FoodMine: Exploring Food Contents in Scientific Literature

Forrest Hooton,[1] Giulia Menichetti,[1,*] Albert-László Barabási[1,2,3,*]

[1]Network Science Institute, Northeastern University, Boston, MA, USA.

[2]Division of Network Medicine, Department of Medicine, Harvard Medical School, Boston, MA, USA.

[3]Department of Network and Data Science, Central European University, Budapest, Hungary.

* Corresponding authors. e-mail: menicgiulia@gmail.com, barabasi@gmail.com

Abstract

Thanks to the many chemical and nutritional components it carries, diet critically affects human health. However, the currently available comprehensive databases on food composition cover only 188 nutritional components that are essential for our health, a tiny fraction of the total number of chemicals present in our food. Indeed, thousands of other molecules, many of which have well documented health implications, remain untracked. To explore the body of knowledge available on food composition, we built FoodMine, an algorithm that uses natural language processing to identify papers from PubMed that potentially report on the chemical composition of garlic and cocoa. After extracting from each paper information on the reported quantities of chemicals, we find that the scientific literature carries extensive information on the detailed chemical components of food that is currently not integrated in databases. Finally, we use unsupervised machine learning to create chemical embeddings, finding that the chemicals identified by FoodMine tend to have direct health relevance, reflecting the scientific community's focus on health-related chemicals in our food.

Introduction

Decades of research in nutrition have documented the exceptional role of diet in health, unveiling the role of selected nutrients, like sugars, fats, proteins, vitamins, and other biochemical factors, as well as factors contributing to non-communicable diseases like deficiency diseases, cardiovascular disease, obesity, and diabetes mellitus. However, our ability to explore how food affects our health is severely limited by the lack of systematic knowledge on food composition. The most accurate data on food composition is maintained by USDA, tracking 188 biochemicals, often called nutritional components.[1,2] Yet, when it comes to the biochemical composition of the food we consume, these nutritional components represent only a tiny fraction of definable biochemicals reported in food. For example, FooDB, a database that integrates food composition data from databases like USDA, DTU, Duke, Phenol Explorer, and others catalogues altogether 26,625 compounds.[3–6] The majority of these compounds are only identified, with no information on their quantities in specific ingredients. For example, sulfides are reported to be present in the Allium family, like garlic or onion, but the precise quantities for important sulfides like diallyl disulfide (garlic) and dipropenyl sulfide (onion) remain unknown, despite their well-documented role in cancer prevention.[7–10] The current incomplete knowledge of the full biochemical composition of food impedes the research community's ability to uncover the mechanistic effect of the thousands of untracked molecules and their ultimate mechanistic roles in health, achieved either through the microbiome,[11] by contributing to the body's metabolism, or by regulating molecular processes in human cells.

The lack of centralized information on the chemical composition of food does not equal a lack of scientific or commercial interest in these chemicals: an exceptional amount of research focuses on identifying and quantifying the presence of certain chemicals in various foods, as well as the health implication and the biochemical role of specific food-borne chemicals. The problem is that data on the chemical composition of food is scattered across the multiple research literatures, spanning different scientific communities, from agriculture to food research, and from health sciences to biochemistry. While we witness notable efforts to mine this extensive literature and catalogue the scattered data into databases, like Phenol Explorer's focus on polyphenols or eBASIS's prioritization of human intervention studies,[12,13] we lack efforts to achieve this across the full food supply and chemicals.

The lack of systematic efforts to map out the existing information on food prompted us to ask how much information is really available on food composition. We developed FoodMine, a pilot project designed to systematically mine the scientific literature to identify and collect all the chemical compositional data for specific ingredients. Hence we demonstrate the capabilities offered by FoodMine by focusing on garlic and cocoa, foods with well documented health effects, which suggests the existence of a sizable yet scattered literature pertaining information on their chemical contents.[14,15] The knowledge gathered here serves as a pilot towards future comprehensive systematic efforts aimed at identifying and organizing the available information on the chemical composition of all food throughout the whole scientific literature.

Results

The FoodMine protocol leveraged the PubMed databases to systematically analyze the title and the abstract of the research papers related to garlic and cocoa (Fig 1).[16] We entered each food as a search term, obtaining 5,676 papers for garlic and 7,620 papers for cocoa. We subset the search results by applying text matching between predefined vocabularies, mesh terms and the abstract of the paper listed in the PubMed entry, narrowing the results to 415 papers for garlic and 475 papers for cocoa. After obtaining the subset of results, we manually accessed the paper if we could access a “full text link”, downloading 299 papers for garlic and 324 papers for cocoa. Finally, we manually evaluated each paper to identify relevant chemical contents and extract information from it. Of the 623 manually evaluated papers, 77 papers contained chemical composition data for garlic and 93 for cocoa, yielding 1,426 and 5,855 individual chemical measurements in total for garlic and cocoa, respectively (See Supplementary Material Section 1). In the resulting FoodMine database a compound is “quantified” when chemical measurements report absolute contents, and “unquantified” otherwise. Supplementary Fig S1 shows that the majority of papers for both foods contained only one or at most a few chemical measurements, referred to as records. However, an outlier for cocoa reported 960 records,[17] measuring the contents of several compounds in cocoa sourced from 15 different origins; the permutations of these variables resulted in the high number of data points. Another outlier examined a large spread of compounds related to human taste perception, reporting 68 unique compounds.[18] For garlic the outlier was a paper reporting 198 records.[19]

We integrated the compound records into single compound entries, and manually divided quantified entries into their respective compound class based on FooDB classifications, as shown in Supplementary Fig S2. We find that 'Carboxylic Acids and Derivatives' contains the most explored compounds for both garlic and cocoa, and the 'Flavonoids' class is in the top three for both ingredients. Compounds from these two classes are common in plant-based food, hence are expected to be present in garlic and cocoa. We also uncovered reports pertaining to various metallic classes, 'Toxins', and 'Pesticides'. Many compounds in the pesticides class came from a paper focusing on the pesticide residues in cocoa products from local markets in Southwestern Nigeria.[20] Despite its local focus, the examined compounds could directly affect health outcomes worldwide, as Nigeria is the world's 3rd largest exporter of cocoa.[21]

The FooDB and USDA databases allowed us to verify if the information recovered from the literature matches or contrasts the existing knowledge on the composition of these ingredients (See Supplementary Material Section 2). For this we merged different variations of garlic and cocoa within the USDA and FooDB databases, like merging "Garlic" and "Soft-necked Garlic" in FooDB when comparing the information to FoodMine. In USDA, all reported compounds are quantified, while FooDB lists both quantified and unquantified compounds. We consider a compound quantified if at least one absolute measurement is reported for the selected foods. Taken together, we find that FoodMine recovered more unique compounds than catalogued by USDA (Fig 2A and 2B), and more quantified compounds than catalogued by FooDB. While only 7-9% of compounds are quantified in FooDB and USDA for garlic and cocoa, through FoodMine we collected quantified information for 70% of garlic compounds and 66% of cocoa compounds (Supplementary Material Section 3). For cocoa and garlic, FooDB and USDA contain more unquantified compounds than quantified. However, we find that ~70% of the information reported in the literature was quantified, indicating that the literature contains an extensive body of information currently not recorded in databases (Supplementary Material Section 3).

Furthermore, 96 quantified garlic compounds and 283 quantified cocoa compounds are novel, meaning that they were not previously linked to the two ingredients in USDA or FooDB. In summary, 48% and 72% of quantified compounds are novel in both garlic and cocoa, respectively, hence the average increase in quantified measurements offered by FoodMine exceeds 137% (Supplementary Material Section 3). These findings suggest that a systematic mining of the information scattered in the scientific literature could significantly improve our current knowledge of food composition.

The most frequently reported compounds (Fig 3) in FoodMine are known to play important roles in health effects and flavor. For example, diallyl disulfide is known to contribute to garlic's smell and taste. More importantly, it is implicated in the health benefits of garlic, in particular garlic allergy.[22,23] Yet, neither USDA nor FooDB offers quantified information for the compound. This is not an isolated case, as Fig 3 shows FooDB and USDA lack information on other frequently explored compounds as well. The need to systemically characterize the nutrient profile of a large number of food items, as USDA does, misses information on those compounds that are specific to a few individual foods, despite the potential role they play in health. Indeed, three of the top ten compounds for cocoa are not quantified in FooDB and one is not listed, while for garlic, five of the top ten compounds are not quantified.

To understand the accuracy of the collected quantified data, we compared the FoodMine compound measurements to their corresponding values in USDA, which is considered the gold standard for measurement reliability. Given the limited nutrient panel reported by USDA, we were able to compare only 11% of the chemical compounds we recovered for garlic, and 5% for cocoa. The recovered information spanned a full spectrum of molecules, mixing compounds with both small and large relative mass (Fig 4). Overall, we find a good agreement between the FoodMine-recovered and the USDA-reported values (see Supplementary Material Section 3 for fit statistics supporting the identity relation $y = x$). Garlic has a logarithmic R-squared of .82, indicating a notable correlation between the known and the FoodMine records, while cocoa has a logarithmic R-squared of only .56. The lower correlation for cocoa is due to a group of amino acids, that came from papers that examined the contents of roasted cocoa, a processing step that systematically alters the quantities of many chemicals, potentially explaining the greater difference from the USDA measurements.[17,18] If we remove the data pertaining to roasted cocoa, we find an R-squared of .75.

To offer a more comprehensive understanding of the classes of chemicals we retrieved and their relation across different databases, we created chemical embeddings using the unsupervised machine learning tool Mol2Vec.[24] Chemical embeddings capture the structural similarity of the chemicals in a low dimensional space. Indeed, as shown in Supplementary Fig S3, when the chemical classification is known, chemicals belonging to the same class tend to be closer in the embedding space defined by Mol2Vec, suggesting that chemical embeddings successfully capture structural information. This process maps the structural knowledge of the compounds we retrieved, and can be used to integrate further information characterizing the compounds. For instance, given the interest in the association between food-borne chemicals and health outcomes, we can start from the Comparative Toxicogenomics Database (CTD) that reports manually curated associations between chemicals and diseases.[25] After matching the total number of health implications to each of the chemicals in FoodMine, FooDB, and USDA, we layered this information on the obtained Mol2Vec embedding (Fig 5). We find that FoodMine covers more chemicals with health effects than FooDB and USDA (see Fig 5 A vs B and C, D vs E and F), a difference particularly clear for cocoa (D vs E and F). Further, we find that the chemicals with health associations are more evenly dispersed throughout the embedding space for FoodMine, implying that FoodMine captures chemicals from rather different chemical classes. Overall, the FoodMine cocoa pilot has detected noticeably more organic, benzenoid, and hydrocarbon compounds, as seen by the absent spaces in E (USDA) and F (FooDB) compared to D (FoodMine, Fig 5) (Supplementary Material Section 3). In summary, compared to the existing databases, FoodMine detects more chemicals with health associations, distributed over a wider range of chemical classes, reflecting a selection bias in the literature: the research community appears to be more focused on chemicals with known health outcomes. Interestingly, there is no overlap between the papers contributing to FoodMine and those manually curated in CTD, meaning that we are recovering information from multiple scientific communities, not only health sciences (Supplementary Material Section 3).

Discussion

Our knowledge pertaining to the more than 26,000 chemicals expected to be present in food, as reported in various databases, is highly incomplete. This incompletion inspired our efforts to examine how much additional uncatalogued knowledge is scattered in the scientific literature. The invisibility of these compounds to experimental, clinical, epidemiological, and demographic studies – the virtual "dark matter" of nutrients – represents a major roadblock towards a systematic understanding of how diet affects our health. The introduced FoodMine pilot systematically scanned the scientific literature, identifying information about a large number of novel, quantified compounds reported by individual papers. We find that the collected information considerably extends our understanding of food composition. Furthermore, many of the recovered compounds have direct relevance to health and nutrition. For instance, the sulfides, quantified by FoodMine, are responsible for garlic's unique health effects, yet are currently not quantified in USDA or FooDB.

Garlic and cocoa are only two of the over a thousand natural ingredients commonly consumed by humans, hence our study supports the hypothesis that there is abundant information in the literature on the composition of other ingredients as well. Indeed, the search terms we used in FoodMine to retrieve papers from PubMed were narrow, and the selection of papers we manually evaluated is small compared to the total body of potential knowledge present in the literature. Consequently, likely there is additional information for garlic and cocoa, not yet captured by FoodMine. Other search terms, focusing on compound classes rather than foods, could uncover an additional body of information about the chemical composition of these ingredients, knowledge that can be generalized to other ingredients as well.

Methods

Literature mining consisted of three steps: search, selection, and information extraction. We began by searching PubMed with the search term 'garlic' and 'cocoa' using the Pubmed Entrez Programming Utilities API.[26] After retrieving the PubMed ID's for search results, we again used the API to retrieve information for each PubMed entry associated with the PubMed ID. We used text matching to scan each PubMed entry's mesh terms and abstract for words relevant to biochemicals, food, and pre-selected measurement methodologies after the API query (Supplementary Material Section 1). The algorithm filtered 5,676 results for garlic and 7,620 papers for cocoa to 415 and 475 results, respectively. We collected papers from the "Full text links" of the combined 900 entries. We skipped an entry if it did not list any "Full text links" or we did not have access to the paper associated with the links, recovering papers for 299 of the 415 and 324 of the 475 PubMed entries for garlic and cocoa, respectively. We also spot-checked search results that fell outside the assessment criteria to quantify the effectiveness of the filtration step. Of those, 0/10 papers contained relevant information for garlic and 3/10 papers contained relevant information for cocoa, indicating that the filtering did eliminate some papers with potentially relevant information.

Papers were individually read by a human assessor to determine whether or not they contained information on the chemical composition of garlic or cocoa. The assessment fell into three categories: "not useful" (papers not containing relevant for chemical contents in food), "quantified" (papers containing quantitative chemical composition that could be translated to unveil its precise contents in samples), and "unquantified" (paper containing chemical composition, but it could not be converted to unveil precise quantities). Examples of unquantified results were compounds detected by a mass spectrometer that only reported relative percentages, hence we could not record their contents in garlic or cocoa. In addition to the human mining procedure, we used machine learning to create a paper classification algorithm, helping to automatize future information collection. This algorithm takes as input the filtered samples and predicts which papers will contain information on the chemical content of cocoa or garlic. We applied the SMOTE sampling technique to balance the labeled data, as only 27% of the reviewed papers contained information on the chemical content of cocoa or garlic.[27] Our algorithm achieved an f1 score of 75.5% on the testing set, better than random. These results are in spite of a limited training set, and could improve with more labels from other foods than garlic and cocoa.

All records for a single unique compound were merged into a single entry. As different papers use different variations of a compound's name, we applied a chemical disambiguation scheme using PubChem CIDs to add keys to the compounds (Supplementary Material Section 2).[28] For each entry, we reported the average content value across all data points standardized in units of mg/100g, and captured additional statistics, such as the highest and lowest reported measurement of the chemical, variance across measurements, and number of measurements. Finally, we leveraged the PubChem CIDs to retrieve a string representation of the structural properties of the molecule (chemical SMILE) which we used as the input for Mol2Vec. Once we learned the vector representation for each chemical, we further reduced the dimensionality using TSNE to obtain the maps shown in Fig 5 and Supplementary Fig S3.[29]

Data Availability:

The raw data from FoodMine and processing code is available on our GitHub page.

https://github.com/fhooton/FoodMine

References

1. USDA. National Nutrient Database for Standard Reference, Release 28 (2015) Documentation and User Guide. **28**, (2015).

2. Bhagwat, S. & Haytowitz, D. B. USDA Database for the Flavonoid Content of Selected Foods Release 3.2 Prepared by. (2015).

3. FooDB. Available at: http://foodb.ca/about. (Accessed: 25th June 2019)

4. *Frida Food Data, version 1*. (2015).

5. U.S. Department of Agriculture, A. R. S. *Dr. Duke's Phytochemical and Ethnobotanical Databases*. (1992). doi:10.15482/USDA.ADC/1239279

6. Rothwell, J. A. *et al.* Phenol-Explorer 3.0: a major update of the Phenol-Explorer database to incorporate data on the effects of food processing on polyphenol content. *Database* **2013**, bat070–bat070 (2013).

7. Munday, R. & Munday, C. M. Relative Activities of Organosulfur Compounds Derived From Onions and Garlic in Increasing Tissue Activities of Quinone Reductase and Glutathione Transferase in Rat Tissues. *Nutr. Cancer* **40**, 205–210 (2001).

8. Nohara, T. *et al.* Antitumor Allium Sulfides. *Chem. Pharm. Bull. (Tokyo).* **65**, 209–217 (2017).

9. Wang, H., Yang, J.-H., Hsieh, S.-C. & Sheen, L.-Y. Allyl Sulfides Inhibit Cell Growth of Skin Cancer Cells through Induction of DNA Damage Mediated G2/M Arrest and Apoptosis. *J. Agric. Food Chem.* **58**, 7096–7103 (2010).

10. Nicastro, H. L., Ross, S. A. & Milner, J. A. Garlic and onions: their cancer prevention properties. *Cancer Prev. Res. (Phila).* **8**, 181–9 (2015).

11. Bashan, A. *et al.* Universality of human microbial dynamics. *Nature* **534**, 259–262 (2016).

12. Kiely, M. *et al.* EuroFIR eBASIS: application for health claims submissions and evaluations. *Eur. J. Clin. Nutr.* **64**, S101–S107 (2010).

13. Plumb, J. *et al.* eBASIS (Bioactive Substances in Food Information Systems) and Bioactive Intakes: Major Updates of the Bioactive Compound Composition and Beneficial Bioeffects Database and the Development of a Probabilistic Model to Assess Intakes in Europe. *Nutrients* **9**, (2017).

14. Garlic and Organosulfur Compounds. *Oregon State University* Available at: https://lpi.oregonstate.edu/mic/food-beverages/garlic. (Accessed: 25th June 2019)

15. Katz, D. L., Doughty, K. & Ali, A. Cocoa and chocolate in human health and disease. *Antioxid. Redox Signal.* **15**, 2779–811 (2011).

16. PubMed. *National Institutes of Health* Available at: https://www.ncbi.nlm.nih.gov/pubmed/. (Accessed: 25th June 2019)

17. J. Serra Bonvehí*, † and & Coll‡, F. V. Factors Affecting the Formation of Alkylpyrazines during Roasting Treatment in Natural and Alkalinized Cocoa Powder. (2002). doi:10.1021/JF011597K

18. Timo Stark, †, Sabine Bareuther, † and & Thomas Hofmann*, ‡. Molecular Definition of the Taste of Roasted Cocoa Nibs (Theobroma cacao) by Means of Quantitative Studies and Sensory

Experiments. (2006). doi:10.1021/JF0608726

19. Lee, J. & Harnly, J. M. Free Amino Acid and Cysteine Sulfoxide Composition of 11 Garlic (*Allium sativum* L.) Cultivars by Gas Chromatography with Flame Ionization and Mass Selective Detection. *J. Agric. Food Chem.* **53**, 9100–9104 (2005).

20. Oyekunle, J. A. O., Akindolani, O. A., Sosan, M. B. & Adekunle, A. S. Organochlorine pesticide residues in dried cocoa beans obtained from cocoa stores at Ondo and Ile-Ife, Southwestern Nigeria. *Toxicol. Reports* **4**, 151–159 (2017).

21. Verter, N. & Bečvářová, V. Analysis of Some Drivers of Cocoa Export in Nigeria in the Era of Trade Liberalization. *AGRIS on-line Pap. Econ. Informatics* **06**, 1–11 (2014).

22. Rao, P. S. S. *et al.* Diallyl Sulfide: Potential Use in Novel Therapeutic Interventions in Alcohol, Drugs, and Disease Mediated Cellular Toxicity by Targeting Cytochrome P450 2E1. *Curr. Drug Metab.* **16**, 486–503 (2015).

23. Garcia-Abujeta, J. L. *et al.* Allergic Contact Dermatitis to Diallyl Disulphide in Spain. *J. Allergy Clin. Immunol.* **117**, S130 (2006).

24. Jaeger, S., Fulle, S. & Turk, S. Mol2vec: Unsupervised Machine Learning Approach with Chemical Intuition. *J. Chem. Inf. Model.* **58**, 27–35 (2018).

25. Davis, A. P. *et al.* The Comparative Toxicogenomics Database: update 2019. *Nucleic Acids Res.* **47**, D948–D954 (2019).

26. Entrez Programming Utilities Help. (2010). Available at: https://www.ncbi.nlm.nih.gov/books/NBK25501/. (Accessed: 26th June 2019)

27. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002).

28. PubChem. *National Center for Biotechnology Information* Available at: https://pubchem.ncbi.nlm.nih.gov/. (Accessed: 25th June 2019)

29. Van Der Maaten, L. & Hinton, G. *Visualizing Data using t-SNE*. *Journal of Machine Learning Research* **9**, (2008).

Acknowledgements:

We thank Daniela Barbery-Flambury for the manual data collection and Michael Sebek for advice on compound classes.

Author Contributions:

FH performed data analysis, programming, and contributed to writing the manuscript. GM designed the project procedure and analysis, and contributed to writing the manuscript. ALB contributed to interpreting the results and writing the manuscript. GM and ALB conceived the project. FH and GM contributed equally to the project.

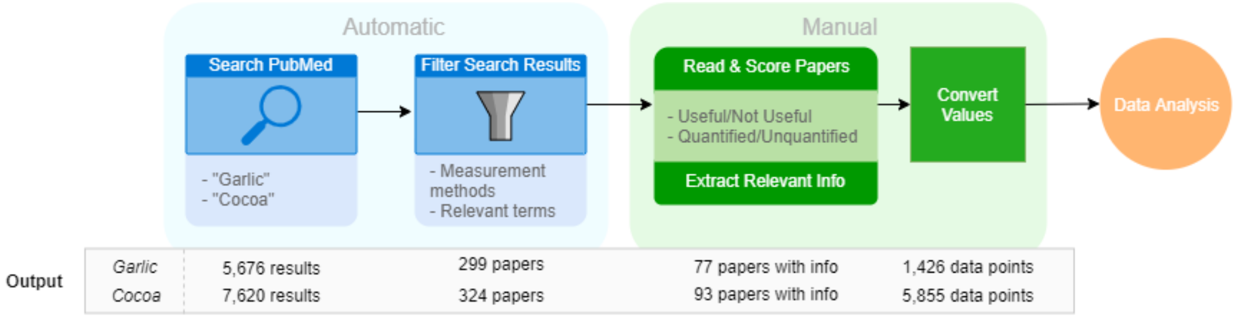
Competing Interests

ALB is the founder of Scipher Medicine, Foodome, and Nomix companies the leverage the application of big data in health.

Additional Information:

See Supplementary Material for more information.

Figures



Output					
	Garlic	5,676 results	299 papers	77 papers with info	1,426 data points
	Cocoa	7,620 results	324 papers	93 papers with info	5,855 data points

Figure 1: Overview of Data Collection Process. *Starting from PubMed, we retrieved a list of paper titles and abstracts using the Pubmed Entrez API, and then applied text matching to automatically filter the search results, obtaining a subset of papers, which were then read and manually evaluated. If papers contained information on the chemical content of cocoa or garlic, we manually extracted the relevant information. Finally, we converted values in comparable units. The "Output" bar shows the result of each step for garlic and cocoa.*

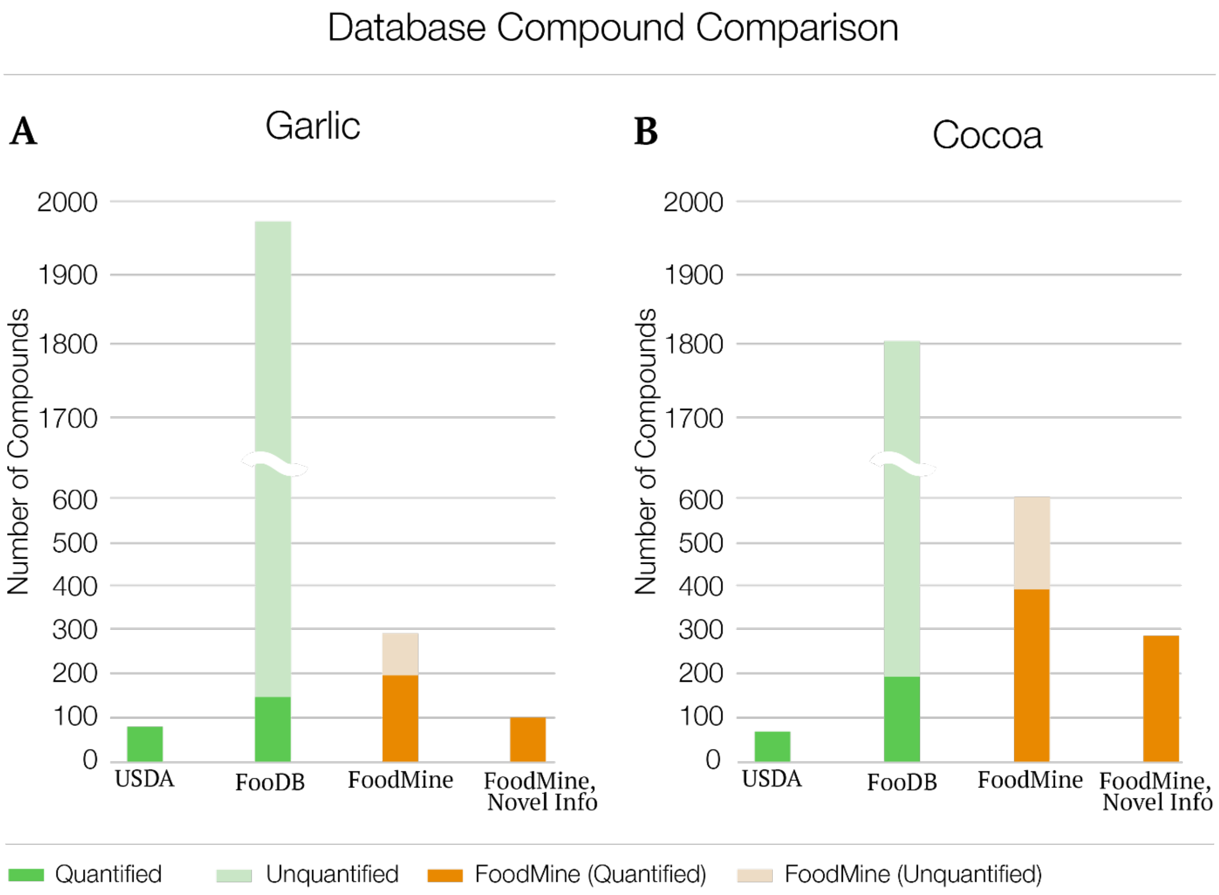


***Figure 2: Number of Unique Compounds Recovered by FoodMine, USDA, and FooDB.** The plots show the number of unique compounds reported by USDA, FooDB, and FoodMine. The columns display 1) the total number of unique quantified compounds, 2) the total number of unique unquantified compounds, and 3) the number of novel quantified compounds found by FoodMine for (A) Garlic and (B) Cocoa.*

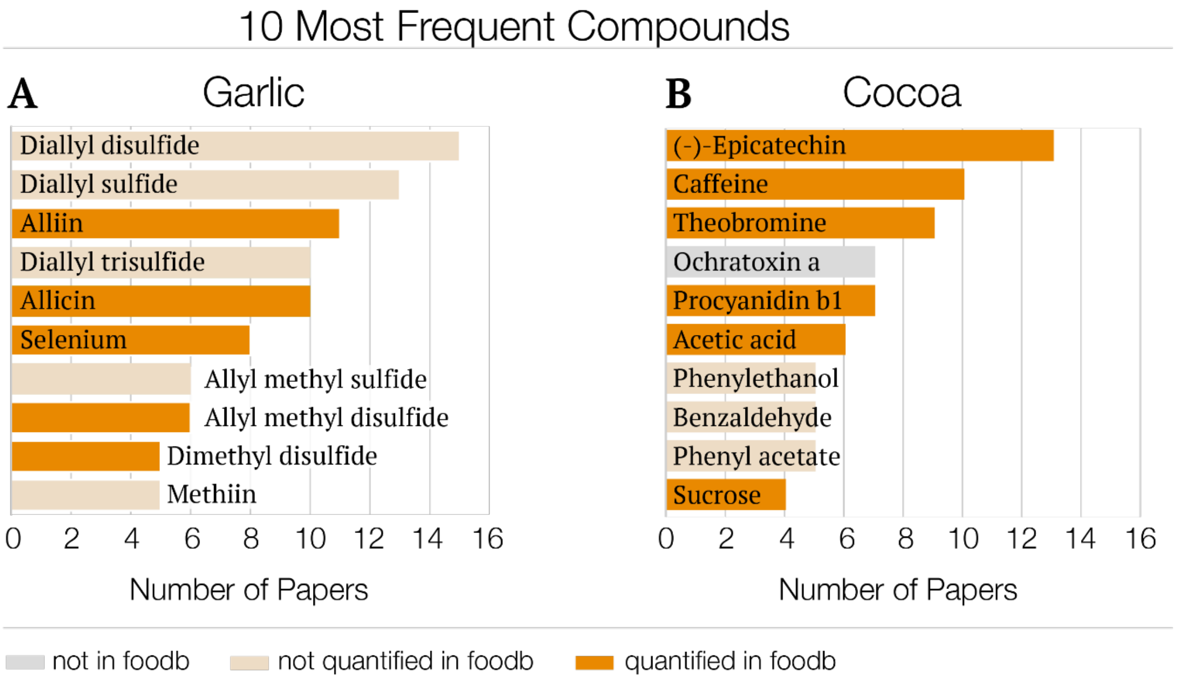


Figure 3: Most Frequently Occurring Compounds in FoodMine. *The graphs show the top 10 most frequently occurring compounds in terms of number of recovered papers for (A) garlic and (B) cocoa, gauging the research interest in each product. The y-axis displays the compound name, and the x-axis shows the number of papers that contain records for the given compound.*

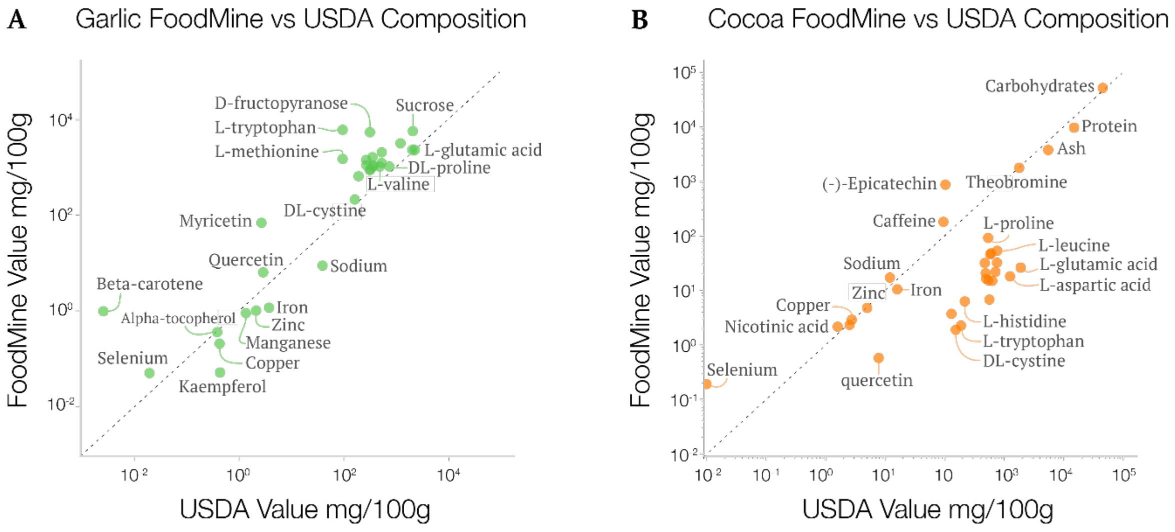


Figure 4: Measurement Comparisons Between FoodMine and USDA. *The nutrient concentrations reported by USDA (x-axis), plotted against the content values of matching compounds in FoodMine (y-axis). The dotted line represents the diagonal. We excluded three and two compounds for (A) garlic and (B) cocoa, respectively, because USDA reported zero values for those compounds.*

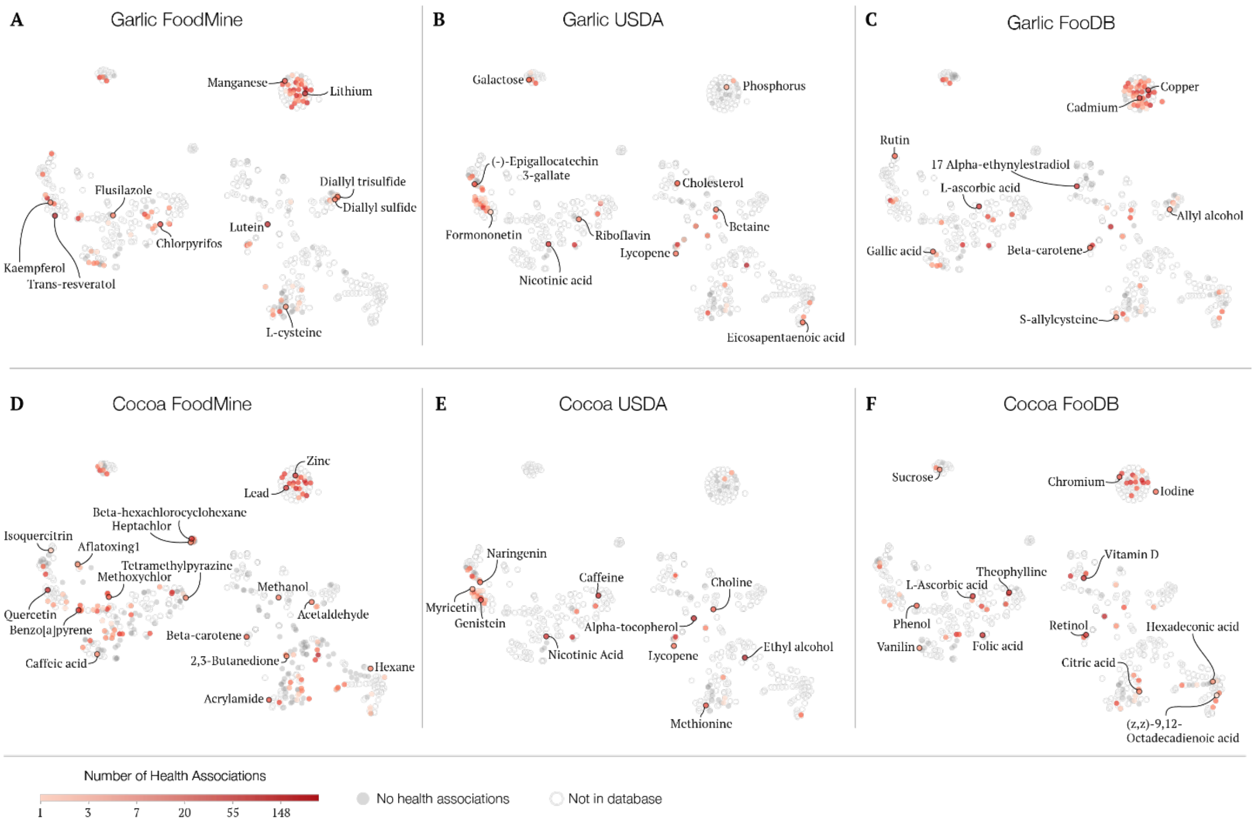
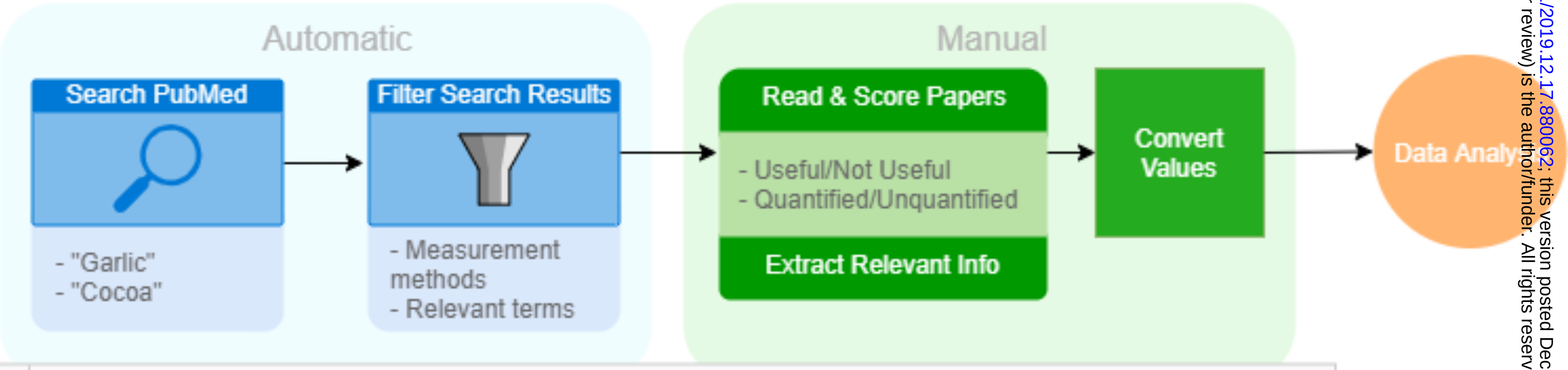
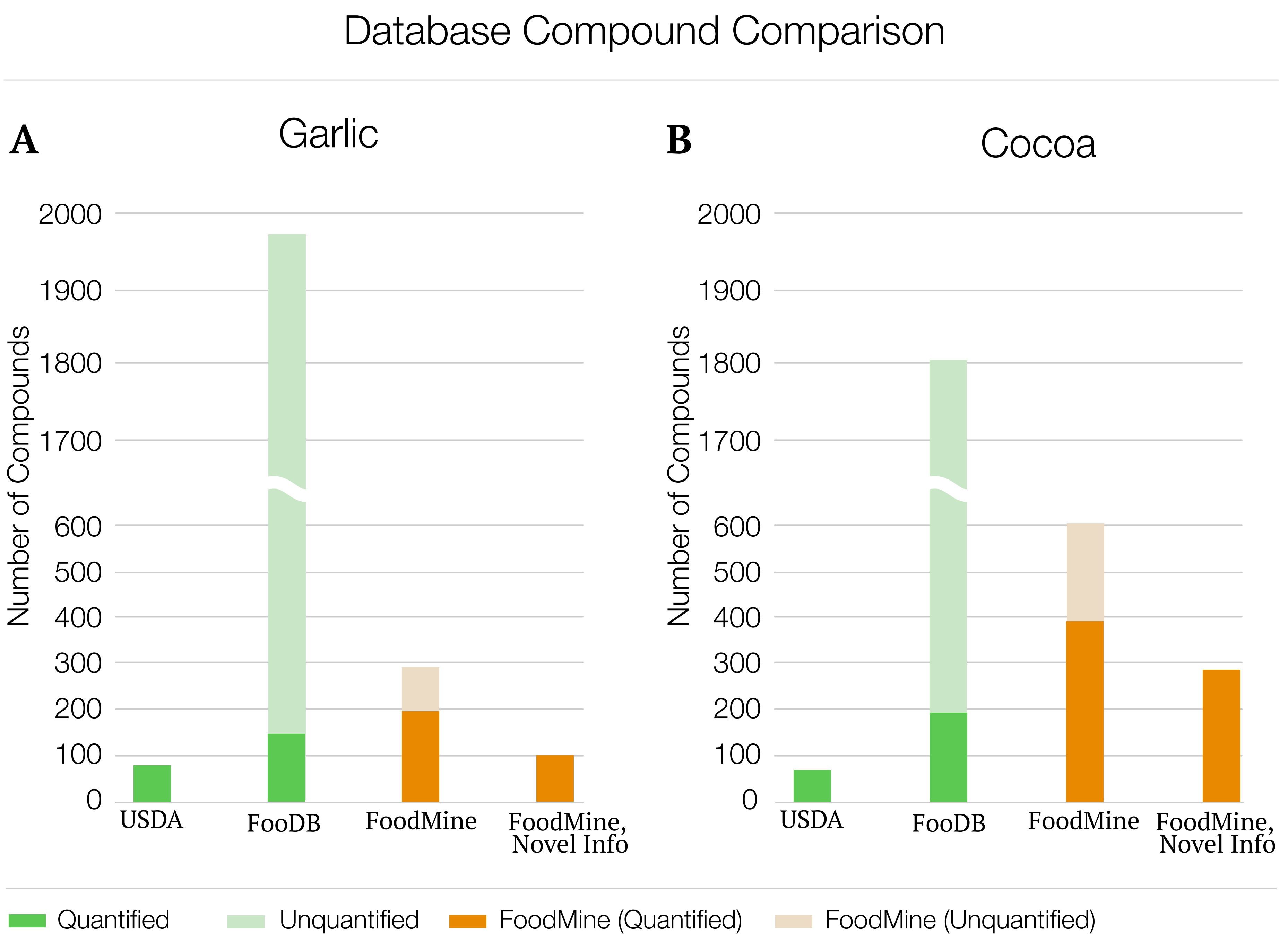


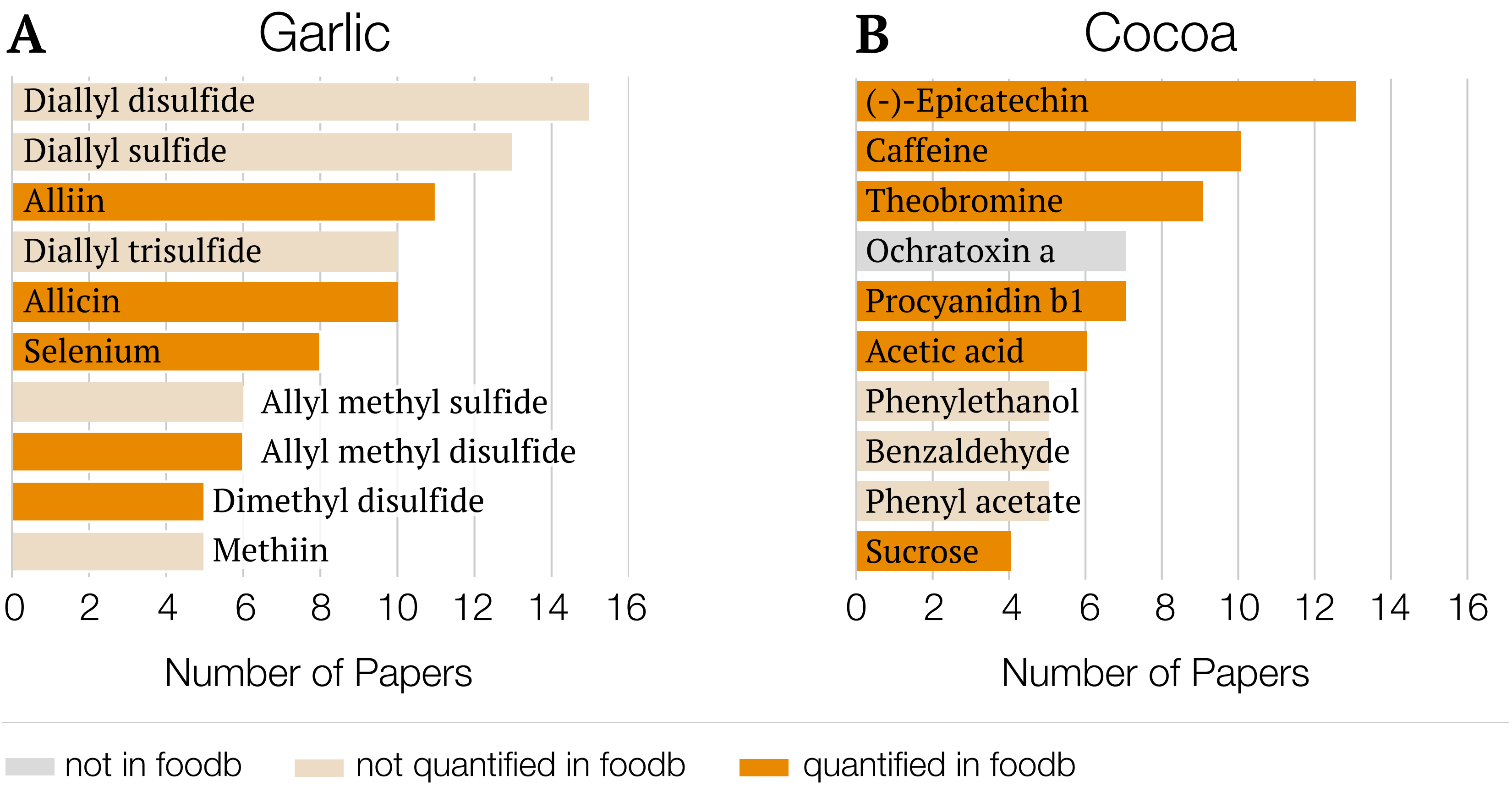
Figure 5: TSNE of Chemical Embeddings with Health Associations. *TSNE plots of Mol2Vec chemical embeddings for garlic (A, B, and C) and cocoa (D, E, and F). The colors of each data point encode the number of health implications associated with those compounds based on the CTD database. Dark gray represents chemicals with 0 health associations. We show chemicals catalogued by each studied database for FoodMine (A & D), USDA (B & E), and FooDB (C & F). Points are filled if the database contains the chemical, and empty if it does not.*

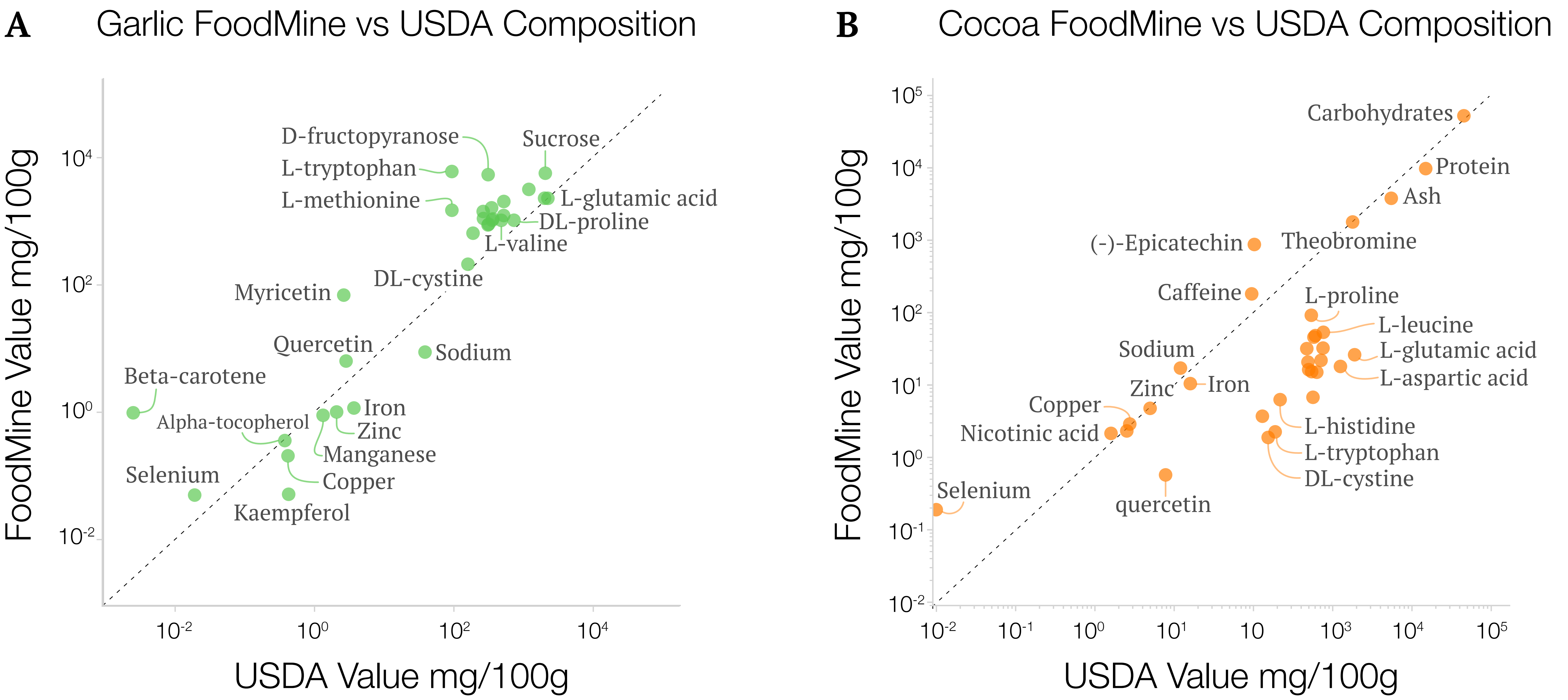


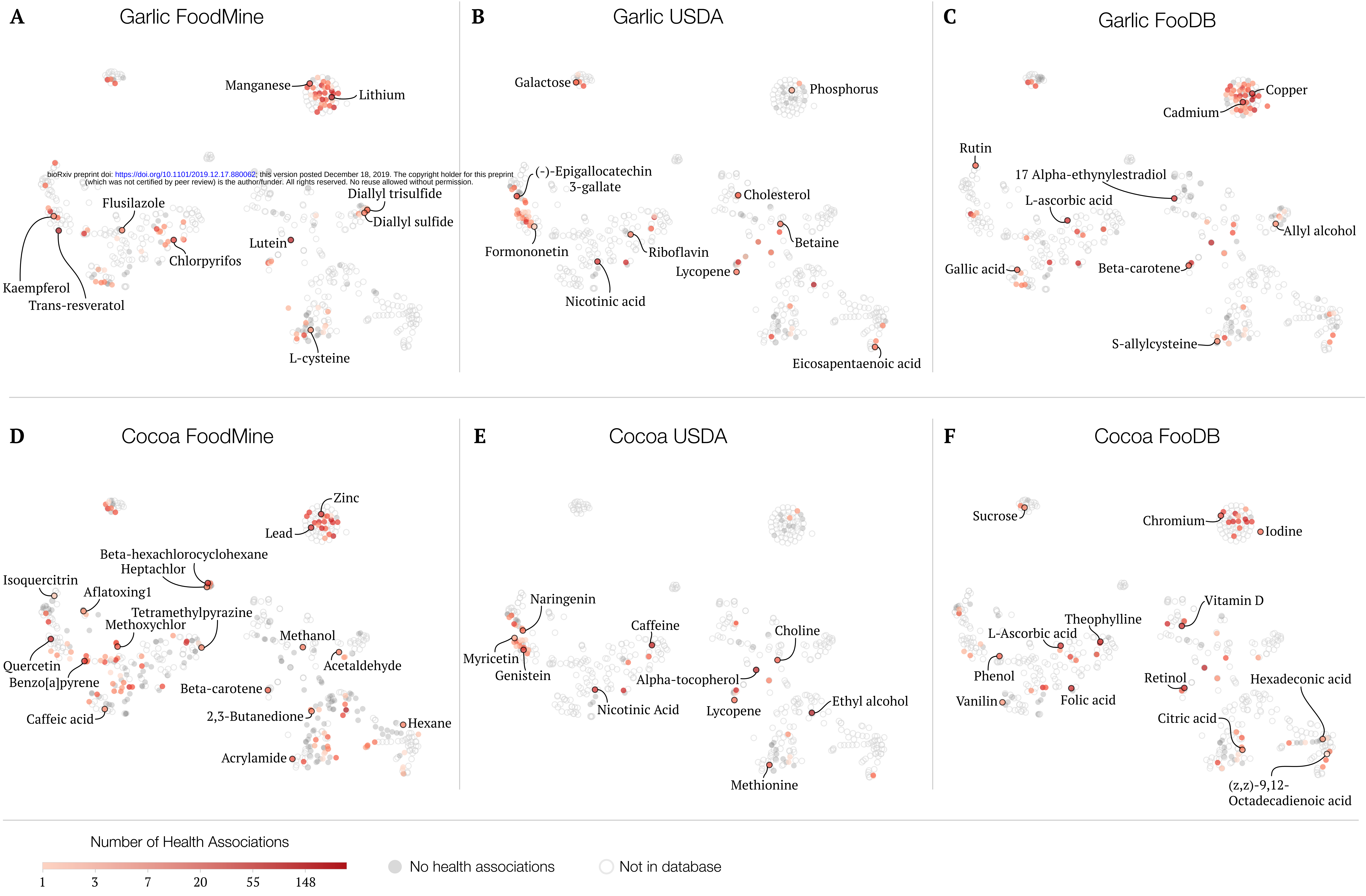
Output	*Garlic*	5,676 results	299 papers	77 papers with info	1,426 data points
	Cocoa	7,620 results	324 papers	93 papers with info	5,855 data points


Database Compound Comparison
A
Garlic
B
Cocoa
Number of Compounds
2000
1900
1800
1700
600
500
400
300
200
100
0
USDA
FooDB
FoodMine
FoodMine, Novel Info
Quantified
Unquantified
FoodMine (Quantified)
FoodMine (Unquantified)

10 Most Frequent Compounds




A
Garlic FoodMine vs USDA Composition
FoodMine Value mg/100g
10^{4}
10^{2}
10^{0}
10^{-2}
USDA Value mg/100g
10^{-2}
10^{0}
10^{2}
10^{4}
D-fructopyranose
Sucrose
L-tryptophan
L-methionine
L-glutamic acid
DL-proline
L-valine
DL-cystine
Myricetin
Quercetin
Sodium
Beta-carotene
Iron
Alpha-tocopherol
Zinc
Manganese
Copper
Selenium
Kaempferol
B
Cocoa FoodMine vs USDA Composition
FoodMine Value mg/100g
10^{5}
10^{4}
10^{3}
10^{2}
10^{1}
10^{0}
10^{-1}
10^{-2}
USDA Value mg/100g
10^{-2}
10^{-1}
10^{0}
10^{1}
10
10^{3}
10^{4}
10^{5}
Carbohydrates
Protein
Ash
Theobromine
(-)-Epicatechin
Caffeine
L-proline
L-leucine
L-glutamic acid
L-aspartic acid
Sodium
Zinc
Iron
Copper
Nicotinic acid
L-histidine
L-tryptophan
DL-cystine
Selenium
quercetin


A
Garlic FoodMine
Manganese
Lithium
Flusilazole
Diallyl trisulfide
Diallyl sulfide
Lutein
Chlorpyrifos
Kaempferol
Trans-resveratol
L-cysteine
B
Garlic USDA
Galactose
Phosphorus
(-)-Epigallocatechin 3-gallate
Cholesterol
Formononetin
Riboflavin
Betaine
Lycopene
Nicotinic acid
Eicosapentaenoic acid
C
Garlic FooDB
Copper
Cadmium
Rutin
17 Alpha-ethynylestradiol
L-ascorbic acid
Allyl alcohol
Gallic acid
Beta-carotene
S-allylcysteine
D
Cocoa FoodMine
Zinc
Lead
Beta-hexachlorocyclohexane
Heptachlor
Isoquercitrin
Aflatoxing1
Tetramethylpyrazine
Methoxychlor
Methanol
Acetaldehyde
Quercetin
Benzo[a]pyrene
Beta-carotene
2,3-Butanedione
Hexane
Caffeic acid
Acrylamide
E
Cocoa USDA
Naringenin
Caffeine
Choline
Myricetin
Genistein
Alpha-tocopherol
Nicotinic Acid
Lycopene
Ethyl alcohol
Methionine
F
Cocoa FooDB
Sucrose
Chromium
Iodine
Vitamin D
Theophylline
L-Ascorbic acid
Phenol
Retinol
Hexadeconic acid
Vanilin
Folic acid
Citric acid
(z,z)-9,12-Octadecadienoic acid
Number of Health Associations
1
3
7
20
55
148
No health associations
Not in database