



OPEN

## In silico identification of natural products from Traditional Chinese Medicine for cancer immunotherapy

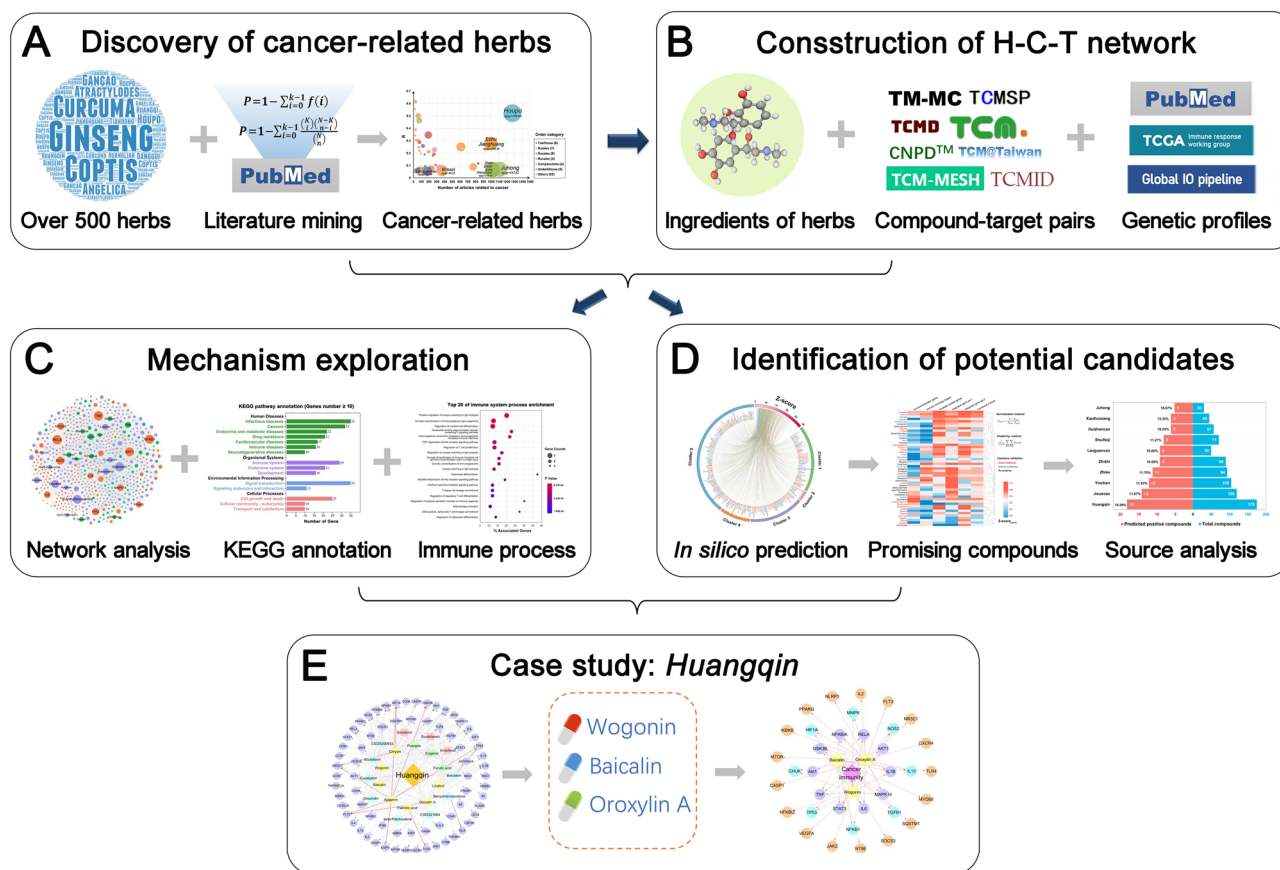
Chuipu Cai<sup>1,2,7</sup>, Qihui Wu<sup>3,7</sup>, Honghai Hong<sup>4,7</sup>, Liying He<sup>2</sup>, Zhihong Liu<sup>5</sup>, Yong Gu<sup>3</sup>, Shijie Zhang<sup>2</sup>, Qi Wang<sup>2</sup>, Xiude Fan<sup>6</sup>✉ & Jiansong Fang<sup>2</sup>✉

Advances in immunotherapy have revolutionized treatments in many types of cancer. Traditional Chinese Medicine (TCM), which has a long history of clinical adjuvant application against cancer, is emerging as an important medical resource for developing innovative cancer treatments, including immunotherapy. In this study, we developed a quantitative and systems pharmacology-based framework to identify TCM-derived natural products for cancer immunotherapy. Specifically, we integrated 381 cancer immune response-related genes and a compound-target interaction network connecting 3273 proteins and 766 natural products from 66 cancer-related herbs based on literature-mining. Via systems pharmacology-based prediction, we uncovered 182 TCM-derived natural products having potential anti-tumor immune responses effect. Importantly, 32 of the 49 most promising natural products (success rate = 65.31%) are validated by multiple evidence, including published experimental data from clinical studies, *in vitro* and *in vivo* assays. We further identified the mechanism-of-action of TCM in cancer immunotherapy using network-based functional enrichment analysis. We showcased that three typical natural products (baicalin, wogonin, and oroxylin A) in *Huangqin* (*Scutellaria baicalensis Georgi*) potentially overcome resistance of known oncology agents by regulating tumor immunosuppressive microenvironments. In summary, this study offers a novel and effective systems pharmacology infrastructure for potential cancer immunotherapeutic development by exploiting the medical wealth of natural products in TCM.

Cancer, as the leading cause of death worldwide, is a serious public health issue<sup>1</sup>. In the past several years, cancer immunotherapy has become a remarkable strategy for cancer treatment and received growing attention worldwide in many cancer types. One of the world-renowned examples is the programmed cell death-1 (PD-1)/PD-1-ligand (PDL-1) pathway, which was recognized with the 2018 Nobel Prize in Physiology or Medicine<sup>2</sup>. The number of active agents in the global immuno-oncology pipeline increased by 91% in the last two years (September 2017 to August 2019)<sup>3</sup>. Cancer immunotherapeutic agents with regulating immune evasion (e.g., co-inhibitory checkpoints) or directly stimulating immunogenic pathways (e.g., agonists of costimulatory receptors) have been demonstrated striking clinical responses in patients with multiple types of advanced cancers<sup>4</sup>. Unfortunately, the response rates of current immunotherapy are still modest that only around 20% of cancer patients are responsive<sup>5</sup>. Therefore, more efforts are clearly required to develop novel agents with new mechanism-of-actions (MOAs) for cancer immunotherapy.

Traditional Chinese Medicine (TCM) has a long history of clinical adjuvant application for cancer treatment<sup>6</sup>. Accumulating clinical evidence and pharmacological studies have demonstrated the great potential of Chinese herbs as well as their extracts as novel and effective cancer interventions<sup>7</sup>. A number of natural products

<sup>1</sup>Department of Computer Science, Key Laboratory of Intelligent Manufacturing Technology of Ministry of Education, Shantou University, Shantou 515000, China. <sup>2</sup>Science and Technology Innovation Center, Guangzhou University of Chinese Medicine, Guangzhou 510000, China. <sup>3</sup>Clinical Research Center, Hainan Provincial Hospital of Traditional Chinese Medicine, Haikou 570100, China. <sup>4</sup>Department of Clinical Laboratory, The Third Affiliated Hospital of Guangzhou Medical University, Guangzhou, China. <sup>5</sup>Guangdong Institute of Microbiology, Guangdong Academy of Sciences, Guangzhou 510000, China. <sup>6</sup>Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44195, USA. <sup>7</sup>These authors contributed equally: Chuipu Cai, Qihui Wu and Honghai Hong. ✉email: FANX@ccf.org; fangjs@gzucm.edu.cn

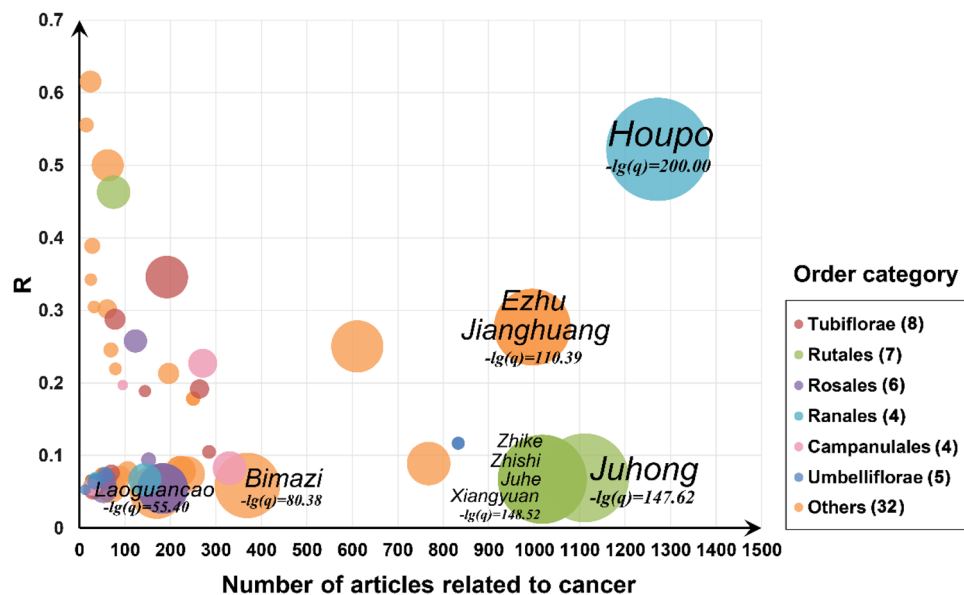


**Figure 1.** Schematic diagram of an integrated systems pharmacology framework for identification of potential cancer immunotherapeutic natural products from Traditional Chinese Medicine (TCM). **(A)** Literature mining-based discovery of cancer-related Chinese Herbs from PubMed database. **(B)** Construction of cancer immune response-related herb-compound-target (H-C-T) network of TCM-derived natural products. **(C)** Mechanism-of-action exploration of TCM in cancer immunotherapy. **(D)** Identification of new cancer immunotherapeutic candidates from natural products. **(E)** *Huangqin* (*Scutellaria baicalensis* Georgi) as a case study to highlight its main active constituents and elucidate the underlying mechanisms of immunotherapeutic against cancer.

derived from TCM, such as berberine<sup>8,9</sup>, baicalin<sup>10</sup>, and resveratrol<sup>11</sup>, have been reported to have regulatory effects on tumor microenvironments, including regulation of macrophage activation, cytokine secretion, and T cell differentiation<sup>12</sup>. Natural products with abundant chemical scaffolds and promiscuous target profiles, are emerging as an invaluable chemical library for discovering bioactive cancer immunotherapeutic candidates<sup>13–15</sup>.

Due to the heterogeneous ingredients and multiple target interactions, conventional experimental assays are labor-consuming and time-costing<sup>16</sup>. Quantitative and systems pharmacology, as a multidisciplinary strategy for the emerging development of efficacious drugs via integration of experimental assays and computational approaches<sup>17</sup>, has been widely applied to identify active natural products and elucidate their underlying mechanisms against complex diseases<sup>18</sup>, such as aging-associated disorders<sup>19,20</sup>, coronary artery diseases<sup>21</sup>, and cancers<sup>16,22</sup>. For example, Huang et al. had successfully uncovered wogonoside as an angiogenesis inhibitor in treating triple-negative breast cancer (TNBC) using systems pharmacology approach and validated it by *in vitro* and *in vivo* assays<sup>23</sup>. Cheng et al. predicted an integrative, network-based systems pharmacology framework for personalized drug repurposing<sup>22</sup>. They successfully identified ouabain, a FDA-approved plant-derived toxic substance in treatment of non-small cell lung cancer<sup>22</sup>. Altogether, quantitative and systems pharmacology approaches offer effective strategies for discovering potential natural products for the development of cancer therapies, including immunotherapy.

In this study, we proposed an integrated systems pharmacology-based framework, as illustrated in Fig. 1, to uncover novel TCM-derived cancer immunotherapeutic agents. Specifically, we firstly identified Chinese herbs highly related to cancer through a large-scale literature mining from PubMed database. Subsequently, a global herb-compound-target (H-C-T) network of TCM derived-natural products was constructed by integrating experimentally reported cancer immune response-related (CIR) genes, herb-compound pairs, and compound-target interactions (CTIs) extracting from our previous consolidated databases<sup>16,23</sup>. On the basis of the H-C-T network, we next systematically inspected the cancer immunotherapeutic mechanisms of TCM, and built *in silico* models with high accuracy to predict natural products for potential immunotherapy. Overall, this study offers a powerful systems pharmacology approach for identification of promising candidates from natural products for development of tumor immunotherapy.



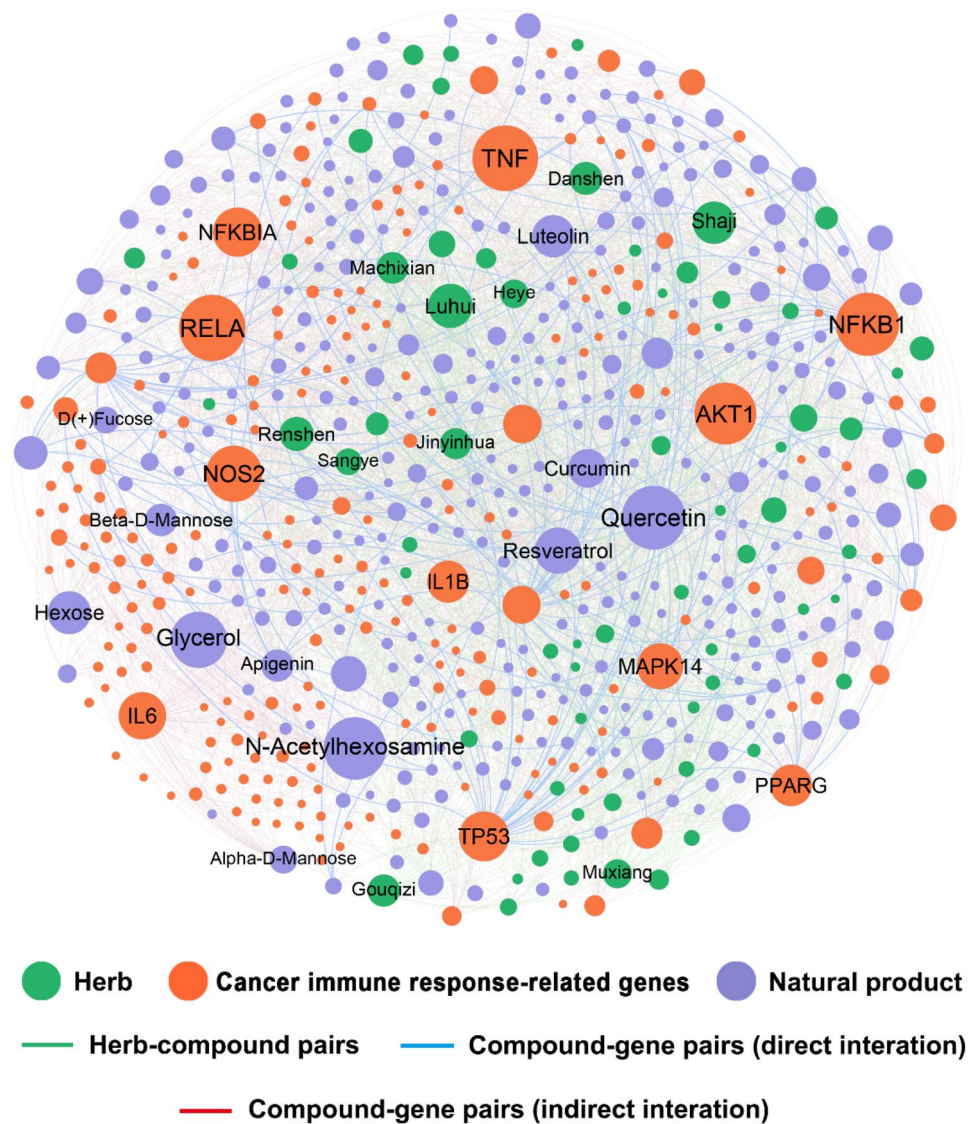
**Figure 2.** Statistics and association analysis results between 66 herbs and cancer based on literature mining from PubMed database. The herbs are prioritized according to the negative logarithm (base 10) of their adjusted- $P$  ( $q$ ), and the size of circles denotes the value of  $-\lg(q)$ .  $R$  represents the ratio of (cancer-herb-related articles)/(herb-related articles). Labels of the top 10 herbs with highest  $-\lg(q)$  values are displayed. The zero value of  $q$  was replaced by an extremely small positive value ( $1E-200$ ) for calculating negative logarithm.

## Results

**A landscape of Chinese herbs in cancer.** Via large-scale literature mining analysis, we found that 66 of the 525 investigated herbs showed significant associations with cancer ( $q < 0.01$ ,  $R > 0.05$ , Supplementary Table S1). As shown in Fig. 2, the top 10 herbs with highest  $-\lg(q)$  values are commonly used in the clinical adjuvant treatment of cancer, while their extracts also have been confirmed to possess anti-tumor effect by *in vivo* and *in vitro* assays<sup>7</sup>. For example, magnolol isolated from the ranking first herb *Houpo* (*Magnolia officinalis*,  $-\lg(q) = 200.00$ ) had been proved to exert broad anti-cancer activity by reducing proliferation, suppressing differentiation, counteracting metastasis, and restraining angiogenesis<sup>24–27</sup>. Besides, *Jianghuang* (*Curcuma longa*,  $-\lg(q) = 110.39$ ), a main source of the polyphenolic compound curcumin, is a natural pleiotropic herb used in treatment and prevention of multiple diseases, including cancer<sup>12</sup>. Put together, the literature mining narrows down the scope of investigation for highlighting the most valuable herbs highly related to cancer, which deserves to further identify their potential ingredients as cancer immunotherapeutic agents.

**Herb-compound-target (H-C-T) network analysis of natural products in cancer immunotherapy.** We firstly acquired the constituent compounds and corresponding targets of the 66 cancer-related herbs from the integrated H-C-T database, resulting in 12,562 interactions connecting 766 unique compounds with 3273 protein targets (Supplementary Table S2). Subsequently, an anti-cancer immune response-related H-C-T network was constructed by extracting the compound-protein interactions related to CIR genes. As shown in Fig. 3, the H-C-T network is composed of 648 nodes (66 herbs, 344 compounds, and 238 targets) and 3270 edges (2095 CTIs and 1175 herb-compound pairs). Among the 2095 CTIs, 186 with direct binding affinity are labeled as direct interactions, while the rest 1909 are regarded as indirect interactions. Network analysis shows that these natural products are connected to multiple cancer immunity-related targets with the average degree ( $K$ ) of 6.1 for each natural product, which indicates that they may have important influence on cancer immunotherapy. For instance, quercetin possessing the third largest number of target connections ( $K = 61$ ), has been reported to promote the immune response and repress the growth of murine leukemia WEHI-3 cells<sup>28</sup>. *In vitro* and *in vivo* studies also suggested that resveratrol ( $K = 56$ ) could suppress lung cancer growth through inhibiting M2 polarization of human monocyte derived macrophage. For the 238 targets, the average degree ( $D$ ) in this network is 8.8. Among them, RELA ( $D = 93$ ) exhibited the highest connections to natural products, followed by TNF ( $D = 92$ ) and NFKB1 ( $D = 88$ ). Relevant studies show that these targets play essential roles in tumor immunity<sup>29,30</sup>.

**Mechanism-of action of natural products in cancer immunotherapy.** To elucidate the molecular mechanisms underlying the cancer immunotherapeutic effects of TCM, 41 CIR genes with degree greater than or equal to 10 ( $D \geq 10$ ) in the H-C-T network (Supplementary Table S3) were selected to perform enrichment analysis (Fig. 4). Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway<sup>31</sup> annotation (Fig. 4A) shows that a large number of targets (T) are involved in cancer ( $T = 32$ ) and immune system ( $T = 29$ ) related pathways, indicating their high correlativity with cancer immunotherapy of TCM. According to the enrichment analysis results

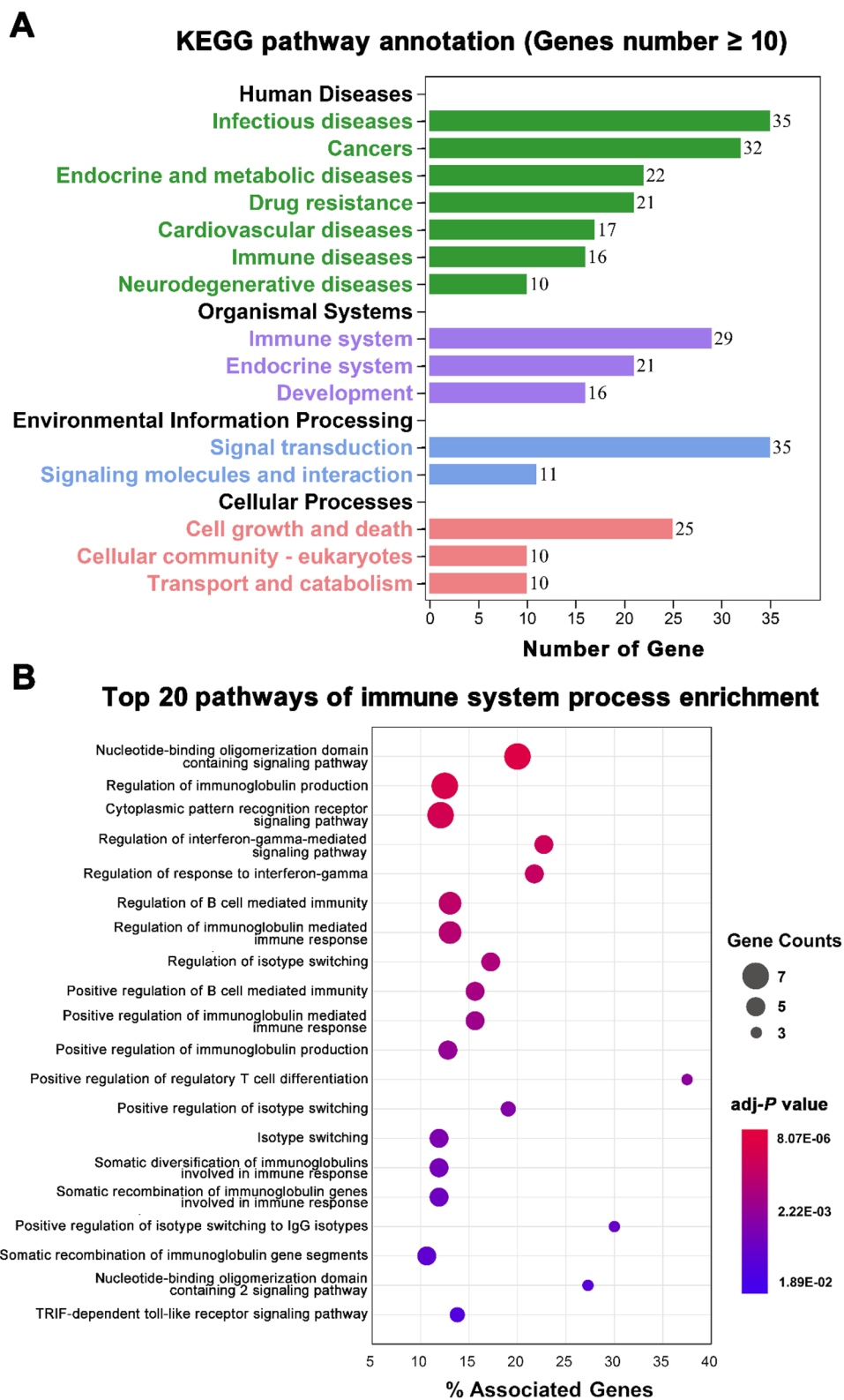


**Figure 3.** Herb-compound-target (H-C-T) network for natural products derived from 66 cancer-related herbs in cancer immunotherapy. The label font size and node size are proportional to degree (connectivity) of the item. Labels of the top 10 herbs, compounds, and targets with the highest degree are displayed. For the compound-target interactions that were labeled as direct and indirect interactions simultaneously, the direct ones were preserved.

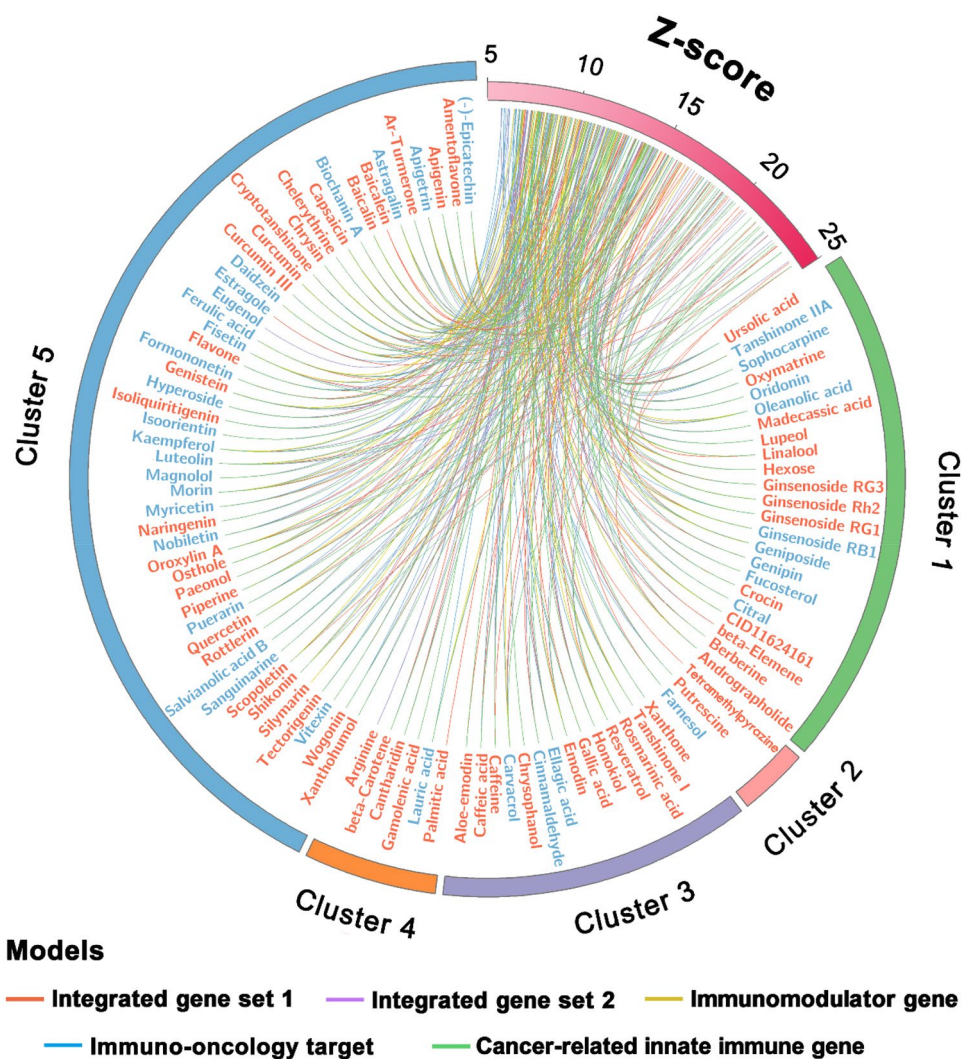
of environment information processing and cellular process, these CIR genes may participate in the regulation of signal transduction and modulate cell growth and death.

In this study, we mainly aim at the immune-related processes and try to dig out the underlying cancer immunotherapeutic mechanisms of TCM. Thus, the immune system process enrichment analysis was further performed using ClueGO<sup>32</sup>. We obtained 31 significant immune system pathways with adjusted *P*-values (*q*) less than 0.05 (corrected with Bonferroni step down) as well as associated genes proportion higher than 10% (Supplementary Table S3). Figure 4B presents the top 20 enriched immune system pathways. Most of these pathways were associated with cancer immunotherapy, which can activate immune responses, including positively regulation of T cell differentiation ( $q = 6.03 \times 10^{-3}$ ), B cell-mediated immunity ( $q = 2.22 \times 10^{-3}$ ), immunoglobulin production ( $q = 5.38 \times 10^{-3}$ ), and secretion of other cytokines ( $q = 0.0499$ ). Besides, tumor immune escape is also an important point. Enrichment analysis demonstrated that these anti-cancer herbs may prevent from tumor immune evasion through regulating immune system processes such as TRIF-dependent toll-like receptor signaling pathway ( $q = 0.0189$ ) and interferon-gamma-mediated signaling pathway ( $q = 3.48 \times 10^{-4}$ ). Detailed immune system process enrichment analysis results of the 41 cancer CIR genes are provided in Supplementary Table S3. In general, TCM-derived natural products may have multi-faceted impact on cancer immunotherapy both via “immune enhancement” and “immune normalization” strategies<sup>4</sup>. Thus, we next turned to identify novel cancer immunotherapeutic agents from TCM-derived natural products.





**Figure 4.** Enrichment analysis of the 41 cancer immune response-related genes with degree higher than or equal to 10 in the herb-compound-target (H-C-T) network. (A) Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway annotation. Different KEGG classes are displayed in various colors. (B) Immune system process enrichment.

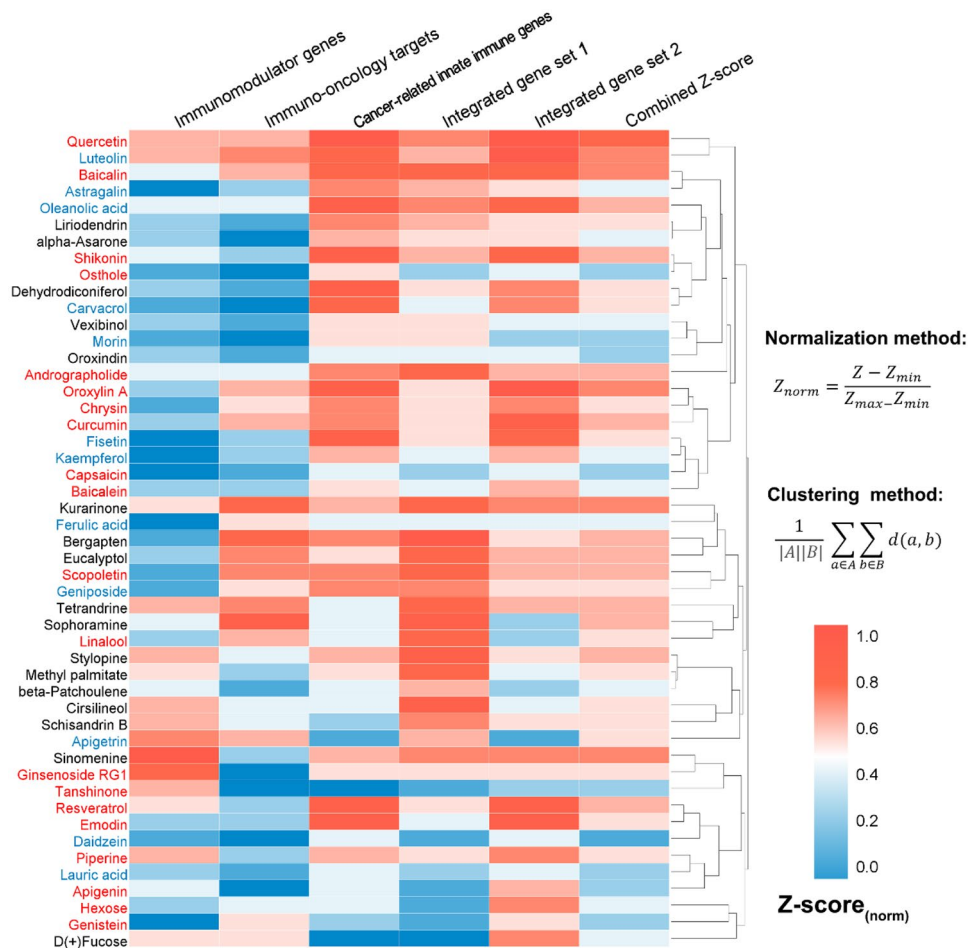


**Figure 5.** Circos plot exhibiting the 95 predicted natural products (adjusted- $P < 0.01$ ) with direct evidence (in red font) or indirect evidence (in blue font). Natural products are grouped by chemical scaffold clustering analysis (Supplementary Fig. S1). The predicted associations from different statistical network models are connected lines in various colors. The full version exhibiting all the 182 predicted natural products (adjusted- $P < 0.01$ ) is provided in Supplementary Fig. S2. The circos plots were drawn using Circos (v0.69)<sup>37</sup>.

**Systems pharmacology-based prediction of potential natural products for cancer immunotherapy.** Through integrating the compound-target network of 66 cancer-related herbs and five sets of CIR gene sets, we further built five statistical network models for discovering novel cancer immunotherapeutic agents from natural products. By applying the threshold of adjusted  $P$ -value ( $q < 0.01$ ), the five *in vitro* models based on different gene sets computationally identified 71–155 natural products as promising candidates, respectively (Supplementary Table S4).

For evaluating the predictive accuracy of the *in vitro* models, we systematically retrieved previous literature evidence from PubMed for the predicted positive natural products in cancer immunotherapy (Supplementary Table S5). Here, the direct evidence refers to a natural product exerting anti-cancer immune responses, while the indirect evidence denotes a natural product enhancing immune response and possessing anti-cancer potential by *in vivo* assays simultaneously. As listed in Supplementary Table S4, all the five models achieve high predictive accuracy with the success rates of over 50%, suggesting the reliability of computational models. The natural products without experimental verification, especially those have indirect evidence, offer new potential immunotherapeutic candidates that deserve to be further experimentally validated.

Due to the similar and satisfactory performance of the five models, we merged their prediction results for further discussion. In total, we obtained 182 natural products for further analysis. Clustering analysis was performed to cluster their chemical scaffolds into 5 groups by calculating the root-mean-square value of the Tanimoto distance based on FCFP<sub>6</sub> fingerprint (Supplementary Fig. S1A). The structures of the five cluster centers, including ginsenoside Rh2, 1-Tetradecanol, chrysophanol, palmitic acid, and apigenin are given in Supplementary Fig. S1B. As presented in Fig. 5 and Supplementary Fig. S2, among the 182 predicted natural products, 59

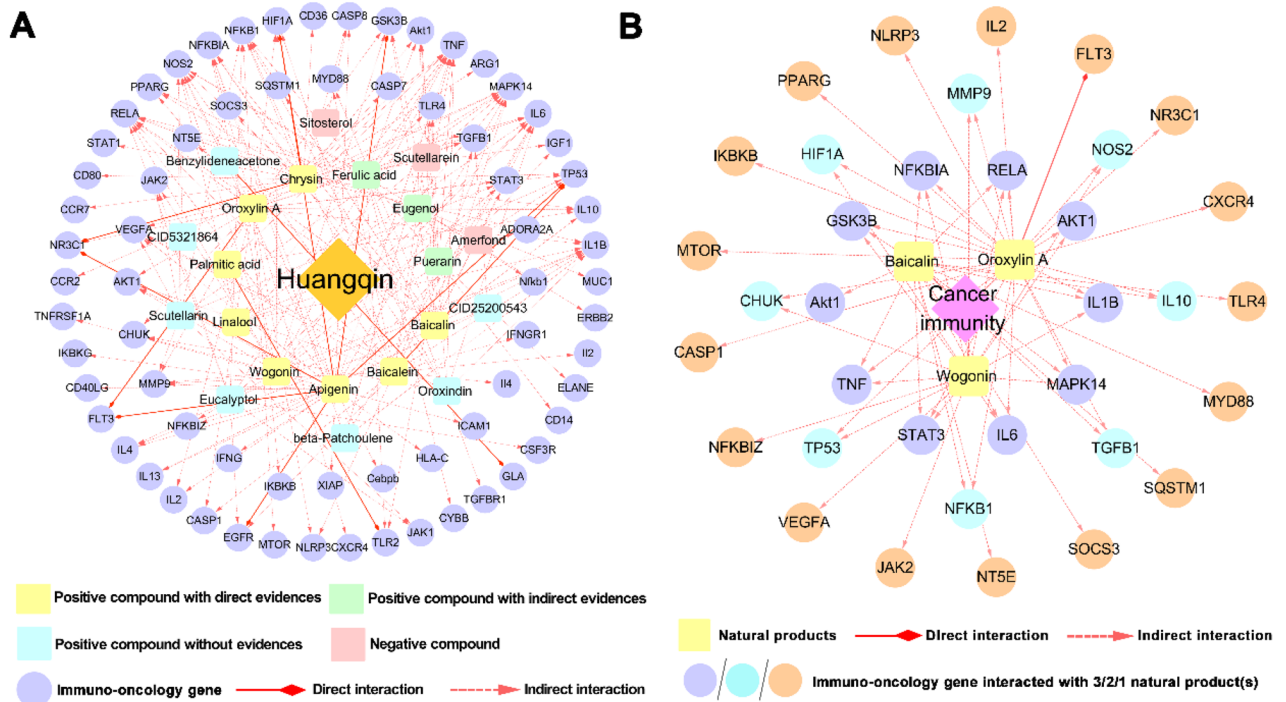


**Figure 6.** Heatmap highlighting the 49 natural products simultaneously predicted as potential cancer immunotherapeutic candidates by all the five statistic network models. The relative metric combined Z-score was calculated for each natural product by averaging the normalized Z-scores predicted by the five models. Natural products validated by direct evidence and indirect evidence are highlighted in red and blue font, respectively. The heatmap was drawn using HemI (v1.0)<sup>38</sup>.

were validated by direct evidence, while 36 were indicated by indirect evidence, with an overall success rate of 52.20% (95/182). Most of the natural products were identified as potential cancer immunotherapeutic agents by more than one statistical network models. We next turned to focus on the 49 natural products that had been simultaneously predicted to be positive by all the five models (Fig. 6). Evidence from literature demonstrates that the success rate is increased to 65.31% (32/49), with 20 and 12 of the 49 natural products possess direct evidence and indirect evidence, respectively. We calculated the combined Z-score to provide a relative metric for comparison by combining the predictions of the five models (see the details in Fig. 6). The top 10 natural products prioritized by combined Z-score are quercetin, luteolin, baicalin, kurarinone, sinomenine, oroxylin A, bergapten, oleanolic acid, tetrandrine, and resveratrol (Supplementary Table S6). A growing body of experimental evidence has indicated their potential anti-cancer effects through cancer immunity such as luteolin<sup>33–36</sup>. Put together, the proposed systems pharmacology-based approach provides promising cancer immunotherapeutic candidates from natural products which deserve further preclinical validation.

**Case study: mechanism-of-action of anti-tumor immunotherapy by *Huangqin* (*Scutellaria baicalensis georgii*).** We systematically analyzed the source of the 182 predicted positive natural products (Supplementary Fig. S3A). We found that there were 16 herbs possessing a positive ingredients proportion higher than 10% (Supplementary Table S7). Among them, *Huangqin* (*Scutellaria baicalensis*) has the highest number ( $n = 18$ ) of potential cancer immunotherapy ingredients (Supplementary Fig. S3B). We thus selected *Huangqin* as a case study to showcase its anti-tumor immunotherapeutic mechanisms.

*Huangqin* and its monomer extracts had been widely validated by clinical trials and experiments for the treatment of various cancers<sup>39</sup>. For example, Skullcapflavone I, a flavone compound extracted from *Huangqin*, has been demonstrated to exert anti-cancer effect on lung cancer cells through inactivating PI3K/AKT/mTOR signaling pathway<sup>40</sup>. Although massive anti-tumor studies of *Huangqin* have emerged<sup>39</sup>, there are few researches



**Figure 7.** Compound-target (C-T) network of *Huangqin* (*Scutellaria baicalensis* Georgi) in cancer immunotherapy. **(A)** C-T network for 204 interactions connecting 21 compounds of *Huangqin* and 66 cancer immune response-related (CIR) genes. **(B)** C-T network for 60 compound-target interactions connecting three main bioactive flavones of *Huangqin* (baicalin, wogonin and oroxylin A) and 34 CIR genes.

focusing on its immunotherapeutic properties. Thus, in this part, we aim to systematically elucidate the anti-cancer mechanism of *Huangqin* and highlight its main active constituents in cancer immunotherapy (Fig. 7).

As illustrated in Fig. 7A, the cancer immunity compound-target network of *Huangqin* consists of 204 interactions connecting 21 natural products and 66 CIR genes. Among the 21 natural products, 18 out of them were predicted as positive, suggesting their high correlation with cancer immunity. Eight of them were validated by direct literature evidence, while three were supported by indirect evidence. For instance, linalool has been reported to exhibit cytotoxic effects by activating anti-tumor immunity<sup>41</sup>. Among the eight predicted natural products with direct evidence, baicalin, wogonin, and oroxylin A are the main bioactive flavones of *Huangqin* detected by ultra-high-performance liquid chromatography (UHPLC) method<sup>42</sup>. *In vivo* studies of *Huangqin* against multiple tumors also revealed that predominantly conjugates and aglycones of them could be found in plasma<sup>39</sup>, suggesting their important role in oncotherapy. Therefore, we selected these three typical natural products (baicalin, wogonin, and oroxylin A) to develop a specific anti-cancer immune response-related compound-target network for further investigating the underlying cancer immunotherapeutic mechanism of *Huangqin*.

As displayed in Fig. 7B, the specific network of the 3 natural products is composed of 60 CTIs. Interestingly, we found that baicalin, wogonin, and oroxylin A shared multiple overlapped CIR genes, including RELA, IL1B, MAPK14, IL6, STAT3, TNF, GSK3B, and NFKBIA, which indicates that *Huangqin* may exert cancer immunotherapeutic effect through the synergistic action of these components. These overlapped targets had been validated to play significant roles in the therapeutic action of *Huangqin* against many types of cancer<sup>43–45</sup>. Overall, the systems pharmacology approach applied here uncovers the main active constituents and key protein targets of *Huangqin* in cancer immunotherapy.

## Discussion

Cancer immunotherapy, which was proven to elicit clinically meaningful and durable responses in multiple tumor types, has led to a new era of cancer treatment. In this study, we proposed a novel integrated quantitative and systems pharmacology infrastructure to identify potential agents of natural products for cancer immunotherapy and decipher the underlying molecular mechanisms. This work explores the following new cancer immunotherapy characteristics of TCM: (i) We identified 66 significant cancer-related herbs through a large-scale literature mining of PubMed database; (ii) A cancer immune response-related H-C-T network was constructed by integrating experimentally reported targets and multiple CIR gene sets. MOAs of various immunotherapeutic effects of TCM-derived natural products are investigated through current network; (iii) *In vitro* models yielding high accuracy are developed to prioritize novel anti-cancer natural products in cancer immunotherapy.

Network analysis showcased the great potential of TCM-derived natural products in immuno-oncology and highlighted the common targets related to tumor immunoregulation. Among the 238 targets in H-C-T network, RELA ( $D=93$ ) and TNF- $\alpha$  ( $D=92$ ) owned the largest connections to natural products. Indeed, emerging evidence has suggested their anti-tumor immunoregulation effect. For example, RELA can drive tumor-associated



macrophages to suppress CD8+ cytotoxic T lymphocytes for tumor promotion, indicating a RELA-targeted immunotherapy for lung cancer<sup>29</sup>. Besides, TNF- $\alpha$ , as the prototype of the TNF superfamily cytokines, has been validated to enhance the differentiation of Th9 cell and exert anti-tumor immunity effect via TNFR2-dependent pathways<sup>30</sup>. Moreover, the *in vitro* models have identified 49 novel anti-cancer natural products that are highly related to cancer immunotherapy. For example, luteolin, a flavonoid ranking second among the 49 predicted natural products (Fig. 6), has not yet been reported for its cancer immunotherapeutic action. Previous *in vivo* study confirmed that luteolin exerted anti-cancer effects in multiple cancer types<sup>33–35</sup>. In addition, a recent study showed that luteolin enhanced immune cell functions, including proliferation, cytotoxic T lymphocyte (CTL) activity and natural killer (NK) cell activity in isolated murine splenocytes<sup>36</sup>, indicating its potential in cancer immunotherapy.

Based on the global H-C-T network and the *in vitro* predictive models, current quantitative and systems pharmacology framework can also be applied to extend the cancer immunotherapy knowledge of single herb. We highlighted three main ingredients of *Huangqin* (baicalin, wogonin, and oroxylin A) that might be able to regulate tumor immunosuppressive microenvironments. Subnetwork analysis indicated their overlapped CIR genes may exert synergistic action against multiple cancer. Taking STAT3 as an example, experimental research had demonstrated that both baicalin and wogonin could inhibit tumor growth via acting on STAT3<sup>43,44</sup>. A recent study revealed that baicalin decreased STAT3 activity and further downregulated IFN- $\gamma$ -induced PD-L1 expression to promote anti-tumor immunity<sup>43</sup>. Furthermore, wogonin was also reported to inhibit tumor growth and enhance immune system activation by suppressing STAT3 signaling<sup>44</sup>. In addition, previous study showed that oroxylin A suppressed the generation of Tregs in lung cancer environment through modulating NF- $\kappa$ B signaling<sup>45</sup>. These experimental results support our *in vitro* findings to some extent.

Several limitations of this work should be recognized. First, although we have integrated a wide range of compound-protein interactions from published literature and publicly available databases, the potential incompleteness of current networks may be inevitable. Recent studies show that the importing of computationally predicted interactions inferred by balanced substructure–drug–target network-based inference may help to improve the performance of current *in vitro* models<sup>14,46</sup>. Second, current approach can only predict the potential immunotherapeutic effect of natural products targeting known CIR gene-encoded proteins. Integrating systems biology resources can facilitate the identification of the growing potential CIR protein targets by indirectly locating their neighbours in the human protein–protein interaction network, gene regulatory network, or biological pathways<sup>19</sup>. Third, a gap may exist between the activity of a specific molecule observed in a model system and the true effect in a living organism. For example, the absorption, distribution, metabolism and excretion (ADME) properties that reflect the pharmacokinetics and disposition of chemicals, are critical related to the real efficacy of chemicals *in vivo* situation<sup>47</sup>. Hence, a tougher standard that integrated multi-dimensional analysis, such as ADME prediction, drug-likeness assessment, or pharmacokinetics and pharmacodynamics (PK/PD) analysis, should be applied for further narrowing the gap. Finally, although recent experimental study had validated some of the predicted positive natural products preliminarily, further *in vivo* experiments are necessary to validate the predicted cancer immunotherapeutic effects and molecular mechanism of natural products before advancement to translational studies or clinical trials.

## Conclusion

In summary, our systems pharmacology-based approaches proposed in this study show promise in the *in vitro* identification of potential anti-cancer natural products acting on immune microenvironment. This network-based approach has provided useful methodology for lead identification purposes of tumor immunotherapy. In combination with in-depth experimental validation, we believe that this framework could serve as a valuable and complementary workflow to shorten the time and improve efficiency for identifying promising anti-tumor immunotherapy candidates, and could be applied in other complex diseases.

## Materials and methods

**Large-scale literature mining of cancer-related Chinese herbs.** We performed a comprehensive literature mining of PubMed database to extract Chinese medicinal herbs that highly correlated with cancer. We initially collected 525 clinically used herbs included in *Pharmacopoeia of the People's Republic of China 2015* and annotated their English scientific terms according to the *Flora Reipublicae Popluaris Sinicae (FRPS)*. To balance the bias of different literature number on various herbs, a variable R was imported to evaluate the correlation between herbs and cancer by calculating the ratio of (cancer-herb-related articles)/(herb-related articles). We further applied *P*-value to statistically analyze the chance probability of co-occurrences of each herb and cancer in at least *k* papers, as Eq. (1) described:

$$P = 1 - \sum_{i=0}^{k-1} f(i) = 1 - \sum_{i=0}^{k-1} \frac{\binom{K}{i} \binom{N-K}{n-i}}{\binom{N}{n}}, \quad (1)$$

where *N* is the total number of papers in PubMed (*N* = 29.0 million, accessed in Jun 10, 2019), *K* stands for the number of literature related to cancer (*K* = 3,874,763), *n* and *k* represent the amounts of papers for a specific herb and its corresponding studies on cancer, respectively. Subsequently, the nominal *P*-values were corrected as adjusted *P*-values (*q*) using R based on the Benjamini–Hochberg approach<sup>48</sup>. In this study, herbs with *q* value lower than 0.01 as well as R values greater than 0.05 are considered significantly correlated with cancer.

**Manual curation and integration of multiple cancer immune response-related (CIR) gene sets.** We collected CIR gene sets from multiple resources: (i) Immuno-oncology targets (IOs) in the current global IO pipeline<sup>49</sup>. We removed targets that possess less than 3 agents, resulting in 138 IO targets; (ii) Cancer immunomodulators (IMs) gene set obtained from a literature review, which consists of 78 genes that had been reviewed and confirmed by tumor immunologists from The Cancer Genome Atlas (TCGA) immune response working group<sup>50</sup>. Since recent accumulating evidence have showed innate immune genes play crucial roles in cancer immunology<sup>51,52</sup>, we further curated: (iii) cancer-related innate immune gene set (INs) containing 226 genes identified by various experimental assays (Supplementary Table S8). All the genes/targets were converted into unified Gene Entrez ID and duplicated data in each gene set were eliminated. Fisher's test method showed that there are significant correlations ( $p < 10^{-5}$ ) among the three gene sets at a biological level (Supplementary Fig. S4).

Furthermore, we integrated two additional gene sets based on current three CIR gene sets. The first integrated gene set (IG-1) contains 57 genes that are included in at least 2 out of 3 CIR gene sets, while the second integrated gene set (IG-2) containing 381 genes is the union set of all the three CIR gene sets.

**Construction of herb-compound-target (H-C-T) networks for natural products.** We built a comprehensive herb-compound network database by integrating data from multiple TCM databases<sup>21</sup>. The constituent compounds of each Chinese herb were manually extracted from six publicly available TCM data sources, including Traditional Chinese Medicine integrated database (TCMID)<sup>53</sup>, Traditional Chinese Medicine Systems Pharmacology (TCMSP)<sup>54</sup>, Traditional Chinese Medicine database (TCMDB)<sup>55</sup>, Traditional Chinese Medicine database@Taiwan (TCM@Taiwan)<sup>56</sup>, TM-MC<sup>57</sup>, and TCM-MESH<sup>58</sup>. All compounds were transformed to canonical SMILES format and compounds with identical structures were merged.

We further constructed a H-C-T database of natural products through combining physical binding (direct) and functional (indirect) targets of natural products from multiple data sources. The physical binding targets were collected from ChEMBL (v21)<sup>59</sup> and BindingDB (accessed in September 2017) (accessed in September 2017) databases, while functional targets were obtained from literature and three databases: STITCH 5 (accessed in September 2017)<sup>60</sup>, TCMID 2.0<sup>61</sup>, and Herbal Ingredients' Targets Database (HIT)<sup>62</sup>. In total, more than 2,000 natural product-specific pharmacological articles (dating from January 2009 to December 2017)<sup>23</sup> were evaluated. For the direct targets, only interactions with binding affinity data [inhibitory constant (Ki), dissociation constant (Kd), half maximal inhibitory concentration (IC<sub>50</sub>) or half maximal effective concentration (EC<sub>50</sub>)] lower than 10  $\mu\text{m}$  were retained<sup>16</sup>. After removing targets without standard UniProt accession number, 38,220 unique interactions connecting 3882 natural products and 5643 human proteins were finally obtained. The detailed descriptions about the CTIs curation are provided in our previous works<sup>16,23</sup>.

**Network-based statistical models.** In this study, we aimed to identify potential immunotherapeutic natural products from cancer-related TCM herbs. Five integrated statistical models were built independently through incorporating compound-target (C-T) networks into the five sets of curated CIR gene sets. It is plausible to hypothesize that a natural product exhibits a high possibility to possess cancer immunotherapeutic effect if its targets are more likely to be CIR gene-encoded proteins. The null hypothesis asserts that the targets of a natural product randomly located at CIR-encoded proteins across the human proteome. We firstly eliminated natural products without any CIR-encoded protein according to the five gene sets. Ions and organic solvents were also removed. Subsequently, permutation test as Eq. (2) given was performed to calculate the statistical significance of a natural product to be prioritized for potential association with cancer immunotherapy:

$$P = \frac{\#\{Sm(p) > Sm\}}{\#\{\text{total permutations}\}} \quad (2)$$

A nominal  $P$  was computed for each natural product by counting the amount of the permutations ( $Sm(p)$ ) larger than observed CIR genes ( $Sm$ ). Here we randomly extracted  $x$  genes ( $x$  refers to the number of genes in each CIR gene set) from protein products at the genome-wide scale that includes 20,462 human protein-coding genes from the National Center for Biotechnology Information database<sup>63</sup> for 100,000 times. Natural products with adjusted  $P$ -values ( $q$ , corrected by Benjamini-Hochberg approach)<sup>48</sup> lower than 0.01 were regarded as significantly related to cancer immunotherapy. Besides,  $Z$ -score was further calculated by Eq. (3) to quantify the potential association of each natural product and cancer immunotherapy during permutation testing:

$$Z = \frac{x - \mu}{\sigma} \quad (3)$$

where  $x$  is the actual number of observed CIR genes targeted by a specific natural product,  $\mu$  is the average number of observed CIR genes targeted by a specific natural product during 100,000 permutations, and  $\sigma$  is the standard deviation (SD).

**Network visualization and statistical analysis.** The networks in this study were displayed and analyzed by Gephi (v0.9.2, <https://gephi.org/>)<sup>64</sup> and Cytoscape (v3.2.0, <http://www.cytoscape.org/>)<sup>65</sup>. The gene enrichment analysis was performed by ClueGO plug-in<sup>32</sup> and OmicShare tools (<http://www.omicshare.com/tools>), and the statistical analysis and graphics were performed by R environment (v3.01, <http://www.r-project.org/>) and Python platform (v3.2, <http://www.python.org/>).

## Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Received: 9 January 2020; Accepted: 20 January 2021

Published online: 08 February 2021

## References

- Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**, 394–424 (2018).
- Fritz, J. M. & Lenardo, M. J. Development of immune checkpoint therapy for cancer. *J. Exp. Med.* **216**, 1244–1254 (2019).
- Xin, Yu. J., Hubbard-Lucey, V. M. & Tang, J. Immuno-oncology drug development goes global. *Nat. Rev. Drug Discov.* **18**, 899–900 (2019).
- Sanmamed, M. F. & Chen, L. A paradigm shift in cancer immunotherapy: From enhancement to normalization. *Cell* **175**, 313–326 (2018).
- Velcheti, V. & Schalper, K. Basic overview of current immunotherapy approaches in cancer. *Am. Soc. Clin. Oncol. Educ. Book.* **35**, 298–308 (2016).
- Fang, J. *et al.* Systems pharmacology-based discovery of natural products for precision oncology through targeting cancer mutated genes. *CPT Pharmacometrics Syst. Pharmacol.* **6**, 177–187 (2017).
- Yan, Z., Lai, Z. & Lin, J. Anticancer properties of traditional Chinese medicine. *Comb. Chem. High Throughput Screen.* **20**, 423–429 (2017).
- Casey, S. C. *et al.* Cancer prevention and therapy through the modulation of the tumor microenvironment. *Semin. Cancer Biol.* **35**, S199–S223 (2015).
- Guo, P. *et al.* An insight into the molecular mechanism of berberine towards multiple cancer types through systems pharmacology. *Front. Pharmacol.* **10**, 857 (2019).
- Orzechowska, B. *et al.* Baicalin from the extract of *Scutellaria baicalensis* affects the innate immunity and apoptosis in leukocytes of children with acute lymphocytic leukemia. *Int. Immunopharmacol.* **23**, 558–567 (2014).
- Sun, L., Chen, B., Jiang, R., Li, J. & Wang, B. Resveratrol inhibits lung cancer growth by suppressing M2-like polarization of tumor associated macrophages. *Cell Immunol.* **311**, 86–93 (2017).
- Bahrami, A., Fereidouni, M., Pirro, M., Bianconi, V. & Sahebkar, A. Modulation of regulatory T cells by natural products in cancer. *Cancer Lett.* **459**, 72–85 (2019).
- Liu, Z. *et al.* TCMIO: A comprehensive database of traditional Chinese medicine on immuno-oncology. *Front. Pharmacol.* **11**, 439 (2020).
- Fang, J., Liu, C., Wang, Q., Lin, P. & Cheng, F. In silico polypharmacology of natural products. *Brief Bioinform.* **19**, 1153–1171 (2018).
- Shen, B. A new golden age of natural products drug discovery. *Cell* **163**, 1297–1300 (2015).
- Fang, J. *et al.* Quantitative and systems pharmacology. 1. In silico prediction of drug–target interactions of natural products enables new targeted cancer therapy. *J. Chem. Inf. Model.* **57**, 2657–2671 (2017).
- Vicini, P. & Van der Graaf, P. H. Systems pharmacology for drug discovery and development: Paradigm shift or flash in the pan?. *Clin. Pharmacol. Ther.* **93**, 379–381 (2013).
- Zhang, W., Huai, Y., Miao, Z., Qian, A. & Wang, Y. Systems pharmacology for investigation of the mechanisms of action of traditional Chinese medicine in drug discovery. *Front. Pharmacol.* **10**, 743 (2019).
- Fang, J. *et al.* Quantitative and systems pharmacology 3. Network-based identification of new targets for natural products enables potential uses in aging-associated disorders. *Front. Pharmacol.* **8**, 747 (2017).
- Wu, Q. *et al.* Systems pharmacology-based approach to investigate the mechanisms of Danggui-Shaoyao-san prescription for treatment of Alzheimer's disease. *BMC Complement. Med. Ther.* **20**, 282 (2020).
- Fang, J. *et al.* Quantitative and systems pharmacology 4. Network-based analysis of drug pleiotropy on coronary artery disease. *Eur. J. Med. Chem.* **161**, 192–204 (2019).
- Cheng, F. *et al.* A genome-wide positioning systems network algorithm for in silico drug repurposing. *Nat. Commun.* **10**, 3476 (2019).
- Huang, Y. *et al.* A Systems pharmacology approach uncovers Wogonoside as an angiogenesis inhibitor of triple-negative breast cancer by targeting Hedgehog signaling. *Cell Chem. Biol.* **26**, 1143–1158.e6 (2019).
- Chei, S. *et al.* Magnolol suppresses TGF- $\beta$ -induced epithelial-to-mesenchymal transition in human colorectal cancer cells. *Front. Oncol.* **9**, 752 (2019).
- Li, M. *et al.* Magnolol inhibits growth of gallbladder cancer cells through the p53 pathway. *Cancer Sci.* **106**, 1341–1350 (2015).
- Ranaware, A. M. *et al.* Magnolol: A neolignan from the magnolia family for the prevention and treatment of cancer. *Int. J. Mol. Sci.* **19**, E2362 (2018).
- Chen, Y., Huang, K., Ding, X., Tang, H. & Xu, Z. Magnolol inhibits growth and induces apoptosis in esophagus cancer KYSE-150 cell lines via the MAP kinase pathway. *J. Thorac. Dis.* **11**, 3030–3038 (2019).
- Yu, C. S. *et al.* Quercetin inhibited murine leukemia WEHI-3 cells in vivo and promoted immune response. *Phytother. Res.* **24**, 163–168 (2010).
- Li, L. *et al.* NF- $\kappa$ B RelA renders tumor-associated macrophages resistant to and capable of directly suppressing CD8 T cells for tumor promotion. *Oncimmunology* **7**, e1435250 (2018).
- Jiang, Y. *et al.* TNF- $\alpha$  enhances Th9 cell differentiation and antitumor immunity via TNFR2-dependent pathways. *J. Immunother. Cancer.* **7**, 28 (2019).
- Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
- Bindea, G. *et al.* ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091–1093 (2009).
- You, Y., Wang, R., Shao, N., Zhi, F. & Yang, Y. Luteolin suppresses tumor proliferation through inducing apoptosis and autophagy via MAPK activation in glioma. *Onco. Targets Ther.* **12**, 2383–2396 (2019).
- Yao, Y., Rao, C., Zheng, G. & Wang, S. Luteolin suppresses colorectal cancer cell metastasis via regulation of the miR-384/pleiotrophin axis. *Oncol. Rep.* **42**, 131–141 (2019).
- Yao, X., Jiang, W., Yu, D. & Yan, Z. Luteolin inhibits proliferation and induces apoptosis of human melanoma cells in vivo and in vitro by suppressing MMP-2 and MMP-9 through the PI3K/AKT pathway. *Food Funct.* **10**, 703–712 (2019).
- Kilani-Jaziri, S. *et al.* Flavones induce immunomodulatory and anti-inflammatory effects by activating cellular anti-oxidant activity: A structure-activity relationship study. *Tumour Biol.* **37**, 6571–6579 (2016).
- Krzywinski, M. *et al.* Circos: An information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
- Deng, W., Wang, Y., Liu, Z., Cheng, H. & Xue, Y. HemI: A toolkit for illustrating heatmaps. *PLoS ONE* **9**, e111988 (2014).

39. Cheng, C. S. *et al.* *Scutellaria baicalensis* and cancer treatment: Recent progress and perspectives in biomedical and clinical studies. *Am. J. Chin. Med.* **46**, 25–54 (2018).
40. Yang, Y. *et al.* Skullcapflavone I suppresses proliferation of human lung cancer cells via down-regulating microRNA-21. *Exp. Mol. Pathol.* **110**, 104285 (2019).
41. Chang, M. Y. & Shen, Y. L. Linalool exhibits cytotoxic effects by activating antitumor immunity. *Molecules* **19**, 6694–6706 (2014).
42. Cui, X. *et al.* Simultaneous determination of 10 flavonoids in crude and wine-processed radix scutellariae by UHPLC. *J. Chromatogr. Sci.* **54**, 312–317 (2016).
43. Ke, M. *et al.* Baicalein and baicalin promote antitumor immunity by suppressing PD-L1 expression in hepatocellular carcinoma cells. *Int. Immunopharmacol.* **75**, 105824 (2019).
44. Xiao, W. *et al.* Wogonin inhibits tumor-derived regulatory molecules by suppressing STAT3 signaling to promote tumor immunity. *J. Immunother.* **38**, 167–184 (2015).
45. Shen, L. *et al.* Oroxylin A inhibits the generation of Tregs in non-small cell lung cancer. *Oncotarget* **8**, 49395–49408 (2017).
46. Wu, Z. *et al.* In silico prediction of chemical mechanism of action via an improved network-based inference method. *Br. J. Pharmacol.* **173**, 3372–3385 (2016).
47. Yang, H. *et al.* admetSAR 2.0: Web-service for prediction and optimization of chemical ADMET properties. *Bioinformatics* **35**, 1067–1069 (2019).
48. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300 (1995).
49. Tang, J., Pearce, L., O'Donnell-Tormey, J. & Hubbard-Lucey, V. M. Trends in the global immuno-oncology landscape. *Nat. Rev. Drug Discov.* **17**, 922 (2018).
50. Thorsson, V. *et al.* The immune landscape of cancer. *Immunity* **48**, 812–830.e14 (2018).
51. Berraondo, P. *et al.* Innate immune mediators in cancer: Between defense and resistance. *Immunol. Rev.* **274**, 290–306 (2016).
52. Vesely, M. D., Kershaw, M. H., Schreiber, R. D. & Smyth, M. J. Natural innate and adaptive immunity to cancer. *Annu. Rev. Immunol.* **29**, 235–271 (2011).
53. Xue, R. *et al.* TCMID: Traditional Chinese Medicine integrative database for herb molecular mechanism analysis. *Nucleic Acids Res.* **41**, D1089–D1095 (2013).
54. Ru, J. *et al.* TCMSP: A database of systems pharmacology for drug discovery from herbal medicines. *J. Cheminform.* **6**, 13 (2014).
55. He, M., Yan, X., Zhou, J. & Xie, G. Traditional Chinese medicine database and application on the Web. *J. Chem. Inf. Comput. Sci.* **41**, 273–277 (2001).
56. Chen, C. Y. TCM Database@Taiwan: The world's largest traditional Chinese medicine database for drug screening in silico. *PLoS ONE* **6**, e15939 (2011).
57. Kim, S. K., Nam, S., Jang, H., Kim, A. & Lee, J. J. TM-MC: A database of medicinal materials and chemical compounds in Northeast Asian traditional medicine. *BMC Complement. Altern. Med.* **15**, 218 (2015).
58. Zhang, R. Z., Yu, S. J., Bai, H. & Ning, K. TCM-Mesh: The database and analytical system for network pharmacology analysis for TCM preparations. *Sci. Rep.* **7**, 2821 (2017).
59. Gaulton, A. *et al.* The ChEMBL database in 2017. *Nucleic Acids Res.* **45**, D945–D954 (2017).
60. Szklarczyk, D. *et al.* STITCH 5: Augmenting protein–chemical interaction networks with tissue and affinity data. *Nucleic Acids Res.* **44**, D380–D384 (2016).
61. Huang, L. *et al.* TCMID 2.0: A comprehensive resource for TCM. *Nucleic Acids Res.* **46**, D1117–D1120 (2018).
62. Ye, H. *et al.* HIT: Linking herbal active ingredients to targets. *Nucleic Acids Res.* **39**, D1055 (2011).
63. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **46**, D8–D13 (2018).
64. Bastian, M., Heymann, S. & Jacomy, M. Gephi: An open source software for exploring and manipulating networks. In *Proceedings of the Third International Conference on Weblogs and Social Media, ICWSM 2009, San Jose, California, USA* (2009).
65. Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

## Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 82074278, 81903912, and 81902693), Natural Science Foundation of Guangdong Province (2018A030310298), China Postdoctoral Science Foundation funded project (2019M662878), and Key laboratory project of colleges and universities in Guangdong province (No. 2019KSYS005).

## Author contributions

J.S.F. conceived and designed the experiments. C.P.C. and Q.H.W. conducted the experiments and wrote the manuscript. H.H.H., L.Y.H., Z.H.L. and S.J.Z. collected and processed the data. Y.G. and Q.W. performed data analysis. H.H.H., X.D.F. and J.S.F. revised the manuscript. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-82857-2>.

**Correspondence** and requests for materials should be addressed to X.F. or J.F.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.





**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021