Engineering a Fluorescent Barcoding System for Highly Multiplexed, Single-cell Analysis of Bimolecular and Cellular Libraries

Stefanie Berges
Advisor: Dr. David W. Colby
Committee: Dr. Maciek R. Antoniewicz, Dr. Wilfred Chen, Dr. Edward R. Lyman

---

A primary goal of biologists is to characterize the dynamic and complex behaviors of biological systems. Large, robust data sets that examine many biological molecules in a variety of conditions are desirable to gain a more multifaceted view of the cell. In addition, single-cell analysis technologies are used to characterize cellular heterogeneity and reduce biological noise that exists within isogenic populations. Noise in gene and protein expression arises from the stochasticity of underlying biochemical reactions, and can confer phenotypic variation which may be advantageous in certain circumstances. However, established technologies for high-throughput, single-cell proteomic analysis have limited throughput.

Multiplexing methods, such as fluorescent barcoding, can dramatically decrease the number of samples and in turn enable collection of more robust data sets including many replicates, conditions, and proteins. Fluorescent barcoding is a powerful tool for identification of different cells within a heterogeneous mixture using a unique fluorescent identifier or 'barcode'. Fluorescent barcoding can potentially reduce the number of samples thousands of fold, thereby facilitating massively-parallel single-cell analysis of biomolecular and cellular libraries. Current fluorescent cell barcoding systems are composed of small numbers of barcodes (~10-100), and in some cases are single-use and have toxicity issues.

The work described here presents the creation of the largest fluorescent barcoding system to date consisting of over 980 unique, genetically-encoded barcodes. We made a library of plasmids encoding protein scaffolds that are composed of different lengths and combinations of epitope tags connected by flexible linkers. Cells expressing protein barcodes were identified by their distinct fluorescence upon immunolabeling. Multiplexing capability was greatly expanded by the discovery that barcodes with four distinct fluorescence intensities can be created by expression of different epitope tag repeat lengths. The effect of barcode expression on cellular growth, and the influence of different promoters and growth conditions on barcode expression was examined. A software package was developed to rapidly analyze barcode flow cytometry data, decreasing analysis time ~10-fold.

The multiplexing power of the fluorescent barcoding system was demonstrated in two applications. Barcode expression did not affect the binding affinity of an α-prion antibody for recombinant prion protein, suggesting barcodes can be used for multiplexed analysis of biomolecular libraries including high-throughput, quantitative protein-protein interaction studies. Barcodes were also used to simultaneously measure the dynamic response of endogenous yeast proteins in single-cells to environmental perturbations. Changes in protein abundance and variability as well as expression distributions were observed, suggesting cells may employ a bet-hedging mechanism to more quickly adapt to fluctuating environments. In addition, long epitope tag repeats facilitated immunodetection of endogenous, low abundance proteins in yeast by increasing the detection limit ~40-fold, potentially enabling analysis of > 1,600 low abundance proteins by flow cytometry.