

Turning single cells into microarrays by super-resolution barcoding

Long Cai

Abstract

In this review, we discuss a strategy to bring genomics and proteomics into single cells by super-resolution microscopy. The basis for this new approach are the following: given the 10 nm resolution of a super-resolution microscope and a typical cell with a size of $(10 \mu\text{m})^3$, individual cells contain effectively 10^9 super-resolution pixels or bits of information. Most eukaryotic cells have 10^4 genes and cellular abundances of 10–100 copies per transcript. Thus, under a super-resolution microscope, an individual cell has 1000 times more pixel volume or information capacities than is needed to encode all transcripts within that cell. Individual species of mRNA can be uniquely identified by labeling them each with a distinct combination of fluorophores by fluorescence *in situ* hybridization. With at least 15 fluorophores available in super-resolution, hundreds of genes in can be barcoded with a three-color barcode (${}^3C_{15} = 455$). These calculations suggest that by combining super-resolution microscopy and barcode labeling, single cells can be turned into informatics platforms denser than microarrays and that molecular species in individual cells can be profiled in a massively parallel fashion.

Keywords: super-resolution microscopy; systems biology; single cells; single-molecule FISH

INTRODUCTION

There had been two transformative technologies in modern systems biology: genomics, which allows all of genes and proteins in an organism to be monitored simultaneously, and single-cell biology, which follows a few specific genes in individual cells with high precision in their native micro-environments. Both techniques are powerful. The genomics technology employing microarrays [1] and next-generation sequencing [2–5] can probe mRNA abundances and DNA–protein interactions across the entire genome. At the same time, the single-cell approach with fluorescent proteins [6–11] and fluorescence *in situ* hybridization (FISH) [12, 13] are highly quantitative and can detect mRNAs and proteins in cells with single-molecule accuracy [14–16]. However, both approaches have complementary limitations: genomics averages over the heterogeneity and spatial complexity of a cell population, and single-cell techniques can only probe a

few genes at a time. Integrating genomics with single cell is the next major challenge in biology.

There have been significant efforts in scaling down high-throughput techniques down to the single-cell level. However, the main challenge is that single cells contain a small amount of material that can be analyzed. For example, nucleic acid contents of single cells need to be amplified in order to be sequenced. However, amplification may introduce biases and distorts the quantitation of molecular species in single cell. Digital PCR [17, 18] partially resolves this problem by spatially separating single molecules of cDNA converted from mRNA molecules into distinct wells and using the number of wells that light up to readout the copy number of mRNAs in the sample. Generalizations of this idea have been recently implemented [19–23] to improve the quantitation of DNA and RNA-seq, by ligating random barcodes to the cDNAs prior to amplification as a way of digitalizing quantification of

Corresponding author. Long Cai, Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena CA 91125, USA. Tel: 626-395-6570; E-mail: lcai@caltech.edu

Long Cai is an Assistant Professor of Chemistry at California Institute of Technology. He received his PhD in chemistry with X. Sunney Xie in 2006 and was a post-doc in the laboratory of Michael Elowitz from 2006 to 2010.

sequencing reads. This method may allow more quantitative RNA-seq from single cells. However, single cells still need to be isolated and extracted from tissues removing the intracellular and intercellular location of the RNAs.

MOTIVATION

Spatial separation underlies the basis of many biochemical and analytical techniques. Gel electrophoresis and affinity columns are routinely used to separate molecules based on their physical properties as well as their binding affinities. Microarray generalizes this in a high-throughput fashion compared to northern blots by spotting different oligonucleotides complementary to different genes on a dense spatial array. Spatial separation can also trade data space for improved accuracy of quantitation, as discussed previously with digital PCR and sequencing.

Resolving molecules natively in individual cells without separation becomes possible with the advent of super-resolution microscopy such as PALM [24], STORM [25], FPALM [26], SSIM [27] and STED [28], as many cellular components can be resolved down to nanometer accuracy. This boon in resolution has made significant impact in cell biology. We propose that super-resolution microscopy also hold high potential for single-cell systems biology: many molecular species can be inherently spatially separated within individual cells. With a typical cell of $(10\ \mu\text{m})^3$, a 3D-STORM microscope with a lateral resolution of 15 nm and an axial resolution of 50 nm can in principle resolve 10^8 such pixels in a cell. In comparison, there are only on the order of 10^6 mRNA molecules per cell [3, 4]. Thus, many messenger RNAs can be spatially resolved and an individual cell can, in essence, serve as a microarray under a super-resolution microscope (Figure 1).

While super-resolution microscope provides the optical space to resolve a large number of molecules in cells, each molecular species still need to be specifically labeled and uniquely identified. Pioneering work in single-molecule FISH (smFISH) by Singer [12] and Raj [13] using short synthetic oligonucleotide have shown that transcriptional active sites and single mRNAs in cells can be detected with high specificity and accuracy. This smFISH technology has been used to multiplex chromosomal loci and transcription active sites by barcoding with a combination of fluorophores [29–31]. We can borrow this approach to labeling single mRNAs. In the

STORM version of super-resolution microscopy, fluorophores are constructed from pairs of organic dyes in an activator and emitter configuration, giving rise to at least nine distinct colors [32]. With this large palette, it can be straightforward to scale up the multiplexing capacity. An alternative to the spectral barcoding used for chromosome labeling involves resolving the spatial order of the barcode on the mRNA in super-resolution. Both spectral and spatial schemes have been demonstrated [33]. The relative advantages and disadvantages of the spatial versus spectral barcoding schemes are that spatial barcoding is more efficient to scale up while spectral coding can be more robustly readout [33].

METHOD

Our work flow for multiplex mRNA detection in single cells is as follows. First, we design probes that will hybridize against each gene specifically. The probes are designed as pairs: one labeled at the 5'-end and the other at the 3'-end, such that when the probe pairs are hybridized on the mRNA, they will bring two dye molecules, an activator and an emitter in close proximity to form a function dye pair for STORM. Second, we will barcode label the probe sets so that each gene in the multiplex is assigned with a unique barcode imparted by the probe. Third, cells are hybridized with this probe set and imaged on a super-resolution microscope (Figure 2). Since the STORM imaging routine involves switching off all of the emitters initially and reactivating them by exciting their neighboring activators, the labeling scheme with pairs of probes allows additional labeling specificity and background rejection. The probes pairs hybridized to target mRNAs can be reactivated, while the probes non-specifically bound to the cell are unlikely to form the dye pairs to be reactivated. Finally, after the imaging, the amount of each barcode detected in individual cells can be counted and used to quantify the abundance of the corresponding mRNAs.

We have recently demonstrated a 32 gene multiplex in single yeast cells, using the spectral super-resolution barcoding technique. To accomplish this multiplex, we used three color barcodes with seven fluorophores (${}_3C_7 = 35$). Three of the barcodes were left intentionally empty to measure the false positive rates. These empty positions were detected at 0.67 ± 0.84 copies per cell, indicating that if more than one copies of a particular mRNA were detected

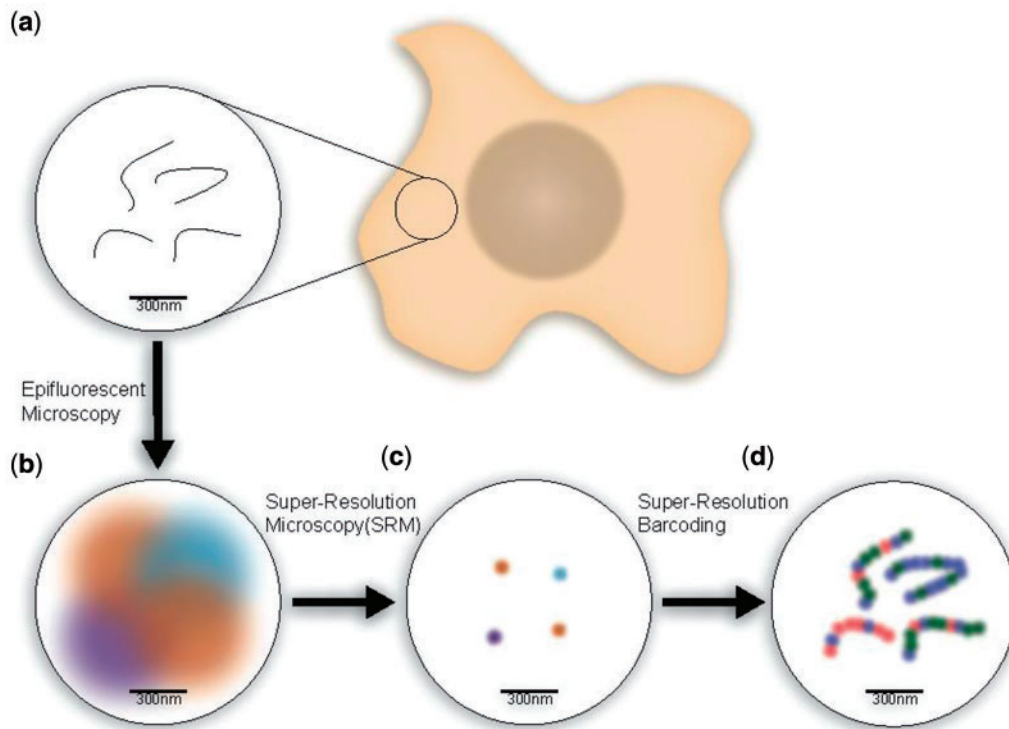


Figure 1: Super-resolution and combinatorial molecular labeling allow multiplex identification and quantification of individual molecules in single cells. (A–B) Individual molecules are difficult to resolve by conventional microscopy due to the diffraction limit of ~ 300 nm. (C) Super-resolution microscopy allows spatial resolution of individual molecules. (D) The identity of molecules can be uniquely addressed by a super-resolution barcode.

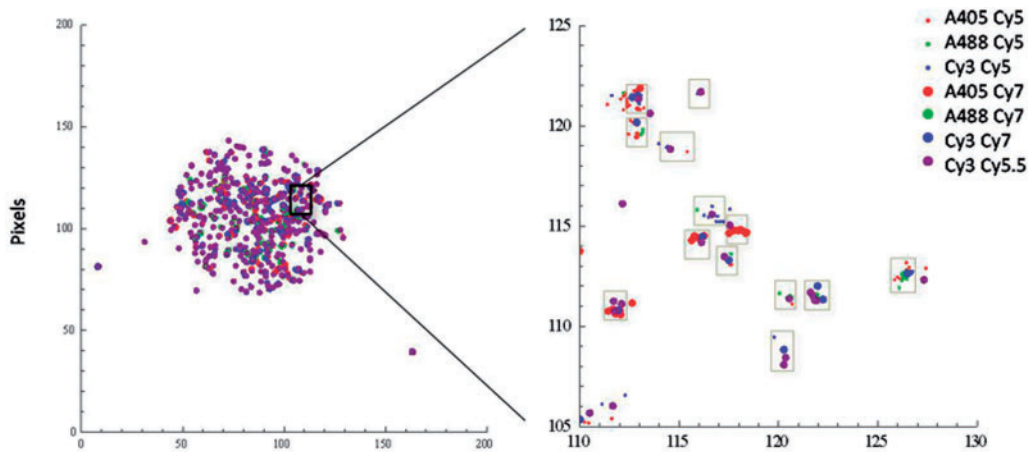


Figure 2: Reconstructions of 32 spectrally barcoded mRNAs in a single cell. Axes are in pixels, each corresponding to 130 nm. The right panel shows a zoomed plot of a region in the cell. Individual mRNAs are shown in boxes. The fluorophores used to barcode are shown in the legend. Each mRNA is coded with three-color barcode from a set of seven dye pairs.

in a cell, they would be confidently measured above the background noise. We performed quantitative comparisons against qPCR and smFISH to show that the super-resolution barcode FISH matches with qPCR and smFISH with a correlation

coefficient of 0.95 in both cases. In addition, we scrambled the barcode order to show that the quantitation is not barcode assignment specific [33]. We used this multiplex to show that gene expression is heterogeneous at the regulon level in individual cells.

OUTLOOK

To further scale up this approach, more fluorophores can be added to the base palette for barcoding. Recent work [34] has shown that there are most likely 15 dye pairs available commercially. A two-color barcode scheme can code for ${}_2C_{15} = 105$ genes and a three-color barcode scheme can allow a ${}_3C_{15} = 455$ gene multiplex. This level of throughput makes this technique a powerful follow-up to RNA-seq experiments with a select group of target genes imaged in cells in their native microenvironments. In addition, considering that most RNAi screens are performed with 100–500 genes, this level of throughput will allow *de novo* discovery of genes with the super-resolution barcoding approach by identify the transcript whose expression correlates the most closely to the phenotype.

The ability to visualize mRNAs *in situ* is powerful for investigating cell–cell interactions in a range of systems from microbial communities to embryos. However, current STORM imaging is confined to thin optical samples. To bring this technology into optically thick samples, light sheet microscopy or Selective Plan Illumination Microscopy (SPIM) [35] is needed to efficiently illuminate the sample and avoid photobleaching. Recent work has shown that combining super-resolution microscopy with SPIM is possible [36]. Further work in applying SPIM to super-resolution barcoding will allow this technique to reach its full potential for single-cell system biology in complex samples involving heterogeneous population of cells.

Super-resolution barcoding can be used beyond multiplex quantitation of mRNA abundance. In principle, any molecule that can be labeled with a high-affinity multi-fluorophore tag can be multiplexed as long as barcode density does not exceed optical resolution. It can be employed to detect transcript diversity in alternative splicing variants as well as to image chromosomal conformation with multiplex labeling of DNA. Recent work has shown the detection of two isoforms of an alternatively spliced mRNA [37, 38]. The super-resolution barcoding can be used to detect more complex splicing patterns, possibly in application to self-avoidance of neurons in flies with DSCAM [39] and proto-cadherins in vertebrates [40]. In addition, we have recently demonstrated efficient DNA FISH for labeling chromosomes with short DNA oligo probes. Multiplexing of DNA FISH will allow us to visualize the native chromosomal structure directly in single cell. Simultaneous

detection of transcription factors or chromatin-modifying proteins with super-resolution imaging on top of the multiplex DNA FISH barcodes serving as landmarks will allow the researchers to perform single-cell versions of chromatin immunoprecipitation experiments. Finally, proteins and other molecular species may be multiplexed in single cells using antibodies or synthetic aptmers [41, 42] with super-resolution barcoding. The vast imaging space available from a single cell under a super-resolution microscope put forward the possibility that single-cell versions of many types of microarray experiments can be implemented with super-resolution barcoding.

CONCLUSION

This technique has the potential transcriptional profile mRNAs with single-molecule sensitivity, generate a high-resolution physical map of chromosome and map protein–DNA interactions, all within individual cells. The preliminary data on resolving barcodes on mRNAs labeled by FISH with super-resolution microscopy demonstrate the feasibility of the super-resolution barcoding approach. This approach offers several distinct advantages: the detection method is high-throughput and single-molecule sensitive, requires only single cells as starting material and the *in situ* labeling preserves the spatial context of molecules in cells as well as cellular contacts in tissues. By mapping regulatory networks within cells and signaling interactions among cells in tissues, we are open to a wide range of biological problems in development and neurobiology in which a individual cells embedded in tissues interact to generate patterns and determine cell fates. The single-cell techniques discussed here have the broad impact from fundamental scientific problems in regulatory networks to medical applications in mapping out the molecular origin of diseases such as cancer and autism, and hold promise as the next-generation diagnostic tools in clinical settings.

Key points

- Super-resolution microscopy allows individual cells to be resolved into millions to billions of pixels, turning individual cells into virtual microarrays.
- Many molecular species can then be barcode labeled and detected simultaneously in single cells with super-resolution imaging.
- We demonstrated using FISH barcode labeling of mRNA of 32 gene multiplex in single yeast cells.

Acknowledgements

The idea for super-resolution barcoding evolved from a discussion with Prof. Barbara Wold about single-molecule techniques and protein-DNA interaction. The mRNA multiplex work was carried out with Eric Lubeck and with a generous equipment loan from Prof. Michael Elowitz.

FUNDING

Funding was initially provided by a Beckman Institute seed grant, and subsequently by a NIH New Innovator Award (1DP2OD008530).

References

- Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995;**270**(5235):467–70.
- Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* 2007;**316**(5830):1497–502.
- Mortazavi A, Williams BA, McCue K, *et al.* Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 2008;**5**(7):621–8.
- Nagalakshmi U, Wang Z, Waern K, Shou C, *et al.* The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 2008;**320**(5881):1344–9.
- Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 2009;**324**(5924):218–23.
- Elowitz MB, Leibler S. A synthetic oscillatory network of transcriptional regulators. *Nature* 2000;**403**(6767):335–8.
- Chalfie M, Tu Y, Euskirchen G, Ward WW, *et al.* Green fluorescent protein as a marker for gene expression. *Science* 1994;**263**(5148):802–5.
- Tsien RY. The green fluorescent protein. *Annu Rev Biochem* 1998;**67**:509–44.
- Süel GM, Garcia-Ojalvo J, Liberman LM, Elowitz MB. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature* 2006;**440**(7083):545–50.
- Rosenfeld N, Young JW, Alon U, *et al.* Gene regulation at the single-cell level. *Science* 2005;**307**(5717):1962–5.
- Lahav G, Rosenfeld N, Sigal A, *et al.* Dynamics of the p53-Mdm2 feedback loop in individual cells. *Nat Genet* 2004;**36**(2):147–50.
- Femino AM, Fay FS, Fogarty K, Singer RH. Visualization of single RNA transcripts in situ. *Science* 1998;**280**(5363):585–90.
- Raj A, van den Bogaard P, Rifkin SA, *et al.* Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods* 2008;**5**(10):877–9.
- Golding I, Paulsson J, Zawilski SM, Cox EC. Real-time kinetics of gene activity in individual bacteria. *Cell* 2005;**123**(6):1025–36.
- Cai L, Friedman N, Xie XS. Stochastic protein expression in individual cells at the single molecule level. *Nature* 2006;**440**(7082):358–62.
- Yu J, Xiao J, Ren X, *et al.* Probing gene expression in live cells, one protein molecule at a time. *Science* 2006;**311**(5767):1600–3.
- Sykes PJ, Neoh SH, Brisco MJ, *et al.* Quantitation of targets for PCR by use of limiting dilution. *Biotechniques* 1992;**13**(3):444–9.
- Kalinina O, Brown J, Silver J. Nanoliter scale PCR with TaqMan detection. *Nucleic Acids Res* 1997;**25**(10):1999–2004.
- Shiroguchi K, Jia TZ, Sims PA, Xie XS. Digital RNA sequencing minimizes sequence-dependent bias and amplification noise with optimized single-molecule barcodes. *Proc Natl Acad Sci USA* 2012;**109**(4):1347–52.
- Casbon JA, Osborne RJ, Brenner S, Lichtenstein CP. A method for counting PCR template molecules with application to next-generation sequencing. *Nucleic Acids Res* 2011;**39**:e81.
- Kinde I, Wu J, Papadopoulos N, *et al.* Detection and quantification of rare mutations with massively parallel sequencing. *Proc Natl Acad Sci USA* 2011;**108**:9530–5.
- Fu GK, Hu J, Wang PH, Fodor SP. Counting individual DNA molecules by the stochastic attachment of diverse labels. *Proc Natl Acad Sci USA* 2011;**108**:9026–9031.
- Kivioja T, Vähärautio A, Karlsson K, *et al.* Counting absolute numbers of molecules using unique molecular identifiers. *Nat Methods* 2011 doi:10.1038/nmeth.1778.
- Betzig E, Patterson GH, Sougrat R, *et al.* Imaging intracellular fluorescent proteins at nanometer resolution. *Science* 2006;**313**(5793):1642–5.
- Rust MJ, Bates M, Zhuang X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat Methods* 2006;**3**(10):793–5.
- Hess ST, Girirajan TP, Mason MD. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys J* 2006;**91**(11):4258–72.
- Gustafsson MG. Nonlinear structured-illumination microscopy: wide-field fluorescence imaging with theoretically unlimited resolution. *Proc Natl Acad Sci USA* 2005;**102**(37):13081–6.
- Dyba M, Hell SW. Photostability of a fluorescent marker under pulsed excited-state depletion through stimulated emission. *Appl Opt* 2003;**42**(25):5123–9.
- Lowenstein MG, Goddard TD, Sedat JW. Long-range interphase chromosome organization in *Drosophila*: a study using color barcoded fluorescence in situ hybridization and structural clustering analysis. *Mol Biol Cell* 2004;**15**(12):5678–92.
- Kosman D, Mizutani CM, Lemons D, *et al.* Multiplex detection of RNA expression in *Drosophila* embryos. *Science* 2004;**305**(5685):846.
- Levsky JM, Shenoy SM, Pezo RC, Singer RH. Single-cell gene expression profiling. *Science* 2002;**297**(5582):836–40.
- Bates M, Huang B, Dempsey GT, Zhuang X. Multicolor super-resolution imaging with photo-switchable fluorescent probes. *Science* 2007;**317**(5845):1749–53.
- Lubeck E, Cai L. Single cell systems biology by super-resolution imaging and combinatorial labeling. *Nat Methods* 2012;**9**:743–8.
- Dempsey GT, Vaughan JC, Chen KH, *et al.* Evaluation of fluorophores for optimal performance in localization-

- based super-resolution imaging. *Nat Methods* 2011; **8**(12):1027–36.
35. Huisken J, Swoger J, Del Bene F, *et al.* Optical sectioning deep inside live embryos by selective plane illumination microscopy. *Science* 2004;**305**(5686):1007–9.
 36. Zanicchi CF, Lavagnino Z, Donnorso PM, *et al.* Live-cell 3D super-resolution imaging in thick biological samples. *Nat Methods* 2011;**8**(12):1047–9.
 37. Waks Z, Klein AM, Silver PA. Cell-to-cell variability of alternative RNA splicing. *Mol Syst Biol* 2011;**7**:506.
 38. Vargas DY, Shah K, Batish M, *et al.* Single-molecule imaging of transcriptionally coupled and uncoupled splicing. *Cell* 2012;**147**(5):1054–65.
 39. Hattori D, Millard SS, Wojtowicz WM, Zipursky SL. Dscam-mediated cell recognition regulates neural circuit formation. *Annu Rev Cell Dev Biol* 2008;**24**:597–620.
 40. Lefebvre JL, Kostadinov D, Chen WV, *et al.* Protocadherins mediate dendritic self-avoidance in the mammalian nervous system. *Nature* 2012;**488**(7412):517–21.
 41. Cho EJ, Lee JW, Ellington AD. Applications of aptamers as sensors. *Annu Rev Anal Chem* 2009;**2**:241–64.
 42. Agnew HD, Rohde RD, Millward SW, *et al.* Iterative in situ click chemistry creates antibody-like protein-capture agents. *Angew Chem Int Ed Engl* 2009;**48**(27):4944–8.