

# Single-cell Analysis of the *Escherichia coli* Proteome with Single-Molecule Sensitivity

Yuichi Taniguchi<sup>1</sup>, Paul J. Choi<sup>1</sup>, Huiyi Chen<sup>2</sup>, Mohan Babu<sup>3</sup>, Andrew Emili<sup>3,4</sup>, X. Sunney Xie<sup>1</sup>

**Short Abstract** — The number fluctuation of proteins between cells yields phenotypic variability in cellular functions, which can be essential for biological processes. However, it has been difficult to obtain proteomic information with high enough sensitivity to determine number distributions in cells with single molecule precision. We developed large-scale single molecule detection (LSMD) system to challenge this problem. The LSMD system combines a 96-well parallel-integrated microfluidic chip and auto-scanning single molecule microscopy, allowing sequential and high-throughput measurements. Over 1,000 proteins were measured via a library of fluorescent protein-tagged *Escherichia coli* strains. The result illuminated a global relationship between cellular functions and phenotypic variability.

**Keywords** — Single-cell proteomics, single molecule detection, DNA library, microfluidics, systems biology.

## I. PURPOSE

THE objective is to characterize protein expression across the entire *Escherichia coli* proteome using sensitive single-cell imaging methods. We aim to characterize the system-wide heterogeneity of gene expression across cell populations, which cannot be measured by traditional ensemble methods such as DNA microarrays or western blots. Although single-cell methods have been applied in the past for other organisms using flow cytometry [1], they lacked the sensitivity to detect many low copy proteins such as transcription factors and regulators. Our novel approach integrates live-cell single-molecule microscopy [2,3] with a high-throughput microfluidic platform to reveal properties, functions, and origins of cell-to-cell variation in the *E. coli* proteome.

## II. RESULTS

We have constructed a library of chromosomal fluorescent protein fusions for over 1,000 ORFs from the *E. coli* genome by an efficient, low-cost conversion of an existing Sequential Peptide Affinity [SPA]-tag library [4]. The library has been measured by a new high-throughput imaging technique, large-scale single molecule detection (LSMD) system, which is a microfluidic platform for high-throughput fluorescence microscopy, coupled with

automated imaging analysis, enabling us to record the protein expression of over 100,000 cells per hour, which is sufficient to describe the statistics of about 100 different reporter strains. We have measured the distribution of protein expression across cell populations and determined the noise properties of each gene with single molecule sensitivity as necessary. In addition, we have imaged the localization of proteins to the membrane, cytoplasm, and DNA, which has been unrevealed by traditional methods.

To determine possible factors affecting the noise of specific genes, we correlate our protein expression data with biological markers and other global data sets. We find that a substantial fraction of the proteome is expressed at low copy numbers, in agreement with previous predictions [4], and these genes are subject to high values of noise. We also observe global properties of protein noise in *E. coli* and find differences in the scaling between noise and average expression for proteins present at low or high copy numbers. Comparing the expression profile between different growth conditions also reveals the associations between functional proteomics and the noise or expression of specific genes.

## III. CONCLUSION

A single molecule measurement platform for a large-scale single cell screening has been developed to examine the entire *E. coli* proteome. The data provides the first comprehensive proteomic resource of expression levels and noise with high sensitivity for the model organism *E. coli*. Also, we expect that LSMD will potentially be a valuable tool for a large-scale *in vivo* or *in vitro* screening of single molecule enzymatic activity for vast number of proteins with an improved statistics.

## REFERENCES

- [1] Newman JRS, et al. (2006) Single-cell proteomic analysis of *Saccharomyces cerevisiae* reveals the architecture of biological noise. *Nature* **441**, 840-846.
- [2] Yu J, et al. (2006) Probing gene expression in live cells, one protein at a time. *Science* **311**, 1600-1603.
- [3] Elf J, Li GW, Xie, XS (2007) Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316**, 1191-1194.
- [4] Butland G, et al. (2005) Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* **433**, 531-537.
- [5] Bon M, McGowan SJ, Cook PR. (2006) Many expressed genes in bacteria and yeast are transcribed only once per cell cycle. *FASEB J.* **20**, 1721-1723.

<sup>1</sup>Department of Chemistry and Chemical Biology, Harvard University.

<sup>2</sup>Department of Molecular and Cellular Biology, Harvard University.

<sup>3</sup>Banting and Best Department of Medical Research, Terrence Donnelly Centre for Cellular and Biomedical Research, University of Toronto

<sup>4</sup>Department of Molecular Genetics, University of Toronto

E-mail: xie@chemistry.harvard.edu