Eat4Genes: A Bioinformatic Rational Gene Targeting App to Address Pathologies using Healthy Diet

Morgan L. Ford
Jessica M. Cooley
Rensselaer Polytechnic Institute
Troy, New York, USA

Dana R. Crawford
Albany Medical College
Albany, New York, USA

John S. Erickson
Kristin P. Bennett
Rensselaer Polytechnic Institute
Troy, New York, USA

ABSTRACT

Eat4Genes is a prototype diet recommendation web app for patients, healthcare providers, and researchers that aids in the selection of a healthy diet to help treat and prevent numerous health conditions. Our approach is focused on the strategic use of diet to regulate key risk gene expression, which we call dietary rational gene targeting (DRGT). To create the source DRGT dataset, we analyzed data of three types: an expert-curated list of conditions and the desirable modulation associated with risk genes to improve those conditions; an expert-curated list of rigorous scientific studies that quantify how the consumption of dietary agents (nutrients) influences gene expression in humans; and gene expression results for each of the studies. For each study, we added analyses of differentially expressed genes, study quality, and nutrient concentration. We developed a ranking system for studies that emphasize in vivo over in vitro studies, as well as whole foods and extracts over isolated phytochemicals at reasonable concentrations of nutrients. The Eat4Genes web app provides an engaging and informative interface that enables users to perform analysis by condition or by gene. We also present user scenarios from physician's and researcher's perspectives. The prototype Eat4Genes DRGT dataset and web app represent important steps towards translating DRGT and dietary research into a precision nutrition approach that is lower cost and healthier compared to pharmaceutical approaches.

CCS CONCEPTS

• **Applied computing** → **Health care information systems**; **Bioinformatics**.

KEYWORDS

dietary recommendation, bioinformatics, gene expression

ACM Reference Format:
Morgan L. Ford, Jessica M. Cooley, Dana R. Crawford, John S. Erickson, and Kristin P. Bennett. 2022. Eat4Genes: A Bioinformatic Rational Gene Targeting App to Address Pathologies using Healthy Diet. In *Proceedings of (ACM BCB '22).* ACM, New York, NY, USA, 10 pages. https://doi.org/XXXXXXX.XXXXXXX

ACM BCB '22, August 07-10, 2022, Chicago, IL

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Almost half of the world's population has one or more chronic diseases with resultant pain and suffering leading to the vast majority of health care spending [10]. Drug treatments are often expensive and can include a wide range of side- and long term-effects [22]. In fact, healthcare costs (per GDP) have more than tripled in the last 60 years [8]. Alternative approaches, such as diets that reduce costs and improve health, have great potential value in health care.

Eat4Genes is a diet recommendation web app for patients, healthcare providers, and researchers to aid in the selection of a healthy diet to help treat and prevent numerous conditions. Here "Condition" is defined as any state of health; i.e., good or bad, with bad then including diseases and disorders. It is based on evaluation of clinically relevant gene expression in response to a healthy diet with an emphasis on whole foods and whole food extracts. Eat4Genes identifies nutrients that may beneficially modulate condition risk genes (and other genes) based on evidence from a curated set of high quality research studies.

Based on previous research on rational gene targeting [13] [25], our approach is focused on the strategic use of diet to regulate key risk gene expression, which we call *dietary rational gene targeting* (DRGT) [2]. The goal of DRGT is to use healthy diet to restore disease-causing gene expression back to normal to treat various conditions. Compared with pharmaceutical drugs, this approach is low-cost and healthy and ultimately emphasizes precision nutrition in the form of personalized confirmation of our suggested diet.

The DRGT dataset at the heart of Eat4Genes has three main sources: an expert-curated list of *conditions* and the desirable modulation associated with *risk genes* to improve those conditions; an expert-curated list of rigorous scientific *studies* that quantify how consumption of *nutrients* influences gene expression in humans; and gene expression for each of the studies. Dr. Crawford's laboratory at Albany Medical College identified the source datasets as part of their research on DRGT [2]. We harmonized and joined this data and embedded it in the Eat4Genes web app, built on the R Shiny framework [3], allowing users to interactively query and visualize the data. In the *By Condition* analysis, Eat4Genes finds risk genes associated with a user-selected condition; determines nutrients that evidence suggests appropriately modulate these genes; and presents the studies supporting this evidence. As shown in Section 3.1, the *By Condition* analysis could be used by a physician or other healthcare provider to determine foods to suggest to a patient to address hypertension. In the *By Gene* analysis, users select a gene, and then Eat4Gene finds the nutrients and associated studies that support the desired modulation of that gene's expression. As discussed in Section 3.2, the *By Gene* analysis could provide valuable leads for researchers with in vitro findings on desirable gene expression;

they could use Eat4Genes to hypothesize how their findings could be translated into practice through diet for further study.

To maximize the potential effectiveness of food suggestions, Eat4Genes emphasizes in vivo studies in which humans actually eat nutrients. Eat4Genes also includes in vitro studies on human cell lines. The data for human cell line studies in Eat4Genes comes from NutriGenomeDB [15], a web-based application that hosts manually curated gene sets defined from gene expression signatures, after differential expression analysis of nutrigenomics experiments performed on human cells available in the Gene Expression Omnibus (GEO) repository [21]. The Eat4Genes app goes beyond the NutriGenomeDB app analysis to examine the relationship between conditions, risk genes, and nutrients.

The primary contributions of Eat4Genes include the following:

- Creation and release of a prototype harmonized and cleaned DRGT dataset that captures conditions, risk genes, nutrients, and studies collected from published high-quality clinical trials and bioinformatics resources with expert curation.
- Creation of the prototype Eat4Genes online app that enables users to find dietary guidance by condition or by genes with powerful visualizations that show the findings and the rigorous scientific evidence supporting them with an emphasis on human whole foods and extracts ingestion studies.
- Development of preliminary ranking system for presentation of DRGT results based on the quality of scientific evidence.
- Demonstrations of how DRGT research could be translated into practical applications to help improve patient health.

To promote research into DRGT, the Eat4Genes web app, code repository, and DRGT dataset are publicly available [19],[18].

2 METHODS

2.1 The DRGT Dataset

The Eat4Genes DRGT dataset (see Figure 1) consists of five linked tables that capture the relationships between nutrients, scientific studies, conditions, genes, and gene expression. The data sources are described in the next section. the `condition2Gene` table links a condition to risk genes. Eat4Genes uses this to enable users to search for appropriate nutrients. `condition2Gene` contains the condition, the condition category, the gene, and the direction of expression desired, i.e. should the gene be up- or down- regulated.

Eat4Genes' dietary suggestions are based rigorous studies of how nutrients change gene expression curated by Dr. Crawford. Nutrients are whole foods or food extracts that can be consumed by humans. The dataset contains 39 different nutrients. The table `nutrientInfo` contains information on the nutrients. It links the nutrient to the category (either whole food, whole food extract, or phytonutrient), a brief description of the nutrient, a link to an external webpage with more information about the nutrient, and a link to a small icon to represent the nutrient.

We harmonized all the gene names in the DRGT dataset to maximize the number of genes matching between conditions and studies. In the DRGT dataset, all unique gene symbols from our data pool were passed through the HGNC multi-symbol checker tool to find their approved gene symbol and full gene name. The `geneNames` table of 8321 gene names was created to form a consistent naming convention. Figure 1 shows its structure.

The DRGT dataset currently contains 67 different studies. The `studyData` table includes the short name of the study; the nutrient; the full name of the study; a brief summary; the GEO accession number used for the expression data (if available); a link to the study; the method used to convert gene expression to significantly up, down, or neutral; whether the study was in vivo or in vitro; the consumption method; the type of subject; the concentration of the nutrient used; the sample size of the study; and the numerical ranking. As discussed in Section 2.4, much of this information is used to indicate the quality of the study. In vivo human studies (that is, clinical trials in which humans ingest nutrients and then have their changes in gene expression analyzed) are considered the highest quality studies because their results are most likely to translate into practice. In vitro human studies (that is, studies of how nutrients change gene expression in human cell lines) are also considered. In vitro studies do not necessarily translate into in vivo effects, so they are considered of lesser quality. We calculated the nutrient concentrations as a percentage of the daily recommended value or serving size. The concentrations were binned into four categories: Low <75%, Moderate 75 % - 150%, High 150% - 250%, and Extremely High >250%. Similarly we binned the samples sizes into three categories: Low 1-5, Moderate 5-24, and High 25+.

`geneData` captures our analysis of which genes are significantly up-regulated, down-regulated, or not significantly changed in each study. Note `studyData` indicates the method used for gene expression analysis since the available data/reported results vary for each paper. The `geneData` table contains the gene name; a short name for the study; the p-value of the significance test; the log2 fold change (L2FC); and the expression in terms of up, down, or neutral.

2.2 DRGT Data Collection

The DRGT dataset has three primary data sources: condition-to-gene mappings created by Dr. Dana Crawford of Albany Medical College; a list of in vivo studies curated by Dr. Crawford's lab; and in vitro studies and results available from NutrigenomeDB. The `condition2Gene` table was created by Dr. Crawford based on his prior work on DRGT [2] and an extensive database search. In general, condition risk genes were identified through independent compilations from PubMed and known pharmaceutical targets as well as using the NCBI resource "Genes and Disease" [6]. They identified 96 key risk genes for a wide range of conditions (e.g., HMGCR for Hypercholesterolemia) for our app.

Crawford Lab in vivo Dataset: We provide a brief description of resources provided by Dr. Crawford's lab and leave a fuller investigation and validation of these resources to future work. To assess the effect of dietary agents/nutrients on gene expression, the Crawford Lab carried out extensive public database mining, searching various databases including the National Center for Biotechnology Information's (NCBI) Gene Expression Omnibus (GEO) [21]; PubMed [9]; Google Scholar [1]; clinical trials published through the U.S. Library of Medicine [17]; the Cochrane library [4]; and the European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI) [14]. To identify modulated genes, they used the Boolean search operator AND in combination with the keywords "human", "gene expression", and, separately, each dietary

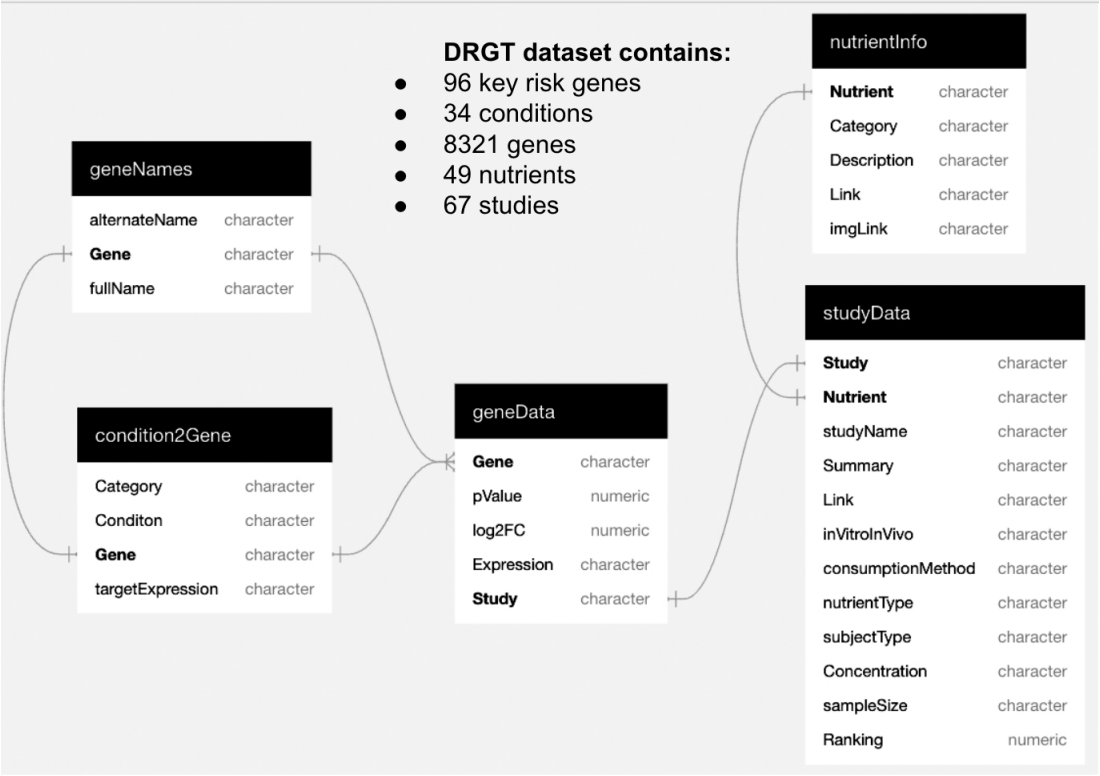


Figure 1: Eat4Genes Dietary Rational Gene Targeting (DRGT) dataset

agent of interest. Where necessary, the NCBI GEO2R web tool was used to analyze noncurated data [7].

Forty-two human whole food or extract gene expression studies were identified in the examined databases. Despite varying conditions (e.g., from four hours to eight-week diets), the number of modulated genes was reasonably consistent (1061 ± 227 SEM) for those studies examining whole genome expression. Human gene expression studies were only found for 13 of 34 searched whole foods, underscoring the need for more such studies. While these analyses have allowed us to construct Eat4Genes, they also revealed a surprising lack of studies translatable to humans in vivo. Instead, beyond the above 42 whole food studies, most dietary-related gene expression studies found used human cells in culture and non-physiological concentrations of dietary agents, and human in vivo ingestion studies with high-dose, poorly bioavailable purified phytonutrient supplements (e.g., curcumin). When conducting their mining analyses, they also noted that the GEO and PubMed databases contained almost all of such data; i.e., this data was not scattered throughout the various databases we utilized.

NutrigenomeDB: The NutrigenomeDB web resource is based on a curated set of studies. Most are in vitro studies, but some in vivo studies are included. We cloned the NutrigenomeDB relational database from the publicly-available github repository[1] and exported its `nutrigenomic` table. The resulting dataset consisted of 48 studies involving two whole foods and 24 nutrients. The NutrigenomeDB team had already performed a differential expression analysis of nutrigenomics experiments performed on human cells found in the Gene Expression Omnibus (GEO) repository. Therefore, we were able to incorporate these results directly into the app.

[1]NutrigenomeDB github URL: https://github.com/rmartin84/NutriGenomeDB

2.3 Data Cleaning and Dataset Preparation

Gene expression data from each study was inspected and harmonized to form the DRGT dataset. For each selected paper, we extracted the results of the gene expression analysis that included the fold change and p-value for tests of differential expression as reported in the paper or its supplementary material. We kept notes of how this was done in `studyData`. The `geneData` table contains the gene expression data for all of the studies. The gene expression data was extracted from the studies and converted to $\log_2$ fold change (L2FC). Then, using the L2FC and the p-value, gene expression was converted to either up, down, or neutral according to the following:

$$\text{Expression} = \begin{cases} \text{Up} & \text{if L2FC} \geq 0.263 \text{ and p-value} < 0.05 \\ \text{Down} & \text{if L2FC} \leq -0.263 \text{ and p-value} < 0.05 \\ \text{Neutral} & \text{if p-value} \geq 0.05 \\ \text{Neutral} & \text{else} \end{cases}$$

For the prototype Eat4Genes app, we chose commonly used values for L2FC and p-value thresholds. The up threshold of L2FC > .263 corresponds to greater than 1.2 fold-change. The down threshold of L2FC < -.263 was selected to be symmetric. In future versions of Eat4Genes, user-defined threshold selection could be enabled.

By "harmonizing" the gene names using `geneNames`, we were able to find more papers on a large number of condition risk genes.

2.4 Ranking System

The general goal of the ranking system is to provide a metric to assess the confidence in potentially achieving a desired risk gene modulation with health dietary agents. The ranking is used as a means to order the presentation of results to the user in the app. Functionally, the ranking is used in the visualization aspect of the app to create visual differentiation between higher-quality data and lower-quality data by having the suggestions that come from higher-quality data presented more prominently to the user. Two formulas were created: one to rank the quality of data found in an individual study and one to rank the quality of a combination of studies that all look at the same gene. The proposed ranking formulas should be regarded as prototypes. In a complete Eat4Genes app, these ranking metrics could be customized to the user's needs. Note that Eat4Genes always provides information and links for each study so that the user can investigate them directly.

A ranking for an individual study is based on six characteristics of the study: whether the study was in vitro or in vivo, the consumption method, the type of nutrient consumed, the subject type, the concentration, and the sample size. The consumption method is only applied to in vivo studies to further classify how the in vivo study was conducted. Most in vivo studies in the dataset are classified as "whole" to signify that a nutrient was consumed directly by a subject; however, some studies have a consumption method of "pill" or "other". For example, in one study, a soy protein power is mixed into water or juice [23].

We account for different types of nutrients in the ranking. These types are different from the three categories of nutrients shown in the dietary guide. From highest to lowest scored: whole food, whole food extracts (such as extra virgin olive oil), simple extract (extracts from whole foods), complex extracts (such as green tea and coffee), dietary supplements, dietary polyphenols, and bacteria.

Thus far, the focus has been on human subjects and cells, but nonhuman studies can be added later if desired. Some studies used very high concentrations of nutrients that may not be realistically consumed. Lower concentrations that achieve effective results are better. The concentration was discretized by comparing the dose of the nutrient given to the subjects and comparing it to the recommended daily value or serving size depending on the study.

This prototype ranking system is subject to change. Obtaining the perfect ranking system would require extensive human double-blind placebo-controlled studies. The Eat4Genes prototype ranking system is intended to show general quality based on factors that were determined as important by the team. The current model is:

$$
\begin{aligned}
\text{Ranking} = &\begin{cases} 50 & \text{if Type = Whole Food} \\ 40 & \text{if Type = Whole Food Extract} \\ 40 & \text{if Type = Simple Extract} \\ 35 & \text{if Type = Complex Extract} \\ 20 & \text{if Type = Dietary Supplements} \\ 20 & \text{if Type = polyphenol} \\ 0 & \text{else} \end{cases} + \\
&\begin{cases} 20 & \text{if VitViv = in vivo} \\ 0 & \text{if VitViv = in vitro} \end{cases} + \\
&\begin{cases} 20 & \text{if Consumption = Whole} \\ 0 & \text{else} \end{cases} + \\
&\begin{cases} 8 & \text{if Concentration = Moderate} \\ 4 & \text{if Concentration = Low} \\ 2 & \text{if Concentration = High} \\ 0 & \text{if Concentration = Extremely High} \end{cases} + \\
&\begin{cases} 2 & \text{if Sample Size = High} \\ 1 & \text{if Sample Size = Moderate} \\ 0 & \text{if Sample Size = Low} \end{cases} + \\
&\begin{cases} 10 & \text{if Subject = Human} \\ 0 & \text{else} \end{cases}
\end{aligned}
$$

The combination ranking formula takes in the number of studies that all look at the same nutrient and the mean of the scores for each study. Nutrients that had more studies supporting evidence for them got higher ranks than nutrients that had fewer studies. The mean of the scores was important to take into consideration because even if a nutrient had many studies carried out, this does not necessarily mean all of the data was high quality. The mean was a good measure of how strong all the studies are as a whole. The mean of the scores for the studies is significantly higher weighted than the number of studies. The combination ranking is as follows:

$$
\begin{aligned}
\text{Combination Ranking} = &.7 * \text{Average Ranking} + \\
&\begin{cases} .3 * 25 & \text{if Number of Studies = 1} \\ .3 * 50 & \text{if Number of Studies = 2} \\ .3 * 75 & \text{if } 3 \leq \text{Number of Studies} \leq 5 \\ .3 * 100 & \text{if Number of Studies} < 5 \end{cases}
\end{aligned}
$$

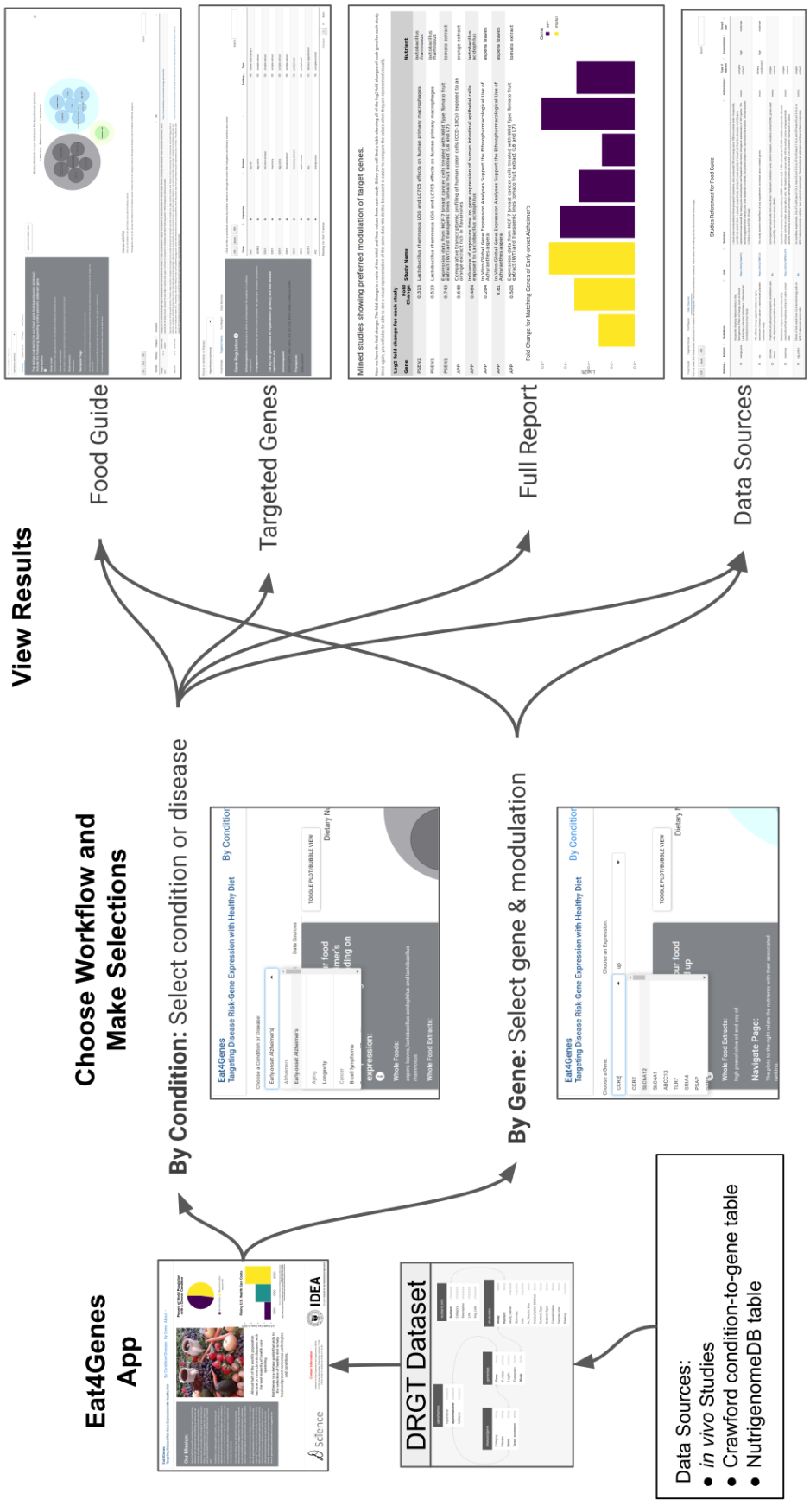


Figure 2: Eat2Genes App Flow: Consuming "Dietary Rational Gene Targeting" data; choosing a workflow; selecting a condition or gene and modulation; and viewing results

2.5 Eat4Genes App Implementation

Eat4Genes is an interactive webpage that serves as a food guide to help treat or prevent specific conditions based on the targeting of key genes. Figure 2 shows the basic structure of Eat4Genes. Users first enter a home page that describes the app. Users then choose to either get nutrient suggestions for a condition by selecting *By Condition* or for a specific gene by selecting *By Gene*. The results are presented in four summary views: *Food Guide*, *Targeted Genes*, *Full Report*, and *Data Sources*. Each view is described below, and screen shots for these pages are provided in the Results Section.

Currently, Eat4Genes is targeted to two different demographics: (1) users seeking fewer details, such as less technically-adept patients or community members, and (2) those seeking more details, such as healthcare provider, interested patients, and biomedical researchers. The app is organized so that the main results are provided on the second page that the user sees; as the user advances through sub-pages, higher-level details based on their selections are provided. The main results page also has two different visualizations for the user to choose from, as well as text results.

Eat4Genes was created using Shiny [3], a framework for easily implementing web applications based entirely on R [20]. The prototype app has been publicly deployed [19] and its source code and DRGT dataset are available via a public github repository [18].

Home Page: The user first sees the Home page. The goal of this page is to relay the purpose of our app as well as to provide sufficient evidence for the need for our food guide. The left panel contains our mission statement, which is also the introduction to this paper. The right column of the home page presents some evidence for the necessity of our food guide.

By Condition Page: Users can search for results based on a specified condition or disease selected through a drop down menu. Results will then show all of the nutrients that modulate the expression of key risk genes and in a direction that benefits a particular condition. This page has four sub pages: Food Guide, Targeted Genes, Full Report, and Data Sources.

Food Guide Page: The left panel of the Food Guide page shows the dietary nutrients in their food guide classified by category. Beneath that, there are instructions on how to navigate the page and a brief explanation of the rankings. On the right side of that, the user is shown the visualization model. Under that, there is instruction on how to interact with the visualization. Underneath the main section of the page, the user can view their food guide in table form. It contains, for each nutrient: the nutrient, the ranking, the category, a brief description, and a link to an external webpage.

Targeted Genes Page: The main result of the Targeted Genes page is a table showing the matches of key risk genes for the condition. It contains, for each matching gene: the gene, the direction of modulation, the nutrient, the ranking, and the type.

On the left side of the page, a panel shows the user's information about the table. It explains to them up and down regulation of gene expression. It also shows the user the key risk genes found in the Eat4Genes database for the condition they have selected.

Full Report Page The Full Report provides a summary of the data given to the user on the app in more detail, intended for healthcare providers and others interested in a more in-depth review of the data pooled to create the food guide.

The user is first shown the key genes analyzed for their selected condition and the target expression. They then see the mined studies showing the preferred modulation of those targeted genes. The generated table provides more information than has been shown thus far in the app. It includes the gene, the L2FC, the name of the study, and the nutrient. Beneath this table is a bar plot of the L2FC for each matching gene. The user is also shown the statistical significance of each matching gene, in a table that includes the gene, the p-value, the study name, and the nutrient. Again, a bar plot is shown beneath the table of the p-values for each matching gene. Next, the user is shown a plot of $-\log_{10}(\text{p value})$ versus L2FC, in which the user can see the statistical significance and fold change all at once. This plot is also color-coded by the nutrient and the points are labeled with the gene. Finally, users see a summary of the studies analyzed. The user has the option to download this report in PDF format.

Data Sources Page: This page provides the user with the studies referenced for the food guide that they have received. Here the user can see, for each study: the ranking, the nutrient, the name of the study, a link to the study, a brief summary, whether the study was in vitro or in vivo, the type of nutrient, the concentration of the nutrient used in the study, and the sample size of the study.

By Gene Page: The users have the option to search for results by specifying a gene and desired expression direction: up or down. The results show only the nutrients that affect that gene's expression in that specified direction. This page has two sub pages: Food Guide and Data Sources. These pages have the same layout as their counterparts in the By Condition page.

About Page: This page provides the user with a summary of the mission of the Eat4Genes app and details such as an explanation of the ranking system presented in Section 2.3 and the definitions we use for each categorization of food types. The key definitions to find there include *whole food extracts* (whole foods that are made using an extraction process or extracts from whole foods, such as olive oil and fish oil) and *phytonutrients* (purified plant nutrients such as commercial polyphenol supplements).

Visualizations: To make results accessible to a wider audience, Eat4Genes provides visualizations of the main results of the app: the food guide and the quality of the food guide. We used bar charts and bubble charts to enable a choice of views (Figures 4 and 4). In the bubble chart, the size of each circle is the value of combined ranking which is computed based on all the evidence for that nutrient and condition combination. To show users the different types of recommended nutrients, the app uses a hierarchical packed bubble graph, which groups the bubbles by category.

The package used to create this visualization is *highcharter* [5], an R-based wrapper of the Highcharts [12] Javascript library. Highcharts uses series to create its plots. To have the results separated by category, each category of nutrients needed to be in its own series. Using `nutrientInfo`, more information was added to our visualization. When the user hovers over a bubble, the name of the nutrient, a small image, the ranking, and a description with important information about the nutrient are shown. When the user clicks on the bubble, they are directed to a website with more nutritional information.

3 RESULTS

3.1 Patient Consultation Scenario

We illustrate the utility of the app in two usage scenarios. In this scenario, the key risk gene responsible for an individual's condition is known. For example, someone diagnosed with hypertension, or a healthcare professional on their behalf, will consult the Eat4Genes app when looking to obtain dietary guidance for hypertension as an alternative or complement to pharmaceuticals. On the app's home page, two workflow options are presented: search by a condition or search by a specific gene. Since this user is searching by a condition, the *By Condition* option is selected at the top of the page and the user is prompted to select their condition.

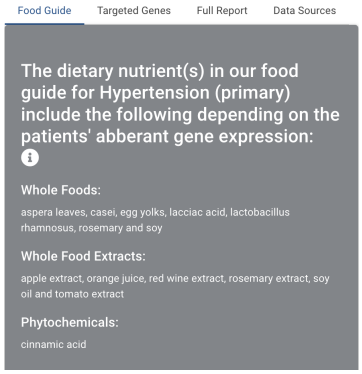


Figure 3: The list of nutrients found for Hypertension.

On the left-hand side of the Food Guide page, the user sees a list of the dietary nutrients, sorted by category (Figure 3). On the right-hand side, the user sees a plot of the suggested foods, ordered by ranking, and colored by nutrient category. For this scenario, orange juice, egg yolks, soy, soy oil and rosemary are among the top suggested foods.

The user can then download an image file of their results if desired (Figure 4). At the bottom of this page, the healthcare provider can view a table of results used to create the food plots. They could use this information to explain the food guide to their patient.

For some users the interactive bubble view, annotated with additional information, may be more effective and compelling. From Food Guide, the user simply selects Bubble View to view nutrients in the form of a bubble plot on the right hand side (Figure 4).

A user, exploring the bubble plot,clicks to remove the "Phytochemicals" category. Only whole foods and whole food extracts are now displayed. They examine their bubble plot, in which the larger bubbles represent the relatively stronger suggestions and the nutrients are sorted by category. They hover over rosemary extract to learn more, where they can view the name of the nutrient, a small image, the relative ranking, and a brief description (Figure 5). Then they click on the bubble, which opens a new tab with even more information about the nutrient. They can download all plots.

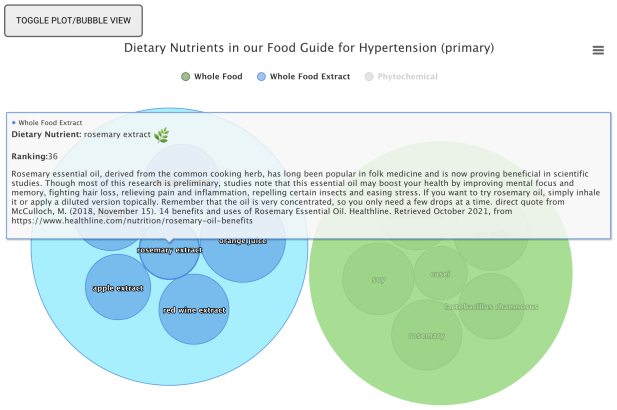


Figure 5: The information box that appears when the user hovers over *rosemary extract* in the bubble plot visualization for hypertension.

The health professional can explore which genes were targeted by these foods on the Targeted Genes page.

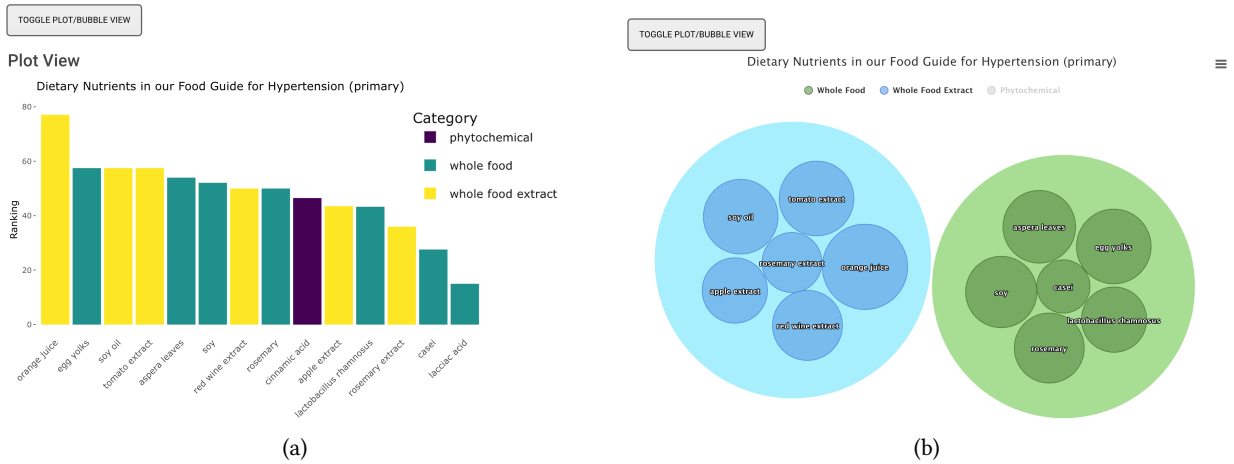
Now, the healthcare provider is interested in the evidence behind the recommendations. All the studies used for the results are shown by navigating to the Data Sources sub-page. Figure 7 shows the Data Sources table generated for hypertension that includes basic information about each study, the study title, and links to the published paper. Here the provider sees that the results are based on fourteen studies, two in vivo and twelve in vitro. The top in vitro study examined orange juice as a potential protective agent for cardiovascular diseases [16], making it more likely to be directly relevant for hypertension. The second in vitro study examined gene expression differences caused by soy in patients with breast cancer [23]. The provider would have to judge whether this study on cancer patients is relevant to address hypertension for the patient being considered. Eat4Genes enables the provider to rapidly identify scientific evidence for DRGT, including key factors on study quality, so that they can make these decisions.

3.2 Researcher Scenario

We now consider a scenario in which the user is a research scientist studying a particular gene, e.g. the hypertension risk gene ACE [11]. They might ask which nutrients may have a beneficial modulation of that gene (reduction in this case) to potentially treat hypertension. Navigating to the top of the page, they can then select the By Gene option. The user will be given the option to search for a specific gene and indicate what expression direction they want to focus on. This in turn leads to the sub-page "Food Guide" again, but this time the results use only the ACE gene to identify healthy dietary agents that modulate its expression down. These results are listed in the panel on the left side of the page (Figure 8).

The researcher now sees nutrients associated with down modulation of ACE. In the bubble chart (Figure 8), the researcher can see each nutrient, as well as the strength of evidence for this nutrient, as displayed by both the relative ranking and the size of the nutrient's bubble. The researcher can also use the toggle button to see results in the plot view.

This can help researchers decide which nutrients to include in human studies. This researcher may want to read the original study



(a) (b)

Figure 4: Food Guide bar (a) and bubble (b) plots showing suggested foods for hypertension along with their rankings and categories.

Click on the up and down arrows by column names to change the order that risk genes and dietary nutrients are sorted

CSV Excel PDF Search:

Gene	Expression	Nutrient	Ranking	Type
ACE	↓	orange juice	78	complex extract
ACE	↓	lactobacillus rhamnosus	19	bacteria
ACE	↓	soy oil	50	whole food extract
ADRB2	↓	lactobacillus rhamnosus	19	bacteria
ADRB2	↓	egg yolks	50	simple extract
AGTR1	↓	soy	53	dietary supplement
AGTR1	↓	cinnamic acid	45	complex extract
EDN1	↓	casei	18	bacteria
EDN1	↓	aspera leaves	45	complex extract

Figure 6: Summary table on *Targeted Gene* tab for user selection of hypertension.

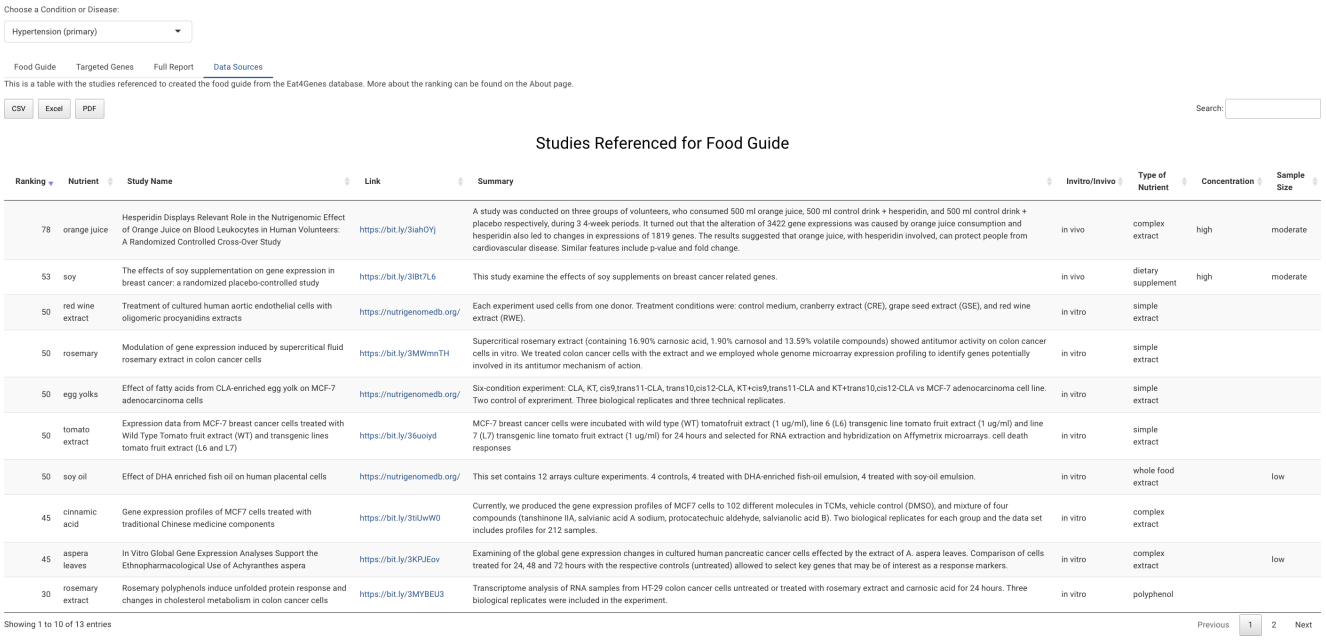
the data came from and view that study's data. The Data Sources sub page (Figure 9) includes a table of relevant characteristics of each study, such as sample size and whether the study was in vivo or in vitro. The user can click to find each study's paper.

4 DISCUSSION

Eat4Genes represents a significant step towards translating the DRGT approach into a clinical deliverable in the form of a dietary guide for healthcare providers, patients, researchers, and the general community. Toward this end, we have carried out a large scale bioinformatic mining of data from multiple public gene expression databases to compile a list of genes modulated by numerous dietary nutraceuticals with a documented history of healthy benefit in rigorous research studies. The prototype DRGT dataset is available as an expandable resource for DRGT research and applications. The prototype Eat4Gene app enables users to find dietary guidance by condition or by genes with visualizations that show the findings along with rigorous scientific evidence supporting them.

The prototype Eat4Genes dataset and app serve as a foundation for future work on a dietary treatment guide for patients. It is the first dietary app, to our knowledge, to focus on how whole foods change gene expression appropriately for conditions based on key condition risk genes supported by rigorous scientific studies. This resource is an extension of the digital web model created by Martin-Hernandez et al [15] and Zheng et al [24]. However, additional research is required. For example, the DRGT dataset should be expanded and validated and usability studies should be performed.

Eat4Genes provides the basic foundation for a dietary suggestion ranking system. We emphasize that implementation of exact ranking is not possible until a significant number of quality in vivo clinical studies are carried out that confirm dietary suggestions both at the level of target gene modulation and associated health benefit. There are simply not close to this number of available such studies - including replicated studies by independent groups - at this time. Nonetheless, we are still able to provide a strategy for such a ranking system that we can provide significant value in eventually ranking dietary suggestions, and is something we envision


Choose a Condition or Disease:

Hypertension (primary)

Food Guide Targeted Genes Full Report Data Sources

This is a table with the studies referenced to created the food guide from the Eat4Genes database. More about the ranking can be found on the About page.

CSV Excel PDF Search:

Studies Referenced for Food Guide

Ranking	Nutrient	Study Name	Link	Summary	Invitro/Invivo	Type of Nutrient	Concentration	Sample Size
78	orange juice	Hesperidin Displays Relevant Role in the Nutrigenomic Effect of Orange Juice on Blood Leukocytes in Human Volunteers: A Randomized Controlled Cross-Over Study	https://bit.ly/3iah0Yj	A study was conducted on three groups of volunteers, who consumed 500 ml orange juice, 500 ml control drink + hesperidin, and 500 ml control drink + placebo respectively, during 3 4-week periods. It turned out that the alteration of 3422 gene expressions was caused by orange juice consumption and hesperidin also led to changes in expressions of 1819 genes. The results suggested that orange juice, with hesperidin involved, can protect people from cardiovascular disease. Similar features include p-value and fold change.	in vivo	complex extract	high	moderate
53	soy	The effects of soy supplementation on gene expression in breast cancer: a randomized placebo-controlled study	https://bit.ly/3lBt7L6	This study examine the effects of soy supplements on breast cancer related genes.	in vivo	dietary supplement	high	moderate
50	red wine extract	Treatment of cultured human aortic endothelial cells with oligomeric procyanidins extracts	https://nutrigenomedb.org/	Each experiment used cells from one donor. Treatment conditions were: control medium, cranberry extract (CRE), grape seed extract (GSE), and red wine extract (RWE).	in vitro	simple extract		
50	rosemary	Modulation of gene expression induced by supercritical fluid rosemary extract in colon cancer cells	https://bit.ly/3MWmnTH	Supercritical rosemary extract (containing 16.90% carnosic acid, 1.90% carnosol and 13.59% volatile compounds) showed antitumor activity on colon cancer cells in vitro. We treated colon cancer cells with the extract and we employed whole genome microarray expression profiling to identify genes potentially involved in its antitumor mechanism of action.	in vitro	simple extract		
50	egg yolks	Effect of fatty acids from CLA-enriched egg yolk on MCF-7 adenocarcinoma cells	https://nutrigenomedb.org/	Six-condition experiment: CLA, KT, cis9,trans11-CLA, trans10,cis12-CLA, KT+cis9,trans11-CLA and KT+trans10,cis12-CLA vs MCF-7 adenocarcinoma cell line. Two control of expreriment. Three biological replicates and three technical replicates.	in vitro	simple extract		
50	tomato extract	Expression data from MCF-7 breast cancer cells treated with Wild Type Tomato fruit extract (WT) and transgenic lines tomato fruit extract (L6 and L7)	https://bit.ly/36uoiyd	MCF-7 breast cancer cells were incubated with wild type (WT) tomatofruit extract (1 ug/ml), line 6 (L6) transgenic line tomato fruit extract (1 ug/ml) and line 7 (L7) transgenic line tomato fruit extract (1 ug/ml) for 24 hours and selected for RNA extraction and hybridization on Affymetrix microarrays. cell death responses	in vitro	simple extract		
50	soy oil	Effect of DHA enriched fish oil on human placental cells	https://nutrigenomedb.org/	This set contains 12 arrays culture experiments. 4 controls, 4 treated with DHA-enriched fish-oil emulsion, 4 treated with soy-oil emulsion.	in vitro	whole food extract		low
45	cinnamic acid	Gene expression profiles of MCF7 cells treated with traditional Chinese medicine components	https://bit.ly/3tiUwW0	Currently, we produced the gene expression profiles of MCF7 cells to 102 different molecules in TCMs, vehicle control (DMSO), and mixture of four compounds (tanshinone IIA, salvianic acid A sodium, protocatechuic aldehyde, salvianolic acid B). Two biological replicates for each group and the data set includes profiles for 212 samples.	in vitro	complex extract		
45	aspera leaves	In Vitro Global Gene Expression Analyses Support the Ethnopharmacological Use of Achyranthes aspera	https://bit.ly/3KPJEov	Examining of the global gene expression changes in cultured human pancreatic cancer cells effected by the extract of A. aspera leaves. Comparison of cells treated for 24, 48 and 72 hours with the respective controls (untreated) allowed to select key genes that may be of interest as a response markers.	in vitro	complex extract		low
30	rosemary extract	Rosemary polyphenols induce unfolded protein response and changes in cholesterol metabolism in colon cancer cells	https://bit.ly/3MYBEU3	Transcriptome analysis of RNA samples from HT-29 colon cancer cells untreated or treated with rosemary extract and carnosic acid for 24 hours. Three biological replicates were included in the experiment.	in vitro	polyphenol		

Showing 1 to 10 of 13 entries Previous 1 2 Next

Figure 7: Summary of studies on *Data Source* tab for user selection of hypertension

Figure 8: Nutrients found for ACE modulated down represented as text (a) or bubble plot (b).

eventually progressing beyond suggestions to recommendations that consider multiple conditions and users preferences.

Eat4genes also supports scientists interested in translational research. These researchers can refer to our app for identification of healthy dietary agents that modulate their research gene(s) and condition(s) of interest. If such dietary agents are found, and they modulate a given gene of interest in the right direction, they have the potential of providing future therapeutic benefit for patients. This provides a new potential line of research for investigators, such as basic preliminary studies to see if the modulation identified in our app also occurs in a relevant cell culture line or in vivo model of interest and, if successful, follow-up clinical-related studies.

An important caveat is that this approach is tailored to conditions that are caused by aberrant expression or activity of one or a small number of key risk genes. Polygenic diseases would be difficult to treat this way due to the need for many modulating dietary agents, some of which may even interfere with the actions of others. Nonetheless, many conditions fit this description as evidenced by the many single genes targeted by pharmaceutical drugs.

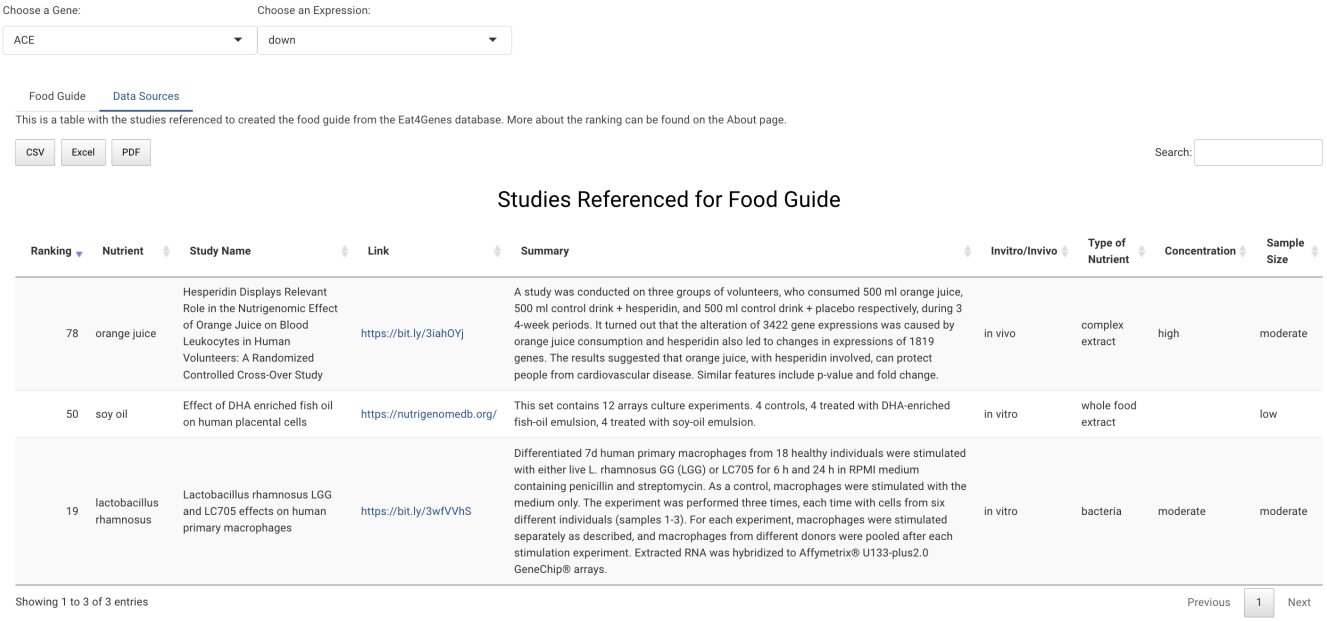


Figure 9: Studies used to produce results for ACE modulated down.

We emphasize that the information in Eat4Genes is intended only as a guide. As part of our overall strategy, we envision individuals consulting our app with guidance from their healthcare provider to obtain healthy dietary suggestions as discussed above. We then recommend considering a next step of testing these healthy dietary suggestions on themselves to see if the same modulation of a key risk gene also occurs in them as hoped. If so, this dietary agent can then be considered as a healthy lifestyle addition for their treatment.

ACKNOWLEDGEMENTS

Tamia Brefo, Jeffrey Chai, Jessica Coco, Shathni Ravindrababu, and Natasha Rosa contributed to earler Eat4Genes versions.

REFERENCES

[1] 2022. *Google Scholar.* Retrieved April 8, 2022 from https://scholar.google.com/
[2] J Bagyi, V Sripada, A M Aidone, H-Y Lin, E H Ruder, and DR Crawford. 2021. Dietary rational targeting of redox-regulated genes. *Free Radical Biology and Medicine* 173 (2021), 19–28. https://doi.org/10.1016/j.freeradbiomed.2021.07.021
[3] W Chang, J Cheng, JJ Allaire, C Sievert, B Schloerke, Y Xie, J Allen, J McPherson, Al Dipert, and B Borges. 2021. *shiny: Web Application Framework for R.* https://CRAN.R-project.org/package=shiny R package version 1.7.1.
[4] Cochrane. 2022. *Cochrane Library.* Retrieved April 8, 2022 from https://www.cochranelibrary.com/
[5] J Kunst et al. 2022. *CRAN - Package highcharter.* Retrieved April 1, 2022 from https://cran.r-project.org/web/packages/highcharter/index.html
[6] Ntnl Ctr for Biotech Info (US). 1998. *Genes and Disease.* Ntnl Ctr for Biotech Info (US). https://www.ncbi.nlm.nih.gov/books/NBK22183/
[7] Ntnl Ctr for Biotech Info (US). 2022. *About GEO2R.* Retrieved May 14, 2022 from https://www.ncbi.nlm.nih.gov/geo/info/geo2r.html
[8] Ctrs for Medicare & Medicaid Services. 2021. *Historical.* Retrieved May 12, 2022 from https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsHistorical
[9] Ntnl Libr for Medicine Ntnl Ctr for Biotech Info. 2022. *PubMed.* Retrieved April 8, 2022 from https://pubmed.ncbi.nlm.nih.gov/
[10] C Hajat and E Stein. 2018. The global burden of multiple chronic conditions: A narrative review. *Preventative Medicine Reports* 12 (Oct 2018), 284–293. https://doi.org/10.1016/j.pmedr.2018.10.008
[11] GB HealthWatch. 2022. *Hypertension and the ACE gene.* Retrieved May 14, 2022 from https://www.gbhealthwatch.com/GND-Hypertension-ACE.php
[12] AS Highsoft. 2022. *Highcharts JS JavaScript charting library.* Retrieved May 10, 2022 from https://www.highcharts.com/
[13] S Hushmendy, L Jayakumar, AB Hahn, D Bhoiwala, DL Bhoiwala, and DR Crawford. 2009. Select phytochemicals suppress human T-lymphocytes and mouse splenocytes suggesting their use in autoimmunity and transplantation. *Nutr. Res.* 29, 8 (Aug. 2009). https://doi.org/10.1016/j.nutres.2009.08.003
[14] European Molecular Biology Laboratory. 2022. *EMBL's European Bioinformatics Institute.* Retrieved April 8, 2022 from https://www.ebi.ac.uk/
[15] Roberto Martín-Hernández, Guillermo Reglero, José M Ordovás, and Alberto Dávalos. 2019. NutrigenomeDB: a nutrigenomics exploratory and analytical platform. *Database* 2019 (2019). https://doi.org/10.1093/database/baz097
[16] D Milkenkovic, C Deval, C Dubray, A Mazur, and C Morand. 2011. Hesperidin Displays Relevant Role in the Nutrigenomic Effect of Orange Juice on Blood Leukocytes in Human Volunteers: A Randomized Controlled Cross-Over Study. *PLOS ONE* (Nov. 2011). https://doi.org/10.1371/journal.pone.0026669
[17] U.S. Library of Medicine. 2022. *ClinicalTrials.gov.* Retrieved April 8, 2022 from https://clinicaltrials.gov/
[18] The Rensselaer Data INCITE Program. 2022. *Eat4Genes code and data repository.* Retrieved May 15, 2022 from https://github.rpi.edu/DataINCITE/Eat4Genes
[19] The Rensselaer Data INCITE Program. 2022. *Eat4Genes web app.* Retrieved May 15, 2022 from http://erickson.shinyapps.io/Eat4Genes/
[20] R Core Team. 2020. *R: A Language and Environment for Statistical Computing.* R Foundation Statistical Computing, Vienna, Austria. https://www.R-project.org/
[21] AE Lash R Edgar, M Domrachev. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30, 1 (Jan 2002), 207–210. https://doi.org/10.1093/nar/30.1.207
[22] SV Rajkumar. 2020. The high cost of prescription drugs: causes and solutions. *Blood Cancer Jrnl* 10, 71 (Apr 2020). https://doi.org/10.1038/s41408-020-0338-x
[23] M Shike, AS Doane, L Russo, R Cabal, JS Reis-Filho, W Gerald, H Cody, R Khanin, J Bromberg, and L Norton. 2014. The Effects of Soy Supplementation on Gene Expression in Breast Cancer: A Randomized Placebo-Controlled Study. *J Natl Cancer Inst.* 106, 9 (Sept. 2014). https://doi.org/10.1093/jnci/dju189
[24] T Zheng, Y Ni, J Li, BKC Chow, and G Panagioutou. 2017. Designing Dietary Recommendations Using System Level Interactomics Analysis and Network-Based Interference. *Front. Physiol.* 8, 753 (Sept. 2017). https://doi.org/10.3389/fphys.2017.00753
[25] PA Zmijewski, LY Gao, AR Saxena, NK Chavannes, SF Hushmendy, DL Bhoiwala, and DR Crawford. 2015. Fish oil improves gene targets of Down syndrome in C57BL and BALB/c mice. *Nutr. Res.* 35, 5 (May 2015). https://doi.org/10.1016/j.nutres.2015.02.007