# **Visualization of Traditional Chinese Medicine Formulas**

## Abstract

**Background:** Traditional Chinese herbal medicine formulas are combinations of Chinese herbal medicines. Understanding classic medicine formulas is the basis of TCM diagnosis and treatment, and is the core for TCM inheritance. The large number and the flexibility of medicine formulas make them difficult to memorize and understand rules of composition. The multifaceted and multidimensional properties of herbal medicines are important for understanding formulas, but are typically separated from the formula information. Furthermore, all information is presented in texts and cannot be analyzed jointly and interactively.

**Objective:** This work aims to devise a visualization method for TCM formulas that shows the composition of medicine formulas and the multidimensional properties of herbal medicines involved, and supports the comparison of medicine formulas.

**Methods:** A TCM formulas visualization method with multiple linked views is proposed and implemented as a web-based tool after a close collaboration between visualization and TCM experts. The composition information of medicine formulas is visualized in a formula view with a similarity-based layout supporting the comparison of compositing medicines of formulas nearby; a shared medicine view complements the formula view by showing all overlaps of pair-wise formulas; a dimensionality reduction plot of medicines enables the visualization of multidimensional medicine properties. Medicines are color-encoded with a perceptual-guided color map that encodes multidimensional TCM attributes and the similarity measure at the same time. The color map is calculated by a data-driven interpolation scheme. With simple interactions, users could flexibly select medicines or formulas of interest, and the corresponding elements in other views are highlighted through brushing-and-linking.

**Results:** Our method is applied to two typical categories of medicine formulas, namely, tonic prescriptions and heat-clearing prescriptions, which contain 20 and 27 formulas, respectively. The numbers of herbal medicines within these prescriptions are 40 and 60, respectively. Each herbal medicine has a XXX-dimensional property. A TCM expert explored the two datasets with our method and quickly gained insight into formulas of interest that are difficult to identify with the traditional text-based method. Moreover, the positive feedback of the expert on our tool indicates the usefulness of our method.

Be sure to include relevant statistics here, such as sample sizes, response rates, *P* values or Confidence Intervals. Be specific (by stating the value) rather than general (e.g., “there were differences between the groups”).

**Conclusions:**

**Keywords:** visual analytics; traditional Chinese herbal medicine formulas; multidimensional data; color map design

## Introduction

Traditional Chinese herbal medicine formulas are combinations of Chinese herbal medicines. Understanding classic medicine formulas is the basis of TCM diagnosis and treatment, and is the core for TCM inheritance. In this paper, we use the term “medicine formulas” and “formulas” to refer to traditional Chinese herbal medicine formulas. In TCM, “syndrome differentiation and treatment” is a core method for treatment and the medicine formulas issued in clinical practice are based on the classic medicine formulas, which may be adjusted at any moment according to symptoms of patients. A typical prescription may contain several medicine formulas. The large number and the flexibility of medicine formulas make them difficult to memorize and understand rules of composition.

The traditional education method is to recite the classical medicine formulas in combination with expert medical records[1, 2]. However, the medicine data of medicine formulas and medical records are only presented in texts (Table 1), and the composition rules could not be intuitively understood. Data mining and some visual presentations are adopted in the existing computerized analysis of TCM medicine formulas[3, 4, 5] . However, these methods are query-based and do not allow users to interactively explore medicine formulas, and the relatively simple visualization cannot provide the overview of a group of medicine formulas nor in-depth comparison of formulas.

In this paper, we propose a visualization method for TCM medicine formulas to address the issues of the existing training method. As a result of a close collaboration between visualization and TCM experts, our method provides compact and clear visualization of hierarchical data of medicine formulas and multidimensional attribute data of medicines. While our method is data-driven, we emphasize the perception aspect of the visualization. Specifically, a

layout algorithm is designed to improve the comparability and reduce visual clutter of tree branches in the icicle chart with data-driven similarity measurement; medicines are color mapped with a 2D color map generated with radial basis function (RBF) interpolation in a perceptual-uniform color space with TCM-concept inspired colors, while their placement in the dimensionality reduction plot is driven by their TCM properties. As shown in Figure 1, our method is realized as a web-based interactive tool, which comprises two linked views: a medicine formula view with an icicle chart (right) and a medicine view with a dimensionality reduction plot (left). With brushing-and-linking, users could flexibly select drugs of interest with lassos in the medicine view, and the corresponding formulas are highlighted in the medicine formulas view; conversely, medicines within selected formula in the medicine formulas view are highlighted in the medicine view. Colors of medicines provide rich information as they encode the multidimensional TCM attributes and the similarity measure at the same time.

These are the contributions of our method:

•A layout algorithm for medicine formula hierarchies to improve comparison capability of medicines in an icicle plot.

•Perception-guided data-driven color encoding for medicines.

•An interactive tool that facilitates the analysis and understanding of TCM medicine formulas as a result of a close interdisciplinary collaboration.

The usefulness of our method is demonstrated by two use cases of typical medicine formulas. A TCM expert analyzed these medicine formulas with our method and considered that the new method could effectively reveal the constitution principle of medicine formulas in an intuitive way, and could assist the learning of the subject in TCM. Furthermore, our method provides a link between medicine attributes and medicine formulas, and, therefore, is a promising tool for designing new formulas.

Table 1: Part of the original text-based medicine formula information summarized from the textbook Medicine Formulas (Tenth Edition)[31][.](#_bookmark32)

|  |  |  |
| --- | --- | --- |
| Formula | Medicines | |
|  |  |
| 八珍汤 | 人参、熟地黄、当归、川芎、白术、茯苓、白芍、炙甘草、生姜、大枣 | |
| 参苓白术散 | 人参、白术、茯苓、莲子、薏苡仁、山药、桔梗、大枣、甘草、砂仁、白扁豆 | |
| 生脉散 | 人参、麦冬、五味子 | |
| 四君子汤 | 人参、甘草、白术、茯苓 | |
| 大补阴丸 | 熟地黄、龟甲、黄柏、知母 | |
| 四物汤 | 熟地黄、白芍、川芎、当归 | |
| 地黄饮子 | 熟地黄、山茱萸、肉苁蓉、巴戟天、麦冬、远志、生姜、附子、茯苓、大枣、五味子、石斛、石蒲、肉桂、薄荷 | |

## Methods

### Background and Data

Classifications of Chinese herbal medicines are multi-faceted and multi- leveled[2]. For example, perimental research that focuses on Siqi (四气), Wuwei (五味), and Guijing (归经) has been an important area in TCM research[1]. Guijing regards the orientation of Chinese herbal medicines, which is to closely connect the functions of medicines with the organs and meridians (脏腑经络) of the human body.

Chinese herbal medicines can be divided into four properties (Siqi) as cold (寒), hot (热), warm (温) and cool (凉) according to their functions on the human body. Cold medicines generally have the function of heat-clearing and detoxifying (清热解毒), which are used to treat febrile diseases. Warm medicines typically have the function of warming Yang to expel coldness (温阳散寒), which are used to treat cold diseases. In addition, a kind of medicines with gentle properties exist, which is namely Ping (平). Wuwei means five flavors: pungent (辛), sweet (甘), sour (酸), bitter(苦) and salty (咸). Pungent medicines, in general, taste spicy (such as ginger) or cool (such as mint and borneol), and are used to sweat and relieve Qi. Sweet medicines typically taste sweet, for example, licorices and wheat. They have the effect of easing and nourishing. They are also used to reconcile medical properties. Sour medicines, such as black plum and gallnut, have an astringent effect and are used to stop sweating and diarrhea. Bitter medicines have functions of clearing heat, drying dampness and catharsis. Finally, salty medicines have the functions of diarrhea and defecation, softening and firmness. It is believed that these factors are associated with body heat production processes or metabolic activities, and may also play a role in the digestive system, nervous system, and cardiovascular system, and so on[30].

Another important concept is Jun-Chen-Zuo-Shi (君臣佐使). Jun- Chen-Zuo-Shi are the principles for the compatibility of TCM prescriptions. Junyao (君药), or namely, sovereign medicines as used hereafter, play a major role to against the main disease or syndrome. It is the primary medicine among the prescriptions. Texts in blue in Table 1 indicate Junyao in the corresponding formulas.

In this work, the Siqi and Wuwei properties are represented as XXX- and YYY-dimensional vectors, respectively.

### Requirement Analysis and Method Overview

The goal of our work is to devise a method that supports the joint visualization of medicine formulas and the attributes of corresponding drugs. The visual design should support comparison of formulas and drugs and also facilitate the classification of drugs based on their SiqiWuwei properties. Visualization and TCM experts work closely together to analyze requirements of the visual analysis method for medicine formulas. The requirements are summarized as follows.

R1: clear visualization of medicine formulas.

R2: comparing different medicine formulas with ease.

R3: sovereign drugs should be highlighted.

R4: associating medicine formulas and attributes of the corresponding drugs.

R5: visual elements should be effectively perceived.

R6: interactions should be easy.

R7: visual designs should reflect general concepts of TCM.

Our method is the result of an iterative development process with quick prototypes. Prototypes were realized based on the requirements and proposed to the TCM expert and improvements were made given the feedback of the TCM expert.

The workflow of our method is shown in Figure [2:](#_bookmark3) the medicine formulas information and the multidimensional medicine attribute data are prepared as the input (Section [4.1);](#_bookmark5) medicine attribute data are projected to the low dimensional space (2D) and pair-wise distances are calculated; medicine formulas data is converted into a hierarchy with a tree layout using our similarity-based layout algorithm and is visualized as an icicle plot (Section 4.2); next, colors are designed for medicines using our perceptual-guided data-driven color encoding method (Section 4.3).

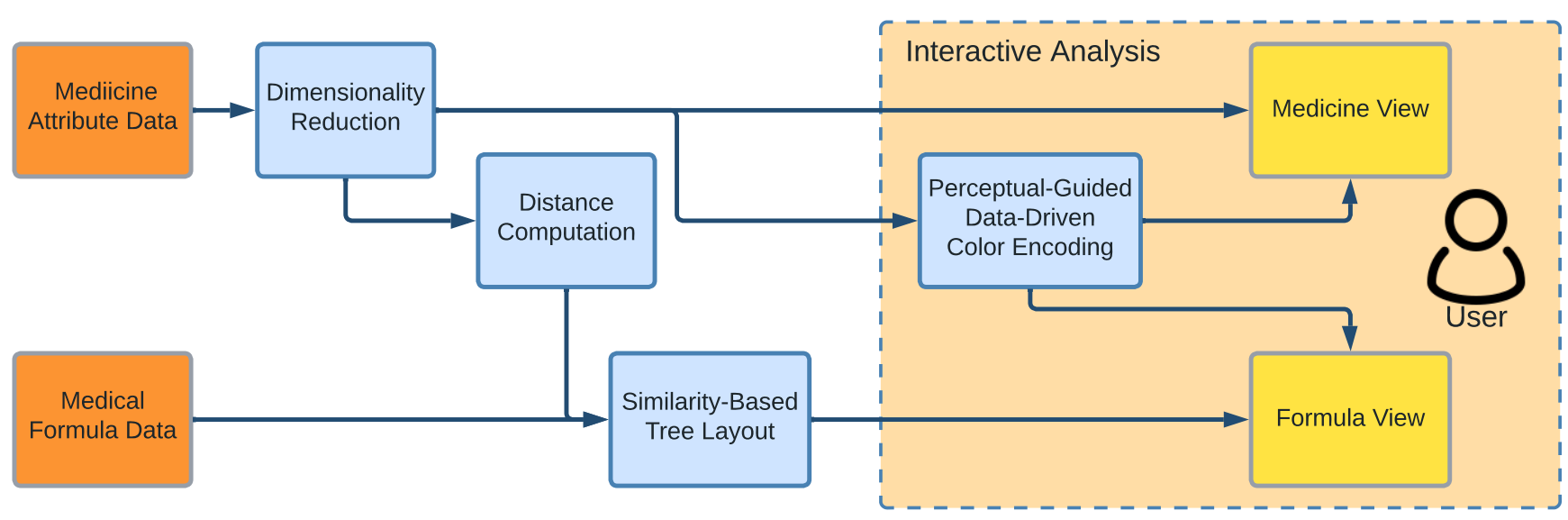


Fig. 2: The workflow of our method.

In this paper, the sources of the input data are as follows. The medicine formulas data are extracted from the key medicine formulas in the textbook Medicine Formulas (Tenth Edition)[31] as shown in Table 1. The multidimensionial medicine attribute data are retrieved from the SymMap database[32].

### Dimensionality Reduction and Distance Computation

Attributes used in our method are Siqi and Wuwei in an M-dimensional space (M=XXX?). The Siqi and Wuwei attributes of an herbal medicine can be written as a vector **P** of binary valued elements:



The M-dimensional space is then dimensionality reduced to 2D, and an herbal medicine can be represented with a 2D vector **p**:



UMAP[23]is used for its structure preservation ability and computational efficiency. The scatterplot of medicines in the dimensionality- reduced space is shown in Figure [1(left).](#_bookmark0)

The distance between medicines is the basis of our subsequent similarity-based hierarchy layout computation and visualization. We define the distance *d(u, v)* between two herbal medicine *u* and *v* as the L2-norm, i.e., Euclidean distance, between their corresponding 2D vectors **pu** and **pv**, respectively:



A distance between **Pu** and **Pv** in the original M-dimensional space is also considered. Therefore, discriminating herbal medicine based on the distance using **P** is more difficult than that of the projected vectors **p**, and, makes the resulting hierarchy visualization more difficult for comparison and comes with more visual clutter.

### Formulas Visualization

#### Domain Expert Evaluation of Set Visualization Methods

Typically, a dozen formulas and even more herbal medicines are involved in a category of formulas. From a set visualization perspective, both the number of sets and set elements are large, and, therefore, a suitable visualization that scales well and is easy to understand is required.

An informal domain expert evaluation of popular sets visualization techniques was performed to drive the design of a proper set visualization method. Examples of a Euler diagram, an overlay (Bubblesets), a node-link diagram, and a matrix-based method that are figures in the survey paper by Alsallakh~\cite{} were shown to the TCM expert. The expert was asked to rank the feasibility of these methods for medicine formulas visualization based on the scalability, the ease of understanding and the support of comparison.

The matrix-based method is ranked first by the TCM expert, followed by the node-link diagram, the Euler diagram, and the overlay. The expert found that the matrix-based method is easy to understand and scales to a large number of sets and set elements. However, she was concerned about the compactness of this method as the matrix is typically large but sparse for medicine formulas data. Also, she found it difficult to trace elements of a set and comparing sets if the matrix is large.

For the node-link diagram, the expert found it easier to understand than the matrix, but the crossings of links make tracing difficult for large number of sets.

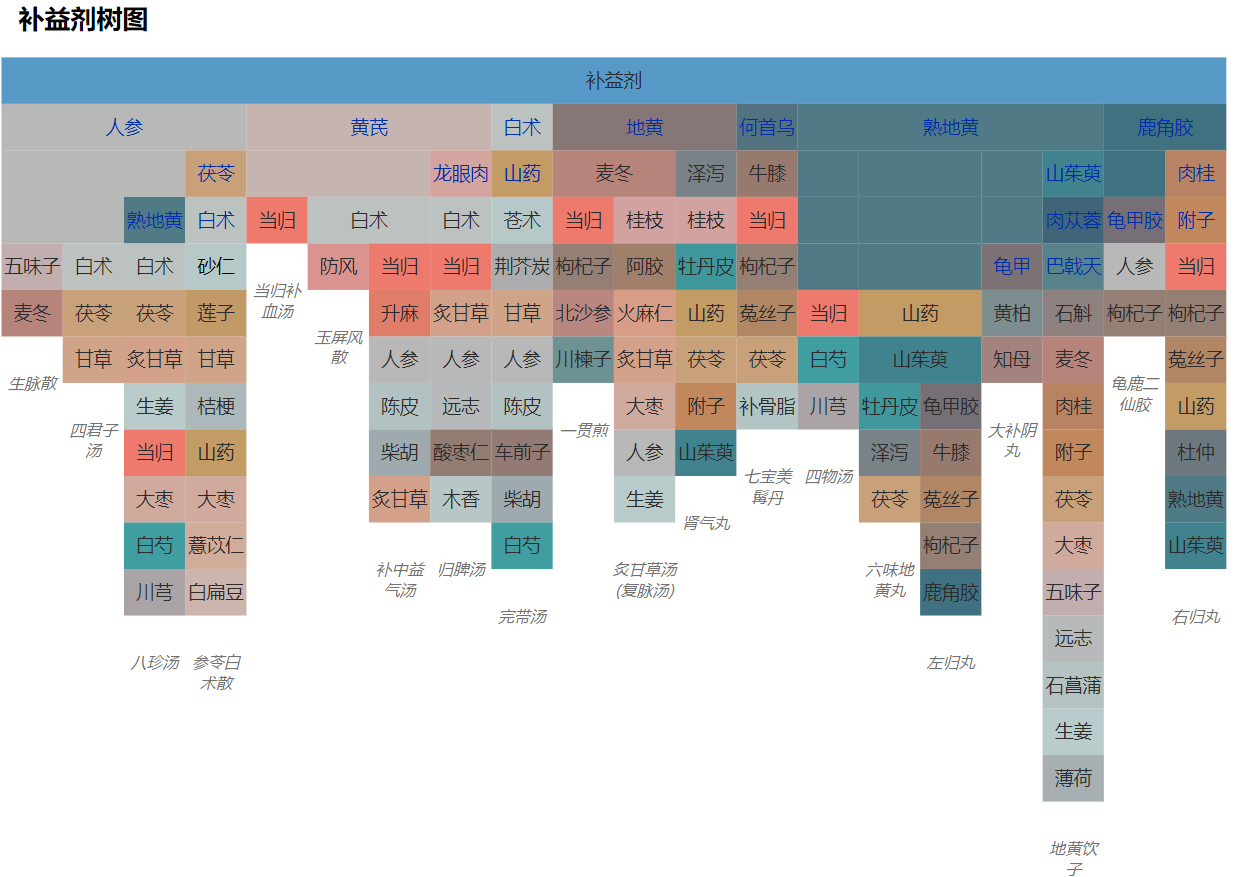
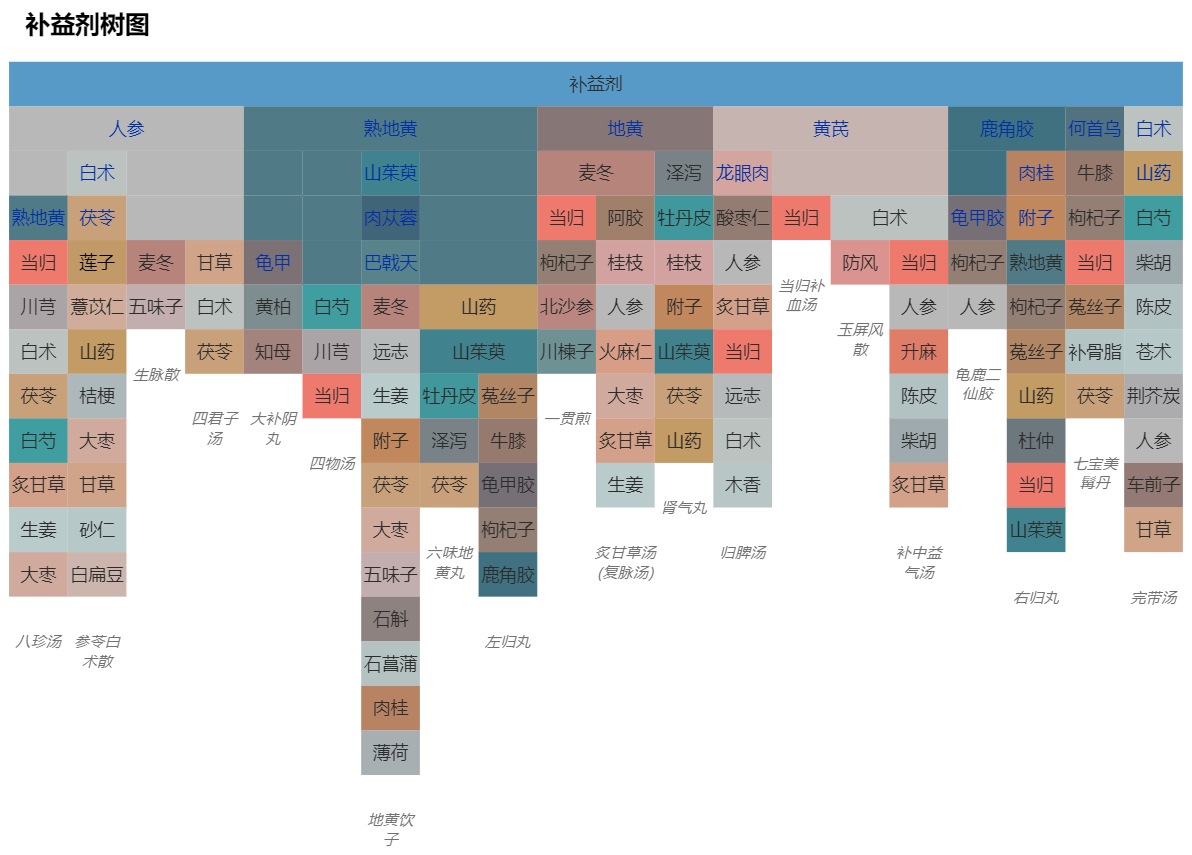
The intuitiveness of the Euler diagram and the overlay were liked by the expert. However, they did not scale well for the number of sets. Especially, overlapping sets which are typically the case of medicine formulas data as formulas in a category share same medicines quickly makes these methods unusable. Therefore, the expert considered the Euler diagram and the overlay infeasible for our case.

According to the evaluation, we decide to devise a sparse matrix-based method to show medicine formulas and meets requirements R1 and R2. With a sparse matrix representation, the set-element matrix can be compressed to rows of non-zero elements. This representation is similar to an icicle plot~\cite{} for hierarchy visualization. If the icicle plot is properly laid out, it has potential to support comparison of similar medicine formulas. Furthermore, the icicle plot provides the opportunity to encode medicines in hierarchy, e.g., the separation of sovereign medicines and other medicines.

Therefore, we use the icicle plot as the basis of our visual design for the medicine formulas view. To support analysis of shared medicines within formulas, a co-occurrence matrix view is used to complement the icicle plot view.

#### Icicle Plot of Medicine Formulas

Each record in the medicine formula data contains the name of the formula, names of medicines, and tags for sovereign medicines (Table [1).](#_bookmark1) We set the content of elements of the icicle plot to names of medicines, and use each column to show a medicine formula as shown in Figures [3](#_bookmark9) and [4.](#_bookmark10)



1. (b)

Fig. 3:Icicle plots with (a) the original order of medicine formulas data and (b) our similarity-based layout.

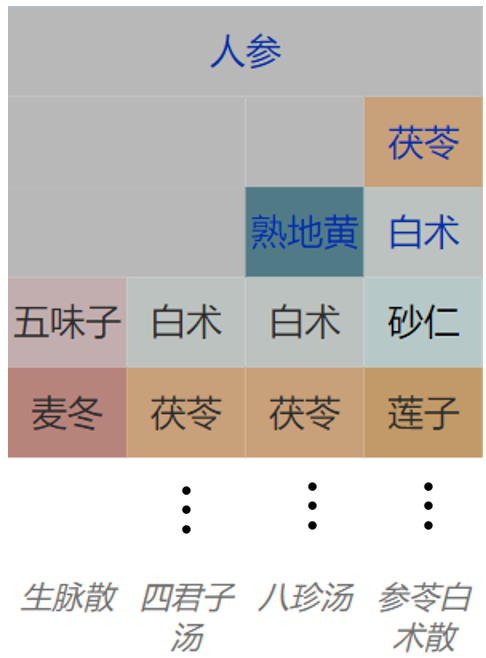


Fig. 4: The design of the icicle plot of medicine formulas. Each column of the icicle plot contains a medicine formula, which comprises sovereign medicines (texts in blue) and other medicines (texts in black). The name of a formula is placed under each column.

In our design, sovereign medicines are highlighted and treated differently than other medicines to meet requirement R3. As shown in Figure 4, sovereign medicines are placed on the top levels of the hierarchy and colored blue. Formulas with common sovereign medicines are grouped together. Rows are padded so that the top of all none- sovereign medicines are aligned for comparison (R2). For example, rows are padded for Renshen (人参) as seen in Figure 4.

The name of a medicine formula is placed under its corresponding column in italic font face with a fixed vertical spacing as shown in Figure 3. This design is simple yet effective: the height of each column is used as an additional cue to the horizontal position for quick alignment of a formula and its name.

Since the set-based formula information has to be converted into columns of the icicle plot—an ordering is needed for medicines in a formula. However, medicines in the original medicine formulas data has no specific ordering: the resulting icicle plot of medicine formulas of tonic prescriptions with the original appearance ordering of medicines is shown in Figure 3 (a). The plot is cluttered and comparing compositions of medicine formulas is difficult as frequent context switch has to be made while searching for a same medicine. Therefore, we propose a similarity-based tree layout to facilitate easier comparison and clearer visualization of medicine formulas than using the original ordering.

#### Similarity-Based Layout Computation

Our goal is to design a layout that is arranged based on the similarity of the corresponding herbal medicine so that those are similar in attributes are grouped together to enable effective comparison and reduce visual clutter in the icicle plot. Our method is an efficient greedy algorithm with two steps: first, the arrangement of sovereign medicines, and then we arrange the remaining medicines.

To facilitate the explanation, we introduce the similarity sequence  for a set of medicines . The element *si* of *S* reads:



where ** is the distance between *s* and *h* with Equation 3, and *t* is a random number between 1 and *n*.

**The Arrangement of Sovereign Medicine Layout**

In this step, columns of the icicle plot are sorted based on the similarity of sovereign medicines. For the set of all the sovereign medicines, if a medicine is the only sovereign medicine in a certain medicine formula, it is assigned as the top-level sovereign medicine. We denote the set of all such medicines as *Hs*. The first element *s1 = ht* is randomly selected from the set *Hs*, and the rest of the sequence is set by finding the medicine *hj* with the shortest distance to the previously ordered element *si*. The sorted top-level sovereign medicines are placed on the first row of the icicle plot.

We now process formulas with more than one sovereign medicines. For a top-level sovereign medicine *hi*, a set  denotes all medicine formulas that have *hi* as a sovereign medicine, and the *j*-th formula in  is denoted as . If its sovereign medicine contains elements of *Hs*, add that formula to set ; if none of the sovereign medicines in a formula is contained in *Hs*, *hj* is selected randomly as the top- level medicine and added to *Hs*. An example is Wandaitang (完带汤) in Figure 3. For each , formulas with single sovereign medicine are sorted from left to right by the number of remaining medicines; formulas with multiple sovereign medicines are sorted by the number of sovereign medicines. For sovereign medicines that are not top-leveled, they are sorted according to the distance and laid out as subsequent children nodes (as rows). Padding is made to ensure that all non- sovereign medicines start at the same row.

For example, Figure 4 visualizes set  of Figure 3, with Renshen (人参) as the top-level sovereign medicine, and Bazhentang (八珍汤) and Shenlingbaizhusan (参苓白术散) have more than one sovereign medicines (columns 2 and 3, respectively). Therefore, the sovereign medicine rows are padded to three rows as Shenlingbaizhusan has the maximum of sovereign medicines which is three.

**The Arrangement of Remaining Medicines**

Next, we arrange the remaining medicines in the medicine formulas data. Define the medicine in the *j*-th column, and *k*-th row in the *i*-th multiple medicine formulas set  in the icicle plot as *hijk*. Each formula column is then the previously introduced set. We construct the position sequence of as from left to right. The leftmost column is sorted by the distance-based ordering using Equation 4. Starting from the second column from the left, medicines are sorted by local similarity. We align the same medicines in adjacent columns even if they are not from the same . Other medicines are sorted according to their distances to medicines on its left column within the set :



If  contains more elements than , construct the hierarchy with similarity ordering as in Equation 4.

Figure 3 (b) shows the icicle plot of tonic prescriptions formulas with the new tree layout. Compared to the original layout (Figure 3 (a)), the alignment of medicines are improved and the same medicines in adjacent columns are aligned vertically. For example, note that how Baizhu (白术), Fuling (茯苓) and Renshen (人参) are aligned as non-sovereign medicines in Figure 3 (b), whereas in Figure 3 (a), such alignments are nonexistent.

### Visualization of Shared Medicines in Formulas

The icicle plot visualization allows effective visualization of the comprising medicines of formulas and the comparison of neighboring formulas. However, comparing formulas that are far apart, e.g., having different sovereign medicines, in the layout is also necessary to provide a deeper understanding of formulas. Therefore, a co-occurrence matrix view of formulas is included to complement the icicle plot. The benefit of using a matrix view is that the complete pair-wise intersection information of all formulas can be effectively represented and easily identified.

As shown in Figure matrix, the matrix has rows and columns as formulas, and an element of the matrix is the number of shared medicines of the corresponding formulas of its row and column. With a sequential color map, this view allows the user to quickly examining the overlapping information of each formula against all others by focusing on a row or a column. Also, the color encoding effectively draws the attention of the user to formulas with the highest number of shared medicines: in this case, Zuoguiwan (左归丸) and Youguiwan (右归丸) as highlighted in red in Figure matrix.

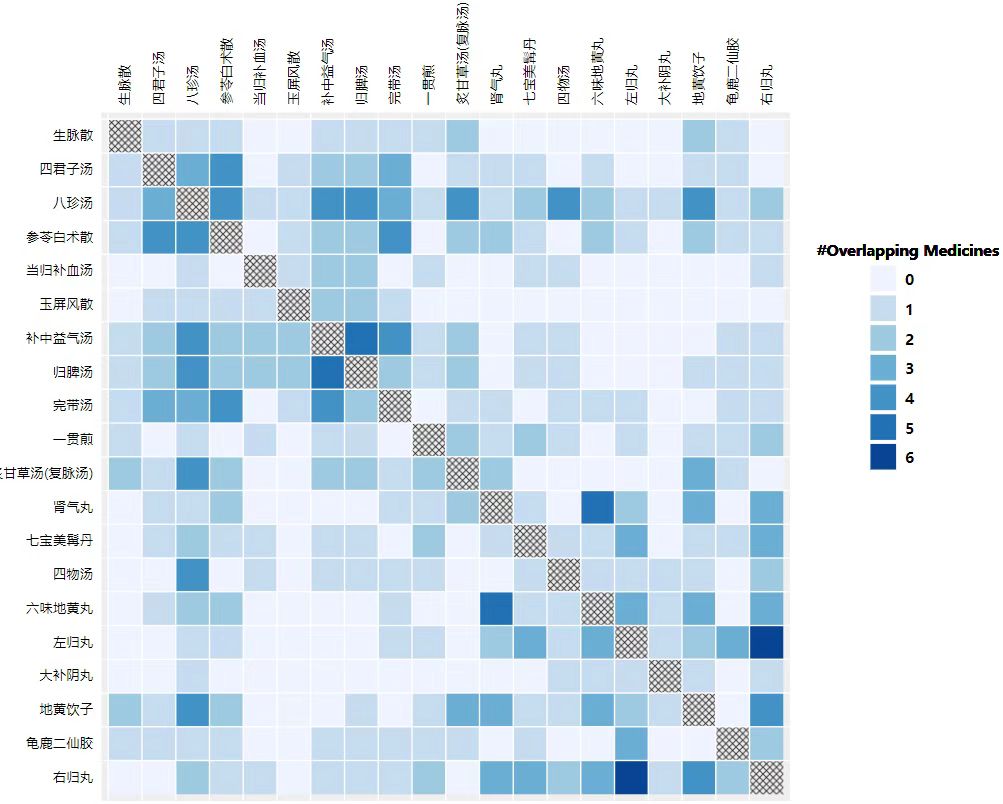


Fig. matrix: the shared medicine matrix view of formulas.

### Perceptual-Guided Data-Driven Color Encoding

Colors are assigned based on the multidimensional data for medicines in both the medicine view and the formula view guided by perception that assists the quick understanding of attributes of medicines and distances between medicines. The workflow of our color encoding method is illustrated in Figure 5. The method is based on the 2D dimensionality reduced space derived from the multidimensional medicine attribute data and requires the knowledge of users to identify representative medicines within it. For a group of medicine formulas, medical experts can identify a number of representative medicines based on their TCM attributes with our TCM-concept inspired colors (R7). These colors are transformed into a perceptual uniform color space and interpolated with radial basis functions in there to get the medicine colors and/or the continuous 2D color map that spans the entire dimensionality reduced attribute space.

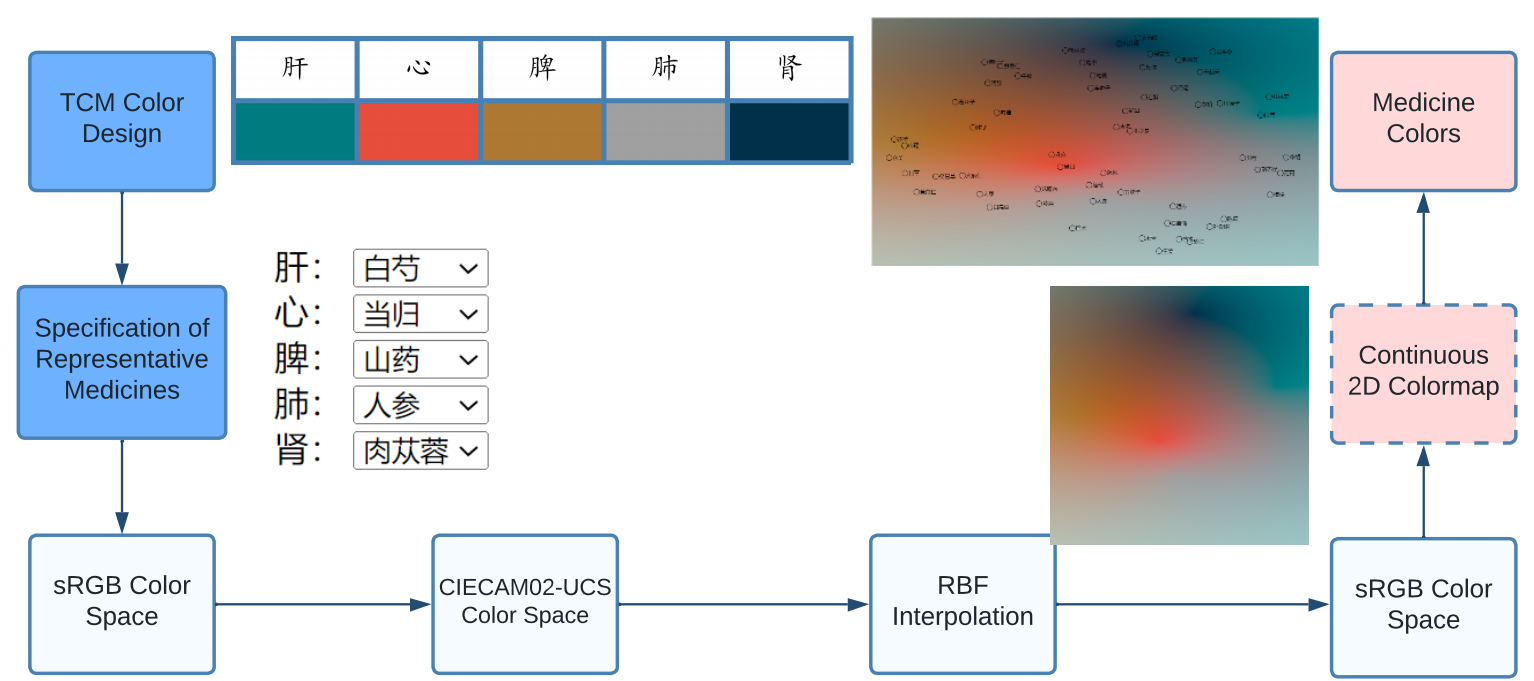


Fig. 5: The pipeline of our color encoding method.

#### TCM-Concept Inspired Representative Color Design

Colors of representative medicine are carefully chosen to show TCM concepts. These TCM concepts include Wuxing (五行), Wuse (五色), and Wuzang (五脏) as summarized in Figure 6. The associated colors are handpicked to show the connection to “Wuse” with perceptual and aesthetic considerations—the luminance of colors should not vary too much, and saturated colors should be avoided. Initially, isoluminant colors that are beneficial for metric comprehension are experimented. However, the TCM expert considers the resulting colors are not distinct enough in the medicine view. As a trade-off, the representative colors are chosen to have a relatively small range of luminance (14 ≤Y ≤ 62).

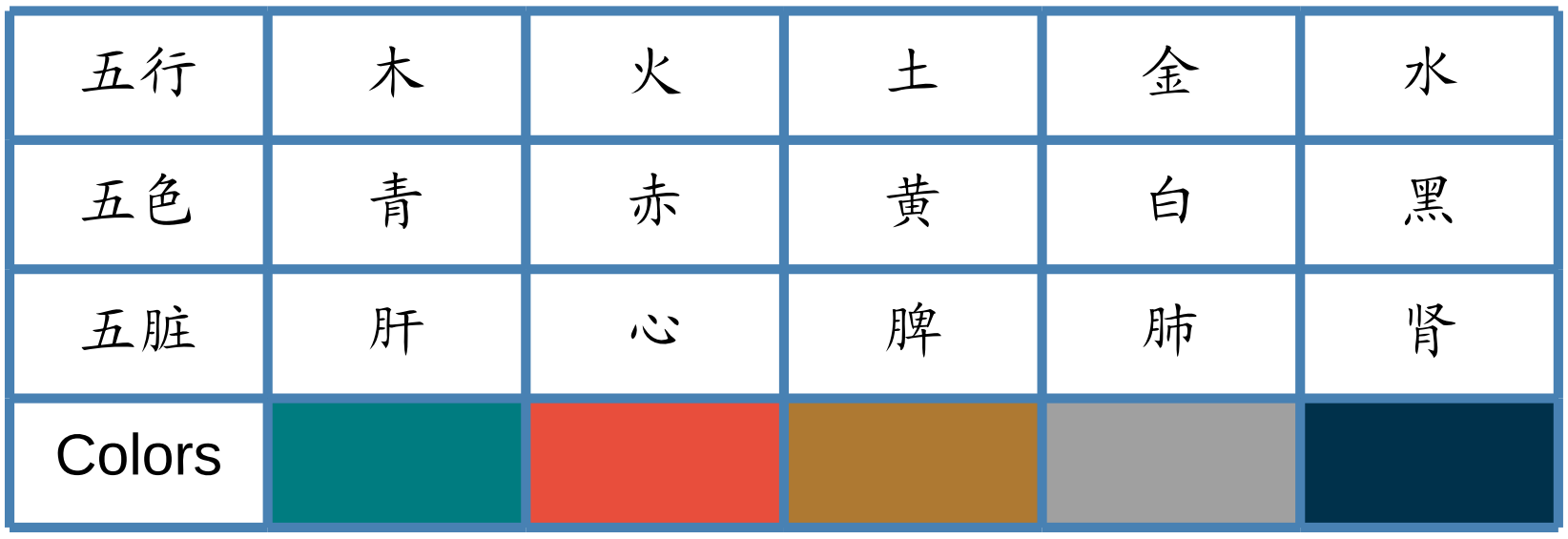


Fig. 6: Colors designed for medicine based on TCM concepts.

#### Perceptual-Uniform Color Space

For perceptual uniformity, we use the CIECAM02-UCS color space[28] to calculate colors of the remaining medicine with color interpolation. This color space is spanned by three parameters  that is calculated by first transforming XYZ color stimuli to the channels of the CIECAM02 color appearance model followed by a transformation from CIECAM02 to CIECAM02-UCS. As shown in Figure 5, we transform the colors of representative medicine from sRGB to CIECAM02-UCS through CIEXYZ. Then, the RBF interpolation is performed for *J*,and  channels, respectively. Next, the interpolated colors are converted back to sRGB for display.

#### RBF Color Interpolation

RBF interpolation enables the interpolation of unstructured data, e.g., a few scattered points or point clouds—making them a natural choice for our method. The RBF interpolation function s(**x**) at location **x** can be written as:

,

where *K* is the number of known data points (*K* = 5 according to Figure 6), **x***k* are the locations of known data points,  is the radial basis function of a distance r between **x** and **x***k* (here, the Euclidean distance), and *wk* are unknown weights. If **x** = **x***k* ,, interpolated values should be data values y at that location: s(**x**) = f (**x**) = y.

We experimented with several radial basis functions, e.g., Gaussian, cubic, thin-plate functions, and choose the linear radial basis function:

.

The choice is made due to two reasons: first, the measure of Euclidean distance which matches the distance of medicines (Section 4.1), and second, it also generates least duplicate colors among typical radial basis functions.

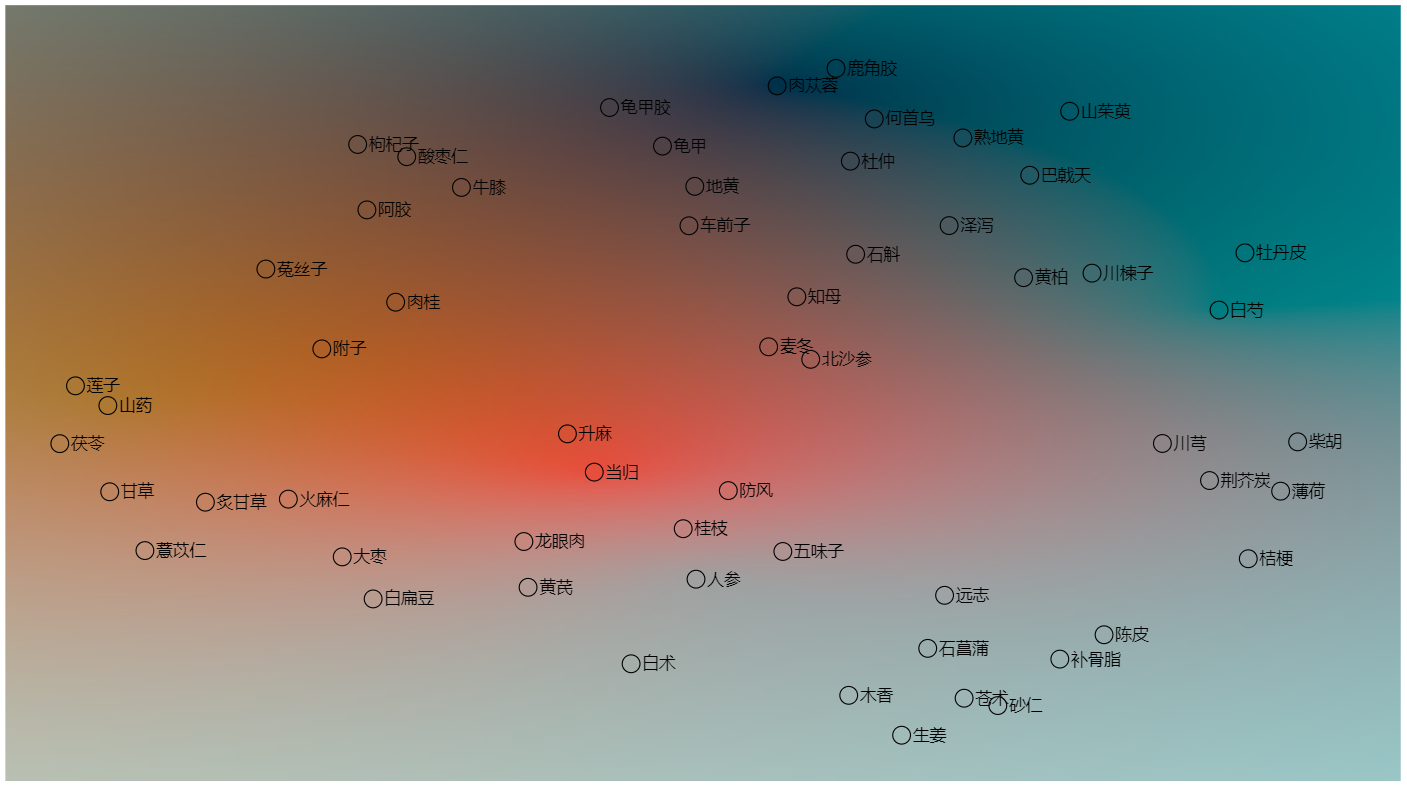
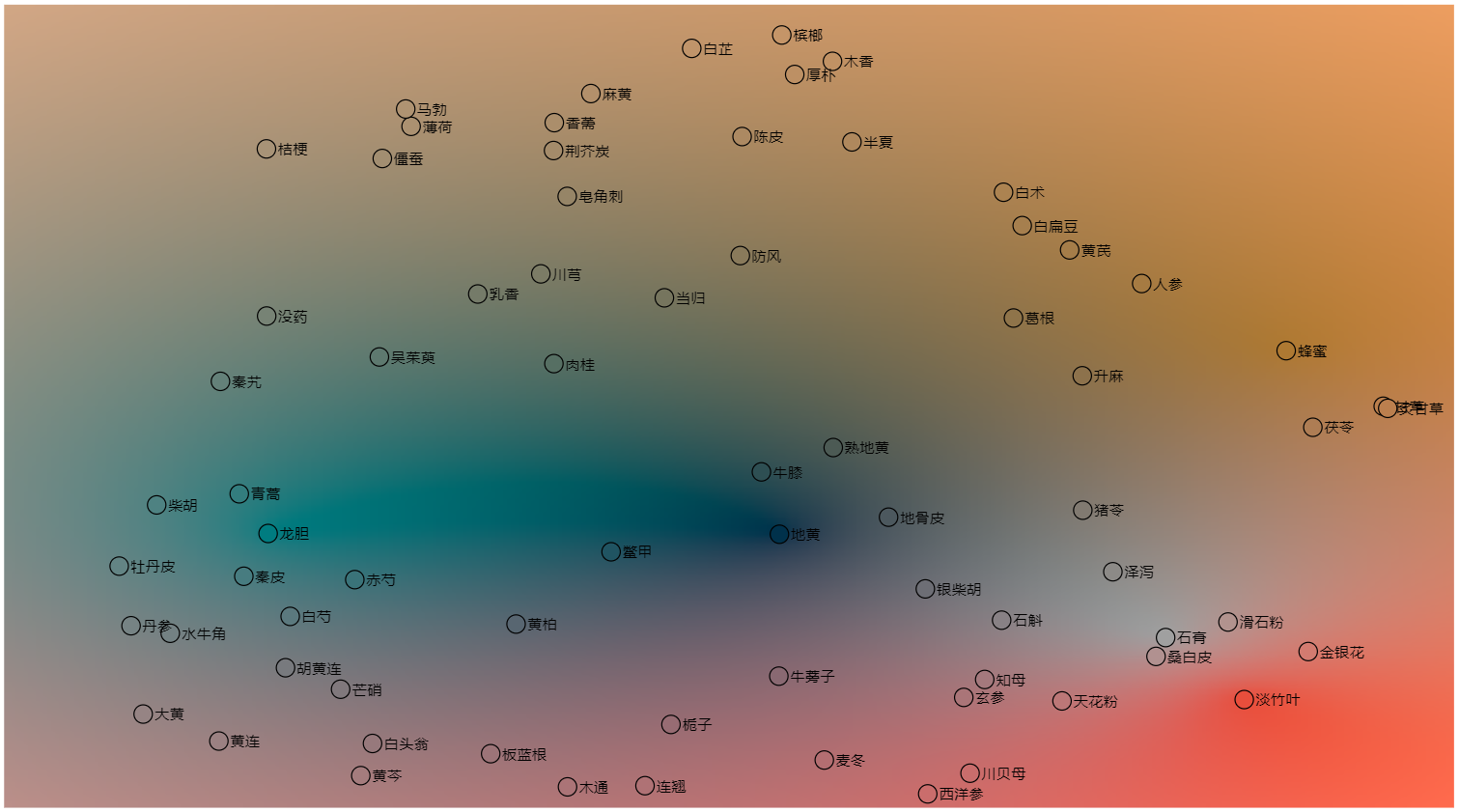
#### Color Assignment

Continuous 2D color maps of two groups of medicine formulas generated by RBF interpolation over the entire 2D domain are shown in Figure 7 (a and b). Smooth transitioning between attributes of medicines can be seen with the 2D color maps, while color differences indicate distances between medicines. Therefore, 2D continuous color maps are a useful tool for examining the distribution of medicines in the multidimensional space of a certain medical formula.

To assign colors to medicines, the 2D location of each medicine in the dimensionality reduced space is used for the interpolation of colors using Equation 6. Medicine colors overlaid on the continuous color map are shown for the two formula groups as shown in Figure 7 (c and d). For efficiency, only colors of points of medicines shown in medicine formulas need to be calculated if the overall trend in the 2D domain is not the focus.

1. (b)

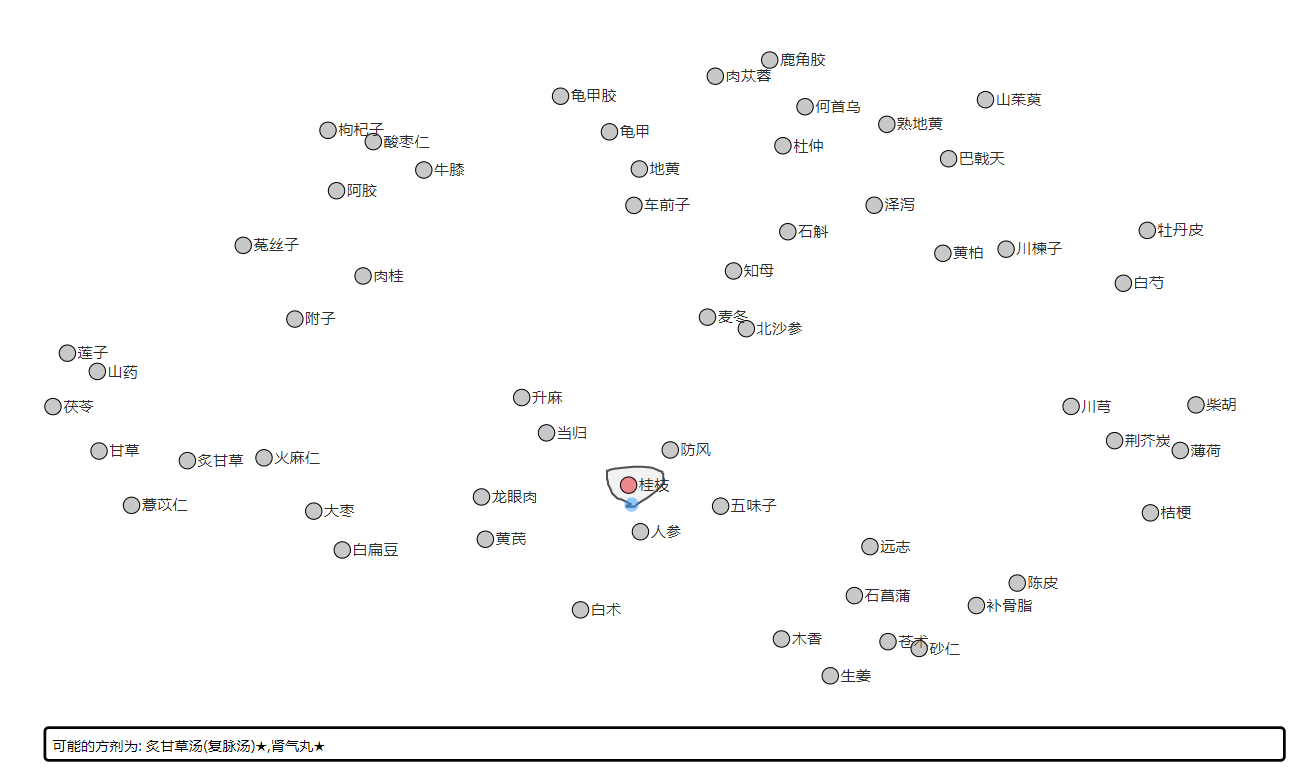
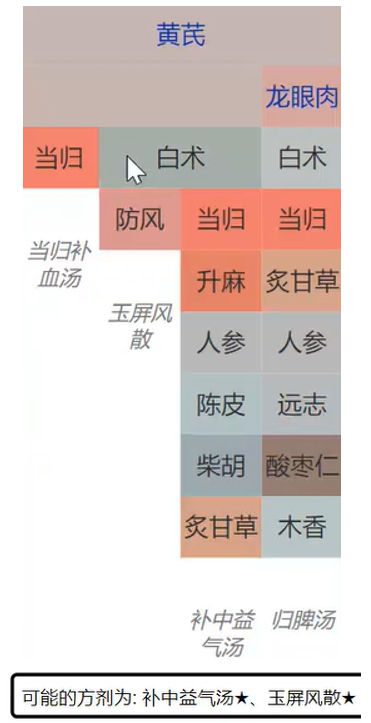
(c) (d)

Fig. 7: Color encoding with our method for tonic prescriptions (the left column) and heat-clearing prescriptions (the right column). Continuous 2D colormaps are shown in (a) and (b), respectively. Medicine colors are calculated based on their positions in the 2D domain (c and d).

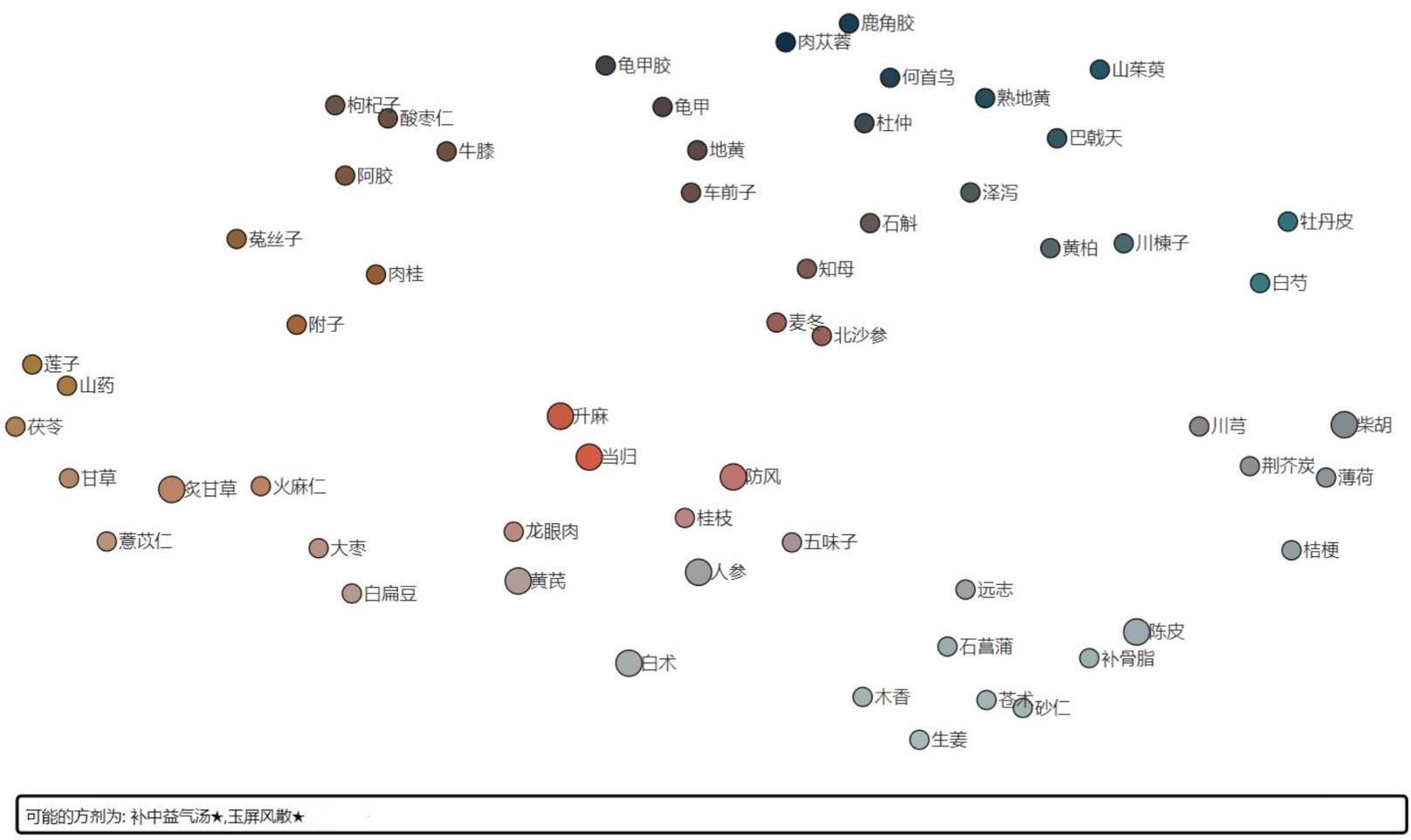
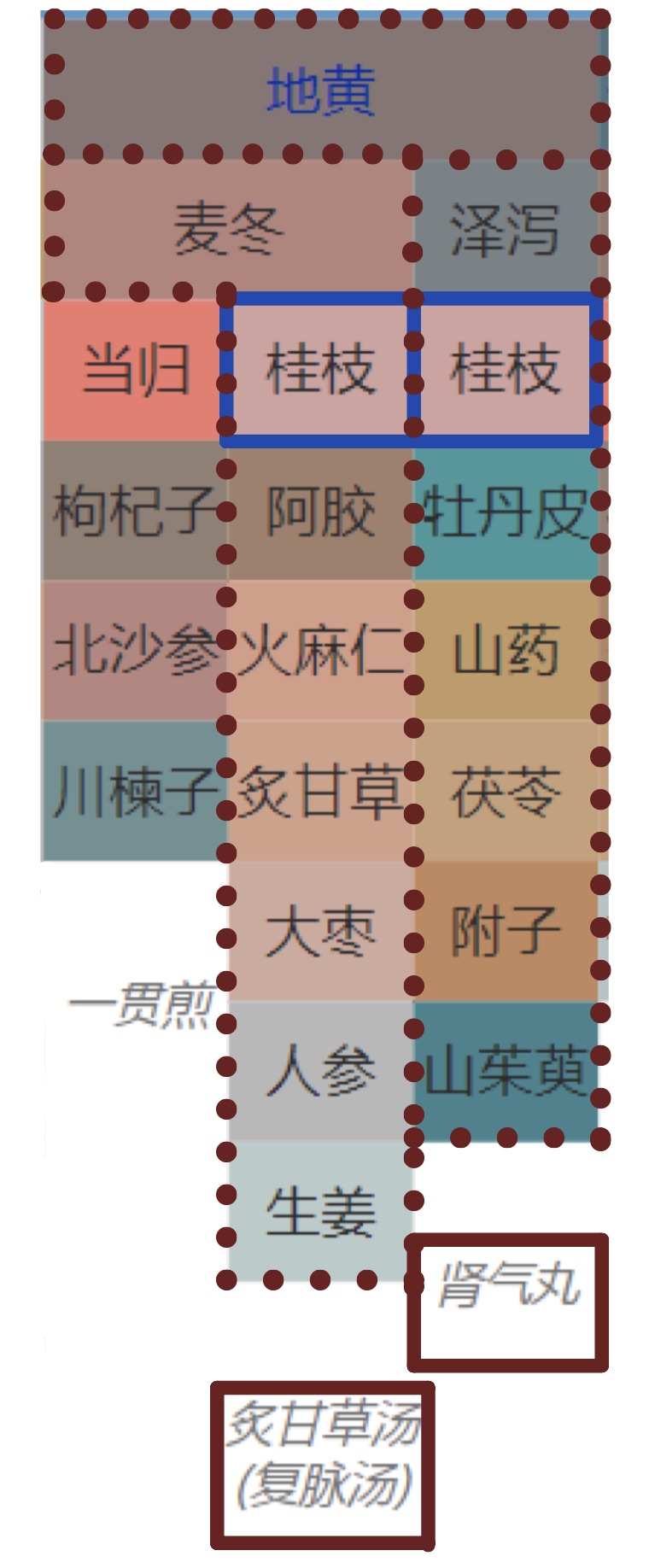
### User Interactions

Our visual analysis method supports interactive exploration within the formula view and the medicine view, and also brushing-and-linking between these two views (R4). In the formula view, names of all formulas are shown whenever the mouse is hovered on a medicine as shown in Figure 8(a). In the medicine view, a lasso tool allows users to flexibly select medicines of interest. All potential formulas are shown as text under the scatterplot of the medicine view (Figure 8(c)). Representative medicines can be assigned and updated through the selection boxes on top of the medicine view (Figures 1 and 9). These user interactions are easy to use and intuitive to users how are not familiar with interactive visualization. Therefore, requirement R6 is satisfied.

Brushing-and-linking enables visual connections between the formula view and the medicine view interactively. All medicines are highlighted in the medicine view with enlarged size (Figure 8(b)) if any formula is selected in the formula view (Figure 8(a)). Conversely, whenever any medicine is selected in the medicine view (Figure 8(c)), the formula view is updated as shown in Figure 8(d). Here, all selected formulas are highlighted with blue solid lines and formulas containing the selected medicines are highlighted with red dashed lines. As a result, brushing-and-linking helps enhancing the understanding of users on the composition of medicine in formulas (R5).



1. (b)

(c) (d)

Fig. 8: User interactions in our method: (a) mouse hovering in the formula view and (b) corresponding updates in the medicine view; (c) lasso selection in the medicine view and (d) corresponding changes in the formula view.

### Implementation

Our method is implemented as a web-based interactive visual analysis tool as shown in Figure 1. Data processing procedures were realized in python aided by the ‘umap’ package for dimensionality reduction, the ‘scipy’ package for RBF interpolation, and the ‘colour’ package for color space transformations, Visualization and user interactions were realized in JavaScript aided by the D3 package, and the communication between python and JavaScript components is achieved using the ‘eel’ package.

## Results

### Overview

Our TCM expert analyzed two groups of typical medicine formulas: namely, tonic prescriptions and heat-clearing prescriptions with our interactive visual analysis tool. The tonic prescriptions contain xxx formulas with yyy herbal medicines, and the heat-clearing prescriptions include zzz formulas with ttt herbal medicines.

### Use Cases

The TCM expert started the analysis by looking at the overall distributions of medicines and used her knowledge to assign representative medicines for each medicine category listed in Figure 6. The resulting continuous 2D colormaps show that the center of the attribute space of tonic prescriptions is red (Figure 7(a)), while heat-clearing prescriptions have the center of its space as green and black (Figure 7(b)). These indicate the different properties of tonic and heat-clearing prescriptions, and are inline with the related TCM concepts.

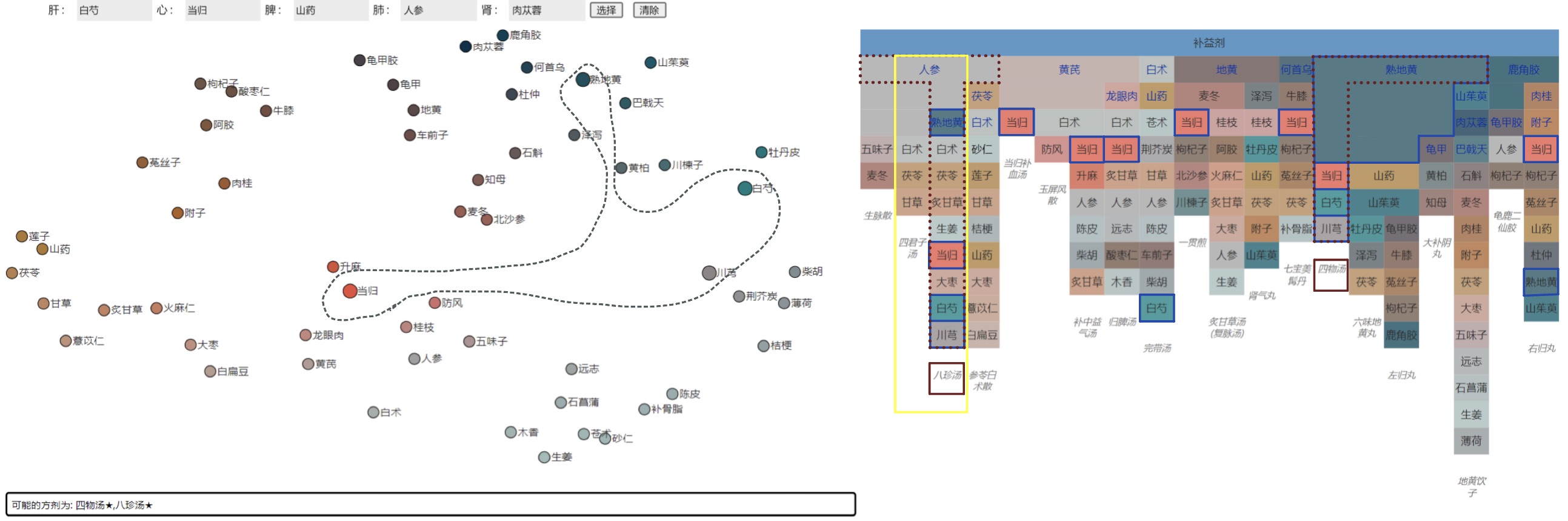


Fig. 1: The web-based tool of our visual analysis method for traditional Chinese herbal medicine formulas. Different aspects of herbal medicine formulas are visualized with two linked views: the medicine view based on dimensionality reduction (left) and the formula view based on an icicle plot with similarity-based tree layout (right). Here, we show a use case of tonic prescriptions.

In the icicle plot of tonic prescriptions (Figure 1(right)), it is easily seen that that two adjacent columns are similar: the Bazhentang (八珍汤) contains the Sijunzitang (四君子汤) as highlighted in the yellow box. The TCM expert then analyzed the difference between these two formulas. She used the lasso tool in the medicine view to select four other medicines in Bazhentang as shown in Figure 1(left). The text below the scatterplot shows that formulas containing these medicines are Bazhentang and Siwutang (四物汤). These two formulas are selected with red dashed lines and the selected medicines are highlighted with blue solid lines in the formula view (Figure 1(right)). A close examination shows that the lasso selected medicines form Siwutang. Moreover, Bazhentang is the combination of Sijunzitang and Siwutang.

The analysis of heat-clearing prescriptions is shown in Figure 9. The TCM expert was interested in Sanhuang (三黄): Huanglian (黄连), Huangqin (黄芩) and Huangbo (黄柏), which is a commonly used medicine combination for clearing heat and detoxification in TCM. The three medicines are relatively close in the medicine view (Figure 9(left)), and the expert used a lasso to select them. Both Huanglian-jiedutang (黄连解毒汤) and Danggui-liuhuangtang (当归六黄汤) contain Sanhuang as suggested by the text below. The expert further examined the formula view (Figure 9(right)) where these two formulas were highlighted. According to the medicine attributes, the function of Huanglian-jiedutang is to clear heat and detoxify. While the composition of Danggui-liuhuangtang contains tonic medicines, meaning that in addition to clearing heat and detoxification, it also has the effect of nourishing Yin (滋阴).

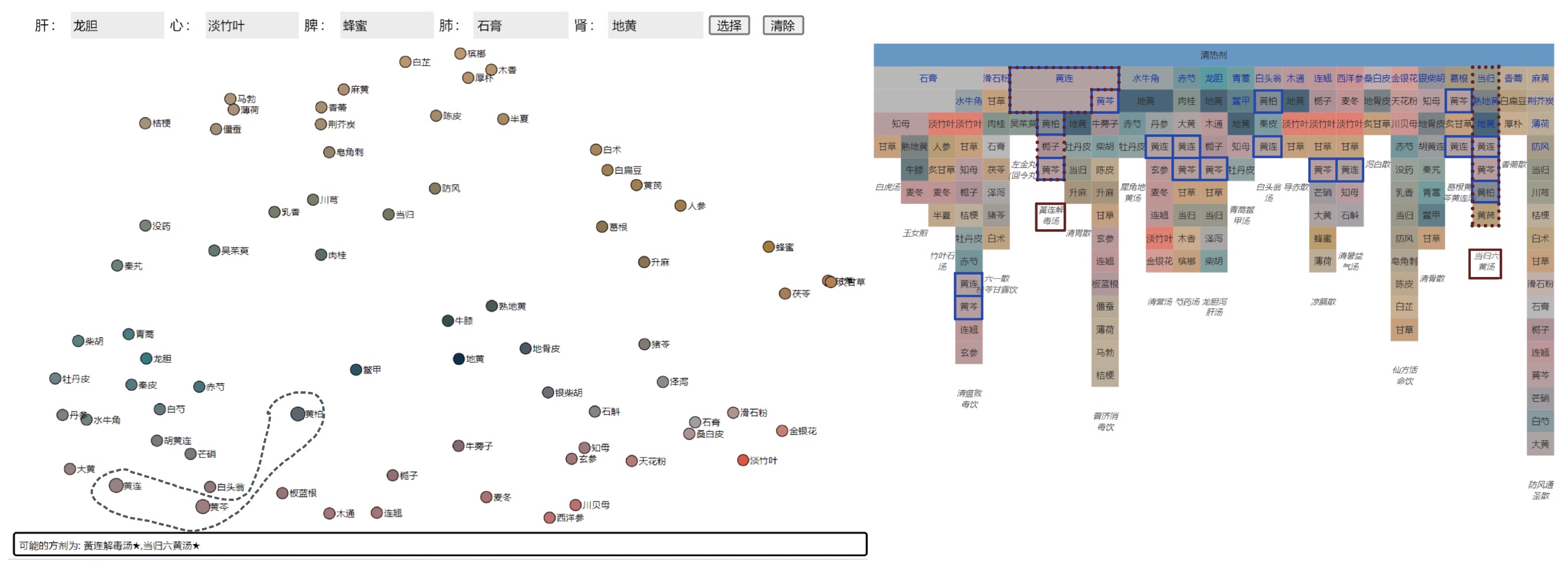


Fig. 9: The analysis of heat-clearing prescriptions with our method.

The expert also made positive comments on the coloring of medicines. For example, Danggui (当归) is a blood tonic medicine and corresponds to red. On the other hand, Shigao (石膏) works on lungs, and is colored white.

Overall, the expert thinks that our method is able to clearly disassemble complex prescriptions and assist the memorization of their functionalities. The interactive visual analysis process is new to TCM students and experts and is helpful for enhancing their understanding of formulas composition theories by making and testing their own hypothesis. The color encoding of medicines allows TCM students and beginners to understand the effect of medicines more intuitively and facilitates the memorization. Moreover, the expert suggests that our method can be extended to a new medical formula design tool.

### TCM Expert Feedback

### User Statistics

These are only examples of possible headings. Please feel free to use different headings to best describe your results.

### Evaluation Outcomes

Please make reference to your Textboxes (Textbox 1), Tables (Table 1), Figures (Figure 1), and Multimedia Appendices (Multimedia Appendix 1) in parenthesis. Please see the examples below for how they should be formatted. Please note the punctuation used in all components, including the caption/title, footnotes etc.

Figures and Multimedia Appendices are uploaded online, while Textboxes and Tables are not uploaded and remain in the body of the manuscript, appearing in the order they are mentioned after the first mention of each Table.

Textbox 1. The caption/title is placed here in a sentence format (capitalization of every word is unnecessary).

|  |
| --- |
| 1. The formatting is actually a 1x1 Table, not an actual “textbox”. 2. Textboxes have no footnotes. 3. Bullet points or numbered lists are allowed in textboxes. |

Table 1. The table caption/title is placed here in a sentence format (capitalization of every word is unnecessary).a-e

|  |  |  |  |
| --- | --- | --- | --- |
|  | Main heading 1 | Main heading 1 | Main heading 1 |
|  | Main heading 2 | Main heading 2 | Main heading 2 |
|  |  |  |  |
| **Subheading** |  |  |  |
| (leave blank) | data | data | data |
| (leave blank) | data | data | data |
| **Subheading** |  |  |  |
| (leave blank) | data | data | data |
| (leave blank) | data | data | data |
| **Subheading** |  |  |  |
| (leave blank) | data | data | data |
| (leave blank) | data | data | data |

aNot all elements are necessary for every table, simply omit the irrelevant sections for your table and keep the formatting of the rest. For further details, please refer to the main Instructions for Authors of JMIR document.

bFootnotes are labeled in superscript lower case a-z. Other symbols are not used.

cAstericks (\*) can only be used if exact *P* values cannot be provided for a specific reason, and are listed after the superscript a-z footnotes.

dplease be conscious of the overall width of the table. Tables will be automatically fitted/resized to the width of a US Letter Small page in portrait configuration during typesetting. Overcrowded Tables or Tables that are too crowded WILL look squished, and should be avoided if possible.

elonger headings can be abridged within the Table, with a full explanation in a footnote.

Figure 1. Captions/titles are inserted online. Try to use Times New Roman for text within the Figure to match the font of the final typeset manuscript when possible. These should be .jpeg or .png files. Please prepare Figures with good resolution – Figures that are predominantly graphics/pictures should have dpi close to 300, while those that are text-dominant can have lower resolution (usually dpi 200). Try to use combinations of color and symbols/line styles to define and refer to different categories. This will help with readability if Figures are printed/viewed in black and white.

## Discussion

### Principal Results

Visualization techniques are able to show TCM formulas information with interactive graphics. With carefully designed visualization methods

Two uses cases of typical groups of medicine formulas analyzed by the TCM expert demonstrate the effectiveness of our method for medicine formula composition learning and TCM inheritance. The expert also suggests that our method could be potentially used for designing new formulas.

### Limitations

Our method does not directly support the visualization of the overlapping of more than two medicine formulas, i.e., intersections of more than two sets. However, such information can be implicitly gained by visual searching in the medicine formula view and by interactively selecting medicines of interest which would highlight all formulas containing the shared medicines.

Another limitation is that the dimensional reduction view does not explicitly show the multidimensional properties but the relative distances between medicines. This could be addressed using additional multidimensional visualization techniques such as parallel coordinates.

### Comparison with Prior Work

Our method is related to TCM visualization in general. Important aspects of our method include set visualization, multidimensional data visualization and color perception.

#### TCM Visualization

Visualization methods are used in the area of TCM. Visual recognition and visualization is proposed for TCM pulse information—the pulse information is quantified and visualized to support more accurate diagnosis [6]. Digital tongue images that are important in TCM are recognized and analyzed with a visualization of tongues[7]. Infrared thermal imaging visualization enables users to see and assess physiological states or pathological conditions intuitively as the temperature of local tissues or the whole body may change due to illness [8]. Visualization based on a 3D human model of Chinese medicine pulses could facilitate the teaching, understanding, and communication of meridians and acupoints[9].

Specific visualization methods for herbal medicine formulas are available. Cold and hot properties are visualized as indicators for herbal medicine formulas in a formula analysis platform[4]. However, this method covers only two properties and does not show the multidimensional attributes of medicines. Knowledge graph visualization is proposed for a large number of medicine formulas through manual processing and natural language processing[3]. Network visualization is used for showing the composition of medicine formulas to assist the construction of medicine formulas databases[5]. However, these methods rely on querying based on text input and do not support interactive visual analysis.

Moreover, most if not all works above are not done through collaborations between TCM and visualization experts. In contrast, a visual analysis method of TCM health records is available recently as a result of the collaboration between TCM and visualization experts[10]. This method supports the analysis of time-varying TCM health records and comparing medicines in prescriptions of different patients. However, a visual analysis method for herbal medicine formulas is not available yet.

#### Set Visualization

Set visualization is an active research area. A survey on set visualization techniques is given by Alsallakhi et al. [3]. Different strategies are employed to visualize set memberships, including Euler and Venn diagrams [4, 27, 30, 36–38, 46], node-link diagrams [13, 28, 39], matrix-based methods [22, 26, 34], and aggregation methods [2, 17, 21]. Note that these methods are dedicated to show the relationship between elements and sets without concerning the context of the data, for example, locations of elements on a geological map or a timeline

#### Hierarchical and Multidimensional Data Visualization

Hierarchical data visualization techniques can be classified into explicit techniques, i.e., trees using node-link diagrams, and implicit techniques that no explicit edges are drawn. Implicit hierarchy visualization techniques are summarized in an extensive survey[11], and the main benefit of implicit techniques is the efficient use of space making them more suitable for large hierarchical data than trees. Popular implicit methods include treemaps[12, 13] and icicle charts[14]. Treemaps use nested sets to show hierarchies by placing the graphical representation of children inside that of their parent. Typically, treemaps are drawn as rectangles[12, 13], and variants of circles and even irregular shapes formed by Gosper curves exist[15]. Research efforts are made to improve the perception[16, 17]or to encode more information, for example, uncertainty[18], while maintaining the compactness of the treemaps. Icicle plots utilize the vertical placement of nodes to show the hierarchical information. Compared to treemaps, icicle plots are not as space-efficient but seem to be more intuitive from our informal empirical study with domain experts. In our case, we choose the icicle plot for hierarchy visualization as our TCM expert considers it easier to understand and allows for quick comparison of formulas.

Multidimensional data can be effectively visualized using dimensionality reduction techniques[19]. Nonlinear dimensionality reduction methods[20] are more suitable to preserve complex high- dimensional structures than linear methods[21]. Currently, t-SNE[22] and UMAP[23] are the most popular nonlinear dimensionality reduction methods as they could preserve the neighboring information in the high-dimensional space. We choose UMAP in our method as it is more efficient and overcomes several limitations of t-SNE.

#### Color Perception in Visualization

Perception is important in visualization, and color perception, among others, is most relevant in our case. With numerous medicines and formulas, colors provide critical classification information that assist the users for the reasoning of formulas composition. A survey of the use of colors in visualization can be found elsewhere[24]. A key concept for effective use of colors is perceptual uniformity, i.e., the perceived color difference should match the data value difference. Perceptual uniformity is used for color map design[25, 26], and an online tool ColorBrewer is readily available[[1]](#footnote-1), but these methods do not support flexible color specification for 2D color maps. Instead, we propose a color encoding method for drugs based on a 2D color map created by RBF interpolation with user-specified representative drugs assigned with TCM concept-inspired colors. To achieve perceptual uniformity, colors have to be computed in a perceptual uniform color space. CIELab is perhaps the most well known perceptual uniform color space[27]. However, studies show that the uniformity performance of CIELab is not very good[28]. Recently, several color spaces based on the CIECAM02 color appearance model[29] with better uniformity than CIELab are available. In our method, we choose the CIECAM02- UCS color space[28] for its good performance.

### Conclusions

We have introduced a visual analysis method for TCM formulas. Our method supports the visualization of medicine formulas data in a formula view with a similarity-based layout complemented by a shared medicine matrix view, and multidimensional attribute data of medicines are visualized in a dimensionality reduction-based medicine view. Requirements and design choices of our method are made through a close collaboration between visualization and TCM experts in an iterative quick prototyping fashion. Effective comparison of medicine formulas is supported with our new similarity-based layout algorithm; colors of visual elements are assigned with a perceptual-guided data-driven color encoding method that focuses perceptual uniformity and TCM concepts of medicine attributes. Interactive analysis of medicine formulas and corresponding medicines is available with brushing-and-linking between the two views. Two uses cases of typical groups of medicine formulas analyzed by the TCM expert demonstrate the effectiveness of our method for medicine formula composition learning and TCM inheritance. The expert also suggests that our method could be potentially used for designing new formulas.

### Acknowledgements

This work was supported in part by …

### Conflicts of Interest

None declared.

### Abbreviations

JMIR: Journal of Medical Internet Research

RCT: randomized controlled trial

## Multimedia Appendix 1

Multimedia appendices are supplementary files, such as a PowerPoint presentation of a conference talk about the study, additional screenshots of a website, mpeg/Quicktime video/audio files, Excel/Access/SAS/SPSS files containing original data (very long tables), and questionnaires. See <https://jmir.zendesk.com/hc/en-us/articles/115003396688> for further information. Do not include copyrighted material unless you obtained written permission from the copyright holder, which should be uploaded together with your Publication Agreement form as supplementary file.

The Multimedia Appendices must be uploaded online, accompanied by a caption. CONSORT-EHEALTH checklists are always uploaded as Multimedia Appendices. Although this is primarily intended for randomized trials, the section of the checklist describing how an intervention should be reported is also relevant for manuscripts with other evaluation designs.

Before submission, authors of RCTs must **fill in the electronic CONSORT-EHEALTH questionnaire at**[**http://tinyurl.com/consort-ehealth-v1-6**](http://tinyurl.com/consort-ehealth-v1-6) with quotes from their manuscript (if you wish to comment on the importance of the items from the checklist for reporting, please also rate each item on a scale between 1-5). BEFORE you press submit, please generate a pdf of the form with your responses and upload this file as supplementary file entitled CONSORT-EHEALTH V1.6.

## References

[1] J. Wang, *Basic Theory of Traditional Chinese Medicine*. Beijing: China Press of Traditional Chinese Medicine, 2016.

[2] X. Gao, *Chinese Pharmacy*. Beijing: The Press of Traditional Chinese Medicine, 2004.

[3] W. Guo, “Research and Implementation of Knowledge Mapping of Traditional Chinese Medicine Prescription,” *Lanzhou University*, 2019.

[4] J. Gao, “Construction of Visual Analysis Platform for Cold and Heat Properties of Formulae Based on Quantitative Study,” Ph.D. dissertation, Beijing University of Chinese Medicine, 2009.

[5] Z. Yan, G. Bo, and C. Meng, “Design and implementation of the analysis system of TCM prescription,” *China Journal of Traditional Chinese Medicine and Pharmacy*, vol. 29, no. 5, p. 4, 2014.

[6] A. C. Y. Tang, “Review of traditional chinese medicine pulse diagnosis quantification,” in *Complementary Therapies for the Contemporary Healthcare*, M. Saad and R. de Medeiros, Eds. Rijeka: IntechOpen, 2012, ch. 4. [Online]. Available: https://­doi.org/­10.5772/­50442

[7] J. Xie, C. Jing, Z. Zhang, J. Xu, Y. Duan, and D. Xu, “Digital tongue image analyses for health assessment,” *Medical Review*, vol. 1, no. 2, pp. 172–198, 2021. [Online]. Available: https://­doi.org/­10.1515/­mr-2021-0018

[8] A. Ovechkin, S.-M. Lee, and K.-S. Kim, “Thermovisual evaluation of acupuncture points,” *Acupuncture & electro-therapeutics research*, vol. 26, no. 1-2, pp. 11–23, 2001.

[9] M. Wei, Z. Chen, G. Chen, X. Huang, Y. Jin, K. Lao, Z. Li, S. Li, F. Zhong, H. Liang *et al.*, “A portable three-channel data collector for chinese medicine pulses,” *Sensors and Actuators A: Physical*, vol. 323, p. 112669, 2021.

[10] X. Hu, S. Peng, H. Hou, N. Yang, Y. Lv, and l. Zhou, “Visual Analysis of Traditional Chinese Medicine Health Records,” *Journal of Computer-Aided Design and Computer Graphics*, vol. 33, no. 12, p. 10, 2021.

[11] H.-J. Schulz, S. Hadlak, and H. Schumann, “The design space of implicit hierarchy visualization: A survey,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 4, pp. 393–411, 2011.

[12] B. Johnson and B. Shneiderman, “Tree-maps: a space-filling approach to the visualization of hierarchical information structures,” in *Proceeding Visualization ’91*, 1991, pp. 284–291.

[13] B. Shneiderman, “Tree visualization with tree-maps: 2-d space-filling approach,” *ACM Trans. Graph.*, vol. 11, no. 1, p. 92–99, 1992. [Online]. Available: https://­doi.org/­10.1145/­102377.115768

[14] J. B. Kruskal and J. M. Landwehr, “Icicle plots: Better displays for hierarchical clustering,” *The American Statistician*, vol. 37, no. 2, pp. 162–168, 1983. [Online]. Available: http://­www.jstor.org/­stable/­2685881

[15] D. Auber, C. Huet, A. Lambert, B. Renoust, A. Sallaberry, and A. Saulnier, “Gospermap: Using a gosper curve for laying out hierarchical data,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 11, pp. 1820–1832, 2013.

[16] J. Van Wijk and H. Van de Wetering, “Cushion treemaps: visualization of hierarchical information,” in *Proceedings 1999 IEEE Symposium on Information Visualization (InfoVis’99)*, 1999, pp. 73–78.

[17] M. Bruls, K. Huizing, and J. J. van Wijk, “Squarified treemaps,” in *Data Visualization 2000*, W. C. de Leeuw and R. van Liere, Eds. Vienna: Springer Vienna, 2000, pp. 33–42.

[18] J. Grtler, C. Schulz, D. Weiskopf, and O. Deussen, “Bubble treemaps for uncertainty visualization,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 719–728, 2018.

[19] L. van der Maaten and G. Hinton, “Visualizing data using t-sne,” *Journal of Machine Learning Research*, vol. 9, no. 86, pp. 2579–2605, 2008. [Online]. Available: http://­jmlr.org/­papers/­v9/­vandermaaten08a.html

[20] J. A. Lee and M. Verleysen, *Nonlinear Dimensionality Reduction*. New York: Springer-Verlag, 2007.

[21] J. P. Cunningham and Z. Ghahramani, “Linear dimensionality reduction: Survey, insights, and generalizations,” *Journal of Machine Learning Research*, vol. 16, no. 1, p. 2859–2900, 2015.

[22] E. P. L. van der Maaten and H. van den Herik, “Dimensionality reduction: A comparative review,” Tilburg University, Tech. Rep., 2009.

[23] L. McInnes, J. Healy, and J. Melville, “Umap: Uniform manifold approximation and projection for dimension reduction,” 2018.

[24] L. Zhou and C. D. Hansen, “A survey of colormaps in visualization,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 8, pp. 2051–2069, 2016.

[25] P. K. Robertson and J. F. O’Callaghan, “The generation of color sequences for univariate and bivariate mapping,” *IEEE Computer Graphics and Applications*, vol. 6, no. 2, pp. 24–32, 1986.

[26] H. Levkowitz and G. Herman, “Color scales for image data,” *IEEE Computer Graphics and Applications*, vol. 12, no. 1, pp. 72–80, Jan 1992.

[27] CIE, “Colorimetry, 4th edition,” Commision Internationale de l’Eclairage, Tech. Rep., 2018.

[28] M. R. Luo, G. Cui, and C. Li, “Uniform colour spaces based on ciecam02 colour appearance model,” *Color Research & Application*, vol. 31, no. 4, pp. 320–330, 2006. [Online]. Available: https://­onlinelibrary.wiley.com/­doi/­abs/­10.1002/­col.20227

[29] N. Moroney, M. Fairchild, R. Hunt, l. Changjun, M. Luo, and T. Newman, “The ciecam02 color appearance model,” vol. 10, 2002, pp. 23–27.

[30] Y. Xuming, Q. Mingyuan, L. Qian, C. Li, Y. Zhongyi, and Y. Lin, “Information integration research on cumulative effect of ‘siqi, wuwei, and guijing’ in traditional chinese medicine,” *Journal of Traditional Chinese Medicine*, vol. 36, no. 4, pp. 538–546, 2016. [Online]. Available: https://­www.sciencedirect.com/­science/­article/­pii/­S0254627216300723

[31] J. Li, *Chinese Herbal Formulas*. Beijing: The Press of Traditional Chinese Medicine, 2016.

[32] Y. Wu, F. Zhang, K. Yang, S. Fang, D. Bu, H. Li, L. Sun, H. Hu, K. Gao, W. Wang *et al.*, “Symmap: an integrative database of traditional chinese medicine enhanced by symptom mapping,” *Nucleic acids research*, vol. 47, no. D1, pp. D1110–D1117, 2019.

## References

1. DO NOT use italics, periods after authors’ initials, and periods after journal abbreviations.
2. DO use a semicolon (;) after a journal title before the year, put volume number in parenthesis, and use a colon (:) before the page numbers.
3. Titles should be in sentence case (do NOT capitalize the first letter of every word).
4. Do not use the footnotes tool to generate the reference list.
5. **Cite only *published* or *accepted* (“in print”) works**. *Submitted* papers (not *accepted*) documents not widely available (personal emails, letters), or oral communications (unless they are published abstracts) should NOT be cited as references. Cite these in the main body of text as **“personal communication by NAME, DATE”** after obtaining permission from the communicator to quote his communication.
6. **Remove OLE elements** from reference management softwares such as Endnote and Reference Manager. Select the entire document (Ctrl+A or Command A), remove field codes (*Ctrl+Shift+F9* or *Command+6*). This is important for correct parsing of your reference list using RefCheck during copyediting. This is an automatic process, but please check for completeness and accuracy of parsed fields for each reference when prompted during copyediting steps after acceptance of your manuscript.
7. **Journal Articles** (examples following)**:** append the PubMed Identifier(**PMID**, eg, "PMID:1234567", where 1234567 is the PubMed identifier) **or DOI** (digital object identifier, eg,doi:10.1136/bmj.331.7529.1391) **after each reference**. Alternatively (as per our old instructions) you could append a [[Medline](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=9062335&dopt=Abstract)] link after each reference, linking to the PubMed abstract of the article you are citing. You may check whether a DOI is correct using the DOI resolver at <http://dx.doi.org/>.
8. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. JAMA 1997;277:927-934. PMID:9062335
9. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. JAMA 1997;277:927-934. [[Medline](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=9062335&dopt=Abstract)]
10. **Websites and Web articles (URLs)** (example following) should be cited as **"webcited®"** references in the reference section at the end of the manuscript - **do not include links to websites in the text**. To webcite® a web reference means to take a snapshot of the cited document and to **cite the archived copy (WebCite link) in addition to the original URL**. JMIR now *requires* that authors use the [WebCite](http://www.webcitation.org/) ® [technology (www.webcitation.org)](http://www.webcitation.org/) to archive cited web references first before they cite them. Do **not** cite uncached "live" webpages and websites in the article or reference section, unless archiving with WebCite has failed. Provide both the original URL and the WebCite link. Note that journal articles in electronic formats are journal articles, not a web reference.
11. Fox S, Fallows D. 2003. Internet Health Resources. http://www.pewinternet.org/pdfs/PIP\_Health\_Report\_July\_2003.pdf. Archived at: http://www.webcitation.org/5I2STSU61
12. **For books, please add the ISBN, if known** (no blanks). (<http://isbndb.com/>; examples below)
13. Iverson CL, Flanagin A, Fontanarosa PB, et al. American Medical Association Manual of Style: A Guide for Authors and Editors. 9th edition. Baltimore, Md: Williams & Wilkins; 1998. ISBN:0195176332
14. Phillips SJ, Whisnant JP. Hypertension and stroke. In: Laragh JH, Brenner BM, editors. Hypertension: pathophysiology, diagnosis, and management. 2nd ed. New York: Raven Press; 1995. p. 465-78.
15. **Conference Proceedings** (example below).If conference proceedings are available through Medline, please use the Medline citation.
16. Kimura J, Shibasaki H, editors. Recent advances in clinical neurophysiology. Proceedings of the 10th International Congress of EMG and Clinical

1. https://colorbrewer2.org [↑](#footnote-ref-1)