Are you a lumper or a splitter? For centuries, biologists have pondered this fundamental epistemological dichotomy: Should they pursue a holistic study of a vast swath of life, or take the reductionist path and dissect it into its most fundamental components?

Molecular biologists have long considered themselves reductionists, but with a frustrating limitation: While they can sequence DNA and RNA with astonishing speed and accuracy, the resulting data represent populations of cells, not individual genomes. A tissue slice, a tumor biopsy, or a sample of a bacterial culture yields a sequence representing the average of all of the cells within it, even though researchers know there can be tremendous variation between those cells.

In the past few years, investigators and equipment makers have finally begun to break that barrier, creating new tools and techniques that can sequence individual cells. Single-cell RNA sequencing can now show the full spectrum of transcriptional activity across a sample, while single-cell DNA data are beginning to reveal subtle differences between cells previously considered identical. The new results are exciting, but some of the techniques, particularly for DNA sequencing, conceal serious pitfalls for newcomers.

Reading the transcript
Of the two techniques, single-cell RNA sequencing is the best supported. Several companies now sell reagents and equipment for isolating RNA from individual cells while keeping the thousands of resulting samples separate. Some also offer the technique as a service, taking customers’ vials of cells and returning complete single-cell RNA sequence data. Researchers who plan to do a lot of single-cell RNA analysis can even buy complete systems for it right off the shelf.

“Anyone can buy [our] system, and it’s intended to be a highly distributed, very accessible platform,” says Ben Hindson, chief scientific officer and cofounder of 10x Genomics in Pleasanton, California. The 10x system uses a microfluidic chip to partition a sample of cells into hundreds of thousands of miniscule droplets. The droplets then mix with gel beads, each of which carries a unique oligonucleotide “barcode.” 10x’s proprietary barcode synthesis and mixing system ensures that nearly every cell gets a separate barcode bead. Standard RNA sequencing techniques then yield millions of short sequences, including both the cells’ RNA and the barcodes. By linking the RNA sequences to the barcodes, the system’s open-source software can reconstruct the RNA sequence pool from each individual cell.

Collating that much data obviously requires sophisticated computing, but Hindson says the company has put a heavy emphasis on keeping its system user-friendly. “Our goal is to offer these turnkey solutions to less-sophisticated bioinformatics users,” he says, adding that “it just makes it easier for a wider audience.” 10x’s software includes not only the basic functions required to deconvolute the single-cell data, but also a suite of tools for analysis and visualization. Because the 10x software is open source, more experienced users can modify it or even export the raw data to their own programs.

Before buying a single-cell RNA system or sending samples to a service company, though, scientists should think carefully about their needs. They should ask themselves, “What kinds of samples are you looking to analyze, what kinds of sensitivity, and what kinds of information are you trying to get from those cells?” explains Hindson. Companies working on the technique are usually happy to discuss those topics with potential customers. Hindson adds that even seemingly simple issues, such as how the cells are isolated and stored before sequencing, can have a major impact on the quality of the data.

Upcoming features
Proteomics: Biomarkers—December 8 ■ Biotechnology: Biosensors and Microfluidics—March 16 ■ Molecular Biology: PCR Anniversary—May 11
Pick a cell, any cell

Careful sample handling is an essential requirement of many single-cell isolation strategies, a family of methods that has grown exponentially in recent years. It’s especially important to choose a good technique when analyzing DNA instead of RNA.

“There are so many options out there, and each of the options has its own advantages and its own disadvantages. It’s very important for researchers to decide depending on their application what type of isolation method they should use,” says Mary Langsdorff, senior market manager for life sciences at Qiagen in Hilden, Germany.

For investigators who know what their target cells look like, it might be best to pick them out manually. To do that as gently as possible, Qiagen’s QIAscout cell isolation system spreads cells across an array of magnetic rafts. Using a standard inverted microscope, researchers can then lift the desired cells and transfer them into individual sample tubes or microtiter wells.

With the cells isolated, the next step is to choose an analytical strategy. For DNA sequencing, that requires first amplifying each cell’s genome in order to construct a library for sequencing. While polymerase chain reaction (PCR) remains the standard DNA amplification technique for most experiments, even high-fidelity PCR techniques can introduce millions of errors when starting from a single genome. That’s why most single-cell DNA sequencing now relies on multiple displacement amplification (MDA).

MDA operates at a constant temperature, doesn’t require sequence-specific primers, and introduces very few errors. “Multiple displacement technology uses an enzyme which has a very high processivity and is a really high-fidelity enzyme, so we can ensure uniform coverage during the process ... and a very high sequence accuracy,” says Langsdorff.

After amplifying the DNA, researchers can then construct libraries and sequence them. Even with prices for genomic sequencing plummeting to below US$1,000 per human genome, sequencing the genomes of dozens or hundreds of individual cells can obviously get expensive, underscoring the importance of selecting and processing the cells carefully beforehand.

Nonetheless, it’s crucial to sequence multiple cells before trying to identify variants. That’s because single-cell DNA sequencing is typically shallow, often having a coverage of only 1x or less. “But then you compare all these ... libraries you’ve generated, and you only call a variant a variant when [it] appears in more than one library,” Langsdorff explains.

Despite its challenges as compared to single-cell RNA studies, single-cell DNA sequencing is clearly gaining popularity. Researchers are already using single-cell analysis to identify genetic variations within tumors and tissues, and Langsdorff says the technique is quickly establishing a foothold in microbiology as well.

Variation under unnatural selection

The need to compare multiple cells’ DNA sequences to distinguish genuine genetic variants from artifacts limits some of the potential of single-cell sequencing. It’s an especially serious problem for researchers who want to identify single-nucleotide variants. Recent reports have found that conventional cell isolation and genome amplification strategies, even using MDA, can yield as many as a million false positive single-nucleotide variants in a human genome, swamping the true positives (J).

Most of those errors stem from the cell-lysis protocols scientists have been using. “When you prepare the cell lysate and [perform] DNA denaturation, it introduces an artifact called cytosine deamination that’s due to the elevated temperature [step of the protocol],” says Joyce Peng, global marketing director for Novogene in Chula Vista, California. Recently, however, researchers at SingulOmics in New York, New York, developed a new technique that avoids heating the cells and their DNA. Combining this method with high-fidelity MDA amplification reduces the false positive rate by two orders of magnitude.

SingulOmics and Novogene now work together to offer a complete single-cell genomics workflow as a service. Researchers can send their cells to SingulOmics for preparation and genome amplification, and then have the amplified DNA sent to Novogene for sequencing. SingulOmics can also provide data analysis services once the sequencing is done. The two halves of the process can also be separated. Investigators who prefer to do their own sequencing and analysis can omit Novogene from the process, and simply receive the amplified DNA from SingulOmics. Those who are comfortable isolating cells and only need sequencing can go directly to Novogene. “There are two kinds of people; one is following our instructions to separate the cells themselves, and the other [lets us] separate the cells for them, so they send us a cell suspension,” says Novogene’s Peng.

Hiring contractors to handle the entire process is especially useful for researchers new to single-cell sequencing. That’s because service providers “made a lot of mistakes and [spent] a lot of
money just to perfect these techniques,” says Peng, adding that scientists should “consider a service provider to save time.”

Newcomers to single-cell DNA work may also need help navigating the complex and evolving landscape of amplification techniques. “There have been a bunch of whole-genome amplification protocols, but [some] people don’t quite understand which protocol would be best for their study,” says Zonghui Peng, field applications scientist for BGI in Cambridge, Massachusetts.

While MDA has become the preferred amplification method for projects seeking single-nucleotide variations, labs looking for genome copy number variations often prefer a method called “multiple annealing and looping-based amplification cycles” (MALBAC). Like PCR, MDA allows secondary copies of the genome to serve as templates for further amplification, but MALBAC amplifies only the original genome’s DNA. That makes it ideal for accurately counting the number of copies of a gene in each cell. The diversity of genome amplification options underscores the importance of having a clear experimental question in mind before starting a single-cell sequencing project.

Another reason for careful planning is cost. “Whole-genome, single-cell [sequencing] costs about the same as bulk-cell sequencing,” says BGI’s Peng. Because investigators inevitably need to sequence dozens or hundreds of cells in a sample, mistakes or poorly designed experiments can quickly devour vast quantities of funding.

Like SingulOmics and Novogene, BGI offers a range of single-cell genomics services. “BGI can handle both types of samples, either isolated single cells or bulk cells,” says Zonghui Peng. Customers can also have the company provide control samples from another set of droplets containing oligonucleotide barcodes. Though these barcodes are synthesized differently from 10x Genomics’ RNA-sequencing barcodes, their purpose is similar: to link each cell’s amplified genome pieces to a unique identifying sequence. Finally, the team sequences the barcoded genomes and analyzes the data.

Though the results of the experiment allowed Abate’s lab to distinguish individual microbial species, identify antibiotic resistance genes, and find virulence factors, he sees substantial room for improvement. “We could only get about 1% coverage per genome, and the distribution was pretty ugly; many cells had far more than 1%, [and] many cells had far less,” he says. Abate’s researchers are now working to improve those statistics, and his company, Mission Bio, in South San Francisco, California, is developing commercial versions of the microfluidic chips to give other scientists access to the technique.

As costs come down and methods become more standardized, experts in the field expect single-cell sequencing to become the new model for both transcriptomic and genomic studies. Instead of choosing between holistic analysis of bulk samples and reductionist glimpses of individual components, scientists will be able to combine both approaches to yield comprehensive maps of entire biological systems. “It’s a truly transformative technology that’s going to have impacts that are just absolutely impossible to understand until well after they’ve occurred,” says Abate.

References