The Foodome Project
Tackling the Complexity of Food Systems

Giulia Menichetti

Channing Division of Network Medicine

Brigham and Women's Hospital, Harvard Medical School

Center for Complex Networks Research

Network Science Institute, Northeastern University

Rockefeller Foundation-Acumen Academy Food Systems Fellow

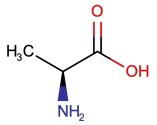
Harvard Data Science Initiative

www.giuliamenichetti.com

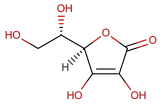
The dark matter of nutrition: a universe in expansion

Known	**Unknown**	
USDA SR	Foodome DB	
188	139,443	92,612 detected 46,831 inferred
nutrients	compounds	
69	6,802	1,984 detected 4,818 inferred
Nutrients	Compounds	

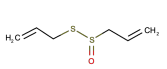
Alanine



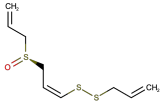
Ascorbic acid



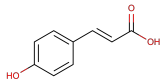
Allicin



Ajoene



p-Coumaric acid



Decoding the Foodome: Molecular Networks Connecting Diet and Health

Giulia Menichetti,[1,2,3] Albert-László Barabási,[1,2,4] and Joseph Loscalzo[1]

Annu. Rev. Nutr. 2024. 44:18.1–18.32

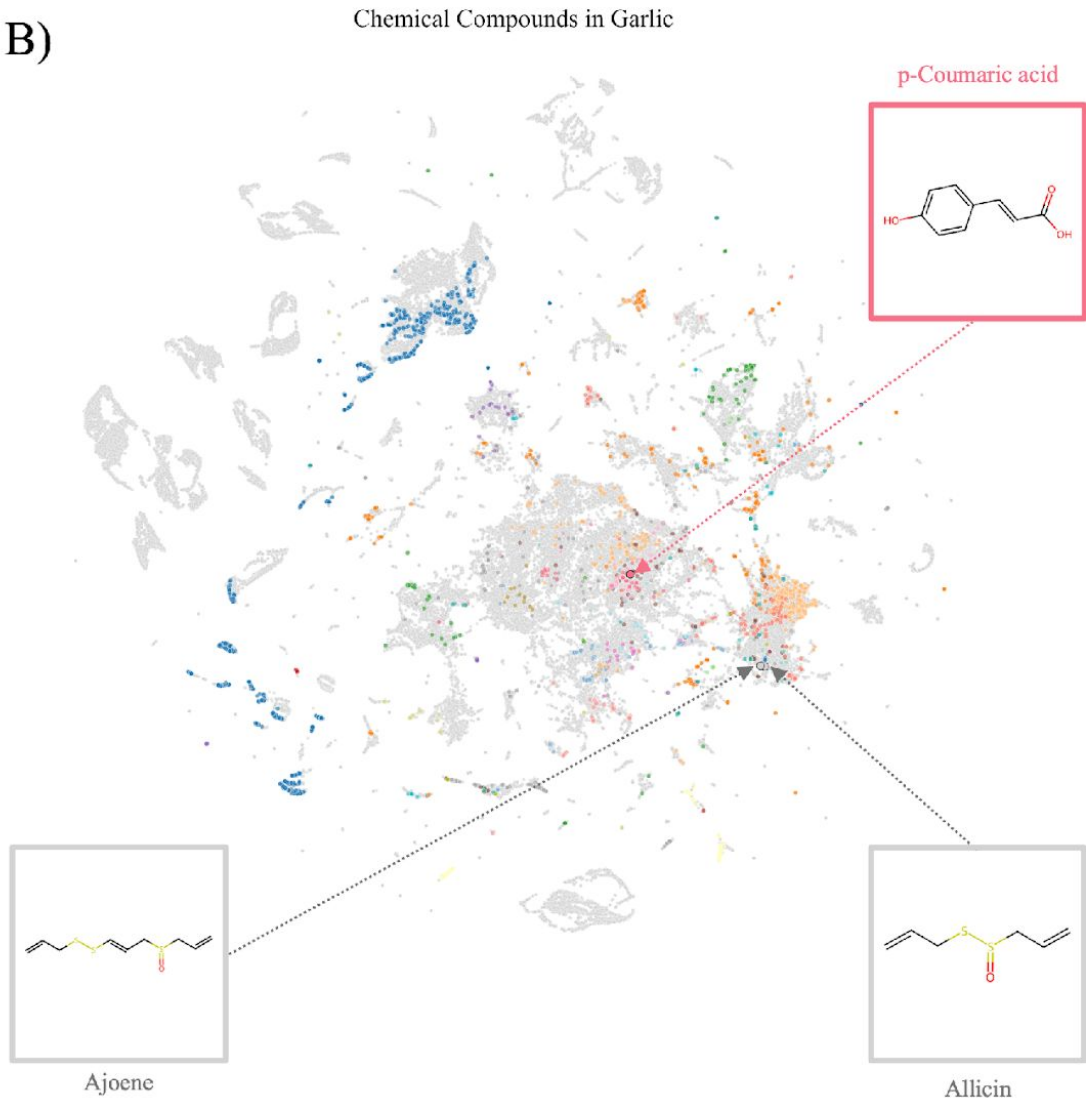
The *Annual Review of Nutrition* is online at nutr.annualreviews.org

Keywords

network science, machine learning, artificial intelligence, complexity, network medicine, systems pharmacology, nutrition

A)

B)



Chemical Subclass

- Other (41240)
- Triradylcglycerols (39198)
- Carbohydrates and carbohydrate conjugates (9792)
- Amino acids, peptides, and analogues (6192)
- Flavonoid glycosides (5811)
- Terpene glycosides (3661)
- Diradylglycerols (2076)
- Triterpenoids (1973)
- Purine ribonucleotides (1626)
- Carbonyl compounds (1456)
- Sesquiterpenoids (1219)
- Benzoic acids and derivatives (1150)
- Fatty alcohols (1129)
- Steroidal glycosides (1028)
- Fatty acyl glycosides (1019)
- Monoterpenoids (995)
- Fatty acid esters (975)
- Glycerophosphocholines (940)
- Fatty acids and conjugates (890)
- Diterpenoids (829)
- Terpene lactones (815)
- Hydroxycinnamic acids and derivatives (744)
- O-methylated flavonoids (735)
- 1-benzopyrans (650)
- Alcohols and polyols (576)
- Ethers (537)

6,802 compounds
1,984 detected
4,818 inferred

"The Chemical Complexity of Food and Implications for Therapeutics"
Menichetti G., Barabási A.-L., Loscalzo. J. – New England Journal of Medicine, 2025 (In press)

Biases of the current AI models

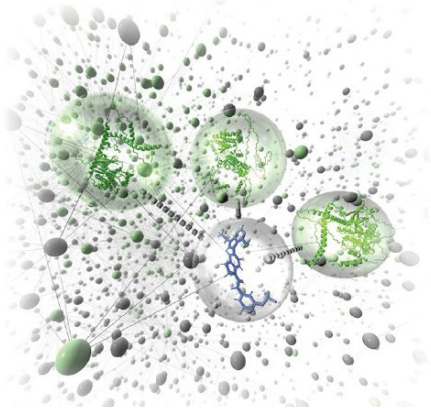
Nature Reviews Physics | https://doi.org/10.1038/s42254-022-00471-1 | Published online: 19 May 2022

TOOLS OF THE TRADE

An AI pipeline to investigate the binding properties of poorly annotated molecules

Binding between proteins and small molecules is an important link-prediction task affecting biological processes such as metabolic reactions, regulatory mechanisms and signal transduction pathways. However, binding annotations are available only for a small subset of the possible combinations of proteins and molecules. For instance, in the Foodome project, we have data for more than 135,000 small molecules found in foods, but only 4.58% of the collected molecules have known binding annotations with proteins. Machine learning (ML) models aim to fill the gap by predicting binding, but — given datasets, which presents an extremely biased picture of binding: molecular structures tend to either have an excessive ratio of positive annotations (binding) to negative (non-binding), or vice versa. In scenarios affected by this annotation imbalance, ML models behave like configuration models, disregarding structural information, and performing poorly in inductive tests.

Inspired by ML models in drug discovery, we created AI-Bind, a deep learning pipeline designed to provide reliable binding predictions for poorly annotated proteins and ligands. In recent years, ML models



nature communications

Article

https://doi.org/10.1038/s41467-023-37572-z

Improving the generalizability of protein-ligand binding predictions with AI-Bind

Received: 16 May 2022

Accepted: 23 March 2023

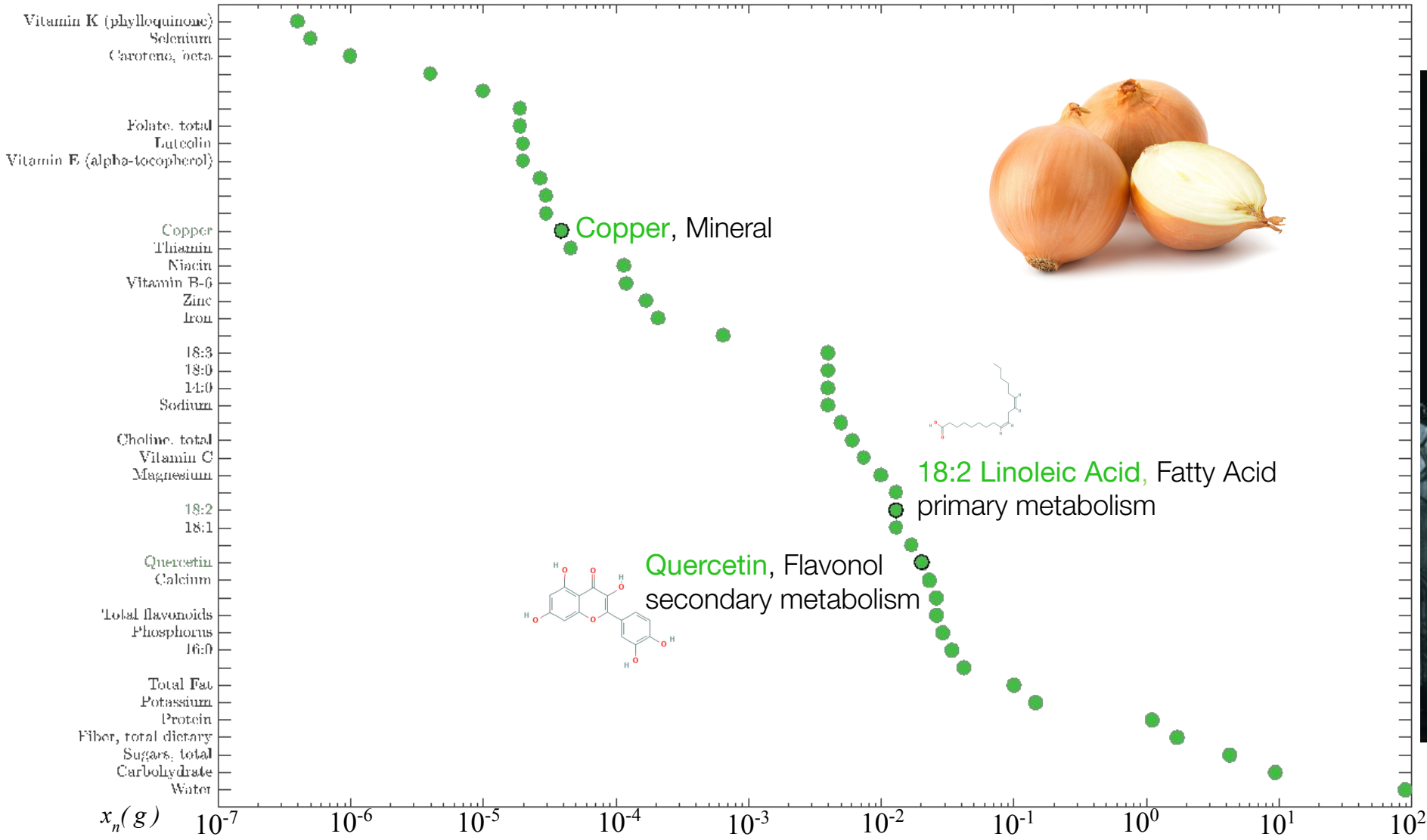
Published online: 08 April 2023

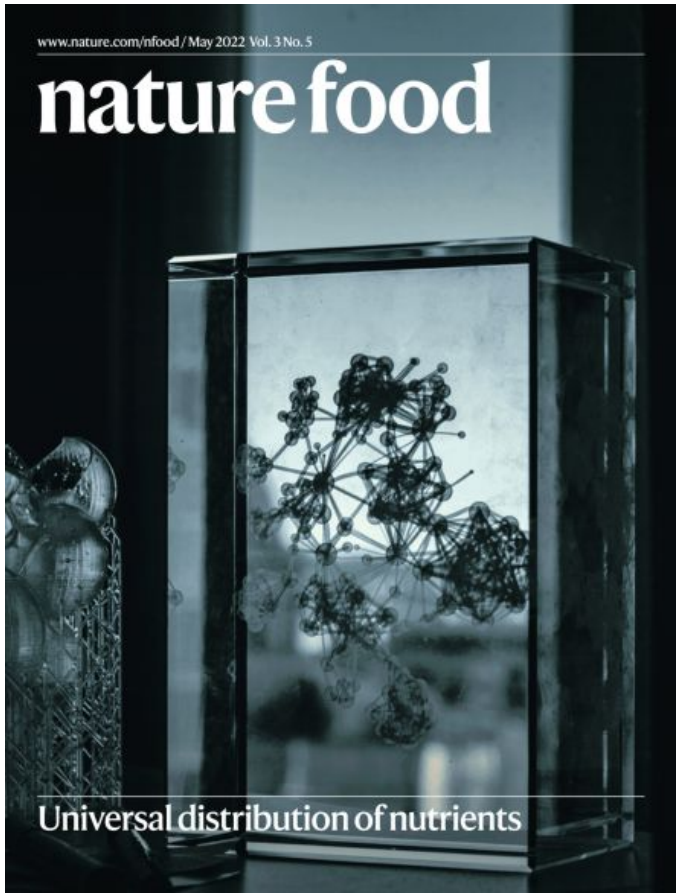
Check for updates

Ayan Chatterjee[1], Robin Walters[2], Zohair Shafi[2], Omair Shafi Ahmed[2], Michael Sebek[1,3], Deisy Gysi[1,3,4], Rose Yu[5], Tina Eliassi-Rad[1,2,6,7], Albert-László Barabási[1,3,8] & Giulia Menichetti[1,3,9]

Identifying novel drug-target interactions is a critical and rate-limiting step in drug discovery. While deep learning models have been proposed to accelerate the identification process, here we show that state-of-the-art models fail to generalize to novel (i.e., never-before-seen) structures. We unveil the mechanisms responsible for this shortcoming, demonstrating how models rely on shortcuts that leverage the topology of the protein-ligand bipartite network, rather than learning the node features. Here we introduce AI-Bind, a pipeline that combines network-based sampling strategies with unsupervised pre-training to improve binding predictions for novel proteins and ligands. We validate AI-Bind predictions via docking simulations and comparison with recent experimental evidence, and step up the process of interpreting machine learning prediction of protein-ligand binding by identifying potential active binding sites on the amino acid sequence. AI-Bind is a high-throughput approach to identify drug-target combinations with the potential of becoming a powerful tool in drug discovery.

How Much Nutrient is in a Food?





"Nutrient concentrations in food display universal behaviour"
Menichetti G., Barabási A.-L. - Nature Food, 2022

Raw Onion

Onion Rings

Nutrients

z-score

z-scor

A

Mini No Sugar Added Cheesecake, 3 oz

3

Cakes | WholeFoods
FPro: 0.7195

Nutrients in 100 grams (all units are grams)

- 305.9 Calories
- 2.4 Sugar
- 0 Fiber
- 24.7 Total Fat
- 0.3 Sodium
- 5.9 Protein
- 20 Carbs
- 0.1 Cholesterol
- 15.3 Sat. Fat
- 0.1 Calcium
- 0 Iron
- -- Vit. A
- -- Vit. C

Edwards Frozen Original Whipped Cheesecake - 24oz

61

Cakes | Target
FPro: 0.9533

Nutrients in 100 grams (all units are grams)

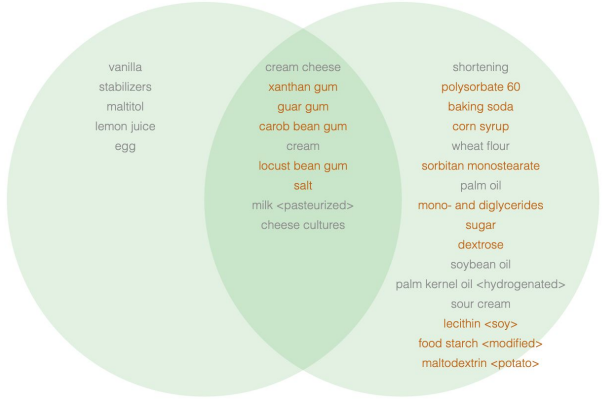
- 341.2 Calories
- 27.1 Sugar
- 0 Fiber
- 18.8 Total Fat
- 0.3 Sodium
- 2.4 Protein
- 40 Carbs
- 0 Cholesterol
- 10.6 Sat. Fat
- 0.1 Calcium
- 0 Iron
- -- Vit. A
- -- Vit. C

B

Mini No Sugar Added Cheesecake, 3 oz

3

Cakes | WholeFoods
FPro: 0.7195

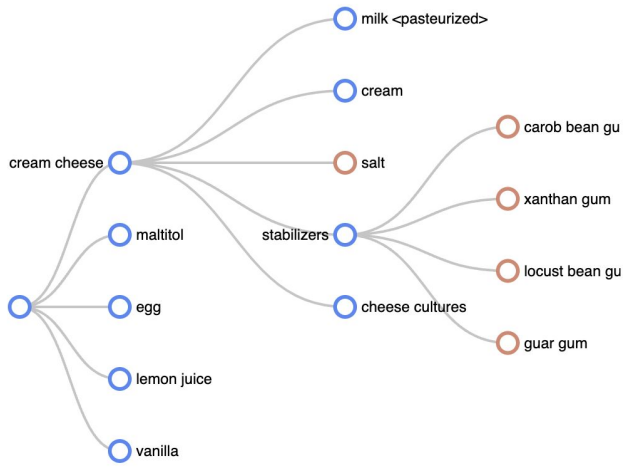


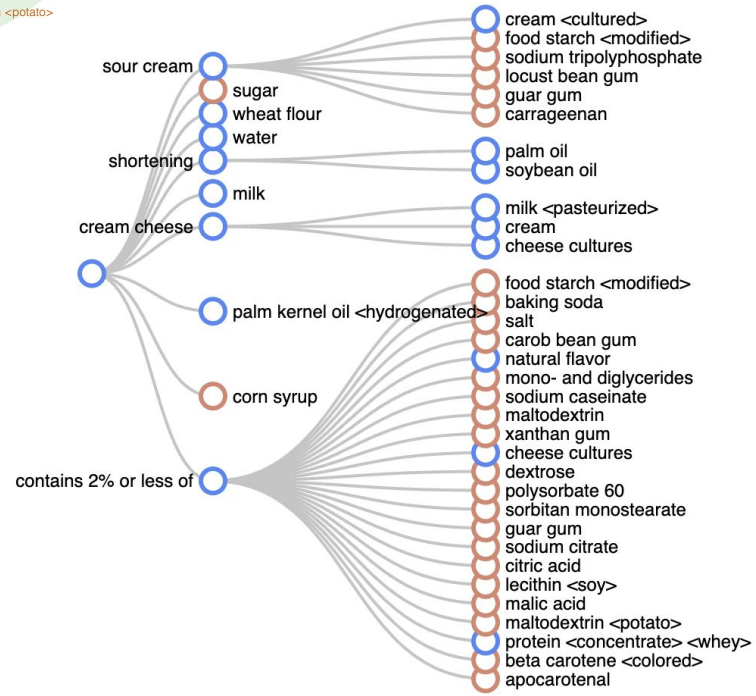
Edwards Frozen Original Whipped Cheesecake - 24oz

61

Cakes | Target
FPro: 0.9533

C





D

Switch Recommendations

3

Mini No Sugar Added Cheesecake, 3 oz
Cakes | WholeFoods

6

Raspberry Cheesecake 2 Inch, 1 oz
Cakes | WholeFoods

8

Philadelphia Salted Caramel Cheesecake Snacks, 2 ct Pack, 3.25 oz Cups
Cakes | Walmart

12

New York Style Cheesecake with Cinnamon Brown Sugar and Honey Crust, Plain, Individual Slice
Cakes | Walmart

13

Freshness Guaranteed Strawberry Swirl Cheesecake, 16 oz
Cakes | Walmart

Key Publications

•Prediction of chemical annotations and concentrations in food
- The unmapped chemical complexity of our diet. (Nature Food, 2019)
- Nutrient concentrations in food display universal behavior. (Nature Food, 2022, Journal Cover)
- The use of genomics-based annotations towards the study of the molecular composition of edible plants. (Under Review in Nature Food, 2023)
- Estimating Nutrient Concentrations in Food Using Untargeted Metabolomics. (Under Review in Nature Food, 2024)

•Text mining of the literature to identify the chemical composition of food ingredients
- Exploring food contents in scientific literature with FoodMine. (Scientific Reports, 2020)

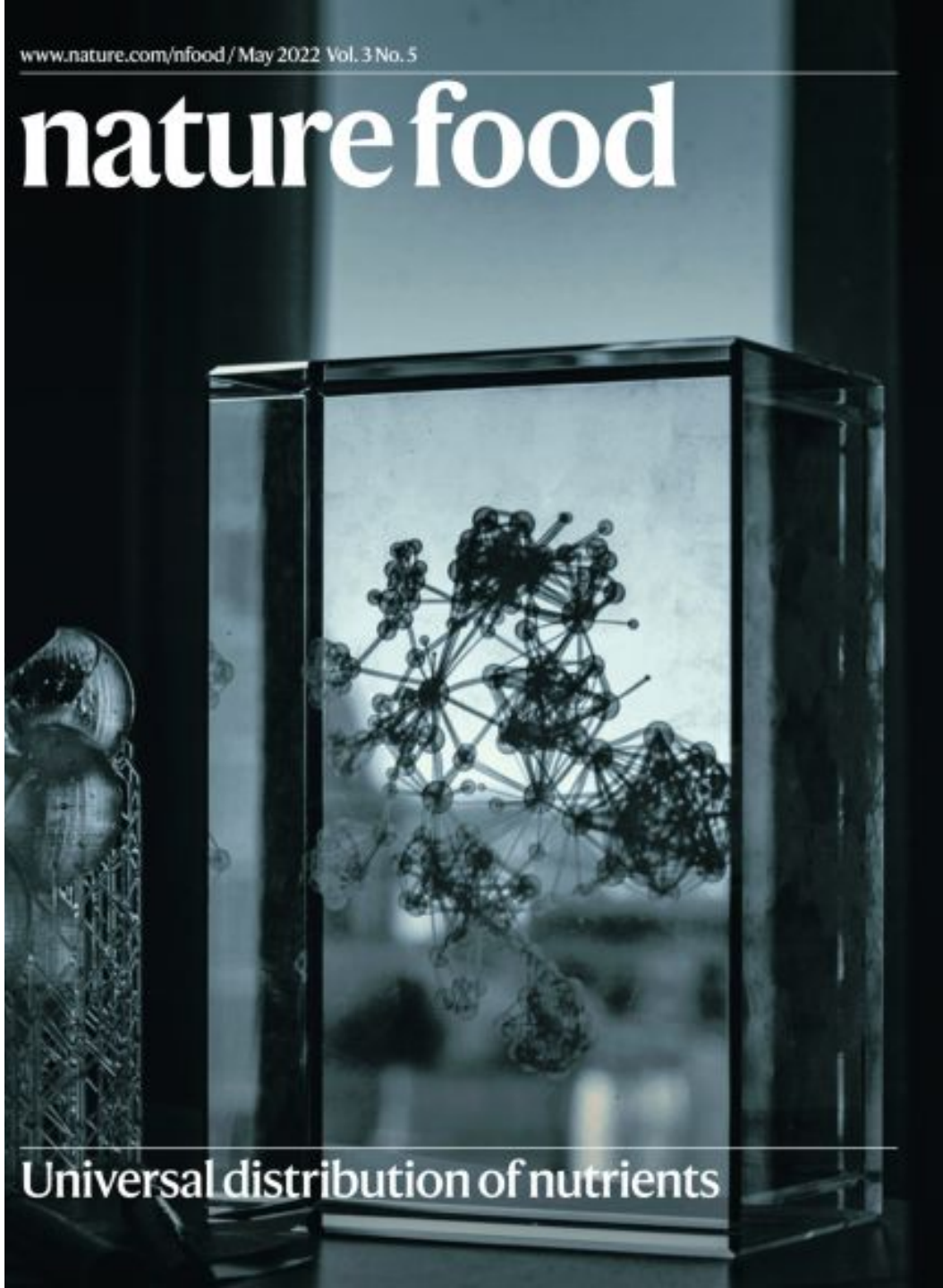
•Embeddings of chemical compounds for structural characterization
- Identification of potent inhibitors of SARS-CoV-2 infection by combined pharmacological evaluation and cellular network prioritization. (iScience, 2022)

•Protein binding prediction for naturally occurring compounds and drugs with network sampling strategies
- Improving the generalizability of protein-ligand binding predictions with AI-Bind (Nature Communications, 2023)
- An AI pipeline to investigate the binding properties of poorly annotated molecules. (Nature Reviews Physics, 2022)
- CPIExtract: A software package to collect and harmonize small molecule and protein interactions. (Under Review in Bioinformatics, 2024)

•Network Medicine and Food
- Molecular interaction networks and cardiovascular disease risk: the role of food bioactive small molecules. (ATVB, 2023)
- NetMedPy: A Python package for Large-Scale Network Medicine Screening. (Under Review in Bioinformatics, 2024)

•Environment-Wide Association Study (EWAS) for foods and nutrients
- A systematic comprehensive longitudinal evaluation of dietary factors associated with acute myocardial infarction and fatal coronary heart disease. (Nature Communications, 2020)

•Quantification of food processing, food supply quality, and related health consequences
- Machine learning prediction of the degree of food processing. (Nature Communications, 2023)
- Prevalence of Processed Food in Major US Grocery Stores. (Nature Food, 2025)



2025 Team

Postdoctoral Fellows

Gordana Ispirova
Computer Scientist

Hao Xu
Computational Chemist

Bonnie Patchen
Nutri-Genomics Researcher

Eleanor Kate Phillips
MD in Critical Care and Pulmonology

Michael Sebek
Chemist

Andres Aldana Gonzalez
Computer Scientist

Nan Lin
Epidemiologist

Salvo Danilo Lombardo
Medical Informatician/MD

Thank you!

giulia.menichetti@channing.harvard.edu
www.giuliamenichetti.com
GitHub: menicgiulia