Corrected: Publisher Correction

The unmapped chemical complexity of our diet

Albert-László Barabási [1,2,3]*, Giulia Menichetti [1] and Joseph Loscalzo[2]

Our understanding of how diet affects health is limited to 150 key nutritional components that are tracked and catalogued by the United States Department of Agriculture and other national databases. Although this knowledge has been transformative for health sciences, helping unveil the role of calories, sugar, fat, vitamins and other nutritional factors in the emergence of common diseases, these nutritional components represent only a small fraction of the more than 26,000 distinct, definable biochemicals present in our food—many of which have documented effects on health but remain unquantified in any systematic fashion across different individual foods. Using new advances such as machine learning, a high-resolution library of these biochemicals could enable the systematic study of the full biochemical spectrum of our diets, opening new avenues for understanding the composition of what we eat, and how it affects health and disease.

The maxim of Jean Anthelme Brillat-Savarin, "*Dites-moi ce que vous mangez et je vais vous dire ce que vous êtes*"—'you are what you eat'—remains as pertinent today, in the era of modern medicine, as it did in 1826. Indeed, the exceptional role of diet in health is well documented by decades of research in nutritional epidemiology, unveiling the role of nutrients and other dietary factors in cardiovascular disease, obesity, type 2 diabetes mellitus (T2DM) and other common diseases[1]. Yet, the bulk of our current understanding of the way food affects health is anchored in the 150 nutritional components that the United States Department of Agriculture (USDA) and other national databases track[2,3], and these nutritional components represent only a subset of the total pool of definable biochemicals in the food supply (see Supplementary Discussion 1).

The dark matter of nutrition

Consider garlic, a key ingredient of the Mediterranean diet: the USDA quantifies 67 nutritional components in raw garlic, indicating that this bulbous plant is particularly rich in manganese, vitamin B$_6$ and selenium[4]. However, a clove of garlic contains more than 2,306 distinct chemical components[5,6]—from allicin, an organosulfur compound responsible for the distinct aroma of the freshly crushed herb, to luteolin, a flavone with reported protective effects in cardiovascular disease[7]—which are listed in FooDB, a database representing the most comprehensive effort to integrate food composition data from specialized databases and experimental data. As of August 2019, FooDB records the presence of 26,625 distinct biochemicals in food[8,9], a number that is expected to increase in the near future (see Supplementary Discussion 2). This exceptional chemical diversity could be viewed as the 'dark matter' of nutrition, as most of these chemicals remain largely invisible to both epidemiological studies, as well as to the public at large.

Where does this remarkable chemical diversity come from? Living organisms require a large number of biochemicals to grow and survive in their limited environments, well beyond the nutritional components that we humans need in our diet. From an evolutionary perspective, plants are characterized by a particularly rich chemical composition, mainly because they are unable to outrun their predators; their defence is occasionally mechanical (for example, through the development of spikes) but is predominantly chemical, exercised through smell, taste and appearance. These chemical defences require an extensive secondary metabolism that produces a wide range of flavonoids, terpenoids and alkaloids. Polyphenols—a highly studied group of chemicals believed to be responsible for the health effects of tea and other plants—are the product of that secondary metabolism. The number of secondary metabolites is estimated to exceed 49,000 compounds, indicating that the 26,000 chemicals currently assigned to food represent an incomplete assessment of the true complexity of the ingredients we consume[10]. Multiple environmental factors, from light to soil moisture, fertility and salinity, can influence the biosynthesis and accumulation of such secondary metabolites[11]. Humans and other animals who can hunt for the necessary food sources do not have the ability to synthesize many molecules our metabolism requires, like ascorbic acid or alpha-linolenic acid, necessitating a source for these essential nutrients.

Overall, an analysis of USDA and FooDB data confirms that plants as a group have the highest chemical diversity, with approximately 2,000 chemicals detected in most examples. Yet, 85% of these chemicals remain unquantified, meaning that while their presence has been detected or inferred, their concentration in specific food ingredients remains unknown (see Supplementary Discussion 2). With garlic, for example, FooDB reports the chemical concentration for just 146 chemical components; the remaining 2,160 chemicals listed in FooDB are not quantified[5,6]. We, therefore, raised the question as to whether the scientific literature contains valuable information on food composition beyond that currently compiled by food databases. Indeed, experimental and analytical projects focused on specific foods and foodborne chemicals are published on a daily basis, and only a small fraction of them inform databases. To unveil this potentially hidden knowledge, we developed a pilot project, FoodMine, that uses natural language processing to mine the full scientific literature for the purpose of comprehensively expanding all available scientific data on the biochemical composition of foods[12].

FoodMine identified 5,676 papers from PubMed that potentially report on chemicals pertaining to the detailed chemical composition of garlic. After filtering this list using machine learning, we manually evaluated 299 papers, of which 77 reported 1,426 individual chemical measurements pertaining to garlic's chemical

[1]Network Science Institute and Department of Physics, Northeastern University, Boston, MA, USA. [2]Channing Division of Network Medicine and Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [3]Department of Network and Data Science, Central European University, Budapest, Hungary. *e-mail: barabasi@gmail.com
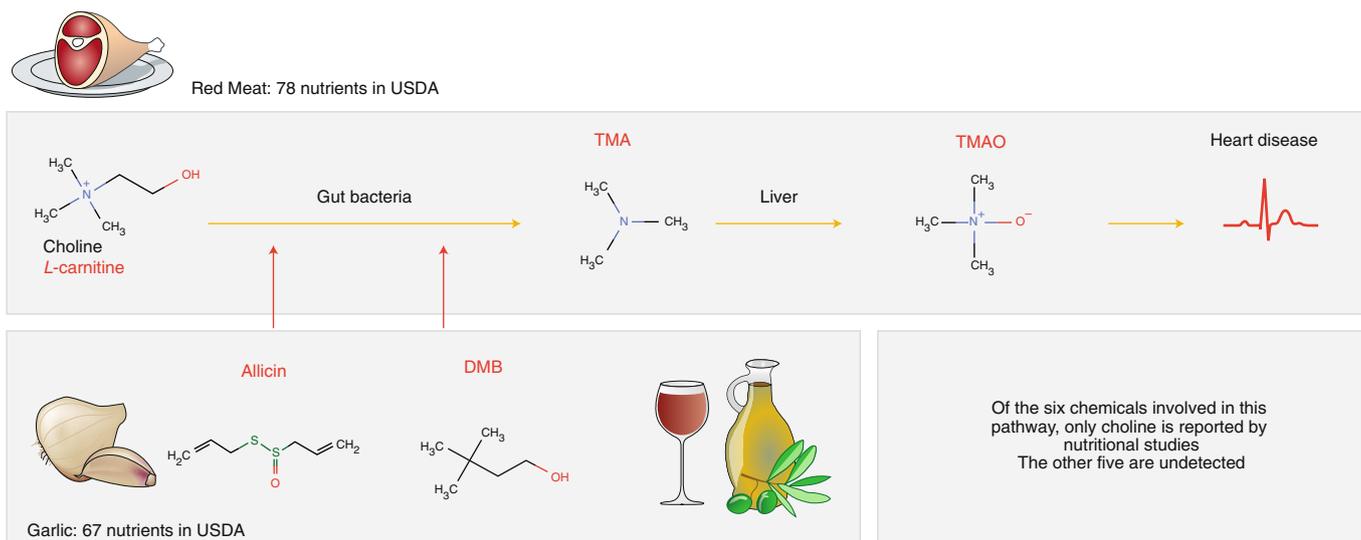
Fig. 1 | Untracked biochemicals and their health implications. Animal products contain *L*-carnitine, choline and choline-contributing compounds[21]. These molecules are metabolized by gut bacteria into trimethylamine (TMA), which is converted in the liver to trimethylamine-*N*-oxide[17] (TMAO), a compound linked to coronary events[16]. Garlic, extra-virgin olive oil and red wine, staple ingredients of the Mediterranean diet, reduce the production of TMAO through allicin and 3,3-dimethylbutan-1-ol (DMB), compounds that block TMA production by gut bacteria. Of the six biochemical compounds involved in this pathway, only one, choline, is tracked in food by the USDA. The other compounds are part of the nutritional 'dark matter' (in red).

composition. Our pilot project recovered more unique quantified compounds than are catalogued by the USDA and FooDB together (see Supplementary Discussion 3 and Supplementary Table 1). For example, diallyl disulfide is known to contribute to garlic's smell and taste, and is implicated in the reported health benefits of garlic, as well as in garlic allergy[13,14]. Although FoodMine found multiple publications reporting on its concentration in garlic, the current databases do not offer quantified information for the compound. Furthermore, FoodMine identified information for 170 compounds that were not previously linked to garlic, either in the USDA or FooDB database (see Supplementary Discussion 3).

Taken together, we find that there is a wealth of exceptionally detailed information about food composition scattered across multiple literature sources. The current incompleteness in coverage within existing food composition databases is not due to a lack of interest in these chemicals or lack of efforts to map these chemical building blocks of food. Rather, it reflects the absence of systematic in-depth efforts to identify and catalogue the data scattered across multiple scientific communities and literature sources. As we discuss below, high-throughput tools required to scan the scientific literature and to overcome these limitations have emerged in the past several years. Mobilizing them could set the stage for an in-depth and systematic understanding of the ways by which our food affects health.

Health implications

The focus on a relatively limited group of nutritional components, including salt, sugar, protein and fat has been justified, given the important role each of them plays in health and disease. Yet, many documented health effects may be linked to untracked chemicals. Consider, for example, trimethylamine *N*-oxide (TMAO)[15]. Recent studies have found that patients with stable coronary heart disease had a fourfold greater risk of dying from any cause over the subsequent five years if they had high blood levels of TMAO[16].

While TMAO and its precursor trimethylamine (TMA) naturally occur in fish and milk, important sources of TMAO in the Western diet are *L*-carnitine and choline, both of which are found in red meat[17,18]. These molecules are metabolized by gut bacteria into TMA, which is then converted in the liver to TMAO[18] (Fig. 1).

The Mediterranean diet[19], which regularly pairs red meat with fresh garlic, derives some of its known health benefits from allicin[20], which blocks TMA production in the gut, ultimately lowering the TMAO concentration in plasma. Taken together, there are at least six distinct biochemicals in our diet involved in the TMAO pathway: *L*-carnitine, choline, TMA, TMAO, allicin and 3,3-dimethylbutan-1-ol (DMB). Yet, only one of them, choline, is tracked and quantified in nutritional databases[21]. The remaining five, despite the key roles they play in health, are effectively nutritional dark matter (Fig. 1).

Overall, 37 nutritional components of garlic can be linked to diseases according to the Comparative Toxicogenomics Database (CTD)[22]. Indeed, garlic carries vitamins B$_1$, B$_6$ and C, and the minerals manganese, copper, selenium and calcium—nutrients whose deficiency or excess have been linked to disease such as T2DM, Parkinson's disease and cardiomyopathies. These links confirm the important role the currently tracked nutrients play in health[23] (see Supplementary Discussion 4). At the same time, the CTD reveals that 485 of the currently unquantified chemicals in garlic can also be linked to multiple therapeutic effects, like the protective action of allicin in cardiovascular disease[24,25] discussed above.

There is a remarkable parallel between pre-genome biology and our current understanding of the health implications of diet. Indeed, in the 1980s, detractors of the Human Genome Project insisted that only the coding regions, representing 1.4% of all base pairs in our DNA, are worth the cost of decoding, labelling the remaining 98.6% 'junk DNA'. Yet, today it is estimated that 66% of disease-carrying variants are, in fact, in these non-coding regions. Similarly, today the 150 nutritional components tracked in food composition tables represent about 0.5% of the 26,625 chemical compounds documented in food. The health implications of these nutritional components are well studied. Yet, more than 99% of the biochemicals present in food, many of which play a role in health and disease, are untracked by national databases, with the health implications of this largely unexplored nutritional dark matter remaining largely unknown. The absence of information on these untracked biochemicals could be responsible for inconsistencies in, and the irreproducibility of, published results as well as for missing health effects, and can also create spurious associations that are not replicable by meta-analysis[26].
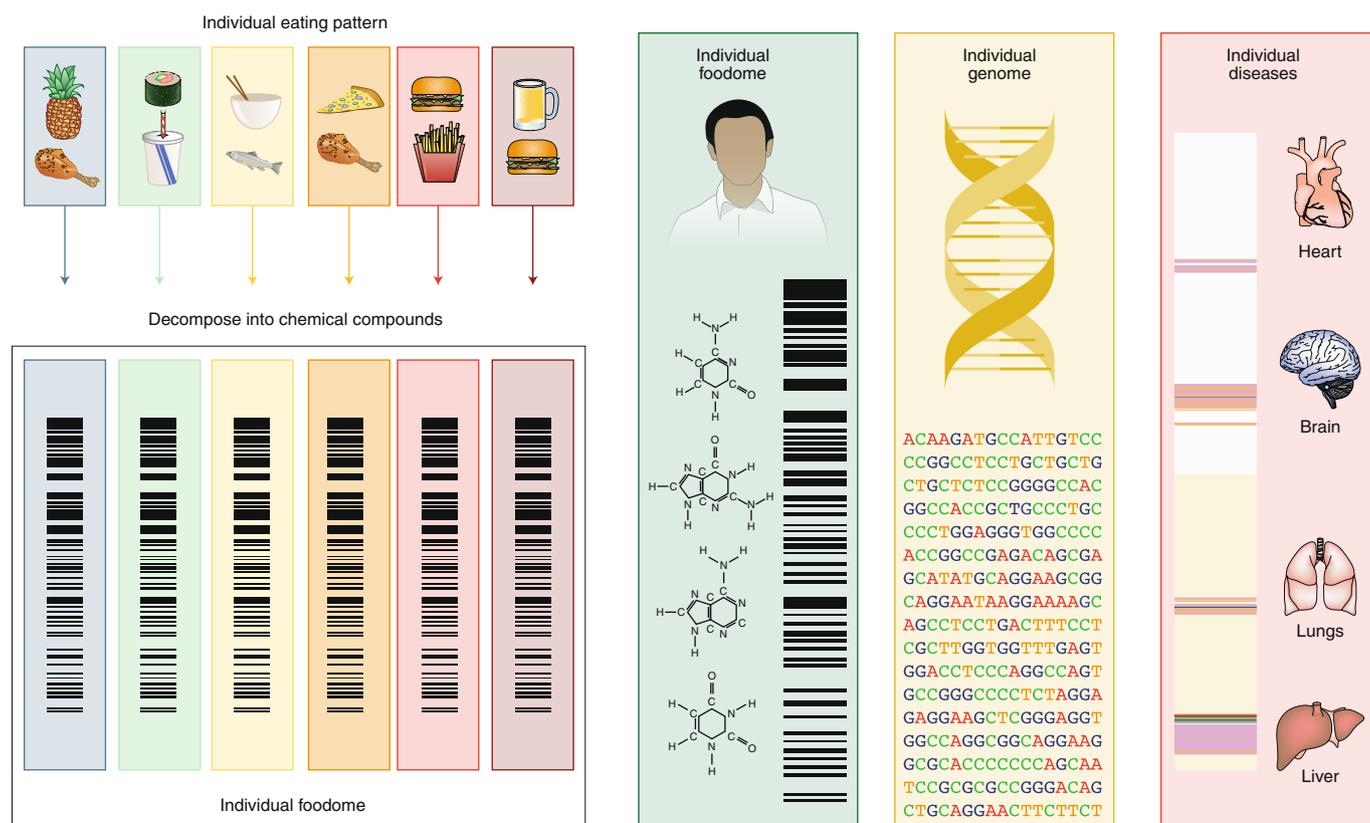
Fig. 2 | Linking the diet to the genome and disease. Our daily eating patterns define a unique biochemical barcode, representing a high-resolution description of each person's individual biochemical exposure through his or her diet, or individual foodome. To assess the individual foodome in a reliable fashion, we can take advantage of the smartphone revolution and collect daily food diaries[59] via image capture. Combined with genomics and disease histories, access to this full biochemical palette could help us expand the widely used genome-wide-association-study-based tools to account for the biochemical composition of our eating patterns, and systematically unveil the linkages between specific food biochemicals, genome variations and health.

Advances in network medicine[27–31]—a post-genome discipline that emphasizes the role of comprehensive molecular interactions (comprising a molecular interaction network, or interactome) in the prevention and treatment of disease—could help us to systematically unveil the mechanistic role of the wide array of molecules found in our diet. Consider, for example, the polyphenol (–)-epigallocatechin 3-O-gallate (EGCG), an abundant biochemical compound in green tea, with potential therapeutic effects in T2DM. Network-based metrics reveal a proximity between 52 human proteins targets of EGCG[32] and 83 proteins associated with T2DM[30,33]. This offers multiple mechanistic pathways by which to account for the relationship between green tea consumption and its many reported effects on health and disease risk[34–36], and its glucose-lowering effects observed using in vitro and in vivo models[37,38]. Unveiling the nutritional dark matter could open up new strategies for discovering the wide array of molecular mechanisms through which food affects health, helping us understand how to use food as therapy, and to aid the identification of food biochemicals with direct therapeutic impact.

Food affects our health through multiple molecular mechanisms: some chemicals serve as a direct source of intermediates for human metabolism, while others, such as polyphenols, play a regulatory role. Yet many food molecules also feed the microbiome in our gut, which metabolizes these compounds into other species that can be further transformed by mammalian metabolism (such as TMA and TMAO)[39,40]. Tracking the full chemical composition of the specific ingredients is also unavoidable if we wish to gain a better understanding of the many ways by which the microbiome responds to the vast diversity of our diet, and how best to alter the microbiome for therapeutic purposes.

Mapping out the foodome

The current incomplete chemical profiling of food poses a number of fundamental scientific and methodological challenges, limiting our ability to systematically explore the health implications of our diet. Yet, the multiple 'known unknowns' of nutrition offer a potential roadmap by which to address them. Indeed, a systematic mapping of the complete chemical composition of the food we consume, although costly, is feasible, and could be greatly accelerated by recent advances in the use of big data and artificial intelligence.

For example, the remarkable governmental and community efforts behind databases like the USDA[3,41,42], FooDB[43], Frida[2], PhenolExplorer[44] and eBasis[45] have already resulted in a wealth of information on food composition. Rapid advances in metabolic reconstructions and biochemical modelling enable us to infer specific pathways from the genome, and through machine learning we can combine metabolic pathway information with the existing food composition databases in a systematic fashion, potentially elucidating the missing chemicals. Indeed, the closer two ingredients are on the phylogenetic tree[46], the more similar is their expected metabolic pathway structure and biochemical composition. Machine learning is ideally suited to combine the known chemical composition of chosen food ingredients over different taxonomical branches with the list of orthologous enzymes in sequenced organisms; the missing chemical information can then be elucidated by learning the appropriate distance metric[47,48] between organisms and

clustering correlated groups of pathways and biochemicals[49]. Such efforts, taking full advantage of existing knowledge, could offer experimentally verifiable predictions about the missing chemicals and their concentration.

Yet, part of the challenge is experimental: the time-consuming, low-throughput, structural chemical tools (spectroscopic methods, nuclear magnetic resonance, mass spectrometry and so on) may need to be fundamentally reengineered into high-throughput methods that can scan food with sufficient chemical resolution and sensitivity, helping to catalogue the presence and the concentration of the vast array of currently unquantified chemical compounds in the food supply. Such efforts are complemented by 'foodomics', a movement aiming to bring omics-technology to the systematic exploration of food[50,51].

Cooking and food processing alter the chemical composition of food, adding chemicals that are absent in raw ingredients and transforming others, from emulsifiers to new lipids. Some of these changes have well documented health implications, like the presence of acrylamide (a carcinogenic compound) in fried and baked goods and in coffee. While the impact of food processing on basic nutritional components is well studied, little is known about the impact of processing on the thousands of chemicals found in the nutritional dark matter. Equally important, we must account for the numerous toxins added to food during cooking, preservation and packaging, or accumulated in food according to the environmental production conditions, and their effect on health, such as the well documented toxicity of highly reactive aldehydes or of persistent organic pollutants[52].

For nutrition to compete with genetics in accuracy, reach and impact, we must organize the information on eating patterns to fit the big-data platform that fuels advances in this digital age of biomedicine. Indeed, our eating patterns are digital—each of us consumes a weighted subset of chemicals found in the food supply. The precise subset of chemicals to which each individual is exposed defines that person's individual nutritional-chemical 'barcode', or his or her 'foodome' (Fig. 2). The determinants of this personal foodome are complex, from food supply to personal choices, and are modulated by geography, culture and socio-economic status. Efforts to ensure the traceability[53] of food, allowing us to track the source and production of the raw material introduced into the food chain, together with the environmental and processing conditions modulating the individual foodome, will also greatly enhance future research in this arena. Our ability to track the nutritional–chemical barcode of each individual, and correlate it with individual genetic variations and health history, could help merge nutrition with a precise digital and statistical platform similar to that which fuelled the spectacular advances in genomics[54,55]. Such a platform could help us scan systematically for novel causal mutation–chemical–health associations that are largely invisible to current hypothesis-driven research in nutrition.

To appreciate the transformative potential of a deeper quantitative understanding of the nutritional dark matter, we must realize that our genetic predispositions to specific phenotypes and patho-phenotypes can conceivably be modified by these food-based molecules. Indeed, while we cannot currently change the genetic basis for disease, we regularly modulate the activity of our subcellular networks through the food we eat, diminishing the impact of some mutations and enhancing the role of others. This differential modulation of subcellular networks explains why individuals with strong genetic predispositions to heart disease can lower the chance of developing the disease by up to 70% with proper lifestyle choices[56], within which dietary changes play a dominant role[56,57]. This finding implies that an accurate mapping of our full chemical exposure through our diet could lead to actionable information to improve health. Recent trends in nutrition research, aiming to explore the synergies, competitions and interactions among the entire matrix of

what constitutes a food product, increasingly acknowledge the complexity of the problem, and the need for new tools to address it[58]. We must embrace this irreducible complexity to be able to integrate changes in the food supply, the role of the microbiome and personalized dietary patterns, so that we can eventually offer individually tailored food-based therapies and appropriate lifestyle choices for disease prevention and lifespan optimization.

References

1. Willett, W. *Nutritional Epidemiology* (Oxford Univ. Press, 1990).
2. National Food Institute, Technical University of Denmark. *Frida Food data* version 2 (DTU, 2016); frida.fooddata.dk
3. *Composition of Foods Raw, Processed, Prepared.* National Nutrient Database for Standard Reference, Release 28. Documentation and User Guide (USDA, 2015).
4. USDA. FoodData Central. *Garlic, raw* https://fdc.nal.usda.gov/fdc-app.html#/food-details/169230/nutrients (2019).
5. FooDB. *Garlic* http://foodb.ca/foods/FOOD00008 (data dump 06/29/2017, ID=8, 2017).
6. FooDB. *Soft-necked garlic* http://foodb.ca/foods/FOOD00850 (data dump 06/29/2017, ID=880, 2017).
7. Luo, Y., Shang, P. & Li, D. Luteolin: a flavonoid that has multiple cardio-protective effects and its molecular mechanisms. *Front. Pharmacol.* **8**, 1–10 (2017).
8. Dagnino, S. & Macherone, A. (eds) *Unraveling the Exposome: A Practical View* (Springer, 2019).
9. FooDB. *Compounds* http://foodb.ca/compounds (accessed 1 August 2019).
10. Wink, M. (ed) in *Annual Plant Reviews Volume 40: Biochemistry of Plant Secondary Metabolism* 2nd edn Ch. 1 (Wiley, 2010).
11. Yang, L. et al. Response of plant secondary metabolites to environmental factors. *Molecules* **23**, 1–26 (2018).
12. Hooton, F., Menichetti, G. & Barabási, A.-L. FoodMine: exploring food contents in scientific literature. Preprint at https://doi.org/10.1101/2019.12.17.880062 (2019).
13. Rao, P. 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).
14. Garcia-Abujeta, J. L. et al. Allergic contact dermatitis to diallyl disulphide in Spain. *J. Allergy Clin. Immunol.* **117**, S130 (2006).
15. Cho, C. E. et al. Trimethylamine-N-oxide (TMAO) response to animal source foods varies among healthy young men and is influenced by their gut microbiota composition: a randomized controlled trial. *Mol. Nutr. Food Res.* **61**, 1–12 (2017).
16. Senthong, V. et al. Intestinal microbiota-generated metabolite trimethylamine-N-oxide and 5-year mortality risk in stable coronary artery disease: the contributory role of intestinal microbiota in a COURAGE-like patient cohort. *J. Am. Heart Assoc.* **5**, 1–7 (2016).
17. Velasquez, M. T., Ramezani, A., Manal, A. & Raj, D. S. Trimethylamine N-oxide: the good, the bad and the unknown. *Toxins* **8**, e326 (2016).
18. Wang, Z. et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**, 57–65 (2011).
19. Estruch, R. et al. Primary prevention of cardiovascular disease with a Mediterranean diet supplemented with extra-virgin olive oil or nuts. *N. Engl. J. Med.* **378**, e34 (2018).
20. Lawson, L. D. & Hunsaker, S. M. Allicin bioavailability and bioequivalence from garlic supplements and garlic foods. *Nutrients* **10**, 4–6 (2018).
21. Patterson, K. Y. et al. *USDA Database for the Choline Content of Common Foods, Release 2* (USDA, 2008).
22. Davis, A. P. et al. *Comparative Toxicogenomics Database* (NC State University, accessed: 25 March 2019); ctdbase.org
23. King, B. L., Davis, A. P., Rosenstein, M. C., Wiegers, T. C. & Mattingly, C. J. Ranking transitive chemical-disease inferences using local network topology in the comparative toxicogenomics database. *PLoS ONE* **7**, e46524 (2012).
24. Varshney, R. & Budoff, M. J. Garlic and heart disease. *J. Nutr.* **146**, 416S–421S (2016).
25. Wu, W. K. et al. Dietary allicin reduces transformation of L-carnitine to TMAO through impact on gut microbiota. *J. Funct. Foods* **15**, 408–417 (2015).
26. Micha, R., Wallace, S. K. & Mozaffarian, D. Red and processed meat consumption and risk of incident coronary heart disease, stroke, and diabetes mellitus: a systematic review and meta-analysis. *Circulation* **121**, 2271–2283 (2010).
27. Goh, K.-I. et al. The human disease network. *Proc. Natl Acad. Sci. USA* **104**, 8685–90 (2007).

28. Barabási, A.-L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* **12**, 56–68 (2011).

29. Panagiotou, G. & Nielsen, J. Nutritional systems biology: definitions and approaches. *Annu. Rev. Nutr.* **29**, 329–339 (2009).

30. Menche, J. et al. Uncovering disease-disease relationships through the incomplete interactome. *Science* **347**, 1257601 (2015).

31. Loscalzo, J., Barabási, A.-L. & Silverman, E. K. *Network medicine: complex systems in human disease and therapeutics* (Harvard Univ. Press, 2017).

32. Szklarczyk, D. et al. STITCH 5: Augmenting protein-chemical interaction networks with tissue and affinity data. *Nucleic Acids Res.* **44**, D380–D384 (2016).

33. Faria do Valle I. et al. *Network-based framework for understanding the health benefits of dietary polyphenols* (in press).

34. Greenberg, J. A., Axen, K. V., Schnoll, R. & Boozer, C. N. Coffee, tea and diabetes: the role of weight loss and caffeine. *Int. J. Obes.* **29**, 1121–1129 (2005).

35. Iso, H. et al. The relationship between green tea and total caffeine intake and risk for self-reported type 2 diabetes among Japanese adults. *Ann. Intern. Med.* **144**, 554–562 (2006).

36. Song, Y., Manson, J. E., Buring, J. E., Sesso, H. D. & Liu, S. Associations of dietary flavonoids with risk of type 2 diabetes, and markers of insulin resistance and systemic inflammation in women: a prospective study and cross-sectional analysis. *J. Am. Coll. Nutr.* **24**, 376–84 (2005).

37. Wolfram, S. et al. Epigallocatechin gallate supplementation alleviates diabetes in rodents. *J. Nutr.* **136**, 2512–2518 (2006).

38. Keske, M. A. et al. Vascular and metabolic actions of the green tea polyphenol epigallocatechin gallate. *Curr. Med. Chem.* **22**, 59–69 (2015).

39. Dodd, D. et al. A gut bacterial pathway metabolizes aromatic amino acids into nine circulating metabolites. *Nature* **551**, 648–652 (2017).

40. Schmidt, T. S. B., Raes, J. & Bork, P. The human gut microbiome: from association to modulation. *Cell* **172**, 1198–1215 (2018).

41. Sebastian, R. et al. *Flavonoid Values for USDA Survey Foods and Beverages 2007–2010* (Food Surveys Research Group, 2016); www.ars.usda.gov/Services/docs.htm?docid=25102

42. Duke, J. A. *Database of Biologically Active Phytochemicals and Their Activity* (CRC Press, 1992).

43. *FooDB* (2017); http://foodb.ca/

44. Neveu, V. et al. Phenol-Explorer: an online comprehensive database on polyphenol contents in foods. *Database* **2010**, bap024 (2010).

45. 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**, 1–15 (2017).

46. NCBI. *Taxonomy* (US National Library of Medicine, 2018); https://www.ncbi.nlm.nih.gov/taxonomy

47. Xing, E. P., Ng, A. Y., Jordan, M. I. & Russell, S. Distance metric learning, with application to clustering with side-information. *in Proc. 15th International Conference on Neural Information Processing Systems* 521–528 (MIT Press, 2002).

48. Davis, J. V., Kulis, B., Jain, P., Sra, S. & Dhillon, I. S. Information-theoretic metric learning. *in Proc. 24th International Conference on Machine Learning* (ed. Ghahramani, Z.) 209–216 (ACM, 2007).

49. Bao, J., Wang, W., Yang, T. & Wu, G. An incremental clustering method based on the boundary profile. *PLoS ONE* **13**, e0196108 (2018).

50. García-Cañas, V., Simó, C., Herrero, M., Ibáñez, E. & Cifuentes, A. Present and future challenges in food analysis: Foodomics. *Anal. Chem.* **84**, 10150–10159 (2012).

51. Capozzi, F. & Bordoni, A. Foodomics: a new comprehensive approach to food and nutrition. *Genes Nutr.* **8**, 1–4 (2013).

52. Jones, K. C. & de Voogt, P. Persistent organic pollutants (POPs): state of the science. *Environ. Pollut.* **100**, 209–221 (1999).

53. Espiñeira, M. & Santaclara, F. J. (eds) in *Advances in Food Traceability Techniques and Technologies: Improving Quality Throughout the Food Chain*. 3–8 (Elsevier, 2016).

54. Patel, C. J. & Ioannidis, J. P. A. Studying the elusive environment in large scale. *JAMA* **311**, 2173–2174 (2014).

55. Milanlouei, S., et al. *A systematic comprehensive longitudinal evaluation of dietary factors associated with coronary heart disease* (in press).

56. Khera, A. V. et al. Genetic risk, adherence to a healthy lifestyle, and coronary disease. *N. Engl. J. Med.* **375**, 2349–2358 (2016).

57. Schroeder, S. et al. We can do better-improving the health of the American people. *N. Engl. J. Med.* **357**, 1221–1228 (2007).

58. Mozaffarian, D., Rosenberg, I. & Uauy, R. History of modern nutrition science—implications for current research, dietary guidelines, and food policy. *BMJ* **361**, k2392 (2018).

59. Bennett, D. A., Landry, D., Little, J. & Minelli, C. Systematic review of statistical approaches to quantify, or correct for, measurement error in a continuous exposure in nutritional epidemiology. *BMC Med. Res. Methodol.* **17**, 1–22 (2017).

Author contributions

A.L.B., G.M. and J.L. conceived the project and wrote the manuscript. G.M. performed the statistical analysis.

Competing interests

A.L.B. is founder of Nomix and Foodome, and J.L. and A.L.B. are founders of Scipher Medicine, companies that explore the role of networks and food in health.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s43016-019-0005-1.

Correspondence should be addressed to A.-L.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.