
By Jeffrey M. Perkel

Surpassing the Law of Averages

How to expose the behaviors of genes, RNA, proteins, and metabolites in single cells.



By necessity or convenience, almost everything we know about biochemistry and molecular biology derives from bulk behavior: From gene regulation to Michaelis-Menten kinetics, we understand biology in terms of what the “average” cell in a population does.

But, as Jonathan Weissman of the University of California, San Francisco, points out, “A lot gets lost in the average.” For instance, “The census might say, the average family has 1.9 children, but no family has 1.9 children.” Similarly, though most bacterial cells in a culture may behave in one way—growing preferentially on glucose, for instance—a small number may exist in a different state, the better to protect the population against the vagaries of future environmental conditions.

“Single-cell genomics appears to be the most straightforward, and at the moment the only way we can assemble the genomes of the uncultured microbes from complex communities.”

Over the past few years, researchers have been eschewing bulk studies and pushing technological limitations to bring their studies of genomics, genetics, RNA transcription and translation, proteins, and metabolites down to this single cell level. *The Scientist* spoke with five such scientists to see how they are breaking the law of averages. Here is what they said:

Improving IVF



RESEARCHER: Joris Vermeesch, Professor of Molecular Cytogenetics and Human Genome Research, Center for Human Genetics, Leuven, Belgium

Project: Developing preimplantation genetic diagnostics to improve in vitro fertilization (IVF) efficiency

Problem: Standard diagnostic methods, such as fluorescence in situ hybridization, analyze just one or a few loci per cell; genome-scale methods require too much DNA. Vermeesch wanted a way to look at the entire genome in individual blastomeres from eight-cell human embryos.

Solution: The emerging genome-scale method for detecting chromosomal abnormalities, array comparative genomic hybridization (aCGH), typically requires DNA from several hundred cells. Vermeesch’s team circumvented that problem by amplifying each blastomere’s DNA before hybridizing it to both a home-made bacterial artificial chromosome array and a commercial single nucleotide polymorphism array (*Nat Med*, 15:577-83, 2009). The two arrays served to validate each other, and concordance between their datasets was high, Vermeesch says. Of the 23 embryos the team analyzed, 21 contained at least one cell with a chromosomal abnormality, whether at the whole chromosome level or smaller genomic segments. That means Vermeesch’s stated purpose—using aCGH on a single blastomere to

improve the “baby take-home rate” of IVF by identifying aneuploidies—cannot work, because no single blastomere can reliably report the chromosomal status of the remainder of the embryo. “We killed our own application,” he says. Yet the study also opens up the possibility of doing preimplantation single-cell SNP screening—a single-cell form of whole-genome genetic and mutational analysis, if you will. “All the other info you can do with 23andMe if you send in a blood sample, I can do on a single cell,” he says.

Considerations: Vermeesch says his technique has broad developmental biology application, from embryogenesis to tumorigenesis. But there are challenges, he says, including both amplification bias and the cell cycle itself. Early-stage embryos divide rapidly, and cells caught in S-phase

Protein Profusion

will have different copy numbers of DNA than cells in either G1 or G2.



RESEARCHER: Jonathan Weissman, Professor of Cellular and Molecular Pharmacology, University of California, San Francisco

Project: Studying cell-to-cell variability in protein abundance on a proteomic scale

Problem: Because many proteins are present at low levels, researchers assume there’s “lots of cell-to-cell variability,” Weissman says. But little hard data exists, so he wanted a way to measure protein abundance per cell.

Solution: Standard proteomics methods wouldn’t help, because they require too much material. So, Weissman’s team opted for flow cytometry, using fluorescence as a proxy for protein abundance. Starting with a library of 4,159 yeast strains, each containing green fluorescent protein fused to the 3’ end of one particular gene, the group measured the fluorescence intensity in each of about 50,000 cells per strain using a high-throughput flow cytometer (*Nature*, 441:840-46, 2006). The team then calculated the range of each protein’s abundance—that is, cell-to-cell variability—a strategy that helped them uncover a kind of logic in normally overlooked biological “noise.” Though cells can neither know

what their neighbors are doing nor predict future conditions, they can, Weissman explains, “hedge [their] bets, and have a fraction, say 10% of cells, expressing stress proteins that may slow down the growth but give it a better chance to survive if conditions change.” And that’s what the team found: Though housekeeping genes varied very little from cell to cell, “proteins that varied from one condition to another were found, under any given condition, to have much more variability.”

Considerations: Just about anyone with a flow cytometer can use this approach, Weissman says. The instrument should be fast (Weissman’s system processed 50,000 cells in 7 seconds), high-throughput (e.g., front-end robotics and a sipper attachment to handle multiwell plates), and multicolor (to read multiply labeled cells). The biggest obstacle, he says, is having a good library of GFP-tagged strains. “Once you have that, you can do many high-precision single-cell measurements rapidly, on the cell type you care about, under different conditions, using different drugs, and in different genetic backgrounds.”

Cracking Microbial Ecology



RESEARCHER: Ramunas Stepanauskas, Senior Research Scientist, Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, Maine

Project: Sequencing the genomes of uncultivated ocean microbes

Problem: Traditional metagenomic strategies, which shotgun-sequence the DNA in a microbial community, typically can neither select for interesting organisms nor produce completely assembled genomes.

Solution: To assemble genomes from uncultured microbes, you must be able both to isolate the organisms, and harvest sufficient DNA for sequencing. Stepanauskas paired fluorescence-activated cell sorting with multiple displacement amplification—a whole-genome amplification method—to do just that. His team then used PCR to determine whether those cells merited further analysis—for instance, based on their ribosomal RNA or protorhodopsin genes—before subjecting them to both Sanger and 454 paired-end sequencing. The result was the near-

complete assembly of two marine microbes from the Gulf of Maine—one 91% complete, the other, 78% (*PLoS ONE*, 4: e5299). “Single-cell genomics appears to be the most straightforward, and at the moment the only way we can assemble the genomes of the uncultured microbes from complex communities,” he says.

Considerations: Stepanauskas says his approach has implications for those studying microbial ecology, evolution, and bioprospecting. But first, some kinks must be worked out. One is amplification bias: the two genomes presented in this study were not quite complete, meaning some segments did not amplify well. A more troubling problem is cell lysis, especially of Archaea and gram-positive bacteria. “There are major limitations on what kinds of methods you can use to lyse the cells for single-cell genomics,” he says. “That method has to be effective enough to open the cell, but at the same time, it has to be mild enough to maintain the integrity of the genome.” Finally, the published technique is too low-throughput to be practical; a facility at the Bigelow to automate this procedure should be online this fall.

RNA in Action



RESEARCHER: Xiaoliang Sunney Xie, Professor of Chemistry, Harvard University, Cambridge, Massachusetts

Project: Unraveling the molecular mechanisms of gene expression and regulation at the single-cell level

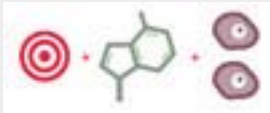
Problem: How to visualize single molecule events of gene expression and gene regulation in bacteria

Solution: Researchers have observed GFP in living cells, but not usually one at a time; instead, they see the aggregate fluorescence of tens to hundreds of molecules. But Xie wanted to watch the synthesis of individual fluorescent protein molecules. To do that, he developed a technique he calls “detection by localization,” exemplified in a trio of *Science* papers between 2006 and 2008. In one example (*Science*,

311:1600-3, 2006), Xie's team built a membrane-targeted yellow fluorescent protein (YFP), coupled it to the lactose operon, and shut it off. Like a leaky faucet, occasional YFP molecules emerged, each the product of a transcript caused by random release of the lac repressor from its operator. Normally, individual YFP molecules cannot be seen; their fluorescence is lost in the cell's autofluorescent background because their rapid diffusion dilutes the signal. But when localized to the membrane, where diffusion is slower, individual molecules emerge from the background and may be counted. Photobleaching, induced by continuous illumination, ensures that each molecule is counted only once. In this way, Xie was able to measure the frequency of "leaky" transcription, and the number of proteins produced per transcript. The result, he says, "is a quantitative description of how these life processes occur in a living cell in real time."

Considerations: According to Xie, his method offers "single-molecule sensitivity, high specificity, millisecond time resolution, [and] nanometer spatial precision." Sensitivity is possible due to the fluorescent protein's localization; specificity is conferred by the genetic encoding of fluorescent proteins; strobe photography (the trick that allows photographers to capture bullets in flight) provides the resolution; and modern CCD cameras' ability to determine the center of fluorescence intensity ensures spatial precision.

Metabolic Distinctions



RESEARCHER: Renato Zenobi, Professor of Analytical Chemistry, Swiss Federal Institute of Technology (ETH), Zurich

Project: Single-cell metabolomics: Identifying and quantifying metabolic differences between seemingly identical cells

Problem: Metabolite levels in individual cells can be vanishingly small, below the limits of detection of mass spectrometry, and masked by more abundant molecules.

Solution: Zenobi and his ETH collaborator, Andreas Hierlemann, are

working to develop an automated microfluidic system that will isolate a cell, lyse it, crudely separate its molecular constituents, and transfer those molecules to a MALDI target plate for mass spec analysis. “With the exception of the lysis, which doesn’t always work very well for every cell type, we have solved all of these problems,” he says. The lab demonstrated that by concentrating metabolites in a small spot on the target plate, it can detect certain metabolites (such as ATP, ADP, and acetyl-CoA) at levels comparable to those found in yeast (*Angew Chem*, 120:5462-65, 2008). More recently, his team measured ATP concentrations in individual single-cell protists called *Euglena*, and found that the concentrations can vary wildly from cell to cell.

Considerations: “Sensitivity,” says Zenobi, “is the issue.” It’s simple geometry: The intracellular volume of a single *E. coli* cell is about 1000 times smaller than that of yeast. That means that bacterial metabolites exist at below the detection limit of many mass specs. At the moment, Zenobi is concentrating on yeast and larger *Euglena*. At the same time, he is working to improve sensitivity by, for instance, reducing chemical background and eliminating suppression effects that can reduce mass spec efficiency.