PEOPLE & IDEAS

Sarah Teichmann: Science always wins if we work together

Lucie Van Emmenis[1]

Sarah Teichmann is a professor at the University of Cambridge, where she leads a research group that uses a combination of genomics, artificial intelligence (AI), and bioinformatics to further understand various aspects of immunity. She is also one of the co-founders of the Human Cell Atlas, a nonexecutive director of 10x Genomics, an advisor for multiple biotech startups, and now a part-time Vice President, Translational Research at GlaxoSmithKline (GSK). We spoke to Sarah about expanding beyond academic science in your career, dealing with uncertainty in the workplace, and being able to carve out time for yourself despite having multiple demands on your time.

Please tell us a little about yourself and how you first became interested in science (i.e., where did you grow up? What was your first experience of science?)

I grew up in a village in southwest Germany called "vineyard" (Weingarten/Baden), which pretty much describes the idyllic childhood that my two sisters and I had. My father was a German electrical engineer, and my mother is an American German literature academic. This meant that we grew up speaking both English and German, and so all three of us sisters were offered a place in the European School Karlsruhe, which is one of the dozen or so schools linked to European Union (EU) institutions to educate the children of the EU civil servants. (The school that we attended is affiliated with an atomic physics research institute funded by the EU in Karlsruhe/Germany, and tuition was free for all students at that time.) The school had a high standard of teachers and teaching in both humanities and sciences, and we were lucky to have a chemistry and biology teacher (Mr. Walter Henderson) who offered an after-school science club with access to the school lab facilities. This was where I was able to embark on a science project studying the biochemistry of the metabolic changes in leaves throughout the seasons—motivated by the super simple question of why leaves

change color during the year. Of course the answer is not simple, but it opened my eyes to the hidden world of invisible cells and molecules that are the basis of life, which really gripped my imagination to this day.

Tell us about your career trajectory, and what led you to becoming a group leader.

The science project mentioned above meant that I won a bunch of competitions, which likely helped me gain a place to read Natural Sciences at Cambridge University. The Cambridge admissions interview at Trinity College (with immunologist Michael Neuberger and physical chemist Iain MacDonald) was tough, and I did not get the impression I did that well, so the science competition success may have been important. Studying at Cambridge opened up a whole new world of science and research opportunities during my BA and PhD years. I carried out my BA research project on labeling patterns on nuclear magnetic resonance (NMR) in the lab of Ernest Laue, where I learned UNIX and started scripting. This led to a PhD in the lab of Cyrus Chothia in the early days of bioinformatic analysis of whole genomes. Then I moved to the Thornton Group, at University College London at that time, funded by a Beit Memorial Fellowship.

Sarah Teichmann.

During that time I was invited to interview at the Structural Biology Department at Stanford and then the Structural Studies Division at the MRC Laboratory of Molecular Biology, where I started my group in October 2001. I was fortunate in the sense that two great scientists and great environments were proactive in approaching me for principal investigator jobs: Michael Levitt at Stanford and Richard Henderson at the MRC Laboratory of Molecular Biology (both later went on to be awarded Nobel prizes). I was too young to be very savvy about jobs and careers at that time, and having interest from two places meant that I got an offer that was not utterly shabby. At the same

---

[1]Senior Scientific Editor, *JEM*, Rockefeller University Press, New York, NY, USA.

Correspondence to Lucie Van Emmenis: lvanemmenis@rockefeller.edu.

time, I want to emphasize that the career path was by no means easy for a young woman in computational biology at the time for at least two reasons: equal opportunities in UK science were not as much in the consciousness of society and of institutions as they are now, and the field of data-driven science was not as central to biomedical research as it is now (as per your question on artificial intelligence [AI] below). Things have changed for the better with respect to both equal opportunities and recognition of the importance of data science/AI in biomedical research over the past 30 years or so, but there is still room for improvement on both fronts in the academic community.

It is important to be aware that I always enjoyed the process of blue skies research and wanted to stay in this career path but did not feel that an academic career was a certainty due to the vagaries of academia. During the early years, I watched my contemporaries exit the academic career pathway and go into management consulting, finance, etc., and this type of option was always a plan B. It is only now that I am much more aware of the full breadth of opportunities that science training provides, which includes the startup and biotech ecosystem, big pharma, venture capital, publishing and communications, and many more options. I have been fortunate to be involved in founding two startup companies (Transition Bio and Ensocell Therapeutics) and now work part-time (30%) at GlaxoSmithKline (GSK) as Vice President (VP) of Translational Research.

How did you first become interested in bioinformatics?

In my final year of my biochemistry bachelor's degree (1995/96), I got into scripting/coding through a project in NMR as mentioned above. At the same time, some of our required reading included, e.g., the "One thousand families for the molecular biologist" Nature News &; Views article by Cyrus Chothia (1992) and "Protein superfamilies and domain superfolds" by Orengo/Jones/Thornton (1994), which estimated the number of protein families based on the available protein sequence and structure data at the time. The scope to answer big-picture questions in biology through data-driven approaches fascinated me and got

me hooked on this way of doing science from that point onward.

What are you currently working on, and what projects are you most excited about?

We are continuing to drive forward the Human Cell Atlas (HCA) project through both data science/artificial intelligence markup language (AIML) as well as data generation focused on the heart, the gut, and the immune system (especially thymus) in development, adult, and disease tissues. Spatial technologies are currently catalyzing exciting new discoveries in tissue architecture, and we are getting closer and closer to building full 3D models of organs by integrating multiple measurements of the same modality or making integrative models of multiple modalities. An example is our model of a quasi-3D thymus cell atlas analysis in Yayon et al. (2024), and there are also other approaches popping up to achieve this. Full 3D cell atlases give us a new way of understanding the function of organs at single-cell resolution and full molecular breadth, which is where we always wanted the HCA to get toward.

In terms of data science/AIML, we are now in the era of data integration of the HCA, and we are actively building reference data objects for single-cell/single-nuclear RNA sequencing data for all 18 of the major organs/systems of the human body as a community. We are releasing these data objects on the HCA Data Portal at https://data.humancellatlas.org as they are completed, and will update them over the coming years with the new version.

This data will form the basis for single-cell foundation models for human cells, which will in future likely also encompass training data from in vitro data (e.g., Billion Cells Project), perturbation experiments (e.g., Virtual Cells Project), and from cells of other organisms (e.g., Biodiversity Cell Atlas). The sheer amount of data and the AI modeling methods will make the HCA data more accessible and also more impactful through modeling directly in context with human in vitro cell data and model (and non-model) organism cell data.

From a biological perspective, integrating HCA data across the whole human body is opening up opportunities, such as gaining a new insight into how our organs communicate with each other across large distances

via hormones, which we are working to address with the Farooqi Group in the Clinical School at Cambridge. This is just one example of the many new ways that we can now approach biomedical questions by interrogating the HCA.

Please tell us about some work in your field that you are currently interested in.

Besides the HCA technologies I have mentioned above (spatial genomics, foundation models, and agentic AI), I am interested in where multiscale modeling of the human body can take us—from the molecular to the population and disease scale. I am co-director of the Canadian Institute for Advanced Research Multiscale Mapping and Modeling of the Human Body program, where we brainstorm around these questions in two meetings per year with a small community of fellows from across the world and across disciplines (https://cifar.ca/). The opportunities for integrating the HCA data with, e.g., 3D structures of protein complexes à la AlphaFold, are one exciting area, for instance, which could help us de-orphanize receptors on the cell surface. Another is systematic integration with genetics data, e.g., HLA and B/T cell receptor sequence variation across single-cell data sets, which may open up our ability to crack the enigmatic antigen–antigen receptor code in human immunology. Integrating the HCA data and what I will loosely call "disease cell atlas" data with medical records and medical diagnostic modalities (e.g., imaging) in a systematic way is a huge future area of opportunity for understanding disease pathology, developing better diagnostics, and identifying the best drug targets. This is a huge active area of interest for pharma and biotech.

Related to this is using the HCA and disease cell atlas data sets for benchmarking and modeling in vitro systems and designing them in an optimal way. We are using this approach for the thymus for instance, using the in vivo thymus cell atlas to identify which in vitro systems best mimic which tissues and cell compartments of the thymus to make diverse T cell subsets in vitro.

In summary, we are at a tremendously exciting juncture in biomedical research, which we could call "the era of human biology": large amounts of data and

integration and modeling with AI are making the human body accessible from the molecular level all the way to whole-organism physiology in a way that is more comprehensive and quantitative than ever before.

There have been big technological advancements in single-cell sequencing and genomics research over the years, and now there is obviously a lot of interest in what AI can do in research. This is a big question, but how do you feel about the increased use of AI, and do you think it will help (or hinder) your work?

AI for biomedical data modeling is undoubtedly a huge enabler. It is also increasingly used for guiding experiments and analyses and for writing code. The formalized version of this are AI agents that interact with data and models in an iterative way. All of these uses of AI are rapidly becoming pervasive in the biomedical research community, and in general I feel this is a huge opportunity—perhaps on par with the introduction of the internet, for example. It is making things faster and easier and allowing us to leverage the vast amounts of data of different types (literature, genomics, protein structure, etc.) distributed across databases and the internet in a way that was not possible before.

You are one of the co-founders of the HCA, which is an amazing project! How important are collaborations like these in science, and are they easy to put together?

The international HCA consortium is really a new way of collaborating across the scientific community through its open membership model. Both my co-founder Aviv Regev and I, as well as the other organizing committee members, were all fiercely committed to building an inclusive and open community. This continues to benefit the HCA project by welcoming scientists from across disciplines and across the world into the community, whether it is into one of our working groups, biological networks, or short-term task forces. The coordination of such a large community is not easy and has worked through a combination of volunteer work by a huge number of people in the various leadership positions over the years. Examples of challenges include misunderstandings across different scientific disciplines (for instance, a "model" means something very different in the compbio/machine learning community versus in cell or tissue biology) and across teams working in different countries in different time zones. What made it easier over time was the establishment of a funded executive office of about a dozen amazing staff members and contractors spread across the UK, EU, USA, and Asia. They provide a professional backbone for the community and are incredibly helpful when it comes to coordination and organization. They have been funded by a variety of sources over the years, ranging from philanthropic funding to our pharma partner program.

The success and impact of the HCA data (which are far from completed) are now evident across both academic and biotech/pharma drug discovery. I would like to think that this shows how important these types of collaborative projects can be, and how much more we can achieve as a community by working together. The concept of team science and recognition of the power of this way of working is growing across the community, including in institutions and funding bodies. The new generation that "grew up" with the HCA way of working are "native collaborators" as well as "native cell atlas and AI technologists," and they will shape the academic community in the coming decades.

How do you deal with uncertainty in the workplace?

There are two types of uncertainty in academic research: scientific uncertainty, which is our raison d'être, and job uncertainty, which is generally not pleasant for most people yet increasingly enters our workplace. We are in more uncertain times from a global political point of view than I have experienced previously, and this does have an impact on science. It is all the more important that we have high-quality governance and leadership in science. One aspect of good leadership means being transparent about how and why there may be uncertainty in the workplace from a funding and work contract point of view and providing certainty about uncertainty (if that makes sense). In my experience, being transparent and providing context to decisions leads to trust in the decision-makers. Personally, after working at a variety of research institutes over the past 25 years, aged 50, I have found myself as the sole provider for my family in recent years. This has meant that I feel comfortable with my recent move to a chair in Cambridge University combined with a VP of Translational Research position in GSK. These are both large organizations that are relatively stable within the greater scheme of things and have been a great learning experience over the past 18 mo.

Dealing with and resolving scientific uncertainty also requires transparency and clarity of communication about what or why there are uncertainties in data or results. This is the nature of many of our daily debates among colleagues. Because there is not necessarily one right answer about how to resolve uncertainty, there can be disagreements about how to proceed, what level of uncertainty to accept prior to submitting data for publication, etc. I feel that it is important to maintain the tone of these discussions to be as fun and friendly as possible, and remember that it is always science that wins if we work together to make new discoveries and resolve open questions.

What do you most enjoy about your work/role as a group leader?

That is an easy one: I love discussing science and ideas with my academic group and our amazing collaborators. The academic environment consistently brings incredibly talented and creative trainees from all over the world to our lab, and it is a pleasure and an honor to be able to spend time with them.

While not in the lab, how do you like to spend your time, or alternatively, how would you like to spend your time?

I work in multiple labs! In addition to being a professor at Cambridge University and running an academic research group, I am a part-time VP of Translational Research at GSK and have also founded and advised startup companies, such as most recently Ensocell Therapeutics. I am also a nonexecutive director of 10x Genomics and am writing this piece while on an airplane to a 10x board meeting in the Bay Area.

So there is not a huge amount of free time for me at the moment, but what there is I spend with my husband and our two daughters, who have sports and outdoors

activities that we support, as well as my extended family. (We are fortunate that my middle sister, artist Esther Teichmann, and her family live nearby; all other relatives live in mainland Europe or USA.) At the weekend, I usually play tennis at a local club with an old friend for singles and a large formalized rotational doubles group, and in the pandemic, I took up standup paddle board and sprint triathlon racing, which I still try to do when the weather and time permit and would love to have more time for.

Reference

Yayon, N., V.R. Kedlian, L. Boehme, C. Suo, B.T. Wachter, R.T. Beuschel, O. Amsalem, K. Polanski, S. Kopley, E. Tuck, et al. 2024. A spatial human thymus cell atlas mapped to a continuous tissue axis. *Nature*. 635:708–718. https://doi.org/10.1038/s41586-024-07944-6