



## Database tool

# HerDing: herb recommendation system to treat diseases using genes and chemicals

Wonjun Choi<sup>1</sup>, Chan-Hun Choi<sup>2</sup>, Young Ran Kim<sup>3</sup>, Seon-Jong Kim<sup>2</sup>,  
Chang-Su Na<sup>2</sup> and Hyunju Lee<sup>1,\*</sup>

<sup>1</sup>School of Information and Communications, Gwangju Institute of Science and Technology, 123 Cheomdangwagi-ro, Buk-gu, Gwangju 61005, Republic of Korea, <sup>2</sup>College of Korean Medicine, Dongshin University, 185 Geonjae-ro, Naju-si, Jeollanam-do 58245, Republic of Korea and, <sup>3</sup>College of Pharmacy, Chonnam National University, 77 Yongbong-ro, Buk-gu, Gwangju 61186, Republic of Korea

\*Corresponding author: Tel: +82 62-715-2213; Fax: +82 62-715-2204; Email: hyunjulee@gist.ac.kr

Citation details: Choi, W.J., Choi, C.H., Kim, Y.R. *et al.* HerDing: herb recommendation system to treat diseases using genes and chemicals. *Database* (2016) Vol. 2016: article ID baw011; doi:10.1093/database/baw011

Received 11 September 2015; Revised 1 December 2015; Accepted 27 January 2016

## Abstract

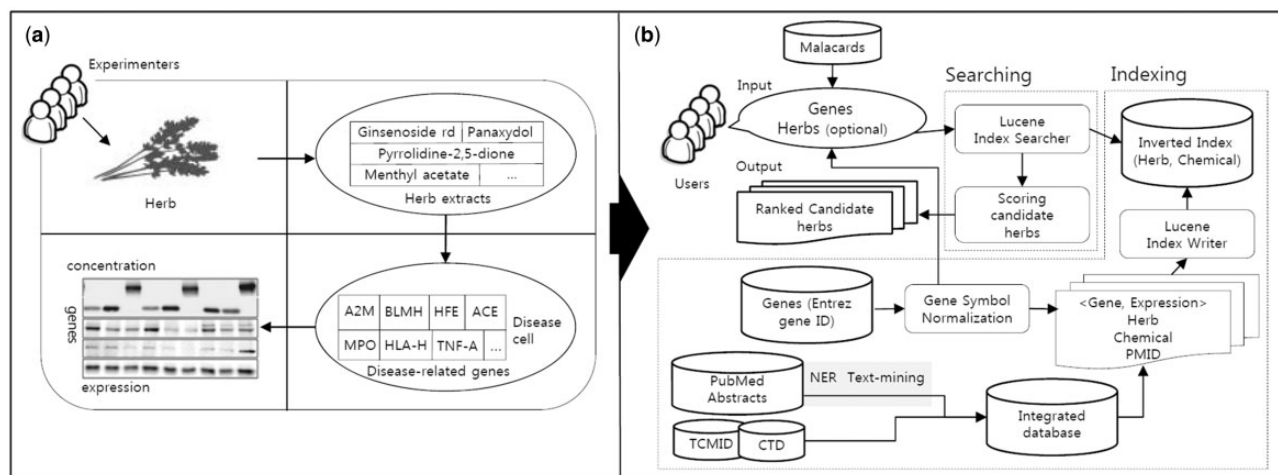
In recent years, herbs have been researched for new drug candidates because they have a long empirical history of treating diseases and are relatively free from side effects. Studies to scientifically prove the medical efficacy of herbs for target diseases often spend a considerable amount of time and effort in choosing candidate herbs and in performing experiments to measure changes of marker genes when treating herbs. A computational approach to recommend herbs for treating diseases might be helpful to promote efficiency in the early stage of such studies. Although several databases related to traditional Chinese medicine have been already developed, there is no specialized Web tool yet recommending herbs to treat diseases based on disease-related genes. Therefore, we developed a novel search engine, HerDing, focused on retrieving candidate herb-related information with user search terms (a list of genes, a disease name, a chemical name or an herb name). HerDing was built by integrating public databases and by applying a text-mining method. The HerDing website is free and open to all users, and there is no login requirement.

**Database URL:** <http://combio.gist.ac.kr/herding>

## Introduction

Natural products such as herbs have drawn increasing attention in finding new drug candidates in recent years because herbs have been practiced for treating diseases for thousands of years and might reduce side effects. With the increasing prevalence of chronic diseases such as cancer,

diabetes, coronary heart disease and hypertension, natural product-based medicines are being popularized and have been used globally. Hence, many studies have been conducted to assess the effectiveness of natural products in treating diseases. For validation of the effects of herbs, experiments such as reverse transcriptase-polymerase chain



**Figure 1.** (a) An experiment to assess the effectiveness of herbs based on the changes of marker genes in a disease cell. (b) The process of the HerDing system consists of searching and indexing parts.

reaction analysis and western blots have been widely used to measure the expression levels of disease-related genes or proteins in a disease cell (Figure 1a). For example, Wang *et al.* (1) assessed the therapeutic efficacy of the Realgar-Indigo naturalis formula in treating human acute promyelocytic leukemia (APL). For the assessment, the expression levels of marker genes of APL, including CD11b/CD14/CDK2, were measured in an APL-induced cell treated by the Realgar-Indigo naturalis formula.

Furthermore, studies for developing databases specializing in natural products have been actively conducted. Traditional Chinese Medicine Integrated Database (TCMID) (2) is one of the largest databases of oriental Chinese medicine, which records TCM-related information collected through a text-mining approach and by integrating different public databases including TCM-ID (3), HIT (4), TCM@Taiwan (5), STITCH (6), OMIM (7) and DrugBank (8). TCMID currently contains 8159 herbs, 25 210 compounds, 6828 drugs, 3721 diseases and 17 521 targets, and provides networks of relationships between various components such as herb-disease, herbal ingredients-targets and herbal-ingredient-target-disease-drug networks. TCM-ID (3) provides general TCM-related data including 3D structures of TCM compounds, which are manually collected from TCM-related books and other printed sources in Medline. TCM-ID currently contains 1313 herbs, 5669 herbal ingredients, 1588 prescriptions and the 3D structure of 3725 herbal ingredients. Also, TCM Database@Taiwan (5) is the world's largest traditional Chinese medicine database specialized for *in silico* drug identification. It has collected >20 000 pure compounds isolated from 453 TCM herbs from Chinese medical texts and scientific publications.

In spite of the increasing interest in TCM-related databases, current Web applications contain a relatively small number of herbs compared to NCBI taxonomy (9), which

contains >151 250 herbs in English name. This is partly because relationships between herbs and diseases were extracted only when they were directly mentioned in research articles. However, relationships between herbs and diseases can be expanded if disease-gene relationships are incorporated with herb information. In this study, we assume that if there is a resource that provides an association between marker genes of diseases and herbs, it would be helpful in recommending natural products for diseases. However, relationships between genes and herbs are not well understood. Hence, we used Swanson's ABC model (10) to infer the influence of natural products on disease-related genes; a gene is targeted by a chemical, and a chemical is contained in a natural product. By bridging each relationship, we defined indirect relationships between herbs and genes (or diseases) and they were stored in our database. Then, we constructed a novel search engine, HerDing, which is a recommendation system to treat diseases based on the precomposed database and simulates the experimental process for measuring the efficacy of herbs in diseases (Figure 1a) with a computational approach (Figure 1b). The objective of HerDing is to assist researchers to reduce their time and experimental costs for proving herbs' efficacy in treating diseases

## Materials and Methods

### Resources

HerDing integrated herb-chemical, chemical-gene and gene-disease relationships, which were obtained from other databases and text-mining of published articles, to find herb-disease relationships.

Among them, herb-chemical relationships were extracted by combining TCMID databases and a text-mining

approach. We extracted herb-chemical relationships from TCMID, which provides 8159 herbs with 25 210 chemicals (as of March 2015). However, the number of herbs from TCMID is small compared to 151250 herbs in NCBI taxonomy. To enrich the relationships, we applied a rule-based text mining model to infer herb-chemical relationships (11). The text mining model consists of the following five rules: a verbal trigger rule, a prepositional trigger rule, a relative trigger rule, an apposition trigger rule and a copula trigger rule. An F-measure of the rule-based model was 0.749 when it was tested with 204 gold standard sentences. The rule-based model was applied to PubMed abstracts. For this task, herb names and chemical names were first identified from abstracts. For herb names, an herb name dictionary including English names, Chinese names and Latin names was constructed using TCMID and NCBI taxonomy, and then LingPipe (12), a dictionary-based exact matching NER tool, was used to locate herb names in abstracts. Chemical names were annotated using ChemSpot (13), which is a named entity recognition tool for locating chemical names including trivial names, drugs, abbreviations and molecular formulas in texts. When the rule-based model was applied to 13408 621 PubMed abstracts, a total of 101 550 herb-chemical relationships were extracted.

To construct the chemical-gene relationships, we integrated two public databases: The Comparative Toxicogenomics Database (CTD) (14) and TCMID. CTD provides detailed information about manually curated 1166896 chemical-gene interactions between 11 406 unique chemicals and 38433 genes in 540 organisms, 197288 chemical-disease relationships and 33814 gene-disease relationships. Because chemicals are presented using MeSH identifiers and CAS identifiers in CTD, whereas TCMID uses STITCH and PubChem (15) identifiers, we again used the ChemSpot NER tool to normalize chemical names. In addition, we normalized gene names to Entrez Gene identifiers using an identifier mapping tool provided by UniProt (16). Through this process, we accumulated 1240308 chemical-gene relationships from TCMID and CTD. Note that CTD provides directions of gene expression changes (i.e. up-regulated and down-regulated genes) in chemical-gene relationships.

Finally, we constructed herb-gene relationships by associating herb-chemical relationships and chemical-gene relationships using Swanson's ABC model. Then, with the input of genes related to the disease, herb-disease relationships can be extracted. In summary, HerDing currently indexes 19 476 herb names, 16 762 gene names, 6655 chemical names and 11 394 diseases and provides 4880575 herb-gene relationships (as of November 2015).

## Indexing and searching

The indexing and searching process of the HerDing system is shown in Figure 1b. Indexing is the process of assigning a numerical value to a database entity to enable the quick search of queries in the search engine, and it is useful when there is a large number of entities. Lucene (17) is a high-performance text search engine library written in Java, and it provides a full text indexing and searching technology, supporting high access speed for multi-user accesses. Thus, we used Lucene to index database entities, which facilitated fast search results for users in HerDing. Through the indexing process, we constructed an inverted index file that indicates the location of relationships among herbs, chemicals and genes in the HerDing system. Thus, it allows rapid searches with input queries.

In the searching process, when the input is given by users, the Lucene index searcher accesses the inverted index file system to search relationships among herbs, chemicals and genes. In other words, herbs and chemicals with relationships with input genes will be collected by the Lucene index searcher. The scoring step is to measure the relevance of the candidate herbs, and the score of the herb was determined by the number of input genes related to a specific herb.

## Results

Users of HerDing can search herbs related to a disease with an input of disease-related genes or a disease name. In addition, users can query a chemical name for retrieving herbs and genes related to the input chemical and herb names for extracting chemicals and genes related to the input herb

### Example queries

The webpage of the HerDing system provides the following examples:

- *TNF, IL1B, IL6, IL8, IL17A, MMP1, MMP3, MMP9, MMP12, MMP13, COX2, IL10 and IL4*: These genes are known as marker genes for rheumatoid arthritis. When they were input to HerDing, 30 (default) herbs were recommended. Among them, top-ranked herbs such as *Radix Cudraniae* (Chuan Po Shi) and *Polygonum cuspidatum* (Hu Zhang) were previously reported to be effective in treating rheumatoid arthritis (18–20).
- *Down-regulation (TNF, IL1B, IL6, IL8, IL17A, MMP1, MMP3, MMP9, MMP12, MMP13, COX2) and up-regulation (IL10, IL4)*: Down-regulation of the former set of genes and up-regulation of the latter set of genes are known to be a therapeutic strategy for rheumatoid arthritis. When they were input to HerDing, 30 (default)

herbs were recommended. Among them, top-ranked herbs such as *Polygonum multiflorum* (He Shou Wu) and *Broussonetia papyrifera* (Gou Shu) were previously reported to reduce inflammation, which may be effective in treating rheumatoid arthritis (21, 22).

- **Hypertension:** When hypertension was given as an input from a user, 50 genes related to hypertension, which were extracted from MalaCards (23), were automatically replaced as the input to the HerDing system. As a result, 30 (default) herbs, including *Holothuria leucospilota* (Hai Shen), *Lycium chinense* (Gou Qi Zi), *Urtica dioica* (Yi Zhu Qian Ma) and soybean, were recommended. Among them, *H. leucospilota* and *L. chinense* were first- and third-ranked because chemicals contained in *H. leucospilota* and *L. chinense* target 22 and 18 genes out of the 50 input genes, respectively. Note that several studies (24–27) reported the effects of *H. leucospilota*, *L. chinense*, *U. dioica* and soybean on hypertension.
- **Saponins:** When a chemical name, saponins, was input to HerDing, 517 herbs that contain saponins and genes targeted by saponins were retrieved.
- **Dried tangerine peel:** When an herb, dried tangerine peel, was given as an input, 34 chemicals contained in the input herb and genes targeted by those chemicals were recommended.

### Web interface of the HerDing systems

We implemented HerDing's web interface using the Apache Tomcat, an open-source software implementation of the Java Servlet and Java Server Pages (JSPs).

- **Input:** Figure 2 shows the screen shots of web pages of HerDing and query interfaces for users, which contain three sections. In Figure 2a, HerDing's query consists of disease-related genes. Genes can be specified in four different ways: (i) a list of up-regulated genes when an herb is treated; (ii) a list of down-regulated genes when an herb is treated; (iii) a list of genes (regardless of the direction of expression changes) and (iv) a single disease name or a MeSH identifier. (A list of genes will automatically be provided from the MalaCards (23) database when an input disease name or a MeSH ID is entered.) HerDing also provides two advanced options. When a list of herbs is provided as an input (optional), the search space of herbs is limited to the input herbs. The maximum number of herbs to be retrieved can be selected, on which the search speed depends. In Figure 2b, users can input a chemical name to retrieve herbs that contain the given chemical and genes targeted by the chemical. In Figure 2c, users can input an herb name to obtain chemicals

that are contained in the given herb and genes targeted by the chemicals.

- **Output:** Figure 3 shows a summary of results for a given user input (i.e. Alzheimer's disease). The results page consists of three components including input information, a summary of retrieved herbs from input genes, and detailed information. From MalaCards, 50 genes related to Alzheimer's disease are automatically retrieved, and they are used as inputs to HerDing. Then, the recommended herbs from HerDing are displayed in the order of the number of genes targeted by chemicals contained in the retrieved herbs. For example, in the second table of Figure 3, HAI SHEN is the top-ranked herb, which targets 25 Alzheimer-related genes. By clicking the 'chemicals' button, users can check a list of chemicals targeting the gene.

Furthermore, detailed information about the herbs is described in the third table of Figure 3. By clicking the 'show all' button, users can closely check detailed information such as active ingredients including the formula and 3D structure, the direction of expression changes of the targeted genes by chemicals, gene' organisms and PMIDs that support chemical-gene relationships and herb-chemical relationships. A list of herb-related diseases is also provided. Herb-disease relationships in the detailed information were extracted based on the frequency of abstracts, where both an herb name and a disease name co-occurred in the same PubMed abstracts. Thus, highly ranked diseases would be the most related diseases that can be treated with a specific herb. Users can optionally restrict results of queries to gene-chemical relationships, which were contained in PubMed abstracts selected by clicking 'apply filters', as shown in the first table of Figure 3. This allows researchers to use HerDing as a search prioritization tool based on articles that are more appropriate for their research design. All the retrieved results are freely downloadable by clicking a button denoted as 'download this result(.tsv)' in the results page and stored in the tab-separated file.

When users input a chemical name or an herb name, the summary and detailed information in Figure 3 are changed. With a query of a chemical name, a list of herbs that contain the input chemical and a list of genes targeted by the chemical are shown. With a query of an herb name, a list of chemicals contained in the given herb and a list of genes targeted by the chemicals are presented.

### Evaluation of the recommended herbs

To evaluate the HerDing system, we examined 20 diseases including Alzheimer's disease and diabetic neuropathies



**(a) Genes** A list of supported genes

Gene  
Gene symbols are separated by a comma.  
(retrieve candidate herbs that are related to input genes.)

**Genes(up-regulated)**

Gene(up-regulated)  
Gene symbols are separated by a comma.  
(retrieve candidate herbs that up-regulate input genes.)

**Genes(down-regulated)**

Gene(down-regulated)  
Gene symbols are separated by a comma.  
(retrieve candidate herbs that down-regulate input genes.)

**Example Queries**

- TNF, IL1B, IL6, IL8, IL17A, MMP1, MMP3, MMP9, MMP12, MMP13, COX2, IL10, IL4
- (Up-regulated) IL10, IL4
- (Down-regulated) TNF, IL1B, IL6, IL8, IL17A, MMP1, MMP3, MMP9, MMP12, MMP13, COX2

**Disease**

Disease name  
Input a single disease name. At least two alphabets are required to see disease lists.  
(When a disease name is given, it will be automatically replaced with genes from MalaCards, which are related to the given-disease.)

**Example Queries**

- Hypertension (D006973)
- Rheumatoid arthritis (D001172)
- Alzheimer Disease (C536597)

**>> Advanced Options**

**Herbs** A list of supported herbs (Chinese/English/Latin names)

Herb name  
When a list of herbs is provided here, the search space of herbs is limited to the input

**(b) Chemical** A list of supported chemicals

Chemical  
Input a single chemical name to retrieve related-herbs and -target genes.

**Example Queries**

- e.g. Saponins
- e.g. Beta-myrcene
- e.g. Palmitic acid

**(c) Herb (Chinese Name)** A list of supported herbs (Chinese/English/Latin names)

Herb  
Input a single herb name (Chinese Name) to retrieve related-chemicals and -target genes.

**Herb (English Name)**

Herb  
Input a single herb name (English Name) to retrieve related-chemicals and -target genes.

**Herb (Latin Name)**

Herb  
Input a single herb name (Latin Name) to retrieve related-chemicals and -target genes.

**Figure 2.** HerDing's input pages. (a) Users can enter four different types of input in the each input box (a list of up-regulated genes, a list of down-regulated genes, genes regardless of the direction of expression changes, and a single disease name or a MeSH identifier). For advanced options, when users enter a list of herbs, the search space of herbs is limited to the input herbs. A list of supported herbs including Chinese/English/Latin names can be checked by clicking the 'A list of supported herbs' button. The number of maximum retrieved herbs can be changed. (b) Users can enter a chemical name to retrieve related herbs and target genes. (c) An herb name can be entered to retrieve chemicals containing the input herb and genes targeted by the chemicals.

(Supplementary Table S1). Genes related to a given disease were collected from MalaCards. When we manually examined the top ten recommended herbs for each disease, research articles supporting the effectiveness of herbs for the disease were retrieved, as shown in Supplementary Table S1. For example, sea cucumber, wolfberry fruit and aloe vera were recommended for Alzheimer's disease (ranked first, second and third, respectively), and their effectiveness has previously been reported. Similarly, it was reported that *Angelica sinensis*, *Curcuma longa* and *Portulaca oleracea* (ranked first, second and third, respectively) were effective for the treatment of diabetic neuropathies.

Jeoung *et al.* (28) reported that Ganghwaljetongyeum (GHJTY), a complex herbal formula consisting of 18 herbs, was effective for the treatment of rheumatoid arthritis. When we searched HerDing using 50 genes related to rheumatoid arthritis and 18 herbs from GHJTY, all 18 herbs were found to target at least one gene out of 50 genes. We also observed that *Chinese angelica* (Dang Gui) and *clematis chinensis* (Wei Ling Xian) were highly ranked herbs, targeting 21 and 10 genes, respectively, whereas both oriental water plantain (*Ze Xie*) and five leaf akebia (*Mu Tong*) target a single gene. This result implies that

further research to discover a simpler herbal formula using a few top-ranked genes can be conducted in treating rheumatoid arthritis.

## Summary and Discussion

HerDing is a search engine for recommending herbs to treat diseases and was constructed based on the assumption that an association between marker genes of diseases and herbs would draw indirect relationships between herbs and diseases. These herb-chemical-gene-disease indirect relationships were extracted by integrating public resources and by using the text-mining approach. Although several Web-based databases in the field of traditional Chinese medicine, including TCMID (2), have been constructed, most of them provide herb-disease relationships when the relationships were directly mentioned in research articles. With the increasing knowledge about molecular mechanisms of diseases, diseases have been researched based on genes. The novelty of HerDing is that users can retrieve candidate herbs for treating diseases based on multiple disease-related genes, which allows users to explore more candidate herbs.

## 1. Your Input

- 50 Genes ("Alzheimer's Disease" related) : A2M, BLMH, HFE, ACE, MPO, HLA-H, APP, MT-ND1, PSEN1, PSEN2, APOE, CTSD, GSK3B, COL25A1, AC091878.1, RP13-143G15.3, PLD3, AD6, AD9, AD7, AD11, AD8, AD16, AD17, AD14, AD15, AD10, AD5, AD12, AD13, MAPT, SORL1, NOS3, PLA2, ADAM10, APBB2, DHCR24, SNCA, PICALM, CTSB, RCAN1, CRH, SLC6A4, GSAP, ITM2B, CD40, PADI2, CALHM1, BCHE, SERPINA3

- Restriction of the results to the articles containing the gene-chemical relationships that appear in the selected PubMed articles.

Click here if you want to restrict the results to the selected PMIDs

19931519  8524236  8631852  18260696  23373633  19619570  18248325  20219449  22649187  4631366  21706413  22960160  
 19135050  22609851  12039845  17404688  18317567  7738427  15637762  18355967  16793930  7909315  15615701  19371625

## 2. Summary (retrieved herbs from input genes)

[Download This Result \(.tsv\)](#)

[chemicals] : Show chemicals that target the input genes among those contained in the herb. ( )

| Herbs     | # of matched genes | A2M              | BLMH             | HFE              | ACE              | MPO              | HLA-H | APP              | MT-ND1 | PSEN1            | PSEN2            | APOE             | CTSD             | GSK3B            | COL25A1          | AC091878.1 | RP13-143G15.3 | PLD3             | AD6 | AD9 | AD7 | AD11 | AD8 | AD1 |   |
|-----------|--------------------|------------------|------------------|------------------|------------------|------------------|-------|------------------|--------|------------------|------------------|------------------|------------------|------------------|------------------|------------|---------------|------------------|-----|-----|-----|------|-----|-----|---|
| HAI SHEN  | 25                 | [2]<br>chemicals | [1]<br>chemicals | [2]<br>chemicals | [1]<br>chemicals | [4]<br>chemicals | -     | [4]<br>chemicals | -      | [1]<br>chemicals | [2]<br>chemicals | [2]<br>chemicals | [1]<br>chemicals | [2]<br>chemicals | [1]<br>chemicals | -          | -             | -                | -   | -   | -   | -    | -   | -   | - |
| GOU QI ZI | 23                 | -                | -                | [1]<br>chemicals | [2]<br>chemicals | [4]<br>chemicals | -     | [5]<br>chemicals | -      | [1]<br>chemicals | [2]<br>chemicals | [2]<br>chemicals | [1]<br>chemicals | [2]<br>chemicals | [1]<br>chemicals | -          | -             | -                | -   | -   | -   | -    | -   | -   | - |
| LU HUI    | 22                 | [1]<br>chemicals | -                | [1]<br>chemicals | [2]<br>chemicals | [3]<br>chemicals | -     | [4]<br>chemicals | -      | [1]<br>chemicals | [2]<br>chemicals | [3]<br>chemicals | [4]<br>chemicals | [2]<br>chemicals | [1]<br>chemicals | -          | -             | -                | -   | -   | -   | -    | -   | -   | - |
| DANG GUI  | 21                 | [2]<br>chemicals | [2]<br>chemicals | -                | [1]<br>chemicals | [6]<br>chemicals | -     | [5]<br>chemicals | -      | [2]<br>chemicals | [2]<br>chemicals | [1]<br>chemicals | [6]<br>chemicals | [2]<br>chemicals | -                | -          | -             | [1]<br>chemicals | -   | -   | -   | -    | -   | -   | - |



## 3. Detailed information

[Download This Result \(.tsv\)](#)

[25] HAI SHEN (\*English: SEA CUCUMBER \*Latin: HOLOTHURIA LEUCOSPILOTA)

\* Targeted Genes : APP, RCAN1, HFE, MPO, PSEN1, BLMH, ACE, NOS3, A2M, GSAP, PSEN2, MAPT, BCHE, CD40, GSK3B, COL25A1, SORL1, CALHM1, PADI2, CRH, APOE, CTSD, CTSB, DHCR24, ADAM10  
[Hide all](#)

| Chemical | Chemical Reference | Expression                                  | Gene Symbol | Gene Reference | Organism | Reference (chemical-gene) | Reference (herb-chemical) |
|----------|--------------------|---|-------------|----------------|----------|---------------------------|---------------------------|
| Iron     | PubChem27264       | CTD: affects*abundance increases*expression | APP         | 351            | NA       | 19818510                  | TCMID                     |
| calcium  | PubChem24448133    | NA  | RCAN1       | 1827           | NA       | TCMID                     | TCMID                     |

\* Related Diseases : neoplasms , multiple sclerosis , van der woude syndrome , hepatitis a , carcinoma etc....  
[Hide all](#)

| Disease Name       | Disease ID   | Frequency (Co-occurrence) |
|--------------------|--------------|---------------------------|
| neoplasms          | MeSH D009369 | 46                        |
| multiple sclerosis | MeSH D009103 | 21                        |

**Figure 3.** HerDing's results page with an input of Alzheimer's disease. HerDing displays user inputs, a summary of retrieved herbs, and detailed information. In the first section, users can check their input and restrict results to the selected articles. In the summary section, retrieved herbs with input genes are briefly presented in the table and the graph. Furthermore, chemicals targeting genes can be checked by clicking the 'chemical' button. In the detailed information section, users can obtain detailed information about herb-chemical-gene relationships and a list of diseases that are known to be related to a specific herb.

Although HerDing provides new insights in searching herb-disease relationships, there is still much room for improvement by providing more abundant and accurate data. To extract information from less-studied herbs, a high-performance text-mining method is required. Thus, if we update the dictionary with more herb names and develop an accurate NER tool, the number of herb-chemical relationships will be increased. In addition, developing the machine learning approaches for extracting herb-chemical relationships will be a future work. To increase the database size, other text resources related to natural products, such as news articles, electronic books, Wikipedia and Google, will be useful.

## Acknowledgements

This work was supported by the Bio-Synergy Research Project (NRF-2013M3A9C4078138) of the Ministry of Science, and by a grant of the Korea Healthcare technology R&D Project, Ministry for Health & Welfare Affairs, Republic of Korea (HI13C2285).

## Supplementary data

Supplementary data are available at *Database Online*.

*Conflict of interest.* None declared.

## References

- Wang, L., Zhou, G.-B., Liu, P. *et al.* (2008) Dissection of mechanisms of Chinese medicinal formula Realgar-Indigo naturalis as

- an effective treatment for promyelocytic leukemia. *Proc. Natl. Acad. Sci. U. S. A.*, 105, 4826–4831.
2. Xue,R., Fang,Z., Zhang,M. *et al.* (2012) TCMID: traditional Chinese medicine integrative database for herb molecular mechanism analysis. *Nucleic Acids Res.*, 41, 1089–1095.
  3. Wang,J.F., Zhou,H., Han,L.Y. *et al.* (2000) Traditional Chinese medicine information database. *Clin. Pharmacol. Ther.*, 78, 92–93.
  4. Ye,H., Ye,L., Kang,H. *et al.* (2011) HIT: linking herbal active ingredients to targets. *Nucleic Acids Res.*, 39, 1055–1059.
  5. Chen,C.Y.C. (2011) TCM Database@Taiwan: the world's largest traditional Chinese medicine database for drug screening in silico. *PLoS One*, 6, e15939
  6. Kuhn,M., von Mering,C., Campillos,M. *et al.* (2008) STITCH: interaction networks of chemicals and proteins. *Nucleic Acids Res.*, 36, 684–688.
  7. Hamosh,A., Scott,A.F., Amberger,J.S. *et al.* (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.*, 33, D514–D517.
  8. Knox,C., Law,V., Jewison,T. *et al.* (2011) DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.*, 39, 1035–1041.
  9. Federhen,S. (2012) The NCBI Taxonomy database. *Nucleic Acids Res.*, 40, 136–143.
  10. Swanson,D.R. (1986) Fish oil, Raynaud's syndrome, and undiscovered public knowledge. *Perspect. Biol. Med.*, 30, 7–18.
  11. Choi,W.J., Lee,H. (2015) A text mining approach for identifying herb-chemical relationships from biomedical articles. DTMBIO'15, October 23–23, 2015, Melbourne, VIC, Australia.
  12. Baldwin,B. and Carpenter,B. (2000) LingPipe. <http://www.alias-i.com/lingpipe/>. [Accessed 23 December 2014].
  13. Rocktaschel,T., Weidlich,M., and Leser,U. (2012) ChemSpot: a hybrid system for chemical named entity recognition. *Bioinformatics*, 28, 1633–1640.
  14. Davis,A.P., Murphy,C.G., Saraceni-Richards,C.A. *et al.* (2009) Comparative Toxicogenomics Database: a knowledgebase and discovery tool for chemical-gene-disease networks. *Nucleic Acids Res.*, 37, 786–792.
  15. Bolton,E., Wang,Y., Thiessen,P.A. *et al.* (2008) PubChem: integrated platform of small molecules and biological activities. *Annu. Rep. Comput.*, 4, 217–241.
  16. UniProt Consortium. (2009) The Universal Protein Resource (UniProt). *Nucleic Acids Res.*, 37, 169–174.
  17. Hatcher,E. and Gospodnetic,O. (2004) *Lucene in Action*. Manning Publications Co., Greenwich, CT.
  18. Han,J.H., Koh,W., Lee,H.J. *et al.* (2012) Analgesic and anti-inflammatory effects of ethyl acetate fraction of *Polygonum cuspidatum* in experimental animals. *Immunopharmacol. Immunotoxicol.*, 34, 191–195.
  19. Lee,E.-G., Lee,S.-I., Chae,H.-J. *et al.* (2010) Ethyl acetate fraction from *Cudrania tricuspidata* inhibits IL-1 $\beta$ -induced rheumatoid synovial fibroblast proliferation and MMPs, COX-2 and PGE2 production. *Biol. Res.*, 43, 225–231.
  20. Lee,E.-G., Lee,S.-I., Yun,H.-J. *et al.* (2010) Ethyl acetate fraction from *Cudrania tricuspidata* inhibits IL-1 $\beta$ -stimulated osteoclast differentiation through downregulation of MAPKs, c-Fos and NFATc1. *Korean J. Intern. Med.*, 25, 93–100.
  21. Lin,L.W., Chen,H.Y., Wu,C.R. *et al.* (2008) Comparison with various parts of *Broussonetia papyrifera* as to the antinociceptive and anti-inflammatory activities in rodents. *Biosci. Biotechnol. Biochem.*, 72, 2377–2384.
  22. Wang,X., Zhao,L., Han,T. *et al.* (2008) Protective effects of 2,3,5,4'-tetrahydroxystilbene-2-O-beta-d-glucoside, an active component of *Polygonum multiflorum* Thunb, on experimental colitis in mice. *Eur. J. Pharmacol.*, 578, 339–348.
  23. Rappaport,N., Nativ,N., Stelzer,G. *et al.* (2013) MalaCards: an integrated compendium for diseases and their annotation. *Database*, 2013, bat018
  24. Esmat,A.Y., Said,M.M., Soliman,A.A. *et al.* (2013) Bioactive compounds, antioxidant potential, and hepatoprotective activity of sea cucumber (*Holothuria atra*) against thioacetamide intoxication in rats. *Nutrition*, 29, 258–267.
  25. Liu,J., Jia,S.H., Kirberger, M. *et al.* (2014) Lunasin as a promising health-beneficial peptide. *Eur. Rev. Med. Pharmacol. Sci.*, 18, 2070–2075.
  26. Jia,Y., Dong,J., Wu,X. *et al.* (1998) The effect of *Lycium barbarum* polysaccharide on vascular tension in two-kidney, one clip model of hypertension. *Sheng Li Xue Bao.*, 50, 309–314.
  27. Legssyer,A., Ziyat,A., Mekhfi,H. *et al.* (2002) Cardiovascular effects of *Urtica dioica* L. in isolated rat heart and aorta. *Phytother. Res.*, 16, 503–507.
  28. Jeoung,BR., Lee,KD., Na,CS. *et al.* (2013) Ganghwaljetongyeum, an anti-arthritic remedy, attenuates synoviocyte proliferation and reduces the production of proinflammatory mediators in macrophages: the therapeutic effect of GHJTY on rheumatoid arthritis. *BMC Complement. Altern. Med.*, 13, 786–792.