(including non-mutant antigens, which can sometimes trigger an anticancer response), neoantigens are emerging as crucial targets that T cells can use to detect and destroy cancer cells, and represent important targets for immunotherapy[6–8]. ■

Siranush Sarkizova *and* **Nir Hacohen** *are at the Center for Cancer Research, Massachusetts General Hospital, Boston, Massachusetts 02114, USA, and at the Broad Institute of MIT and Harvard, Cambridge, Massachusetts.* **S.S.** *is also in the Department of Biomedical Informatics, Harvard Medical School, Boston. e-mail: nhacohen@broadinstitute.org*

1. Balachandran, V. P. *et al. Nature* **551,** 512–516 (2017).
2. Łuksza, M. *et al. Nature* **551,** 517–520 (2017).
3. Robins, H. *Curr. Opin. Immunol.* **25,** 646–652 (2013).
4. Sharma, P. & Allison, J. P. *Science* **348,** 56–61 (2015).
5. Schumacher, T. N. & Hacohen, N. *Curr. Opin. Immunol.* **41,** 98–103 (2016).
6. Carreno, B. M. *et al. Science* **348,** 803–808 (2015).
7. Ott, P. A. *et al. Nature* **547,** 217–221 (2017).
8. Sahin, U. *et al. Nature* **547,** 222–226 (2017).

MICROBIOLOGY

Crowdsourcing Earth's microbes

A large-scale study has been assessing microbial diversity by analysing DNA sequences from samples submitted by scientists around the globe. The initial results are now being used to create an open-access resource. SEE ARTICLE P.457

JEROEN RAES

A simple, though dauntingly ambitious idea of sampling the microbial genetic diversity across Earth is the driving force behind the Earth Microbiome Project. On page 457, Thompson *et al.*[1] report the results of this experimental tour de force.

The project began life at a meeting in Snowbird, Utah, in 2010 at which a group of scientists from a wide range of disciplines discussed the goals, challenges and practicalities of such an enterprise[2]. Seven years on, the authors now report the microbial compositional profiles of a whopping 27,751 samples from 97 independent studies, providing insights into the diversity of microorganisms — from the bacterial and archaeal domains — in a wide range of geographic and environmental ecosystems, both terrestrial and aquatic. From these samples, Thompson and colleagues generated 2.2 billion DNA sequence reads of a highly variable region of a universally evolutionarily conserved gene called 16S rRNA, which encodes a component of the ribosome (the cell's protein-synthesis machinery).

The remarkable nature of this study lies not only in its scale and in the breadth of the environmental samples analysed (Fig. 1), but also in its methodology. The project involved a massive, global crowdsourcing effort in which scientists raided their collection freezers for samples to share with the project.

The approach was straightforward. A call was made for scientists to contribute well-preserved environmental samples collected during specific research projects, and the Earth Microbiome Project offered to sequence the DNA of the 16S rRNA gene in the microbial samples and to make the data available as open access.

This project is a prime example of a trend in the adoption of scientific approaches involving widespread engagement, in which the ease of electronic communication and the power of social media are harnessed to generate useful resources. In the same spirit, in the Polymath Project, mathematicians collaborate to tackle challenging mathematical problems.

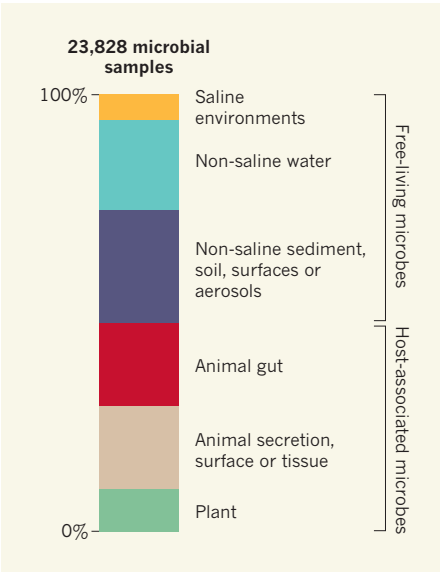

Figure 1 | Environmental sources of microbial samples analysed by the Earth Microbiome Project. Thompson *et al.*[1] report data generated from their work in a project to assess microbial diversity across Earth. Scientists from 43 countries sent the project's authors 27,751 DNA samples of free-living and host-associated microorganisms collected from various environments. After a quality-control step, the authors analysed part of the sequence of a gene called 16S rRNA in 23,828 of the samples.

Such approaches to generating crowdsourced experimental data usually work by first getting the project under way and obtaining funding later as the project gathers momentum, perhaps by crowdfunding. Examples of crowdsourced projects include those analysing bacteria in the human gut, such as the Flemish Gut Flora Project[3], or the Personalized Nutrition Project[4]. Such studies contrast with conventional research collaborations that begin once a grant is obtained from a funding agency.

Challenges inevitably arise in the type of work conducted by Thompson *et al.*, particularly from having to handle samples from many collection sites. A common frustration in microbial research is that sampling procedures, storage, transport conditions, DNA extraction and amplification protocols often result in a 'lab-of-origin' effect that makes it difficult to compare data generated by different research groups. To address this, the Earth Microbiome Project developed a range of protocols[5] and standards for sample collection, DNA extraction, transport and the formatting of the associated auxiliary data (such as temperature or location), as well as data-analysis procedures. These protocols were used for the project itself, but have been rapidly adopted by the wider research community, and more than 2,000 papers have already been published that use them[6]. By having a single protocol for all samples, and running all analyses in one laboratory, Thompson and colleagues have tried to remove as many potential technical confounding factors as possible.

The results seem to confirm that they have succeeded, revealing that sample microbial profiles cluster by environment — those from a specific type of environment are more similar to each other than to those from other types of environment, irrespective of the research group that collected the sample. This approach also has a drawback, because a single DNA-extraction protocol cannot be expected to perform equally well across the wide chemical and biological variability of the samples collected in this type of broad survey, and might be less effective than a targeted approach in which extraction protocols are optimized for the environment being sampled. Thompson and colleagues have favoured generalizability over sensitivity, a choice that can surely be defended in these circumstances.

Another limitation of the study is its lack of hypothesis-driven experimental design,

because it deliberately positions itself as an exploratory data analysis across different environments and sample types. This produces certain constraints on the inferences that can be made, because environmental data collected for the samples were not always measured in the same way in different environments.

The debate about the relative merits of data-driven and hypothesis-driven experimental approaches is not new, and there are examples of each of these approaches providing scientific insights. This study is an excellent example of the former, even if concessions had to be made regarding the selection of variables that could be used for analyses across all the environments.

Thompson and colleagues made several findings. For example, they investigated whether existing theories about the relationship between species richness (as monitored by the diversity of 16S rRNA sequences) and temperature and pH across environments were consistent with their data. For example, there is a model that proposes a steady logarithmic rise of microbial richness with increasing temperature[7,8]. Surprisingly, in contrast to this theory, the authors found that microbial biodiversity peaks at a relatively narrow pH and temperature range and then drops again.

The authors also observed an unexpectedly high amount of 'nestedness' among samples from different environments: samples showing low biodiversity were always present as microbial subsets of other, high-biodiversity samples, irrespective of the sample origin. Notably, this pattern of nestedness was mostly observed for microbial analyses above the level of genus — when analysed at the level of species, or when different strains of the same species were analysed, a strong decrease in nestedness was observed.

The value of the Earth Microbiome Project will extend far beyond what is reported in the present paper. The project provides a resource that will keep microbial ecologists and evolutionary biologists busy for years. More than 60 publications have already been published using subsets of the data that had been released previously [6]. By implementing and fiercely pursuing this open-access model, Thompson and colleagues emphasize the value of collaboration and sharing over competition, which is unfortunately still too frequent in the scientific community. ■

Jeroen Raes *is in the Department of Microbiology and Immunology, KU Leuven–Rega Institute, and in the VIB, Center for Microbiology, Leuven 3000, Belgium.*
e-mail: jeroen.raes@kuleuven.be

1. Thompson, L. R. *et al. Nature* **551,** 457–463 (2017).
2. Gilbert, J. A., Jansson, J. K. & Knight, R. *BMC Biol.* **12,** 69 (2014).
3. Falony, G. *et al. Science* **352,** 560–564 (2016).
4. Zeevi, D. *et al. Cell* **163,** 1079–1094 (2015).
5. www.earthmicrobiome.org/protocols-and-standards
6. www.earthmicrobiome.org/publications
7. Allen, A. P., Brown, J. H. & Gillooly, J. F. *Science* **297,** 1545–1548 (2002).
8. Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M. & West, G. B. *Ecology* **85,** 1771–1789 (2004).

This article was published online on 1 November 2017.

ORGANIC CHEMISTRY

Super-reactive catalyst for bond cleavage

Carbon–hydrogen bonds in organic molecules can be cut to install other chemical groups on the carbon atom, but these reactions have been limited. A catalytic palladium complex opens up fresh opportunities. SEE LETTER P.489

JOANNA WENCEL-DELORD & FRANÇOISE COLOBERT

Compounds such as drugs, agrochemicals and plastics are prepared from simple chemical precursors through multistep synthetic routes. Accordingly, strategies that permit straightforward conversion of simple starting materials into the desired molecular structures, avoiding the additional steps and fancy tricks often needed for chemical transformations, are urgently needed. On page 489, Wang *et al.*[1] report a remarkable advance that addresses this issue using a strategy known as non-directed C–H functionalization.

The basic components of all organic molecules are carbon and hydrogen atoms. The strong C–H bonds that form between these atoms account for the stability of organic molecules, but they also make it difficult to modify such molecules by selectively replacing hydrogen atoms with other chemical groups. Moreover, replacing a single hydrogen can be difficult without destroying the whole molecular system, because of the 'harsh' reaction conditions that are generally required. Solving these problems has been a real challenge for organic chemists, and has led to the establishment of a field known as C–H bond functionalization[2,3]. The most extensively explored solution involves using transition metals — particularly the noble metals, which under certain conditions are sufficiently active to cleave C–H bonds.

Another fundamental issue is how to target one hydrogen selectively in the presence of many others that have very similar chemical

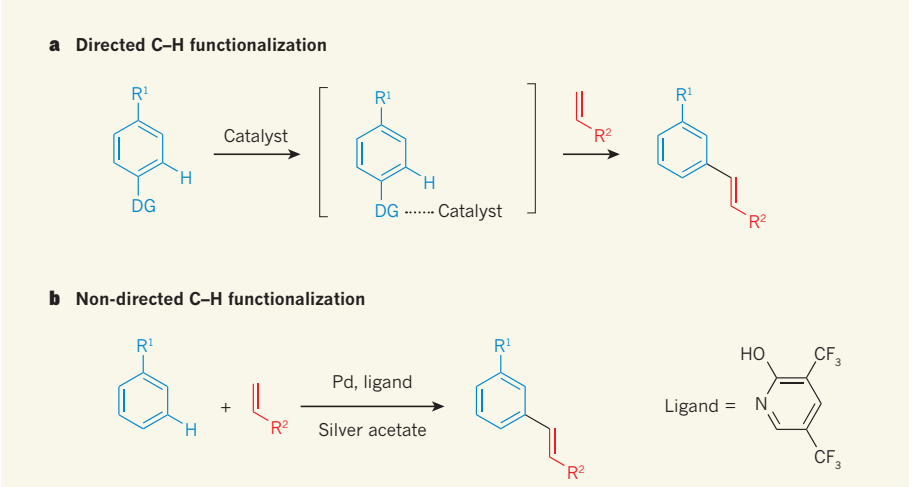

Figure 1 | A non-directed, palladium-catalysed C–H functionalization. a, In directed C–H functionalization reactions, such as the one depicted here, a directing group (DG) on the starting material guides (dotted line) the catalyst to a single hydrogen atom; only the reactive hydrogen atom on the benzene ring is shown. R^1 and R^2 represent any chemical group. Square brackets indicate a transiently formed state. The catalyst is typically a palladium, rhodium or ruthenium complex. **b**, Wang *et al.*[1] report a catalyst for non-directed C–H functionalizations. The highly active and robust catalyst is formed *in situ* from a palladium (Pd) source and a finely tuned ligand (a 2-pyridone molecule), and cleaves C–H bonds on both electron-rich and electron-poor substrates in the presence of silver acetate. The selectivity of the reactions for the hydrogen atom shown is moderate to good.