Studying the Individual Cell

March 01, 2010

Byline: Ciara Curtin
Newsletter: Genome Technology
Genome Technology - March 2010

A cell is just one individual in a population, perhaps making up a tissue or part of a bacterial colony living in the human gut or in the ocean. As an individual, it can be slightly different than its neighbors — and that can lead to some interesting findings, as researchers focusing on single cells have been discovering. "One [reason for studying single cells] is the heterogeneity of cells. They are never quite the same, no matter what you are talking about," says Mike White at the University of Liverpool.

When researchers study a population of cells, those slight differences among individual cells are averaged out. "The information you get from a population average does not represent the single cell," adds Kun Zhang from the University of California, San Diego.

Also, some cells, such as cancer cells in the bloodstream, are just rare and researchers are lucky to find a few of them; these samples don't have enough to them to be studied as a population.

Furthermore, other cells, particularly bacteria, do not do well in culture. "Some 99 percent of microbes in the ocean and in most other environments — including our own bodies — cannot be cultured. At the moment there is no way of generating pure cultures of the vast majority of existing microbial types and to use such cultures in genomic analyses," says Ramunas Stepanauskas from the Bigelow Laboratory for Ocean Sciences, adding that metagenomic approaches have not been very successful in assembling marine genomes.

In the upcoming pages, *GT* takes a look at the tools and technologies researchers are using to study single cells.

Imaging: Developing single-cell models

The signaling molecule NF-kB has long been studied, but the University of Liverpool's Mike White says much is left to learn about its behavior, particularly at the single-cell level.

NF-kB is an important and highly regulated signaling pathway — it has been implicated in pathogen and stress responses — but the current models of how and why it shuttles between the nucleus and cytoplasm in a cell are too simplistic. "One minute it was in the cytoplasm and the next these genes got switched on. But actually if you think about the numbers, it didn't completely make sense that it could be so simple," White says. "I'd argue that the dynamics have been relatively ignored very often."

As White and his lab imaged cells over time, labeling NF-kB to watch its movements, they noticed that it would oscillate in and out of the nucleus about every hour and a half when stimulated. That regular movement had not been noticed before, as the cells were out of sync with one another, and previous work looking at cell population could only detect the population average. "That was a real eye-opener from the point of view of why you'd want to look at single cells," White says.

Other groups have since found similar behavior in more signal transduction pathways. Steve Wiley's team at the Pacific Northwest National Laboratory reported in December — again using fluorescent imaging — that the growth regulator ERK oscillates in and out of the nucleus every 15 minutes. Another lab showed that p53 oscillates every five or six hours. "In each case, the cells are out of phase with each other so, as a result, there've been massive numbers of people studying these pathways for many years, [but] you couldn't really see it," White says.

As for NF-kB, White and his group developed a model of how this oscillation could occur. In this signaling system, NF-kB turns on its inhibitor, ikB, which is normally in the cytoplasm. So, ikB heads into the nucleus and picks up NF-kB, taking it out to the nucleus. Then, ikB is degraded and NF-kB heads back to the nucleus where it switches ikB back on. There's a lag, though, between the degradation and re-making of ikB. "That makes a feedback loop. Basically, in engineering systems, that leads to oscillations," he says.

But White is still working out just what these oscillations are for. "The obvious hypothesis is the idea of using it as a timing system to control processes," he says. Calcium, he notes, has been known to oscillate; it's also known that the frequency of the calcium oscillations control downstream processes. Last April, White and his team published a paper in Science showing that when they manipulated the timing of the NF-kB oscillations, downstream gene expression was altered.

White's particular aim is to understand how transcription is affected by these dynamics and how the NF-kB pathway interacts with other oscillating pathways to influence cell fate decisions. He is developing better constructs to watch the process as well as moving into primary cells rather than cell lines.

A challenge, however, is putting all the information coming out of single cell studies — linking imaging with single cells manipulated through microfluidics with transcriptomic and proteomic data — together in a coherent way. "Linking these data sets together, thinking about how to do that, is a real challenge," White says.

Sequencing: Multiple displacement amplification

Small amounts of DNA such as that from a single cell may be amplified through a process other than PCR called multiple displacement amplification. This method, pioneered by the J. Craig Venter Institute's Roger Lasken, then at Molecular Staging, used f29 polymerase along with random primers. This polymerase has extreme processivity and does not drop the template strand for long. Amplicons average 12 kilobases in length and can get up to more than 100 kb long. A drawback, however, is that chimeras may form during the amplification process. The Bigelow's Ramunas Stepanauskas estimates that there is, on average, one chimerical rearrangement for every 20kb of amplicon.

But since Stepanauskas works with marine prokaryotes, those rearrangements have not been much of an issue. By combining MDA and fluorescence-activated cell sorting, Stepanauskas and his colleagues sequenced two uncultured marine microorganisms from the Gulf of Maine.

"Genomics that relies on classical microbiology, which is based on pure cultures, is not applicable to most microbes that have significant ecological roles," he says. Instead, they use a flow cytometer to separate individual cells, which they then break open to amplify the DNA using the MDA approach. Once amplified, that DNA may be screened by PCR or sequenced. "We now have several publications demonstrating that it is feasible and that we are gaining information on microbes that actually are of ecological significance," Stepanauskas says.

One of those publications was the sequencing of the two microbes. They were proteorhodopsin-containing flavobacteria and are closely related to bacteria from marine and Antarctic locations that have been PCR-cloned; they are also distantly related to cultured bacterial strains.

Stepanauskas and his team are now about a year into a new project to create a library of single-cell amplified genomes from the mesopelagic zone, which extends from about 200 meters to 1,000 meters below sea level. Once they have those genomes, they will be using PCR to screen for genes that are of interest from an ocean ecology perspective. "We will use that information to choose a handful of single amplified genomes, or SAGs, for whole genome sequencing," he says. "We will try to get complete genomes, although we may end up with partial assemblies."

His work isn't limited to the ocean. Another project is looking at freshwater bacteria that contain photorhodopsins or bacterial chlorophylls. "That's two different photometabolic systems which allow some of the microbes to gain energy from solar light, and it's not photosynthesis. It doesn't allow them to fix carbon, but it gives them a little extra kick from the energy that they capture," Stepanauskas says. They'll be searching through the single cell sequencing to find microbes with those metabolic genes.

There is, however, another drawback with the MDA method, namely that it doesn't always give even amplification. "When you start with a single copy of the genome and apply MDA reaction on it, the stochastic nature of reaction priming becomes important. Some genome regions may start amplifying earlier that others, resulting in more copies produced from that region than others," Stepanauskas says. "That means that when the MDA products are shotgun sequenced, there is a lot of variation in the coverage depth. So you are going to have to sequence more to cover the underrepresented regions and may still end up with an incomplete assembly."

Sequencing: Polymerase cloning

Another approach to amplifying small amounts of genomic information is polymerase cloning, which came out of George Church's lab a few years ago. "Ploning" is an optimized version of MDA, says Kun Zhang, who was a postdoc in Church's lab. This approach used real-time isothermal amplification to form the plones. Zhang and Church published a proof-of-concept paper in Nature Biotechnology in which they amplified two E. coli genomes and cyanobacteria genomes. They then sequenced those genomes using a Sanger approach. "We published a paper showing this is possible. … In that work we used cyanobacteria from ocean, called *Prochlorococcus*," Zhang says. "We amplified two genomes and we sequenced two genomes. That was part of a proof of concept, that it was possible."

Now at the University of California, San Diego, Zhang is working to make ploning more efficient, scalable, and optimized for next-generation sequencing. "We have been trying to push the envelope," he says. Once the approach is more efficient, Zhang hopes that his team will be able to "process a lot of single cells and come up with good sequencing library for use in next-gen sequencing."

To that end, Zhang and his lab are developing a device to handle polymerase cloning. Initially, he says, everything was done manually. "To amplify a good cell, it takes many, many hours and because everything involves manual handling — human beings are error-prone. If you are going to handle hundreds of reactions, it's very difficult to do that consistently," he says.

While others are trying to use robotics to do the ploning reactions in 96-well plates or 384-well plates, Zhang is exploring microfabrication with some help from others in UCSD's engineering department. "We're trying to come up with a platform that we can amplify a least a few thousand single cells in parallel, with only tiny amounts of reagent and starting material," he says.

That's an ongoing effort, Zhang says, and it is part of the technology development effort for the Human Microbiome Project. It's a $1.8 million, three-year grant to develop a lab-on-a-chip that sequences single microbial cells. "We have had a lot of progress, but we haven't finished all the technology development yet," he says.

Once the single cell genomes are sequenced, there are a few challenges to assembling it all. As Stepanauskas pointed out, there is amplification bias. "Certain parts of the genome will have many, many copies, have very high sequencing depth — and the other part will have very low sequence depth, even missing," Zhang says. He adds that "it is quite challenging to assemble complete genome from single cells simply because the issue is amplification bias in this region. But it is feasible to assemble enough contigs to cover 98 or 99 percent of the genome. I think that should be very informative."

To get that coverage, Zhang says that the existing algorithms for assembly have to be tweaked so they take into consideration the facts that some regions are over-sampled and that there are variations in the sequencing coverage.

Copy number variation: Studying cancer in single cells

While some researchers such as Stepanauskas and Zhang are looking to bacterial communities to find single cells to analyze, others are looking within human cell populations.

Paul Dear at the MRC Laboratory of Molecular Biology says that single-cell analysis is going to explode — and test the paradigm that all cells in a body have identical genomes. "I think we are going to see a lot more cell-cell variation," he says. Right now, though, Dear is focusing on cancer as those individual cells have lots of genomic differences.

"Let's take an extreme case. Suppose you are looking for rare mutants in cell population or suppose you are looking at a cancer and you want to know how the copy number of certain genes has changed in cells. If you take a biopsy and do mass DNA extraction from that, you are getting a mixed population of cells and you will get a mixed signal as a result," he says. "Ideally what you would like to do is take 100 individual cells, one at time, and look at the copy number in those cells and then maybe you find there are completely different populations of cells in there or you see other variation between the cells."

In cancer samples, Dear is delving into the very early stages of the disease. There, he and his team are searching for copy number variations from early, very small biopsies. "In different types of cancers, cells tend to acquire extra copies of certain key genes that are involved in the development of that cancer," he says. "Initially, we were working with tens of cells and we are trying to get down to single cell copy-number analysis, which is quite tricky to do, but we are moving in that direction." They hope to be able to detect CNVs of oncogenes that can then be used to detect cancer in its earliest stages. In particular, they are studying lung and colon

cancer. There's a region on chromosome 3 that has long been known to be amplified in lung cancer, and they are trying to pinpoint the gene at the focus of that amplification.

In a related area, Dear and his lab are beginning to look at circulating tumor cells. People with cancer have a very small number of tumor cells circulating in their bloodstream. "The trick is to be able to recover those very, very rare cells from blood and do analyses on those," Dear says. "We're trying to develop tools to do that, recovery and the analysis of the cells."

In addition to single cell analysis, Dear's lab is also pursuing single-molecule analysis. The researchers are developing a single-molecule, microfluidic-based sequencing platform that will remove the bases one at a time so they can be detected. "We can literally just read ACAGTCC," he says. "The only way to do that in terms of the optics and the physics of it is in a microfluidic channel. It's conveniently small plumbing."

Proteomics: Technology development

The protein composition of cells has long been known to vary from cell to cell and over time, but tracking those changes on a proteomic level is still difficult. "We broadly don't know very much about the variation of protein copy number from cell-to-cell and how important it is. That's partly because the tools aren't available, so of course nobody has much information, but there are these hints from microscopy that this may be more significant than it was first thought," says Imperial College London's David Klug, who directs the Single Cell Analysis Project.

That project is a £5 million program that aims to develop tools to study the proteomics of single cells. "The primary goal is to develop two platform technologies — one of which is labeled and one of which is label-free — to study, to be able to count protein copy number, but also to get extra added-value information, such as information on phosphorylations and information on protein-protein interaction and maybe even small molecule-protein interaction," Klug says. "And drive that down toward single-cell sensitivity."

Using microfluidic chips, Klug and his colleagues separate cells into single chambers where they lyse the cell contents. The contents are then pulled down onto the chip surface and read out with single-molecule spectroscopy.

One of their label-free spectroscopy approaches is two-dimensional infrared spectroscopy — Klug and his colleagues published a proof-of-principle paper in the Proceedings of the National Academies of Sciences in late 2008 that showed the technology could be used to identify proteins, based on their vibrational signals.

Another project that is part of the program is to write the whole proteome onto a one-centimeter by two-centimeter chip. "We are trying to get the whole proteomes distributed on that chip in an orderly fashion," Klug says. To do that, the researchers used capillary zone electrophoresis to write protein tracks on a substrate. The protein may be focused down to a 100 micron spot, or even smaller, according to Klug. If the capillary is drawn across the substrate in a controlled way, he says that protein tracks that are 50 microns to 100 microns wide may be written, and focus the proteins into 100 micron pixels. "Now we can do that. We're trying to automate that and make it more reliable and faster and better so that eventually we can write a hundred of these tracks," he says, adding that they could get to a 10,000 pixel resolution.

One project that Klug describes as "slightly off the wall" is their work in spatially selecting membrane proteins. Their approach uses microemulsions trapped in lasers to select proteins from a patch of a cell membrane. The emulsion droplet is about a micron in diameter and is

brought up to the cell, while trapped in the laser. "If you do that in the right way, with the right emulsion, what happens is the membrane proteins will come off and go into the droplet," Klug says. "We're sure we can do the removal, now we have to try to learn a bit more about whether we are getting protein selectivity or whether we are taking all the proteins in that patch and make sure we can read out and identify what we have harvested."

So far, the Single Cell Analysis Project is only in the technology development stage, but Klug is hopeful that the first biological data will be coming out within the year. "Once that data starts coming, it should come out in a pretty steady trickle," he says.

The future

Single cell analysis is just kicking off. As the tools and technologies to obtain that level of resolution are refined, even more questions may be asked about cell variation. One such question, the University of Liverpool's White says, deals with stem cells and discovering why some cells can be induced to become pluripotent stem cells while others can't. "That opens a really interesting question to these sorts of studies to understand what happens in those sorts of cells, what are the probability steps that occur," he says.

Single cell analysis could also shed light on human development, adds UCSD's Zhang. "Every human being came from a single fertilized egg and that single fertilized egg divided into two and those two are very different and they divided into four — and those are slightly different — and then one of the fundamental questions is where does this asymmetry come from?" he asks. "I think that [in] single cell analysis, transcriptional and genomics analysis will play a very important role in helping us to understand the development of very complex structures in higher organisms."

Imperial College London's Klug adds that he hopes that new technology developments will allow them to study currently hidden phenomena. "We hope that when we can make the technology work reliably, we can uncover biology that's previously been simply unobserved," he says. "It may turn out that cell-to-cell variation is very boring, there's nothing going on at all, but obviously we believe otherwise."

-->