



# THE SINGLE LIFE

*Sequencing DNA from individual cells is changing the way that researchers think of humans as a whole.*

BY BRIAN OWENS

**A**ll Nicholas Navin needed was one cell — the issue was how to get it. It was 2010, and the postdoctoral fellow at Cold Spring Harbor Laboratory in New York was exploring the genetic changes that drive breast cancer. Most of the cancer-genome studies before then had ground up bits of tumour tissue and sequenced the DNA en masse, giving a consensus picture of the cancer's genome. But Navin wanted to work out the sequence from individual cells to see how

they had mutated and diverged as the cancer grew.

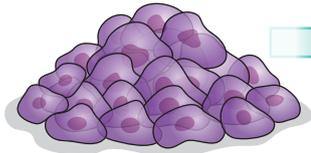
He ran into trouble almost immediately. "Cells like to stick together," he says. He tried the most advanced microdissection techniques, which use robots to peel cells apart or suck them into the tips of hair-thin glass pipettes. But he could never be sure that a second cell hadn't come along for the ride. Eventually, he settled on using chemicals to dissolve the cell's outer membrane and release the dense nucleus. Then he separated out the nucleus

VISUALS UNLIMITED/CORBIS

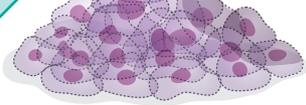
## ONE GENOME FROM MANY

Sequencing the genomes of single cells is similar to sequencing those from multiple cells — but errors are more likely.

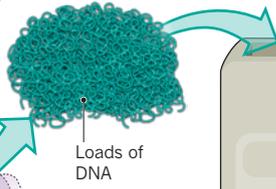
### ► Standard genome sequencing



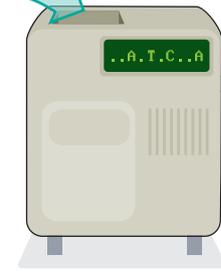
A sample containing thousands to millions of cells is isolated.



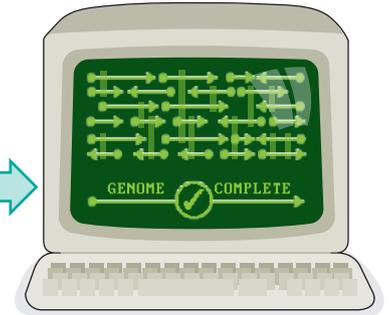
DNA is extracted from all the nuclei.



Loads of DNA

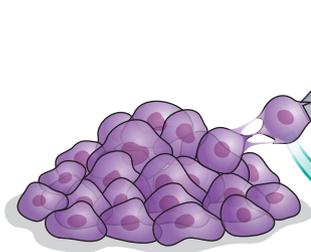


DNA is broken into fragments and then sequenced.

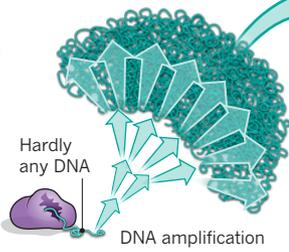


The sequences are assembled to give a common, 'consensus' sequence.

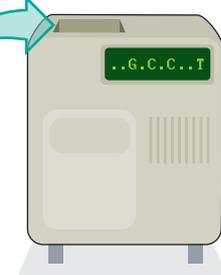
### ► Single-cell sequencing



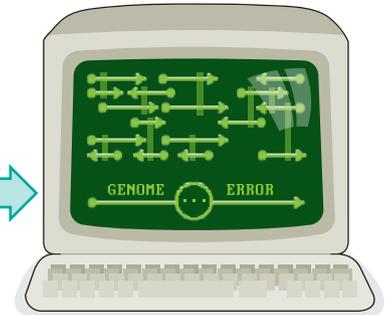
A single cell is difficult to isolate, but it can be done mechanically or with an automated cell sorter.



Hardly any DNA  
DNA amplification



Amplified DNA is sequenced.



Errors introduced in earlier steps make sequence assembly difficult; the final sequence can have gaps.

using an automated cell sorter, and extracted its DNA. He repeated the process for around 100 cells, and the sequences he obtained revealed how the tumour had evolved from a few rogue cells into a complex mélange of genetically distinct ones<sup>1</sup>.

The ability to sequence 100 human cancer genomes was unthinkable a decade ago, and it is still a remarkable feat. Technology has moved apace, dramatically reducing costs and making genome sequencing fairly routine. But most human genomes, cancer or otherwise, are still sequenced from DNA extracted from multiple cells, which misses differences between cells that could be crucial in controlling gene expression, cell behaviour and drug response.

## “I THINK IT’S THE NEXT LEVEL OF COMPLEXITY.”

“People are becoming very interested in what is the variation from cell to cell,” says Navin, now at the University of Texas MD Anderson Cancer Center in Houston. Last month, the US National Institutes of Health (NIH) announced that it would be funding single-cell studies, including genome sequencing, to the tune of US\$90 million over 5 years.

But challenges abound. Amplifying the tiny amount of DNA in a single cell until there is enough to sequence without introducing too many errors is still difficult (see ‘One genome from many’). The bioinformatics required to stitch the data together and deal with artefacts can be fiendishly complicated. And, as Navin found, even isolating a cell can be tough. For this reason, several research groups have started with cells that are easily separated, such as sperm, or those that are likely

to have dramatic genomic differences, such as tumour cells.

But as the techniques are refined, scientists hope to work out ever more subtle differences between cells, such as the tiny genomic rearrangements that happen in many neurons and may serve a purpose in organizing information flow. One of the most startling facets of such research is not the differences between cells, but how tissues and organs manage to work coherently despite them. To get a handle on these issues, it makes sense to start at the smallest possible unit. “The cell is kind of the ultimate denomination of an organism,” Navin says.

### TUMOUR DIVERSITY

In his first effort with breast-cancer tumours<sup>1</sup>, Navin was able to sequence only about 10% of the DNA — not enough to see individual point mutations, but good enough to study larger segments that are commonly duplicated or deleted, called copy number variants.

The results suggested that the tumour was made up of three major populations of cells, which emerged from the root tumour population in leaps and spurts at different times during the tumour’s growth. “It suggested a model of evolution where, instead of having lots of gradual intermediates, we saw hundreds of chromosomal rearrangements that occurred probably in very short periods of evolutionary time,” he says.

Since moving to Texas, Navin has started a research group focused entirely on single-cell genomics. He has been improving his methods, and can now piece together up to 90% of a cell’s genome, he says, which allows him to study the mutations in individual cancer cells in much more detail.

The team has also looked at another type of breast cancer. Sequencing the tumour as a whole, the group found six mutations in cancer-associated genes. “It seemed like a very simple genome,” says Navin. But when his group sequenced four individual cells they identified “hundreds of additional mutations,

many of which were unique or 'private' to that individual cell". In addition to tumours, his group has studied a breast-cancer cell line that has been maintained in the laboratory. The cells in such lines might be expected to contain identical genomes, but in data that have not yet been published, the team found that about 1% of the mutations, involving 12,000–20,000 base pairs, differed from cell to cell and that these variations could not be detected when the cells were sequenced together. In both studies, many of the new-found mutations were in cancer genes, or in other regions expected to disrupt protein function.

"Nicholas's work was very nice," says Timour Baslan, a graduate student studying cancer genetics at Cold Spring Harbour Laboratory and a former colleague of Navin's in Michael Wigler's lab. "It was very informative, but it was extremely expensive" — \$1,000 or more per cell. Baslan and his colleagues are trying to bring down the cost. They add genetic barcodes — short, easily traceable strands of DNA — to a cell's DNA, which allows them to sequence cell genomes en masse then identify the sequences from individual cells. Use of the barcodes, in combination with improved bioinformatics approaches, brings the cost of a genome down to about \$60 per cell. At that price, more researchers can work with the number of cells necessary to dissect a tumour-cell population as it proliferates and evolves. "Now we're doing hundreds and we're thinking in the future we'll be able to do thousands," says Baslan.

Baslan's group is using the techniques to study which tumour cells are left behind after chemotherapy treatment and why those cells were resistant to the drugs. Such analysis could guide treatment, says Navin. "By sequencing a few single cells you can get an idea of the heterogeneity in a tumour before chemotherapy and that might affect your choice of which agent to use, or whether to subject the patient to chemotherapy at all," he says.

### EVERY SPERM IS SACRED

The tendency of sperm to swim alone makes the cells ideal for single-cell genomics. Adam Auton, a statistical geneticist at Albert Einstein College of Medicine in New York is using sperm to study recombination, the process that shuffles genes during the formation of germ cells and therefore influences which genes are inherited. "Recombination is one of the fundamental forces that shapes genetic diversity," he says. "In recent years we've learned that there is considerable variation in the recombination rate between different populations, between the sexes and even between individuals." But pinning down the rate in people once seemed impossible because it would have required finding individuals with hundreds of children and sequencing their genomes.

The ability to sequence single cells meant that researchers could take another approach. Working with a team at the Chinese sequencing powerhouse BGI, Auton sequenced nearly 200 sperm cells and was able to estimate the recombination rate for the man who had donated them. The work is not yet published, but Auton says that the group found an average of 24.5 recombination events per sperm cell, which is in line with estimates from indirect experiments<sup>2</sup>. Stephen Quake, a bio-engineer at Stanford University in California, has performed similar experiments in 100 sperm cells and identified several places in the genome in which recombination is more likely to occur. The location of these recombination 'hotspots' could help population biologists to map the position of genetic variants associated with disease.

Quake also sequenced half a dozen of those 100 sperm in greater depth, and was able to determine the rate at which new mutations arise: about 30 mutations per billion bases

per generation<sup>3</sup>, which is slightly higher than what others have found. "It's basically the population biology of a sperm sample," Quake says, and it will allow researchers to study meiosis and recombination in greater detail.

Perhaps the most intriguing potential use of single-cell sequencing lies in neuroscience. Alysson Muotri, a neuroscientist at the University of California, San Diego, would like to study how long interspersed nuclear elements (LINEs) — 'jumping' genes that can move around the genome — cause each neuron to differ from its neighbours. His group has compared the number of LINEs in human brain, heart and liver tissue, and found that brain tissue contains significantly more jumping genes than the others<sup>4</sup>. Each human neuron probably has between 80 and 300 unique insertions, he says, differences

## “EVERY TIME WE LOOK AT OUR DATA WE DISCOVER SOMETHING NEW.”

that could affect a person's susceptibility to neurological disorders<sup>5</sup>, or provide the brain with a reservoir of diversity with which to respond to challenges. He says that it would be useful to sequence individual neurons and work out what effect this heterogeneity is having on brain function and even on personality. "I think it's the next level of complexity," he says. "We look at the brain and we think about the tissue, but actually it seems like lots of tissues in one, because the cells are so heterogeneous. It's almost like every cell was there for a purpose."

To make good on these plans, however, the techniques will need to improve further. Ramunas Stepanauskas, director of the Single Cell Genomics Center at the Bigelow Laboratory for Ocean Sciences in East Boothbay, Maine, says that the processes involved will continue to be integrated and miniaturized, until eventually they will form an off-the-shelf technology that involves just the press of a button — although he cautions that this is still many years off. In the meantime, his centre offers single-cell sequencing for labs that lack the necessary equipment and expertise to do it themselves. The US Department of Energy's Joint Genome Institute in Walnut Creek, California, offers a similar service.

With its new funding for single-cell studies, the NIH is attempting to spur innovation in the field. So is Life Technologies, a sequencing company based in Carlsbad, California, which is offering a \$1 million prize to the first researchers who can sequence the whole genome and all the RNA in a single human cancer cell using the company's technology. The deadline for the challenge is the end of this year, and the company says that there has been an enthusiastic response.

Baslan, for one, is intent on breaking down tissues into their components. "Every time we look at our data we discover something new," he says. "There's just so much analysis to be done that it's quite daunting." ■

**Brian Owens** is the assistant news editor for *Nature* in London.

1. Navin, N. *et al. Nature* **472**, 90–94 (2011).
2. McVean, G. A. T. *et al. Science* **304**, 581–584 (2004).
3. Wang, J., Fan, H. C., Behr, B. & Quake S. R. *Cell* **150**, 402–412 (2012).
4. Coufal, N. G. *et al. Nature* **460**, 1127–1131 (2009).
5. Muotri, A. R. *et al. Nature* **468**, 443–446 (2010).