https://doi.org/10.1038/s41577-026-01340-z

How Big Data and AI are revolutionizing human immunology



By Sarah A. Teichmann

Check for updates

Understanding human biology in health and disease has entered a new era, one defined not just by better tools but by a fundamental shift in how biological knowledge is generated.

For most of its history, immunology has advanced through hypothesis-driven experiments: a researcher forms an idea, designs a test and interprets the result. This approach has been enormously productive but is inherently limited by what scientists already suspect to be true. 'Big Data' and artificial intelligence (AI) are changing this by enabling unbiased, comprehensive interrogation of biological systems — surfacing principles and patterns that no previous hypothesis would have anticipated.

> "AI is … not merely a convenience … it is a necessary partner in extracting meaning from data that would otherwise remain intractable"

The scale of data now available to researchers is qualitatively different from anything that came before. Technologies that profile individual cells across entire tissues, organs and bodies — capturing gene activity, molecular identity and spatial position simultaneously — generate datasets of extraordinary richness. No human researcher can hold this complexity in mind at once. AI is therefore not merely a convenience here; it is a necessary partner in extracting meaning from data that would otherwise remain intractable and opaque. Crucially, this is not AI operating in isolation: it works in an iterative loop with scientists, generating new concepts and identifying patterns that are tested experimentally, refined and then fed back into the models. The result is a new kind of scientific reasoning; one that moves fluidly between the computational and the empirical, and that can biological principles at a scale and resolution previously out of reach. In response to a previous article by Ruslan Medzhitov[1], in which he discusses the imbalance between generating data and developing theories, here I discuss how a new approach to immunology will bridge and strengthen both the empirical and theoretical pillars of scientific progress.

A central resource driving this transformation is the Human Cell Atlas, an international effort to map every cell type in the human body across tissues, developmental stages (including pregnancy) and the human lifespan. What makes this resource so powerful is precisely its comprehensiveness and lack of prior bias: it does not set out to prove a specific point about immunity, but to document biology as it actually is. When AI tools are applied to these data, they can reveal how immune cell identity shifts with tissue context, how regulatory circuits are organised, how cells communicate, and how these patterns change with age or disease. These are not insights extracted from targeted experiments — they are principles that emerge from the data itself.

This unbiased quality proved especially valuable during the COVID-19 pandemic. Instead of speculating about how a new virus might infect human tissue, we could query existing cell atlas data to identify, objectively and rapidly, which cells across the body carry the molecular machinery needed for viral entry[2]. This approach identified the nasal epithelium, corneal epithelial cells and salivary glands as key sites of infection and transmission — findings that were unexpected, rapidly validated and directly translated into public health guidance. The atlas functioned not as a tool for answering pre-specified questions but as an open framework to generate new ones. This principle continues to be deployed as new emerging viral threats surface and it has transformed the speed and accuracy with which we can assess horizontal and vertical mechanisms of viral transmission.

The same logic applies to understanding immune diversity across individuals. Blood is the most accessible human tissue, and large-scale profiling across thousands of people of different ages, sexes and ancestries has shown that individuals carry stable, characteristic immune profiles — distinct setpoints that seem to influence resilience and disease susceptibility throughout life. Children and supercentenarians[3] each have immune signatures associated with protection, but of entirely different kinds. These are not findings that would have emerged from small, targeted studies; they required the statistical power and molecular resolution that only large-scale data and AI analysis can provide. They point towards a new framework for understanding immune homeostasis, not as a fixed state but as a dynamic, individually variable landscape shaped by genetics, environment and time.

Beyond identifying cell types and states, one of the deepest ambitions of this AI-guided approach is to decode the rules that govern how the immune system recognizes threats. The adaptive immune system operates through an astronomically large repertoire of receptors, each subtly different, collectively capable of recognising almost any molecular pattern. Understanding the logic of this recognition — what structural and sequence features determine which targets a receptor will bind — is one of immunology's great unsolved problems. Large datasets of receptor sequences, combined with AI tools that can model receptor–target interactions structurally, are beginning to make this problem tractable. In this case, AI is not just accelerating known biology but enabling a genuinely new class of question: moving from cataloguing what receptors exist to understanding the molecular grammar that governs their function.

The translational consequences of this shift in approach are profound. In drug development, the high failure rate of clinical trials reflects in large part the difficulty of choosing the right targets; a decision that has traditionally been made with incomplete information. AI systems that integrate large-scale single-cell data with clinical trial outcomes are now able to identify which drug targets, when selected based on cell type-specific expression patterns, are substantially more likely to succeed and less likely to cause harm. This is not

CREDIT: IMAGE COURTESY OF CÉCILE BURBAN

incremental improvement; it represents a principled, data-driven approach to a problem that has resisted solution for decades.

Similarly, laboratory models of human lymphoid tissue have historically been difficult to validate in that it has been unclear how closely they recapitulate the biology they are meant to represent. Cell atlas data provide, for the first time, a rigorous molecular reference against which in vitro models can be benchmarked to assess how accurately they recapitulate our bone marrow, thymus, spleen or lymph nodes. AI tools that compare gene activity across cell development and differentiation trajectories can identify precisely where a laboratory model diverges from real tissue, and point towards specific interventions — adding a missing signalling molecule, for instance — that bring it into closer alignment. This closes a loop that has long been open: data from real human tissue improve laboratory models, and improved models generate better experimental data, which in turn refine our biological understanding.

Looking ahead, these converging capabilities point towards something more ambitious: a computational model of the human immune system capable of predicting individual responses to infection, vaccination or therapeutic intervention. Early versions of this vision are already taking shape. Machine learning applied to immune receptor sequences can now identify a range of conditions — infectious, autoimmune, metabolic — from a single blood sample, without any prior knowledge of which specific immune targets are involved[4]. This is a striking demonstration of the latent information encoded in immune receptor repertoires, and of the power of AI to extract it.

> "biological understanding is deepened, not diminished, when we let data speak before hypothesis"

The common principle that unites all these developments is that biological understanding is deepened, not diminished, when we let data speak before hypothesis. The immune system is vastly more complex than any single investigator's intuition can encompass. Big data and AI do not replace scientific thinking; they expand its scope, providing an unbiased scaffold on which new principles can be discovered, tested and ultimately translated into medicine. We are at the beginning of understanding what a truly comprehensive, predictive model of human immunity might look like. The foundations are being laid now.

Sarah A. Teichmann [1,2,3] ✉
[1]Cambridge Stem Cell Institute, Department of Medicine, Jeffrey Cheah Biomedical Centre, University of Cambridge, Cambridge, UK. [2]Department of Medicine, University of Cambridge, Cambridge, UK. [3]CIFAR Macmillan Multi-scale Human Programme, CIFAR, Toronto, Ontario, Canada.
✉e-mail: sat1003@cam.ac.uk

Published online: 23 July 2026

References

1. Medzhitov, R. On the balance of knowledge. *Nat. Rev. Immunol.* **26**, 163–165 (2026).
2. Sungnak, W. et al. SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nat. Med.* **26**, 681–687 (2020).
3. Hashimoto, K. et al. Single-cell transcriptomics reveals expansion of cytotoxic CD4 T cells in supercentenarians. *Proc. Natl. Acad. Sci. USA* **116**, 24242–24251 (2019).
4. Zaslavsky, M. E. et al. Disease diagnostics using machine learning of B cell and T cell receptor sequences. *Science* **387**, eadp2407 (2025).

Acknowledgements

The author thanks O. Dufva, Y. Gu, L. Kretschmer and I. Sarroupoulos for constructive comments and feedback, and G. Alderton for writing support.

Competing interests

S.A.T. is a scientific advisory board member of Bioptimus, ForeSite Labs and Xaira Therapeutics; a co-founder, board observer and equity holder of TransitionBio; a co-founder, consultant and board director of Ensocell Therapeutics; a non-executive director of 10x Genomics; and a part-time employee of GlaxoSmithKline.