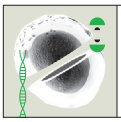
THE CELL SEEKER

Aviv Regev uses genomic techniques to examine single cells. Now she is part of an effort to map every cell in the human body.

BY ANNA NOWOGRODZKI

Aviv Regev likes to work at the edge of what is possible. In 2011, the computational biologist was collaborating with molecular geneticist Joshua Levin to test a handful of methods for sequencing RNA. The scientists were aiming to push the technologies to the brink of failure and see which performed the best. They processed samples with degraded RNA or vanishingly small amounts of the molecule. Eventually, Levin pointed out that they were sequencing less RNA than appears in a single cell.

To Regev, that sounded like an opportunity. The cell is the basic unit of life and she had long been looking for ways to explore how complex networks of genes operate in individual cells, how those networks can differ and, ultimately, how diverse cell populations work together.



SINGLE-CELL BIOLOGY
A *Nature* special issue
go.nature.com/singlecell

The answers to such questions would reveal, in essence, how complex organisms such as humans are built. "So, we're like, 'OK, time to give it a try'," she says. Regev and Levin, who both work at the Broad Institute of MIT and Harvard in Cambridge, Massachusetts, sequenced the RNA of 18 seemingly identical immune cells from mouse bone marrow, and found that some produced starkly different patterns of gene expression from the rest[1]. They were acting like two different cell subtypes.

That made Regev want to push even further: to use single-cell sequencing to understand how many different cell types there are in the human body, where they reside and what they do. Her lab has gone from looking at 18 cells at a time to sequencing RNA from hundreds of thousands — and combining single-cell analyses with genome editing to see what happens when key regulatory genes are shut down.

The results are already widening the spectrum of known cell types — identifying, for example, two new forms of retinal neuron[2] — and Regev is eager to find more. In late 2016, she helped to launch the International Human Cell Atlas, an ambitious effort to classify and map all of the estimated 37 trillion cells in the human body (see 'To build an atlas'). It is part of a growing interest in characterizing individual cells in many different ways, says Mathias Uhlén, a microbiologist at the Royal Institute of Technology in Stockholm: "I actually think it's one of the most important life-science projects in history, probably more important than the human genome."

Such broad involvement in ambitious projects is the norm for Regev, says Dana Pe'er, a computational biologist at Memorial Sloan Kettering Cancer Center in New York City, who has known Regev for 18 years. "One of the things that makes Aviv special is her enormous bandwidth. I've never met a scientist who thinks so deeply and so innovatively on so many things."

UNDECIDED

When Regev was an undergraduate at Tel Aviv University in Israel, students had to pick a subject before beginning their studies. But she didn't want to decide. "Too many things were interesting," she says. Instead, she chose an advanced interdisciplinary programme that would let her look at lots of subjects and skip a bachelor's degree, going straight to a master's.

A turning point in her undergraduate years came under the tutelage of evolutionary biologist Eva Jablonka. Jablonka has pushed a controversial view of evolution that involves epigenetic inheritance, and Regev says she admired her courage and integrity in the face of criticism. "There are many easy paths that you can take, and it's always impressive to see people who choose alternative roads."

Jablonka's class involved solving complicated genetics problems, which Regev loved. She was drawn to the way in which genetics relies on abstract reasoning to reach fundamental scientific conclusions. "I got hooked on biology very deeply as a result," she says. "Genes became fascinating, but more so how they work with each other. And the first vehicle in which they work with each other is the cell."

Regev did a PhD in computational biology under Ehud Shapiro from the Weizmann Institute of Science in Rehovot, Israel. In 2003 she moved to Harvard University's Bauer Center for Genomics Research in Cambridge, through a unique programme that allows researchers to leapfrog the traditional postdoctoral fellowship and start their own lab. "I had my own small group and was completely independent," she says. That allowed her to define her own research questions, and she focused on picking apart genetic networks by looking at the RNA molecules produced by genes in cells. In 2004, she applied this technique to tumours and found gene-expression patterns that were shared across wildly different types of cancer, as well as some that were more specific, such as a group of genes related to growth inhibition that is suppressed in acute lymphoblastic leukaemias[3]. By 2006, at the age of 35, she had established her lab at the Broad Institute and the Massachusetts Institute of Technology in Cambridge.

SHATTERING SIMILARITIES

At Broad, Regev continued working on how to tease complex information out of RNA sequencing data. In 2009, she published a paper on a type of mouse immune cell called dendritic cells, revealing the gene networks that control how they respond to pathogens[4]. In 2011, she developed a method that could assemble a complete transcriptome[5] — all the RNA being transcribed from the genes in a sample — without using a reference genome, important when an organism's genome has not been sequenced in any great depth.

It was around this time that Levin mentioned the prospect of sequencing the RNA inside a single cell. Up to that point, single-cell genomics had been almost impossible, because techniques weren't sensitive enough to detect the tiny amount of RNA or DNA inside just one cell. But that began to change around 2011.

The study with the 18 immune cells — also dendritic cells — was meant to test the method. "I had kind of insisted that we do an experiment to prove that when we put the same cell types in, everything comes out the same," says Rahul Satija, Regev's postdoc at the time, who is now at the New York Genome Center in New York City. Instead, he found two very different groups of cell subtypes. Even within one of the groups, individual cells varied surprisingly in their expression of regulatory and immune genes. "We saw so much in this one little snapshot," Regev recalls.

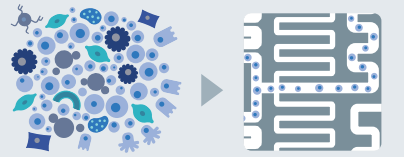
"I think even right then, Aviv knew," says Satija. "When we saw those results, they pointed the way forward to where all this was going to go." They could use the diversity revealed by single-cell genomics to uncover the true range of cell types in an organism, and find out how they were interacting with each other.

In standard genetic sequencing, DNA or RNA is extracted from a blend of many cells to produce an average read-out for the entire population. Regev compares this approach to a fruit smoothie. The colour and taste hint at what is in it, but a single blueberry, or even a dozen, can be easily masked by a carton of strawberries.

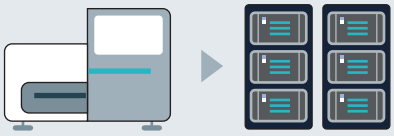
By contrast, "single-cell-resolved data is like a fruit salad", Regev says. "You can distinguish your blueberries from your blackberries from your raspberries from your pineapples and so on." That promised to expose a range of overlooked cellular variation. Using single-cell genomics to sequence a tumour, biologists could determine which genes were being expressed by malignant cells, which by non-malignant cells and which by blood vessels or immune cells — potentially pointing to better

TO BUILD AN ATLAS

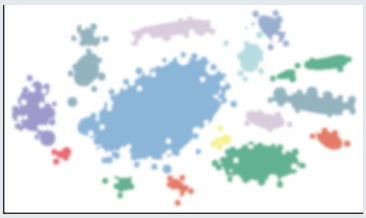
Scientists wishing to put together a 3D map of the thousands of cell types and subtypes in the human body will face challenges at every step.



Sophisticated devices will be required to isolate different kinds of human cells from a range of tissues and prepare them for study in a way that does not stress them or change their nature.



Sequencing must account for variability in the amount and quality of RNA or other molecules in different cell types, and yet computational approaches need to be standardized to ensure compatibility.



Multidimensional maps based on sequencing data will reveal the relative types, subtypes and abundances of cells in tissues, but in many cases these must be mapped back to where they reside in the body, using different spatial methods.

CASEY ATKINS FOR *NATURE*

ways to attack the cancer.

The technique holds promise for drug development in many diseases. Knowing which genes a potential drug affects is more useful if there's a way to comprehensively check which cells are actively expressing the gene.

Regev was not the only one becoming enamoured with single-cell analyses on a grand scale. Since at least 2012, scientists have been toying with the idea of mapping all human cell types using these techniques. "The idea independently arose in several areas of the world at the same time," says Stephen Quake, a bioengineer at Stanford University in California who co-leads the Chan Zuckerberg Biohub. The Biohub, which has been funding various biomedical research projects since September 2016, includes its own cell-atlas project.

THE HUMAN CELL ATLAS

Around 2014, Regev started giving talks and workshops on cell mapping. Sarah Teichmann, head of cellular genetics at the Wellcome Trust Sanger Institute in Hinxton, UK, heard about Regev's interest and last year asked her whether she would like to collaborate on building an international human cell atlas project. It would include not just genomics researchers, but also experts in the physiology of various tissues and organ systems.

Regev leapt at the chance, and she and Teichmann are now co-leaders of the Human Cell Atlas. The idea is to sequence the RNA of every kind of cell in the body, to use those gene-expression profiles to classify cells into types and identify new ones, and to map how all those cells and their molecules are spatially organized.

The project also aims to discover and characterize all the possible cell states in the human body — mature and immature, exhausted and fully functioning — which will require much more sequencing. Scientists have assumed that there are about 300 major cell types, but Regev suspects that there are many more states and subtypes to explore. The retina alone seems to contain more than 100 subtypes of neuron, Regev says. Currently, consortium members whose labs are already working on immune cells, liver and tumours are coming together to coordinate efforts on these tissues and organs. "This is really early days," says Teichmann.

In co-coordinating the Human Cell Atlas project, Regev has wrangled a committee of 28 people from 5 continents and helped to organize meetings for more than 500 scientists. "I would get stressed out of this world, but she doesn't," Jablonka says. "It's fun to have a vision that's shared with others," Regev says, simply.

It has been unclear how the project would find funding for all its ambitions. But in June, the Chan Zuckerberg Initiative — the philanthropic organization in Palo Alto, California, that funds the Biohub — contributed an undisclosed amount of money and software-engineering support to the Human Cell Atlas data platform, which will be used to store, analyse and browse project data. Teichmann sees the need for data curation as a key reason to focus on a large, centralized effort instead of many smaller ones. "The computational part is at the heart of the project," she says. "Uniform data processing, data browsing and so on: that's a clear benefit."

> "I WOULD GET STRESSED OUT OF THIS WORLD, BUT SHE DOESN'T."

In April, the Chan Zuckerberg Initiative had also accepted applications for one-year pilot projects to test and develop technologies and experimental procedures for the Human Cell Atlas; it is expected to announce which projects it has selected for funding some time soon. The applications were open to everyone, not just scientists who have participated in planning meetings.

BRAIN DRAIN

Some scientists worry that the atlas will drain both funding and effort from other creative endeavours — a critique aimed at many such international big-science projects. "There's this tension," says Atray Dixit, a PhD student in Regev's lab. "We know they're going to give us something, and they're kind of low-risk in that sense. But they're really expensive. How do we balance that?"

Developmental biologist Azim Surani at the University of Cambridge, UK, is not sure that the project will adeptly balance quantity and depth of information. With the Human Cell Atlas, "you would have a broad picture rather than a deeper understanding of what the different cell types are" and the relationships between them, he says. "What is the pain-to-gain ratio here?"

Surani also wonders whether single-cell genomics is ready to converge on one big project. "Has the technology reached maturity so that you're making the best use of it?" he asks. For example, tissue desegregation — extracting single cells from tissue without getting a biased sample or damaging the RNA inside — is still very difficult, and it might be better for the field, some say, if many groups were to go off in their own directions to find the best solution to this and other technical challenges.

And there are concerns that the project is practically limitless in scope. "The definition of a cell type is not very clear," says Uhlén, who is director of the Human Protein Atlas — an effort to catalogue proteins in normal and cancerous human cells that has been running since 2003. There may be a nearly infinite number of cell types to characterize. Uhlén says that the Human Cell Atlas is important and exciting, but adds: "We need to be very clear, what is the endpoint?"

Regev argues that completion is not the only goal. "It's modular: you can break this to pieces," she says. "Even if you solve a part of a problem, it's still a meaningful solution." Even if the project just catalogues all the cells in the retina, for example, that's still useful for drug development, she argues. "It lends itself to something that can unfold over time."

Regev's focus on the Human Cell Atlas has not distracted her from her more detailed studies of specific cell types. Last December, her group was one of three to publish papers[6–8] in which they used the precision gene-editing tool CRISPR–Cas9 to turn off transcription factors and other regulatory genes in large batches of cells, and then used single-cell RNA sequencing to observe the effects. Regev's lab calls its technique Perturb-seq[6].

The aim is to unpick genetic pathways very precisely, on a much larger scale than has been possible before, by switching off one or more genes in each cell, then assaying how they influence every other gene. This was possible before, for a handful of genes at a time, but Perturb-seq can work on 1,000 or even 10,000 genes at once. The results can reveal how genes regulate each other; they can also show the combined effects of activating or deactivating multiple genes at once, which can't be predicted from each of the genes alone.

Dixit, a co-first author on the paper, says Regev is indefatigable. She held daily project meetings at 6 a.m. in the weeks leading up to the submission. "I put in this joke sentence at the end of the supplementary methods — a bunch of alliteration just to see if anyone would read that far. She found it," Dixit says. "It was 3 a.m. the night before we submitted."

Regev's intensity and focus is accompanied by relentless positivity. "I'm one of the fortunate people who love what they do," she says. And she still loves cells. "No matter how you look at them, they're just absolutely amazing things." ■

Anna Nowogrodzki *is a science journalist based in Boston, Massachusetts.*

1. Shalek, A. K. *et al. Nature* **498,** 236–240 (2013).
2. Shekhar, K. *et al. Cell* **166,** 1308–1323.e30 (2016).
3. Segal, E. Friedman, N., Koller, D. & Regev, A. *Nature Genet.* **36,** 1090–1098 (2004).
4. Amit, I. *et al. Science* **326,** 257–263 (2009).
5. Grabherr, M. G. *et al. Nature Biotechnol.* **29,** 644–652 (2011).
6. Dixit, A. *et al. Cell* **167,** 1853–1866.e17 (2016).
7. Adamson, B. *et al. Cell* **167,** 1867–1882.e21 (2016).
8. Jaitin, D. A. *et al. Cell* **167,** 1883–1896.e15 (2016).