Single molecules meet systems biology

Application of single-molecule methods to assay genome-wide mRNA and protein levels in single bacteria provides a systems-level view of their relationship.

In biology, working with large samples of cells helps ensure a good signal, but this comes with a drawback: the measurement is an average that will mask differences in individual cells. A prominent example of this is single persistent bacteria that resist antibiotic killing of a larger genetically identical population. This antibiotic resistance must come from differences in the repertoire of proteins expressed by these cells but population-based methods cannot be used to examine expression differences between single bacteria in a population.

Sunney Xie at Harvard University has spent years working with methods to understand biological phenomena in single molecules and cells, from measuring the activity of enzyme pairs to visualizing the activity of a single gene in a single bacterium. “The process of transcription and translation is a single-molecule problem,” says Xie. “There is only one copy of the gene, and the processes of gene regulation and transcription initiation are much slower than subsequent steps, so they are rate-limiting, and ‘stochasticity’ comes into play.” This stochastic nature can result in phenotype switching, a possible explanation for the existence of persisters.

Xie previously visualized the expression of individual proteins produced from a single gene tagged with yfp in a growing bacterium in real time, but he wanted to apply these techniques across the genome. This requires a library of bacterial strains in which each gene is tagged at its 3’ end with yfp. Xie says, “my graduate student Paul Choi came up with a great idea of swapping yfp into an affinity-tagged library, which was previously constructed by Andrew Emili’s group at University of Toronto, with only one primer using λ-Red-mediated recombination. Huiyi Chen worked with him to do this cheaply and quickly.”

Once Xie and his group had their library, they needed a high-throughput way of quantifying YFP expression in each strain. For this they turned to microfluidics. Postdoc Yuichi Taniguchi designed and built a microfluidic device that could image 96 strains at a rate of 4,000 cells in 25 s per strain. By careful calibration, normalization for cell size and deconvolution to remove autofluorescence, they obtained accurate measurements of protein abundance. Xie says, “we have five decades of dynamic range from 0.1 to 10^4 copies per cell.”

The researchers describe a thought experiment that should allow one to do this. “Imagine that you grow cells in the condition you’re interested in, and in parallel you grow them in the same conditions except you now increase the fitness differences between cells by some constant multiple,” explains Kussell. This should now accentuate the fitness differences between the cells so that selection can act on them. “At the end of the day, if the population structure is different between these two experiments, then you know selection is in play under those conditions.” In other words, you disentangled the effect of selection from that of cells sensing and responding to environmental change.

The problem is, it is difficult if not impossible to actually translate this experiment into reality. How would one practically increase the fitness of all individuals in a population by a small constant factor? What Kussell and Leibler found, however, is that by measuring a property that they call the variance of the historical fitness of cells, the results of the above thought experiment can be inferred. This property can be measured by examining independent individual lineage histories of cells in a growing population and by simply recording the number of times a cell divides over time under different conditions of interest.
Because Xie wanted to investigate the relationship between transcription and translation system-wide in single molecules, he also needed a way to count the mRNA molecules encoding the YFP-tagged protein in each strain. Using fluorescence in situ hybridization, graduate student Gene-Wei Li could detect an individual yfp mRNA labeled with a single red fluorophore in a fixed cell.

While Taniguchi collected protein expression data in live cells for all 1,018 strains, Li used the library and RNA counting to simultaneously measure mRNA and protein levels in single fixed cells for 137 of the more highly expressing strains.

“This is the first time the proteome has been characterized at single-molecule sensitivity in any organism,” remarks Xie. They found that half of all the proteins are present at fewer than ten molecules per cell and 20% of proteins are expressed at under one copy per cell. Xie admits that some were leakages, but others were functioning proteins. They found that the protein expression is represented by a gamma distribution and, perhaps surprisingly, there was no correlation between the mRNA copy number and protein copy number in the same cell. This is explained by the stochastic nature of transcription and the fact that most of the proteins had a lifetime longer than the cell cycle, whereas the mRNAs had a much shorter lifetime.

These results promise to be the first of many biological insights to come. Xie is making the library available for other people to use and is looking forward to applying single-molecule systems biology to questions such as bacterial persistence.

**Daniel Evancho**

**RESEARCH PAPERS**


Not only does the formulation of Kussell and Leibler provide a tool to study the effective strength of selection in a particular system, it also allows one to infer whether a cell or unicellular organism is using a stochastic switch to adapt to a particular environmental fluctuation. “For stochastic switches, there is a very pronounced peak in the historical fitness variance, which occurs at fluctuation periods that are longer than the generation time,” says Kussell. “So you can infer the existence and the type of an internal mechanism when you essentially don’t know what the organism is doing, and you don’t know what genes are important.” Undoubtedly, it will be of interest to detect that some were leakages, but others were functioning proteins. They found that the protein expression is represented by a gamma distribution and, perhaps surprisingly, there was no correlation between the mRNA copy number and protein copy number in the same cell. This is explained by the stochastic nature of transcription and the fact that most of the proteins had a lifetime longer than the cell cycle, whereas the mRNAs had a much shorter lifetime.

These results promise to be the first of many biological insights to come. Xie is making the library available for other people to use and is looking forward to applying single-molecule systems biology to questions such as bacterial persistence.

**Natalie de Souza**

**RESEARCH PAPERS**