

Linking Traditional Chinese Medicinal Herbs to Cancer Related Pathways

By
Lawrence Hsu

A Captstone

Presented to the Department of Medical Informatics and Clinical Epidemiology
and the Oregon Health & Science University
School of Medicine in partial
fulfillment of the requirements
for the degree of

Master of Biomedical Informatics

June 2018

TABLE OF CONTENTS

Introduction and Purpose.....	5
Background.....	6
Database.....	10
Methodology Journey.....	12
Reflection.....	19
References	20
Appendix	23

List of Illustrations

Figure 1.....	12
Table 1.....	17

ACKNOWLEDGMENTS

I'd like to take the time to thank my academic advisor, Dr. Eilis A. Boudreau, for her support and guidance on this capstone project as well as her assistance through the OHSU graduate guidance. I had a wonderful time working on this project learning to discuss with other researchers, adjusting objectives due to obstacles found through the initial phase exploring the dataset. I am very fortunate to have had the opportunity to work with her and I am extremely grateful for her encouraging and calming nature when things became stressful.

I'd like to take the time to thank my academic advisor, Dr. Eilis A. Boudreau, for her support and guidance on this capstone project as well as her assistance throughout my time in the OHSU graduate program. I had a wonderful time working on this project learning to discuss with other researchers, and adjusting objectives due to obstacles found through the initial phase exploring the dataset. I am very fortunate to have had the opportunity to work with her and I am extremely grateful for her encouraging and calming nature when things became stressful.

I would also like to thank Dr. Ted Laderas for teaching me the process of developing a data visualization through R Shiny and all the great things it can do for the future of this research. I would also like to thank all of the students, faculty and staff members in the bioinformatics and medical informatics program at OHSU for being such caring and wonderful people to talk to and work with.

ABSTRACT

The purpose of this project is to link the traditional Chinese medical herbs to cancer-related pathways as proof of concept. There is evidence that compounds in Chinese

medical herbs have a synergetic effect when used together rather than just using them alone. The goal of this project was to facilitate understanding of these synergetic effects. This work may assist physicians when they are developing a protocol to supplement their patient's treatment with traditional Chinese medicinal herbs. . The secondary goal is to gain experience with data visualization. Using information from public available databases and R studio, a heatmap was produced to show which hallmark cancer pathways were targeted by the chemical compounds contained in these herbs. Cancer hallmarks are the six traits that govern the transformation of a normal cell to a malignant one and each hallmark consists of pathways related to those traits. The six common traits are sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, and activating invasion and metastasis.¹The progress to reaching this goal was hampered by several obstacles including: lack of access to all of the information contained in key Chinese herb databases, and the need to adjust the project objectives in accordance with the available data. In the end, I was able to develop a heatmap using only a small subset of chemical compounds targeting three cancer hallmarks: inflammatory, angiogenesis, and sustaining proliferative signaling. The inflammatory hallmark was suggested by Dr.Laderas as it was a common feature among all cancers. The other two hallmarks were chosen at random.

Introduction and Purpose

Since the early 2000s, the use of Complementary and Alternative Medicine (CAM) has grown worldwide¹. Currently, 30-60% of patients use CAM^{2,3}. However, some researchers believe the number of patients using CAM is higher than is being disclosed as many patients may not disclose their use of CAM for fear of their treating provider's disapproval.⁴ The other problem is most of these remedies are publicly available through the internet and many of these concoctions have no evidence to back up their claims, allowing patients access to potentially harmful treatments⁵. This is also a problem with Traditional Chinese Medicine (a subcategory of CAM). One potential solution to this problem is to investigate the ingredients in these remedies and identify their potential biological targets. Recent advances in TCM include investigating the metabolites found in TCM and incorporating them into the western biomedical approach. One path of interest that researchers are looking into is how compounds may act synergistically to create a greater effect. A better understanding of how these compounds work synergistically may permit the development of better treatments that have fewer side-effects and are less likely to interfere with traditional cancer treatments that patients are prescribed. The other benefit of this would allow physicians to have an open dialogue with patients using CAM and advise them on potential benefits and harms.

Capstone Project

The purpose of this capstone project is to link the herbal ingredients (chemical compounds) found in plants to the cancer-related biological pathways they may impact.

The main objective is to create a database where researchers may use this information, hopefully gaining insight on therapeutic benefits and developing a tool for future research in this area. The secondary objective is to inspire future researchers to begin investigating “full plants effect.” The full plants effect is an phenomena whereby compounds in a plant act synergistically to create a stronger effect than when used separately⁶. This project includes an overview of issues encountered, how the problem was resolved and if it was not resolved, what other approaches were tried.

Background

Complementary and Alternative Medicine (CAM)

Complementary and Alternative Medicine (CAM) is used worldwide¹. According to a survey conducted by NIH⁷ in 2008, about 38% of people over the age of 18 used some form of CAM, primarily to alleviate chronic diseases symptoms. CAM, defined by the National Center for Complementary and Alternative Medicine⁷, is a group of diverse medical practices that are not part of the conventional medicine. The “alternative medicine” part of the term implies “instead of,” and the “complementary” part implies “to support conventional medicine.” Some examples of practices include yoga, acupuncture, and herbal medicine. The biggest issue with CAM is many of these treatments have little to no scientific evidence to back up their validity⁵. The reasons why people continued to use CAM as a form of treatment is complex. Patients may use CAM based on cultural preferences and/or social context. Patients afflicted with a severe illness like cancer may have the mentality of “leave no stone unturned” which drives them to use CAM when modern medicine cannot help.

Traditional Chinese Medicine (TCM)

TCM uses remedies derived from plants and animals, as opposed to being made in a laboratory. As mentioned above, TCM is often used as alternative medicine by those who are unable to afford treatment from health care providers. Very similar to the situation with all of alternative medicine, many of these treatments have unfounded claims which could be harmful to patients. Many of the compounds, particularly remedies made from animals' extremities and organs, are not properly processed. Cooper et al.⁸ found evidence of lead, arsenic and other metalloids in selected TCM remedies. To combat this problem, researchers have begun investigating the chemical compounds in TCM treatments.

The medical field has been using metabolomics not only to research TCM but connect it to Western Medicine. Metabolomics is the study of metabolites found within an organism, cell, or tissue⁹. TCM involves a lot of eating of certain foods to create a healthy lifestyle. Indirectly, by researching and connecting TCM to modern approaches, it may be possible to incorporate TCM practices into more standardly used western medical approaches.

TCM has been used for thousands of years and some TCM approaches already mirror western practices. For example, van der Greef et al.¹⁰ investigated how TCM practitioners treat rheumatoid arthritis. TCM practitioners ask for the patient's medical history, perform a physical checkup and then compare the results to past observations to

determine the treatment plan. However, TCM categorizes into different syndromes, called Bi-syndromes.”¹⁰“Bi” stands for blockage or obstruction of Qi and blood and what causes “Bi-syndromes” is an external attack by four different factors: “Heat”, “Cold”, “Dampness”, and “Wind”. These external attacks affect the “Fu organs” or “Five Tissues”. “Fu organs” or “Five Tissues” refer to the different body systems (ie central nervous system, digestive system, cardiovascular system etc). Furthermore, rheumatoid arthritis is a chronic disease which practitioners refer to as “bridge symptoms or bifurcation points.” This refers to the “Bi-syndrome” ability to switch from one state to another state (ie Heat to Wind). Using metabolomics, they found when practitioners refer to “Heat”, patients had apoptosis-stimulating genes, which refer to genes stimulating cell death. “Cold” patients had more genes that resisted apoptosis activity. Li et al¹¹ investigated what treatments a “Cold” and “Heat” patient would need. “Cold” patient suffered from hormone imbalance whereas “Heat” patients suffered from immune system imbalance. This would suggest that a “Cold” patient would response better to hormone treatment and a “Heat” patient might respond better to immune-related treatment. This example illustrates how TCM and Western Medicine approaches can be mapped to each other. Recent studies of TCM treatments have been on hypertension, diabetes, and, cancer^{12,13,14}. The common finding with TCM in these articles is that they take a holistic approach to treating the disease. Many of the herbs target multiple areas which could mediate symptoms. For example, when treating for hypertension, the herb formula not only reduces the symptoms but it may improve insulin resistance, lipid and glucose metabolism, decrease cardio hypertrophy, and block

the renin-angiotensin system¹³. The renin-angiotensin system regulates acute and chronic of BP. Excessive activation of the system leads to high BP¹³.

Synergy of Chemical Compounds

Over the last decade, there has been a shift in development of multi-drug therapy from mono-substance therapy¹⁵. This is due to mono-drugs becoming ineffective and resistance problems. The concept behind multi-drug therapy is identifying multiple different pathways that are being altered by a disease and targeted simultaneously. As shown by previous research, this is common in cancer as many pathways are involved to varying degrees; thus, it is difficult to treat one gene or single pathway^{15,16,17}. Due to the shift in development of multi-drug therapy, researchers have turned their attention to Traditional Chinese Medicine as the compounds in treatment may target multiple symptoms.

Traditional Chinese Medicine (TCM) is widely used in China due to their reduced side effects, high efficacy and wide range of pharmaceuticals activities¹⁵. Furthermore, TCM is becoming well known as a modern alternative to western medicine or complementary to treating diseases or supporting health. One reason why TCM is popular may be due to the herb extracts acting synergistically with one another to create potent treatments. This may serve as a valuable resource for network-based multi-target drug discovery. A case example is polyphenols and terpenoids, two common group of compounds found in plant extracts. The former has a high affinity to bind to different molecular structures such as proteins or glycoproteins. On the other hand, Terpenoids have a high affinity to bind to cell membranes¹⁸. Chen et al.¹⁹ investigated how to improve the treatment for

glioblastoma patients and discovered by combining the temozolomide with polyphenols, terpenoids, and other active compounds the overall efficacy of the treatment was increased.

Consistent with this trend, this project is contributing to the growing knowledge of multi-drug treatment by linking compounds in TCM to cancer-related pathways. By having a understanding of what compounds target these pathways, we may be able to utilize that information to supplement the main treatment plan to accelerate the healing process.

Database

Data Acquisition

Data for the project was acquired from two sources. The data regarding traditional Chinese medicine was extracted from the Traditional Chinese Medicine Integrated Database (TCMID)²⁰. It contains five datasets, however, the only table that was used was the one containing "[ingredient_targets_disease_drug-TCMID.v2.03](#)". The term "ingredients" refer to the chemical components found in the plant of interest. The version of this table used was v2.03. The other data was extracted from the online database reactome. Reactome²¹ is a public database that holds information about biological pathways for humans and other animals that are updated daily. Three pieces of information were gathered from this database. The first was the lowest level of the pathways. In the biological pathways, there is a top level and a bottom level. Top level pathways refer to the major pathways that are commonly seen. The bottom levels are the

pathways that are in between the major pathways that make up the top layer. The second piece of information was the top level of the pathways. The pathway data had a UniProt identifier which was used to link the two different data sources. The last piece of data was the protein's UniProt code associated with the pathways of interest: inflammation, sustained proliferative signaling, and angiogenesis.

Tools

R version 3.4.3 and R studio 1.1.383 were used. The libraries that were used were: tidyverse (1.2.1)²², RSQLite (2.0)²³, and splitstackshape (1.4.2)²⁴ and plotly (4.7.1)²⁵.

Data Visualization

The visualization prepared was an interactive heatmap made using the program plotly. The rows contain the pathway names and the columns contain the ingredient names. Each of the cells display whether or not the compound of interest targets that pathway. Cells filled in black means this compound targets this pathway and vice versa if it fills in white. When the mouse is hovering over a cell, additional information about the pathway is provided. The additional information includes pathway name, herbal ingredient that

targets the pathway, the associated cancer hallmark, and the pathway ID.

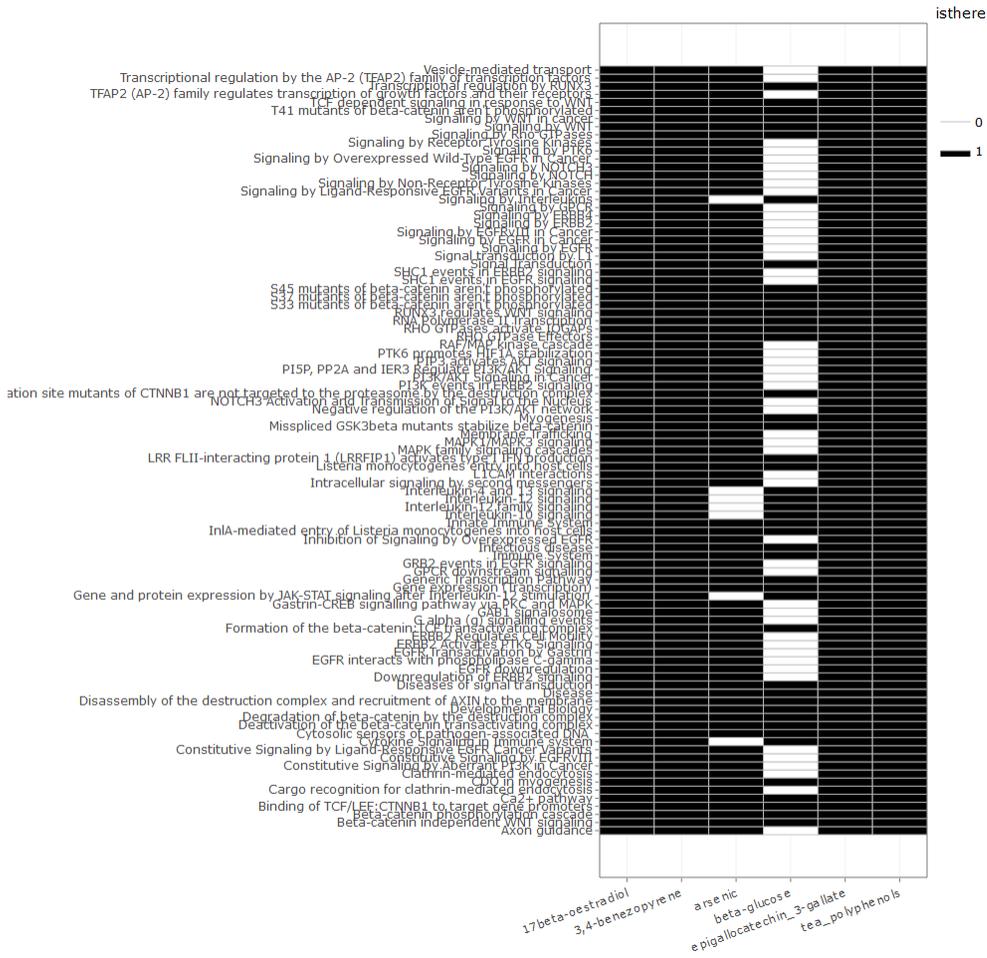


Figure 1. Rows are the pathways that were explored and the columns are the chemical compounds that were used to match with the pathway. 0 (white) means that this chemical compound did not target this pathway and 1 (black) means this chemical compound targets this pathway.

Methodology Journey

Introduction

In this section, the journey of how the data visual was developed will be discussed.

Discussions will include obstacles faced, solutions to these obstacles, and questions that were discussed that drove the process. For detailed methods, please see the index.

Data Selection

A database had to be selected as the first step and it had to be a publicly available. We made the decision to work with a database related to Traditional Chinese Medicine as I had interest in the ongoing effort of marrying Eastern and Western medicine. We requested assistance from Dr. Steven Chamberlin, a researcher who studies alternative medicine, to help with choosing the most informative databases for the project. He recommended two different publically available databases: TCM Database@Taiwan²⁶ and Traditional Chinese Medicines Integrated Database (TCMID)²⁰. Both databases contain entries of medicines including their names, chemical compounds, and usage. We selected the TCMID for extracting the information as data from Database@Taiwan was not as easily compared to TCMID. On the Database@Taiwan website, the user must manually selected items to download. This made it difficult as I did not have any prior information on the subject. On the other hand, TCMID allowed users to download all entries without having to select entries of interest.

Exploring the Database

During the first term, we explored the TCMID database quite a bit before we decided on a a specific question. The TCMID database contains five tables: prescription, herb, disease target, herbal ingredients links to GNSP, and Mass Spectra information. From initial data exploration, two of the five tables were useful. The issue with the prescription table was it was written in symbols. My assumption was the table was written in Mandarin given the nature of database origin. The tables that related to GNSP and Mass Spectra were not particularly useful as they did not seem to have an identifier that linked to the other tables.

The herb table displayed information about the ingredients including how the items were used in traditional Chinese medicine. The target network contained information about what chemical compounds were found in the database. The main problem with this table was the lack of labels on the column fields. To overcome this obstacle, I explored each column by searching the first value in each column on Google. These column labels were UniProt ID, Gene ID, OMIM ID, and drugbase ID. There were two columns that I was unable to identify thus I eliminated them from the future investigation. I then filtered out the rows that weren't complete (complete as in all fields were filled) for both tables. In the herb table, only 336 out of 8203 rows were returned as complete and in the network target table, only 18814 out of 162077 rows were returned. This was an issue as only a small number were returned, which could cause a problem later on with future analysis. In both tables, I did a group by group analysis to see how many of these herbs targeted a disease. In the herb target, I did a group by groups on the Meridians. Meridians are signs that practioner of TCM use to diagnosis an aliment. Many of the herbs targeted the liver. In the disease network table, I grouped the rows based on chemical compound. Since one of the fields is the Online Mendelian Inheritance in Man(OMIM) ID²⁷, we were able to see which compounds were linked with specific diseases. It seemed half of them were specific (based on a arbitrary rule of compounds that have more than 10 items were considered not specific). Although this initial data exploration gave us a general direction for formulating a question, we still were far away from an reasonable question to pursue. We made the decision to meet with Dr.Chamberlin to discuss what kind of visual would be useful for a researcher in this field.

Discussing with Researcher

With Dr. Chamberlin, we discussed the databases that are currently available and that might be helpful to explore. For instance, Dr. Chamberlin mentioned Supernatural 2²⁸, a database contains 325,508 natural compounds (NCs), however this database was proprietary. We also discussed the differences between what Western and Naturopathic physicians might need. Western physicians are more interested in chemical to target interactions and Naturopathic physicians are more interested in the targets of the natural products. We then needed to decide who the target user for the project would be, Western physicians or Naturopathic physicians? Once we decided that the target would be Western physicians, we needed to decide whether we would focus on chemical compounds or whole plant extracts. The advantage of the latter is that ingredients in the whole plant extracts can work synergistically. Information about this synergy is lost when only looking at individual chemical compounds. Lastly, we decided to link the full plant to the cancer related pathways. We chose cancer related pathways because it was of interest to me and there is research interest in using TCM to treat cancer patients. We discussed that I should link plants that are associated with inflammation in the inflammatory pathways by using the Uniprot ID to link to the reactome pathway. We briefly talked about how to visualize this information via a heatmap where the pathways would be the rows and the columns would be the compounds. The major contribution of this work would be to look at how many hallmarks of cancers are targeted by the chemical compounds in TCM.

Linking Issues

I gathered the UniProt ID related to inflammation from reactome and crosslinked them back to the target network table to see which UniProt ID would be returned. The next steps were to link the UniProt ID to the pathways and plant. I ran into no issues with the former step. However with the latter step, I ran into a problem where the tables did not have identifiers in TCMID tables to allow me to link to the UniProt ID. When I brought this to the committee, we had to change the goal of the project slightly. Instead of trying to link the plant to the pathway, we decided to link UniProt ID to the pathway as that was viable and we still could use this information to show how TCM ingredients target multiple pathways. However one suggestion that was made was to at least show proof of concept that linking plant to the pathways was viable. Taking the most common UniProt ID from three cancer hallmarks (inflammation, sustained proliferative signaling, and inducing angiogenesis), I performed a manual search on the TCMID database online where I had to reverse search the UniProt ID back to the plant and then look through the list of herbal ingredients to find any that matched.

Herb name	Herbal ingredient	Uniprot:P22301 (Inflammatory)	Uniprot:P00533 (sustain proliferative signaling)	Uniprot:P35222 (angiogenesis)
Green Tea	Tea polyphenols	X	x	X
Common Tea	Epigallocatechin 3-gallate	X	X	X
Human placenta	17beta-oestradiol	X	X	X
Hairy antler	17beta-oestradiol	X	X	X
abelmusk	17beta-oestradiol	X	X	X

Sunflower seeds	3,4-benzopyrene	X	X	X
Cortex Vignae radiatae	Beta-glucose	X		X
Chinese Hawthorn Leaf	Arsenic		X	X

Table 1. Manual search for herbal ingredient that targeted multiple pathways. Herbal ingredients were checked individually to see if they target each of the UniProt ID. This resulted in some herbal ingredients targeting multiple pathways.

With this new information and change in project goal, the question that was asked of me was what is this capstone contributing to the current knowledge. One of the two contributions is the exploration of new pathways. Mohamad et al.²⁹ had only focused on digestive related pathways while I focused more on the cancer related ones. The second is the proof of concept that linking plants to the pathway is worth exploring. As these herbal ingredients target multiple pathways, it might be important for physicians to considered a different alternative medicine if the treatment targets pathways that may hinder the healing process.

Final Table and Visualization

I presented an initial heatmap of what the final product would look like to the committee. The rows were the herbal ingredients and the Uniprot ID were the columns. One of the concerns was the columns. Physicians had an idea of what the pathways do and what to watch out for. UniProt ID can link to several different pathways including ones outside of the cancer realm. So we discussed what needed to be done to improve on the initial heatmap. Instead of the UniProt ID acting as the pathway identifier, it was suggested to put in the pathways that were linked to that UniProt ID instead. As for the

rows, it would be the chemical compounds. We also discussed creating a final table that contained all the information needed for the heatmap visual. In the final table, we discussed it should contain the UniProt ID, pathway name, pathway ID, cancer hallmark, and a vector containing 1 and 0. The 1 and 0 respectively is represent compounds that do and do not target a specific pathway.

After developing this final table, I noticed that this table looked like any other target to disease network. There was not any additional information being added. So I modified the table so it would contain the herbal compounds linking the pathway to the compound like the original plan. In the final table, it contains the UniProt ID that targets the pathway, pathway id, pathway name, hallmark, a vector that contains whether or not the herbal ingredients target the pathway, and the herbal ingredients.

Originally, the plan was to use the final table to visualize a heatmap of components that target different pathways and see which ones overlap. The problem that was encountered was certain packages that create interactive visualization did not accept the final table. Heatmaply, an R package built off plotly, only accepts matrix as inputted. My final data only contains information if the compound targets the pathway. This is shown as 1 and 0, which heatmaply recognizes as characters rather than numbers. Heatmaply needs some columns to be numeric to visualize the data. After some consideration, I made the decision that heatmaply may not be appropriate for my visualization as my data is only 1 and 0. Furthermore, a legend that ranges for 1 to 0 increments would mislead users about the information being displayed. Exploring other options, I decided to use plotly to

display the information. Plotly is a R package that is built off of ggplot2 (another r package primarily used to display data) to create interactive visualization. It has the option of displaying

Reflection

We see some of the chemical compounds target all three cancer hallmarks, while others target only one of the pathways or even alternative pathways such as Beta-glucose and arsenic. This information is important when doctors are considering supplementing cancer treatment. For instance, if physicians are considering using an supplement that contains arsenic, he/she may consider if arsenic targets pathways not intended. This could hamper the healing process. Although this is a small dataset, it contributes to growing knowledge of linking TCM treatments to the pathways they impact. Currently, only pathways related to the digestive system have been linked²⁰. If there was more time available, more pathways could have been explored and linked.

As for my journey through this project, there were many obstacles that hindered progress and changed the objective of the project slightly. The particular instance that changed the objective of the project was when it was discovered we could not link the herb to the pathway as the public database did not have the identifiers to link the tables together. This eventually forced us to change the project goal to link the compounds to the pathways. The lesson learned from this experience is to be adaptive to the situation. Particularly, in research, not everything will go as planned and sometimes one has to be innovative.

While I worked on linking the compounds to the pathways, Dr. Boudreau suggested at least trying to link the herb to the pathways to show proof of our original goal. This suggestion eventually led to the final goal of this project: proof of concept that compounds in the herbs may target multiple pathways and have a synergetic effect. Through this project I have learned a great deal of how development of the project is done and how a data visual needs to contain enough information to be useful to someone but at the same time, it cannot be too stimulating.

For future considerations, different hallmark pathways need to be linked to chemical compounds to understand how compounds are targeting pathways but furthermore assist us in creating precise treatment. The second consideration is more chemical compounds need to be explored and curated. The database used had many chemical compounds with unknown traits as shown by the data exploration piece of this project. Thus, it might be necessary to investigate chemical compounds before any linking can be done.

References

1. Molassiotis A, Fernandez-Ortega P, Pud D, Ozden G, Scott JA, Panteli V, et al. Use of complementary and alternative medicine in cancer patients: a European survey. *Annals of Oncology*. 2005;16(4):655-63.
2. Evans M, Shaw A, Thompson EA, Falk S, Turton P, Thompson T, et al. Decisions to use complementary and alternative medicine (CAM) by male cancer patients: information-seeking roles and types of evidence used. *BMC Complementary and Alternative Medicine*. 2007;7(1):25.

3. Robinson A, McGrail MR. Disclosure of CAM use to medical practitioners: a review of qualitative and quantitative studies. *Complement Ther Med*. 2004;12(2-3):90-8.
4. Tasaki K, Maskarinec G, Shumay DM, Tatsumura Y, Kakai H. Communication between physicians and cancer patients about complementary and alternative medicine: exploring patients' perspectives. *Psycho-oncology*. 2002;11(3):212-20.
5. Ernst E, Schmidt K. 'Alternative' cancer cures via the Internet? *British Journal of Cancer*. 2002;87(5):479-80
6. Gilbert B, Alves LF. Synergy in plant medicines. *Current medicinal chemistry*. 2003;10(1):13-20.
7. NIH. The Use of Complementary and Alternative Medicine in the United States [Internet]. 2008 [cited May 25 2018]. Available from:https://nccih.nih.gov/research/statistics/2007/camsurvey_fs1.htm
8. Cooper K, Noller B, Connell D, Yu J, Sadler R, Olszowy H, et al. Public Health Risks from Heavy Metals and Metalloids Present in Traditional Chinese Medicines. *Journal of Toxicology and Environmental Health, Part A*. 2007;70(19):1694-9.
9. Kim HK, Wilson EG, Choi YH, Verpoorte R. Metabolomics: A tool for anticancer lead-finding from natural products. *Planta Medica*. 2010;76(11):1094-102.
10. van der Greef J, van Wietmarschen H, Schroen J, Wang M, Hankemeier T, Xu G. Systems biology-based diagnostic principles as pillars of the bridge between Chinese and Western medicine. *Planta Medica*. 2010;76(17):2036-47.
11. Li S, Zhang ZQ, Wu LJ, Zhang XG, Li YD, Wang YY. Understanding ZHENG in traditional Chinese medicine in the context of neuro-endocrine-immune network. *IET systems biology*. 2007;1(1):51-60.
12. Tong X-L, Dong L, Chen L, Zhen Z. Treatment of Diabetes Using Traditional Chinese Medicine: Past, Present and Future. *The American Journal of Chinese Medicine*. 2012;40(05):877-86.
13. Xiong X, Yang X, Liu Y, Zhang Y, Wang P, Wang J. Chinese herbal formulas for treating hypertension in traditional Chinese medicine: perspective of modern science. *Hypertension research : official journal of the Japanese Society of Hypertension*. 2013;36(7):570-9.
14. Li-Weber M. Targeting apoptosis pathways in cancer by Chinese medicine. *Apoptosis Targeting Drugs in Cancer*. 2013;332(2):304-12.
15. Yang Y, Zhang Z, Li S, Ye X, Li X, He K. Synergy effects of herb extracts: Pharmacokinetics and pharmacodynamic basis. *Fitoterapia*. 2014;92:133-47.

16. Parsons DW, Jones S, Zhang X, Lin JC-H, Leary RJ, Angenendt P, et al. An Integrated Genomic Analysis of Human Glioblastoma Multiforme. *Science*. 2008;321(5897):1807-12.
17. Jones S, Zhang X, Parsons DW, Lin JC-H, Leary RJ, Angenendt P, et al. Core Signaling Pathways in Human Pancreatic Cancers Revealed by Global Genomic Analyses. *Science*. 2008;321(5897):1801-6.
18. Wagner H, Ulrich-Merzenich G. Synergy research: Approaching a new generation of phytopharmaceuticals. *Phytomedicine*. 2009;16(2):97-110.
19. Chen TC, Wang W, Golden EB, Thomas S, Sivakumar W, Hofman FM, et al. Green tea epigallocatechin gallate enhances therapeutic efficacy of temozolomide in orthotopic mouse glioblastoma models. *Cancer Letters*. 2011;302(2):100-8.
20. Xue R, Fang Z, Zhang M, Yi Z, Wen C, Shi T. TCMID: Traditional Chinese Medicine integrative database for herb molecular mechanism analysis. *Nucleic acids research*. 2013;41(Database issue):D1089-95.
21. Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L, et al. Reactome: a database of reactions, pathways and biological processes. *Nucleic acids research*. 2011;39(Database issue):D691-7.
22. Wickham H. tidyverse: Easily Install and Load the 'Tidyverse' [Internet]. 2017 [cited June 7 2018]. Available from: <https://CRAN.R-project.org/package=tidyverse>
23. Muller K, Wickham H, James DA, Falcon S. RSQLite: 'SQLite' Interface for R [Internet]. 2017 [cited June 7 2018]. Available from: <https://CRAN.R-project.org/package=RSQLite>
24. Mahto A. splitstackshape: Stack and Reshape Datasets After Splitting Concatenated Values [Internet]. 2014 [cited June 7 2018]. Available from: <https://CRAN.R-project.org/package=splitstackshape>
25. Sievert C, Parmer C, Hocking T, Chamberlain S, Ram K, Corvellec M, et al. plotly: Create Interactive Web Graphics via 'plotly.js' [Internet]. 2017 [cited June 7 2018]. Available from: <https://CRAN.R-project.org/package=plotly>
26. Sanderson K. Databases aim to bridge the East-West divide of drug discovery. *Nature Medicine*. 2011;17:1531.
27. John Hopkins University. Online Mendelian Inheritance in Man [Internet]. 2018 [cited June 7 2018]. Available: <https://www.omim.org/>.
28. Banerjee P, Erehman J, Gohlke B-O, Wilhelm T, Preissner R, Dunkel M. Super Natural II—a database of natural products. *Nucleic acids research*. 2015;43(Database issue):D935-D9.

29. Mohamad Zobir SZ, Mohd Fauzi F, Liggi S, Drakakis G, Fu X, Fan TP, et al. Global Mapping of Traditional Chinese Medicine into Bioactivity Space and Pathways Annotation Improves Mechanistic Understanding and Discovers Relationships between Therapeutic Action (Sub)classes. Evidence-based complementary and alternative medicine : eCAM. 2016;2016:2106465.

Appendix

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1
## .2.1 --

## v ggplot2 2.2.1      v purrr  0.2.4
## v tibble  1.4.2      v dplyr  0.7.4
## v tidyr   0.7.2      v stringr 1.2.0
## v readr   1.1.1      v forcats 0.2.0

## -- Conflicts ----- tidyverse_conflic
## ts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(data.table)

##
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':
##
##   between, first, last

## The following object is masked from 'package:purrr':
##
##   transpose

library(splitstackshape)
```

Initial Analysis:

There are five tables in this database. Some of the tables are usable and others are not.

Tables that are not usable: The prescription table is filled with weird symbols assuming this table is only in Mandarin.

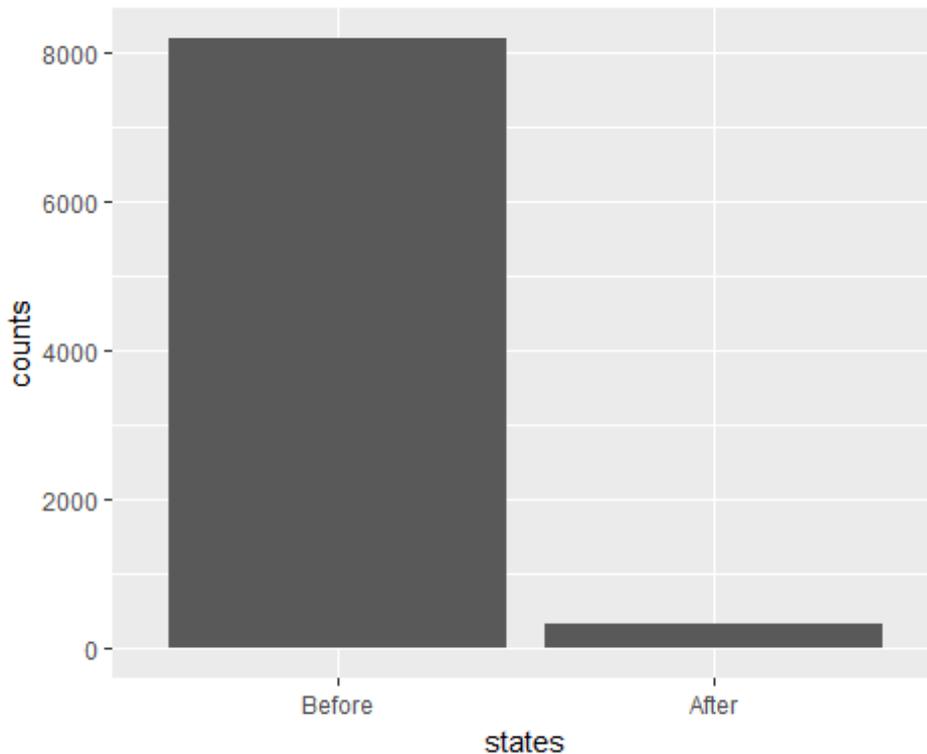
Tables that may potentially useful: Ingredient_MS.TCMID.v2.01 HERB_MS

Tables that were analyzed: herb.TCMID.v2
ingredient_targets_disease_drug.TCMID.v2.03

Investigation of the herb.TCMID.v2 table:

```
## [1] "Pinyin.Name" "Chinese.Name" "English.Name" "Latin.Name"  
## [5] "Properties"    "Meridians"    "Use.Part"    "Effect"  
## [9] "Indication"
```

Table 1. This is a table depicting the before and after state after asking R to return only complete rows. There are 336 complete rows after filtering.



```
##   states counts  
## 1: Before  8203  
## 2: After   336  
  
##  
##   English.Name  
## 10                               Cochin  
##   chinese Asparagus  
## 23  
##   Shiny Bugleweed
```

## 52			C
airo Morningglory			
## 73	Prepared Common Monkshood Daughter Root Equivalent plant:	Aconitum carmichaeli cv	
## 97			Cochinhi
na Momordica Seed			
## 124			
	Orientvine		
## 125			
	China Ixeris		
## 182			Chinese A
rborvitae Kernel*			
## 204			C
ommon Anemarrhena			
## 208			
	Bunge Corydalis		
##		Latin.Name	
## 10	Asparagus cochinchinensis [Syn. Asparagus lucidus]		
## 23		Lycopus lucidus	
## 52	Ipomoea cairica [Syn. Ipomoea palmata]		
## 73		Aconitum carmichaeli	
## 97		Momordica cochinchinensis	
## 124		Sinomenium acutum	
## 125		Ixeris chinensis	
## 182		Biota orientalis	
## 204		Anemarrhena asphodeloides	
## 208		Corydalis bungeana	
##		Properties	Meridians
## 10	Sweet, bitter,Extremely cold		Lung,kidney
## 23	Bitter, pungent,slightly warm		Liver,spleen
## 52	Cold,Sweet,Bitter	Lung, Large Intestine, Liver	
## 73	Extreme Hot,Pungent,Sweet	Spleen,Heart,Kidney	
## 97	Warm,Slightly Sweet,Bitter	Spleen,Stomach,Liver	
## 124	Warm,Pungent,Bitter	Spleen,Liver	
## 125	Cool,Bitter	Lung,Liver	
## 182	Mild,Sweet	Large Intestine,Heart,Kidney	
## 204	Cold,Sweet,Bitter	Lung,Stomach,Kidney	
## 208	Cold,Pungent,Bitter	Liver,Heart	
##		Use.Part	
## 10		tuberoid	
## 23		aerial parts	
## 52		Root or stem-leaf	
## 73		daughter root	
## 97		ripe seed	
## 124		stem	
## 125		whole herb or root	
## 182		kernel	
## 204		rhizome	
## 208		whole herb	

##

Effect

10 To enrich yin and moisten dryness, clear lung and downbear fire.

23 To quicken blood and transform stasis, move water and disperse swelling, resolve toxin and eliminate welling abscess.

52 To clear heat and resolve toxin, disinhibit water and free strangury.

73 To return yang and treat collapse, supplement fire and reinforce yang.

97 To dissipate bindings and disperse swelling, attack toxin and cure sores.

124 To dispel wind and free network vessels, dispel damp and relieve pain.

125 To clear heat and resolve toxin, disperse swelling and expel pus, cool blood and stanch bleeding.

182 To nourish heart and quiet spirit, constrain sweat, moisten intestines and free stool.

204 To clear heat and drain fire, enrich yin and moisten dryness, eliminate vexation and allay thirst.

208 T
To clear heat toxin, disperse swollen welling abscess.

##

Indication

10 Yin vacuity fever, cough with blood ejection, lung wilting, pulmonary welling abscess, swelling pain in throat, diabetes mellitus, constipation.

23 Amenorrhea, concretion and conglomeration, postpartum blood stasis abdominal pain, edema in body and face, knocks and falls, incised wound.

52 Lung heat cough, inhibited urination, strangury, edema, swollen welling abscess and toxin of clove.

73 Yang-collapse vacuity desertion, cold limbs and faint pulse, impotence, uterus cold, cold pain in heart and abdomen, vacuity cold chronic diarrhea and dysentery, yin cold edema, yang vacuity external contraction, wind-cold-damp impediment, yin flat abscess and sores.

97 Swollen welling abscess, mammary welling abscess, scrofula, hemorrhoids and fistulas, dry lichen, bald sores.

124

Wind-damp impediment p

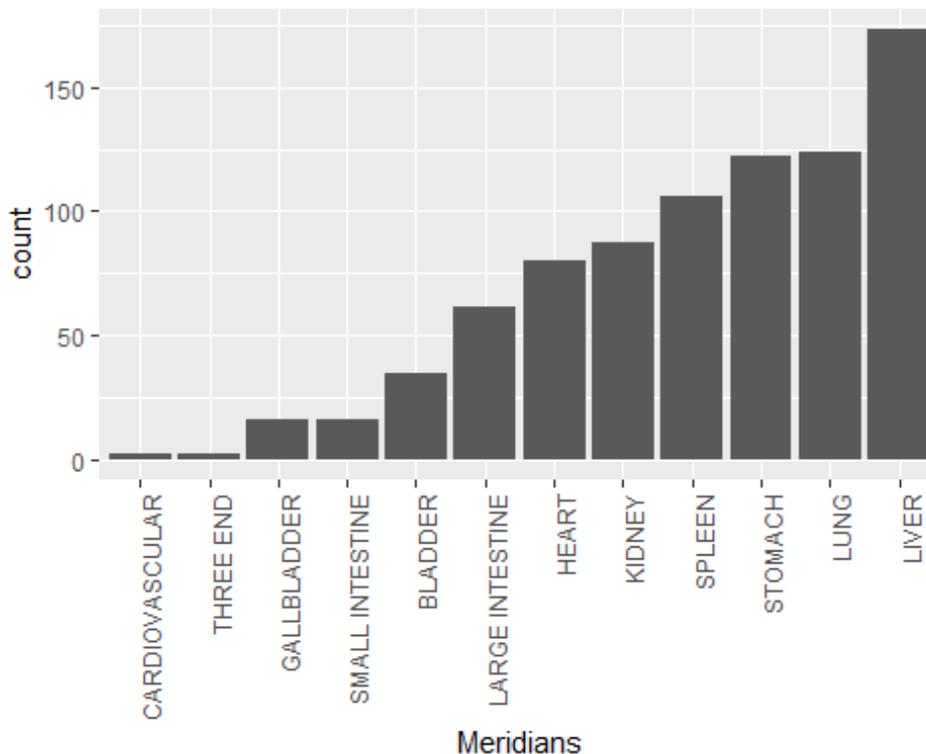
ain, pain from arthritis, swelling in joints, paralytic and pruritus.
 ## 125 Intestinal welling abscess, pulmonary welling abscess, lung heat cough, enteritis, dysentery, cholecystitis, pelvic inflammation, swelling toxin of sore and boil, scrotal eczema, blood ejection, spontaneous external bleeding, flooding, knocks and falls.
 ## 182

Fright palpitation and fearful throbbing, sleepless and amnesia, night sweating, intestinal dry and constipation.
 ## 204

Diabetes mellitus, warm heat disease, ardent fever with vexation and thirst, cough and asthma, dry cough, constipation, steaming bone tidal fever, vacuity vexation and insomnia, strangury-turbidity.
 ## 208
 I
 nfluenza, infection of upper respiratory tract, tonsillitis, infective hepatitis, enteritis, dysentery, nephritis, parotitis, conjunctivitis, acute appendicitis, clove sore and swollen welling abscess, scrofula.

Graph. The data was group by the Meridians to show which category was being heavily used to treat.

Table. A table version of the group by meridians.



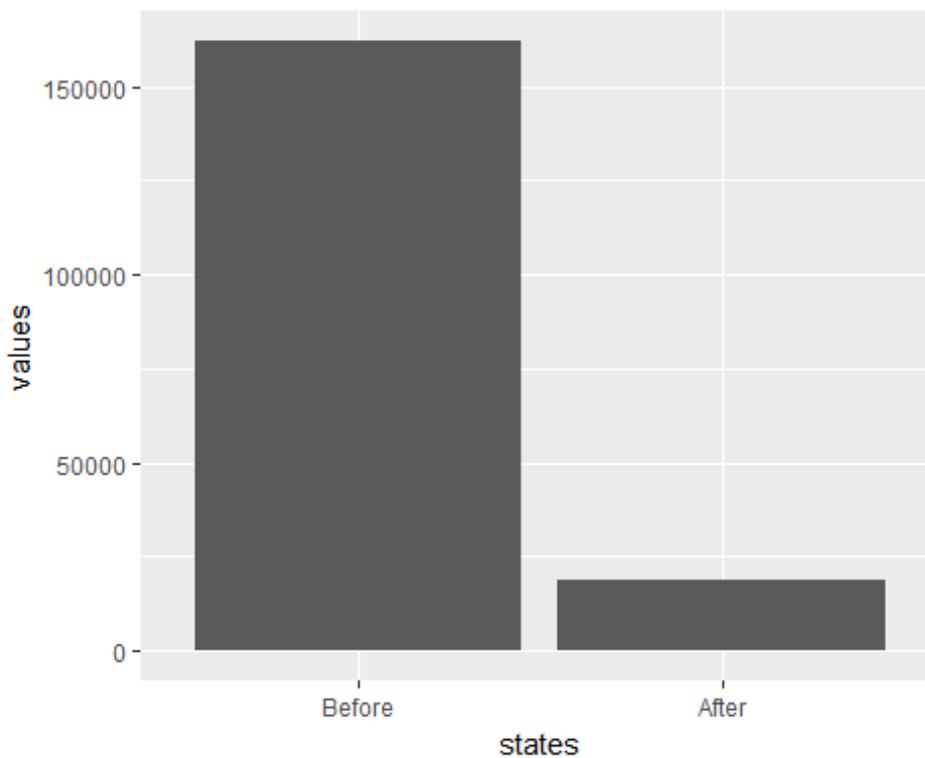
```
## # A tibble: 12 x 2
##   Meridians      count
##   <chr>         <int>
## 1 CARDIOVASCULAR     2
```

```

## 2 THREE END          2
## 3 GALLBLADDER       16
## 4 SMALL INTESTINE   16
## 5 BLADDER           35
## 6 LARGE INTESTINE   61
## 7 HEART              80
## 8 KIDNEY             87
## 9 SPLEEN            106
## 10 STOMACH           122
## 11 LUNG              124
## 12 LIVER             173

```

Target-disease table analysis:



```

## states values
## 1: Before 162077
## 2: After 18814

```

Guess: V1=compound name V2=?? V3=uniprot protein sequence and functional information v4=information about the gene v5=OMIM ID (diseases that this compound has been used to treat) v6=Drug database (compound found in these products)

```

## [1] "V1" "V2" "V3" "V4" "V5" "V6"
##
##          V1    V2    V3    V4
## 1  abietic acid 3145 Q9Y6L6 SLC01B1
## 3  abietic acid 3323 P37231  PPARG

```

```

## 14  acanthoicacid  2461 P01375    TNF
## 25  acetophenone  12415 P15428    HPGD
## 38  acetophenone  12479 Q8TC12    RDH11
## 62  acetovanillone 10835 P04839    CYBB
## 82  acetovanillone  8585 P03956    MMP1
## 85  acetovanillone 13061 P99999    CYCS
## 103 acetylcholine  2323 P00734     F2
## 105 acetylcholine  8113 Q13255    GRM1
##

```

```

V5
## 1 614332;2
37450
## 3 609338;604367;601552;608594;612526;226200;219100;613949;269700;2
16360
## 14 184100;157300;6
16311
## 25 119900;6
08978
## 38 6
16108
## 62 300645;1
67755
## 82 606963;6
09583
## 85 612004;212350;188000;242840;2
73900
## 103 188050;614415;615917;6
13679
## 105 602491;6
14831
##

```

V6

```

## 1
DB00655;DB00950;DB06403;DB002
79;DB00220;DB01045;DB00224;DB00227;DB00286;DB04881;DB00091;DB08912;DB00
917;DB01092;DB00912;DB08860;DB00503;DB00563;DB00859;DB01095;DB00509;DB0
0520;DB00688;DB00412;DB01232;DB01132;DB05804;DB06589;DB00451;DB08884;DB

```

00275;DB00661;DB01076;DB01053;DB06772;DB00783;DB00762;DB00669;DB06626;DB06210;DB00583;DB06151;DB01393;DB00213;DB05239;DB00390;DB06290;DB00877;DB00908;DB00973;DB04348;DB01892;DB11362;DB13153;DB13139;DB09137;DB01098;DB00641;DB09102;DB02659;DB02691;DB09054;DB00175;DB00177;DB08864;DB09272;DB00439;DB00197;DB01583;DB01241;DB00345

3

DB07509;DB07724;DB00159;DB01067;DB00328;DB07723;DB00731;DB07111;DB07842;DB04689;DB05187;DB07863;DB00912;DB07172;DB08402;DB08560;DB01132;DB01014;DB04270;DB01252;DB01050;DB00966;DB01393;DB07208;DB07209;DB07675;DB05490;DB08121;DB07053;DB07302;DB04971;DB08435;DB00412;DB00795;DB08760;DB08915;DB06926;DB06908;DB00244;DB00197;DB05854;DB08302

14

DB02325;DB01411;DB00065;DB08910;DB01041;DB00608;DB05303;DB05676;DB05207;DB00852;DB05066;DB01407;DB00051;DB05869;DB05758;DB05744;DB05250;DB00668;DB01427;DB01296;DB05412;DB08904;DB05218;DB05767;DB05470;DB04956;DB05017;DB06674;DB00005;DB05879

25

DB00157

38

62

DB00162

82

DB00514

3;DB08491;DB00786;DB08482;DB07556;DB07926
85

DB0840

7;DB03977;DB02110;DB01017;DB03934;DB04249
103 DB07277;DB07131;DB07279;DB07278;DB07376;DB07665;DB07660;DB08422;
DB04786;DB06996;DB06858;DB06859;DB06850;DB06853;DB06854;DB04772;DB04771
;DB01767;DB01766;DB07095;DB07091;DB07718;DB07016;DB03865;DB07659;DB0765
8;DB07143;DB06865;DB06866;DB06861;DB06869;DB06868;DB07549;DB00170;DB074
61;DB07088;DB07527;DB03159;DB07522;DB04591;DB07083;DB07005;DB06878;DB00
278;DB07796;DB07550;DB07548;DB07120;DB00100;DB07128;DB07639;DB08624;DB0
9228;DB04697;DB07400;DB08187;DB07508;DB07133;DB03136;DB07741;DB07027;DB
06695;DB06911;DB08546;DB00055;DB07944;DB07946;DB02723;DB07105;DB06919;D
B07521;DB06942;DB06947;DB07897;DB01123;DB04136;DB07353;DB07440;DB08152;

DB0331

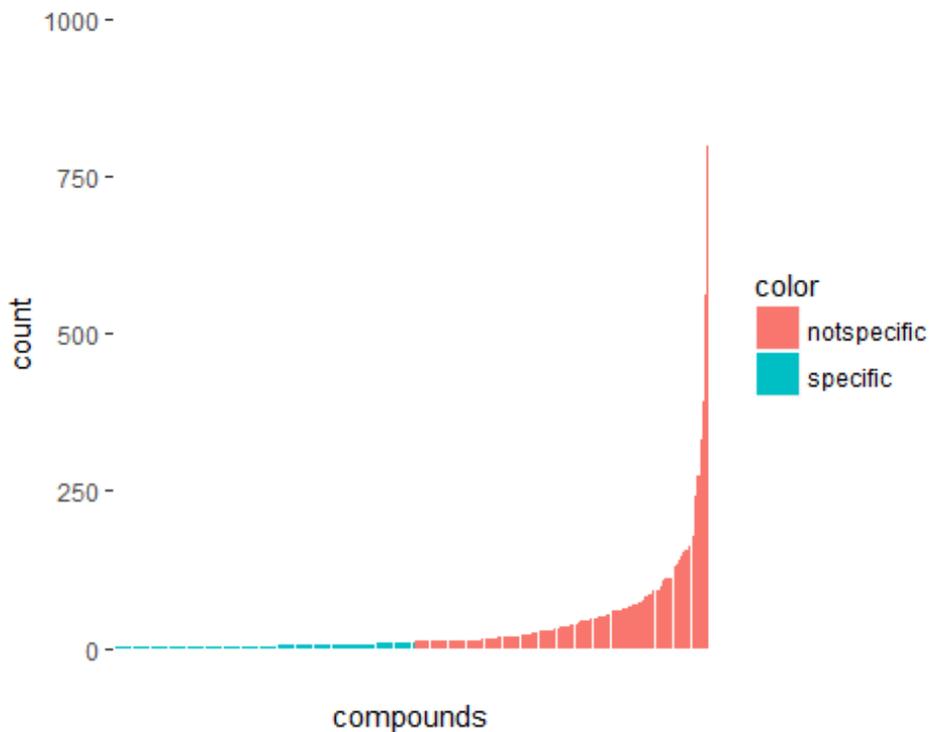
```
DB07190;DB02193;DB07211;DB04722;DB05777;DB07809;DB06936;DB07515;DB06838
;DB08254;DB07366;DB04898;DB07165;DB06404;DB00006;DB00001;DB08062;DB0806
1;DB06841;DB03847;DB07934;DB06929;DB06845
## 105
```

DB00142;DB04256

Graph. ingredient_targets_disease_drug dataset was group by compound. It shows which compounds are specific to a particular disease. Color coded to show how many compounds are specific and not specific.

Table. Shows the compound name and how many diseases it targets.

Table. Shows how many compounds are specific and not specific to a disease(s)



```
## # A tibble: 1,844 x 2
##   V1                                coun
##   <fct>                             <int>
```

```

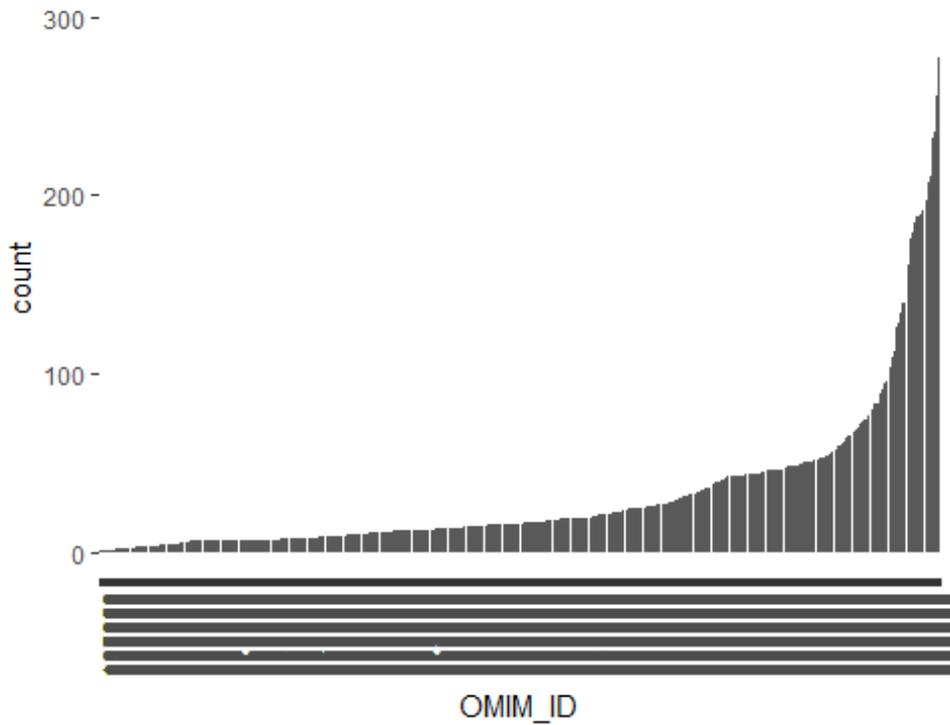
>
## 1 " (-)-cis-Carveol"
1
## 2 (-)-terpinene-4-ol
1
## 3 (e,e)-farnesol
1
## 4 1-(1,4-dimethyl-3-cyclohexen-1-yl)-ethanone
1
## 5 1-[1-oxo-3(3,4-methylenedioxyphenyl)-2e-propenyl]pyrrolidine
1
## 6 1-terpinen-4-ol
1
## 7 1,3-Dichlorobenzene
1
## 8 1,4-dihydroxy-2-methoxybenzene
1
## 9 2'-methoxy-3,4,4'-trihydroxychalcone
1
## 10 2-aminobenzoic acid
1
## # ... with 1,834 more rows

## # A tibble: 2 x 2
##   color      count
##   <chr>      <int>
## 1 notspecific  906
## 2 specific    938

```

Table.Shows how many compounds target each disease

Graph. Display many compounds targeted that disease in ascending order.



```
## # A tibble: 1,907 x 2
##       V5 count
##   <int> <int>
## 1 113100     1
## 2 167250     1
## 3 185800     1
## 4 186300     1
## 5 200700     1
## 6 209950     1
## 7 219750     1
## 8 219800     1
## 9 219900     1
## 10 222448     1
## # ... with 1,897 more rows
```

Top 10 disease: Familial hypercholanemia Periodic fever, menstrual cycle-dependent
 Cranioectodermal dysplasia Cyclic neutropenia Agammabglobulinemia heme oxygenase-
 1 deficiency Nephrogenic diabetes insipidus Nasopharyngeal carcinoma Vitamin D-
 dependent rickets type 2B with normal vitamin D receptor Adrenocortical carcinoma

Pathways_Second_Pass

Lawrence Hsu

Apr 2, 2018

Summary of sustaining proliferative signaling pathways

```
query<-"select V3 as Uniprot,count from proteins_egfr order by count de
sc"
dbGetQuery(test,query)

## Uniprot count
## 1 P00533 154
## 2 P01112 152
## 3 P01116 100
## 4 P42336 78
## 5 P0CG47 56

#247 distinct pathways (simple)
uni_simple_pathway<-xy%>% group_by(V4) %>% rename(simple_pathways=V4) %
>% summarise(count=n())
uni_simple_pathway

## # A tibble: 247 x 2
## simple_pathways count
## <chr> <int>
## 1 ABC-family proteins mediated transport 1
## 2 Activated NOTCH1 Transmits Signal to the Nucleus 2
## 3 activated TAK1 mediates p38 MAPK activation 1
## 4 Activation of IRF3/IRF7 mediated by TBK1/IKK epsilon 1
## 5 Activation of NF-kappaB in B cells 1
## 6 Activation of RAS in B cells 2
## 7 Amyloid fiber formation 1
## 8 Antigen processing: Ubiquitination & Proteasome degradation 1
## 9 APC-Cdc20 mediated degradation of Nek2A 1
## 10 APC/C:Cdc20 mediated degradation of Cyclin B 1
## # ... with 237 more rows

# 462 unique pathways (complex)
unique.pathway.complex.group
```

```

## # A tibble: 462 x 2
##   complex_pathways
##   count
##   <chr>
##   <int>
## 1 ABC-family proteins mediated transport
##   1
## 2 ABC transporter disorders
##   1
## 3 Activated NOTCH1 Transmits Signal to the Nucleus
##   2
## 4 activated TAK1 mediates p38 MAPK activation
##   1
## 5 Activation of APC/C and APC/C:Cdc20 mediated degradation of mito~
##   1
## 6 Activation of IRF3/IRF7 mediated by TBK1/IKK epsilon
##   1
## 7 Activation of NF-kappaB in B cells
##   1
## 8 Activation of NMDA receptor and postsynaptic events
##   1
## 9 Activation of RAS in B cells
##   2
## 10 Adaptive Immune System
##   5
## # ... with 452 more rows

```

summary of inflammatory pathways

```

query<-"select V3 as Uniprot, count from proteins_inflame order by coun
t desc"
dbGetQuery(test,query)

##   Uniprot count
## 1  P22301    720
## 2  P01375    528
## 3  P19438     50
## 4  P29460     36
## 5  P18510     16

#14 distinct pathways (simple)
uni_simple_pathway<-xy%>% group_by(V4) %>% rename(simple_pathways=V4) %
>% summarise(count=n())
uni_simple_pathway

```

```

## # A tibble: 14 x 2
##   simple_pathways
##   count
##   <chr>
##   <int>
## 1 Gene and protein expression by JAK-STAT signaling after Interleu~
##   1
## 2 Interleukin-1 signaling
##   1
## 3 Interleukin-10 signaling
##   5
## 4 Interleukin-12 signaling
##   2
## 5 Interleukin-23 signaling
##   2
## 6 Interleukin-4 and 13 signaling
##   3
## 7 Regulation of TNFR1 signaling
##   2
## 8 TNF signaling
##   2
## 9 TNFR1-induced NFkappaB signaling pathway
##   2
## 10 TNFR1-induced proapoptotic signaling
##   2
## 11 TNFR1-mediated ceramide production
##   2
## 12 TNFR2 non-canonical NF-kB pathway
##   1
## 13 TNFs bind their physiological receptors
##   1
## 14 Transcriptional regulation of white adipocyte differentiation
##   1

```

22 unique pathways (complex)

unique.pathway.complex.group

```

## # A tibble: 22 x 2
##   complex_pathways
##   count
##   <chr>
##   <int>
## 1 Cytokine Signaling in Immune system
##   6
## 2 Death Receptor Signalling
##   2
## 3 Developmental Biology
##   1
## 4 Gene and protein expression by JAK-STAT signaling after Interleu~

```

```

    1
## 5 Immune System
    6
## 6 Interleukin-1 family signaling
    1
## 7 Interleukin-1 signaling
    1
## 8 Interleukin-10 signaling
    5
## 9 Interleukin-12 family signaling
    3
## 10 Interleukin-12 signaling
    3
## # ... with 12 more rows

```

Summary of inducing angiogenesis pathways

```

query<-"select V3 as uniprot, count from proteins_vegf order by count d
esc"
dbGetQuery(test,query)

##      uniprot count
## 1  P35222    368
## 2  P19878    252
## 3  P01112    152
## 4  P01116    100
## 5  P42336     78
## 6  P17252     64
## 7  Q14643     60
## 8  P04792     52
## 9  P04839     42
## 10 P31751     24
## 11 Q14571     22

#151 distinct pathways (simple)
uni_simple_pathway<-xy%>% group_by(V4) %>% rename(simple_pathways=V4) %
>% summarise(count=n())
uni_simple_pathway

## # A tibble: 151 x 2
##   simple_pathways
##   count
##   <chr>
##   <int>
## 1 Acetylcholine regulates insulin secretion
##   1
## 2 Activation of AKT2
##   1
## 3 "Activation of BAD and translocation to mitochondria "

```

```

1
## 4 Activation of RAS in B cells
2
## 5 AKT-mediated inactivation of FOXO1A
1
## 6 AKT phosphorylates targets in the cytosol
2
## 7 AKT phosphorylates targets in the nucleus
1
## 8 Antigen activates B Cell Receptor (BCR) leading to generation of~
4
## 9 AUF1 (hnRNP D0) binds and destabilizes mRNA
1
## 10 Beta-catenin phosphorylation cascade
1
## # ... with 141 more rows

# 290 unique pathways (complex)
unique.pathway.complex.group

## # A tibble: 290 x 2
##   complex_pathways          count
##   <chr>                 <int>
## 1 Acetylcholine regulates insulin secretion          1
## 2 Activation of AKT2                                1
## 3 "Activation of BAD and translocation to mitochondria " 1
## 4 Activation of BH3-only proteins                   1
## 5 Activation of NMDA receptor and postsynaptic events 1
## 6 Activation of RAS in B cells                       2
## 7 Adaptive Immune System                            10
## 8 AKT-mediated inactivation of FOXO1A                1
## 9 AKT phosphorylates targets in the cytosol         2
## 10 AKT phosphorylates targets in the nucleus         1
## # ... with 280 more rows

```

Herb name	Herbal ingredient	Uniprot:P22301 (Inflammato ry)	Uniprot:P00533 (sustain proliferative signaling)	Uniprot:P35222(angiog enesis)
Green Tea	Tea polyphenols	X	x	X
Common Tea	Epigallocate chin 3-gallate	X	X	X
Human placenta	17beta-oestradiol	X	X	X

Hairy antler	17beta-oestradiol	X	X	X
abelmusk	17beta-oestradiol	X	X	X
Sunflower seeds	3,4-benzopyrene	X	X	X
Cortex Vignae radiatae	Beta-glucose	X		X
Chinese Hawthorn Leaf	Arsenic		X	X

Manual search for herbs that targeted multiple pathways.

Heatmap

Lawrence Hsu

April 23, 2018

```
## Warning: package 'tidyverse' was built under R version 3.4.4
```

```
## -- Attaching packages -----
----- tidyverse 1.2.1 --
```

```
## v ggplot2 2.2.1    v purrr    0.2.4
## v tibble  1.4.2    v dplyr    0.7.4
## v tidyr   0.7.2    v stringr  1.2.0
## v readr   1.1.1    v forcats  0.2.0
```

```
## -- Conflicts -----
----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
## Parsed with column specification:
## cols(
##   herbal_ingredient = col_character(),
##   `uniprot:P22301` = col_integer(),
```

```
## `uniprot:P00533` = col_integer(),  
## `uniprot:P35222` = col_integer()  
## )  
ggplot(data=x)+theme_bw()+geom_tile(color="gray",aes(x=herbal_ingredient, y=uniprot,fill=expression))+theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

