

Single-cell genome and transcriptome sequencing methods are generating a fresh wave of biological insights into development, cancer and neuroscience. Kelly Rae Chi reports.

One of several groups applying single-cell genome sequencing to IVF, Sunney Xie at Harvard University and his collaborators have tested their new whole-genome amplification methods on the first and second polar bodies, small cellular castoffs of the fertilized donor egg that reflect its chromosomal health. In a recent paper, Xie's team showed that in eight female donors, polar-body

ago, says Xie, adding that people desperate to have a baby free of a devastating genetic disorder have been e-mailing him. The study's first baby could be born within the year. "I didn't anticipate that [our technique] would be used so quickly for patients," he says.

Single-cell sequencing is no small feat. The amount of DNA or RNA in a single cell starts at a few picograms—not even close to the quantity that today’s sequencing machines demand. So scientists must amplify these molecules and do so in ways that minimize technical errors while surveying sequences as broadly and evenly as possible. Until recently, many researchers doubted that sequencing of single cells could be reliably conducted by any but a few experts.

Updated protocols for DNA and RNA amplification, especially those disseminated in the last two years, have given new



In 2013, single-cell sequencing methods made their way to the mainstream.

Xie and his collaborators on the paper, Fuchou Tang of Peking University and Jie Qiao of Peking University Third Hospital, have launched a clinical study of women undergoing IVF. The team will amplify and sequence whole genomes of the polar bodies of participants' embryos to see whether they are fit for transfer. Such a step toward the clinic seemed impossible only 2 years



Z. He

Sunney Xie may soon see his group's genome amplification method used for preimplantation genetic diagnosis.

users greater choice for their experiments. Industry has also contributed countless kits for amplifying genetic material from single cells, and readout technologies have lowered in price. Fluidigm introduced the first single-cell automated prep system for RNA-seq in 2013. All these advancements are lower-

ing barriers for beginners. "People have been wanting to do this for decades," says Rickard Sandberg of Karolinska Institutet in Sweden, referring to single-cell RNA sequencing. "It's just that technology is now allowing us to do it in a much cheaper and much better way than before. It's becoming really accessible for lots of labs."

At the heart of the approach is the question: why go to the single-cell level? The rationale is that the alternative—pooling cells by the thousands or millions—blurs potential insights into the heterogeneity of complex systems such as the brain, blood and immune system, or even their component cell types. "When you go to the level of the single cell, you lose the information in the total system," says James Eberwine of the University of Pennsylvania. "But if you can do multiple cells within that system, then you can build up that system, I think, in a more informative way."

Numerous fields in which bulk tissue approaches may be insufficient are beginning to benefit from the new tools. And not only are single-cell sequencing methods helping define heterogeneity among cells, they also are allowing a level of comparison that is expected by many to redefine what a cell type is.

Tempering some of the enthusiasm are myriad challenges inherent to the process, from the isolation of cells, to amplification of their genomes or transcriptomes, to making sense of the data. Cost is also a consideration—single cells typically need to be sampled at higher numbers than tissues do—leaving good reason to carefully select situations that justify going to the single-cell level. "Do we need to analyze single cells to meet the objective? If the answer is no, you shouldn't do single cells. It's hard,

expensive, and you start encountering a lot of variability," says Paul Blainey at the Broad Institute and MIT.

From a few molecules of RNA

Sequencing a cell's transcriptome hinges on the ability to amplify large amounts of the complementary DNA (cDNA) that is synthesized from RNA. Capturing small amounts of RNA as cDNA and amplifying the cDNA extensively are difficult to do evenly and efficiently.

In 1990, transcriptome analysis at the resolution of single cells was made possible by Norman Iscove's group, who amplified cDNAs exponentially using PCR. In the early 1990s, Eberwine and his colleagues came up with a technique that generated cDNA from single live neurons and performed linear amplification by transcribing RNA from the cDNA. With the advent of microarrays, scientists used both linear and exponential amplification strategies to identify differences in gene expression among single cells.

High-throughput RNA sequencing (RNA-seq) came onto the scene in 2008, and shortly after, researchers coupled it to such amplification techniques to get a more detailed look at single-cell transcriptomes. For a 2009 study, Tang, then working in M. Azim Surani's laboratory at the Gurdon Institute at the University of Cambridge, showed that it was possible to detect—from a single mouse blastomere—the expression of thousands more genes than had been revealed using microarrays (*Nat. Methods* 6, 377–382, 2009).

That same year, Cold Spring Harbor Laboratory hosted its first single-cell meeting, and fewer than 50 scientists—developers and early adopters—attended. "I remember everybody was trying to do RNA-seq and trying to figure out what they had, how to believe what was real and figure out reproducibility," says Mike McConnell, now at the University of Virginia.

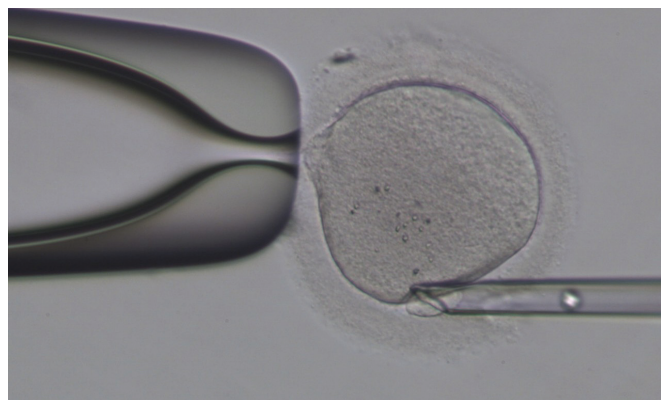
Methods development has since come a long way, researchers say. Now that there are

protocols and product offerings for single-cell sequencing, says Sten Linnarsson at Karolinska Institutet, "the phase of pure method development has culminated this year, and it's now become possible to actually use these methods at a pretty large scale to address biological questions." Rather than hundreds of cells, some groups are aiming to analyze tens of thousands.

For example, as part of the Single Cell Analysis Program supported by the US National Institutes of Health Common Fund, Kun Zhang's team will generate full transcriptomes from 10,000 cells in three areas of the human cortex. They will group the transcripts into cell types—perhaps redefining those cell types in the process—and map the transcripts back to cortical slices of the brain. Single-cell RNA-seq itself is no longer a barrier. "If you have a good cell, and you want to get a measure of the transcriptome, there is more than one option that can lead you to that goal," Zhang says. In general, however, extracting the neurons posthumously, minimizing RNA degradation and preserving some of the neuronal spatial information is challenging, and the group is evaluating several approaches, Zhang says.

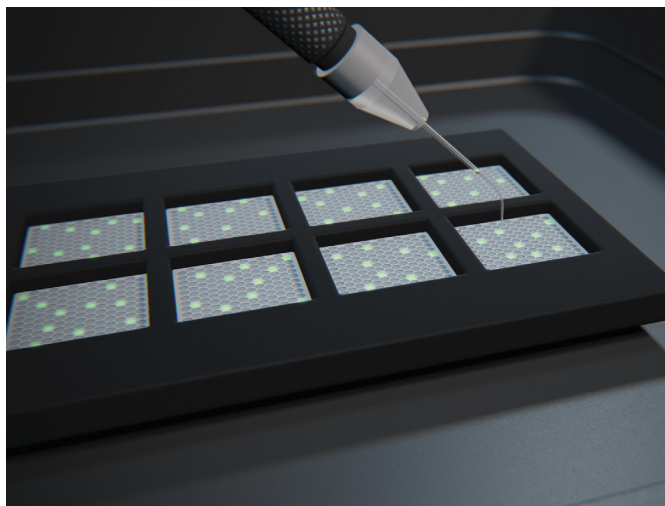
Amplifying the genome

Developing a way to amplify whole genomes of single cells took a bit longer because only one or two unique copies of DNA exist in the cell. The method lagged behind RNA amplification until 2005, when Roger Lasken's group became the first to amplify and sequence DNA from a single cell, that of an *Escherichia coli* bacterium, using the multiple displacement amplification (MDA) method that they had developed. That sparked a vigorous effort by microbiologists to generate reference genomes for diverse,



F. Tang, L. Yan

Sequencing the genomes of polar bodies associated with an *in vitro*-fertilized egg can help with preimplantation diagnostic screens.



K. Zhang

MIDAS is a scaled-up strategy for isolating, amplifying and sequencing genomes from single cells.

uncultivable bacterial species.

One of the most common strategies still used today, MDA is carried out by polymerases such as Phi29, which elongates random primers that have annealed throughout the genome. Each polymerase can dis-



M. Nemchuk/Broad Institute

Single-cell RNA-seq worked on the first try, says Aviv Regev.

Working in Michael Wigler's group at Cold Spring Harbor Laboratory, Nicholas Navin profiled—at 50-kilobase resolution—large deletions or duplications of DNA called copy-number variants (CNVs) across the genomes of breast tumor cells from two individuals (*Nature* **472**, 90–94, 2011).

One of the biggest challenges in single-cell sequencing of genomes is that some portions of a string of DNA get amplified more than others. In 2012, Xie's group described a new strategy called MALBAC, or multiple annealing and looping-based amplification cycles, that involves five cycles of MDA 'pre-amplification' during which newly amplified fragments form closed loops (*Science* **338**, 1622–1626, 2012). The loops prevent the fragments from being copied again, and

the amplification thus stays linear. Normal PCR follows the preamplification but is less prone to bias because of the more evenly amplified starting template. Using MALBAC, Xie's group obtained enough coverage to sequence 93% of the human genome and detect CNVs in a single cancer cell.

Scientists will soon be able to probe genomes more deeply in each cell, which will allow

them to see smaller deletions and duplications or even single-nucleotide variations. Amplifying the genome evenly still poses a challenge, but experts believe that scaling down reaction volumes will help reduce error.

For example, Zhang, at the University of California, San Diego, and his colleagues recently described MIDAS (micro-well displacement amplification system), a strategy for conducting MDA reactions in thousands of nanoliter-sized compartments etched onto a glass slide (*Nat. Biotechnol.*, **31**, 1126–1132, 2013). Researchers extract amplified fragments manually or with a robot and then sequence them. MIDAS allowed the group to detect single-copy-number changes in human neurons, with very little sequencing, at 1- to 2-megabase resolution.

Cells expressing their differences

At the Broad Institute, Aviv Regev, Joshua Levin and their colleagues were comparing RNA-seq methods for low-quantity and degraded bulk samples, when it occurred to them to try RNA-seq on a single cell. They decided to use a protocol called Smart-Seq on bone marrow-derived dendritic cells, postmitotic immune cells known to generate strong transcriptional responses to antigens.

That pilot study used 18 single cells and Regev allotted a week for the experiment. "You try out many things and they fail," but this worked on the first try, she says. Each cell uniformly expressed a set of 'house-keeping' genes, but the individual cells also revealed a surprise: genes important for immune regulation were expressed either

at high levels or not at all. Such bimodality had never before been seen in dendritic cells because differences among cells are averaged out when populations are sequenced. The results, published last June, suggested the presence of a cryptic cell type—a rare 'first responder' among what was thought of as a highly pure population (*Nature* **498**, 236–240, 2013). More broadly, the findings help reshape our understanding of these cells' identity, signaling and behavior.

Single-cell transcriptome sequencing is also helping researchers study gene expression and regulation in early development, and in far greater detail than what was previously possible for such rare samples. For a study published last August, Guoping Fan from the University of California, Los Angeles, and his collaborators in China sequenced transcriptomes from 33 single cells in multiple stages of development, identifying the order in which clusters of genes are expressed through the initial stages of development and how the timing of gene expression differs between early human and mouse embryonic development (*Nature* **500**, 593–597, 2013).

Meanwhile, Tang's group carefully dis-



Univ. Texas MD Anderson Cancer Center

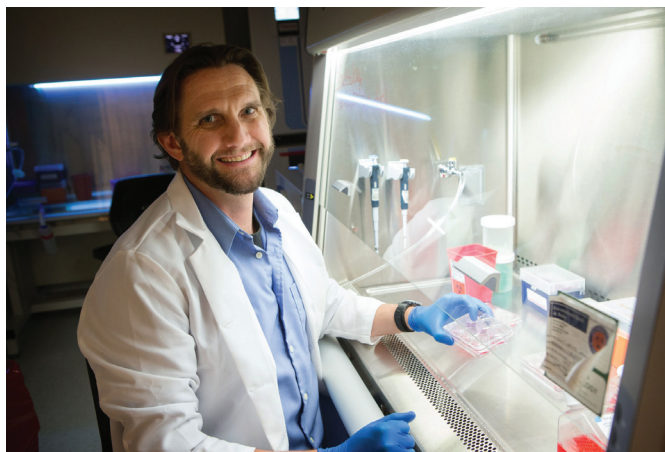
Single-cell sequencing is a powerful tool for understanding genomic variation in cancer cells, says Nicholas Navin.

sociated each cell of several early human embryos and sequenced their transcriptomes individually. "It's quite stressful. It's so important and a rare sample," he says. But the pressure paid off: they discovered more than 2,700 new long noncoding RNAs in the embryos that may play roles in early gene regula-

tion (*Nat. Struct. Mol. Biol.* **20**, 1131–1139, 2013). Before this, all single-cell RNA-seq work had analyzed known genes or, at most, novel alternative splicing isoforms of known genes, Tang says.

The cellular patchwork of cancer

From prognostics to disease monitoring, cancer research stands to benefit enormously from single-cell sequencing approaches. Cancer cells often undergo high mutation rates, and tumors tend to be heterogeneous. Identifying which subsets of cells, called clones, are present and evolve into metas-



Mike McConnell found single neurons in the human brain with large DNA deletions or duplications.

D. Addison

tases or respond in a certain way to chemotherapy is critical to understanding and fighting the disease. In particular, circulating tumor cells (CTCs)—which break off from a tumor and seed a cancer's metastasis—are those rare cells whose genomes or transcriptomes might offer clues for diagnosis, monitoring or treatment.

In Navin's 2011 *Nature* study, for example, profiling the genomes of single cells for CNVs revealed a punctuated model of tumor evolution: bursts of genomic instability following a stable expansion of tumor mass. "That was surprising, because people believed ... mutations gradually accumulate over time," says Navin, now at the University of Texas MD Anderson Cancer Center. "It showed how powerful these single-cell tools could be for understanding at least copy-number alterations in human cancers." He and his collaborators have continued to study copy-number evolution in triple-negative breast cancers—a heterogeneous and aggressive group of cancers—and also hope to better understand metastasis.

Besides Navin's, several other groups are applying single-cell sequencing approaches to cancer. For example, Xie, working with Fan Bai at Peking University and Jie Wang at Harvard, found a shared pattern of CNVs among CTCs of people with one subtype of lung cancer but not another (*Proc. Natl. Acad. Sci. USA*, doi:10.1073/pnas.1320659110, 9 December 2013). This recent finding offers potential for early diagnosis, Xie says.

Transcriptional differences may also hold the key to understanding cancer progression. Sandberg's group used their Smart-Seq method to sequence RNA from a single CTC as a proof of concept for their meth-

ods. Using their new version, Smart-Seq2, they can look at many more cells at a fraction of the cost. The technical noise plaguing CTC studies will greatly improve with more cells. "We are really hoping to make a more systematic effort, to better understand the heterogeneity in CTCs and to better understand their gene expression programs when

they go into circulation," he says.

More elusive still than the genome or transcriptome is the epigenome, the chemical marks on the genome that guide gene expression. Although current techniques have proven insufficient for the single cell—the traditional methods for detecting epigenetic marks on DNA also tend to degrade it—researchers are eager to see what the epigenome might reveal about cancer.

Tang's group has developed a tube-based method to reveal genome-wide DNA methylation in a single cell (*Genome Res.* 23, 2126–2135, 2013). "[Epigenomes] really need the single-cell approach" for researchers to see how tumor cells are different from their neighbors, whether that's through methylation or other mechanisms, he says. Wolf Reik's team at the Wellcome Trust Sanger Institute has taken methylome analysis to 50–100 cells, and, he says, "We are very interested in pushing the boundaries further."

Uncharted territory in the brain

Neurons are among the newest cell types to receive the single-cell treatment, and scientists aren't exactly sure what to expect from them. Experimental support for the idea that neurons harbor diverse genomes came about only recently. Even with these results, the diversity is baffling. In 2001, Jerold Chun's group, then at the University of California, San Diego, found aneuploidy in the mouse brain (and in 2005, in human neurons). "No one knew what to do with it," says McConnell, a graduate student in Chun's lab at the time. "The tip of the iceberg was how I saw it. If there's aneuploidy, there's got to be a lot more changes in these genes or in these genomes."

In the meantime, scientists demonstrated that there are 80–100 potentially active L1s—bits of DNA that copy and paste themselves throughout genomes—in every human genome and that L1s are active in neurons. Those studies and others delivered potential suspects in genomic diversity, but how extensive the variation is remains unclear today.

"We are just beginning to understand the molecular diversity of cells in the brain," says Thomas Insel, director of the US National Institute of Mental Health. "Single-cell methods will be critical, not only for defining the taxonomy of neurons and glia but for revealing the effects of experience or development on profiles of expression within a brain region."

Researchers are unlocking single-cell genomic variation in a few different ways. Christopher Walsh's team at Harvard Medical School scanned the genomes of 300 single human neurons isolated posthumously for L1 insertions (*Cell* 151, 483–496, 2012). They found few, a result suggesting that L1 is not a major player in genomic diversity, at least in the cortex and caudate nucleus.

In 2013, other groups scanned entire genomes of single human neurons. For a study published in November, researchers



Babraham Institute/Wellcome Trust Sanger Institute

Wolf Reik hopes to see epigenetic methods reach single-cell resolution.

took 110 frontal cortex cells from three healthy human brains, sequenced their genomes and found a surprisingly high proportion of neurons with large CNVs (*Science* 342, 632–637, 2013). Neurons derived from the skin cells of healthy people also had more

CNVs than did the skin cells themselves; these findings suggest that human neurons derived from induced pluripotent stem cells might work well for studying the functional implications of cell heterogeneity.

Indeed, neuroscientists are still wrapping their heads around what somatic variation could mean. It might make the brain more robust to perturbation, says geneticist Ira Hall at the University of Virginia, a corresponding author on the *Science* study. On the other hand, genomic mosaicism could affect risk for cancer or other diseases, he adds. To know whether some brain regions are more affected than others, or how much

variability there is from one individual's brain to another's, researchers will have to look at many more cells. "I don't know if we know yet how many," adds McConnell, who led the study.

Beyond proof of concept

Although the single-cell field is abuzz with biological questions, methods development is by no means over. Researchers must sort out biological variation from technical noise in their own systems. Single-cell RNA and DNA sequencing technologies are not yet robust, notes Joakim Lundeberg of the KTH Royal Institute of Technology in Sweden, whose group has developed a preliminary approach to sequencing RNA in the tissue context. "We will need many more single cells to be analyzed within a single experiment to address biological and stoichiometric noise or at least to achieve a better understanding of cell-to-cell variation within tissue," he adds.

Because of the challenges faced on many fronts, including cell isolation and computation, and in different fields of application,

Blainey expects more waves of technological advancements in the next few years.

For newcomers, the choice of which transcriptome-sequencing method to use might be daunting. The answer likely depends on one's research goals—whether it's analyzing many cells, looking for transcript isoforms or detecting low-abundance RNA. But "there's more than one way to skin a cat, which is good," Quake says. In October, his group showed that single-cell qPCR and RNA-seq perform similarly when sample prep is done in nanoliter-scale reaction volumes—in this case, with Fluidigm's C1 system (*Nat. Methods* **11**, 41–46, 2014). "That's a really good thing in terms of trust in the measurement," he says.

The selection of genome amplification strategies will improve, too, as new platforms become commercially available and single-cell researchers introduce their own versions of protocols. On the other hand, because everyone does amplification differently, it's hard to directly compare strategies across different studies. "In our hands

MALBAC is much better than MDA, but it depends on how you do MDA," Xie says. With further development both methods will be obsolete, he says, and he plans to be part of that effort. "By no means is MALBAC the end game. We're trying to do better."

Meanwhile, research in cancer, neuroscience, microbiology, drug development and a host of other fields will continue to reap the rewards as new and improved methods impart the ultimate, cellular resolution of transcripts and genomes. The approaches will attract newcomers such as Reik, who is well established in the epigenetics field. Reik attended his first single-cell meeting in 2013. "I haven't been part of the single-cell community before," he says. "I thought it was hugely exciting to see the field galvanized by these technologies. It is primarily still technology driven, and that's great, but very quickly now we'll get to some super exciting biology questions."

Kelly Rae Chi is a freelance writer based in Cary, North Carolina, USA.