then the FPF of presumed compounds was clustered based on Ward's linkage criteria (Supporting Information Figs. S9 and S10) to obtain the Feature (or the presumed compound) whose cytological profile was close to the reference. The coefficient was calculated by the K-means algorithm which was visualized as a network. In Fig. 4, blue nodes represent Fractions (387 Fractions), and red nodes represent the Features (486 potential hits) detected in those Fractions. The 12 purple nodes represent the reference compounds (Table S1), and the thickness of the line in the network represents the strength of the relationship between two nodes. In Fig. 4, the Feature = 717.1458 @12.106, identified as salvianolic acid B (Table S7), was connected to stachydrine hydrochloride (NF-κB signaling pathway inhibitor) and asperosaponin VI (decreased the expression of active caspase-3) by the thick lines (Fig. 4, Supporting Information Fig. S9A and B). The predicted result was consistent with previously reported that salvianolic acid B inhibited myocardial apoptosis through the NF-κB[26] and Caspase-3 pathways[27]. For Feature = 1007.5427 @6.938, identified as vina-ginsenoside R_4 by comparing with the commercial reference standard, showed a strong correlation with Mdivi-1, a Drp1 inhibitor for mitochondrial autophagy, mitochondrial synthesis, and ATP production. Moreover, according to the clustering results, ginsenoside Rf (Feature = 799.4861 @22.954) may remove superoxide disputants and affect mitochondrial membrane potential because it close to Mito-TEMPO. ginsenoside Rg_6 (Feature = 765.4725 @12.453) may act on voltage-dependent anion channel-1 (like AKOS-22-VDAC1 inhibitor) and affects mitochondrial calcium ions (Fig. 4, Ru360- Inhibits Ca^{2+} uptake into mitochondria). For Feature = 845.4931 @9.698, identified as ginsenoside Rg1 by comparing with the commercial reference standard, showed a strong correlation with BMS-191095 and resveratrol (Fig. 4). Previous studies have shown that Ginsenoside Rg1 protect cardiomyocytes by maintaining ATP and other pathways[28]. In summary, the potential mechanism of the 81 lead compounds were predicted, and the top 5 ordered compounds of SAF (Danshensu, salvianolic acid E, salvianolic acid B, 9″-methyl lithospermate B, and tanshinone IIA) or SAM (ginsenoside Rf, vina-ginsenoside R_4, ginsenoside Rg_6, ginsenoside Rg_1, ginsenoside Rb_1) rankings, respectively, were both selected for the repeated high-content imaging Pharmacodynamic experiments. The results were shown in Supporting Information Figs. S11 and S12. The top 5 compounds of regulating skeleton were all derived from Danshen (Fig. S11), and the five saponins candidate lead compounds had a significant protective effect on hypoxic H9c2 cells, in which 4 were derived from Renshen and 1 was from Sanqi (Fig. 2A). The results were also consistent with previous reports that ginsenosides, the main Fractions of Renshen and Sanqi, have been extensively studied to play an important role in cardiovascular diseases[29,30]. Furthermore, there's a lot of evidence indicating that mitochondrial dysfunction has a distinctly important role in the pathogenesis of multiple cardiovascular disorders[31]. Based on the results of ML and the existing mechanism research, we preliminarily studied the predictive mechanism of 3 kinds of ginsenosides. Subsequently, we combined the screening results with the prediction mechanism of ginsenoside Rf, vina-ginsenoside R_4 and ginsenoside Rg_6 which was further verified by immunofluorescence, flow cytometry, and Western blotting.

3.6. *Verification of TCM-CFA prediction*

In order to verify the validity of the results, experimental verification was carried out for the new prediction, such as ginsenoside Rf, vina-ginsenoside R_4 and ginsenoside Rg_6 were verified. The production of adenosine triphosphate (ATP) was an important function of mitochondria, together with the regulation of apoptosis. Meanwhile, cluster analysis and K-means clustering showed that vina-ginsenoside R_4 was related to mitochondrial ATP production (Fig. 4 and Supporting Information Fig. S10D). The primary functions of the mitochondria are to produce the

mitochondria. (E) Average robust Z-scores (left) and the cumulative distributions (right) for SA of mitochondria. (F) A score of activity and Score of cluster distribution of Feature. (G) The three-way Venn diagram of the number of screened Features according to SA (of F-action and mitochondria, respectively) and SC.

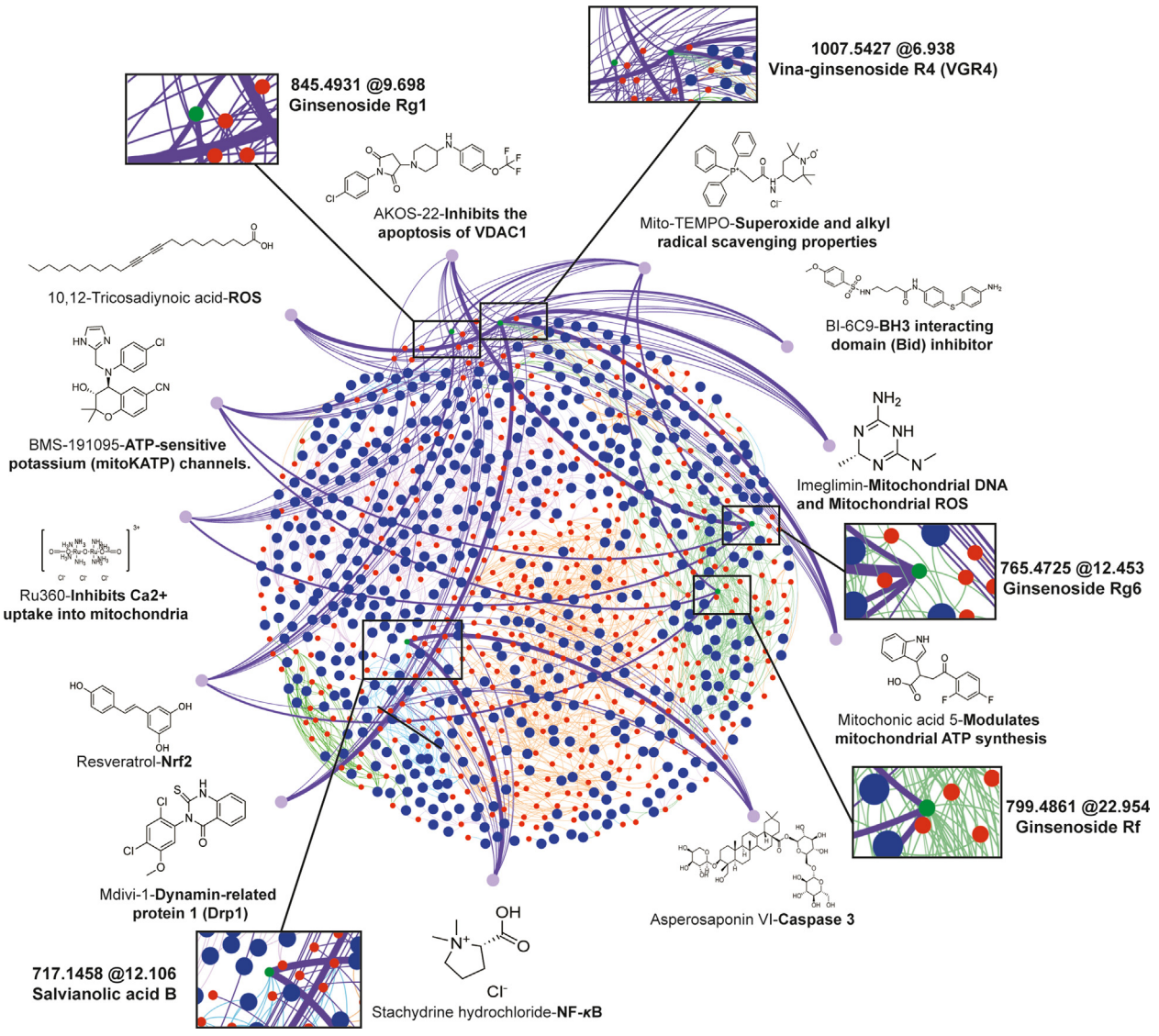

Figure 4 The network of efficient annotation of functional compounds. The thickness of the line represents the similarity of the PFP that between the Feature (green and red nodes) and the reference compounds (purple nodes). The green nodes are the highlight of the features (compounds) which strongly correlated to the reference compounds.

energy currency of the cell, ATP, *via* respiration, and to regulate cellular metabolism[32]. To verify that vina-ginsenoside R_4 can regulate the level of ATP, the ATP content in H9c2 cells was determined by the ATP content commercial kit. As the Fig. 5A, ATP content in H9c2 cells decreased significantly after 2 h of hypoxia. However, ATP content increased in the vina-ginsenoside R_4 group compared with the control group. Moreover, vina-ginsenoside R_4 was associated with mitochondrial dynamin-related protein 1 (Drp1) by ML analysis (Fig. 4 and Fig. S10C). As a regulatory protein of mitochondrial division, Drp1 plays an important role in mitochondrial health and division. Previous studies have shown that the expression of the Drp1 protein leads to myocardial cell apoptosis under hypoxia conditions[33]. As shown in Fig. 5B, compared with the control, the expression level of Mito-Drp1 was increased in the hypoxia. Simultaneously, the Total-Drp1 protein was constant in cells (Fig. 5C). The results showed that a large amount of activated Drp1 was translocated to the mitochondria under hypoxia conditions. Vina-ginsenoside R_4 group decreased the expression of Drp1 protein in mitochondria (Mito-Drp1, Fig. 5B), which had a protective effect on mitochondria. Consistently, Drp1 of NRVMs was downregulated by vina-ginsenoside R_4 dose-dependently in mitochondrial (Mito-Drp1, Fig. 5D and E).

Moreover, in order to prove that compound vina-ginsenoside R_4 may also have a direct interference on Drp1 protein. Vina-ginsenoside R_4 was diluted in PBS to obtain a series of concentrations (*i.e.*, 48−780 nmol/L). The affinity of vina-ginsenoside R_4 to Drp1 protein was assessed using the Biacore T200 Evaluation Software 3.0. The assessment results showed that vina-ginsenoside R_4 could interact with Drp1 protein. The equilibrium dissociation constant (K_D) was used to evaluate the binding activity of each compound. The K_D value for the interaction between vina-ginsenoside R_4 and Drp1 was 1.39×10^{-6} mol/L (Fig. 5F). This indicates that vina-ginsenoside R_4 has a strong binding capacity with Drp1 protein. To quantify the number of mitochondria, we treated H9c2 cell with NAO and chose the mean fluorescence

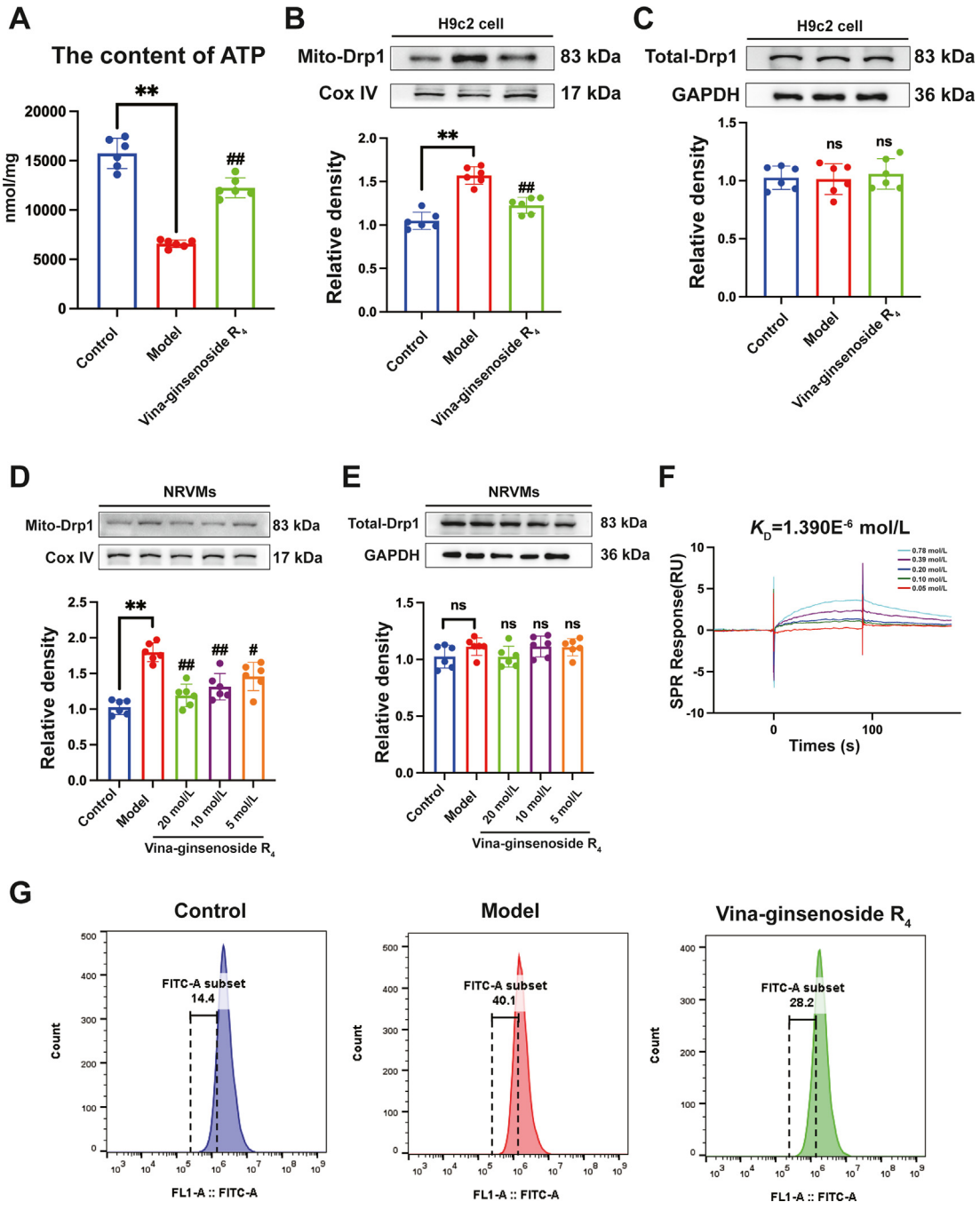

Figure 5 Validation of prediction for vina-ginsenoside R_4. (A) The level of ATP in the H9c2 cell. (B) The analysis of Mtio-Drp1 protein expression in mitochondria of H9c2 cell. (C) The analysis of Total-Drp1 protein expression in the H9c2 cell. (D) The analysis of Mtio-Drp1 protein expression in mitochondria of NRVMs. (E) The analysis of Total-Drp1 protein expression in NRVMs. (F) The analysis of affinity between vina-ginsenoside R_4 and Drp1 by SPR. A series of concentrations (0.05−0.78 mol/L) of vina-ginsenoside R_4 were tested to obtain the affinity between VR_4 and Drp1 by kinetic analysis. (G) The mitochondrial number labeled with nonyl acridine orange (NAO) was viewed using flow cytometry in the H9c2 cell. All data are repeated in six independent experiments. *P* values are determined by one-way ANOVA followed by Tukey's test. All data are presented as mean ± SEM ($n = 6$); $^{*}P < 0.05$, $^{**}P < 0.01$. $^{\#}P < 0.05$, $^{\#\#}P < 0.01$.

intensity of 8×10^6 as an arbitrary reference point. In the control, only 14.4% of cells had a signal below this threshold; however, the ratio raised to 40.10% after anoxic conditions (Fig. 5G). Contemporary, the protective effect of vina-ginsenoside R_4 on mitochondria of H9c2 cardiomyocytes under hypoxia conditions was verified again using high-content system fluorescence quantification (Supporting Information Fig. S13 A and S13B).

As acute hypoxia was capable of boosting superoxide production of mitochondrial, especially superoxide[34]. As shown in Fig. 4, Fig. S10A and S10G, cluster analysis showed that ginsenoside Rf mitochondrial reactive oxygen species production and mitochondrial membrane potential ($\Delta\psi m$). The result showed that ginsenoside Rf, decreased superoxide of mitochondrial generation in response to hypoxia (Fig. S13C and S13D). As expected,

ginsenoside Rf prevented the collapse of mitochondrial membrane potential ($\Delta\psi$m) in H9c2 cardiomyocytes exposed to hypoxia (Fig. S13E and S13F). Tetramethylrhodamine methyl ester (TMRM) results were consistent with those of flow cytometry (Fig. S13G and S13H). Related studies have shown that activation of related proteins contributes to myocardial protection and regulation of reactive oxygen species (ROS) under hypoxia[35]. According to machine learning analysis of candidate compounds and target compounds (Fig. 4 and Fig. S10I), ginsenoside Rg_6 may act on VDAC1 and be related to mitochondrial calcium. It plays a crucial role in both mitochondrial metabolism and cell death, which was the mechanism by which mitochondria of cells in hypoxia, protect from apoptosis[36]. In hypoxia, the fluorescence of calcium ions in mitochondrial co-location was enhanced significantly. However, fluorescence decreased in the ginsenoside Rg_6 group (Fig. S13I and J). Moreover, the mitochondrial membrane potential in the ginsenoside Rg_6 treatment group was also shown to be somewhat regulated (Fig. S13K and S13L). Meanwhile, according to the results of VDAC1 protein expression, ginsenoside Rg_6 effectively inhibited the expression of VDAC1 protein after hypoxia (Fig. S13M). Consistently, mitochondrial VDAC1 of NRVMs was down-regulated by ginsenoside Rg_6 dose-dependently (Fig. S13N).

4. Discussion

The anthology of ancient medicine still served as a useful guide in pharmacy practice and has remained an important source of new drug discovery[3,6]. The knowledge base grid of TCMs practices improves the success rate of screening and makes the purpose of screening more explicit and possible activity. As a typical example, Youyou Tu et al. conducted an in-depth review of the literature. The only reference relevant to the use of qinghao (the Chinese name of *A. annua* L.) for alleviating malaria symptoms appeared in Ge Hong's "*A Handbook of Prescriptions for Emergencies*"[5,37]. Therefore, we explored the knowledge of ancient Chinese medicine prescriptions and established the knowledge base of TCMs (Fig. 1). Meanwhile, the screening results show that of 81 lead compounds, 33 were derived from Renshen (Table S7). Renshen was consistent with the statistical results of the clinical use of TCMs and was the most frequent (Fig. 2A, 129 and 27.15%). Although there have been many studies on the protective effect of ginseng on cardiomyocytes, we have found new active components to protect cardiomyocytes (ginsenoside Rf, vinaginsenoside R_4 and ginsenoside Rg_6). Therefore, we have established a TCMs-CFA platform following the application of ancient Chinese medicine, which can improve the efficiency of new drug discovery.

Besides, chemical profiling of TCMs and their secondary metabolites has been widely reported, which provides multi-dimensional information for the complex system of TCMs[9]. However, the inability to discover active ingredients through chemicalome alone was its limitation. Therefore, we constructed TCMs-CFA platform, which combines the chemome and cytological database by Python. The effective correlation and integration of the two data will facilitate the discovery of active compounds, or lead compounds, in the complex systems of TCMs. It mainly included: (1) selection of herbal drugs of target based on TCMs knowledge base; (2) chemome profiling of TCMs extract library by LC−MS; (3) cytological profiling of TCMs extract library by high-content cell-based imaging; (4) active compounds discovery by combining each mass signal and multi-parametric cell phenotypes; (5) construction of functional annotation map for predicting the potential mechanisms of lead compounds. Compared to the methods established by previous researchers[16], we use the form of a cumulative distribution function when screening features, which are defined by reasonable statistical methods (Fig. 3B−E). Furthermore, high-content imaging technology was different from the previous fluorescence imaging, it has a more excellent image analysis system. It outputs hundreds of morphological parameters through deep learning of imaging[38−40], which was conducive to the discovery of tiny disturbances of cell phenotype caused by trace compounds in TCMs Fractions and the study of pharmacodynamic substances of TCMs. However, the focus of cell phenotypic research was demonstrating changes in cell phenotypic parameters in previous studies[17,41]. No correlation has been shown between changes in these phenotypic parameters and modeling. Hence, correlation coefficients were introduced to demonstrate that the selected phenotypic parameters were associated with the protection of H9c2 cells in our research (Fig. S7).

Different from traditional activity screening, multidimensional parameters of cell mapping can be obtained based on the above screening method, and morphological and phenotypic data of cells can be collected more comprehensively[42]. However, previous studies have added these multidimensional parameters for screening, which leads to the loss of a large number of active compounds[43]. In our approach, we focus on the regulation of subcellular organelle parameters, thus obtaining different active compounds in different subcellular organelles (Fig. 3G). Moreover, we used ML to integrate the above data for activity screening. TCMs-CFA platform integrates ML such as Pearson correlation coefficient, cumulative distribution function, 'Wald' method clustering analysis, method of K-means center distance, and network visualization for data processing and screening. As for the mechanism prediction of the screened compounds, the method of clustering with the target compounds was used to solve the mechanism prediction and exploration of the screened compounds. Finally, the process was completed by Python, integrated with a program, and each step through the Python package output results and image analysis. The codes supporting this study are available on GitHub: https://github.com/CPU-XING/TCMs-High-content-data-and-chemicalome-integration-platform.git. Subsequently, we will also integrate this platform to develop a new open online platform in the future.

Previously, the data for functional annotation maps of TCM active ingredients came from different databases, such as the network pharmacology, in which the data lacked corresponding experimental data of components and activities and which rely on database analysis[44]. The TCMs-CFA platform was constructed activity annotation maps by correlating the chemomics with high-content imaging data of various components to predict active components and their potential mechanism of action, in which the high-content imaging data of the drug with a clearly target were used as reference. To verify this strategy, we take TCMs with myocardial protection as a case. Furthermore, we integrated the "Fraction-Feature (Compound)-Prediction target" information of 7 kinds of TCMs through the TCMs-CFA platform and built a relationship network. Then, the most common clustering method used in phenotypic screening was introduced to predict potential targets. By introducing the above data into the atlas, the rapid annotation of the active ingredients of TCMs was realized. The results showed that predicted targets of some compounds have been confirmed in previous studies. For example, salvianolic acids

(salvianolic acid A, salvianolic acid B, etc.) from Danshen protect cardiomyocytes through NF-κB pathway[26,27]. Saponins from Renshen, Sanqi, and Huangqi (ginsenoside Rb1[45], astragaloside IV[46], etc.), and hydroxysafflor yellow A in Honghua protect cardiomyocytes[47] by influencing mitochondrial function. The known results validate the reliability of our method. Furthermore, many compounds (vina-ginsenoside R_4, etc.) have not been reported. To prove the reliability of our method, we verify three Fractions (ginsenoside Rf, vina-ginsenoside R_4, and ginsenoside Rg_6). Based on established high-content imaging data, the TCMs-CFA platform predicted that the cell phenotype of vina-ginsenoside R_4 was similar to that of Mdivi-1, a Drp1 inhibitor[48] (Fig. 4). In previous studies, Mdivi-1 treatment was found to reduce Drp1 expression in mitochondrial[49,50]. Therefore, the Mito-Drp1 protein expression in vina-ginsenoside R_4-treated cells was measured (Fig. 5B). Vina-ginsenoside R_4 block the mitochondrial translocation of Drp1 protein, while Total-Drp1 protein content unchanged (Fig. 5C), which is consistent with the previous studies that have shown that Mdivi-1 markedly inhibited Drp1 translocation to the mitochondria[51,52]. However, it is unclear whether vina-ginsenoside R4 is an absolute target inhibitor of Drp1 or influences Drp1-related phenotypes (phenotypes similar to Mdivi-1, Fig. 4). Therefore, we further verified the binding of vina-ginsenoside R_4 to Drp1 by SPR. As the result shows, vina-ginsenoside R_4 was bonded to Drp1 directly (Fig. 5F). Besides, phenotypic prediction does not accurately identify pharmacologically related targets of compounds of interest, which only predicts underlying similar mechanisms[39,53]. Therefore, further research is needed to determine the true target of vina-ginsenoside R_4 therapy.

5. Conclusions

In our study, a novel strategy of a basic knowledge-oriented compounds functional annotation map of TCMs-CFA platform was constructed and developed for rapidly discovering candidate compounds. Without the need for chemical separation and comparison one by one, the strong active ingredients are directly predicted from the complex system of TCMs, and the automation was realized through Python. This platform is available at GitHub: https://github.com/CPU-XING/TCMs-High-content-data-and-chemicalome-integration-platform.git. To verify this strategy, active fractions of myocardial protection were rapidly annotated with potential molecular mechanisms from 7 high-frequency drugs to promote blood circulation and remove blood stasis, and the related mechanisms of the three compounds (ginsenoside Rf, vina-ginsenoside R_4, and ginsenoside Rg_6) were also verified. Together, this platform reduces the time consumption to discover the interest molecules, overcomes the bottleneck of traditional target-based and phenotypic assays in TCMs-based drug discovery, and allows driven TCMs library exploration by providing a global view of compound diversity. In the future, the TCMs-CFA could be applied to screen the lead compounds with various activities.

Acknowledgments

This work was supported by National Key R&D Program of China (2022YFC3500302), National Natural Science Foundation of China (No. 82173963), Key R&D Project of Shandong Province (No. 2021CXGC010507, China), and Open Research Fund of Chengdu University of Traditional Chinese Medicine State Key Laboratory Southwestern Chinese Medicine Resources (No. 2022ZYXK2011015, China).

Author contributions

Xudong Xing, Mengru Sun, and Zifan Guo performed experiments, analyzed data, and drafted the manuscript. Ping Zhou and Huiying Wang provided reagents to help design experiments and Mass spectrum data processing. Yongjuan Zhao and Mengru Sun prepared and analyzed the components of TCMs. Xudong Xing and Yuru Cai were used to extract and isolate primary cardiomyocytes from neonatal rats. Yongjuan Zhao and Yuru Cai, helped edit the manuscript. Hua Yang, Ping Li, and Wen Gao, the principal investigators, conceived the scientific ideas, oversaw the project, designed the experiments, and refined the manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

Appendix A. Supporting information

Supporting data to this article can be found online at https://doi.org/10.1016/j.apsb.2023.06.002.

References

1. Xu HY, Zhang YQ, Liu ZM, Chen T, Lv CY, Tang SH, et al. ETCM: an encyclopedia of traditional Chinese medicine. *Nucleic Acids Res* 2018;**47**:976−82.
2. Nemoto S, Morita K, Mizuno T, Kusuhara H. Decomposition profile data analysis for deep understanding of multiple effects of natural products. *J Nat Prod* 2021;**84**:1283−93.
3. Newman DJ, Cragg GM. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J Nat Prod* 2020;**83**:770−803.
4. Atanasov AG, Zotchev SB, Dirsch VM. International natural product sciences taskforce, Supuran CT. Natural products in drug discovery: advances and opportunities. *Nat Rev Drug Discov* 2021;**20**:200−16.
5. Tu Y. Artemisinin-A gift from traditional Chinese medicine to the world (nobel lecture). *Angew Chem, Int Ed* 2016;**55**:10210−26.
6. Huang K, Zhang P, Zhang ZH, Youn JY, Wang C, Zhang HC, et al. Traditional Chinese medicine (TCM) in the treatment of COVID-19 and other viral infections: efficacies and mechanisms. *Pharmacol Therapeut* 2021;**225**:107843.
7. Isgut M, Rao M, Yang C, Subrahmanyam V, Rida P, Aneja R. Application of combination high-throughput phenotypic screening and target identification methods for the discovery of natural product-based combination drugs. *Med Res Rev* 2018;**38**:504−24.
8. Harvey AL, Edrada-Ebel R, Quinn RJ. The re-emergence of natural products for drug discovery in the genomics era. *Nat Rev Drug Discov* 2015;**14**:111−29.
9. Guo MS, Zhu FD, Qiu WQ, Qiao G, Law Betty YK, Yu L, et al. High-throughput screening for amyloid-β binding natural small-molecules based on the combinational use of biolayer interferometry and UHPLC−DAD-Q/TOF-MS/MS. *Acta Pharm Sin B* 2022;**12**:1723−39.
10. Song HP, Chen J, Hong JY, Hao HP, Qi LW, Lu J, et al. A strategy for screening of high-quality enzyme inhibitors from herbal medicines based on ultrafiltration LC−MS and in silico molecular docking. *Chem Commun* 2015;**51**:1494−7.
11. Chen LD, Lv DY, Wang SZ, Wang DY, Chen XF, Liu Y, et al. Surface plasmon resonance-based membrane protein-targeted active ingredients recognition strategy: construction and implementation in

ligand screening from herbal medicines. *Anal Chem* 2020;**92**:3972−80.
12. Erb MA. Processing for destruction. *Nat Chem Biol* 2020;**16**:3−4.
13. Varkuti BH, Kepiro M, Liu Z, Vick K, Avchalumov Y, Pacifico R, et al. Neuron-based high-content assay and screen for CNS active mitotherapeutics. *Sci Adv* 2020;**6**:8702.
14. Bryce NS, Failes TW, Stehn JR, Baker K, Zahler S, Arzhaeva Y, et al. High- content imaging of unbiased chemical perturbations reveals that the phenotypic plasticity of the actin cytoskeleton is constrained. *Cell Syst* 2019;**9**:496−507.
15. Mattiazzi UM, Styles EB, Verster AJ, Friesen H, Boone C, Andrews BJ. High-content screening for quantitative cell biology. *Trends Cell Biol* 2016;**26**:598−611.
16. Kurita KL, Glassey E, Linington RG. Integration of high-content screening and untargeted metabolomics for comprehensive functional annotation of natural product libraries. *Proc Natl Acad Sci U S A* 2015;**112**:11999−2004.
17. Pang HQ, Zhou P, Meng XW, Yang H, Li Y, Xing XD, et al. An image-based fingerprint-efficacy screening strategy for uncovering active compounds with interactive effects in yindan xinnaotong soft capsule. *Phytomedicine* 2021;**96**:153911.
18. Lang P, Yeow K, Nichols A, Scheer A. Cellular imaging in drug discovery. *Nat Rev Drug Discov* 2006;**5**:343−56.
19. Bray MA, Singh S, Han H, Davis CT, Borgeson B, Hartland C, et al. Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat Protoc* 2016;**11**:1757−74.
20. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 2006;**313**:1929−35.
21. Amdi A. Pearson's correlation coefficient, *p*-value, and lithium therapy. *Biol Psychiatr* 1987;**22**:926−8.
22. Akao M, Ohler A, Orourke B, Marban E. Mitochondrial ATP-sensitive potassium channels inhibit apoptosis induced by oxidative stress in cardiac cells. *Circ Res* 2001;**88**:1267−75.
23. Hao PP, Jiang F, Chen YG, Yang J, Zhang K, Zhang MX, et al. Traditional Chinese medication for cardiovascular disease. *Nat Rev Cardiol* 2015;**12**:115−22.
24. Liu C, Yu H. Chinese herbal medicine on cardiovascular diseases and the mechanisms of action. *Front Pharmacol* 2016;**7**:469.
25. Malo N, Hanley JA, Cerquozzi S, Pelletier J, Nadon R. Statistical practice in high-throughput screening data analysis. *Nat Biotechnol* 2006;**24**:167−75.
26. Zheng X, Liu H, Ma M, Ji J, Zhu F, Sun L. Anti-thrombotic activity of phenolic acids obtained from *Salvia miltiorrhiza* F. alba in TNF-α-stimulated endothelial cells *via* the NF-κB/JNK/p38 MAPK signaling pathway. *Arch Pharm Res (Seoul)* 2021;**44**:437−8.
27. Ren DC, Du GH, Zhang JT. Inhibitory effect of salvianolic acids on endothelial cells damage induced by hydrogen peroxide. *Chin J Pharmacol Toxicol* 2003;**17**:333−7.
28. Xu H. Ginsenoside Rg1 protects H9c2 cells against nutritional stress-induced injury *via* aldolase/AMPK/PINK1 signalling. *J Cell Biochem* 2019;**120**:18388−97.
29. Sarhene M, Ni JY, Duncan ES, Liu Z, Li S, Zhang J, et al. Ginsenosides for cardiovascular diseases; update on pre-clinical and clinical evidence, pharmacological effects and the mechanisms of action. *Pharmacol Res* 2021;**166**:105481.
30. Zheng Q, Bao XY, Zhu PC, Tong Q, Zheng GQ, Wang YQ, et al. Ginsenoside Rb1 for myocardial ischemia/reperfusion injury: preclinical evidence and possible mechanisms. *Oxid Med Cell Longev* 2017:6313625.
31. Bravo-San Pedro JM, Kroemer G, Galluzzi L. Autophagy and mitophagy in cardiovascular disease. *Circ Res* 2017;**120**:1812−24.
32. Nishimura A, Shimauchi T, Tanaka T, Shimoda K, Toyama T, Kitajima N, et al. Oxygen is the high-energy molecule powering complex multicellular life: fundamental corrections to traditional bioenergetics. *ACS Omega* 2020;**5**:2221−33.
33. Nishimura A, et al. Hypoxia-induced interaction of filamin with Drp1 causes mitochondrial hyperfission-associated myocardial senescence. *Sci Signal* 2018;**11**:556.
34. Zhao RZ, Wang XB, Jiang S, Ru NY, Yu ZB, Jiao B, et al. Elevated ROS depress mitochondrial oxygen utilization efficiency in cardiomyocytes during acute hypoxia. *Pflug Arch Eur J Phy* 2020;**472**:1619−30.
35. Hernansanz-Agustín P, Ramos E, Navarro E, Parada E, Sánchez-López N, Peláez-Aguado L, et al. Mitochondrial complex I deactivation is related to superoxide production in acute hypoxia. *Redox Biol* 2017;**12**:1040−51.
36. Mazure NM. News about VDAC1 in hypoxia. *Front Oncol* 2016;**6**:193.
37. Tu Y. The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine. *Nat Med* 2011;**17**:1217−20.
38. Yang SJ, Berndl M, Michael Ando D, Barch M, Narayanaswamy A, Christiansen E, et al. Assessing microscope image focus quality with deep learning. *BMC Bioinf* 2018;**19**:77.
39. Lin S, Schorpp K, Rothenaigner I, Hadian K. Image-based high-content screening in drug discovery. *Drug Discov Today* 2020;**25**:1348−61.
40. Krentzel D, Shorte SL, Zimmer C. Deep learning in image-based phenotypic drug discovery. *Trends Cell Biol* 2023;**1**:17.
41. Abraham VC, Taylor DL, Haskins JR. High content screening is applied to large-scale cell biology. *Trends Biotechnol* 2004;**22**:15−22.
42. Beghin A, Kechkar A, Butler C, Levet F, Cabillic M, Rossier O, et al. Localization-based super-resolution imaging meets high-content screening. *Nat Methods* 2017;**14**:1184.
43. Lee S, Van Santen JA, Farzaneh N, Liu DY, Pye CR, Baumeister TUH, et al. NP analyst: an open online platform for compound activity mapping. *ACS Cent Sci* 2022;**8**:223−34.
44. Zhou W, Cheng X, Zhang Y. Effect of liuwei dihuang decoction, a traditional Chinese medicinal prescription, on the neuroendocrine immunomodulation network. *Pharmacol Therapeut* 2016;**162**:170−8.
45. Kong HL, Li ZQ, Zhao YJ, Zhao SM, Zhu L, Li T, et al. Ginsenoside Rbl protects cardiomyocytes against $CoCl_2$-induced apoptosis in neonatal rats by inhibiting mitochondria permeability transition pore opening. *Acta Pharmacol Sin* 2010;**6**:9.
46. Luo Y, Wan Q, Xu M, Zhou Q, Chen X, Yin D, et al. Nutritional preconditioning induced by astragaloside IV on isolated hearts and cardiomyocytes against myocardial ischemia injury *via* improving Bcl-2-mediated mitochondrial function. *Chem Biol Interact* 2019;**309**:108723.
47. Feng J, Guo J, Yan J, Zhang X, Qu H, Yang T, et al. Luhong formula and hydroxysafflor yellow A protect cardiomyocytes by inhibiting autophagy. *Phytomedicine* 2023;**11**:154636.
48. Tanaka A, Youle RJ. A chemical inhibitor of DRP1 uncouples mitochondrial fission and apoptosis. *Mol Cell* 2008;**29**:409−10.
49. Aishwarya R, Alam S, Abdullah CS, Morshed M, Nitu SS, Panchatcharam M, et al. Pleiotropic effects of mdivi-1 in altering mitochondrial dynamics, respiration, and autophagy in cardiomyocytes. *Redox Biol* 2020;**36**:101660.
50. Wang P, Li Y, Yang ZF, Yu T, Zheng GH, Fang XS, et al. Inhibition of dynamin-related protein 1 has neuroprotective effect comparable with therapeutic hypothermia in a rat model of cardiac arrest. *Transl Res* 2018;**194**:68−78.
51. Ding M, Dong Q, Liu Z, Liu Z, Qu Y, Li X, et al. Inhibition of dynamin-related protein 1 protects against myocardial ischemia−reperfusion injury in diabetic mice. *Cardiovasc Diabetol* 2017;**16**:19.
52. Rogers MA, Hutcheson JD, Okui T, Goettsch C, Singh SA, Halu A, et al. Dynamin-related protein 1 inhibition reduces hepatic PCSK9 secretion. *Cardiovasc Res* 2021;**117**:2340−53.
53. Conway LP, Li W, Parker CG. Chemoproteomic-enabled phenotypic screening. *Cell Chem Biol* 2021;**28**:371−93.