

UCSF

UC San Francisco Previously Published Works

Title

Building a chemical blueprint for human blood

Permalink

<https://escholarship.org/uc/item/5fk7f6m3>

Journal

Nature, 588(7836)

ISSN

0028-0836

Authors

Bradley, Patrick H
Pollard, Katherine S

Publication Date

2020-12-03

DOI

10.1038/d41586-020-03122-6

Peer reviewed

not only in the field of vision restoration but also in those looking to understand epigenetic reprogramming of neurons and other cell types generally. For decades, it was argued that understanding normal neural developmental processes would one day lead to the tools to repair the aged or damaged brain. Lu and colleagues' work makes it clear: that era has now arrived.

Andrew D. Huberman is in the Department of Neurobiology and the Department of Ophthalmology, Stanford University School of Medicine, Stanford, California 94305, USA. e-mail: adh1@stanford.edu

1. D.-B., D. *Nature* **125**, 230–231 (1930).
2. Lu, Y. *et al. Nature* **588**, 124–129 (2020).
3. Goldberg, J. L., Klassen, M. P., Hua, Y. & Barres, B. A. *Science* **296**, 1860–1864 (2002).
4. Laha, B., Stafford, B. K. & Huberman, A. D. *Science* **356**, 1031–1034 (2017).
5. Horsburgh, G. M., Lund, R. D. & Hankin, M. H. *J. Comp. Neurol.* **327**, 323–340 (1993).
6. Takahashi, K. & Yamanaka, S. *Cell* **126**, 663–676 (2006).
7. Abad, M. *et al. Nature* **502**, 340–345 (2013).
8. Oberdoerffer, P. *et al. Cell* **135**, 907–918 (2008).
9. López-Otin, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. *Cell* **153**, 1194–1217 (2013).
10. Horvath, S. *Genome Biol.* **14**, 3156 (2013).
11. Wang, M. & Lemos, B. *Genome Res.* **29**, 325–333 (2019).
12. Rasmussen, K. D. & Helin, K. *Genes Dev.* **30**, 733–750 (2016).
13. Laugesen, A., Højfeldt, J. W. & Helin, K. *Mol. Cell* **74**, 8–18 (2019).

Clinical research

Building a chemical blueprint for human blood

Patrick H. Bradley & Katherine S. Pollard

What determines the chemical make-up of human blood? Measurement of the contributions of factors ranging from genetics to lifestyle has now identified diet and gut microbes as key predictors of blood's molecular composition. **See p.135**

Our blood transports many chemicals besides oxygen and carbon dioxide. Some of these molecules provide useful indicators of the state of our health. Indeed, measuring such biomarkers is a common feature of clinical blood tests. Other molecules present, such as hormones and drugs, directly affect health by modulating processes such as metabolism and immune responses. On page 135, Bar *et al.*¹ shed light on the factors that affect the recipe for human blood's chemical brew.

The origin of most blood-borne molecules, and why they vary in concentration between individuals, is unknown. The list of possible regulators is long: for any given molecule, diet, drugs, medical conditions and history, genetic variants and gut microorganisms might all have a role. Furthermore, these factors can interact, as is the case for trimethylamine oxide. This molecule, which promotes the artery-narrowing disease atherosclerosis, is generated as a result of the metabolism, by both microbes and their host, of certain dietary compounds that are abundant in red meat². For molecules such as this, which directly affect health, understanding their metabolic regulation might help to yield new clinical treatments.

Bar *et al.* describe their efforts to tackle the question of what factors govern the molecules present in blood. This work requires not only measurement of the many variables potentially

involved, but also the use of analytical methods that can capture complexity – such as the interactions between variables – while still ensuring that valid predictions can be made for individuals outside the study population.

The authors began by characterizing

blood samples from a group of 491 healthy individuals in great detail. They quantified the molecules in serum – the liquid component of blood that remains after the proteins needed for clotting have been removed. The study participants provided detailed health information, and answered questionnaires about diet and lifestyle. They also gave stool samples, which were used for DNA sequencing, to determine the genetic signatures of the gut microbes present (also known as the microbiome).

As the authors acknowledge, this is a small study group by the standards of genome-wide association studies, which seek to find connections between genes and disease. Bar *et al.* are also not the first to link serum molecules to genetic variation or the microbiome^{3,4}. However, the authors' analysis of this group of individuals is unique in the number of data types that were systematically collected to investigate serum composition.

Next, Bar *et al.* used a machine-learning approach to link factors such as human genetics and microbiome information to the molecules in the blood. By carrying out many analyses omitting different data subsets, the authors found that diet, the microbiome and clinical variables such as prescription-drug use and blood pressure had the most associations with serum molecules. Although the authors found some genetic associations, confirming 46 previously reported gene–metabolite links, they concluded that the association effects for genetic factors were smaller than were those for diet, clinical variables and the microbiome. These various data types are not exactly comparable, but the authors' estimates of the genetic effects are in line with results from

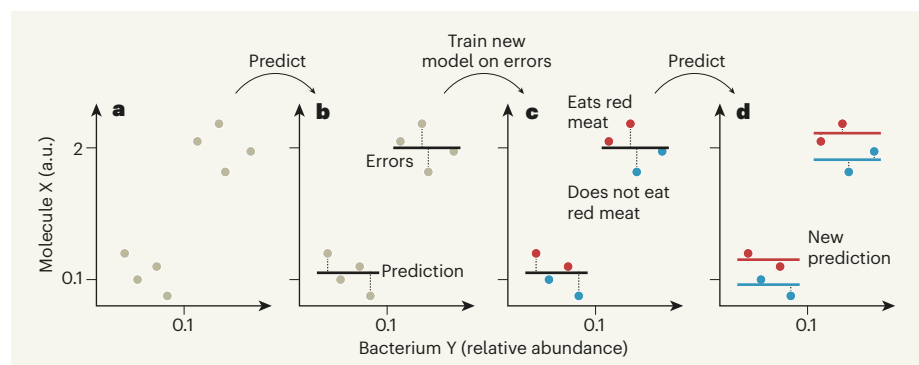


Figure 1 | A way to predict blood's molecular composition. Bar *et al.*¹ obtained human blood samples and identified many of the molecules present. The authors also gathered information about a range of factors, such as diet and gut microbes, that might have affected the molecules found. Using a computational method called gradient-boosted decision trees, Bar and colleagues predicted the molecular composition of an individual's blood. **a**, In this hypothetical example, data points show an individual's concentration of molecule X in arbitrary units (a.u.) and the relative abundance of a type of gut bacterium Y. **b**, The model uses an 'if–then' classification to predict (black horizontal lines) the relationship between bacterial abundance and the concentration of X. The prediction in this case is that if the bacterial abundance is more than 0.1, the concentration of X is 2, and if this abundance is less than 0.1, the concentration of X is 0.1. Dotted lines show the prediction errors. **c**, The model is then refined by taking into account another factor, such as whether the person eats red meat (red) or not (blue). **d**, After another 'if–then' classification that includes this dietary factor, the model generates refined predictions (red and blue horizontal lines) with lower errors that link the predicted concentration of X to dietary and bacterial factors.

previous work, providing support for their conclusion that diet and the microbiome have larger and more pervasive influences on serum composition than do genetic factors.

Diet and the microbiome could predict the data for some molecules with similar levels of accuracy, as would be expected, given that diet can affect microbiome composition. But Bar and colleagues showed that these data types provide non-overlapping information, too. For example, dietary information uniquely predicted particular metabolites associated with the consumption of citrus fruit, whereas the presence of a type of microbe belonging to the Lachnospiraceae family strongly predicted the presence of indoxyl sulfate – a bacterial breakdown product of the amino acid tryptophan, previously linked to diseases of the kidney and vasculature⁵.

To make predictions about the concentrations of molecules present in blood samples, Bar *et al.* used a machine-learning method called gradient-boosted decision trees, which can capture complex interactions. Decision trees learn simple ‘if-then’ rules to make predictions (Fig. 1). This method layers individual decision trees, successively improving them by training new models that focus specifically on reducing the prediction errors of the older ones.

Bar and colleagues interpreted these models using an approach called feature-attribution analysis. This yields specific hypotheses about how individual factors, such as microbes, foods and genetic variants, influence a particular prediction, here, the molecular composition of blood. More-complex models can be prone to ‘overfitting’ – making erroneous predictions that are based on noise or irrelevant details. The authors therefore fitted and evaluated their models conservatively, but, even more importantly, they confirmed many of their predicted microbe-to-metabolite links in two large, independent study groups. Finally, Bar *et al.* tested one set of their predictions in a smaller study, identifying molecules (cytosine and betaine) associated with the consumption of wholewheat bread, and then showing that individuals randomly assigned to eat the bread had the expected changes in these metabolites.

This study is comprehensive, but plenty of room remains for future exploration. The authors used the well-validated and standardized Metabolon platform to measure serum metabolites, but no such metabolomic analysis method can cover the full range of blood-borne compounds. Certain types of molecule, such as blood lipids, might therefore be under-sampled compared with others. This might explain why the authors mostly detected metabolite associations with only one of the two most abundant lineages of gut bacteria^{6,7}. Metabolomics can detect molecules whose identity is unknown

beyond their molecular weight, and, indeed, the authors report several associations with such unknown metabolites. Although these might point to previously unknown aspects of biology (interestingly, for example, one such association was linked to the age of the participant), without metabolite identification, only limited conclusions can be drawn.

The authors’ microbiome data provide DNA information for all the genomes present in stool extracts. However, Bar *et al.* distil these data down to the level of abundances of bacterial species, excluding non-bacteria such as yeasts or protozoan organisms. Limiting analyses to the species level also obscures the fact that strains of the same bacterial species can differ in gene content. For example, the metabolism of the drug digoxin *in vivo* by the bacterium *Eggerthella lenta* requires a gene that is present in only certain strains of *E. lenta*⁸. Finally, the authors were unable to link serum metabolites to specific bacterial enzymes responsible for their generation, which would have helped to connect the associated links to the underlying molecular mechanisms.

These limitations should not detract from the most useful aspect of this paper. By making the full data set available to the research community, Bar and colleagues could help

enable the development of future computational methods, potentially resolving some of these limitations, or even providing ways to answer new questions. Their data are likely to be a rich and valuable resource for scientists interested in the mechanisms by which diet, the microbiome and genetics affect our biochemistry and physiology.

Patrick H. Bradley is in the Department of Microbiology and the Infectious Diseases Institute, Ohio State University, Ohio 43210, USA. **Katherine S. Pollard** is at the Gladstone Institute of Data Science and Biotechnology, University of California, San Francisco, San Francisco, and the Chan Zuckerberg Biohub, San Francisco, California 94158, USA. e-mails: bradley.720@osu.edu; kpollard@gladstone.ucsf.edu

1. Bar, N. *et al.* *Nature* **588**, 135–140 (2020).
2. Koeth, R. A. *et al.* *Nature Med.* **19**, 576–585 (2013).
3. Visconti, A. *et al.* *Nature Commun.* **10**, 4505 (2019).
4. Shin, S.-Y. *et al.* *Nature Genet.* **46**, 543–550 (2014).
5. Hung, S.-C., Kuo, K.-L., Wu, C.-C. & Tarnag, D.-C. *J. Am. Heart Assoc.* **6**, e005022 (2017).
6. Nemati, R. *et al.* *J. Lipid Res.* **58**, 1999–2007 (2017).
7. Farrokhi, V. *et al.* *Clin. Transl. Immunol.* **2**, e8 (2013).
8. Koppel, N., Bisanz, J. E., Pandelia, M.-E., Turnbaugh, P. J. & Balskus, E. P. *eLife* **7**, e33953 (2018).

This article was published online on 11 November 2020.

Precision measurement

Fine-structure constant tests standard model

Holger Müller

A highly precise measurement of a physical constant known as the fine-structure constant provides a stringent test of the standard model of particle physics, and sets strong limits on the existence of speculative particles. **See p.61**

Every physicist knows the approximate value (1/137) of a fundamental constant called the fine-structure constant, α . This constant describes the strength of the electromagnetic force between elementary particles in the standard model of particle physics and is therefore central to the foundations of physics. For example, the binding energy of a hydrogen atom – the energy required to break apart the atom’s electron and proton – is about $\alpha^2/2$ times the energy associated with an electron’s mass. Moreover, the magnetic moment of an electron is subtly larger than that expected for a charged, point-like particle by a factor of roughly $1 + \alpha/(2\pi)$. This ‘anomaly’ of the magnetic moment has been verified to ever-increasing accuracy, becoming ‘the

standard model’s greatest triumph”¹. On page 61, Morel *et al.*² report a measurement of α with an accuracy of 81 parts per trillion (p.p.t.), a 2.5-fold improvement over the previous best determination³.

The measurement of α involves three steps. First, a laser beam makes an atom absorb and emit multiple photons and, in doing so, recoil (Fig. 1a). The mass of the atom is deduced by measuring the kinetic energy of this recoil. Second, the electron’s mass is calculated using the precisely known ratio of the atom’s mass to the mass of an electron^{4,5} (Fig. 1b). Third, α is determined from the electron’s mass and the binding energy of a hydrogen atom, which is known from spectroscopy⁶ (Fig. 1c).

However, the recoil energy is tiny and