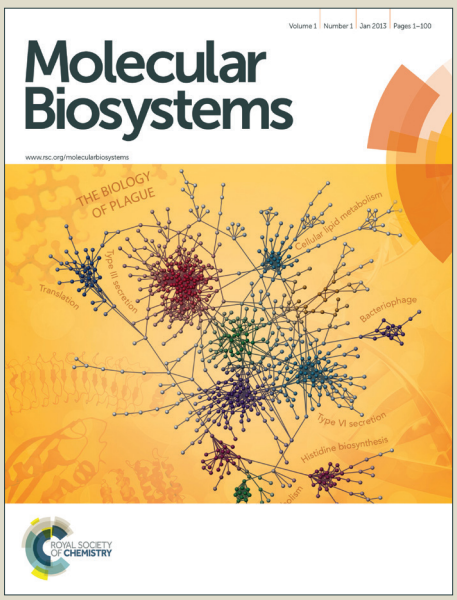
Molecular BioSystems

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

Molecular BioSystems

RSCPublishing

ARTICLE

A novel network pharmacology approach to analyse traditional herbal formula: Liu-wei-di-huang Pill as a case study

Cite this: DOI: 10.1039/x0xx00000x

Received 00th January 2012,
Accepted 00th January 2012

DOI: 10.1039/x0xx00000x

www.rsc.org/

Xujun Liang,[a]† HuiyingLi,[a]† Shao Li[a]*

Understanding the mechanisms of the pharmacological effects of herbal formulae from traditional Chinese medicine (TCM) is important for their appropriate application. However, this understanding has been impeded by the complex nature of herbal formulae. An herbal formula is a mixture of hundreds of chemical ingredients with multiple potential targets. The effects produced by an entire herbal formula cannot be adequately explained by isolatedly considering each ingredient in it. This is a recognised problem that remains in need of methods to solve it. Here we introduce a holistic analysis method to decipher the molecular mechanisms of herbal formulae. This method combines chemical and pharmacokinetic properties with network pharmacology, using a novel approach to evaluate the importance of the targets and ingredients of herbal formulae. We used Liu-wei-di-huang (LWDH) pill, a classic herbal formula, as an example to illustrate our method and validated some results by a following experiment. We revealed the core molecular targets and bioprocess network of the pharmacological effects of LWDH and inferred its therapeutic indications. This method provides a strategy to understand the mechanisms of herbal formulae in a holistic way and implies new applications of classic herbal formulae.

Introduction

Traditional Chinese medicine (TCM) is a comprehensive medicinal system that has been used in clinical practice for thousands of year. It is still regarded as an important part of complementary and alternative medical systems and a rich source for drug discovery. For these reasons, it is important to understand the scientific basis and action mechanism of TCM.

TCM treats diseases primarily with herbal formulae, which consist of several herbs and other natural products. Determining the pharmacological effects of herbal formulae is important to the study of TCM mechanism, but this study presents a substantial challenge for several reasons. First, herbal formulae are mixtures of hundreds of chemical compounds. Second, unlike the majority of current drugs, which are designed to selectively act on single target, most of the active chemical ingredients in herbs may weakly or moderately act on multiple cellular targets[1]. Third, herbal formulae are formulated according to special syndromes or patterns ("ZHENG" in Chinese) instead of targeting disease as modern medicine does. These facts make it difficult to systematically study the mechanisms of herbal formulae with routine methods. Previous studies of herbal formulae mainly focused on particular effects of herbal formulae produced in animal models of diseases or in clinical trials[2, 3]. With the advance of high throughput technology, studies have also utilised gene expression microarrays[4], proteomics[5], and metabolomics[6] to explore the mechanisms of herbal formulae. However, these experiments still cannot fully overcome the limitations imposed by the complex nature of herbal formulae. To achieve a comprehensive and systematic understanding of the mechanisms of herbal formulae, new methods and strategies such as the computational ways are urgently needed.

In our previous works, we proposed that network pharmacology is a promising way to understand multiple-component drugs such as herbal formulae[7]. We then presented a novel concept, the "network target", which extended the concept of a drug target from its effect on a single component to its systematic effect on the biological network, to address multi-component therapeutics such as herbal formulae[8]. We also constructed an integrative platform for herbal formula network pharmacology and successfully applied it to analyse the anti-rheumatoid arthritis formula *Qing-Luo-Yin*[9]. Recently explaining the mechanism of herbal formulae with network pharmacology methods is well-accepted[10], however, there remains a need for methods that holistically consider herbal formulae as a gestalt whose emergent properties cannot be explained by analysing each chemical ingredient separately.

To obtain a better understanding of how herbal formulae affect different biological processes and treat different diseases, in this work we propose a new method for network pharmacology analysis of herbal formula. This method was used to analyse Liu-wei-di-huang (LWDH), a classic herbal formula that has been used since the 11th Century in China. LWDH was formulated to tonify the "Yin deficiency pattern" in TCM. More recently, it has been applied in clinical settings to treat various complex diseases such as hypertension and oesophageal carcinoma[11, 12].

In our previous work, we found a multilayer herb-biomolecule-disease network by manually curating the genes and diseases related to LWDH[13]. We also revealed a metabolism-immune network imbalance underlying TCM Cold Syndrome and Hot Syndrome, a pair of typical Syndrome reflecting the Yin-Yang imbalance of human body[14]. Here, as shown in **Fig.1**, we used purely computational methods to predict the effects and mechanisms of LWDH. First, we analysed the chemical group composition of the ingredients in LWDH, explored their chemical characteristics and distribution in chemical space, and measured their drug-likeness properties. Second, the biological targets of LWDH ingredients were predicted. We hypothesised that if a biological molecule was an effective target of an herbal formula, there should be multiple ingredients in the herbal formula that target this molecule (**Fig.1A**). Under this hypothesis, we found the most represented target molecules of LWDH compared to the null model with Poisson binomial statistics and ranked the ingredients in LWDH by defining an efficacy score. The diseases potentially treated by LWDH were also predicted. Third, we constructed a network that integrated compound-target relationships and target-disease relationships and a network that reflected the common biological processes related to compounds and diseases. We found that the predicted targets of LWDH were highly connected in a protein-protein interaction (PPI) network, and LWDH treated diseases through both shared and separate biological molecules and processes via different groups of compounds. A following experiment was conducted and validated the effects of seven compounds belonging to different chemical groups. This work provides a method for explaining the action mechanisms by which herbal formulae such as LWDH can treat a variety of diseases.

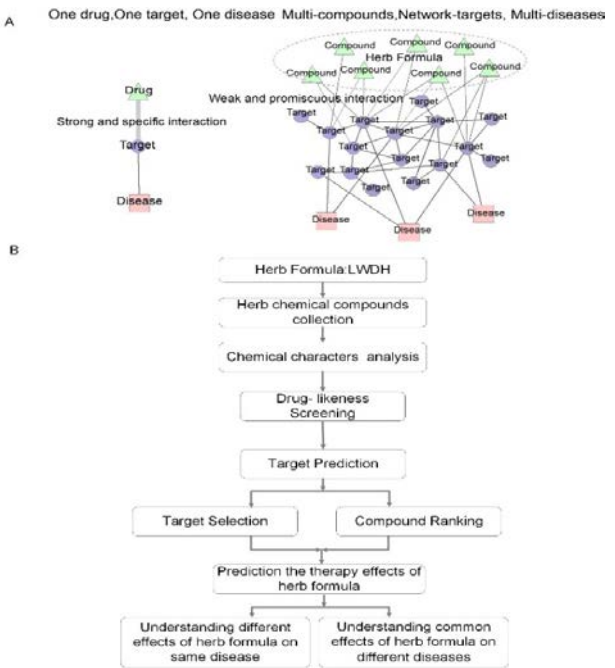


Fig 1. Network pharmacology analysis model. (A) An herbal formula is a multi-component complex system with effects on a variety of targets and with therapeutic potential for a variety of diseases, in contrast with most modern approved drugs, which are designed for one specific target and to treat one specific disease. (B) The steps for our analysis model construction.

Methods

Curation and clustering analysis of chemical ingredients in LWDH

A total of 311 chemical constituents of LWDH were retrieved from HerBioMap, a database built constructed by our group, and from TCM Database @Taiwan[15] as well as literatures[16-18] (Table S1).The names of these chemicals were used to search PubChem[19] to obtain chemical structures. Next, chemical structure clustering was conducted for the ingredients in each herb separately and for all ingredients in LWDH combined using the Chemical Structure Clustering Tool in PubChem. The clustering results were visualised by Multi-Dimensional Scaling (MDS) in the form of scatter plots.

Chemical space and drug-likeness calculation

The physicochemical properties of ingredients in LWDH were calculated using the Chemistry Development Kit[20]. The properties included molecular mass (MW), octanol–water partition coefficient (ALogP), the numbers of hydrogen bond donors (HBDs), the number of hydrogen bond acceptors (HBAs), molecular polar surface area (PSA), the number of rotatable bonds (ROTBs), and the number of aromatic rings (AROMs). 771 approved drugs were collected as described in previous study[21] , and the same properties were calculated as for the herb ingredients. Principal component analysis (PCA) was conducted with the calculated properties, and the three largest principal components were chosen for the scatter plot. Next, the drug-likeness properties of the herb ingredients and approved drugs were calculated using the same method described in previous report[21]. In brief, seven properties above pulsing the number of structural alerts in each compound were used to calculate the weighted quantitative estimate of drug-likeness (QED_w). For this work, ingredients in LWDH with $QED_w \geqslant 0.3$ were included for further study. The threshold of 0.3 was chosen because it is the optimal operating point of the ROC curve[21]. Some ingredients, such as some tannins and glycosides are hydrolysable; the drug-likenesses of these ingredients were calculated after deglycosylation. In order to provide a negative control, 30 compound sets that had the same number of compounds as in LWDH were randomly selected from PubChem database and analyzed for the seven chemical properties. These compounds sets were also screened with the same drug-likeness criteria for following analyses.

Target profile prediction and target selection

The compound target profiles were predicted using our drugCIPHER-CS step[22]. To robustly predict compound-target relationships, both the PPIs and drug-target relationships were randomly shuffled 1000 times to calculate concordance scores while keeping the number of drugs interacting with each target fixed. For each ingredient, predicted targets with P-value≤0.001, as calculated from the random permutations, were considered to be high precision targets for the ingredients. Some target proteins may appear in the target profiles of many ingredients in an herbal formula. We infer that such target proteins may have a high probability of being key players in the pharmacological effects of the herbal formula and represent the combined effects of multiple ingredients in herbal formula. To assess the probability of target proteins being related to the herbal formula's pharmacological effects, we compared the

number of occurrences of each target protein in the target profiles of all ingredients in LWDH to a pure random process represented by a Poisson binomial statistical mode:

$$\Pr(K = k) = \sum_{A \epsilon F_k} \prod_{i \in A} p_i \prod_{j \in A^c} (1 - p_j)$$

Pr(K=k) is the probability that a target protein occurs in the target profiles of k ingredients, F_k is the set of all subsets of k ingredients, A is one particular subset of k ingredient, and A^c is the complement of A. p_i and p_j are the probabilities of a target protein being contained in the target profiles of an ingredient. By random chance, p is m/n, where m is the number of target proteins in the ingredient's target profile and n is the total number of target proteins in the drugCIPHER library. The P-value, Pr (K>k), after being adjusted by the Bonferroni method, measures the probability of a target protein occurring in more than k ingredients' target profiles by random chance. This P-value indicates the relative importance of target proteins for LWDH (they are significant when P-value < 0.01). A total of 128 target proteins were selected with this threshold as a target protein set.

Computation of ingredient efficacy scores

To measure the effectiveness of ingredients in LWDH, we define an efficacy score for ingredients as follows:

$$Score_{ingredient\ i} = \frac{1}{N_i} \sum_{j=1}^{N_i} -\frac{1}{r_{ij}} Log_{10}[P(k)_j] \times I_j$$

Where N_i is the number of target proteins in the target profile of ingredient i, r_{ij} is the rank of target protein j in the target profile of ingredient i, $\Pr(k)_j$ is the target protein's P-value calculated from the Poisson binominal model, k is the number of ingredients that have target j, and I_j is an indicator showing whether target protein j was in selected target:

$$I_j = \begin{cases} 1, & If\ the\ j\ th\ target\ protein\ is \\ & \text{in the selected target protein set,} \\ 0, & otherwise. \end{cases}$$

This score considers both the ingredient specificity and its relationship with important target proteins. We ranked the ingredients according to their scores and found that the top 25% ingredients could cover 90% of the selected targets. So we chose these chemicals as representative ingredients in LWDH (**Table S2**) for network construction.

Biological function and disease ontology enrichment analysis

KEGG pathway profiling and GO biological processes (BP) were used for functional enrichment analysis of the sets of selected target protein and of disease related genes. The functional enrichment tool DAVID[23] was used to calculate both the KEGG pathway and GO BP enrichment. Only GO terms with P-values<0.01 and KEGG pathways with P-values<0.05 were included (both were corrected using the Benjamini method). Disease ontology enrichment analysis was conducted according to previous method[24]. Diseases with P-value<0.05 (adjusted by the Bonferroni method) were included.

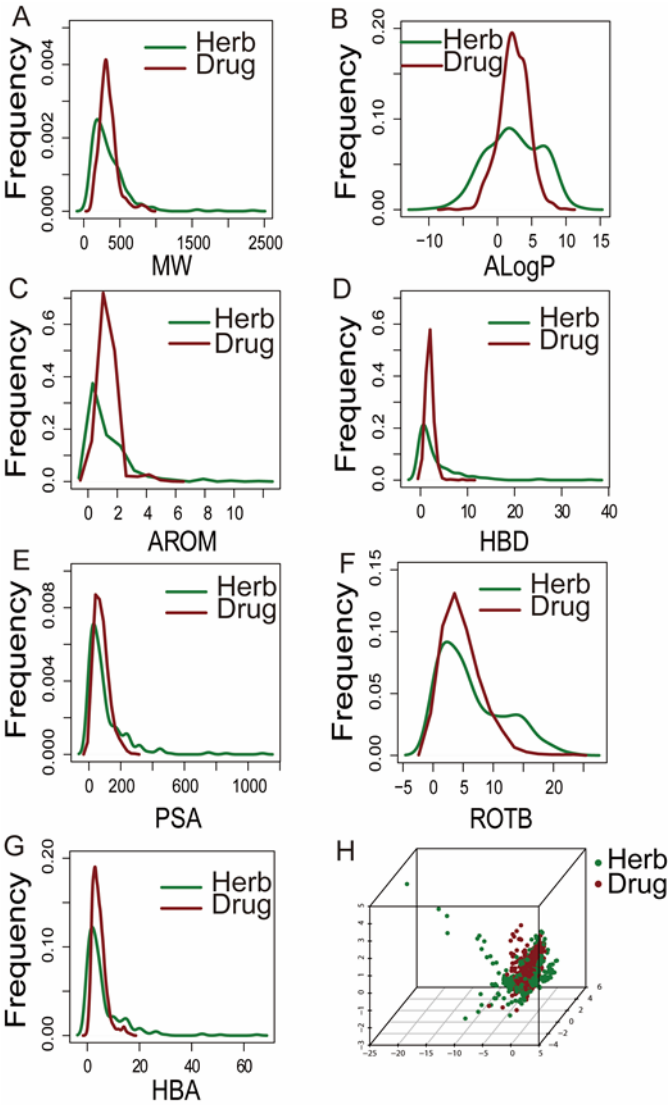


Fig 2. Comparing chemical characteristics of ingredients in LWDH (A)-(G) are distributions of seven chemical characteristics of ingredients in LWDH and approved drugs. MW: Molecular mass; ALogP: octanol–water partition coefficient; HBDs: number of hydrogen bond donors; HBA: number of hydrogen bond acceptors; PSA: molecular polar surface area; ROTB: number of rotatable bonds; AROM: number of aromatic rings. (H) PCA of ingredients in LWDH and approved drugs calculated from these seven chemical characteristics.

Construction of the target network of herbal formula

For the compound-target-disease network, first, the selected targets were mapped to PPI. Hyper-geometric tests were conducted for the genes directly linked to the selected targets in PPI[25]. The tests aimed to access if genes connected to the selected targets just by chance. The target protein network which included selected target and genes statistically significantly linked to them via PPI (P-value<0.01) was constructed. Second, compounds were connected to their targets. Third, an edge was added between a target protein and a disease if the target protein's gene was in the related gene list for the disease.

Table 1 Enriched KEGG pathways of LWDH selected targets.

Pathway class	KEGGID	Term	P-value
Endocrine system	ko03320	PPAR signalling pathway	5.20E-05
	ko04914	Progesterone-mediated oocyte maturation	0.0285
	ko04920	Adipocytokine signalling pathway	0.0323
Excretory system	ko04960	Aldosterone-regulated sodium reabsorption	0.0347
Immune system	ko04062	Chemokine signalling pathway	0.0321
	ko04666	Fc gamma R-mediated phagocytosis	0.0334
	ko04662	B cell receptor signalling pathway	0.0485
Nervous system	ko04722	Neurotrophin signalling pathway	0.0136
Signal transduction	ko04370	VEGF signalling pathway	0.0158

For the compound-bioprocess-disease network, the enriched GO BP terms were identified for disease gene lists and the selected target set. If a GO term was shared between diseases and the selected target set, then this term was retained. To give more specific information and to keep the final network concise, we chose GO terms at level 4~6 and discarded others. If a GO term was enriched for a disease's gene list, the GO term and the disease were connected; if a compound's targets were annotated by a GO term, the compound and the GO term were connected; and if two GO terms had similarity>0.7, which was calculated by the method previously described[26], these two GO terms were connected. Networks were visualised using Cytoscape[27].

Experimental validation

To evaluate the predicted compound-target relationships, seven compounds of interest in were purchased commercially (**Table S3**). For Western blot experiments, HT29 cells were seeded and cultured at a density of 1.5×10^6 in a 6-wellplate with DMEM supplemented with 10% FBS, then incubated with the 10mMtestcompoundsat 37°C. After incubation for 24 h, cells were harvested and Western blotting was conducted. The following antibodies were used: anti-CCR2 (ARP58409_P050, Rabbit, Aviva systems biology), anti-ER1 (sc-542, Rabbit, ZSGB-BIO), anti-PPAR (ZS-72730, Mouse, ZSGB-BIO), and anti-RARA (sc-15040, Goat, Santa Cruz). Beta-actin (TA-09, Mouse, ZSGB-BIO) was used loading control. Experiments were repeated three times.

Results and discussion

Compound families, chemical space and drug-likeness properties of compounds in LWDH

LWDH consists of 6 herbs, *Radix Rehmanniae Preparata, Fructus Corn, Rhizoma Dioscoreae, Rhizoma Alismatis, Cortex Moutan* and *Poriacocos*. To give an overview of the compound families in LWDH, chemical clustering was conducted (**Fig. S1**).The ingredients in LWDH could be roughly divided into three major groups: iridoid glycosides, phenylpropanoid-based aromatic compounds such as tannins, sterols and terpenes (**Fig. S1G**). The clustering results show that although there are numerous different ingredients in LWDH, many of them have similar chemical structures and belong to particular compound families.

The physicochemical characteristics of a compound are important for its absorption, distribution, metabolism, excretion and toxicity (ADMET) properties[28]. Comparing the physicochemical characteristics of ingredients in LWDH with FDA-approved drugs will provide insight into the ADMET properties of these ingredients. Here, seven physicochemical characteristics of the ingredients in LWDH were compared with approved orally administered drugs (**Fig. 2A-G**). The overall shapes of the distributions of these characteristics are similar between the ingredients in LWDH and approved drugs, but the standard deviations for ingredients in LWDH are larger than for approved drugs (**Table S4**). The AlogP distribution of ingredients in LWDH is much wider than for approved drugs (**Fig. 2B**).This result indicates that the water solubility of ingredients in LWDH is more variable than in approved drugs. Furthermore, as shown in **Fig. 2F**, the proportion of compounds with more than 10 rotatable bonds (ROTB) in LWDH is more than in approved drugs, which means the structures of ingredients in LWDH are more flexible. We also calculated the seven physicochemical characteristics for randomly selected chemical compound sets. It appears that both the approved drugs and herbal ingredients have different chemical property distributions from random compounds (**Fig. S4**).

To comprehensively compare the properties of ingredients in LWDH with approved drugs, principal component analysis was conducted using all seven physicochemical characteristics. The combination of these characteristics spans a chemical space. As seen in **Fig. 2H**, the distribution of approved drugs in the chemical space is more compact, while the distribution of ingredients in LWDH is scattered more widely in chemical space. However, there is still considerable overlap between the ingredients in LWDH and approved drugs, which indicates that many ingredients in herbs have drug potential.

To assess the probability for each ingredient in LWDH to be a drug, drug-likeness analysis was conducted. Drug likeness is a composite description that evaluates the bioavailability of a chemical compound[29]. Here we employed a method using a concept called weighted quantitative estimate of drug-likeness (QED_w) to measure drug-likeness[21]. The closer the QED_w value of a compound is to 1,

Table 2 Enriched disease of LWDH selected targets.

Disease class	Disease Name	P-value	Supporting Literatures
Cardiovascular Diseases	Atherosclerosis	7.41E-16	30
	Heart failure	5.50E-08	31
	Hypertension	5.58E-10	32
Digestive System Diseases	Cholelithiasis	1.79E-08	NA
	Esophagitis	0.0452	33
Eye Diseases	Glaucoma	1.91E-07	34
Neoplasms	Esophagus cancer	3.82E-05	12
	Endometrium cancer	0.00238	NA
	Colon cancer	0.0356	NA
Nutritional and Metabolic Diseases	Obesity	1.01E-09	35
	Hyperlipidemia	0.00891	36
Mental Disorders	Bipolar disorder	2.26E-05	NA
	Panic disorder	0.00461	NA
	Depression	0.000139	37
Musculoskeletal Diseases	Osteoporosis	2.09E-05	38,39
	Arthritis	2.09E-05	40,41
Urogenital Diseases	Endometriosis	1.77E-10	42
	Infertility	2.15E-06	43
Stomatognathic Diseases	Dental plaque	0.000194	44

NA: not available.

the more favourable drug-likeness properties it has. The median of the QED_w values of ingredients in LWDH is 0.38, and the interquartile range is 0.272. The ingredients in LWDH were screened based on QED_w values as described in method. Only the ingredients passed the screening left for further study. To reveal the pharmacological effects of LWDH, the targets of its ingredients were predicted and selected as described in the Methods. These targets were used to represent the bio-molecular basis for the pharmacological effects of LWDH.

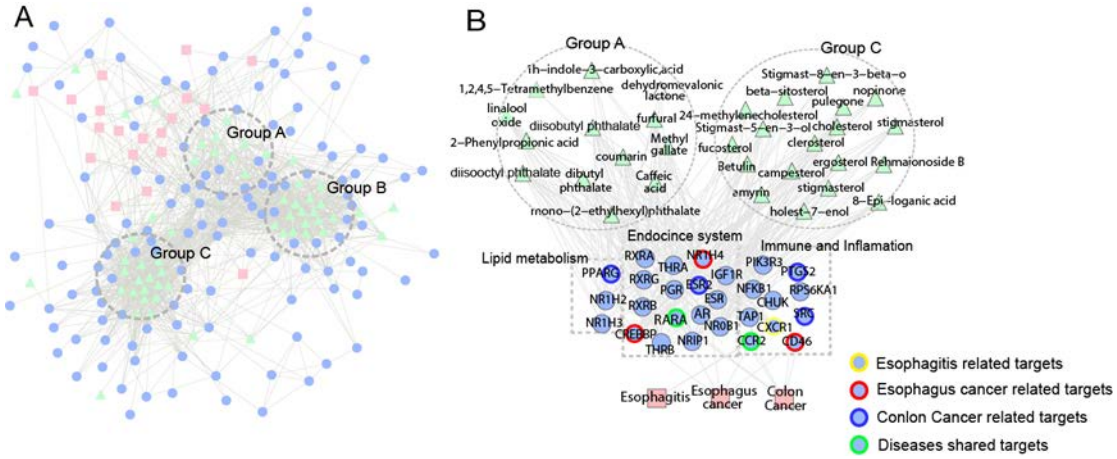


Fig 3. Compound-target-disease network (A) The whole network of selected ingredients (green), selected targets (blue) and diseases (red). Ingredients were roughly clustered as three groups based on different targets that these ingredients linked to. (B) A sub-network from (A) that contains the targets related to colon cancer, esophagitis and esophageal cancer, and ingredients linked to these targets. Targets were grouped according to their function. The nodes with colored edges are targets directly related to these three diseases.

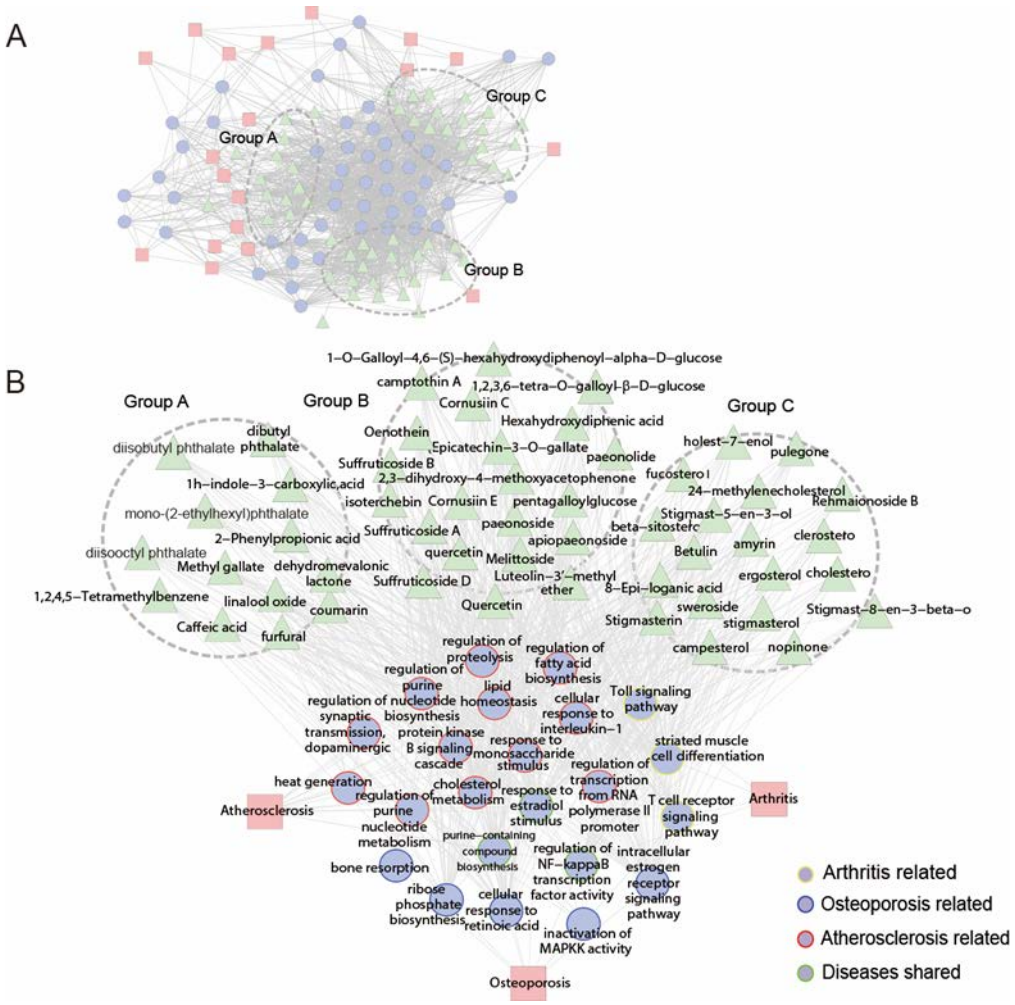


Fig 4. Compound-bioprocess-disease network (A) the network of selected ingredients (green), common enriched biological process GO terms of selected targets and diseases genes (blue) and diseases (red). Three groups of ingredients were circled by grey dashed ellipses based on the different bioprocesses that these ingredients linked to. The members in these three groups here are mostly the same as in Fig 3. (B) A sub-network from (A) that contains the bioprocesses related to atherosclerosis, arthritis and osteoporosis and ingredients related to these bioprocesses. The colored edges of the nodes indicate the relationships between these bioprocesses and these three diseases.

Inference of the pathways and diseases affected by LWDH

To elucidate the biological pathways that LWDH might impact, the significantly overrepresented KEGG pathways were identified (**Table 1**). The result contains endocrine related pathways such as the PPAR signalling pathway (P-value=5.20e-5),which controls lipid metabolism, cell proliferation and blood glucose uptake[45]; excretory system pathways such as Aldosterone-regulated sodium reabsorption (P-value=0.0347), which is important for regulation of blood pressure[46]; and immune system pathway such as the Chemokine signalling pathway (P-value=0.0321). We also found that both the selected targets and the pathway enrichment results overlapped with the results in our previous work[13]. Furthermore, the targets of random compound sets were predicted and selected under the same procedure. We compared the KEGG enrichment P-value distributions of the targets of LWDH and the targets of random compound sets, and found that the P-value distribution of LWDH targets was different from the random background (**Fig. S5**).

To uncover the therapeutic potential of the selected targets, disease ontology enrichment was conducted. The result of disease enrichment is shown in **Table 2**. For most of the enriched diseases (73.7%), there have been clinical reports or experimental evidence that LWDH can be taken as a therapy or an adjuvant therapy. Some other diseases traditionally or clinically targeted by LWDH, such as diabetes mellitus and kidney failure, were not significantly enriched, but they had related genes overlapped with the selected targets of LWDH. The pathway enrichment and disease ontology enrichment results imply that our method does indeed capture the underlying molecular basis and therapeutic effects of LWDH.

The target network reveals the effects of LWDH on different diseases

To understand how a multicomponent treatment system such as LWDH exerts pleiotropic effects on different diseases, a compound-target-disease network was constructed to uncover the relationships among herb ingredients, target proteins and diseases. First, we mapped the selected target proteins to PPI and found that the selected targets were highly connected. The size of the largest component of the sub-network formed by the selected targets in PPI is significant bigger than the largest component generated by randomly selected gene sets of the same gene number (P-value<0.001, 5000 iterations).This result implies that the selected targets reflect the core molecular basis of LWDH effects. Second, to measure the effectiveness of ingredients in LWDH, an efficacy score was defined and used to select representative ingredients. We found that the selected ingredients included many compounds with known pharmacological effects. For example, Paeonolide from *Cortex Moutan* is a maker compounds of LWDH that has broad pharmacological effects, and sweroside from *Fructus Corni* is another marker compound of LWDH[47]. Subsequently, the selected ingredients were connected to their predicted targets. Finally, the diseases from the enrichment result were added to the network. It is noteworthy that the ingredients with similar connections to their targets in the network cluster together to form 3 groups (**Fig.3A** and detailed in **Fig. S2**).The ingredients in each cluster have similarities in chemical structure, and some of the ingredients belong to same chemical family. These clusters were assigned to group A, most of which were heterocyclic and aromatic compound; group B, most of which were tannins; or group C, which were triterpenes and steroids. Target proteins that relate to hormone signalling, such as ESR1, NCOA1 and AR are not only the hubs of the PPI network, but also have high connection numbers to ingredients of LWDH (**Table S5**). ESR1 and AR are also highly connected to disease. This result implies that the regulation of hormone homeostasis is one of the key factors of LWDH pharmacological effects.

To comprehensively reveal the biological processes underlying the therapeutic effects of LWDH, a compound-bioprocess-disease network was constructed (**Fig. 4A** and detailed in **Fig. S3**). This network connects ingredients and diseases by the biological processes that are common to both. From this network, we can make following observations: 1) Most of the ingredients form three clusters, as in the compound-target-disease network above. 2) The biological progresses in the centres of the networks are impacted by most of ingredients of all three clusters. These biological progresses include response to hormones such as estrogen and retinoic acid, lipid and cholesterol metabolism, inflammation and immunity. Interestingly, LWDH is traditionally used for Yin deficiency pattern. Yin deficiency pattern is a holistic description of a special group of symptoms in TCM and could be a common character underlying different diseases in modern medicine. Our results suggest that LWDH can regulate the imbalance of hormone, metabolism and immunity, which agree with the reports that Yin deficiency pattern is related to dysfunction of sterol metabolism, energy consumption, endocrinal and immune functions[48, 49]. 3) One disease might relate to many biological processes that are targeted by different groups of ingredients. For example, osteoporosis is related to bone resorption, the estrogen receptor pathway and NF-κB activity (**Fig. 4B**). Bone resorption is targeted by ingredients of group A and group B, the estrogen receptor pathway is targeted by ingredients from all groups ,and NF-κB activity is targeted mainly by group C.4) One biological process may be related to different diseases. Our disease enrichment results and previous reports showed that LWDH could target a wide range of diseases (**Table 2**).These diseases may share the same phenotypic characters belonging to Yin deficiency pattern. And as observed in **Fig. 4B** and **Fig. 3B**, some of these diseases share similar molecular mechanisms and targeted by same ingredients. For example, the NF-κB signalling pathway is related to osteoporosis, arthritis and atherosclerosis (**Fig. 4B**). In our network, triterpenes(group C) target the NF-κB pathway, and previous studies show that some triterpenes treat osteoporosis[50], arthritis[51] and atherosclerosis[52] by controlling NF-κB pathway activity.

Experimental validation of prediction results

In order to confirm our predictions of the molecular mechanism of LWDH experimentally, four proteins were chosen for test. PPARG, RARA and CCR2 were selected as they denote different functions, targeted by different groups of ingredients (**Fig. 3B**) and related to esophageal cancer, peptic esophagitis and colon cancer (**Fig. 3B** and **Fig. 5A**). These diseases are potential and particular therapeutic indications of LWDH and deserve our concern. ESR1 as a hub of the network was also chosen (**Fig. 3B** and **Table S5)**. It is interesting to examine if ESR1 could be affected by different groups of ingredients. Seven compounds that were related to these proteins were chosen. We chose these compounds based on their ingredients scores, the groups they belong to (**Fig.3B**) as well as their herb sources to see how ingredients from different herbs affect the same proteins. The effects of these compounds on the expression levels of their related proteins were analysed by Western blot.

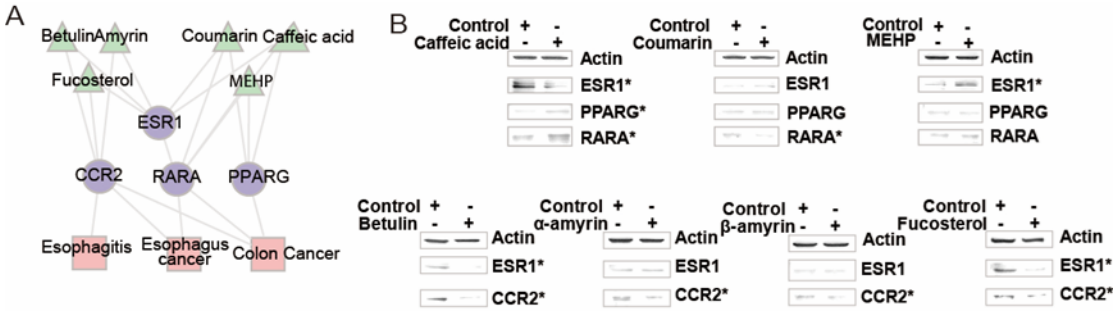


Fig 5. Compound-targets relationship validation by Western blot (A) The links between two groups of compounds and four targets. (B) The influence of these compounds on the expression levels of the four genes was assayed by Western blot. Amyrin has two isoforms, alpha and beta, both of which were tested. Genes with changed expression were labelled with *.

LWDH has been reported to prevent the development of esophageal cancer from esophagitis[12]. Retinoids play an important role in glandular differentiation of the esophagus[53]. On the other hand, retinoids could inhibit the growth of tumour cells[54]. As **Fig.5B** shows, coumarin decreased the expression of RARA while caffeic acid increased the expression of RARA. Activation of PPARG could inhibit growth of colon cancer cell[55], and caffeic acid increased the expression of PPARG. CCR2may promote cancer and may be a therapeutic target in cancer[56]. Betulin, α-amyrin, β-amyrin and fucosterol all down regulated the expression of CCR2. Betulin, fucosterol and caffeic acid also down regulated the expression of ESR1. These results imply that our method could accurately predict the relationships of ingredients to the proteins and could give hints of the therapeutic mechanisms of LWDH. The results also indicate that there are complex interactions between the effects of the different ingredients in herbal formulae.

Conclusions

In this work, we proposed a novel analysis method for identification of the molecular mechanisms of herbal formulae based on integration of multi-component effects. Using this method, we analysed LWDH, a classic herbal formula containing hundreds of ingredients, and derived the target network affected by the ingredients of LWDH. From this network, we inferred the links between herb ingredients and diseases through molecular targets and the biological processes. Some of these links have been reported in previous studies. Our method also could help to discern the core mechanism of the whole herbal formula. For LWDH, the target network and bioprocess network imply that the key pharmacological effects and therapeutic indications of LWDH may lie in maintaining homeostasis in the endocrine system, the immune system and metabolism.

There are some limitations in the present method. First, the omission of ingredients in herbal formulae may produce bias and incomplete results. Fortunately, our method could easily integrate newly found ingredients and then give a more comprehensive understanding of the given herbal formula. Second, our method could not discriminate whether ingredient could directly bind a target or just affect the target indirectly. Meanwhile, there still need to be more tests for the effects of ingredients on different diseases in the future work.

In summary, this work focuses on the multi-ingredient and multi-target property nature of herbal formulae. Based on a network pharmacology approach, we revealed the possible therapeutic mechanisms of LWDH, which traditionally treats Yin deficiency pattern in TCM, for diseases defined by modern medicine.

Acknowledgements

We thank Dr. Gu Jin for helpful discussion. This work is supported by NSFC (91229201, 81225025 and 60934004) and Tsinghua-Giti project.

Notes and references

[a]MOE Key Laboratory of Bioinformatics and Bioinformatics Division, TNLIST, Department of Automation, Tsinghua University, Beijing 100084, China

*To whom correspondence should be addressed.

†The first two authors should be regarded as joint First Authors

‡Electronic Supplementary Information (ESI) available. See

DOI: 10.1039/b000000x/

1. H. F. Ji, X. J. Li and H. Y. Zhang, *EMBO reports*, 2009, **10**, 194-200.
2. S. Kummar, M. S. Copur, M. Rose, S. Wadler, J. Stephenson, M. O'Rourke, W. Brenckman, R. Tilton, S. H. Liu, Z. Jiang, T. Su, Y. C. Cheng and E. Chu, *Clinical colorectal cancer*, 2011, **10**, 85-96.
3. W. Lam, S. Bussom, F. Guan, Z. Jiang, W. Zhang, E. A. Gullen, S. H. Liu and Y. C. Cheng, *Science translational medicine*, 2010, **2**, 45ra59.
4. J. Li, R. G. Wu, F. Y. Meng, Z. Wang, C. M. Wang, Y. Y. Wang and Z. J. Zhang, *PloS one*, 2012, **7**, e45811.
5. X. Liu and D. A. Guo, *Biochemical Society transactions*, 2011, **39**, 1348-1352.
6. A. Zhang, H. Sun, Z. Wang, W. Sun, P. Wang and X. Wang, *Planta medica*, 2010, **76**, 2026-2035.
7. S. Li, *Journal of Chinese integrative medicine*, 2007, **5**, 489-493.
8. S. Li, *China journal of Chinese materia medica*, 2011, **36**, 2017-2020.
9. B. Zhang, X. Wang and S. Li, *Evidence-based complementary and alternative medicine : eCAM*, 2013, **2013**, 456747.
10. E. L. Leung, Z. W. Cao, Z. H. Jiang, H. Zhou and L. Liu, *Briefings in bioinformatics*, 2013, **14**, 491-505.
11. H. J. Cheng Yuling, Chen Wenyuan, Zhang Guoqing, *Clinical Journal of Chinese Medicine*, 2011, **3**, 41.
12. T. L. Jiang, S. C. Yan and L. F. Zhao, *Gan To Kagaku Ryoho*, 1989, **16**, 1511-1518.

13. S. Li, B. Zhang, D. Jiang, Y. Wei and N. Zhang, *BMC Bioinformatics*, 2010, **11 Suppl 11**, S6.
14. R. Li, T. Ma, J. Gu, X. Liang and S. Li, *Scientific reports*, 2013, **3**, 1543.
15. C. Y. Chen, *PloS one*, 2011, **6**, e15939.
16. J. L. Rios, *Planta medica*, 2011, **77**, 681-691.
17. Z. Wu, G. Luo, Y. Huo and X. Zhong, *Zhong Yao Cai*, 1999, **22**, 499-502.
18. G. Zhou and Q. H. Lv, *China journal of Chinese materia medica*, 2008, **33**, 2070-2073.
19. Y. Wang, J. Xiao, T. O. Suzek, J. Zhang, J. Wang and S. H. Bryant, *Nucleic acids research*, 2009, **37**, W623-633.
20. C. Steinbeck, Y. Han, S. Kuhn, O. Horlacher, E. Luttmann and E. Willighagen, *J Chem Inf Comput Sci*, 2003, **43**, 493-500.
21. G. R. Bickerton, G. V. Paolini, J. Besnard, S. Muresan and A. L. Hopkins, *Nat Chem*, 2012, **4**, 90-98.
22. S. Zhao and S. Li, *PloS one*, 2010, **5**, e11764.
23. W. Huang da, B. T. Sherman and R. A. Lempicki, *Nature protocols*, 2009, **4**, 44-57.
24. N. H. Shah, T. Cole and M. A. Musen, *PLoS computational biology*, 2012, **8**, e1002827.
25. E. Cerami, E. Demir, N. Schultz, B. S. Taylor and C. Sander, *PloS one*, 2010, **5**, e8918.
26. C. Pesquita, D. Faria, A. O. Falcao, P. Lord and F. M. Couto, *PLoS computational biology*, 2009, **5**, e1000443.
27. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, *Genome Res*, 2003, **13**, 2498-2504.
28. M. P. Gleeson, A. Hersey, D. Montanari and J. Overington, *Nat Rev Drug Discov*, 2011, **10**, 197-208.
29. T. H. Keller, A. Pichota and Z. Yin, *Curr Opin Chem Biol*, 2006, **10**, 357-361.
30. L. L. J. X. G. Sun, Y. Cai, Y. Y. Zhao, *China Journal of Traditional Chinese Medicine and Pharmacy*, 2006, **21**. 68
31. W. J. Li, *Shaanxi Journal of traditional Chinese Medicine*, 1999, **20**, 340.
32. Z. W. Z. Y. L. Zhou, M. X. Tang, *Guiding Journal of Traditional Chinese Medicine and Pharmacy*, 2003, **9**, 15.
33. J. Y. Z. W. H. Fang, Z. X. Li, B. T. Li, *Journal of Traditional Chinese Medicine*, 1999, **40**, 243.
34. H. S. Y. Zhao, Q.L.Liang, *Journal of Traditional Chinese Medicine*, 2007, **22**. 35
35. B. Perry, J. Zhang, C. Sun, T. Saleh and Y. Wang, *Evidence-based complementary and alternative medicine : eCAM*, 2012, **2012**, 847167.
36. L. J. Y. J. J. Wang, M. Chen, Y. X. Lu, R. G. Wang, *Journal of Fujian University of TCM*, 2012, **22**, 36.
37. W. H. L. D. Z. Fan, F. X. Kong, *Henan Traditional Chinese Medicine*, 2008, **28**, 65.
38. Y. Chen, C. Qu, H. Zhong, Y. Xue, C. Zhou, W. Li and X. Cheng, *Journal of traditional Chinese medicine*, 1994, **14**, 298-302.
39. C. Ma, *China medicine and pharmacy*, 2011, **9**, 74.
40. J. Fang, Y. X. Zhang, X. B. Ru and X. L. Wei, *China journal of Chinese materia medica*, 2001, **26**, 128-131.
41. L. W. B. K. Long, *China's Naturopathy*, 2007, **15**, 34.
42. Y. Q. Z. H. Hong, X. R. Huang, J. Y. Wang, S. J. Guo, *Chinese Journal of Convalescent Medicine*, 2010, **19**.
43. B. Fu, X. Lun and Y. Gong, *Journal of traditional Chinese medicine*, 2005, **25**, 186-189.
44. P. Y. B.Han, *Henan Traditional Chinese Medicine*, 2013, **33**, 731.
45. J. N. Feige, L. Gelman, L. Michalik, B. Desvergne and W. Wahli, *Progress in lipid research*, 2006, **45**, 120-159.
46. S. Viengchareun, D. Le Menuet, L. Martinerie, M. Munier, L. Pascual-Le Tallec and M. Lombes, *Nuclear receptor signaling*, 2007, **5**, e012.
47. X. J. Wang, W. J. Sun, N. Zhang, H. Sun, F. Geng and C. Y. Piao, *Chinese Journal Of Natural Medicines*, 2007, **5**, 277.
48. B. Han, S. Wang, L. Li, Y. Wang and H. Zhao, *Journal of traditional Chinese medicine*, 2013, **33**, 378-383.
49. Q. Wang, X. J. Ren, S. L. Yao and H. D. Wu, *Chinese journal of integrative medicine*, 2010, **16**, 28-32.
50. C. Li, Z. Yang, Z. Li, Y. Ma, L. Zhang, C. Zheng, W. Qiu, X. Wu, X. Wang, H. Li, J. Tang, M. Qian, D. Li, P. Wang, J. Luo and M. Liu, *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research*, 2011, **26**, 644-656.
51. S. M. Nanjundaiah, B. Astry and K. D. Moudgil, *Evidence-based complementary and alternative medicine : eCAM*, 2013, **2013**, 518094.
52. H. Chen, J. Yang, Q. Zhang, L. H. Chen and Q. Wang, *Circulation journal : official journal of the Japanese Circulation Society*, 2012, **76**, 995-1003.
53. C. L. Chang, P. Lao-Sirieix, V. Save, G. De La Cueva Mendez, R. Laskey and R. C. Fitzgerald, *Gut*, 2007, **56**, 906-917.
54. P. Fitzgerald, M. Teng, R. A. Chandraratna, R. A. Heyman and E. A. Allegretto, *Cancer research*, 1997, **57**, 2642-2650.
55. B. S. Tan, O. Kang, C. W. Mai, K. H. Tiong, A. S. Khoo, M. R. Pichika, T. D. Bradshaw and C. O. Leong, *Cancer letters*, 2013, **336**, 127-139.
56. I. Conti and B. J. Rollins, *Seminars in cancer biology*, 2004, **14**, 149-154.