Towards *in situ* Single Cell Systems Biology

Thesis by

Eric Lubeck

In Partial Fulfillment of the Requirements for
the degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2016
(Defended November 20th, 2015)

I dedicate my thesis to my family, without your support none of this would be possible.

Thank you Helen, love you.

Eric Lubeck
ORCID: 0000-0002-5457-0258

Acknowledgments

Thanks to Sheel Shah and Wen Zhou for being fantastic scientific partners and friends. Your assistance and insight are essential and a continual pleasure. Thanks also to my advisor Long Cai for looking out for me and being my friend through this long march.

Abstract

Systems-level studies of biological systems rely on observations taken at a resolution lower than the essential unit of biology, the cell. Recent technical advances in DNA sequencing have enabled measurements of the transcriptomes in single cells excised from their environment, but it remains a daunting technical problem to reconstruct *in situ* gene expression patterns from sequencing data. In this thesis I develop methods for the routine, quantitative *in situ* measurement of gene expression using fluorescence microscopy.

The number of molecular species that can be measured simultaneously by fluorescence microscopy is limited by the pallet of spectrally distinct fluorophores. Thus, fluorescence microscopy is traditionally limited to the simultaneous measurement of only five labeled biomolecules at a time. The two methods described in this thesis, super-resolution barcoding and temporal barcoding, represent strategies for overcoming this limitation to monitor expression of many genes in a single cell. Super-resolution barcoding employs optical super-resolution microscopy (SRM) and combinatorial labeling via-smFISH (single molecule fluorescence *in situ* hybridization) to uniquely label individual mRNA species with distinct barcodes resolvable at nanometer resolution. This method dramatically increases the optical space in a cell, allowing a large numbers of barcodes to be visualized simultaneously. As a proof of principle this technology was used to study the S. cerevisiae calcium stress response. The second method, sequential barcoding, reads out a temporal barcode through multiple rounds of oligonucleotide hybridization to the same mRNA. The multiplexing capacity of sequential barcoding increases exponentially with the number of rounds of hybridization, allowing over a hundred genes to be profiled in only a few rounds of hybridization.

The utility of sequential barcoding was further demonstrated by adapting this method to study gene expression in mammalian tissues. Mammalian tissues suffer both from a large amount of auto-fluorescence and light scattering, making detection of smFISH probes on mRNA difficult. An amplified single molecule detection technology, smHCR (single molecule hairpin chain reaction), was developed to allow for the quantification of mRNA in tissue. This technology is demonstrated in combination with light sheet microscopy and background reducing tissue clearing technology, enabling whole-organ sequential barcoding to monitor *in situ* gene expression directly in intact mammalian tissue.

The methods presented in this thesis, specifically sequential barcoding and smHCR, enable multiplexed transcriptional observations in any tissue of interest. These technologies will serve as a general platform for future transcriptomic studies of complex tissues.

PUBLISHED CONTENT AND CONTRIBUTIONS

E. Lubeck, L. Cai, Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nature Methods*. **9**, 743–748 (2012). doi:10.1038/nmeth.2069

E. Lubeck participated in the conception of the project, collected and prepared data, and participated in the writing of the manuscript.

E. Lubeck, A. F. Coskun, T. Zhiyentayev, M. Ahmad, L. Cai, Single-cell in situ RNA profiling by sequential hybridization. *Nat Meth*. **11**, 360–361 (2014). doi: 10.1038/nmeth.2892

E. Lubeck participated in the conception of the project, collected and prepared data, and participated in the writing of the manuscript.

B. Yang *et al.*, Single-Cell Phenotyping within Transparent Intact Tissue through Whole-Body Clearing. *Cell* (2014). doi:10.1016/j.cell.2014.07.017

E. Lubeck collected and prepared a subset of data for this manuscript.

Contents

Chapter 1

Introduction

1.1 Intro

Biological investigation has been revolutionized by the ability to abstract large data sets from sequencing. This deluge of data has allowed for the large-scale inference of gene regulatory networks, yet only the strongest regulatory links have been extracted from these datasets. One reason for this disconnect is that most of the data so far acquired is not on a single cell basis, meaning that it averages over much of the variation seen in individual cells. This averaging obscures many important sub states in data and dilutes correlations between co-regulated genes, weakening the ability to make inferences about biology.

Several technologies exist to enable single cell sequencing *in vitro*. All of these technologies currently suffer from the same problem: they require removal of cells from their native environment. This obscures the spatial relationships between cells in the environment or inside of a given tissue. Although useful, all *in vitro* single cell technologies limit their scope by reducing spatial resolution.

For my thesis projects I have developed a suite of technologies that allow researchers to monitor gene expression *in situ* in a highly multiplexed fashion. These technologies will enable scientists to identify rare cells in their environment, allowing new questions to be asked. Through the technology I've developed it will become routine to address a researcher's hypothesis inside of the system of interest, instead of outside the native environment in a controlled model system. The work in this thesis will also facilitate medical diagnosis; the ability to monitor gene expression inside of a tissue of interest will augment the abilities of pathologists and other clinicians to make a diagnosis on complex tissue biopsies.

1.2 Research Applications

In the following sections selected areas of research where *in situ* single cell transcriptional profiling will be useful are highlighted.

1.2.1 Cell Type Identification and Mapping

Identifying and mapping the spatial distribution of cell types is an important problem touching all areas of multi-cellular biology. The relationship between diverse cell types inside of a tissue ultimately dictates a tissues function, so it is essential to map the distribution of cell types. Although many cell types have been well characterized *in vitro* through cell culture models and flow sorting, the behavior of these cells in their native environments is often little understood.

Identifying and mapping cell types is of particular importance to neurobiology. Many of the cell types understood by neurobiology were discovered either through clear morphological differences, knockout experiments, or through Cre lines that label a specific population expressing a known cell-type transcription factor cite:taniguchi$_{\mathrm{resource2011}}$. Once these cell types are discovered electrophysiologists can then go in and make single cell measurements of the neuronal activity of that specific cell type. By mapping the spatial distribution of cell types with diverse neural activity patterns, quantitative predictions of neural circuits can be constructed.

Single cell sequencing is a powerful technology for discovering cell types. Transcriptional differences underlie much of the variation between cell types cite:usoskin$_{\mathrm{unbiased2015,zeiselcell2015,grunsinglecell2015}}$. Instead of relying on generating specific genetic markers for cell types or identifying clear morphological differences between cells, sampling a large amount of the transcriptome allows unbiased clustering of cells into different cell types.

For cell types to be effectively discovered and mapped in tissue, single cell *in situ* measurements are necessary. The technologies later developed in this thesis are being directly applied to this problem. In combination with other high-throughput ways of mapping live cell activity, the technology developed in this thesis will play a powerful role in characterizing the distribution of cell types.

1.2.2 Gene Regulatory Circuits

The behavior of cells is partially dictated by genetic programs that modulate expression of specific genes. Gene regulatory networks map the input of various regulatory factors and produce an output in the production of proteins of interest cite:davidson$_{\mathrm{gene2006}}$. A large amount of a cell's ability to adapt to varying stimulus is encoded and regulated in the genome, so the study of the logic and regulation of gene regulatory networks is important for dissecting phenomenon in almost all fields of biology cite:davidson$_{\mathrm{gene2006,shaleksinglecell2014,singerdynamic2014,wanggene}}$.

Single cell technologies play an important role in elucidating regulatory circuits. Cells in clonal populations are often in a diverse array of cell states, making elucidation of regulatory factors difficult for all but the most inducible phenotypes cite:baharhalpern$_{\mathrm{bursty,kumardeconstructing2014,buganimsinglecell2012}}$. Single Cell transcriptional measurements remove this averaging effect, allowing for a fair evaluation of the heterogeneous factors affecting gene regulatory network output throughout the population of cells.

Current single cell technologies make a trade-off between either dynamics and fidelity, or sampling depth. Imaging techniques allow for extremely quantitative and dynamic measurements of a relatively small number of factors cite:singer$_{\text{dynamic2014,caifrequencymodulated2008,dunlopregulatory2008}}$. By contrast, single cell sequencing techniques collect broad data on a large amount of the transcriptome, but suffer from bias and coverage issues limiting the accuracy of their measurements cite:klein$_{\text{droplet2015,usoskinunbiased2015,macoskohighly2015}}$.

In situ single cell transcriptional profiling technologies can serve as a nice middle ground between sequencing and imaging. The technology developed in this thesis allows for monitoring several hundred genes simultaneously *in situ* and is compatible with live cell imaging approaches. This combination of live cell imaging and transcriptomics will allow for powerful new studies of gene regulatory circuits.

1.2.3 Rare Cell Characterization

Many biomedically important processes are the result of a small number of cells. Adult stem cells, in spite of their relatively limited numbers, play an important role in organ regeneration, aging, and malignancy. Neural circuit activity relies on an interplay of many different and low-abundance cell types. Despite the outsize importance of rare cells, methods for performing high-throughput studies of these cell *in situ* have been severely lacking.

Adult stem cells play many important and diverse roles in regenerating tissues. Cancer stem cells are a rare cell population hypothesized to be one of the root causes of cancer malignancy cite:mani$_{\text{epithelialmesenchymal2008}}$. For cancer stem cells it is not yet clear what combination of factors leads these cells to undergo an epithelial to mesenchymal transition and become metastatic cancer cells. It's believed that the local micro-environment of the cancer stem cells plays an important role in this transition, but it's necessary to find these transitions *in situ* to validate this hypothesis and fully characterize the transition cite:lu$_{\text{breast2014}}$-1. The stem cell populations underlying normal tissue maintenance have been little characterized outside of skin and colon systems cite:blanpain$_{\text{plasticity2014,grunsinglecell2015}}$.

In neural biology many rare cell types can only be meaningfully studied *in situ.* Many studies in neurobiology use a combination of Cre markers for a specific cell type or transcriptional state, and use a combination of electrophysiology or optogenetic methods to characterize the activity of the cell of interest cite:liu$_{\text{optogenetic2012,madisentoolbox2012}}$. Although these studies provide compelling data about the function of cells inside of a neural circuit, they remain limited in their ability to dissect the regulatory circuits underlying the specification of a cells activity. Higher-dimensional imaging should circumvent this limitation, and make it possible to study the genetic circuits underlying neural activity states.

In stem cell and immunological studies many of these rare cell types have been traditionally studied via flow sorting for cell surface markers known to be present on these cell types. This has enabled the characterization of these cells through culturing and transplantation approaches and more recently via sequencing. Because these cells are so rare, it has traditionally been very difficult to identify and study these cells *in situ*. Imaging allows these cells to be dissected in their native environment so questions about extrinsic factors effecting cells can be addressed, but traditional imaging approaches can monitor only a small number of factors at a time.

The technology developed in this thesis will allow for rare cells and their microenviroments to be efficiently identified and transcriptionally profiled *in situ*. By augmenting the transcriptional profiling technology developed in this thesis with common commercially available high-throughput imaging technology, large tissue volumes can be scanned as a number of serial sections. Such an approach will allow scientists to identify and characterize these rare cell populations inside their environments.

1.2.4 Ecology

The ecology of microbial communities is a vastly understudied problem. Most microbes are studied as cultured isolates outside of their natural environment. Very few of these microbes can even be cultured and of those that can be a large amount of the transcriptome is never expressed under culture conditions cite:donia$_{systematic2014}$. These unexpressed genes presumably play important roles in their native environment, but the lack of an ability to study them *in vitro* has hampered the ability to characterize these genes.

Interspecies cooperation between humans and commensal microbes may be important for human health. Many compounds have so far been identified from commensal microbial genomes that modulate the activity of the host immune system cite:donia$_{systematic2014,atarashith172015,sharonspecialized2014}$. Although the biochemistry of this phenomenon is starting to become elucidated, little knowledge exists about the regulation of small molecule biosynthetic gene clusters *in situ*.

Current approaches to studying biofilms include metagenomic sequencing and studying cultured isolates from biofilms. Metagenomic sequencing has advanced tremendously in the last few years, now routinely enabling a survey of the taxa present in a sample along with the genes being expressed cite:nielsen$_{identification2014,koremgrowth2015}$-1. While the metagenomic data provides a useful survey of transcriptional activity of a biofilm, it makes it difficult to ascertain spatial patterns and interspecies cooperation inside the biofilm.

To truly understand the behavior of individual microbes an *in situ* picture must be developed of the microbes. The technology developed in this thesis will allow for single cell *in situ* resolution of gene expression inside of a biofilm. This should make it possible to generate a hypothesis about the contributions of individual microbes to the metabolic activity of the overall biofilm and ascertain novel methods of interspecies cooperation.

1.3 Single Cell Technologies

Recently a number of technologies have emerged for enabling both *in situ* and *in vitro* readouts of single cell transcriptional state. All of these technologies make a fundamental trade-off between spatial resolution and information content.

1.3.1 Sequencing

Single cell RNA sequencing has made tremendous strides in the last few year, now routinely delivering quantitative and high-quality data on 10s to 1000s of cells simultaneously cite:shalek$_{\text{singlecell2014,kleindroplet2015,macoskohi}}$ A combination of microfluidics and barcode-sequence ligation has turned single cell sequencing from a lossy and expensive endeavor, into a significantly more quantitative and accessible technology. Despite these advances, there still exists fundamental limitations preventing sequencing from becoming the ultimate technology.

Quantitative single cell sequencing technologies use a barcode identifier covalently ligated to the cDNA from a single cell. This barcode enables many single cells to be pooled together in one sequencing reaction, allowing deconvolution of the pooled sequencing data into unique single cells through sequencing of the barcode cite:islam$_{\text{quantitative2014}}$. In addition many systems also place a unique barcode on every mRNA molecule, enabling the facile quantification of individual mRNAs by counting the unique barcodes attached to them cite:islam$_{\text{quantitative2014}}$. This circumvents problems with amplification bias seen in early next-generation sequencing experiments cite:aird$_{\text{analyzing2011}}$.

Many strategies exist for single cell isolation and subsequent barcoding by microfluidics and manual partitioning of cells. Of particular note are microfluidic approaches to single cell isolation made commercially available by Fluidigm capable of capturing data on 96 single cells cite:hong$_{\text{nanoliterscale2004}}$. More recently Drop-Seq has enabled the cost-effective profiling of many thousands of single cells in a single sequencing reaction, greatly reducing the cost and increasing the throughput of single cell sequencing cite:klein$_{\text{droplet2015,macoskohighly2015}}$. These technologies will continue to improve for the foreseeable future, and will provide a robust method of quantifying single cell data outside of it's native spatial context.

Several efforts are underway to adapt single cell sequencing into an *in situ* technology. FISSEQ is an attempt to directly sequence mRNA *in situ*, but suffers from a very low efficiency(~0.1%) cite:lee$_{\text{highly2014}}$. Efforts are underway to probabilistically infer a cells position in its environment cite:satija$_{\text{spatial2015}}$ and to directly attach position dependent barcodes to all mRNA in cells of interest *in situ*, followed by reconstruction of tissue morphology after barcode sequencing. These indirect methods will partially bridge the gap between true *in situ* sequencing and high-throughput *in vitro* sequencing, but will not provide the capacity for discovery possible with direct *in situ* measurements.

1.3.2 Imaging

Fluorescent imaging technologies have intrinsic single cell resolution. For this reason many *in situ* assays are performed with a fluorescent microscope, but they are traditionally limited in the number of species that can be monitored simultaneously and the accuracy of these *in situ* measurements. Herein I briefly review imaging technologies that facilitate single cell analysis.

FISH. Single molecule fluorescence *in situ* hybridization (smFISH) allows for the single molecule identification and quantification of pre-defined genes. By designing many fluorescent oligonucleotides targeted to an mRNA molecule of interest, a bright easily identifiable spot can be visualized with a standard wide-field fluorescent microscope cite:raj$_{\text{imaging2008}}$. The position of the mRNAs in the fixed cell is perfectly preserved, allowing mRNAs to be identified and quantified in their native spatial context.

The primary challenges for using smFISH for mRNA identification lie in its lack of multiplexing ability, and the difficulty of designing specific probes against many targets. The lack of multiplexing arises from the limited spectral pallet of easily-distinguishable dyes, making labeling of more than five species simultaneously impossible if each species is to be distinguished with a single flurophore. The lack of multiplexing ability of smFISH is addressed by the research in this thesis. Despite the limitations of designing complementary oligonucleotide probes, good probesets can be designed for simultaneous use on at least half the mouse genome and can be synthesized using microarrays in a relatively cost effect fashion cite:kosuri$_{\text{largescale2014}}$. Analyzing small sequence variations is a problem for FISH technology. Techniques exist for SNP and alternative splicing visualization cite:levesque$_{\text{visualizing2013,levesquesinglechromosome2013}}$, but these technologies have yet to be demonstrated to be multiplexable.

Tissue Clearing. One of the common problems with performing smFISH studies in thick tissues is the lack of fluorescent signal that diffuses through high-background samples such as bone and brain. This problem can be ameliorated by two complementary approaches: fluorescent signal of individual smFISH foci can be amplified to obtain much higher gain signal, or photon scattering background can be removed from tissue samples, leaving the tissue transparent and amenable to deep imaging.

Samples can be rendered transparent by embedding in hydrogels followed by stripping away of non-target components such as lipids and proteins. Several techniques have been developed to accomplish this goal, including Clarity cite:chung$_{\text{structural2013}}$ and PACT cite:yang$_{\text{singlecell2014}}$. These methods enable smFISH signal to be seen relatively deep into the tissue, and in combination with amplification techniques make it simple to quantify fluorescent molecules. Hydrogel embedding techniques also permit the expansion of tissue, enabling normally optically un-resolvable features to be physically separated and rendering many sub-diffraction details visible. cite:chen$_{\text{expansion2015}}$. All of these hydrogel embedding methods greatly expand both the imaging depth and resolution of fluorescent microscopy, enabling technologies like the ones developed in this thesis to scale up to organ level observations of transcriptional networks with single cell resolution.

Lightsheet. Advanced microscopes are also necessary for highly multiplexed deep *in situ* imaging. The current state-of-the-art in microscope technology is light sheet microscopy. These scopes allow large tissue volumes to be imaged relatively fast with minimal photo-bleaching. Confocal microscopes allow a single molecule signal to be visualized relatively deep in tissue, but photobleaching becomes a problem for imaging particularly thick samples. In principal light sheet microscopes in combination with tissue clearing should be the optimal technology for addressing this problem, but the lack of adequate and affordable commercial light-sheet microscopes has led to the ad-hoc development of very different setups between labs. In the near future as the market need for light sheet microscopes is addressed by conventional microscope manufacturers, this technology will become commonplace for *in situ* single molecule studies of gene expression of complex tissues.

1.4 Conclusion

In this thesis I will provide a description of a series of methods to merge the advantages of single cell microscopy with the data depth of single cell sequencing techniques. This technical work is still ongoing, but already provides a nice middle ground between sequencing and imaging; about 100 genes can be robustly quantified at one time inside of complex tissue samples. This technological development will open up many questions about *in situ* biology to researchers. As this technology is further developed into commercial products and core facilities, it will become a routine methodology for studying complex biological processes *in situ*.

Bibliography

[1] Hiroki Taniguchi et al. “A Resource of Cre Driver Lines for Genetic Targeting of GABAergic Neurons in Cerebral Cortex”. In: *Neuron* 71.6 (Sept. 2011), pp. 995–1013. ISSN: 0896-6273. DOI: `10.1016/j.neuron.2011.07.026`. URL: `http://www.sciencedirect.com/science/article/pii/S0896627311006799` (visited on 08/25/2014).

[2] Dmitry Usoskin et al. “Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing”. en. In: *Nature Neuroscience* 18.1 (Jan. 2015), pp. 145–153. ISSN: 1097-6256. DOI: `10.1038/nn.3881`. URL: `http://www.nature.com/neuro/journal/v18/n1/full/nn.3881.html` (visited on 01/05/2015).

[3] Amit Zeisel et al. “Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq”. en. In: *Science* (Feb. 2015), aaa1934. ISSN: 0036-8075, 1095-9203. DOI: `10.1126/science.aaa1934`. URL: `http://www.sciencemag.org/content/early/2015/02/18/science.aaa1934` (visited on 02/20/2015).

[4] Dominic Grün et al. “Single-cell messenger RNA sequencing reveals rare intestinal cell types”. en. In: *Nature* advance online publication (Aug. 2015). ISSN: 0028-0836. DOI: `10.1038/nature14966`. URL: `http://www.nature.com/nature/journal/vaop/ncurrent/full/nature14966.html` (visited on 08/19/2015).

[5] Eric H. Davidson and Douglas H. Erwin. “Gene Regulatory Networks and the Evolution of Animal Body Plans”. en. In: *Science* 311.5762 (Oct. 2006), pp. 796–800. ISSN: 0036-8075, 1095-9203. DOI: `10.1126/science.1113832`. URL: `http://www.sciencemag.org/content/311/5762/796` (visited on 10/24/2015).

[6] Alex K. Shalek et al. “Single-cell RNA-seq reveals dynamic paracrine control of cellular variation”. en. In: *Nature* 510.7505 (June 2014), pp. 363–369. ISSN: 0028-0836. DOI: `10.1038/nature13437`. URL: `http://www.nature.com/nature/journal/v510/n7505/full/nature13437.html` (visited on 02/04/2015).

[7] Zakary S. Singer et al. “Dynamic Heterogeneity and {DNA} Methylation in Embryonic Stem Cells”. In: *Molecular Cell* 55.2 (2014), pp. 319 –331. ISSN: 1097-2765. DOI: `http://dx.doi.`

org/10.1016/j.molcel.2014.06.029. URL: `http://www.sciencedirect.com/science/article/pii/S1097276514005632`.

[8] Sui Wang et al. "A Gene Regulatory Network Controls the Binary Fate Decision of Rod and Bipolar Cells in the Vertebrate Retina". In: *Developmental Cell* (). ISSN: 1534-5807. DOI: 10.1016/j.devcel.2014.07.018. URL: `http://www.sciencedirect.com/science/article/pii/S1534580714004845` (visited on 08/27/2014).

[9] Keren Bahar Halpern et al. "Bursty Gene Expression in the Intact Mammalian Liver". In: *Molecular Cell* (). ISSN: 1097-2765. DOI: 10.1016/j.molcel.2015.01.027. URL: `http://www.sciencedirect.com/science/article/pii/S1097276515000507` (visited on 03/03/2015).

[10] Roshan M. Kumar et al. "Deconstructing transcriptional heterogeneity in pluripotent stem cells". en. In: *Nature* 516.7529 (Dec. 2014), pp. 56–61. ISSN: 0028-0836. DOI: 10.1038/nature13920. URL: `http://www.nature.com/nature/journal/v516/n7529/full/nature13920.html` (visited on 12/03/2014).

[11] Yosef Buganim et al. "Single-Cell Expression Analyses during Cellular Reprogramming Reveal an Early Stochastic and a Late Hierarchic Phase". en. In: *Cell* 150.6 (Sept. 2012), pp. 1209–1222. ISSN: 00928674. DOI: 10.1016/j.cell.2012.08.023. URL: `http://linkinghub.elsevier.com/retrieve/pii/S0092867412010215` (visited on 08/24/2014).

[12] Long Cai, Chiraj K. Dalal, and Michael B. Elowitz. "Frequency-modulated nuclear localization bursts coordinate gene regulation". en. In: *Nature* 455.7212 (Sept. 2008), pp. 485–490. ISSN: 0028-0836. DOI: 10.1038/nature07292. URL: `http://www.nature.com/nature/journal/v455/n7212/full/nature07292.html` (visited on 02/24/2015).

[13] Mary J Dunlop et al. "Regulatory activity revealed by dynamic correlations in gene expression noise". In: *Nature Genetics* 40.12 (Dec. 2008), pp. 1493–1498. ISSN: 1061-4036. DOI: 10.1038/ng.281. URL: `http://www.nature.com/doifinder/10.1038/ng.281` (visited on 08/24/2014).

[14] Allon M. Klein et al. "Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells". In: *Cell* 161.5 (May 2015), pp. 1187–1201. ISSN: 0092-8674. DOI: 10.1016/j.cell.2015.04.044. URL: `http://www.sciencedirect.com/science/article/pii/S0092867415005000` (visited on 08/29/2015).

[15] Evan Z. Macosko et al. "Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets". In: *Cell* 161.5 (May 2015), pp. 1202–1214. ISSN: 0092-8674. DOI: 10.1016/j.cell.2015.05.002. URL: `http://www.sciencedirect.com/science/article/pii/S0092867415005498` (visited on 08/29/2015).

[16] Sendurai A. Mani et al. "The Epithelial-Mesenchymal Transition Generates Cells with Properties of Stem Cells". In: *Cell* 133.4 (May 2008), pp. 704–715. ISSN: 0092-8674. DOI: `10.1016/j.cell.2008.03.027`. URL: `http://www.sciencedirect.com/science/article/pii/S0092867408004443` (visited on 10/26/2015).

[17] Haihui Lu et al. "A Breast Cancer Stem Cell Niche Supported by Juxtacrine Signaling from Monocytes and Macrophages". In: *Nature cell biology* 16.11 (Nov. 2014), pp. 1105–1117. ISSN: 1465-7392. DOI: `10.1038/ncb3041`. URL: `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4296514/` (visited on 11/01/2015).

[18] Cédric Blanpain and Elaine Fuchs. "Plasticity of epithelial stem cells in tissue regeneration". en. In: *Science* 344.6189 (June 2014), p. 1242281. ISSN: 0036-8075, 1095-9203. DOI: `10.1126/science.1242281`. URL: `http://www.sciencemag.org/content/344/6189/1242281` (visited on 04/04/2015).

[19] Xu Liu et al. "Optogenetic stimulation of a hippocampal engram activates fear memory recall". en. In: *Nature* 484.7394 (Apr. 2012), pp. 381–385. ISSN: 0028-0836. DOI: `10.1038/nature11028`. URL: `http://www.nature.com/nature/journal/v484/n7394/full/nature11028.html` (visited on 08/23/2014).

[20] Linda Madisen et al. "A toolbox of Cre-dependent optogenetic transgenic mice for light-induced activation and silencing". en. In: *Nature Neuroscience* 15.5 (May 2012), pp. 793–802. ISSN: 1097-6256. DOI: `10.1038/nn.3078`. URL: `http://www.nature.com/neuro/journal/v15/n5/full/nn.3078.html` (visited on 10/27/2015).

[21] Mohamed S. Donia et al. "A Systematic Analysis of Biosynthetic Gene Clusters in the Human Microbiome Reveals a Common Family of Antibiotics". In: *Cell* 158.6 (Sept. 2014), pp. 1402–1414. ISSN: 0092-8674. DOI: `10.1016/j.cell.2014.08.032`. URL: `http://www.sciencedirect.com/science/article/pii/S0092867414011027` (visited on 01/07/2015).

[22] Koji Atarashi et al. "Th17 Cell Induction by Adhesion of Microbes to Intestinal Epithelial Cells". English. In: *Cell* 163.2 (Oct. 2015), pp. 367–380. ISSN: 0092-8674. DOI: `10.1016/j.cell.2015.08.058`. URL: `http://www.cell.com/article/S0092867415011101/abstract` (visited on 10/27/2015).

[23] Gil Sharon et al. "Specialized Metabolites from the Microbiome in Health and Disease". In: *Cell Metabolism* 20.5 (Nov. 2014), pp. 719–730. ISSN: 1550-4131. DOI: `10.1016/j.cmet.2014.10.016`. URL: `http://www.sciencedirect.com/science/article/pii/S155041311400463X` (visited on 08/29/2015).

[24] H. Bjørn Nielsen et al. “Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes”. en. In: *Nature Biotechnology* 32.8 (Aug. 2014), pp. 822–828. ISSN: 1087-0156. DOI: `10.1038/nbt.2939`. URL: `http://www.nature.com/nbt/journal/v32/n8/full/nbt.2939.html` (visited on 10/09/2015).

[25] Tal Korem et al. “Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples”. en. In: *Science* (July 2015), aac4812. ISSN: 0036-8075, 1095-9203. DOI: `10.1126/science.aac4812`. URL: `http://www.sciencemag.org/content/early/2015/07/29/science.aac4812` (visited on 07/30/2015).

[26] Saiful Islam et al. “Quantitative single-cell RNA-seq with unique molecular identifiers”. en. In: *Nature Methods* 11.2 (Feb. 2014), pp. 163–166. ISSN: 1548-7091. DOI: `10.1038/nmeth.2772`. URL: `http://www.nature.com/nmeth/journal/v11/n2/full/nmeth.2772.html` (visited on 10/27/2015).

[27] Daniel Aird et al. “Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries”. en. In: *Genome Biology* 12.2 (Feb. 2011), R18. ISSN: 1465-6906. DOI: `10.1186/gb-2011-12-2-r18`. URL: `http://genomebiology.com/2011/12/2/R18/abstract` (visited on 10/27/2015).

[28] Jong Wook Hong et al. “A nanoliter-scale nucleic acid processor with parallel architecture”. en. In: *Nature Biotechnology* 22.4 (Apr. 2004), pp. 435–439. ISSN: 1087-0156. DOI: `10.1038/nbt951`. URL: `http://www.nature.com/nbt/journal/v22/n4/full/nbt951.html` (visited on 10/27/2015).

[29] J. H. Lee et al. “Highly Multiplexed Subcellular RNA Sequencing in Situ”. en. In: *Science* 343.6177 (Mar. 2014), pp. 1360–1363. ISSN: 0036-8075, 1095-9203. DOI: `10.1126/science.1250212`. URL: `http://www.sciencemag.org/cgi/doi/10.1126/science.1250212` (visited on 08/24/2014).

[30] Rahul Satija et al. “Spatial reconstruction of single-cell gene expression data”. en. In: *Nature Biotechnology* 33.5 (May 2015), pp. 495–502. ISSN: 1087-0156. DOI: `10.1038/nbt.3192`. URL: `http://www.nature.com/nbt/journal/v33/n5/full/nbt.3192.html` (visited on 10/27/2015).

[31] Arjun Raj et al. “Imaging individual mRNA molecules using multiple singly labeled probes”. en. In: *Nature Methods* 5.10 (Oct. 2008), pp. 877–879. ISSN: 1548-7091. DOI: `10.1038/nmeth.1253`. URL: `http://www.nature.com/nmeth/journal/v5/n10/full/nmeth.1253.html` (visited on 10/27/2015).

[32] Sriram Kosuri and George M. Church. "Large-scale de novo DNA synthesis: technologies and applications". en. In: *Nature Methods* 11.5 (May 2014), pp. 499–507. ISSN: 1548-7091. DOI: `10.1038/nmeth.2918`. URL: `http://www.nature.com/nmeth/journal/v11/n5/abs/nmeth.2918.html` (visited on 10/05/2014).

[33] Marshall J. Levesque et al. "Visualizing SNVs to quantify allele-specific expression in single cells". en. In: *Nature Methods* 10.9 (Sept. 2013), pp. 865–867. ISSN: 1548-7091. DOI: `10.1038/nmeth.2589`. URL: `http://www.nature.com/nmeth/journal/v10/n9/abs/nmeth.2589.html` (visited on 08/13/2015).

[34] Marshall J. Levesque and Arjun Raj. "Single-chromosome transcriptional profiling reveals chromosomal gene expression regulation". en. In: *Nature Methods* 10.3 (Mar. 2013), pp. 246–248. ISSN: 1548-7091. DOI: `10.1038/nmeth.2372`. URL: `http://www.nature.com/nmeth/journal/v10/n3/full/nmeth.2372.html?WT.ec_id=NMETH-201303` (visited on 08/02/2015).

[35] Kwanghun Chung et al. "Structural and molecular interrogation of intact biological systems". en. In: *Nature* 497.7449 (May 2013), pp. 332–337. ISSN: 0028-0836. DOI: `10.1038/nature12107`. URL: `http://www.nature.com/nature/journal/v497/n7449/full/nature12107.html` (visited on 02/24/2015).

[36] Bin Yang et al. "Single-Cell Phenotyping within Transparent Intact Tissue through Whole-Body Clearing". In: *Cell* (2014).

[37] Fei Chen, Paul W. Tillberg, and Edward S. Boyden. "Expansion microscopy". en. In: *Science* 347.6221 (Jan. 2015), pp. 543–548. ISSN: 0036-8075, 1095-9203. DOI: `10.1126/science.1260088`. URL: `http://www.sciencemag.org/content/347/6221/543` (visited on 10/29/2015).

Chapter 2

Single cell systems biology by super-resolution imaging and combinatorial labeling

Eric Lubeck[1]and Long Cai[1,2]

1. Program in Biochemistry and Molecular Biophysics 2. Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena CA 91125 USA

2.1 Abstract

Fluorescence microscopy is a powerful quantitative tool for exploring regulatory networks in single cells. However, the number of molecular species that can be measured simultaneously is limited by the spectral separability of fluorophores. Here we demonstrate a simple but general strategy to drastically increase the capacity for multiplex detection of molecules in single cells by using optical super-resolution microscopy (SRM) and combinatorial labeling. As a proof of principle, we labeled mRNAs with unique combinations of fluorophores using Fluorescence *in situ* Hybridization (FISH), and resolved the sequences and combinations of fluorophores with SRM. We measured the mRNA levels of 32 genes simultaneously in single yeast cells. These experiments demonstrate that combinatorial labeling and super-resolution imaging of single cells provides a natural approach to bring systems biology into single cells.

2.2 Introduction

Systems biology seeks to understand interactions between biological components. Two distinct scales of experimentation are employed: large-scale approaches use sequencing, [1, 2, 3, 4] microarray [5] and mass-spec technologies, while small-scale techniques employ fluorescence microscopy [6, 7, 8]. Large-scale techniques provide genome-wide information, but lack sensitivity and average over heterogeneity and spatial information within a large population of cells. In contrast, microscopy-based techniques can track cells and molecules *in situ* [9, 10, 11], with single molecule sensitivity [12, 13, 14, 15]. Previous works have reported that chromosome loci [16] and cellular proteins [17] can be multiplexed up to 10 species per cell using combinatorial coding schemes or spectral unmixing methods. However, the maximum capacity of existing combinatorial approaches is limited by the number of spectrally distinct fluorophores (~4-5) and the diffraction limit of light.

To overcome these limitations, we employ super-resolution microscopy [18, 19, 20, 21]. Under a super-resolution microscope, individual molecules can be resolved with a resolution of 10-20 nm.

---

[1] Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of in vivo protein-DNA interactions. Science. 2007 316(5830):1497-502.

[2] Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008. 5(7):621-8.

[3] Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science. 2008. 320(5881):1344-9.

[4] Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science. 2009 Apr 10;324(5924):218-23.

[5] Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science. 1995 Oct 20;270(5235):467-70.

[6] Elowitz MB, Leibler S. A synthetic oscillatory network of transcriptional regulators. Nature. 2000 Jan 20;403(6767):335-8.

[7] Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC. Green fluorescent protein as a marker for gene expression. Science. 1994 Feb 11;263(5148):802-5.

[8] Tsien RY. The green fluorescent protein. Annu Rev Biochem. 1998;67:509-44.

[9] Süel GM, Garcia-Ojalvo J, Liberman LM, Elowitz MB. An excitable gene regulatory circuit induces transient cellular differentiation. Nature. 2006 Mar 23;440(7083):545-50.

[10] Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB. Gene regulation at the single-cell level. Science. 2005 Mar 25;307(5717):1962-5.

[11] Lahav G, Rosenfeld N, Sigal A, Geva-Zatorsky N, Levine AJ, Elowitz MB, Alon U. Dynamics of the p53-Mdm2 feedback loop in individual cells. Nat Genet. 2004 36(2):147-50.

[12] Golding I, Paulsson J, Zawilski SM, Cox EC. Real-time kinetics of gene activity in individual bacteria. Cell. 2005 Dec 16;123(6):1025-36.

[13] Cai L, Friedman N, Xie XS. Stochastic protein expression in individual cells at the single molecule level. Nature. 2006 Mar 16;440(7082):358-62.

[14] Yu J, Xiao J, Ren X, Lao K, Xie XS. Probing gene expression in live cells, one protein molecule at a time. Science. 2006 Mar 17;311(5767):1600-3.

[15] Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA synthesis in mammalian cells. PLoS Biol. 2006 Oct;4(10):e309.

[16] Fan, Y., Braut, SA, Lin, Q., Singer, RH , Skoultchi, AI. Determination of transgenic loci by expression FISH. Genomics. 2001 Oct 2; 71(1): 66-9.

[17] Dickinson ME, Bearman G, Tille S, Lansford R, Fraser SE. Multi-spectral imaging and linear unmixing add a whole new dimension to laser scanning fluorescence microscopy.. Biotechniques. 2001 Dec; 31(6):1272, 1274-6,

[18] Dyba M, Hell SW. Photostability of a fluorescent marker under pulsed excited-state depletion through stimulated emission. Appl Opt. 2003 Sep 1;42(25):5123-9.

[19] Betzig E, et al. Imaging intracellular fluorescent proteins at nanometer resolution. Science. 2006; 313(5793):1642-5.

[20] Rust MJ, Bates M, Zhuang X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). Nat Methods. 2006 Oct;3(10):793-5.

[21] Hess ST, Girirajan TP, Mason MD. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. Biophys J. 2006 Dec 1;91(11):4258-72.

For a typical cell size of 10um^3, this resolution translates into effectively 10^8-10^9 independent volume elements (voxels) per cell. Thus, even highly abundant molecules, such as all of the mRNA molecules in a cell (10^6 molecules per cell[2,3]), can be resolved individually under super-resolution imaging. Each molecular species can then be uniquely labeled with a distinct combination of fluorescent probes. For example, with only nine of the super-resolution fluorophore pairs currently available, over a hundred distinct combinations of quadruplet (${}_9C_4$=126) barcodes can be used. If the spatial ordering of the barcode can also be resolved, a geometrically larger pool of codes is available with a fixed number of fluorophores. By counting the number of times a particular barcode is observed, the abundances of the corresponding molecule can be quantitated directly from a super-resolution image with single molecule accuracy (Fig 2.1). Thus, many molecular species can be simultaneously imaged and quantified in single cells, while preserving the native cellular and inter-cellular context of the molecular interactions.

2.3 Results

To demonstrate the feasibility of this approach, we performed proof-of-principle experiments to detect multiple mRNA species in single *Saccharomyces cerevisiae* cells. We based our approach on the single molecule FISH (smFISH) techniques developed by Singer [22] and refined by Raj and Tyagi [23]. We use smFISH to combinatorialy barcode transcripts, taking advantage of the high specificity of combinatorial labeling with oligo probe sets. As 20mer probes are approximately 7nm in length, regions of the mRNA separated by 100 nucleotides and hybridized by a group of 4-5 probes can in principle be resolved by super-resolution microscopy with its 10-20 nm resolution. By labeling the probes with different fluorophores that hybridize to mRNA in a specific pattern, a spatial nanoscopic barcode can be resolved on each transcript (Fig 2.1). Our approach directly barcodes single mRNAs in comparison to previous multiplex FISH approaches that labeled chromosomal loci [24], [25] and transcriptional active sites [26].

---

[22] Femino AM, Fay FS, Fogarty K, Singer RH. Visualization of single RNA transcripts in situ. Science. 1998 Apr 24;280(5363):585-90.

[23] Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S. Imaging individual mRNA molecules using multiple singly labeled probes. Nat Methods. 2008 Oct;5(10):877-9.

[24] 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 Dec;15(12):5678-92.

[25] Kosman D, Mizutani CM, Lemons D, Cox WG, McGinnis W, Bier E. Multiplex detection of RNA expression in Drosophila embryos. Science. 2004 Aug 6;305(5685):846

[26] Levsky JM, Shenoy SM, Pezo RC, Singer RH. Single-cell gene expression profiling. Science. 2002 Aug 2;297(5582):836-40.

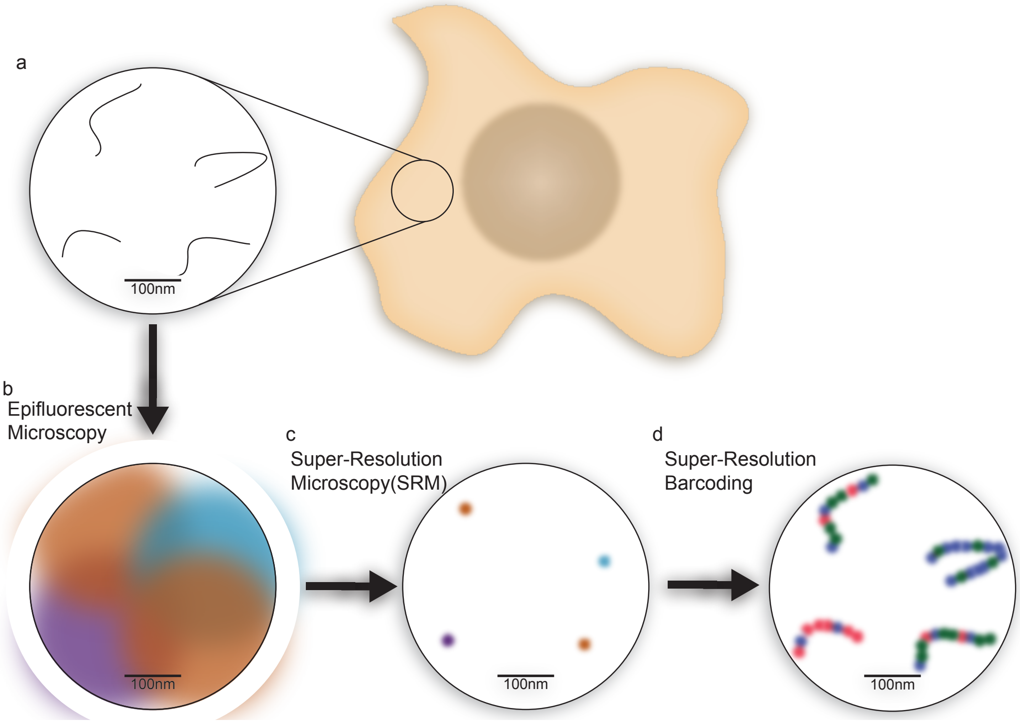


Figure 2.1: Super-resolution and combinatorial molecular labeling allow multiplex identification and quantification of individual molecules in single cells. A-B. Individual molecules are difficult to resolve by conventional microscopy due to the diffraction limit of ˜300 nm. C. Super-resolution microscopy allows spatial resolution of individual molecules. D. The identity of molecules can be uniquely addressed by a super-resolution barcode.

Spatial coding of single mRNAs

In the spatial coding scheme (Fig 2.2) discrete regions of mRNA are labeled with different super-resolution fluorophores, generating a unique spatial sequence. First, we show that individual mRNAs can be combinatorial labeled with conventional fluorophores and their spatial ordering resolved *in situ*. We target the YLR414c mRNA with 3 sets of oligo probes labeled with different fluorophores (Fig 2.2b). These probes are tiled along the mRNA in a 5' to 3' spatially ordered fashion. Hybridized mRNAs appeared as co-localized and diffraction-limited spots (Fig 2.2). We observed that 96±2% (N=29 molecules) of spots co-localized in all three channels of YLR414c probes, indicating efficient hybridization of the probes. If probes were poorly hybridized, then a larger fraction of the mRNAs would only appear in one or two of the channels. We quantified the hybridization efficiency of a single FISH probe to be H_{single} =67.5±9.1*%* (Figure S2.1A,B). The effective hybridization efficiency of the barcode, or the probability that at least one probe is hybridized for each coding position, is $H_{\text{barcode}}=(1-(1-H_{\text{single}})^{R})^{N}$, where R is the probe redundancy at each coding position and N is the number of coding positions. With four probes in each code position, this translates to a 99%

labeling efficiency for each position and a 96% coincidence of all three coding positions, suggesting that individual mRNAs can be readily labeled with a small number of FISH probes and observed with high fidelity.

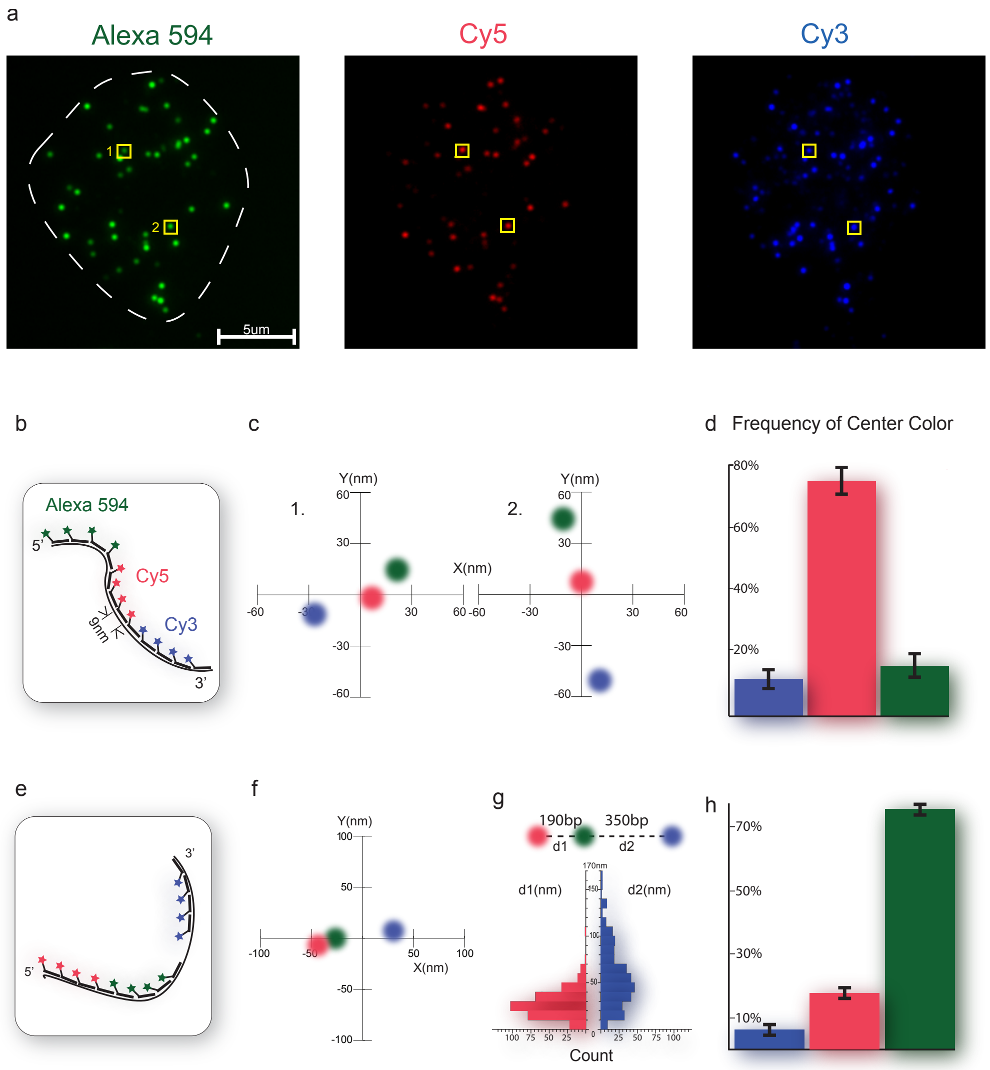


Figure 2.2: Spatial ordering of fluorophores on mRNAs can be resolved by gaussian centroid localization. A. Fluorescence images of YLR414c probes hybridized in a single budding yeast cell, shown in each channel. B. Probe Schematic. Labeled 25mer oligonucleotides are hybridized to YLR414c mRNA. C. Reconstructions of the centroids of spots 1 and 2 following localization by

Gaussian fitting and image alignment. D. The percentage of co-localized YLR414C three-color dots that can be reconstructed in the above image (A) with the correct barcode (n=28, Correct Order=74±8%). E. Schematic of probe-set hybridized to GFP mRNA with different order and distances between the probes positions. F. Gaussian fitting reconstruction of this probe set. G. The distance between the resolved centroids positions (d1=27.93±14 nm, d2= 56±33 nm) is proportional to the intramolecular distance between barcode positions (190 and 350bp). H. The frequency of barcode identification for this probe set (n=327, correct order=76±2%).

Even though each mRNA appears as a co-localized diffraction limited spot, the position of labeled probes can be determined to a much higher resolution by Gaussian fitting of the diffraction-limited signal [27, 28, 29]. We observed the correct spatial order in 74±8% (N=28) of three-color co-localized mRNAs following alignment, independent of barcode order (Fig 2.2d). Error in detecting the order of the barcode (26%) may result from a combination of factors, including localization error, lack of z resolution, and mRNA secondary structure (supplementary discussion). Barcode readout fidelity is unaffected by switching the barcode order (Fig2) and is robust regardless of mRNA identity (Fig S2, S4).

Super-resolution imaging of mRNA barcodes

Barcoding with conventional fluorescence microscopy is only useful in cases where transcript levels are low. When the density of transcripts is high, diffraction-limited fluorescent spots will overlap and make barcode readout impossible. Super-resolution imaging is essential for single-cell microscopy to achieve high-density multiplex barcoding.

To perform super-resolution barcoding, we turned to the Cyanine dye based photoswitchable dye pairs developed by Bates et al [30]. In this scheme an activator dye is placed in close proximity to a Cy5 emitter, which can be switched off by imaging and re-activated by exciting the activator[20]. We designed the oligo-fluorophore conjugate probes such that two probes are used: one labeled at its 3' position with the activator and the other at the 5' position with the emitter (Fig 3a). The probes were designed to hybridize on the mRNA at adjacent positions separated by 2 bp, bringing the dye pairs within 1 nm. This labeling scheme has several advantages: it avoids the high rate of non-specific blinking of covalently linked dye pairs [31] and also increases labeling specificity. As both probes were required for the fluorophore to be re-activated, the background from non-specifically bound emitter probes in the cell was reduced because these probes do not reactivate. We used

---

[27] Thompson RE, Larson DR, Webb WW. Precise nanometer localization analysis for individual fluorescent probes. Biophys J. 2002 May;82(5):2775-83

[28] Yildiz A, Forkey JN, McKinney SA, Ha T, Goldman YE, Selvin PR. Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization. Science. 2003 Jun 27;300(5628):2061-5.

[29] Churchman LS, Flyvbjerg H, Spudich JA. A non-Gaussian distribution quantifies distances measured with fluorescence localization techniques. Biophys J. 2006 Jan 15;90(2):668-71

[30] Bates M, Huang B, Dempsey GT, Zhuang X. Multicolor super-resolution imaging with photo-switchable fluorescent probes. Science. 2007 Sep 21;317(5845):1749-53

[31] Conley N, Biteen JS, Moerner W. Cy3-Cy5 covalent heterodimers for single-molecule photoswitching. The Journal of Physical Chemistry B. 2008 July 28; 112(38): 11878–11880.

the activator-emitter dye pairs to construct our barcodes, with Cy5, Alexa 680, and Alexa 750 as emitters and Alexa 405, FITC, and Cy3 as activators[34]. With this large pool of fluorophores, two schemes of barcode labeling are possible, spatial and spectral, each with distinct advantages (Table 2.1).

Table 2.1: Comparative advantages of SRM barcoding techniques. Spectral Barcoding has a reconstruction fidelity dependent only on the hybridization of fluorophores, does not require linearization and has low photon requirements and can be hybridized in a distributed pattern to increase robustness. These advantages make it an excellent technique for barcoding mRNA and other molecules with unknown structures with the existing SRM dye palette. In comparison, spatial barcoding has a localized hybridization pattern, a lower reconstruction fidelity due to the much higher photon requirements and the tertiary structure of the barcoded molecule. It has the advantage of scaling and may be more applicable to chromesome and splice isoform barcoding. Calculations of multiplex scaling use p as the number of fluorophores and n as the number of positions.

	Spectral Barcoding	**Spatial Barcoding**
Hybridization Pattern	distributed	localized
Spatial Reconstruction Fidelity	100%	74%
Resolution Requirement	100nm	20nm
Minimum required Flurophore Emission	~400 photons	~3000 photons
Linearization required	No	Yes
Multiplex Scaling	p!/(p-n)!/n!	p!/(p-n)!/2

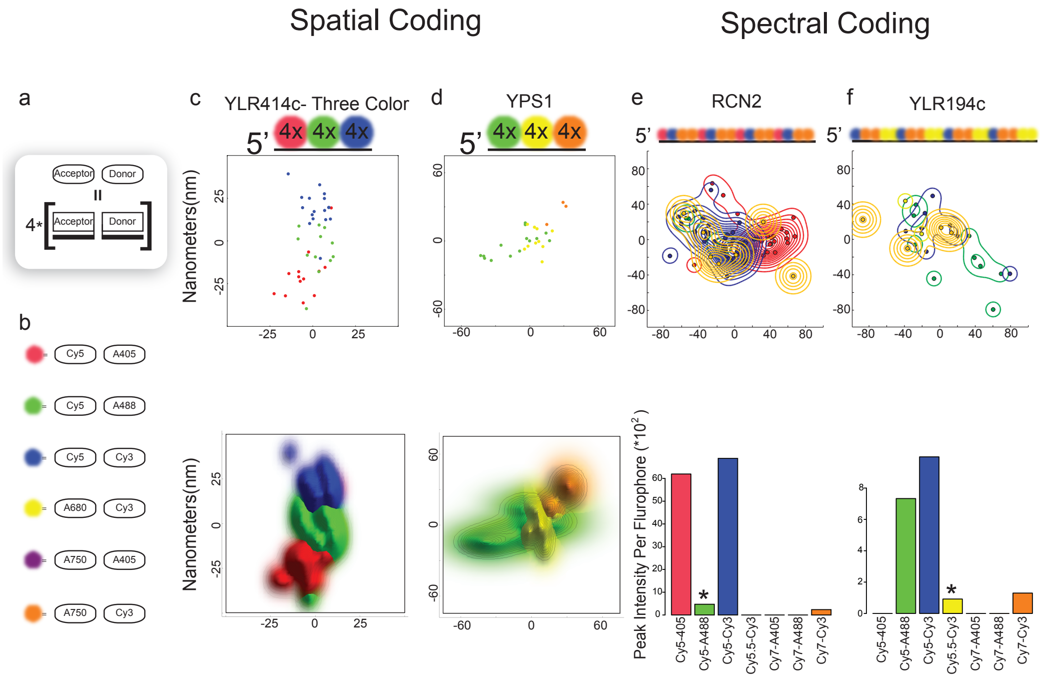


Figure 2.3: Super-resolution imaging enables combinatorial labeling of individual transcripts. A. Schematic of super-resolution barcoding scheme. For each super-resolution color, four acceptor/emitter pairs are hybridized in sequence for redundancy. B. Each barcode color consists of an activator (Alexa 405, 488, and Cy3) labeled oligo adjacent to a 5' emitter (Cy5, Alexa 680, and Alexa 750) labeled oligo. C. YLR414c mRNA 3 position barcode. The order of the probes is shown schematically in the cartoon. A histogram of the super-resolution reconstruction of a single barcode is shown along with a localization scatterplot in which each dot represents an activation of a super-resolution fluorophore. D. YPS1 mRNA 3 position barcodes with 3 different emitters. E. RCN2 mRNA spectral 3 Position Barcode. Probe Positions are scattered throughout the mRNA, enabling robust hybridization and identification by unique combinations of fluorophores. Histograms show the integrated peak intensity for each fluorophore pair in the barcode for all specific activation frames. Bar height in histogram is equal to the number of integrated peak photons detected for each fluorophore pair over time. Cy5-A405, Cy5-A488, Cy5-Cy3, and A750-Cy3, are detected with 6195, 471, 6881 and 235 peak pixel counts integrated over time respectively. Cy5-A488 (green and asterisk) is a false-positive detected due to cross-talk from Cy5-Cy3 (blue), based on the threshold measurements in Fig S3, S4, and is rejected in the barcode determination (supp). Note the A750 based dye pairs give significantly fewer photons than Cy5 dye pairs, but are readily detected with less crosstalk. F. YLR194c mRNA Combinatorial 3 Position barcode. Cy5-A488, Cy5-Cy3, A750-Cy3 and A680-Cy3 are detected with 773, 999, 130, and 92 peak counts integrated over time, respectively. A680-Cy3 (yellow and asterisk) was determined to be a false positive due to the low

amount of photon emission and presence of Cy5-Cy3, a color it can significantly cross-talk with.

We labeled the YLR414c mRNA (Figure2.3a) using Cy5-based super-resolution probe pairs. The effective hybridization rate of SRM barcodes is lower than smFISH because single probe positions require binding of independent probe pairs, giving an the effective hybridization efficiency of $\mathrm{H}_{\mathrm{eff}}=\mathrm{H}^2_{\mathrm{single}}=46\%$ for a single probe pair. To improve the barcode identification rate, we used redundant probe pairs. With 4 probe-pair redundancy, we observed a 3-color hybridization rate of 61±8% (N=85), consistent with the theoretical hybridization efficiency of $\mathrm{H}_{\mathrm{superbar}}=75\%$ (supplementary and Fig S2.1). Hybridization efficiency can be improved by using a higher redundancy in coding: for example, with a 6-probe redundancy, the hybridization rate increase to 92% for a 3-position barcode and 85% for a 6-position barcode.

Of the barcodes that were completely hybridized, 72±10% (N=50) were reconstructed correctly, consistent with the conventional fluorophore results. This rate is independent of mRNAs species (Figure 3 a,b) and consistent with the reconstruction rate of conventional fluorophores. The spatial barcoding scheme necessitates the linearization of mRNA so barcodes can be read with high-fidelity. To do so, we compressed cells between coverglass slips, extending the mRNA through the lateral force of compression. Such treatment can be readily applied to single cell organisms and embryos, but may destroy spatially complex samples (Supplementary Discussion).

The principle drawback of spatial barcoding is its high-resolution requirement. It is difficult to use dyes that emit lower numbers of photons, such as Alexa 680 and Alexa 750, when high (~20nm) intramolecular resolution is necessary. We anticipate that the continued development of improved fluorophores, 3D microscopy, and nucleic acid self-assembly [32] will enable spatial coding to reach its full potential.

Spectral Coding

The spectral barcoding scheme is immediately applicable and robust with current technology. In spectral barcoding (Figure 2.3e,f) the identity of mRNA is coded for by the combination of fluorophores, ignoring spatial order. Spectral barcoding fidelity is completely dictated by the hybridization efficiency and cross-talk amongst probes, unlike spatial barcoding, which has a much higher resolution requirement (Table 2.1). As long as all the colors are present (as determined by the integrated photon counts over the super-resolution imaging cycle) and can be identified above the cross-talk tolerances, the barcode assignment can be confidently made even if the total amount of collected photons is low. This comparatively low resolution requirement enables the use of relatively dim fluorophores, as probes need only to be localized to within the ~100nm area of a single mRNA (Table 2.1). Probes labeled with the same fluorophores can be distributed throughout an mRNA, making spectral coding more robust to heterogeneities in hybridization and partial degradation of

---

[32] Barish RD, Schulman R, Rothemund PWK, Winfree E. An information-bearing seed for nucleating algorithmic self-assembly. Proceedings of the National Academy of Sciences of the United States of America. 2009 Jan 7; 106(15): 6054-9.

mRNA. Additionally, in contrast to spatial coding, molecules do not need to be linearized to be faithfully identified in spectral coding. This potentially permits other molecules, such as proteins, to be multiplexed in single cells under super-resolution imaging.

The drawback of spectral barcoding is that it scales n!/2 times slower than spatial coding, with n being the number of positions in the barcode. However, with the currently available nine dye pairs, a 4-position spectral coding scheme already allows _{9}C_4=126 genes to be multiplexed at the single cell level. Recent work [33] suggests that at least one additional emitter may be available to pair with the activators, giving three additional photoswitchable fluorophore pairs and enabling _{12}C_5=792 genes to be multiplexed in single cells. These characteristics make spectral coding ready for immediate use multiplexing RNAs, with spatial coding holding potential for applications such as barcoding splice isoforms and chromosomes.

Profiling stress response genes in single cells

It has been recently observed that many transcription factors (TFs), particularly those involved in stress responses, activate gene expression in discrete pulses [34, 35]. The TF pulses occur in a wide range of organisms, such as SOS in bacteria [36], Msn2 [37] and Crz1 [38] in budding yeast, and p53[11] and NF-kB [39] in mammalian cells. Since many of the pulses occur stochastically in single cells and are averaged out in population experiments, we set out to determine if bursting in many of the stress response genes are correlated and in doing so ascertain their modes of combinatorial regulation by upstream TFs. We hypothesize that co-regulated genes are more likely to bursts together.

As a proof of principle of our technique, we profiled 32 known stress responsive genes (Table S2.1) in single *S. cerevisiae* cells in response to extracellular Calcium stress. We have previously shown the master transcription factor Crz1 translocates in and out of the nucleus in short (2-3 minute) well-defined pulses in the presence of extracellular Calcium. These pulses occur stochastically in time and involve most of the Crz1 molecules in the cell. In addition to Crz1, calcium stress also triggers the Msn2 pathway, a general stress response regulator that also pulses in its localization[40]. We selected 14 genes that are regulated by Crz1 [40], 5 general stress response genes [41], as well as

---

[33] Dempsey GT, Vaughan JC, Chen KH, Bates M, Zhuang X. Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging. Nat Methods. 2011 Nov 6;8(12):1027-36.

[34] Larson DR, Zenklusen D, Wu B, Chao JA, Singer RH. Real-time observation of transcription initiation and elongation on an endogenous yeast gene. Science. 2011 Apr 22;332(6028):475-8.

[35] Suter DM, Molina N, Gatfield D, Schneider K, Schibler U, Naef F. Mammalian genes are transcribed with widely different bursting kinetics. Science. 2011 Apr 22;332(6028):472-4.

[36] Friedman N, Vardi S, Ronen M, Alon U, Stavans J. Precise temporal modulation in the response of the SOS DNA repair network in individual bacteria. PLoS Biol. 2005 Jul;3(7):e238

[37] Jacquet M, Renault G, Lallet S, De Mey J, Goldbeter A. Oscillatory nucleocytoplasmic shuttling of the general stress response transcriptional activators Msn2 and Msn4 in Saccharomyces cerevisiae. J Cell Biol. 2003 May 12;161(3):497-505.

[38] Cai L, Dalal CK, Elowitz MB. Frequency-modulated nuclear localization bursts coordinate gene regulation. Nature. 2008. 455(7212):485-90.

[39] Nelson DE, et. al. Oscillations in NF-kappaB signaling control the dynamics of gene expression. Science. 2004 Oct 22;306(5696):704-8.

[40] Yoshimoto H, Saltsman K, Gasch AP, Li HX, Ogawa N, Botstein D, Brown PO, Cyert MS. Genome-wide analysis of gene expression regulated by the calcineurin/Crz1p signaling pathway in Saccharomyces cerevisiae. J Biol Chem. 2002. 277(34):31079-88

[41] Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D., Brown, P.O.

13 other aging and stress markers. To ensure that we are observing the products of individual TF pulses [42], we fixed cells under conditions where the average interval between pulses*[38]* was longer than the typical mRNA lifetime.

To label these genes, we used spectral coding of combinations of 3 out of 7 super-resolution dye pairs (${}_3C_7$=35). Barcode assignments are shown in Table S2.1. Cells were imaged by super-resolution microscopy and transcript levels in individual cells quantified based on the abundances of the corresponding barcodes. All genes are assigned barcodes containing combinations of 3 fluorophore pairs. A typical super-resolution reconstruction is seen in Fig S2.4c.

The main concern with multiplexing a large probe set is the crosstalk among the barcodes. This can arise from several sources: fluorescent background from the cell, blinking of nonspecifically bound probes, crosstalk amongst the fluorophores, and errors in computation analysis. We control for each of these sources of errors (supplementary, Figure S2.4d-g). To rule out crosstalk and non-specific blinking, we thresholded the fluorophore assignments in our clusters so only colors more than 3 standard deviations from the measured crosstalk values are counted (supplementary, Figure S2.4d-g). To rule out the possibility that random autofluoresence from cells would skew the measurements, we implemented an imaging scheme containing frequent frames without activation so high-frequency false-activation could be determined (Fig S2.4h). In addition, we intentionally left certain barcodes unassigned and measured the rate of false positives from those barcodes. We measured the frequency of false detection events to be 0.67±0.84 copies per cell, which represents the additive noise floor of our measurements.

To determine if there is significant bias introduced by a particular barcode scheme, we compared the expression levels measured by two different scrambled barcode schemes of the 20 genes with a range of average copy numbers (from 1 to 10 copies per cell) (Table S2.2). Some genes (Cmk2, Npt1) are expressed at levels comparable or higher than the abundant tubulin subunits, as determined by TAP and GFP measurements [43]. The 12 remaining genes with low mean copy numbers (~1 per cell) were not measured since they are within the noise floor of our measurements, making an informative regression with these species unlikely. The two measurements agree with an R^2=0.88 (N=1871 for the standard code, N= 1523 for the switched code. Fig 4c, Table S2) in 19 of the 20 compared genes. This control demonstrates that there was no significant bias introduced in our experiments either from the photophysical properties of different dyes or endogenous background in the cells. As switching scrambles the barcode assignments for genes with different expression levels, the consistent results in this experiment show that barcodes of the high copy # genes do not leak into the low

---

(2000). Genomic expression programs in the response of yeast cells to environmental changes. Mol. Biol. Cell 11(12):4241-4257

[42] Yulei Wang, Chih Long Liu, John D. Storey, Robert J. Tibshirani, Daniel Herschlag, and Patrick O. Brown. Precision and functional specificity in mRNA decay (2002) PNAS Vol. 99:5860-5865.

[43] Huh WK, Falvo, JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea, EK. (2003). Global analysis of protein localization in budding yeast. Nature, 425(6959): 686-91.

copy # barcodes and bias their readings. As an additional check on our barcoding accuracy, we compared the mean expressions level with qPCR as well as single molecule FISH measurements, and obtained $R^2=0.95$ for both experiments (Figure S2.4a,b). We found a 2-fold difference between mean copy number measured with smFISH vs super-resolution barcodes, which can be accounted for by the hybridization efficiency of our probe set and the high threshold levels used to reject false activations. For future probe set designs, a higher redundancy in probes will significantly increase the effective hybridization rate.

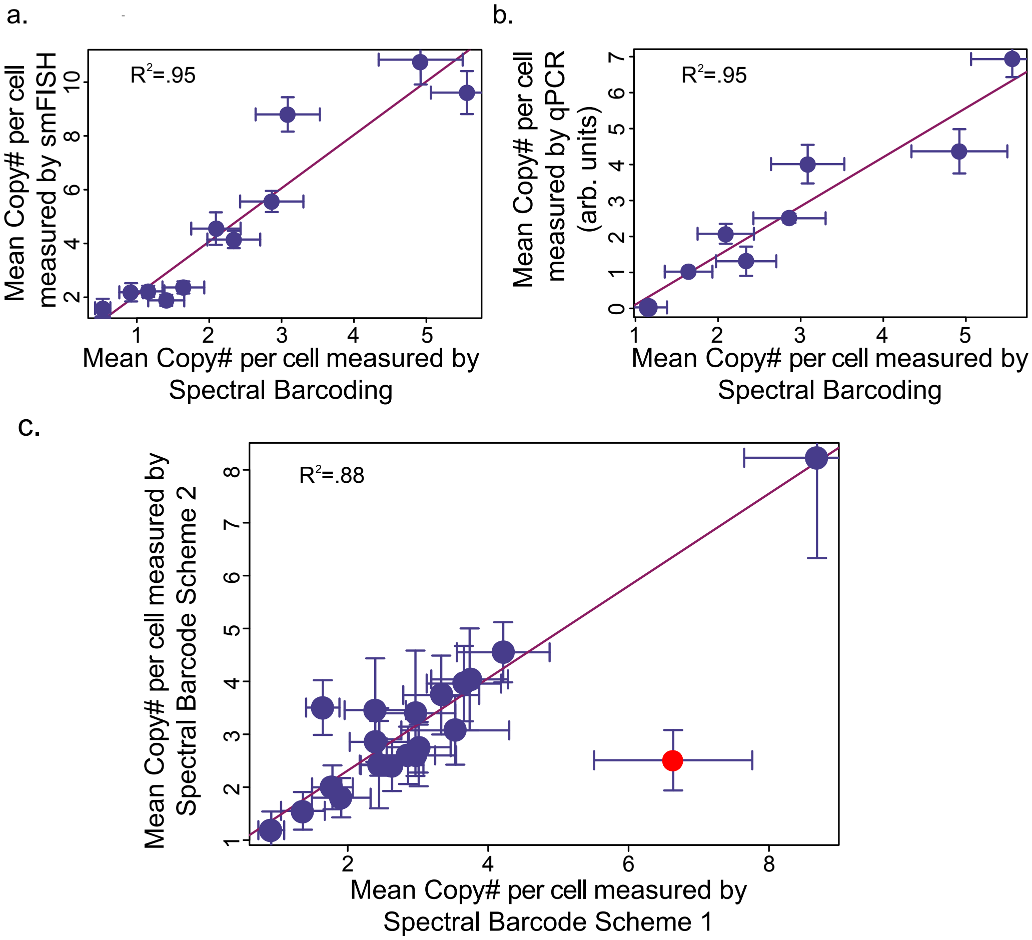


Figure 2.4: Validation of mRNA quantitation by super-resolution barcoding. A). Comparison of smFISH results with super-resolution barcode gives an $R^2=0.95$, with a slope of 2.05. The mean copy number is determined from averaging over a sample size of >50 cells for each gene. The 2-fold difference observed between the super-res barcoding and smFISH results was likely due to hybridization efficiency of the complete barcode as well as the high thresholds used to eliminate false activations. 11 genes were FISHed, including 8 crz1 specific genes, 1 Msn2 target genes, and 2 aging and stress genes. For the qPCR experiment, 8 Crz1 genes were quantified. C. Robustness of mRNA quantitation measured by two different barcode schemes. For twenty genes, we scramble the

barcodes so the same pool of fluorophore combinations would be applied to different genes. Mean copy-number measurements for barcoding schemes are displayed as points in the scatterplot. A regression with an R^2 value of 0.88 was obtained following removal of the one outlier connoted by a hollow circle. The outlier was removed due to its high cook's distance of 2.08226 (Fig S2.10). These measurements indicate that spectral barcoding, regardless of the scheme used, accurately and robustly measures the copy numbers of mRNA in single cells.

Combinatorial regulation in the calcium response transcriptional network

Cells often need turn on the expression of a large battery of genes in response to external stress. To understand this behavior we analyzed the expression of many genes in the calcium stress regulon that respond to pulses of the calcium responsive transcription factors Crz1 and Msn2. In Figure 5, each column represents a single cell and the mRNA abundances measured for each gene under 50mM Ca2+ stress. From this data it is clear that the transcriptional response to calcium varies widely among individual cells.

To characterize the heterogeneity in regulon expression in single cells, we first examine the variability in the abundances of each mRNAs species. The distribution of expression levels for most genes included a low basal state and a long-tailed high-expression mode corresponding to transcriptional bursts (Fig S6), similar to those observed in previous smFISH experiments in yeast [44]. Since these transcriptional bursts are triggered by a TF pulse in our experiment, we can determine whether target genes downstream of the TFs burst synchronous by examining the pair-wise correlation coefficients of their expression levels from the single cell data. We find a range of correlation coefficients from -0.11 to 0.78 (Fig S7), indicative of heterogeneous coordination throughout the regulon. Previous two-color smFISH experiments in yeast found that targets of a constitutively active TF in the galatcose response network [45] as well as metabolic genes [46] were well correlated (R^2=0.5-0.7). Our results show that genes controlled by pulsing TFs within the same regulon can have a large range of correlations values.

Given this heterogeneous regulon expression profile, we next ask whether there are subgroups of the genes that tend to burst together and whether they correspond to different regulatory architecture. By agglomerative hierarchal clustering of the correlation distances between genes with Ward's criterion, we resolved two distinct clusters for the Crz1 responsive genes (p=0.09 and 0.08, respectively) (Fig 2.5). In the first cluster, Crz1 genes were preferentially clustered based on their coexpression with Msn2 genes. The promoter sequences of these genes, such as YLR414c, YLR194c,

---

[44] Zenklusen, D., Larson, D.R., and Singer, R.H. (2008). Single-RNA counting reveals alternative modes of gene expression in yeast. Nat Struct Mol Biol 15, 1263-1271.

[45] Gandhi, S.J., Zenklusen, D., Lionnet, T., and Singer, R.H. (2011). Transcription of functionally related constitutive genes is not coordinated. Nat Struct Mol Biol 18, 27-34.

[46] Silverman, S., Petti, A., Slavov, N., Parsons, L., Briehof, R., Thiberge, S., Zenklusen, D., Gandhi, S., Larson, D., Singer, R., et al. (2010). Metabolic cycling in single yeast cells from unsynchronized steady-state populations limited on glucose or phosphate. Proceedings of the National Academy of Sciences of the United States of America 107, 6946-6951.

Rcn2 and Npt1 contain Msn2 binding sites along with Crz1 binding sites [47]. Genes in the other cluster contain predominantly only Crz1 binding sites, with the exception of Yps1 and Pmc1, which also contain Msn2 sites. The lower average correlations between pairs of Crz1 and combinatorialy regulated genes (R^2 = 0.36, SD=0.16, N=62) compared to those among pure Crz1 targets (R^2 = 0.55, SD=0.17, N=62) or among the combinatorial targets (R^2 = 0.52, SD=0.11, N=62) indicate different sources of inputs on the target genes.

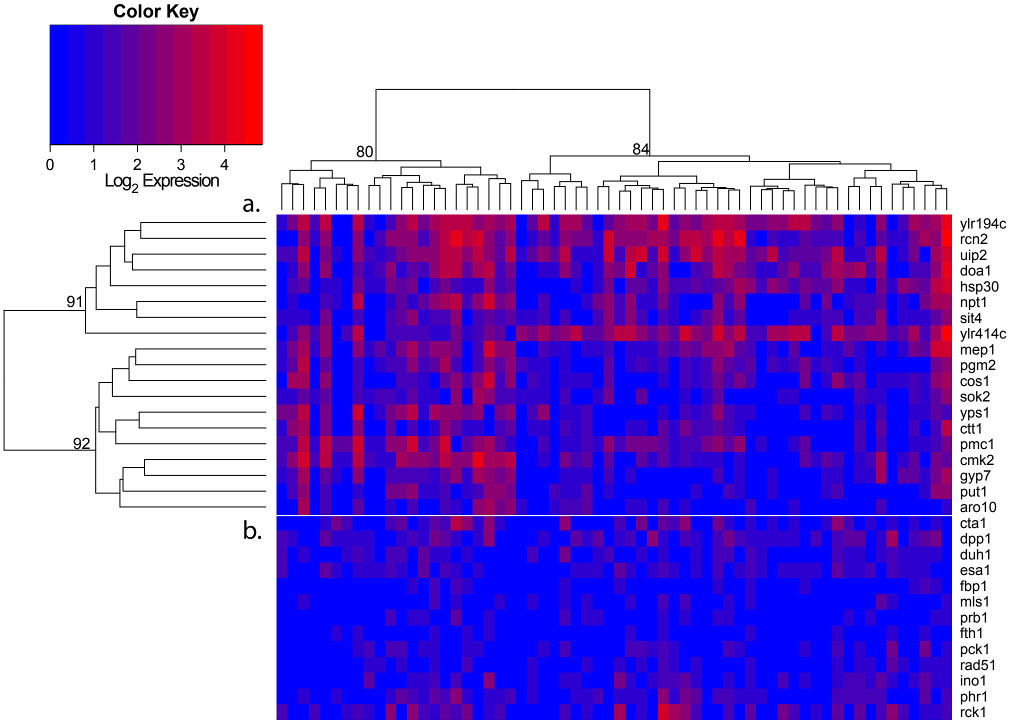


Figure 2.5: Msn2 and Crz1 combinatorially regulate target regulons. All genes in this analysis derive from the data in Fig5a, genes included in Fig5b are not included in these analyses. A. Total number of regulon-specific mRNAs in single cells shown in box plots. Pure Crz1 targets are expressed in WT and in the Msn2/4 deletion cells, but are repressed in FK506 treated cells. WT cells are clustered into two groups (Fig 5a), corresponding to cells with Crz1 or Msn2/4 pulses. Combinatorial targets are repressed in the Msn2/4 deletion as well as in the FK506 treated cells. B.C. Averaged correlation coefficients between pairs of genes in the combinatorial or pure Crz1 target clusters. The detailed pairwise gene expression scatter plots are shown in FigS7. In WT cells (B), each gene cluster correlates well with itself, but is less correlated with the other cluster. In Msn2/4 deletion cells (C), because Crz1 is the dominant input, combinatorial genes are as well correlated to pure

---

[47] Miguel C. Teixeira, Pedro Monteiro, Pooja Jain, Sandra Tenreiro, Alexandra R. Fernandes, Nuno P. Mira, Marta Alenquer, Ana T. Freitas, Arlindo L. Oliveira, and Isabel Sá-Correia, The YEASTRACT database: a tool for the analysis of transcription regulatory associations in Saccharomyces cerevisiae, Nucl. Acids Res., 2006, 34: D446-D451, Oxford University Press

Crz1 genes as to themselves. The distinct expression patterns seen in B suggest that Msn2/4 act as additional input to modulate the expression of combinatorial targets.

Since Crz1 and Msn2 pulse are not synchronized in individual cells[38] at the low calcium (50mM) levels used in this experiment, we hypothesize that the two clusters of cells in the data correspond to cells responding to either a Crz1 pulse or a Msn2 pulse (Fig 2.5, supp). We test this by knocking out the Crz1 and Msn2 pathway separately and visualizing the loss of the associated expression clusters (Fig S8A,B). In cells treated with FK506, an inhibitor of the Crz1 phosphotase Calcineurin, the overall expression level is lower in cells (Fig 2.6a). Conversely, in the Msn2/Msn4 deletion cells, all targets genes should be expressed in response only to Crz1 pulses. This is consistent with the observed coordination between pure Crz1 and combinatorial targets (Fig 6e and Fig S7C), as well as the disappearance of the cell cluster with only combinatorial genes expressed (Fig S8B). The loss of distinct clusters and the homogeneity of coordination throughout the calcium stress regulon in the Msn2/4 deletion cells is indicative of a modulatory effect for Msn2/4 on Crz1 bursts.

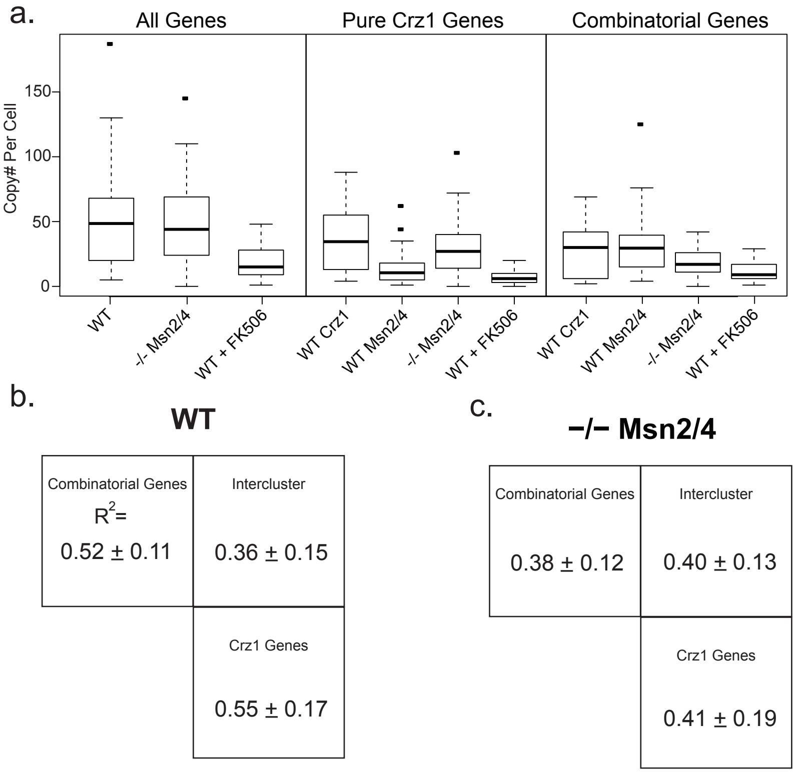


Figure 2.6: Msn2 and Crz1 combinatorially regulate target regulons. All genes in this analysis derive from the data in Fig5a, genes included in Fig5b are not included in these analyses. A. Total number of regulon-specific mRNAs in single cells shown in box plots. Pure Crz1 targets are expressed in WT and in the Msn2/4 deletion cells, but are repressed in FK506 treated cells. WT cells are clustered into two groups (Fig 5a), corresponding to cells with Crz1 or Msn2/4 pulses. Combinatorial

targets are repressed in the Msn2/4 deletion as well as in the FK506 treated cells. B.C. Averaged correlation coefficients between pairs of genes in the combinatorial or pure Crz1 target clusters. The detailed pairwise gene expression scatter plots are shown in FigS7. In WT cells (B), each gene cluster correlates well with itself, but is less correlated with the other cluster. In Msn2/4 deletion cells (C), because Crz1 is the dominant input, combinatorial genes are as well correlated to pure Crz1 genes as to themselves. The distinct expression patterns seen in B suggest that Msn2/4 act as additional input to modulate the expression of combinatorial targets.

These experiments elucidated the combinatorial role of Msn2 and Crz1 on the stress response network and show that single cell transcriptional correlations can be a valuable tool for identifying combinatorial control architecture in regulatory networks.

2.4 Discussion

Under a super-resolution microscope, cells can become virtual microarrays with giga-pixel information density. Since the average size of the reconstructed barcode is 100 nm, a typical yeast cell with a diameter of 5 um can readily accommodate 100,000 barcodes, which is sufficient to accommodate most transcripts in yeast[3]. For larger mammalian cells of 10um^3, up to 10^6 transcripts, each occupying a 100nm^3 volume, can be resolved *in situ* in single cells. As a proof-of-principle, we have demonstrated two barcoding methodologies, spectral and spatial, and shown that by using 7 of the current super-resolution fluorophore pairs, 32 genes can be detected simultaneously in single cells using spectral barcoding. Both spectral and spatial barcoding have the potential to barcode large numbers of transcripts. A back of the envelop calculation shows with just one additional emitter, close to 1000 genes can be multiplexed in single cells using that spectral coding scheme. As multiplex capacity increases exponentially with the number of fluorophores available, the robust spectral coding scheme has the potential to achieve transcriptome level multiplexing (supplementary discussion) if additional emitters can be synthesized (${}_{18}C_6$=18,532).

Super resolution barcoding provides several advantages over existing transcriptional profiling techniques. First, direct imaging of the sample preserves the spatial information both within cells and amongst cells. With the application of light sheet microscopy [48], the technique can be extended into optically thick samples. This advantage makes it a powerful tool in studying signaling in heterogeneous systems such as microbial ecosystems, tissue and embryos, where interactions among different cellular populations play an essential role in cellular decisions. Second, because of the single-molecule and *in situ* nature of the technique, the method is quantitative and avoids intrinsic

---

[48] Cella Zanacchi F, Lavagnino Z, Perrone Donnorso M, Del Bue A, Furia L, Faretta M, Diaspro A. Live-cell 3D super-resolution imaging in thick biological samples. Nat Methods. 2011 Oct 9;8(12):1047-9.

bias in RNA extraction and conversion to cDNA [49]. Lastly, many cells can be imaged simultaneously under a microscope quickly and throughput can be scaled up without significant costs. Compared to the high cost and long waiting time for sequencing single cells, the super-resolution FISH approach is economical to scale up. After the initial cost of the probe set synthesis, the probe set can be hybridized many thousands of times to wild-type and mutant organisms. We envision that super-resolution barcoding at its current throughput can be a useful follow-up to existing high-throughput transcriptomics techniques. After genes of interest are identified by these techniques, they can be monitored with single cell resolution in spatially complex samples with super resolution barcoding.

The combinatorial labeling scheme can be applied to many types of molecules *in situ*. It is a short leap to consider combinatorial labeling of chromosomes and proteins [50, 51], for single cell proteomics and ChIP experiments. We hypothesize that for many types of biochemical techniques, such as microarrays, there is an equivalent *in situ* single-cell experiment possible through super-resolution barcoding, removing the need for spatial separation traditionally performed by gels or dilution on a chip.

Acknowledgement

We credit B. Wold with discussions that led to this work. We thank M. Elowitz for allowing us to carry out a significant number of experiments in his laboratory. We thank A. Eldar, S. Fraser, G.W. Li, J. Levine and J. Locke for discussion and reading of the manuscript. This work was supported by a Beckman Institute seed grant and a NIH New Innovator Award 1DP2OD008530.

Methods

Probes design, Purification and Hybridization. 25mer oligonucleotide probes were designed to match melting temperature whenever possible. FISH probes were designed with 2 base pair spacing between probes to allow efficient reactivation of the pair dyes, often leading to varying TM between probes. Alexa 405, Fluorescein, and Cy3 were used as the activators and Cy5 as the switchable dye. Labeling and purification of the probes followed the protocol at www.singlemoleculefish.com. Yeast cells were grown in minimal media with 50mM Calcium and fixed in log growth phase following the Singer lab protocols [52] with the addition of 0.1% $NaBH_4$ treatment before the ethanol permeabilization step. We found the NaBH4 treatment significantly decreased the auto-fluorescent background of fixed yeast cells. Cells were stored at -20C in Eppendorf tubes and aliquoted out for hybridization experiments. Cells were hybridized with the probes overnight at room temperature in 20% Formamide and 10% dextran sulfate. For all smFISH experiments, 12 probes were hybridized to mRNA in sets of 4. For super resolution barcoding, sets of between 4 and 5, 6 to 7 probe pairs were

---

[49] Tang F, Lao K, Surani MA. Development and applications of single-cell transcriptome analysis. Nat Methods. 2011 8(4 Suppl):S6-11.

[50] * Cho EJ, Lee JW, Ellington AD. Applications of aptamers as sensors. Annu Rev Anal Chem. 2009; 2:241-64.

[51] Agnew HD, Rohde RD, et al. Iterative in situ click chemistry creates antibody-like protein-capture agents. Angew Chem Int Ed Engl. 2009; 48(27):4944-8.

[52] In Situ Hubridization of Yeast Cells. Accessed 20 Sept 09. `http://www.einstein.yu.edu/labs/robert-singer/protocols/insitu_yeast.htm`. Published 21 Jul 98.

used for Cy5, A680, and A750, respectively. After hybridization, cells were washed in Formamide and .2X SSC solution 3 times and imaged.

Imaging. For conventional epi-fluorescence microscopy, images were acquired on an Olympus IX81 with a 100x sapo objective with laser illumination at 532nm, 594nm, and 640nm. Images were acquired with Micromanager software and an Andor Ikon-M DU934 BV CCD. Conventional fluorescence images were acquired in three different fluorescence channels (Semrock zero line filters). The centroids of the FISH PSFs were calculated in each channel and the images were aligned by center of mass alignment of co-localized fluorophore PSFs between channels. This was sufficient for alignment without correcting for rotation and dilation.

Super-resolution imaging was performed on a Nikon TI-eclipse microscope with PFS autofocus lock. Standard super-resolution imaging buffers were used[20]. The imaging lasers, a 640nm laser along with a 30mW 691nm and a 30mW 730nm laser (Coherent Lasers), were brought to the sample through a 100x TIRFM objective. 405nm, 473, and 556 lasers were used as activation lasers and imaging automation was controlled by Micromanager software.

For the spatial coding images, samples were first imaged with only the 640nm laser for 100 frames to switch off Cy5 and to determine the non-specific blinking rate. Then, 100 frames were acquired in each activation channel by co-illumination of the activation and the imaging lasers. This reduced the cross-talk among the different activation channels.

For the spectral coding images, the samples were first bleached in all imaging channels for 6 frames. Samples were imaged in order of the activators, starting at 556nm through 473nm and to 405nm. For each activator, the microscope cycled through imaging with the 730nm followed by either the 640nm or 691nm laser for 48 cycles. Samples are only illuminated with the activation light in the A750 channel, with either the 640nm and 691nm channel imaged next without activation. This imaging pattern minimized photobleaching of A750 while still allowing us to observe adequate activation of the A680 and Cy5 probe pairs. Activation powers were selected to maximize activation rate while avoiding crosstalk among the channels. The 640nm and 691nm excitation lasers alternated after every emitter cycle. This scheme reduces the photobleaching due to the spectral overlap between the Cy5 and A680 emitters. Following this imaging cycle, the 640nm lasers were continuously used to image Cy5 for another 30 frames of specific activation. Throughout the imaging routine, for every four cycles of the imaging lasers, two cycles without the activator lasers were acquired. This enabled us to rule out much of the false positive and nonspecific blinking events in the images. The acquisition sequence is illustrated in Fig S4H.

The activation lasers are controlled by an Arudino microcontroller board and a servo motor shutter. Fluorescent beads (Invitrogen F-8810) were used as fiducial markers to correct for stage drifts. The microscope stages (Prior and ASI) were automated and controlled by acquisition software to enable multi-position imaging. Buffers used in super-resolution imaging followed the protocol in

Bates et al, with Glucose Oxidase as the oxygen scavenger and BME as the reducing agent.

Analysis. Images from diffraction limited and super-resolution experiments were analyzed with a custom written Mathematica script, available upon request. In the analysis, the beads were first aligned to determine the stage drifts. Beads emit on the order of 50,000 photons per image, and could be localized to a few nanometers. Beads close to the cells were eliminated from analysis as the switching of fluorophores in cells can disrupt the bead alignment. Then, fluorophores were selected from each image by intensity thresholding (600 photons for cy5, 400 photons for Alexa 680 and 250 photons for alexa 750), and their centroid calculated by Gaussian fitting. We did not reject activation events that involved multiple fluorophores in the analysis since they came predominantly from single clusters and allowed us to preserve more of the photons collected for the analysis. In principle, multiple activation frames can be compared to neighboring frames to extract localization information. To determine which of the barcode colors are present instead of nonspecific activation, we compare the number of activation events occurring in the specific activation channel versus the activations observed in frames with no activation. If the frequency of nonspecific activation events exceeded that of specific activation events, then that channel is rejected. As several of the activators and emitter channels can crosstalk into other channels, we quantified the crosstalk ratio and reject activation events if they fall below our thresholds (further discussion in the supplementary materials). Switching events that spatially cluster together were grouped to display the resolved barcode. For spectral barcodes, activations were clustered together on a 184nm grid for simplicity since most RNAs are contained within a diameter of 100nm. Activation events near grid vertices were assigned to a neighboring region containing localizations of the same fluorophore pair. For three-color spatial barcodes, the center color was determined by finding the position that was not one of the two localized colors separated by the longest distance. Cell positions were determined by manual segmentation. Barcodes were collected and tabulated for each single cell. Cross-correlation was calculated using the correlation function in Mathematica and the standard errors were calculated from re-sampling the data 100 times.

2.5 Supplementary

Hybridization efficiency

To determine the hybridization efficiency of the probes, we used photobleaching to measure the number of bleaching steps. 12 27mer probes targeting Cmk2 were coupled to Cy3 and imaged with a 532nm laser. Discrete photobleaching steps were observed corresponding to bleaching of single fluorophores (2.7). The average step size was ~300 cts. Using this value as the average fluorophore intensity, we estimated the number of probes bound per mRNA based on the dot intensities in the image before photobleaching. Some variations in intensity are likely due to unevenness in illumination and homo-quenching effects of closely spaced fluorophores. We found that on average 8.1 probes are bound out of the total of 12 probes, suggesting a hybridization efficiency of 67.5±9.1*%* (Figure S1a,b) per probe. The observed distribution is consistent with a binomial distribution with the probability of each probe binding at 67%. Less than perfect hybridization efficiency may be due to the tertiary structure of the mRNA molecule and heterogeneities in bound ligands such as proteins on the mRNA. We demonstrate a more robust spectral coding scheme in which single-colored probes are distributed throughout the mRNA. If occluding molecules are bound to a small region of mRNA, they will only block a subset of the probes in every color, as opposed to removing a single color completely.

This hybridization efficiency implies that mRNAs tagged with four probes have a 99% chance of being detected with at least one probe bound. This is consistent with our observation that 96±2% (N=29) of molecules spatially overlapped in all three channels in the three-color YLR414c probes.

3 color STORM spatial reconstructions.

For the three-color STORM YLR414c centroid reconstruction, we observed that 74±8% (N=28) of codes reconstruct correctly. With the cy3 cy5 pair for super-resolution imaging, this labeling scheme improves background rejection, as both probes are required for the fluorophore to be reactivated, and therefore non-specifically bound Cy5 probes in the cell cannot be reactivated. In comparison, directly labeling oligos with Cy3-Cy5 covalently-linked pairs will have the same non-specific background as standard FISH and have drastically increased blinking rate [1] due to the complex photophysical properties of the Cy3-Cy5 heterodimer. Indeed, we observed prior to inactivation, cells contain a hazy background of singly bound probes in addition to the hybridized FISH spots. After imaging with the 640nm laser for 4-5 frames, these non-specifically bound molecules switched off and blinked at the non-specific activation rate of cy5. Upon activation with 405, 473, or 532nm lasers, these background probes did not reactivate. It is highly unlikely that probes with an activator would be bound close enough (<1 nm) to activate a Cy5. The majority of spots that reactivated were specific mRNA targets, although noise was observed from cellular auto-fluorescence and probe complexes. Some of this noise was due to x-talk among specifically bound dyes. We observed x-talk ratios of around 7% for the most egregious Cy5- dye pairs (Figure S2.4f,g). To ensure proper coding fidelity we only selected identified barcode colors that activated at least three standard deviations above the x-talk ratios. It has been previously reported that 473nm laser can activate cy3-cy5 pairs with 10-20% efficiency. We adjusted the 473nm laser power such that it is higher than the non-specific blinking rate, but less than the power needed to consistently activate Cy3 Cy5 probe pairs.

The cost of this background rejection of using the cy5 pair-probes is reduced effective hybridization rate. As both probes are required for a functional dye pair, the effective hybridization efficiency is $(67\%)^2$=45%. Thus the probability of having at least one pair formed out of a redundant set of four probes pairs is 1-$(0.45)^4$=0.9. With a three-color barcode, the theoretical probability of having all three colors present is 0.9^3=0.72. We observed a 61±8% probability (50 out of 85 reconstructions) that three colors were present on a given mRNA, and a 33±6% probability (28/85) of resolving only two colors. We observed no differences in hybridization efficiency between different mRNAs, and the relative ratios between average expression level were recapitulated in our qPCR and smFISH experiments (Figure S2.4a,b). This data suggests that our STORM results accurately capture the relative ratios of mRNA expression, and can be much improved through increased probe redundancy.

The effective hybridization efficiency can be improved by using more probe pairs. With eight probe pairs per position, the three-color spatial proximity rate is increased to >95%. The long term solution is the development of super-resolution fluorophores with improved contrast ratios. With reduced blinking, fluorophores can be directly labeled to oligos. As in the experiments, only four redundant probes are needed to co-localize in three colors with 96% probability and six colors with 93% probability.

Our use of physical compression allowed us to image most RNAs in a single focal plane. This simple physical treatment permitted us to forgo axial resolution of barcodes. Currently, two approaches are available to improve axial resolution. Some approaches, such as dumbbell shaped point-spread-function [2], can improve axial resolution down to 50nm. Such a resolution is sufficient to resolve overlapping molecules, but won't be high enough to resolve discrete barcode regions on a single RNA. Interferometric PALM [3] and dual objective astigmatic STORM [4] would allow us to reach axial resolutions below 20nm. Such a high resolution would allow us to accurately capture the 3D structure of barcodes, dramatically improving the fidelity of spatial barcoding.

Scaling up multiplexing capacity with spatial and spectral barcoding

Spatial and spectral barcoding schemes have different strategies for scaling up their throughput. Spatial barcoding is efficient. In principle, five-position barcodes allow $9^5/2$= 29525 genes to be tagged simultaneously in single cells. In practice, super-resolution barcode readout accuracy and labeling density are constrained by the non-specific 'blinking' of the Cy5 dyes (i.e., contrast ratio), occurring at 1 in 200 frames per molecule [5]. This means that each diffraction limited spots can only accommodate 10-20 mRNA molecules, limiting the number of total mRNAs resolvable in cells to <10,000. While multiplexing hundreds of genes is feasible with existing fluorophores, further increasing the multiplex capacity will require development of fluorophores that have high contrast ratios and can be directly coupled to oligos to accommodate dense and high-fidelity labeling. In addition, spatial barcoding requires mRNAs to be stretched out to resolve the spatial sequence of colors. We experimented with different fixation conditions and methods to extend mRNAs, but found compressing cells to be the best method for consistently stretching out transcripts. As thick sample are routinely squeezed to reduce optical sectioning for FISH imaging [6], spatial barcodes may be readily resolved in compact and compressible systems such as embryos.

However, not all biological samples can be compressed, such as tissue samples or biofilms. Spectral coding provides an alternative labeling scheme suitable for such experiments. In this scheme, we note that the multiplex capacity increases exponentially with the number of fluorophores available. In principle, cyanine dyes can be extended further into the infrared region to act as additional emitters. Recent works [7] suggest that two additional emitters may be already available, allowing potentially close to a thousand ($_{12}C_5=792$) genes to be multiplexed, which is likely close to the limit imposed by the contrast ratio of the dye pairs.

Terminology used to describe STORM Imaging

Throughout the paper we refer to three sources of STORM activation, specific, non-specific, and false. We have used this terminology as it provides a direct analogy to potential sources of error in our imaging routine. Specific activation refers to the desired activation of the intended dye by excitation of the activator dye. Non-specific activation refers to excitation of the emitting dye due to absorption of the imaging light, often referred to as blinking. False activation refers to activation of the emitter due to absorption of the wrong wavelength by the activator dye.

Crosstalk and accuracy of the barcode readout.

The spectral coding approach is more robust because errors associated with identifying spatial positions can be avoided. However, crosstalk among different fluorophores can impede the identification of the proper barcode and result in leakage amongst the barcodes. To control for crosstalk, we performed several control experiments.

First, we imaged individual dye pairs with the full imaging routine, going through all activators and emitters, to quantify the amount of leakage from each dye pair into the others. By examining all seven dye pairs used in our study, we found that the most leakage occurs from Cy3 activators, which can be activated by the 473 and 405 nm lasers. However, Alexa 488 and Alexa 405 cannot be activated with the 555nm laser, so the crosstalk only appears in one direction. From the single dye pair experiments we quantified the idealized x-talk with 12 of the target dye pairs, imaged in exactly the same routine as our barcode quantification. There is a small amount (~1%) of non-specific activation in Alexa 405 and 473 with 556nm activation, due to non-specific blinking of the dye pair[6]. The probe pairs that can be x-talked to A488 and Cy3, X-talk at a higher rate than previously reported (6-6.5%) (Figure S2.4f,g). This may be due to the close proximity of the dyes to each other in our probe design. This x-talk was still clearly separable from signal. For A680, we only use the Cy3-A680 pair, and thus no crosstalk between A680 dyes can occur. To reject the false activation of cy3 by 473 laser, we discard activations in the 473 channel that are less than 30% of the activations observed in the 555nm laser channel. Similarly, we set the threshold for rejection at 15% and 30%, respectively, for 405nm activation of A488 and Cy3. In addition, there is crosstalk between the cy5 and A680 emitter channels. Since this crosstalk only occurs in the cy3 activation channel, we compare the activation intensity in cy3-cy5 vs cy3-A680 channels and found about 12%

crosstalk between the 2 emitters. Thus, any activation in A680 that is less than 30% of the intensity in cy5 is rejected. All of these thresholds are more than 3SDs from the measured cross-talk values (Figure S2.4f,g). We confirmed that the tertiary structure of mRNA did not result in any probe crosstalk. We observed no false activation from probe positions not directly proximal to the emitter probes, in agreement with the photoswitching distance of ~1 nm.

Second, to test the accuracy of the three color barcode readout, we used the barcode that is the most prone to crosstalk, which is the cy3 activator paired with all three emitters. This probe set against yps1 is hybridized and the correct barcode is picked up, while several barcodes are also observed at 20% crosstalk rate. However, this represents the worst-case scenario for crosstalk, since Cy3 can be activated by both 405 and 473nm lasers. In addition, the gene targeted with this probe set is a relatively low copy number, so false barcodes due to cellular background and nonspecific blinking appear at relatively higher frequency compared to the correct barcode (Figure S2.4e). A different three-color barcode with 405, 488, and cy3 as activator and cy5 as emitter shows a much lower crosstalk ratio (Figure S2.4d). Most of the extraneous barcodes observed are due to background blinking in the cell and do not scale with the copy number of the genes probed. Thus, there is a constant background of barcodes that is additive but not multiplicative to the real barcodes.

Third, when we analyzed data for the full dataset with 32 genes, we examine the frequency of observing the barcode position that was not coded. With a total of 35 possible coding positions in our current scheme, there are three empty code positions which should not show up. The mean false identification frequency is 0.67±0.84 copies per cell, suggesting that our entire barcode set imaged simultaneously is not significantly affected by false positives. In addition, we performed analysis on the full data set with a single gene barcode dropped out, and we observe that the empty position which is normally present at 4.9±2.3 copies per cell is now present at 0.75±0.84 copies per cell, indicating small amounts of crosstalk into that position from other barcodes.

Fourth, we took a 20 gene probe set containing cy5 and A750 emitters, and circularly permuted the activators (405->488, 488->cy3, cy3->405). This effectively scrambled the barcode assignment since the emitters remained in the same position. We observed a strong correlation between genes measured amongst both probe sets, indicating that no significant bias is introduced by a particular assignment of the barcode (Figure S2.4c). One significant outlier existed in our analysis, ylr194c. This outlier was dropped based upon its high Cook's distance of 2.08226 (2.26). A regression with an R^2 value of .88 was obtained following removal of the one outlier connoted in red. The other large outlier with a high copy number also has a high cook's distance of 3.5515. Despite its leverage, the point fits well on a regression where both of the aforementioned points are dropped, so it was retained in our analysis.

Fifth, we performed single molecule FISH experiments measuring the expression of 11 genes, including 8 crz1, 1 msn2, and 2 aging genes. We observed a R^2=0.95 correlation between the mean levels measured by smFISH and the barcode approach (Figure S2.4a).

Sixth, we also performed q-PCR experiments measuring the mean copy number of 8 crz1 target genes. We observed a correlation of .95 between the qPCR and the barcode data. The qPCR experiments were performed in triplicates and quantitated using 1x, 10x, and 100x serial dilutions (Figure S2.4b). The PCR amplication efficiency for each gene was determined from linear fitting of serial dilutions.

Spatial capacity of super-resolution barcoding

The size of mRNAs that we reported is approximately 100nm (Figure S2.2a-d). At this size, a typical cell of 10um^3 can fit in about 10^6 of these 100nm^3 volume elements, sufficient to accommodate a significant fraction of the transcriptome simultaneously in the cell. In the reconstruction shown in Figure S2.4c it is apparent that there is at least 20x more space left to accommodate additional barcodes. With the implementation of 3D storm, there will be 10x more room in the axial direction. In mammalian cells, the much larger available space, 20-50x larger than yeast, will compensate for the higher transcript levels (also 20-50 fold).

These calculations do depend on the assumption that transcripts are homogenously distributed throughout the cell. In our observations, and in that of other authors using smFISH in eukaryotes, a homogeneous spatial distribution of RNAs outside of the nucleus is usually observed. If certain transcripts are found to cluster, they can be removed from the large dataset and analyzed separately. We imagine such a combination of techniques should cover the vast bulk of *in situ* transcriptional profiling experiments.

In the case of the nucleus, we and other authors have found dense transcriptional sites to be rare in FISH experiments. In yeast, we found transcriptional sites to be in <5% of cells at any given time. Given the density of very highly expressed genes and the large size of the mammalian nucleus, we expect that most of these sites can still be identified and barcoded using our technology. In the cases where the nuclear density is too high to distinguish the transcripts of interest, we can still use the intensity ratio to infer the most likely barcode and assign its abundance to transcriptional active sites.

In addition, a recent paper from the Tyagi group [8] showed that mRNAs in neurons are packed singly into transport granules separated by at least 200nm, rather than compacted in dense granules as previously thought. While non-uniform distribution of mRNAs in cells could in principle pose a challenge to the barcode readout, few existing smFISH experiments in eukaryotes suggest that the spatial distribution of transcripts will be a problem.

In our current 32 gene multiplex experiment, we need super-resolution imaging to spatially separate transcripts, as shown in Figure S2.4c. In this figure, the axis units are pixels, which on our setup correspond to 130nm. A diffraction limited spot would span 6-7 pixels end-to-end. One can see that the average RNA is confined within 1 pixel. Many mRNAs reside within one diffraction limited region. With FIONA imaging, the neighboring barcodes would be almost impossible to distinguish.

Our single cell profiles cover a wide gamut of copy #'s, demonstrating the wide dynamic range of SRM barcoding. Several of the 32 genes we profiled are highly expressed, on the level of major structural proteins. For example, Cmk2 is expressed at 7500 copies of proteins compared to 5000-10k copies for tubulin, as measured by Weissman and O'Shea [9] using quantitative westerns and GFP. Many other genes, such as YLR414c and YLR194c, are expressed at similar levels (2.20). The average mRNA copy number might seem low because the expression is bursty and heterogeneous among cells, but is often not low in individual cells. In our experiments Cmk2, YLR414c, and YLR194c were additionally induced by extracellular calcium, so we expected the protein copy # to be well above these reference measurements.

The available space should be more than sufficient to accommodate ~1000 mRNAs at a time. We imagine the typical use case of single cell barcoding will follow-up traditional high-throughput techniques such as RNA-seq, so many uninteresting highly expressed housekeeping genes will not be probed in the barcoding experiment. By measuring highly-multiplexed gene expression *in situ* with spatial information, our technique will make a significant contribution to systems biology not possible with existing technologies.

Single Cell Profiling and Correlation

We measured the copy number of probed genes by tabulating the barcode reconstructions in single cells (barcodes in 2.2). We did so for 62 cells. On average 2-3 cells were observed per field of view. Each set of STORM images took ~10 minutes to acquire; the imaging schematic is shown in 2.17. We manually found the positions of 10-15 cells and set up an automated stage to scan through the positions for STORM imaging. It takes approximately 5-6 hours to collect 60 cells. The field of view (FOV) was kept small to reduce auto-fluorescent background from glass. In principle, quartz slides can decrease background and allow the FOV to expand, dramatically increasing throughput.

Additional discussions of analysis protocol

For this proof-of-principle we performed a simple analysis that gridded the data into a fixed width to initially find barcodes. After barcodes were found, the grid was removed and a local center of mass calculation was used to find the center of a barcode cluster. A fixed grid size of 184nm was used for simplicity, since most barcodes fall within a 100nm diameter. In the future, more intelligent clustering algorithms will be needed to identify barcode clusters. We note that this is a well-studied problem in computational image analysis and a variety of tools are available to address this problem, but they are beyond the scope of the current paper.

In our current analysis protocol, we do not exclude multiple activation events since they are infrequent in our movies and largely constitute emission from a single RNA. We are able to use more of the photons collected in the experiment to determine barcodes more accurately. However, for denser samples, a more stringent application of thresholding, or an application of the ideas in the photobleaching approach (Selvin 2005, and Lippincott-Schwartz, PNAS 2011) and fluctuation imaging (SOFI) approach (Weiss, Optical Express 2010) can be implemented for more efficient reconstructions with minimum information loss. These approaches do not require single activations, but rather use intensity fluctuations, for super-resolution imaging. Incorporating them in our future analysis protocol will result in more efficient utilization of photons and better barcode reconstructions. Although these methods are not yet implemented, there is no fundamental limitation to incorporating them. Our current protocol was optimized to resolve a multiplex of 32 barcodes from each other, as demonstrated in 2.15. We anticipate this protocol will carry over for multiplexing $<$100 genes. We expect to incorporate the additional changes in the next iteration of barcoding when we go above 100 genes.

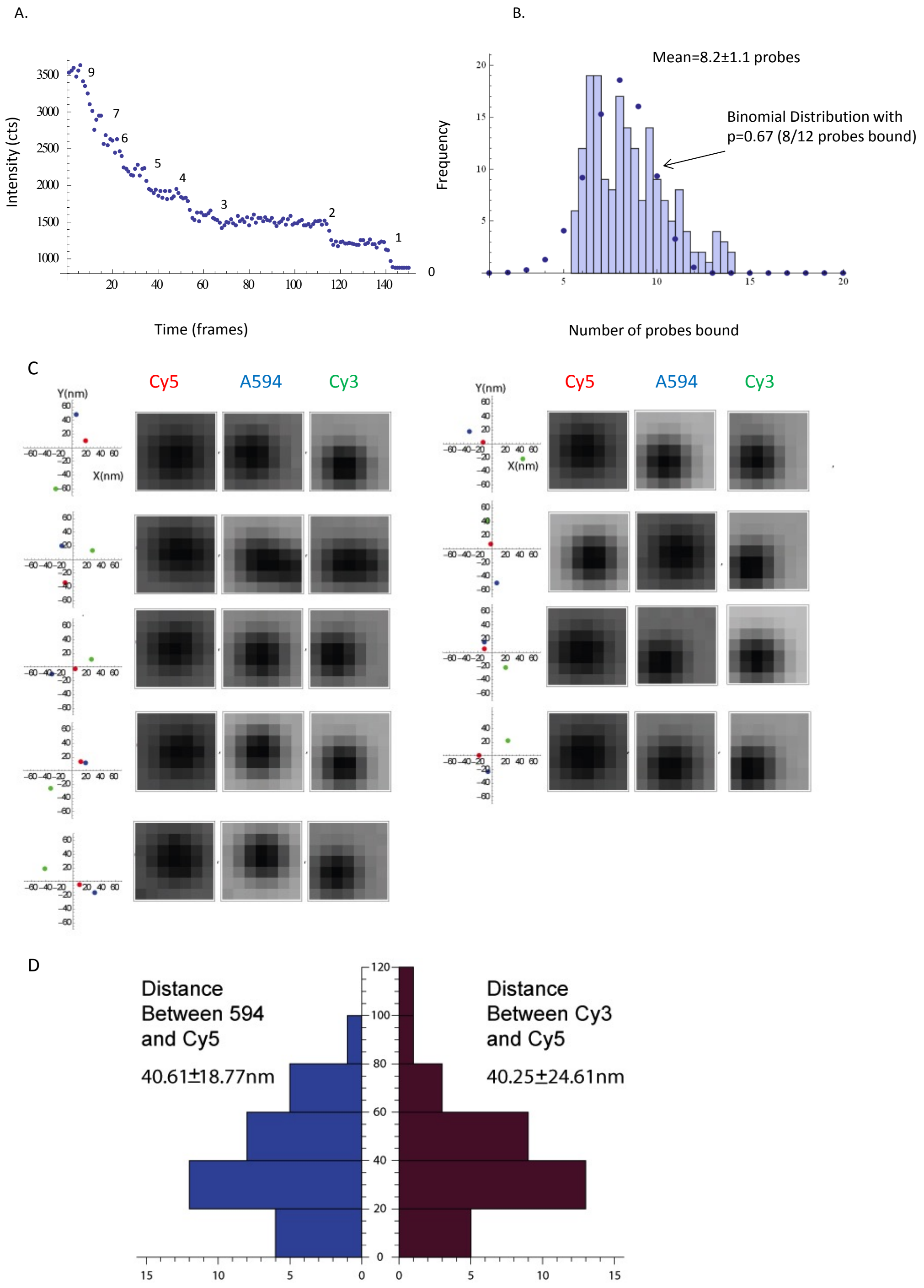


Figure 2.7: A) Sample photobleaching traces. Cmk2 mRNA was hybridized with 12 27mer probes labeled with Cy3. The sample was illuminated with a 532nm laser for 150 frames. No antibleaching buffer was used. Stepwise drop in fluorescence intensity corresponds to photobleaching of a single fluorophore. The intensities of fluorophores were not uniform, possibly due to micro-environment and homo-FRET quenching. On average each step corresponds to ~300 cts, with a background of

~900 cts. The initial intensities suggest that in both traces, 8-9 probes out of the 12 probes were bound to the mRNA, corresponding to ~2/3 hybridization efficiency for each probe. B) Distribution of hybridization efficiencies for the Cmk2 probe set. The number of probes bound is determined from the initial intensities of dots observed prior to photobleaching divided by the average step size. The mean number of probes bound was 8.2±1.1. This distribution is overlayed with a binomial distribution with probability of each probe bound at 67%, corresponding to 8 out 12 probes bound on average. C. FIONA reconstructions of barcodes on YLR414c mRNAs in a single cell in Fig2. 5' modified YLR414c probes were used. The intensity profiles of the dots in each channel are shown in the right panels, corresponding to Cy5, 594, and Cy3 channels. The reconstructions from Gaussian fitting of the intensity profiles are shown in the left. mRNAs with intensities above thresholds in all three channels are selected in order to increase localization accuracy. D. Spatial separation between terminal and center positions of the barcode. The distances between the Alexa594 -Cy5 and Cy3-Cy5 probe positions was both ~240bps, reflected in the symmetrical mean physical distances observed.

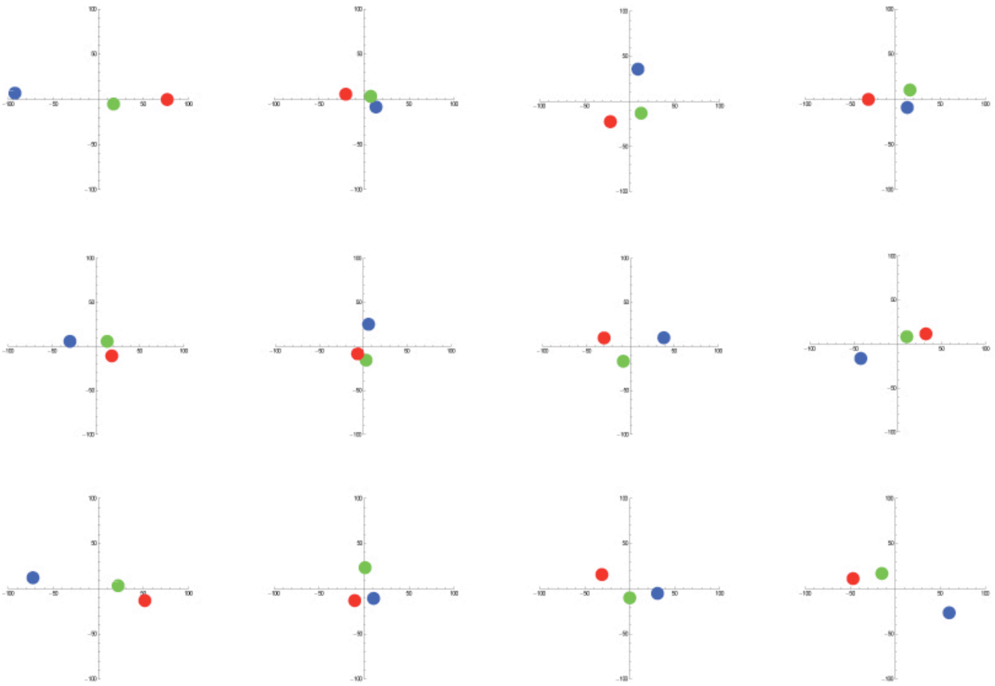


Figure 2.8: Gaussian fitting reconstructions of MDN1-GFP (14kb) mRNA tagged with the barcode Red-Green-Blue at 3' end 1kb region. Axes are in nm. Reconstructions are taken from all of the dots from one image. Reconstruction rate is 92±4.4%.

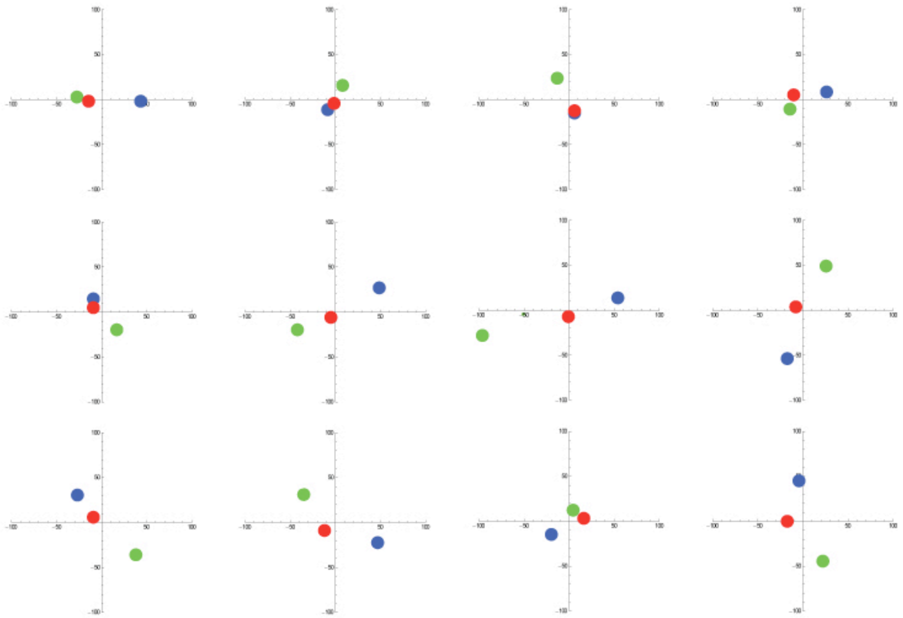


Figure 2.9: Centroid reconstructions of MDN1-GFP mRNA tagged with the barcode Green-Red-Blue, order switched from S17a. Axes are in nm. Reconstructions are taken from all of the dots from one image. Reconstruction rate is 72±7.5% of the time.

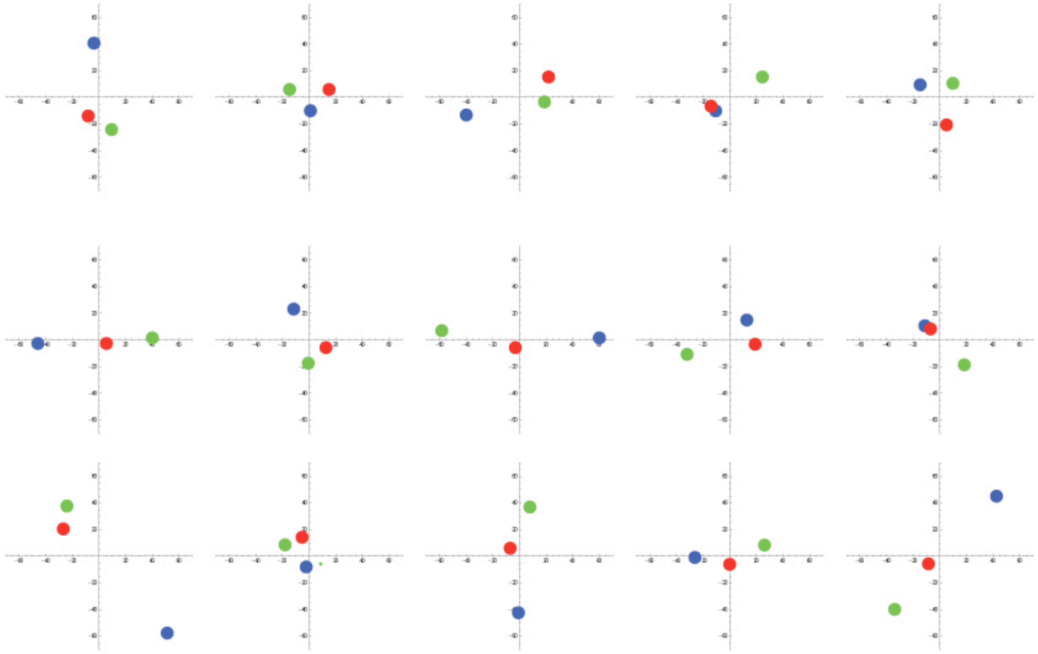


Figure 2.10: Centroid reconstructions of YLR414c mRNA (700bp) tagged with the barcode Green-Red-Blue.

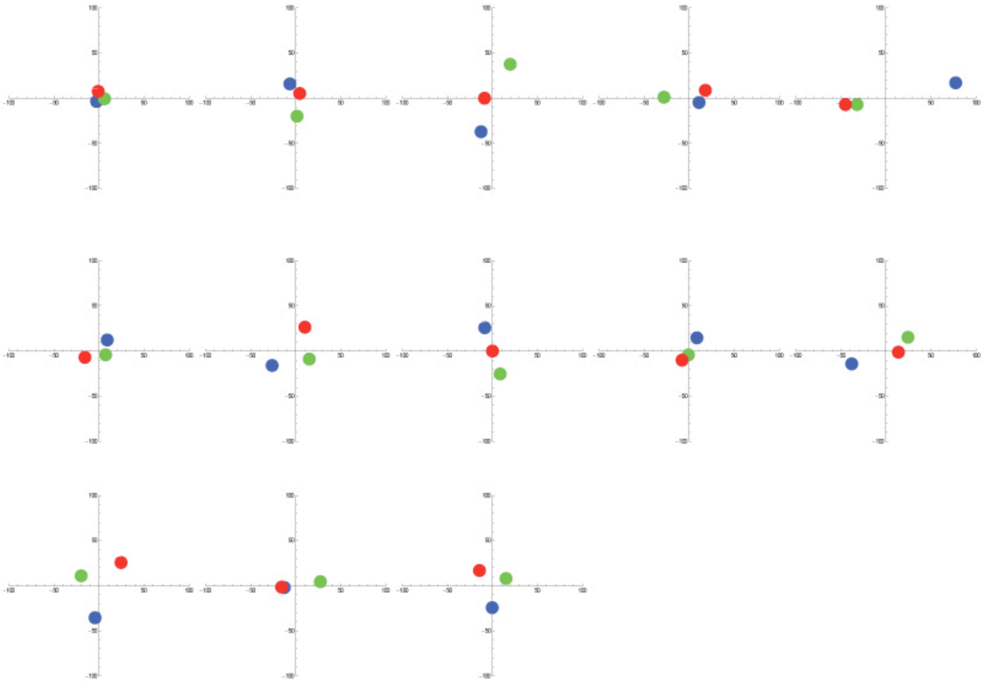


Figure 2.11: Centroid reconstructions of YLR414c mRNA tagged with the barcode Green-Red-Blue, but with cells that are not compressed. Note that the order of the barcode is scrambled. In compressed cells the YLR414c barcode identification rate was 77± 7.3%. In the un-compressed cells the YLR414c barcode identification rate fell to 34± 9.3%, which is essentially random.

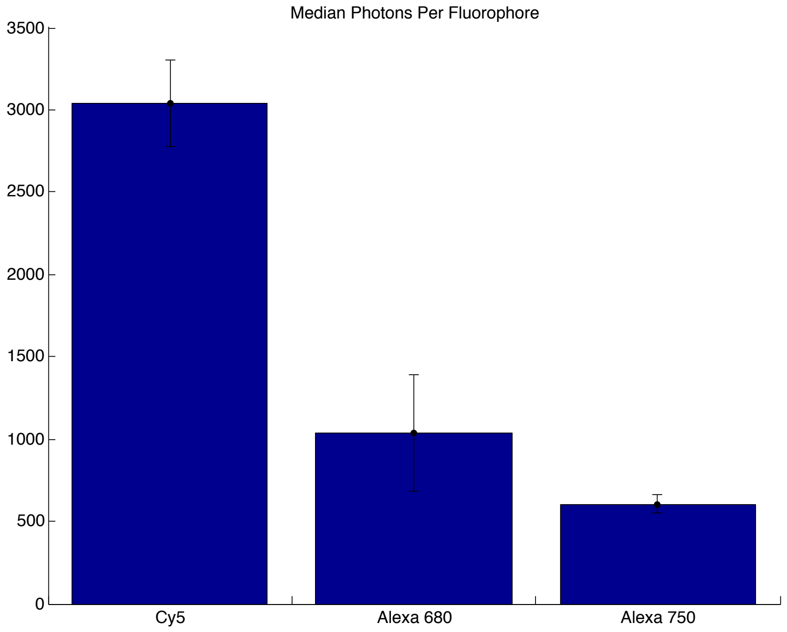


Figure 2.12: Median Photons per flurophore. Error bars determined by the 1000 replicates of the bootstrap standard deviation. SD of 160nm for the PSF, Cy5, A680, and A750 should give a localization accuracy of 2.9, 5.7, and 6.8nm. In practice the localizaiton accuracy was significantly lower than this due to the additive error of aligning multiple fluorescent beads(SD ~5-15nm). N=197, 71 and 250 for Cy5, A680, and A750.

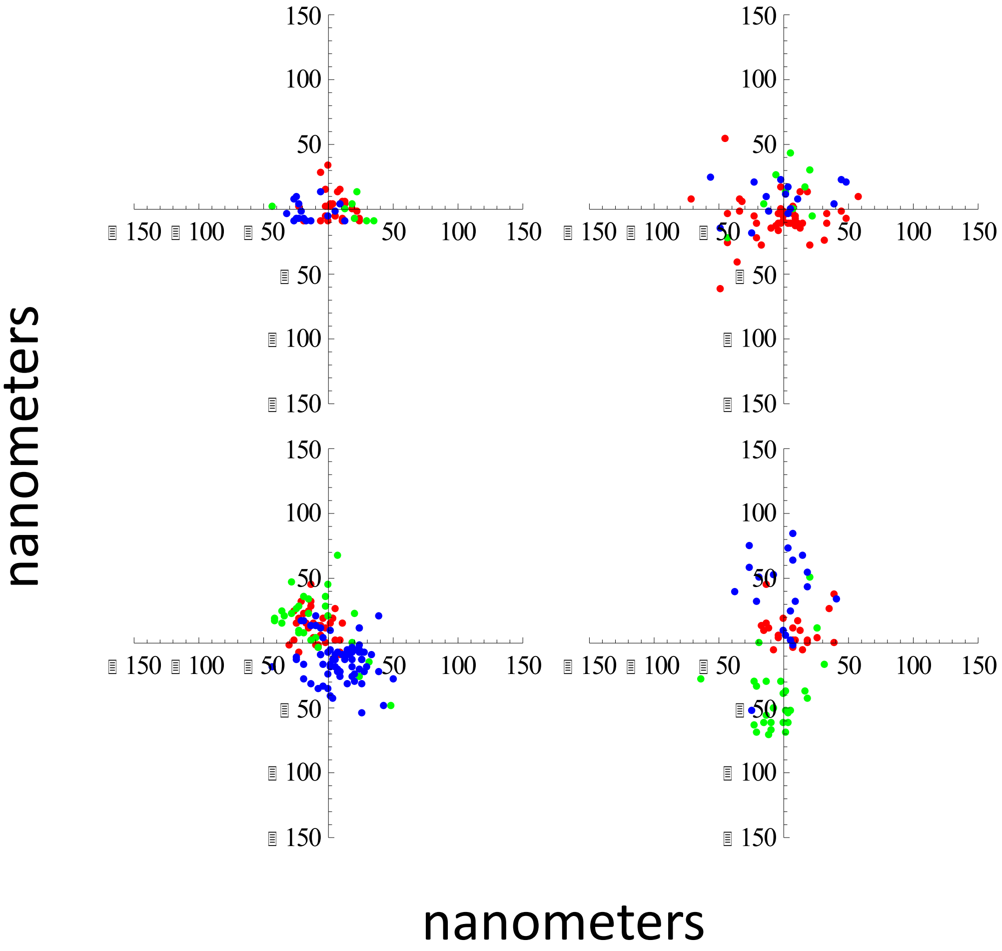


Figure 2.13: Single cell STORM reconstructions of PMC1 three-color barcodes (Blue-Red-Green).

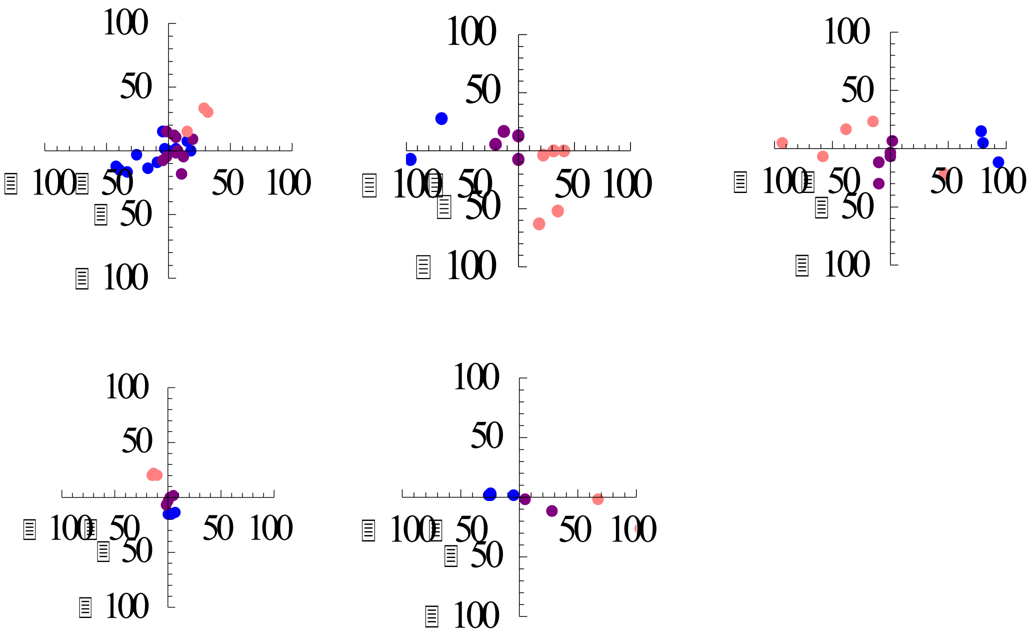


Figure 2.14: Single cell STORM reconstructions of Yps1 3 color barcode with Cy3 paired with Cy5 (blue), A680 (purple), and A750 (pink). The order is pink-purple-blue. Axes are in nm.

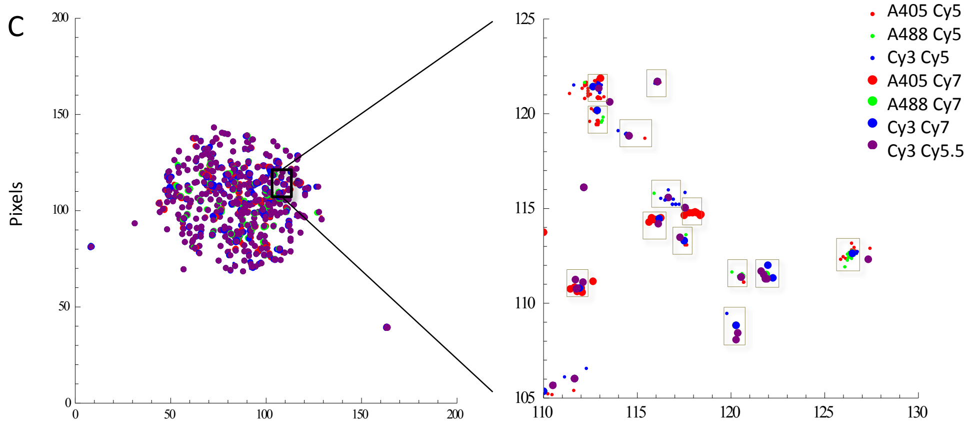


Figure 2.15: Reconstructions of barcode labeled mRNAs in single cells. Each pixel corresponds to 130nm. The right panel shows a zoomed plot of a region in the cell. Individual mRNAs are shown in boxes.

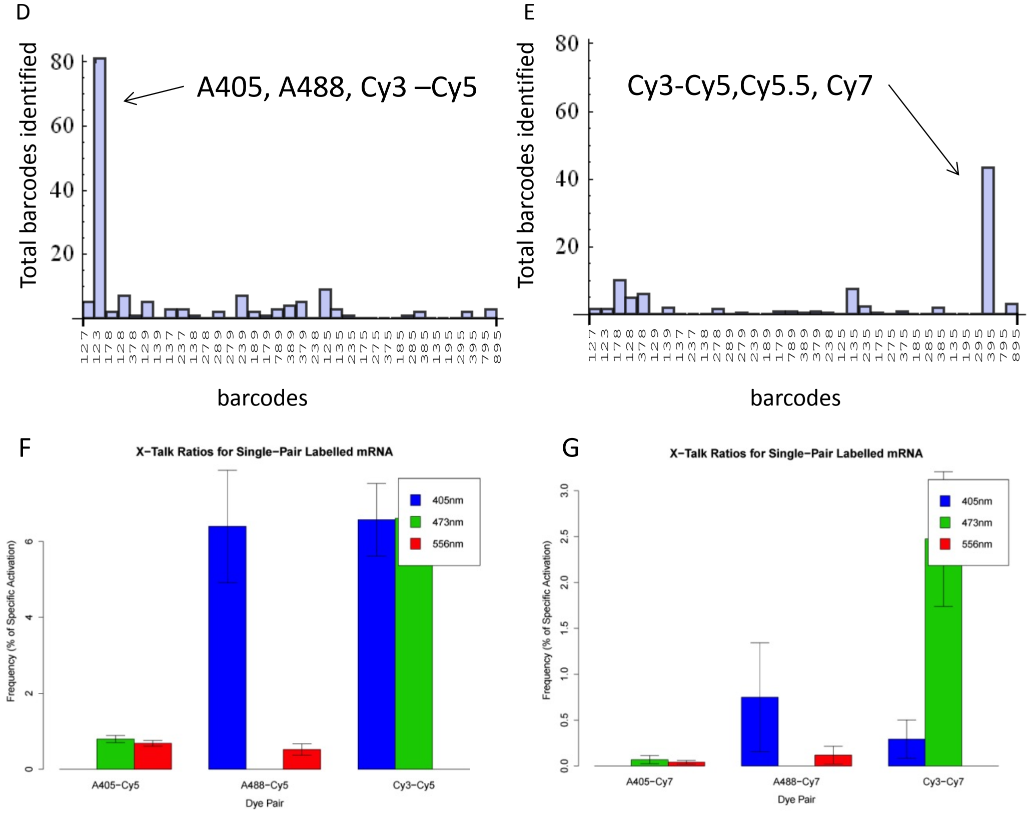


Figure 2.16: Barcode crosstalk measurements. Three-color barcode is hybridized and imaged. The leakage of that barcode into other barcodes is shown on the histogram, representing the errors in detection and analysis. A total of 20 cells are counted in each case. A) a barcode with Cy5 emitters and all three activators, hybridized against YLR414c. B) the worst case scenario, with Cy3 activators and all emitters hybridized against YPS1 which is present at lower abundances than YLR414c. Because Cy3 can be activated by 405 and 473nm lasers, there is more crosstalk into those channels. We observe that there is a relative low level but uniform background of barcodes observed due to autofluorescence in the cells and nonspecific blinking events. This background is additive to the barcode quantitation and does not scale with the copy number of the genes. F-G. Single dye pair crosstalk ratios. 12 probe pairs are hybridized against YLR414c coupled with each combinations of fluorophores. Then the false activation rate in different STORM channels are measured for Cy5 (F) and A750 (G) emitters. Crosstalk from Cy3-Cy5 into Cy3-A680 is 11.6% and negligible in the reverse direction.

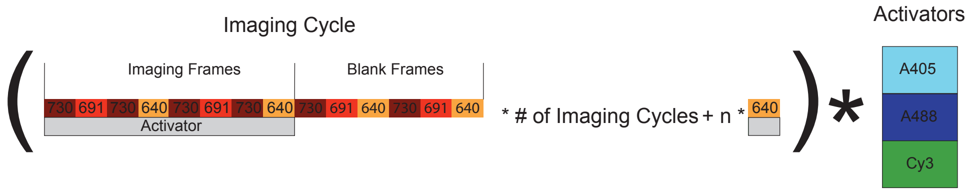


Figure 2.17: Imaging Scheme. For each activator we sequentially went through one cycle consisting of looping the pattern of imaging frames followed by blank frames 20 times. We followed this pattern by imaging just Cy5 for two blank frames every five frames until cy5 was completely exhausted after 30 additional frames.

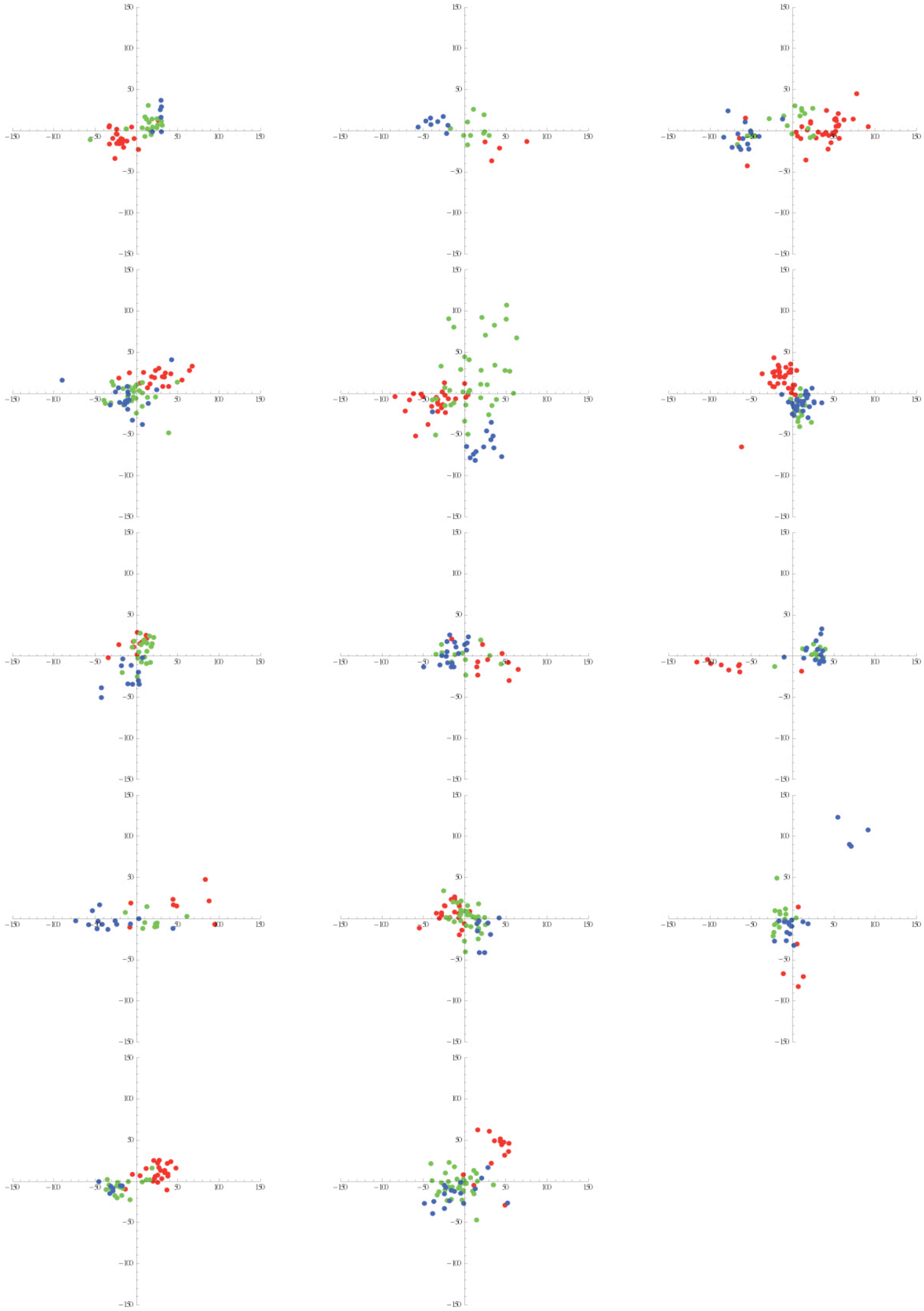


Figure 2.18: STORM reconstructions. B. Each barcode color consists of an activator (Alexa 405, 488, and Cy3) labeled oligo adjacent to a 5' emitter (Cy5, A680 and A750) labeled oligo. The axis for each figure go from -150nm to 150nm in the X and Y dimensions.

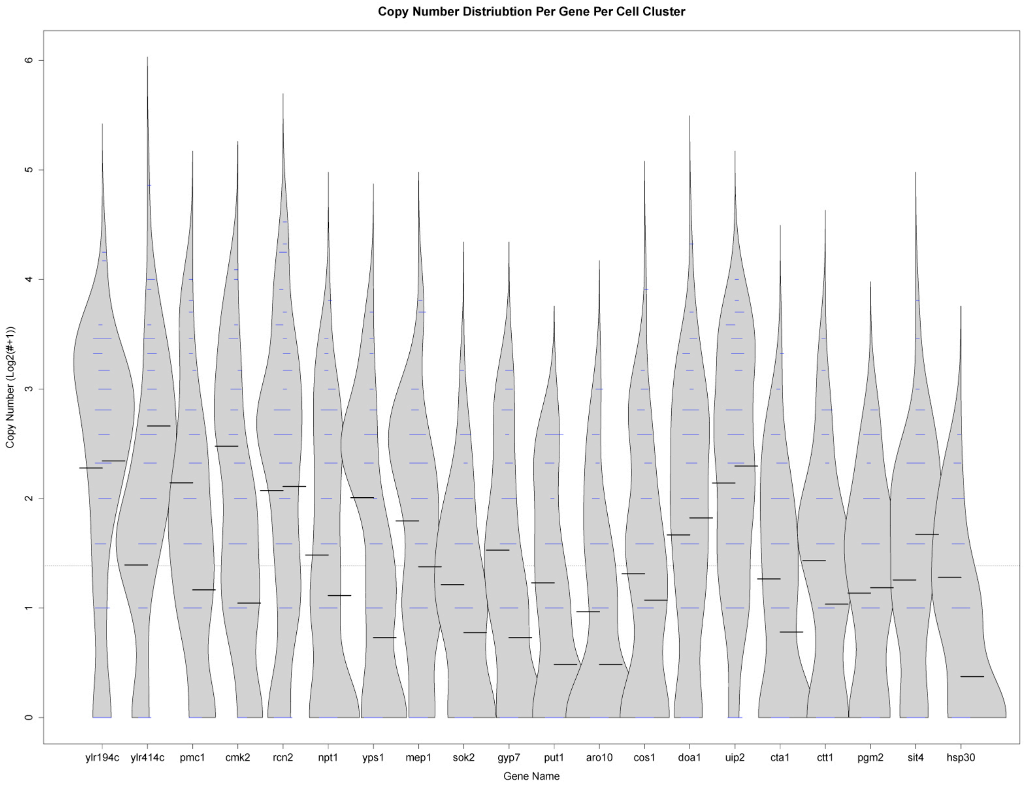


Figure 2.19: Distribution of expression levels for each Crz1 and Msn2 gene. Expression levels are shown in log2 values Bean plots. For each gene, two distributions are shown. On the left are the single cell expression profiles in cell cluster 2: cells with only the combinatorial targets are active. On the right are the distributions for that gene in cell cluster 1: cells with all Crz1 target genes on. Black lines indicate the mean values of the distribution. Blue lines mark the integer number of transcripts.

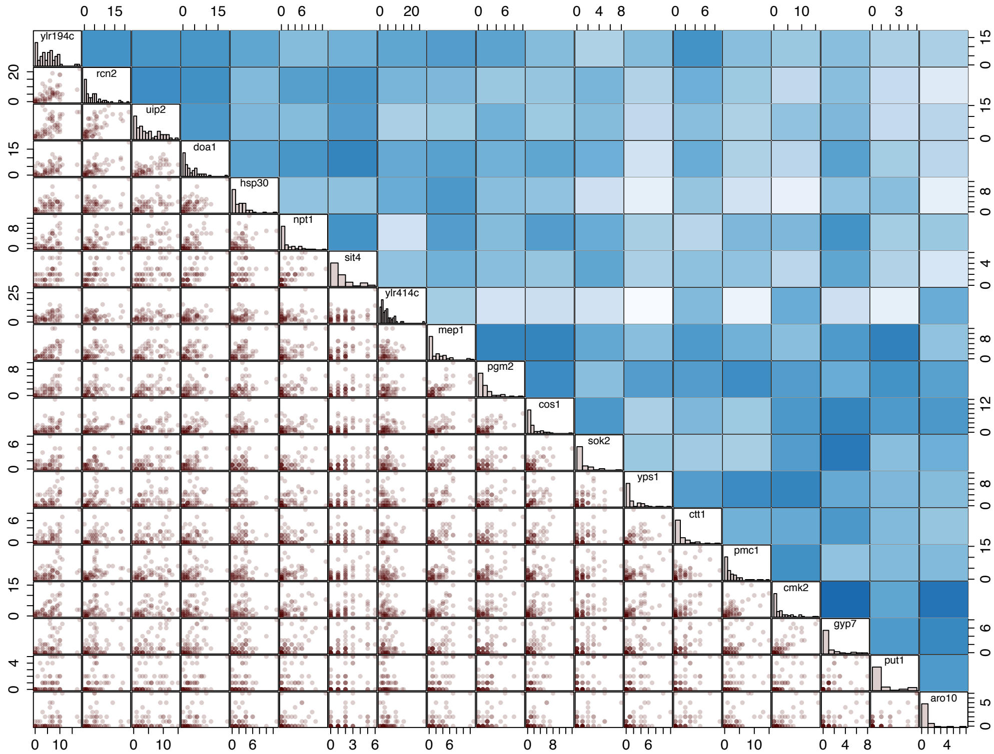


Figure 2.20: WT pairwise gene correlations. Genes are organized in the same order as Fig 5. The gene name and a histogram of each genes copy # distribution is shown in the diagonal. The lower triangle displays the raw pairwise correlation data, while the darkness of the upper triangular quadrant corresponds to the correlation coefficient of each pair.

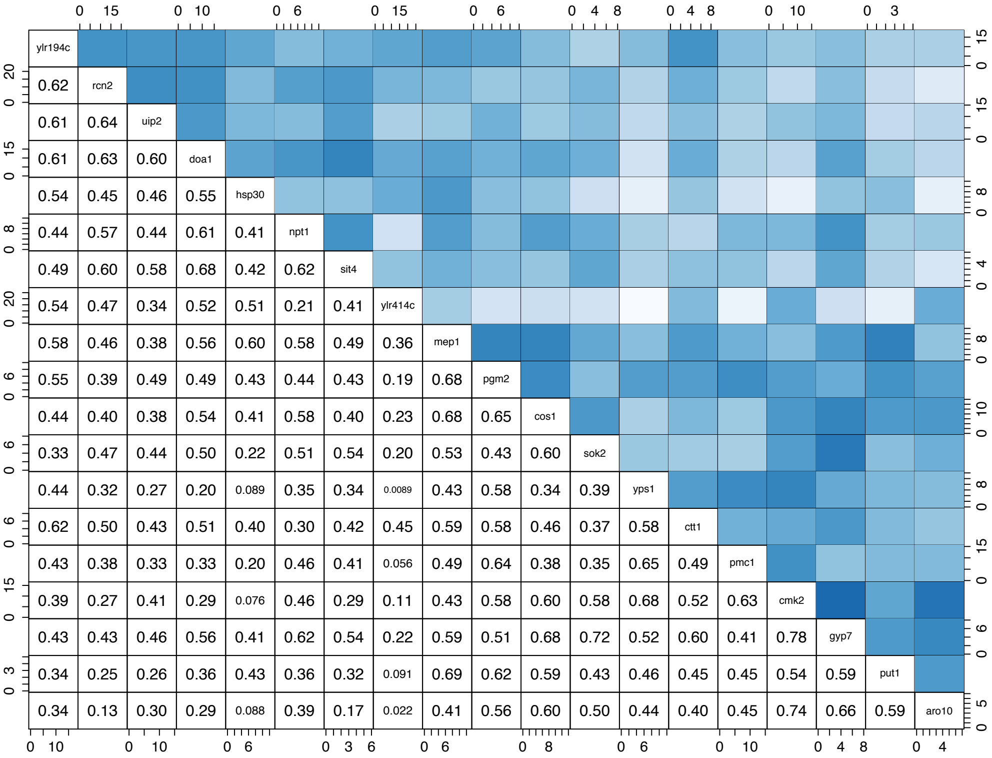


Figure 2.21: Numeric reference for shaded WT correlations shown in FigS7a.

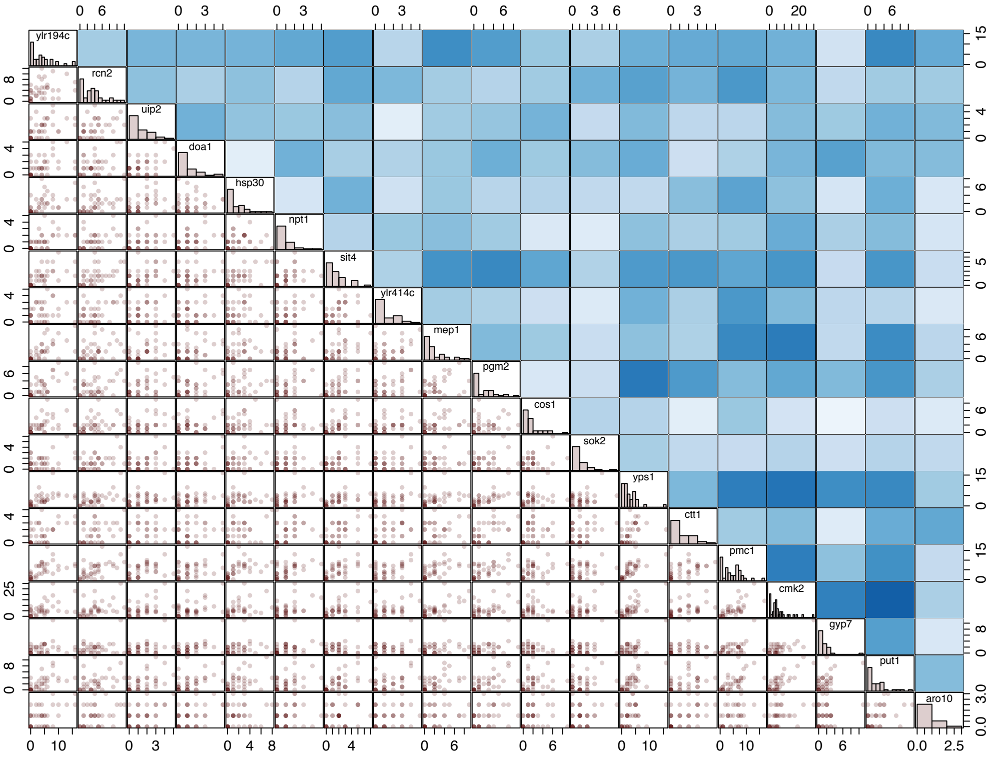


Figure 2.22: Msn2/4 deletion pairwise gene correlations. Genes are organized in the same order as 2.5Fig 5. The gene name and a histogram of each genes copy # distribution is shown in the diagonal. The lower triangle displays the raw pairwise correlation data, while the darkness of the upper triangular quadrant corresponds to the correlation coefficient of each pair. Not the overall smaller difference in correlations between clusters of genes compared to the WT plot.

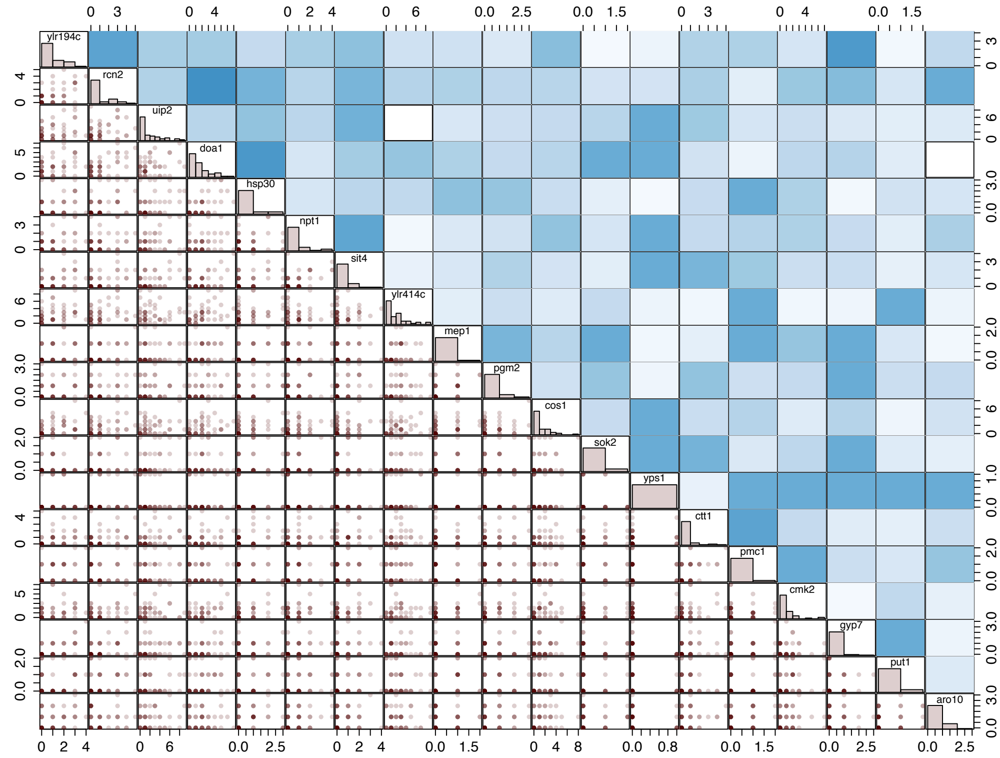


Figure 2.23: FK506 pairwise gene correlations. Genes are organized in the same order as 2.5Fig 5. The gene name and a histogram of each genes copy # distribution is shown in the diagonal. The lower triangle displays the raw pairwise correlation data, while the darkness of the upper triangular quadrant corresponds to the correlation coefficient of each pair. Note the overall low levels of expression.

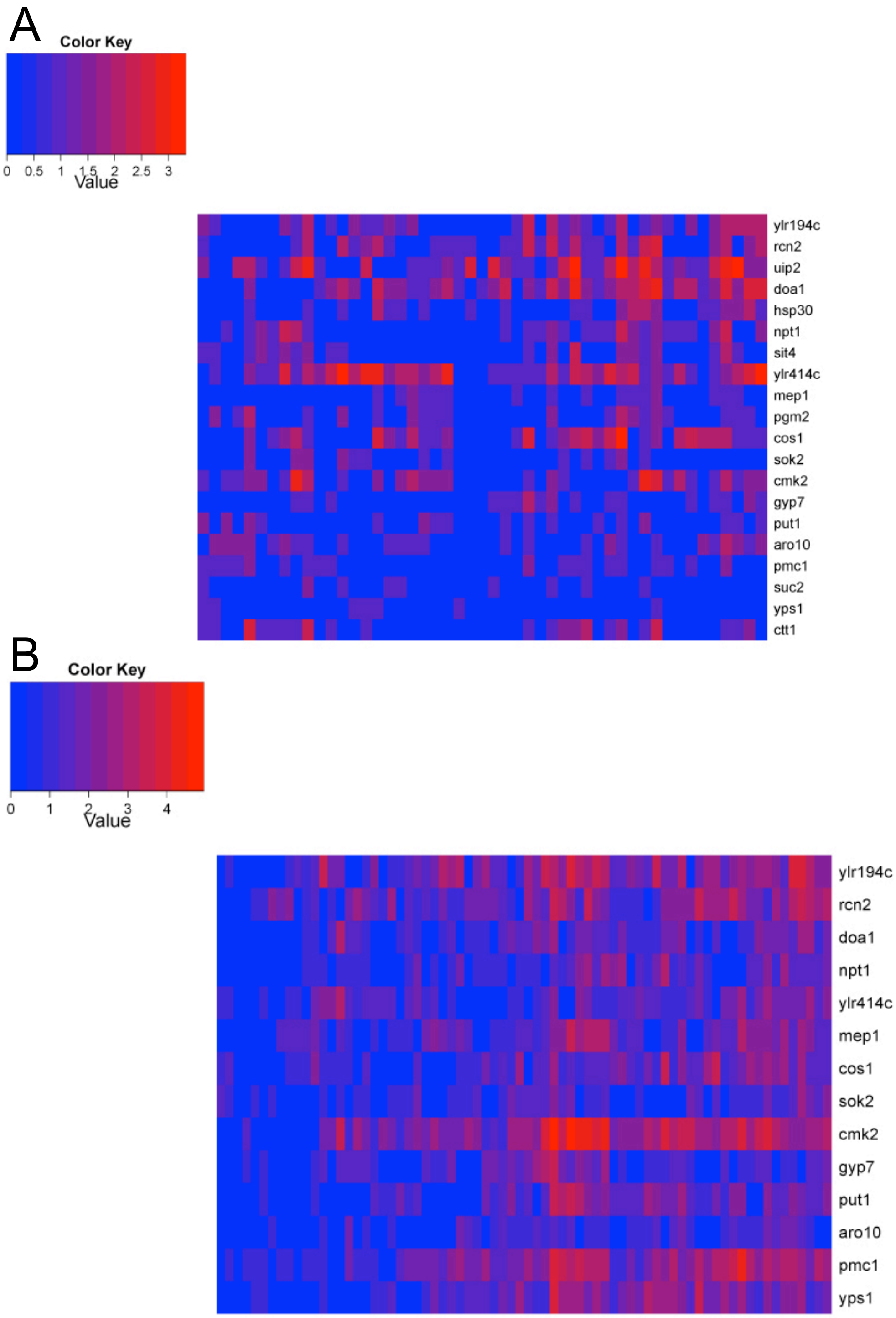


Figure 2.24: Heat maps of single cell gene expression levels under different conditions. In all figures, the cells are treated with 50mM $CaCl_2$ and the combinatorial genes are plotted in the upper half of the figure. Cells are plotted in columns. Note the values are log2 of expression level and different

for each of the heat maps. A) FK506 treated cells. Combinatorial targets are active while the pure Crz1 targets are inactive, indicating non-Crz1 inputs can drive combinatorial target expression. B) Msn2/4 deleted cells. Most cells show coordinated expression among Crz1 and combinatorial target genes, suggesting that Msn2 was the major factor in driving heterogeneous expression of combinatorial genes from the pure Crz1 genes.

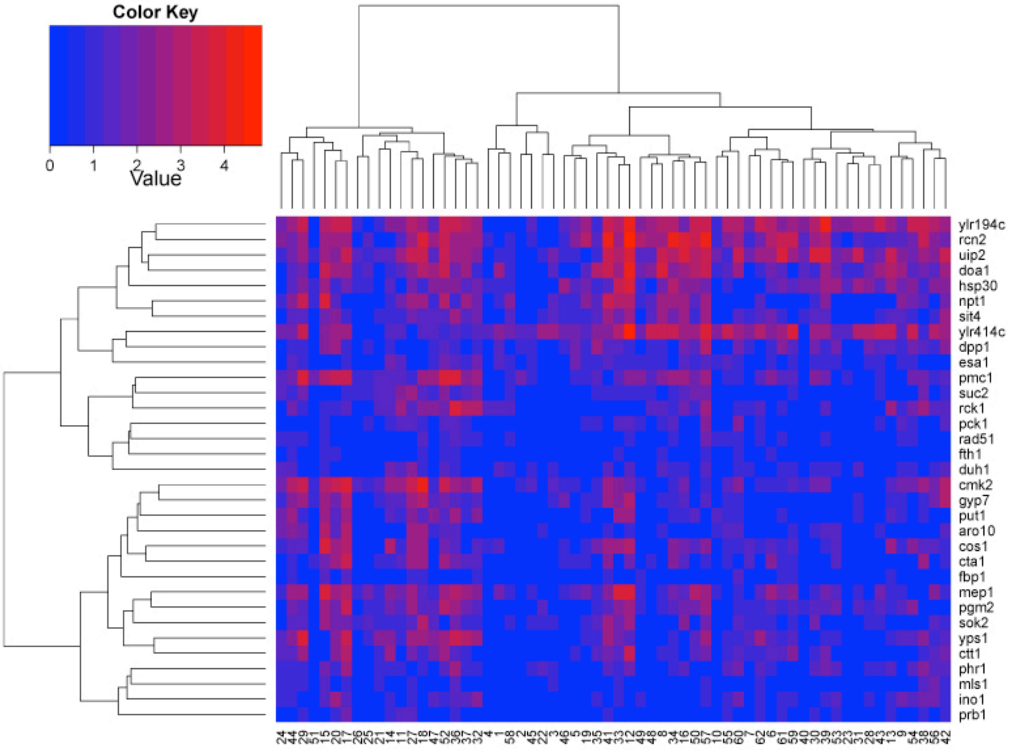


Figure 2.25: Clustering with the aging genes included. The same two clusters of pure and combinatorial Crz1 target genes are preserved. Aging and stress genes fall roughly into two separate clusters, apart from the Crz1 and Msn2 genes.

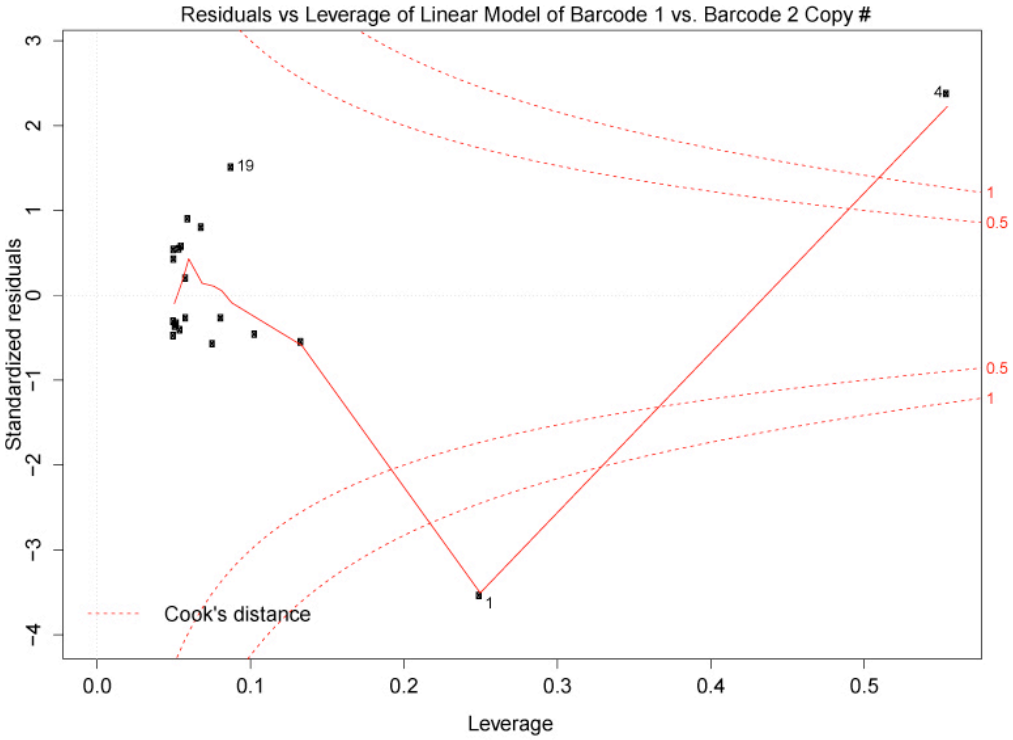


Figure 2.26: Statistical analysis of barcode scramble experiment. Two outliers are present with Cooks distances greater than 1 that greatly skew the correlation. The first point was dropped for our regression, as removing it placed point 4 back into a tight regression with the remaining data.

Table 2.2: Barcode Assignments for the 32 genes profiled. Codes 135, 235, and 895 are left empty. In the barcode scramble experiments, the activators are permuted. 1->2, 2->3, 3->1. 1= 405+Cy7, 2=488+Cy5, 3=Cy3+Cy5, 7=404+Cy7, 8=488+Cy7, 9=Cy3+Cy7

Gene	**Barcode**
YLR414c	123
YLR194c	239
Cmk2	789
Pmc1	389
Cos1	127
Mep1	289
Npt1	189
Put1	378
Yps1	379
Sok2	279
Gyp7	278
Aro10	178
Doa1	128
Rcn2	139
Ctt1	137
Hsp30	238
Pgm2	237
Sit4	138
Uip2	129
Cta1	179
Dpp1	125
Duh1	795
Esa1	895
Fbp1	157
Fth1	158
Ino1	159
Mls1	257
Pck1	258
Phr1	259
Prb1	785
Rad51	358
Rck1	359

Table 2.3: Circularly permutated barcodes used in switched barcode experiment (Figure 4 and S5). Barcodes are permutated following the pattern: (405->488, 488->cy3, cy3->405).

Gene	**Barcode**
YLR414c	123
YLR194c	137
cmk2	789
pmc1	179
cos1	238
mep1	379
npt1	279
put1	189
yps1	178
sok2	378
gyp7	389
ao10	289
doa1	239
rcn2	127
ctt1	128
hsp30	139
pgm2	138
sit4	129
uip2	237
cta1	278

Table 2.4: Probes for smFISH 5' amine modified.

Sequence	**Name**
gatctcacgctacaccatagaatgaa	ylr414c-1
catcaaaccctggtagttcctaccaa	ylr414c-2
tatgctttaggatgtatttgatgtat	ylr414c-3
actaatagggcggcaaaggcgaaaaa	ylr414c-4
ccttatgtggatgatccagcgcaata	ylr414c-5
caataccaataagaatggtaatgaac	ylr414c-6
attttactttttagtttttcgggcaa	ylr414c-7
cagagcctcattgttgttgatattgt	ylr414c-8
ggataccgtgaggcgaagaacatgat	ylr414c-9
tacgaccaaagccctatatttatata	ylr414c-10
agaactcaaagaagggagcaccgtcg	ylr414c-11
cacagtaaattttatttatgggactg	ylr414c-12
acggacgctaccttaccgttgactg	ylr194c-1
tgtagaacctgacgtagtggtataa	ylr194c-2
ttgattccggttttgatgaggatcc	ylr194c-3
tcagttgtggctgaggacggtagcc	ylr194c-4
cgaattcgtggtagttactatagta	ylr194c-5
aggaggatgcggagttggtgattcc	ylr194c-6
gcagttgaagttgtgcttacggcag	ylr194c-7
tgtcgtggttttgccttgtgcatcc	ylr194c-8
cataggtgttgctgacgacgttgct	ylr194c-9
acagttgatgcgctttcttgggctt	ylr194c-10
gtttgagctttccttttgtgagcta	ylr194c-11
agtctttttgagcagcggctagagt	ylr194c-12
atgcagacttcaatttcatttgctc	cmk2-1
tgcagacgtaaatcatccaacgaat	cmk2-2
ggaattctcttctatatcgttatcg	cmk2-3
acctcttaattctattattaagctt	cmk2-4
cgcaaagaaaaccctttcttaacgt	cmk2-5
tgaagtaatccatggatcgtccagc	cmk2-6
tcaatctcaatgccttcaagatgaa	cmk2-7
tatggcatatggaaggttaccgggt	cmk2-8

Continued on next page

Continued from previous page

Sequence	**Name**
ttcaacgctttcggcaataaaagga	cmk2-9
caccaatggaccatatatcacaagg	cmk2-10
ggtgccacataacccaacgatccgg	cmk2-11
caattgtttagctataccgaagtcc	cmk2-12

Table 2.5: qPCR Primers

Primer	Sequence
Cmk2 F	TCGCCTCTGGTAATTGCGGAC
Cmk2 R	TAACCCAACGATCCGGCTGC
Pmc1 F	TTGTTGCGGTCACTGGCGAT
Pmc1 R	AAGCCTCTCTGGCAACCTCC
Ylr414c F	GCTACGCTATCTTCGTTGGGC
Ylr414c R	CTGGATACCGTGAGGCGAAGA
Ylr194c F	AGCAACTCTGCCGTAAGCACA
Ylr194c R	GTCGTTGAGGAGGATGCGGA
Npt1 F	GGGAGATCCTGCCACTGTGA
Npt1 R	AGGTCCATCTGTGCGCTTCG
Gyp7 F	ACGATGGGAGGCTGAGGGTC
Gyp7 R	ACCCCAAACTTTCCCTCGCA
Put1 F	GGCGATAAAACGGGCACTGA
Put1 R	AGGCGACAACCAAGTGACCAA
Yps1 F	TTGACGGGAACGGGCAGTG
Yps1 R	CCGAAGCAGGCACGGATTGA
Actin F	ACGTTTCCATCCAAGCCGT
Actin R	GGAACGACGTGAGTAACACCA

Chapter 3

Single cell *in situ* RNA profiling by sequential hybridization

3.1 Main Text

Eric Lubeck[1,2], Ahmet F. Coskun[1,2], Timur Zhiyentayev[1], Mubhij Ahmad[1], Long Cai[1]

1. Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA, 91125. 2. Equal Contribution

Correspondence should be addressed to L.C. (lcai@caltech.edu).

To the editor: In our previous paper, Lubeck and Cai [1], we used super-resolution microscopy to resolve a large number of mRNAs in single cells. In this correspondence, we present a sequential barcoding scheme to multiplex different mRNAs.

Here, the mRNAs in cells are barcoded by sequential rounds of hybridization, imaging, and probe stripping (3.1, 3.2). As the transcripts are fixed in cells, the corresponding fluorescent spots remain in place during multiple rounds of hybridization, and can be aligned to read out a fluorophore sequence. This sequential barcode is designed to uniquely identify an mRNA.

---

[1] Lubeck, E. & Cai, L. Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nat. Methods* **9**, 743–748 (2012).

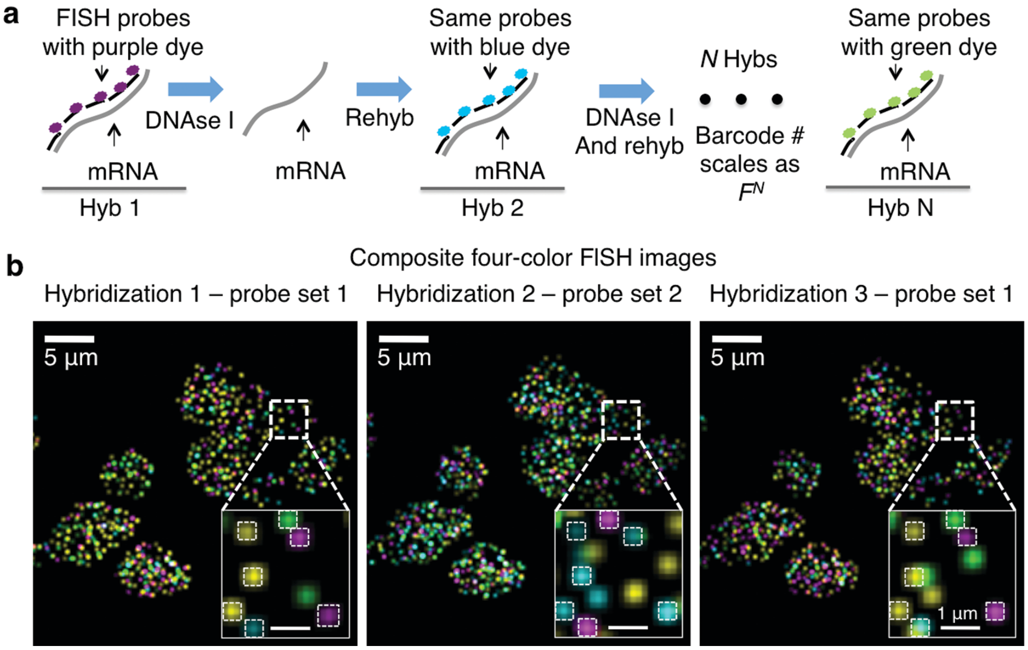


Figure 3.1: Sequential barcoding. (a) Schematic of sequential barcoding. In each round of hybridization, 24 probes are hybridized on each transcript, imaged, and then stripped by DNAse I treatment. The same probe sequences are used in different rounds of hybridization, but probes are coupled to different fluorophores. (b) Composite four-color FISH Data from three rounds of hybridizations on multiple yeast cells. Twelve genes are encoded by two rounds of hybridization, with the third hybridization using the same probes as hybridization 1. The boxed regions are magnified in the bottom right corner of each image. The matching spots are shown and barcodes are extracted. Spots without colocalization are due to nonspecific binding of probes in the cell as well as mis-hybridization. The number of each barcode can be quantified to provide the abundances of the corresponding transcripts in single cells.

During each round of hybridization, we targeted each transcript by a set of FISH probes labeled with a single type of fluorophore. We imaged the sample and then treated it with DNAse I to remove the FISH probes. In a subsequent round the mRNA was hybridized with the same FISH probes, but now labeled with a different dye. The number of barcodes available scales as F^N, where F is the number of fluorophores and N is the number of hybridization rounds. For example, with four dyes, eight rounds o=f hybridization can cover the entire transcriptome (4^8=65,536).

As a demonstration, we barcoded 12 genes in single yeast cells with four dyes and two rounds of hybridization (4^2=16, with four barcodes left out). Cells are immobilized on glass surfaces (**Supplementary Methods**). The DNA probes are hybridized, imaged, and then removed by DNAse I treatment (88.5±11.0% (SE) efficiency 3.3). The remaining signal is photobleached (3.4). Even after six hybridizations, mRNAs were observed at 70.9±21.8% (SE) of the original intensity

(3.5). We observed that 77.9±5.6% (SE) of the spots that co-localized in the first two hybridizations also co-localize with the third hybridization (3.1, 3.6, 3.7). The mRNA abundances are quantified by counting the occurrence of corresponding barcodes in the cell (3.8, 3.9 n=37 cells). We also show that mRNAs can be stripped and re-hybridized efficiently in adherent mammalian cells (3.10, 3.11).

Sequential barcoding has many advantages. First, it scales up quickly; with even two dyes the coding capacity is in principle unlimited. Second, during each hybridization, all available FISH probes against a transcript can be used, increasing the brightness of the FISH signal. Lastly, barcode readout is robust, enabling full Z-stacks on native samples.

This barcoding scheme is conceptually akin to sequencing transcripts in single cells with FISH. Compared to Ke et al [2], our method takes advantage of the high hybridization efficiency of FISH (>95% of the mRNAs are detected [53, 3]) and the fact that base pair resolution is usually not needed to uniquely identify a transcript. We note that FISH probes can also be designed to resolve a large number of splice-isoforms, SNPs [55], as well as chromosome loci [4] in single cells. In combination with our previous report of super-resolution FISH [53], the sequential barcoding method will enable the transcriptome to be directly imaged at single cell resolution in complex samples, such as the brain.

3.2 Supplementary

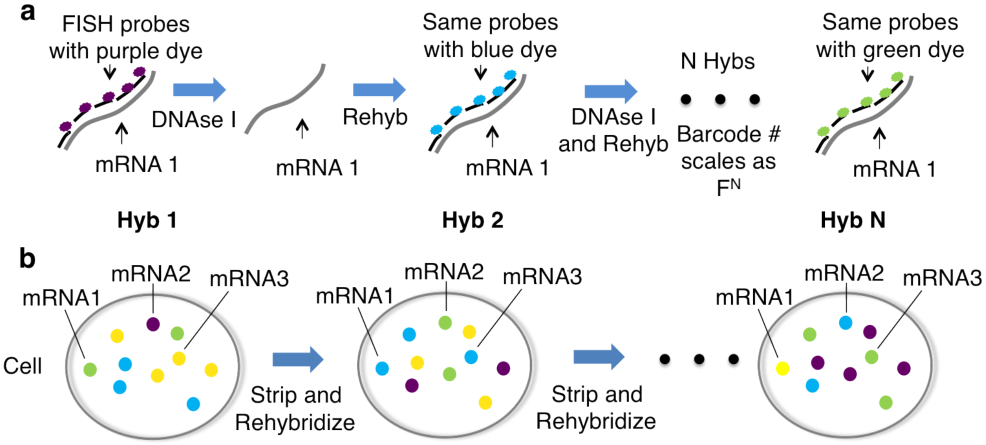


Figure 3.2: (**A**) Schematic of sequential hybridization and barcoding. (**B**) Schematic of the FISH images of the cell. In each round of hybridization, the same spots are detected, but the dye associated with the transcript changes. The identity of the mRNA is encoded in the temporal sequence of dyes hybridized.

Supplementary Methods

---

[2] Ke, R. *et al.* In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10**, 857–860 (2013).

[3] Levesque, M. J., Ginart, P., Wei, Y. & Raj, A. Visualizing SNVs to quantify allele-specific expression in single cells. *Nat. Methods* **10**, 865–867 (2013).

[4] Levesque, M. J. & Raj, A. Single-chromosome transcriptional profiling reveals chromosomal gene expression regulation. *Nat. Methods* **10**, 246–248 (2013).

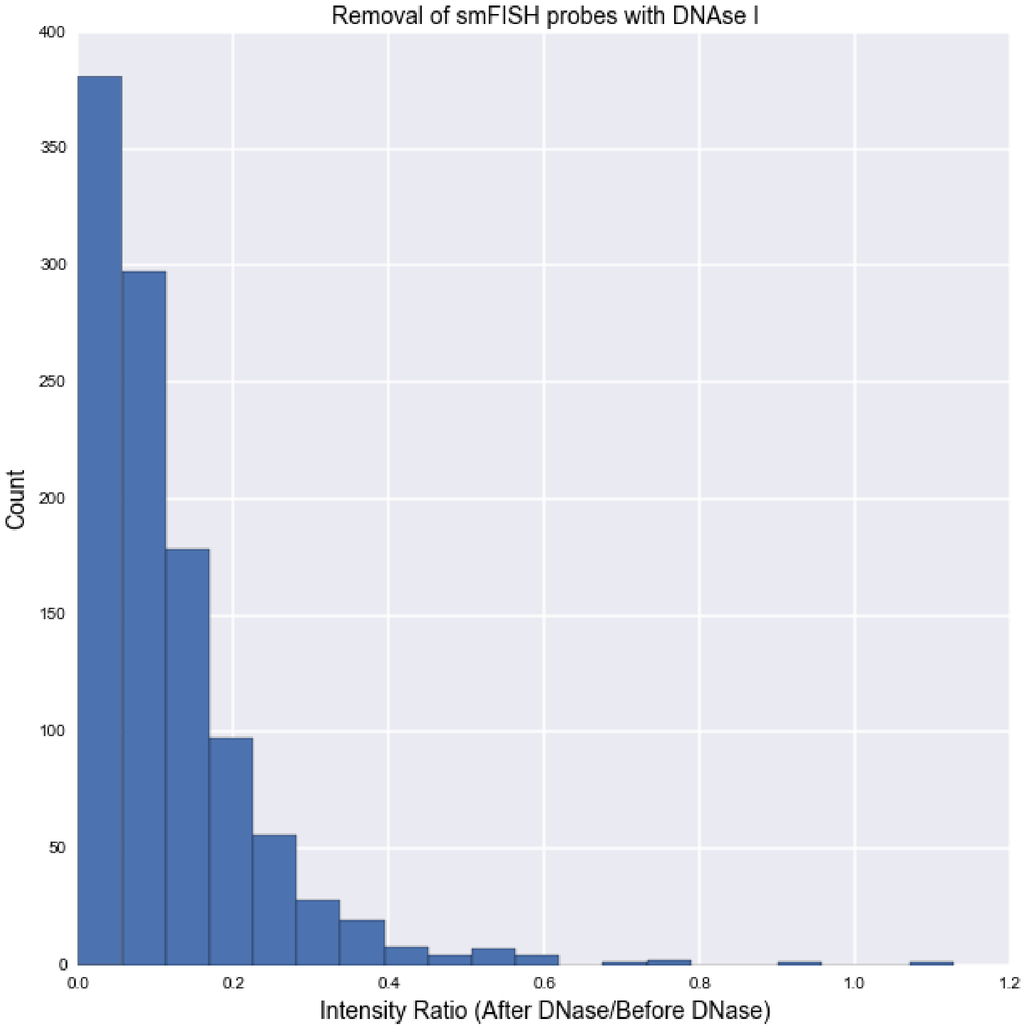


Figure 3.3: DNase I efficiently removes smFISH probes bound to mRNA. Spots were imaged before and after a 4 hour DNAse I treatment in anti-bleaching buffer. The mean, median, and STD of the intensity ratio after treatment are 11.5%, 8.3%, and 11%. The ratio of the spot intensities after and before DNase I treatment is plotted for each spot. n = 1084 spots.

Sample Preparation

MDN1-GFP yeast cells were grown in YPD supplemented with 50mM $CaCL_2$ to OD 0.3. Cells were fixed in 1% Formaldehyde 5% Acetic Acid for 5 minutes, rinsed 3X in Buffer B, and spheroplasted for 1 hour at 30°C. Cells were stored in 70% EtOH at -20°C for up to two weeks.

Coverslips were prepared by sonicating 3X with alternating solutions of 1M NaOH and 100% EtOH followed by a final round of sonication in acetone. A 2% solution of (3-Aminopropyl) triethoxysilane(Sigma 440140) was prepared in acetone and the cleaned coverslips were immediately submerged in it for two minutes. Amine-modified coverslips were rinsed and stored in ultra pure water at room temperature.

Fixed yeast cells were pre-treated with a 0.5U/uL solution of DNase I (Roche 04716728001) for

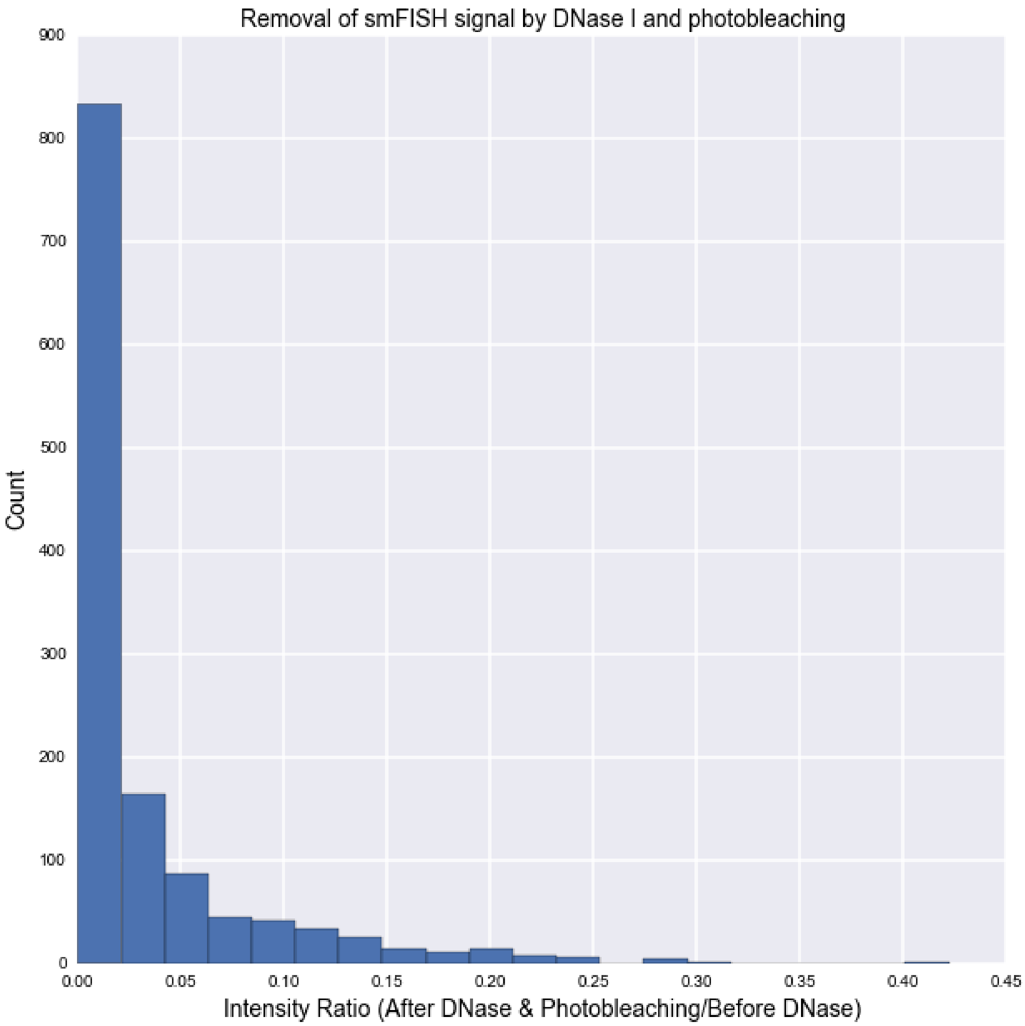


Figure 3.4: Photobleaching removes residual intensity following DNase I treatment. Spots were bleached by 10 seconds of excitation following a 4 hour DNAse I treatment. The mean, median, and STD of the intensity ratio after bleaching are 0.03%, 0.01%, and 0.049%. The ratio of the spot intensities after and before DNase I treatment is plotted for each spot. n = 1286 spots.

30 minutes at 23°C. Following treatment, yeast cells were adhered to coated coverslips by physically compressing a dilute solution of yeast between two amine-modified coverslips. The coverslips were than carefully pealed apart and immediately submerged in a 1% Formaldehyde solution for 2.5 minutes. Following fixation coverslips were dried and a flow cell was constructed by adhering an adhesive coated flow cell to the coverslip (GraceBio Labs SA84-0.5-SecureSeal). FluoSphere 365nm fluorescent beads were added to the coverslip to measure drift over multiple hybridizations (Life F8805). Flow cells were stored at 4°C covered with parafilm.

Probe Preparation

Probes were prepared according to the method in our previous paper[53]. For each gene, 24 probes are used. All 24 probes for each set of genes are coupled to one of the four dyes used: Alexa 532, 594, Cy5, and Cy7.

Hybridization

Flow cells were hybridized at a concentration of 2nM/probe overnight in a hybridization buffer of 10% Dextran Sulfate (Sigma D8906), 10% Formamide, 2X SSC. Following hybridization, samples were washed in a 30% Formamide, 0.1% Triton-X 100 buffer pre-heated to 37°C before adding to room temperature samples for 10 minutes. Samples were washed several times with 2X SSC to remove diffusing probes.

Imaging

Samples were immersed in an anti-bleaching buffer[5] consisting of 20mM Tris-HCL, 50mM NaCL, 0.8% Glucose, Saturated Trolox(Sigma: 53188-07-1), Pyranose oxidase(Sigma P4234) at an $\mathrm{OD}_{405\mathrm{nm}}$ of 0.05, and catalase at a dilution of 1/1000 (Sigma:9001-05-2).

Probe Displacement

Following imaging, cells were washed in DNase I buffer (Roche) and allowed to sit in 0.5U/uL DNase I (Roche) for 4 hours. To inhibit DNase cells were washed 2X with 30% Formamide, 0.1% Tritíon-X 100, 2X SSC. Following DNase treatment cells were imaged once more in anti-bleaching buffer to determine DNase I probe stripping rates. To remove remaining probe signal samples were bleached with 10 seconds of excitation in all imaging channels and imaged once more with standard excitation times to record residual signal.

Re-hybridization

Samples were re-hybridized on the microscope according to the previously outlined conditions. Samples were covered with parafilm during hybridization on the scope to prevent evaporation.

Up to six rounds of hybridizations were carried out on the same sample. Each round of hybridization took place overnight on the microscope, with DNAse treatment and imaging occurring during the day. In the iterative hybridization scheme applied in this correspondence, we used two rounds of hybridization to barcode the mRNAs. The barcode scheme was then repeated, such that hyb1 and hyb3 were performed using the same probes, while hyb2 and hyb4 were done with another set of probes. The co-localization between hyb1 and hyb3 gave a calibration for transcripts that were detected, while hyb1 and hyb2 yielded the barcoding data.

Data Analysis

Data analysis was carried out with ImageJ, Python and Matlab. Since the sample drifted during the experiments, the raw images were aligned using cross-correlation based registration method that was determined from the DAPI channel of each imaging position. The drift-correction was then propagated to the other four-color channels corresponding to the same position. The images were then deconvolved to decrease the overlap between adjacent FISH spots. We note that we rarely observed spots overlap in individual channels, but spots in different channels can overlap

---

[5] Swoboda, M. *ACS Nano* **6**, 6364–69 (2012).

in their point spread functions (PSFs) when the images are overlaid. Therefore, we processed the raw data based on an iterative Lucy-Richardson algorithm[6,7]. The PSF of the microscope was estimated by averaging the measured bead images (~200 nm diameter) in the DAPI channel of the microscope. Using this measured point spread function with the Lucy-Richardson algorithm, we performed maximum-likelihood estimation of fluorescent emitter distribution in the FISH images after computing this process over ~20 iterations. The output of this deconvolution method provides resolved FISH data and increases the barcode assignment fidelity.

Dots corresponding to FISH signals in the images were identified using a local maximum function[8]. Dots below a threshold were discarded for further analysis. The value of the threshold was determined by optimizing the co-localization between hyb1 and hyb3 images, which were hybridized with the same probe sets. The maximum intensity pixel for each PSF was used as a proxy for the location of that mRNA molecule.

The barcodes were extracted automatically from the dots corresponding to mRNAs in hyb1 and hyb2. The algorithm calculated the pairwise distances between each point identified in hyb1 with all the points identified in hyb2. For each point in hyb1, the closest neighbor in hyb2 was identified. If that distance were 0 or 1 pixel and the closest neighbor of the point in hyb2 were also the original point in hyb1, then the barcode pair was confirmed. The symmetrical nearest neighbor requirements decreased the false assignment of barcodes. To reduce false positives in cy7, we required points detected in hyb 1 cy7 to reappear in hyb 3 in cy7.

Supplementary Note

For this publication we chose to strip off probes with DNase I due to its low cost and rapid activity. In principal any method that removes probes from mRNA and leaves it intact could be used in our barcoding approach. Other available methods include strand-displacement[9] and high temperature or formamide washes. Compared to these other methods DNase I has the advantage of not requiring probe redesigns from standard smFISH probes, and not perturbing the sample with harsh washes.

We saw a rapid loss of DAPI signal from dsDNA within seconds, while smFISH probes took a substantially longer period of time (10s of minutes) to be degraded. Although the efficiency of DNase I probe removal is low relative to the dsDNA cleavage rate, the process is still observed in a short amount of time.

In our experiments we see that 11.5% of the fluorescent signal remains on mRNA after DNase I treatment. The remaining signal is reduced almost to zero by bleaching. We note that our rehybridization technique would not work without the DNase I treatment. If probes were still bound

---

[6] Lucy, L.B. *The Astronomical Journal.* **79**, 745 (1974).
[7] Richardson, W.H. *J. Opt. Soc. Am.* **62**, 55-59 (1972).
[8] Shaffer, S. M. *PLoS ONE*, **8**, e75120 (2013).
[9] Duose, D. Y. *Nucleic Acids Research*, **40**, 3289-3298 (2012).

to mRNA but bleached, the following rounds of hybridization would not work. Photobleaching is not necessary for barcoding, but does simplify the process by removing residual signal that might give false positives in further rounds of barcoding.

Some of the 11.5% of residual probes bound to mRNA may inhibit further rounds of hybridization. We note that residual probes are not significantly inhibiting progressive rounds of hybridization, as our data shows a minor drop in hybridization efficiency for five hybridizations.

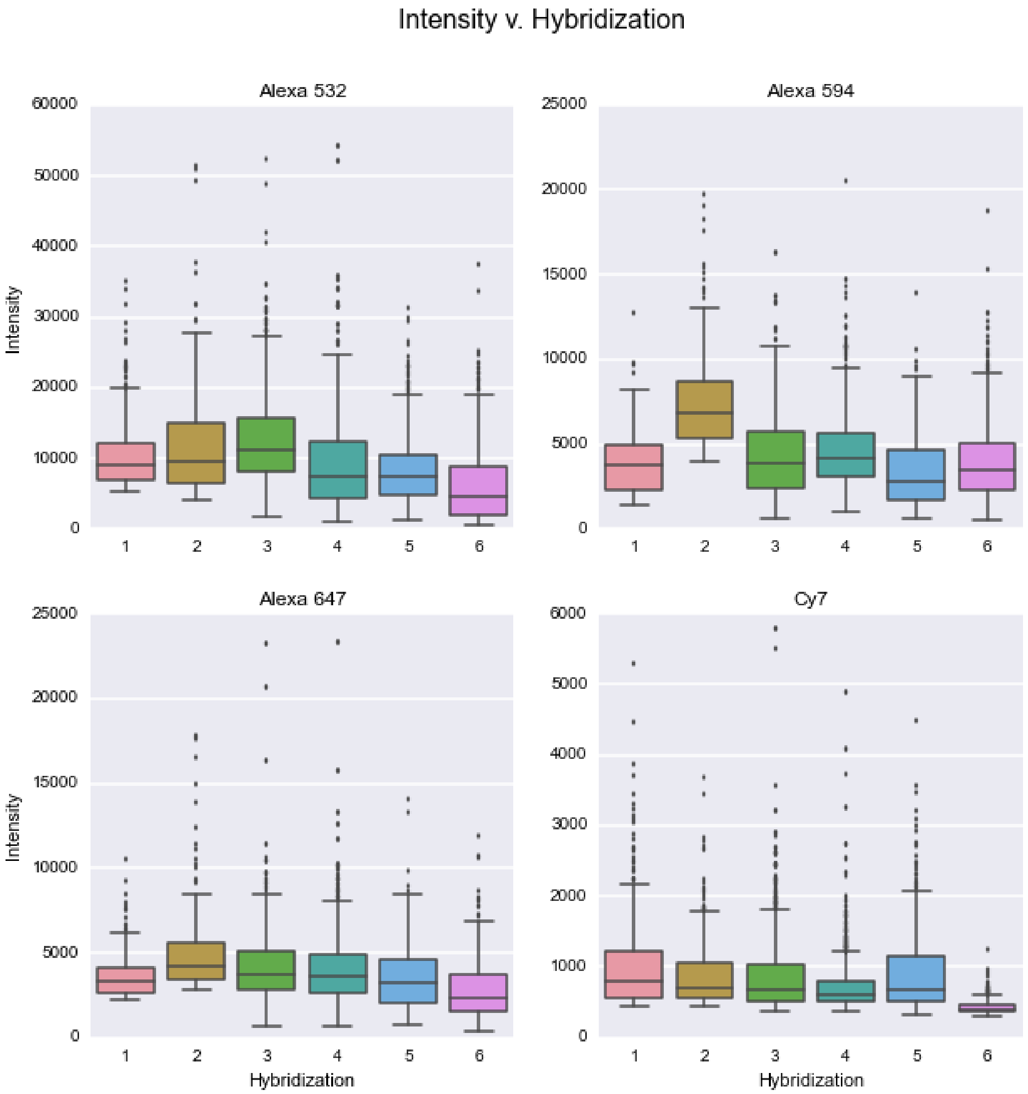


Figure 3.5: mRNAs are stable over multiple rounds of re-hybridization. The intensity distributions of smFISH spots are plotted over six hybridizations. Two hybridizations were repeated three times to make six total hybridizations. Spots were identified by their co-localization with spots in the next identical hybridization. For each boxplot the number of spots counted was between 191 and 1337.

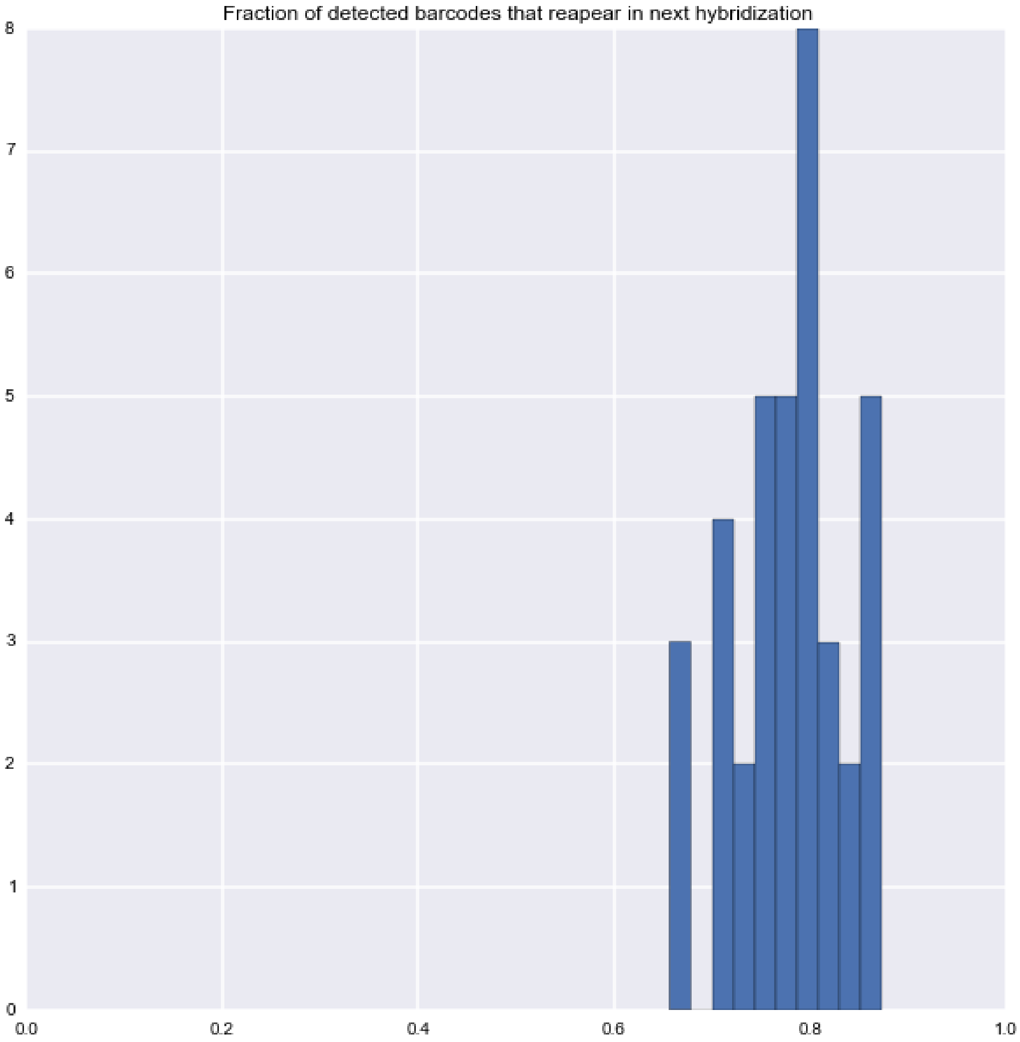


Figure 3.6: Fraction of barcodes identified from first two rounds of hybridization that reoccur in following round of hybridization per cell. Barcodes were identified by co-localization through all three hybridizations. 77.9 ± 5.6% of barcodes reoccur. n = 37 cells.

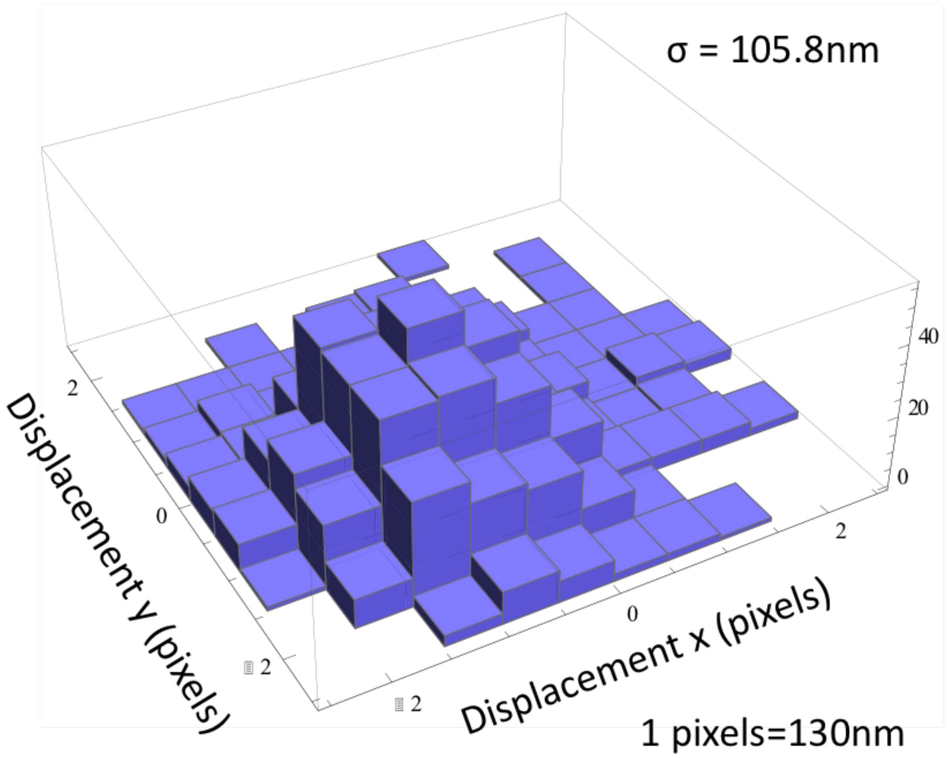


Figure 3.7: Point-wise displacement between FISH points in Hybridizations 1 and 3. FISH dots in the Cy5 images in Hybridization 1 and 3 are extracted and fitted with 2D Gaussians. The point-wise displacements are shown in the 3D histogram. The standard deviation is 105.8nm, indicating that mRNAs can be localized to 100nm between two rounds of hybridizations. n = 1199 spots.

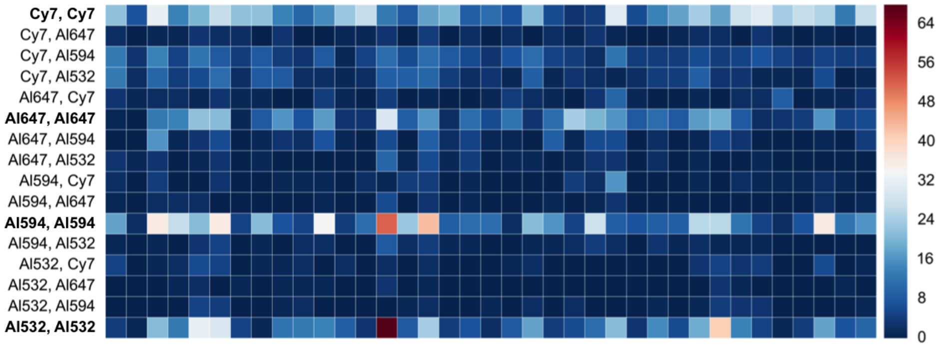


Figure 3.8: Barcodes identified between repeat hybridizations of the same probe set (hybridization 1 and 3). Barcodes were identified by co-localization between the hybridizations. Each column corresponds to an individual cell. Each row corresponds to a specific barcode identified between hybridization 1 and 3. Bolded row names correspond to repeated color barcodes that should co-localize between hybridization 1 and 3. Non-bolded row names correspond to false positive barcodes. For example, a large number of barcodes are detected for (Alexa 532, Alexa 532), indicating co-localization of spots in the Alexa 532 channels. n = 37 cells. Al532 = Alexa 532, Al594 = Alexa 594, and Al647 = Alexa 647.

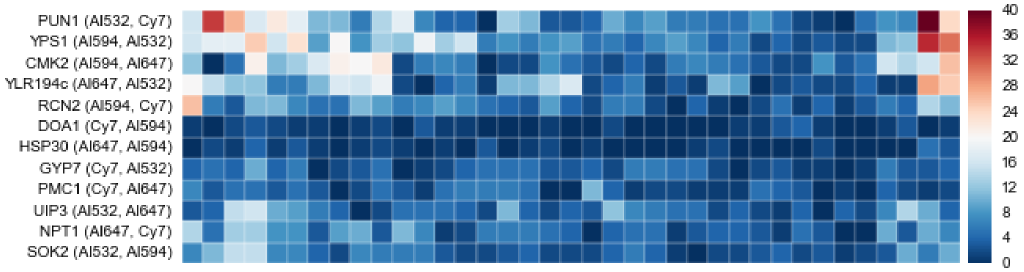


Figure 3.9: Single cell mRNA levels from barcode extraction. Barcodes were identified by co-localization between hybridizations 1 and 2. Each column corresponds to an individual cell. n = 37 cells. Al532 = Alexa 532, Al594 = Alexa 594, Al647 = Alexa 647 Al647 = Alexa 647.

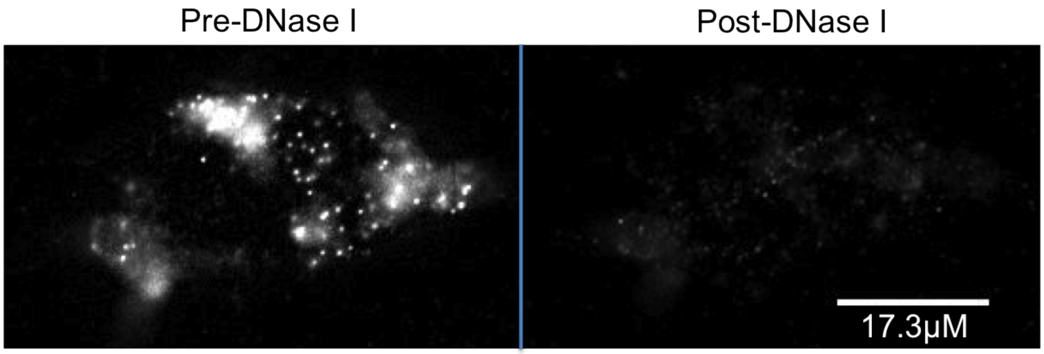


Figure 3.10: DNase I stripping of Nanog Alexa 647 probes in mouse embryonic stem cells (mESCs). Forty-eight probes targeting Nanog were hybridized in mESCs. Probes were stripped off by 30 minutes of DNase I incubation at a concentration of 3 Units/L.

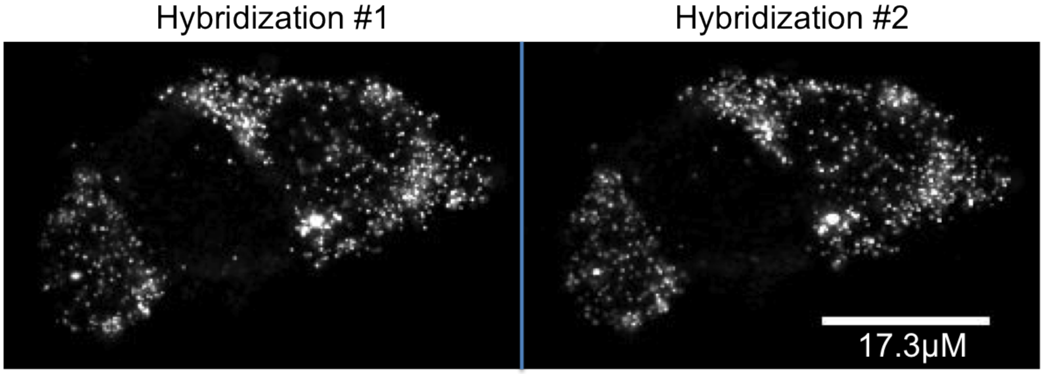


Figure 3.11: Re-Hybridization of Nanog mRNA in Mouse Embryonic Stem Cells (mESCs). Probes were stripped off by 30 minutes of DNase I incubation at a concentration of 3 Units/L. Nanog Alexa 647 probes were re-hybridized for 12 hours and imaged. Images are 2D maximum projections created from z stacks of 11 images taken every 1.5m.

Chapter 4

Single-molecule RNA detection at depth via hybridization chain reaction and tissue hydrogel embedding and clearing

Sheel Shah[1,4,†], Eric Lubeck[1,†], Maayan Schwarzkopf[2,†], Ting-fang He[1,†], Alon Greenbaum[2], Chang ho Sohn[1], Antti Lignell[1], Harry M. T. Choi[2], Viviana Gradinaru[2,*], Niles A. Pierce[2,3,*], and Long Cai[1,*]

Affiliations: 1. Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA, USA 91125. 2. Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA 91125. 3. Division of Engineering & Applied Science, California Institute of Technology, Pasadena, CA, USA 91125. 4. UCLA-Caltech Medical Scientist Training Program, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA, USA 90095

†These authors contributed equally.

*Correspondence to: lcai@caltech.edu, niles@caltech.edu, viviana@caltech.edu

4.1 Abstract

Accurate and robust detection of mRNA molecules in thick tissue samples can reveal gene expression patterns in single cells within their intact environment. Preserving spatial relationships while accessing the transcriptome of selected cells is a crucial feature for advancing many biological areas, from developmental biology to neuroscience. However, because of the high autofluorescence background of many tissue samples, it is difficult to detect single-molecule fluorescence *in situ* hybridization (smFISH) signals robustly in opaque thick samples. Here, we developed a robust method for multi-color, multi-RNA imaging in deep tissues using single-molecule hybridization chain reac-

tion (smHCR) compatible with epi-fluorescence, confocal and selective plane illumination microscopy (SPIM). We show that smHCR has high sensitivity in detecting mRNAs in cell culture and zebrafish embryos. By combining SPIM and PACT (PAssive CLARITY Technique) tissue hydrogel embedding and clearing, smHCR can detect single mRNA transcripts in thick brain slices. With a ~20-fold signal amplification, smHCR overcomes the long-standing barriers of high background and photobleaching encountered with smFISH methods and therefore offers a robust and versatile approach to detecting single mRNAs *in situ* in a variety of biological samples.

4.2 Introduction

Imaging gene expression levels with single-cell resolution in intact tissues is essential for understanding the genetic programs in many systems, such as developing embryos and dynamic brain circuits. Single-molecule fluorescence *in situ* hybridization (smFISH) has been the standard tool for detection of individual RNAs in cells (1-5). Using smFISH, an mRNA is detected by a probe set containing 20-40 DNA probes, each carrying one or more fluorophores, and each complementary to a different short subsequence (20-50 nt) along the mRNA target. This approach ensures that multiple probes bind the mRNA, generating bright puncta that can be discriminated from background staining resulting from non-specific binding of individual probes. However, background due to sample autofluorescence is significantly higher in tissue samples than in cell culture, making it difficult to robustly detect smFISH signals in tissue. In addition, light scattering caused by deep tissue imaging necessitates probes with higher photon counts than for thin section imaging. While we have recently shown that PACT clearing (PAsive CLARITY Technique) (6-8) can alleviate autofluorescence background and light scattering problems (8), a more robust signal amplification strategy is needed to enable multi-color mapping of single mRNAs in deep tissues.

4.3 Results

Attempts have been made to specifically amplify a single mRNA signal, but these tend to suffer from low efficiency and/or cumbersome protocols (9). Our solution provides a simple and efficient method for multiplexed signal amplification based on the mechanism of hybridization chain reaction (HCR) (10,11). With this approach, short DNA probes complementary to mRNA targets trigger chain reactions in which metastable fluorophore-labeled DNA hairpins self-assemble into tethered fluorescent amplification polymers (Fig 4.1a). As with smFISH, each target mRNA is addressed by 20-40 probes complementary to different subsequences along the target to enable discrimination between mRNAs with multiple probes bound and dots resulting from non-specific binding of individual probes. In contrast to previous *in situ* HCR methods (10,11), we limit the HCR amplification time to achieve a mean polymer length of ~20-40 hairpins, generating puncta that are bright enough for high sensitivity yet small enough for diffraction-limited resolution. Using orthogonal HCR amplifiers programmed to operate independently, straightforward multiplexing is achieved for up to five channels simultaneously (Fig 4.1b). We term this method single-molecule HCR (smHCR).

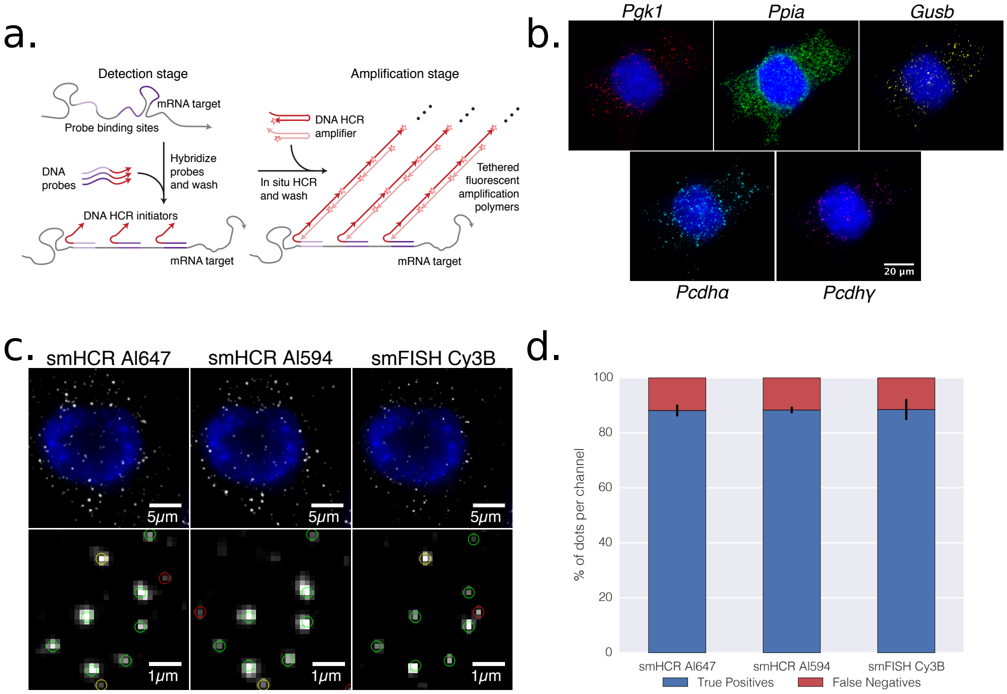


Figure 4.1: Single-molecule hybridization chain reaction (smHCR). **(a)** smHCR protocol. Detection stage: an mRNA target is detected by a probe set containing 20-40 short DNA probes, each binding a 20-30 nt subsequence of the target; each probe in the probe set carries an initiator for the same HCR amplifier. Amplification stage: metastable fluorophore-labeled DNA HCR hairpins penetrate the sample and self-assemble into fluorescent amplification polymers tethered to their initiating probes. The same two-stage protocol is used for multiplexed studies: during the detection stage, all probe sets are introduced simultaneously, each carrying an initiator for an orthogonal HCR amplifier; during the amplification stage, all HCR amplifiers are introduced simultaneously, each labeled with spectrally distinct fluorophores. **(b)** Simultaneous mapping of five target mRNAs in cultured CAD cells using five spectrally distinct HCR amplifiers (DAPI in blue): *Pgk1* (Cy7), *Ppia* (Alexa 647), *Gusb* (Alexa 594), *Pcdh(Cy3b), /Pcdh(Alexa 488).* ***(c)*** *Comparison of smHCR and smFISH for detection of /Pgk1* via dot colocalization in three channels. Ch1: smHCR (Alexa 647), Ch2: smHCR (Alexa 594), Ch3: smFISH (Cy3B). Dots are classified as triple-detected true positives (present in all three channels; green circles), double-detected true positives (present in two out of three channels; yellow circles), or false positives (present in only one channel; red circles). **(d)** True-positive and false-negative rates for each channel in panel (c). Probe sets: 22 probes per set, each

addressing a 20 nt target subsequence. See Supplementary Figures 4.1-14 for additional data.

To characterize the sensitivity and selectivity of smHCR in cells, we performed a colocalization experiment in which a target mRNA was simultaneously detected using three probe sets of 22 probes each (one smFISH set and two smHCR sets), with the probes alternating amongst the three sets along the target. Dots were identified in each channel by applying a threshold following standard methods for smFISH data analysis (Supplementary Fig. 4.1) (2). We interpret dots that are colocalized in at least two of the three channels as representing true mRNA signals (Fig 4.1c). Thus, a dot that is absent in only one channel is interpreted as a false negative in that channel; a dot that is present in only one channel is interpreted as a false positive in that channel. For the two-smHCR channels and the smFISH channel, the false negative rates are all 12% (Fig 4.1c), and the false positive rates are 36%, 27% and 20%, respectively (Supplementary Fig. 4.2). Hence, smHCR demonstrates comparable sensitivity and somewhat worse selectivity to smFISH in cultured cells where background autofluorescence is minimal. Comparing intensities for true and false positive dots, smHCR provides signal amplification of approximately a factor of 20-40 relative to smFISH and false positives are considerably dimmer than true positives (Supplementary Figs 4.3-11), a feature that will become crucial in samples with higher levels of autofluorescence.

Interestingly, in experimental designs where two channels can be allocated to each target mRNA, near-quantitative single-molecule mapping can be achieved (Supplementary Figs 4.12-14). Using this approach, the threshold for dot identification is lowered in each channel to achieve a lower false negative rate ($< 5\%$) at the cost of a higher false positive rate ($> 60\%$). Then, dot colocalization between channels is used to identify the subset of dots that represent true mRNA signals. On the other hand, if each target is detected in only a single channel and colocalization is not used to confirm dot classification, then the threshold should be set high to reject false positives at the cost of rejecting some true positives similar to the procedure used in smFISH (2).

To examine the performance of smHCR within the more challenging imaging setting of an intact vertebrate embryo, we repeated our 3-channel colocalization study in a whole-mount zebrafish embryo. The target mRNA, *kdrl*, is detected using three smHCR probe sets of 39 probes each, with probes alternating between the three sets along the target mRNA. Here, we found that the false negative rates are 16%, 15%, and 14% (Fig 4.2b), and the false positive rates are 36%, 21%, and 31% for channels 1, 2, and 3, respectively, using the same thresholding method as described above (Supplementary Figs 4.15-16). smFISH has been previously demonstrated to be feasible in zebrafish embryos (12), so here we compared the sensitivity and selectivity of smHCR in whole-mount zebrafish embryos to smFISH (Supplementary Figs 4.17-29). To do so, we disabled HCR amplification in channel 2 by introducing only the first HCR hairpin species, enabling only one hairpin carrying one fluorophore to bind to each probe as a proxy for smFISH (Fig 4.2c; compare dots in Ch2 of panels a and c). We find the amplification gain of smHCR compared to smFISH in whole-mount zebrafish embryos to be approximately a factor of 15 (Fig 4.2d).

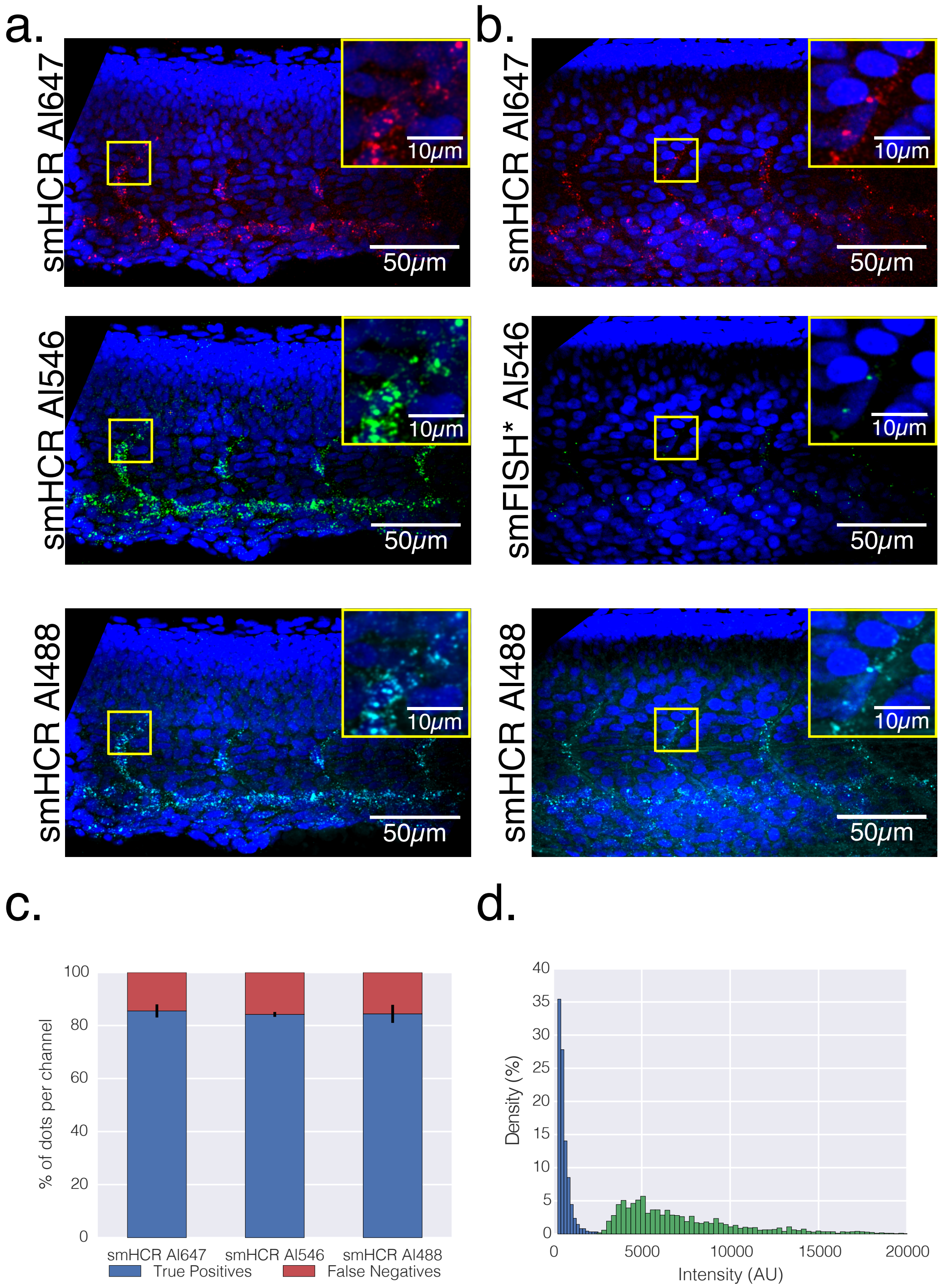


Figure 4.2: Mapping single mRNAs within whole-mount zebrafish embryos using smHCR. **(a)** Validating smHCR via dot colocalization in 3-channels. Ch1: smHCR (Alexa 647), Ch2: smHCR (Alexa 546), Ch3: smHCR (Alexa 488). **(b)** True-positive and false-negative rates for each channel

in panel (a). **(c)** Comparison of smHCR and smFISH* via dot colocalization in 3 channels. Ch1: smHCR (Alexa 647), Ch2: smFISH* (Alexa 546; as a proxy for the unamplified performance of smFISH, smFISH* uses only one of the two HCR hairpins so that only the first polymerization step is possible), Ch3: smHCR (Alexa 488). Channel pairs between panels (a) and (c) are shown with the same contrast, with Ch 2 illustrating the difference in intensity between amplified smHCR dots and unamplified smFISH* dots. **(d)** Comparison of amplified smHCR and unamplified smFISH* dot intensities for Ch 2 of panels (a) and (c). Target mRNA: *kdrl* expressed in the blood vessels. Microscopy: spinning disk confocal. Probe sets: 39 probes per set, each addressing a 30 nt target subsequence. Embryos fixed: 27 hpf. See Supplementary Figures 4.15-29 for additional data.

To image thick brain slices (>100 µm), which better preserve morphology and connectivity, we used selective plane illumination microscopy (SPIM) (13-15) on PACT cleared brain slices (7,8,16). PACT turns tissues optically transparent and macromolecule-permeable by removing light-scattering lipids, which are replaced by a porous hydrogel to both preserve the morphology and lock in elements of interest such as proteins or nucleic acids. However, even with optimal clearing, imaging deep into ex vivo tissues requires a high-resolution imaging modality that rejects scattered light from out-of-focus planes. While confocal microscopy filters this background noise, image acquisition is slow, and out-of-plane excitation can photobleach the sample. SPIM (13-15, 17) offers a fast alternative (~100 times faster than confocal) that rejects out-of-focus noise by illuminating and capturing images only from a thin selective plane, typically on the order of 1-10 µm. By combining smHCR, PACT clearing, and SPIM, phenotypical studies can be conducted quickly while preserving the natural long-range architecture of the sample.

In order to examine the selectivity of smHCR in thick PACT cleared sections with SPIM, 1-mm thick brain slices from a Thy1-YFP mouse were PACT cleared (7,8) and hybridized with HCR probes against YFP mRNA (see Supplementary Fig. 4.30 for flow diagram). As PACT cleared tissue retains endogenous YFP fluorescence, we were able to directly test the selectivity of HCR staining without the need for parallel antibody staining. To generate a 3D reconstruction (Fig 4.3a, Supplementary Fig. 31, Supplementary Movie 1), 1,000 optical sections with 1 µm spacing were taken (25 frames per second), and the intensity was normalized between the different sections. Supplementary figure 29 illustrates the change of the mean intensity as a function of imaging depth, without the use of intensity normalization.

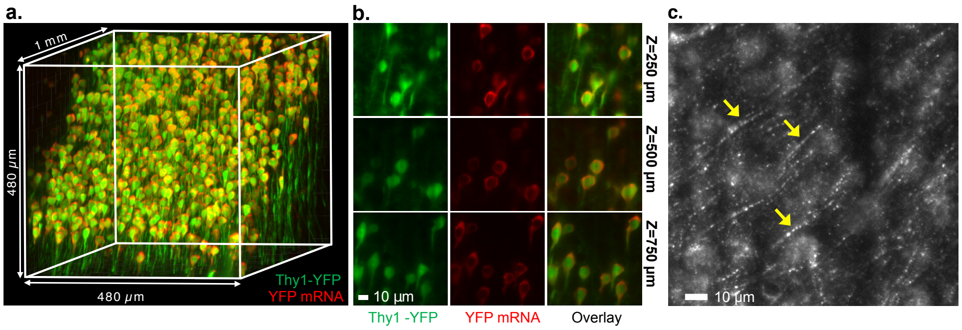


Figure 4.3: Mapping mRNAs in 1-mm PACT-cleared mouse brain section using smHCR and SPIM. **(a)** Three-dimensional reconstruction depicting colocalization of *Thy1-YFP* protein (expressed in ~5% of cells) and *Thy1-YFP* mRNA (smHCR; Cy3B). **(b)** Representative optical sections demonstrating that smHCR signal is detected throughout the brain slice (z = 250 m, 500 m, 750 m). **(c)** Visualization of *Ppia* mRNA decorating neuronal projections using smHCR probes, similar to previously reported mRNA packages in projections [Batish 2012, Buxbaum 2014]. See Supplementary Figures 4.30-33 for additional data.

The Thy1-YFP expressing cells (green) account for less than 5% of the cellular population (Supplementary Fig. 4.32) and we observe a one-to-one correspondence between cells stained by YFP protein fluorescence and cells stained for YFP mRNA by smHCR (Fig 4.3b). For this transcript, the expression level is sufficiently high that we do not observe individual dots. For less abundant mRNAs such as Peptidylprolyl Isomerase A (Ppia), individual mRNA transcripts are again observed (Fig 4.3c).

4.4 Discussion

smHCR provides a robust method to quantify mRNAs in whole-mount zebrafish embryos, while retaining the favorable performance of smFISH in cell culture. Furthermore, smHCR combines well with PACT clearing and SPIM to image transcripts deep in brain samples. Noting that whole bodies and a wide range of tissues have been successfully cleared (7,8), we expect the combination of smHCR, PACT, and SPIM to enable molecular profiling of a wide variety of samples with single-cell resolution and preserved geometry and connectivity. As smHCR is compatible with sequential hybridization methods that we have previously developed (18,19), it should be possible to perform highly multiplexed studies within thick autofluorescent samples, mapping single mRNAs at depth.

References:

1. Femino, A.M., Fay, F.S., Fogarty, K., Singer, R.H. Visualization of single RNA transcripts in situ. Science *280,* 585-590 (1998).

2. Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. Nat Methods **5**, 877-879 (2008).

3. Levsky, J.M., Shenoy, S.M., Pezo, R.C., Singer, R.H. Single-cell gene expression profiling. Science **297**, 836-840 (2002).

4. Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., Tyagi, S. Stochastic mRNA synthesis in mammalian cells. PLoS Biol. **4**, e309 (2006).

5. Fan, Y., Braut, S.A., Lin, Q., Singer, R.H., Skoultchi, A.I. Determination of transgenic loci by expression FISH. Genomics **71**, 66-69 (2001).

6. Chung, K., Wallace, J., Kim, S.Y., Kalyanasundaram, S., Andalman, A.S., Davidson, T.J., Mirzabekov, J.J., Zalocusky, K.A., Mattis, J., Denisin, A.K., Pak, S., Bernstein, H,, Ramakrishnan, C., Grosenick, L., Gradinaru, V., Deisseroth, K. Structural and molecular interrogation of intact biological systems. Nature **497**, 332-337 (2001).

7. Treweek, J.B., Chan, K.Y., Flytzanis, N.C., Yang, B., Deverman, B.E., Greenbaum, A., Lignell, A., Xiao, C., Cai, L., Ladinsky, M.S., Bjorkman, P.J., Fowlkes, C.C. and Gradinaru, V. Whole-Body Tissue Stabilization and Selective Extractions via Tissue-Hydrogel Hybrids for High Resolution Intact Circuit Mapping and Phenotyping. Nature Protocols (2015; In Press).

8. Yang, B., Treweek, J.B., Kulkarni, R.P., Deverman, B.E., Chen, C.-K., Lubeck, E., Shah, S., Cai, L., and Gradinaru, V. Single-Cell Phenotyping within Transparent Intact Tissue through Whole-Body Clearing. Cell **158**, 1-14 (2014).

9. Player, A.N., Shen, L.P., Kenny, D., Antao, V.P., Kolberg, J.A. Single-copy gene detection using branched DNA (bDNA) in situ hybridization. J Histochem Cytochem. **49**, 603-612 (2001).

10. Choi, H.M.T., Chang, J.Y., Trinh, L.A., Padilla, J.E., Fraser, S.E., and Pierce, N.A. Programmable in situ amplification for multiplexed imaging of mRNA expression Nature Biotechnol. **28**,1208-1212 (2010).

11. Choi, H.M.T., Beck, V.A., Pierce, N.A. Next-generation in situ hybridization chain reaction: higher gain, lower cost, greater durability. ACS Nano. **8**, 4284-4294 (2014).

12. Oka, Y., Sato, T.N. Whole-mount single molecule FISH method for zebrafish embryo. Sci Rep. **5**, 8571 (2015).

13. Huisken, J., Swoger, J., Del Bene, F., Wittbrodt, J., and Stelzer, E. H. Optical sectioning deep inside live embryos by selective plane illumination microscopy. Science **305**, 1007-1009 (2004).

14. Dodt, H.U., Leischner, U., Schierloh, A., Jährling, N., Mauch, C.P., Deininger, K., Deussing, J.M., Eder, M., Zieglgänsberger, W., and Becker, K. Ultramicroscopy: three-dimensional visualization of neuronal networks in the whole mouse brain. Nature Method **4**, 331-336 (2007).

15. Keller, P.J., Schmidt, A.D., Santella, A., Khairy, K., Bao, Z., Wittbrodt, J., and Stelzer, E.H. Fast, high-contrast imaging of animal development with scanned light sheet-based structured-illumination microscopy. Nature methods **7**, 637-642 (2010).

16. Tomer, R., Ye, L., Hsueh B., and Deisseroth K. Advanced CLARITY for rapid and high-resolution imaging of intact tissues. Nat. Prot. **9**, 1682-1697 (2014).

17. Baumgart, E., and Kubitscheck, U. Scanned light sheet microscopy with confocal slit detection. Optics express **20**, 21805- 21814 (2012).

18. Lubeck, E., Coskun, A., Zhiyentayev, T., Ahmed, M., Cai, L. Single cell in situ RNA profiling by sequential hybridization. Nature Methods 11, 360–361 (2014).

19. Lubeck, E. Cai, L. Single cell systems biology by super-resolution imaging and combinatorial labeling. Nature Methods **9**, 743–748 (2012).

4.5 Supplementary

4.5.1 Methods

1. smHCR in cultured cells

   (a) Sample Preparation

   CAD cells were cultured as described previously (Suri et al., 1993). CAD cells were grown in DMEM/F-12 medium (ThermoFisher, Cat. # 12634), supplemented with 10% FBS (Gibco Cat. # 16000) and 1% penicillin-streptomycin (Gibco, Cat. # 15140) on standard tissue culture flasks in a humidified 5% $C0_2$ incubator. CAD cells were passaged every 3 to 4 days and cells were plated at a 1:10 dilution. For microscopy experiments, 22 mm × 40 mm no. 1.5 coverslips (EMS, Cat. #

      i. were coated with 0.01% poly-D-lysine (Sigma, Cat. # P7280) for 2

   hours at 37 °C in standard tissue culture dishes. Coverslips were washed with UltraPure water (ThermoFisher, Cat. # 10977) and cells were plated on the coverslip at a 1:6 dilution. Cells were grown overnight and fixed the following day with 4% formaldehyde in 1× PBS for 15 minutes at room temperature. Formaldehyde was then washed out with RNAse-free 2× SSC and stored in 70% ethanol and stored at -20°C.

(b) *In situ* hybridization

i. Probe design and synthesis*

DNA 20-nt probes complementary to the target mRNA were designed using Stellaris Designer (Supplementary Table 1 and 2). For smFISH probes 5' amine modified oligos were purchased and were directly coupled to a fluorophore and purified as described in Lubeck 2012. smHCR probes were synthesized with an HCR initiator to the 3' end of an oligo, and a short linker between the probe sequence and the initiator.

ii. Probe Hybridization

Coverslips coated with cells were removed from 70% EtOH storage at -20 *°*C. Coverslips were air-dried to remove all traces of EtOH and rehydrated in 2× SSC for 5 minutes. Samples were then hybridized overnight at 1 nM probe concentration in 10% hybridization buffer at 37 *°*C. 10% hybridization buffer is composed of 10% formamide (Ambion, Cat. # AM9342), 10% Dextran Sulfate (Sigma-Aldrich, Cat. # D6001) and 2× SSC (Sigma-Aldrich, Cat. # 93017) in RNAse-Free H_2O (Life Technologies, Cat. # 10977-015). While higher concentrations of probe often generate more signal per mRNA molecule, the corresponding increase in background generally negates any advantages.

iii. Probe Wash

Following hybridization, samples were washed in a solution of 30% formamide, 2× SSC, 0.1% Triton-X 100 (Sigma-Aldrich, Cat. # T8787) for at least 15 minutes. Increasing the wash duration appeared to have little effect on signal or background. All traces of wash buffer and unbound probe were removed by multiple rinses in 2× SSC.

iv. HCR amplification*

During the wash step above, fluorophore-labeled HCR hairpins purchased from Molecular Instruments (molecularinstruments.org) were thawed out of -20 *°*C storage and snap-cooled (heat at 95 *°*C for 2 minutes and cool to room temperature for 30 minutes) before use. Amplification was performed for 45 minutes at room temperature at a concentration of 120 nM per hairpin in an amplification buffer consisting of 10% Dextran Sulfate and 2× SSC in RNAse-Free water. Amplification times can be varied to modify polymer length. Following amplification, samples were washed in the wash buffer described above for at least 10 minutes to remove unbound hairpins.

(c) Imaging

After amplification, samples were briefly stained with DAPI (Life Technologies, Cat. #

D1306), rinsed in 2× SSC and placed in an enzymatic anti-bleaching buffer described in Lubeck 2014. Samples were imaged on Nikon Ti-Eclipse microscopes. Due to the high gain of smHCR, samples could be easily imaged on microscopes equipped with either a solid state laser or a xenon arc lamp. All publication data was obtained using laser illumination. All microscopes used high quantum efficiency (QE) Andor iKon-M cameras, but in principle, lower QE cameras can be used to image smHCR.

2. smHCR in whole-mount zebrafish embryos

   The following protocol is adapted from Choi *et al.* 2014.

   (a) Sample Preparation

   Zebrafish embryos were collected and grown in a petri dish with egg H_2O at 28 °C. Prior to fixation, embryos were dechorionated, transferred to an eppendorf tube, and excess egg H_2O was removed. Embryos were fixed at ~27 h post-fertilization (27 hpf) in 1 mL of 4% freshly prepared paraformaldehyde (PFA) for 24 h at 4 °C. Fixation was terminated by three 5 min washes in 1 mL of 1× phosphate-buffered saline (PBS). Embryos were permeabilized and dehydrated using a series of 1 mL methanol (MeOH) washes: 4 × 10 min followed by 1 × 50 min. Embryos were then rehydrated with a series of graded MeOH / PBST washes (1 mL each):

   - 75% MeOH / 25% PBST for 5 min
   - 50% MeOH / 50% PBST for 5 min
   - 25% MeOH / 75% PBST for 5 min
   - 100% PBST for 5 × 5 min.

   Embryos were stored at 4 °C prior to use.

   (b) *In situ* hybridization

   i. Probe design and synthesis
   smHCR probes for the mRNA target (*kdrl*) are 71 nt long (30-nt mRNA recognition sequence, 5-nt spacer, 36-nt initiator). Each probe set contains 39 1-initiator DNA probes (Supplementary Table 3; sequences listed 5' to 3'). Probes were designed to have 5 nt gaps between each other along the target mRNA. Within a given probe set, each probe initiates the same DNA HCR amplifier. Probes were synthesized by Molecular Instruments.

   ii. Probe hybridization
   Approximately six embryos per sample were pre-hybridized in pre-hybridization buffer for 30 mininutes at 65 °C. Pre-hybridization buffer is composed of 50% formamide

(Ambion, Cat. # AM9342), 5× SSC (Invitrogen, Cat. # 15557-044), 9 mM citric acid pH 6 (Mallinckrodt Chemicals, Cat. # 0627-12), 0.1% Tween-20 (Bio-Rad, Cat. # 161-0781), 50 g/mL heparin (Sigma, Cat. # H3393), 1× Denhardt's solution (Invitrogen, Cat. # 750018), and 10% dextran sulfate (Sigma, Cat. # D6001). Embryos were hybridized overnight at 2 nM probe concentration in hybridization buffer at 37 °C. Hybridization buffer composition is identical to the pre-hybridization buffer with 30% formamide instead of 50%.

iii. Probe wash

Following hybridization, samples were washed at 37 °C in a series of graded probe wash buffers and 5× SSCT:

- 75% probe wash buffer / 25% 5× SSCT for 15 minutes
- 50% probe wash buffer / 50% 5× SSCT for 15 minutes
- 25% probe wash buffer / 75% 5× SSCT for 15 minutes
- 100% 5× SSCT for 15 minutes
- 100% 5× SSCT for 30 minutes

iv. HCR amplification

Fluorophore-labeled hairpins and amplification buffer were purchased from Molecular Instruments. Samples were pre-amplified in amplification buffer at room temperature. Meanwhile, hairpins were snap-cooled individually at the provided concentration (3 µM) by heating to 95 °C for 90 seconds and cooling to room temperature in a dark drawer for 30 mininutes. Prior to amplification, hairpins were mixed together in amplification buffer to achieve a final concentration of 60 nM for each hairpin. The pre-amplification solution was removed and replaced with the hairpin solution. Samples were incubated for 1 hour at room temperature in the dark. Following amplification samples were washed in 5× SSCT: 2 × 5 minutes wash, 2 × 30 minutes wash, 1 × 5minutes wash.

(c) Imaging

Prior to imaging, the 5× SSCT was replaced with SlowFade Gold Antifade Mountant with DAPI (ThermoFisher Scientific, Cat. # S36937). The head and yolk sac were resected and fish tails were mounted sandwiched between two no. 1 coverslips (VWR, Cat. # 48393 106). Images were collected on an Andor CSU-W1 spinning disk confocal on a Nikon Ti-E with Perfect Focus System microscope equipped with a Plan Apo 60×/1.4 and an Andor iXon ULTRA 888BV camera.

3. Dot Identification and Classification for Cell and Zebrafish Studies

All image analysis was performed on three-dimensional image stacks. Before image analysis, the images were processed by performing a rolling ball background subtraction with a radius of 3 pixels. Local maxima within the image were found by thresholding the image based on previously published smFISH thresholding methods (Reference??) and comparing the resulting image with a 1-pixel radius dilated image to find local maxima. Once dots were found in all images, the local maxima were matched using a maximum 1-pixel tolerance nearest neighbor search.

4. PACT clearing

   (a) PACT clearing and smHCR in 1-mm adult mouse brain slices

   All procedures were done in an RNase-free environment. Adult mice were subjected to standard transcardiac perfusion with 4% paraformaldehyde (PFA; all solutions cold) followed by brain extraction and post-fixation in 4% PFA for 3.5-4 hours at room temperature (RT). Using a mouse brain matrix (Kent Scientific Corp., Cat. # RBMS-200C), the brain was cut into 1-mm thick slices, and further post-fixed for an additional 3-4 hours at RT in order to better preserve mRNA transcripts. The PFA-fixed brain slices were then PACT cleared following the previously reported protocol [Yang 2014, Treweek, in press]. Briefly, the slices were incubated in hydrogel monomer solution, 4% acrylamide supplemented with 0.25% photoinitiator 2,20 -Azobis[2-(2-imidazolin-2-yl) propane] dihydrochloride (Wako Chemicals USA, Cat. # VA-044) at 4 °C overnight. The following day the samples were rigorously degassed using nitrogen for 8-10 minutes, and the tissue-gel hybrid was constructed at 37 °C for 2-3 hours. After multiple rounds of PBS washes, the samples were transferred to 50 mL conical tubes pre-filled with detergent solution (8% SDS in PBS, pH 7.6), and incubated at 37 °C for 2-3 days. Before HCR hybridization, the 1-mm brain slices were extensively washed with PBS for 2-3 days in order to remove residual SDS in the sample.

   (b) smHCR

   smHCR was performed in 6-well plates. First, the samples were pre-incubated in 2× SSC for 1 hour (RT) and then incubated (37 °C) for an additional hour in the hybridization buffer (10% formamide, 10% dextran sulfate, 2× SSC). Next, the pre-warmed (37 °C) hybridization buffer containing DNA HCR probes (10 nM per probe) was added to the sample and incubated overnight (sequences in Supplementary Table 4). The following day, the samples were washed (30% formamide, 2× SSC at 37 °C for 1.5 hours) to remove free and nonspecifically-bound probes, and then washed with 2× SSC. The HCR amplification stage was performed according to the next-generation HCR protocol (Choi 2014). However, for smHCR, we used a higher concentration of each amplifier hairpin (120 nM) and performed amplification for only 6 hours at RT. The HCR brain slices were placed in RIMS (refractive index of 1.47) and degassed by nitrogen for 20 minutes prior to SPIM imaging.

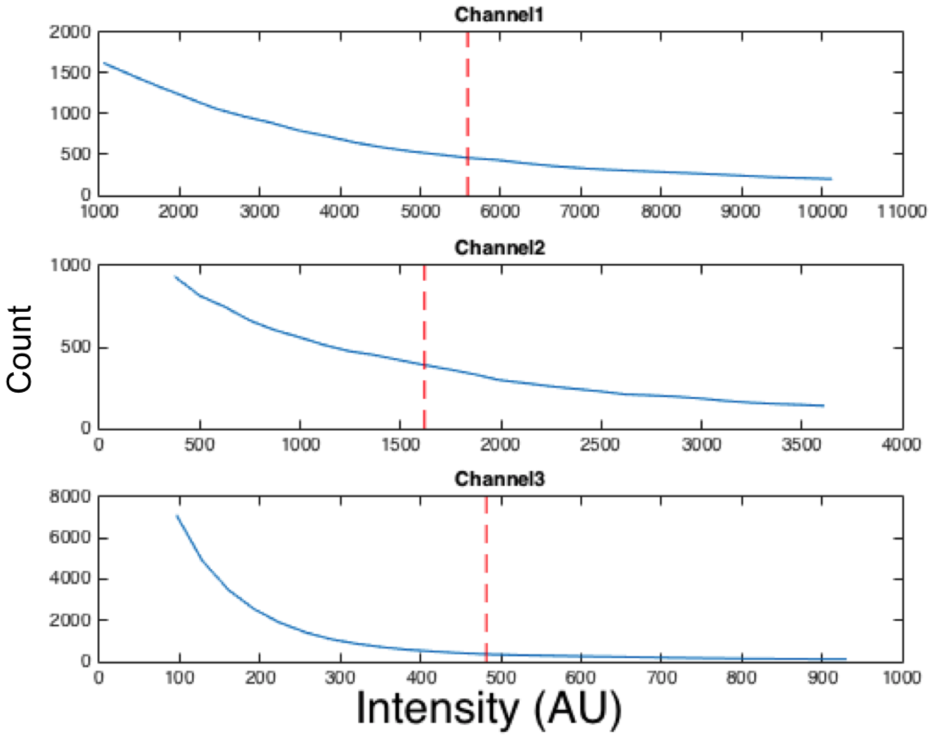
5. SPIM

   SPIM images were acquired using a custom built microscope for tissue clearing applications (Tomer 2014, Keller 2010, Treweek, in press). The SPIM design, and its complete list of components can be found in Treweek, in press.1-mm thick brain sections were positioned on thin glass slides and placed in a Spectrosil Quartz cuvette (Fisher scientific, Cat. # NC9520499), which was filled with a refractive index matching solution (RIMS, refractive index of 1.47; Treweek, in press). Thereafter, the cuvette was attached to a translation stage that lowered the sample to the imaging chamber, which was pre-filled with 100% glycerol. YFP and Cy3B fluorophores were excited using 473 nm and 561 nm lasers, respectively. The emitted light was collected using a 25× CLARITY objective, with a numerical aperture of 1.0 (Olympus), and a working distance of 8 mm. The image was digitized using a 4-megapixel cMOS camera (Andor Zyla 4.2), at a frame rate of 25 frames per second. The light-sheet microscope field-of-view was 480 × 480 m^2, and the voxel size was 0.234 m × 0.234 m × 1 m (16 bits). The two colors (YFP and Cy3B) were recorded sequentially where the YFP image stack (1000 images, ~8GB) was recorded first, followed by the Cy3b image stack (1000 images, ~8 GB). Imaris software (Bitplane) was used to compensate for the shift between the two color channels as well as for data analysis. In order to maximize the signal-to-noise ratio, a confocal slit detection method was utilized via the synchronization of the excitation Gaussian beam (full-width at half-maximum = 7-8 m) with the camera unidirectional readout (Baumgart 2012).

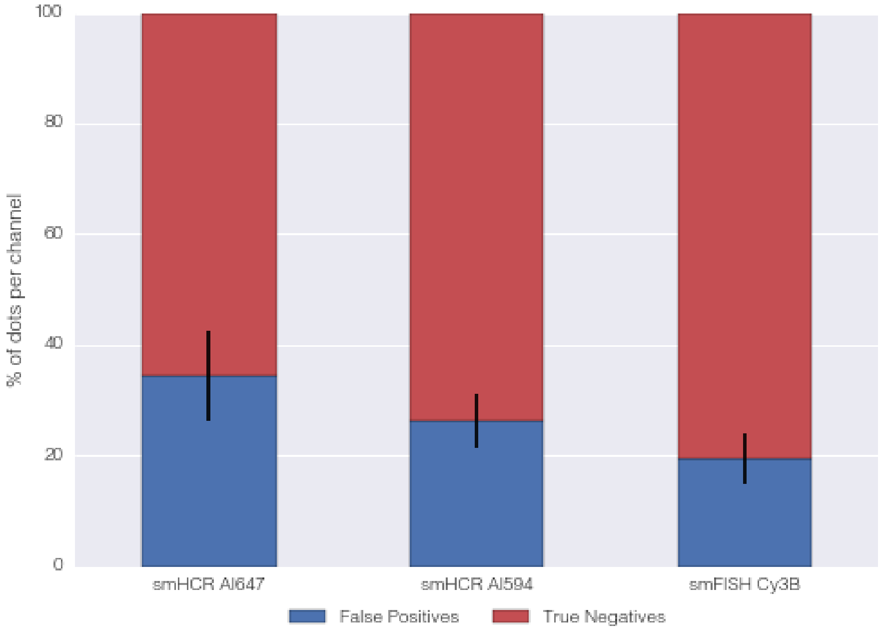
6. Calculating the percentage of YFP expressing cells

   11 field-of-views of cortical areas (60× oil objective numerical aperture of 1.42, wide field microscopy) were processed from a 1-mm Thy1-YFP brain slice. First, a group of YFP expressing cells was located and an image stack (~50 images with 1 µm spacing) was captured (DAPI and YFP). The volume of each stack was 74 m × 74 m × ~50 m. Then the number of DAPI stained cells and YFP expressing cells was counted manually (Supplementary Figure 33).

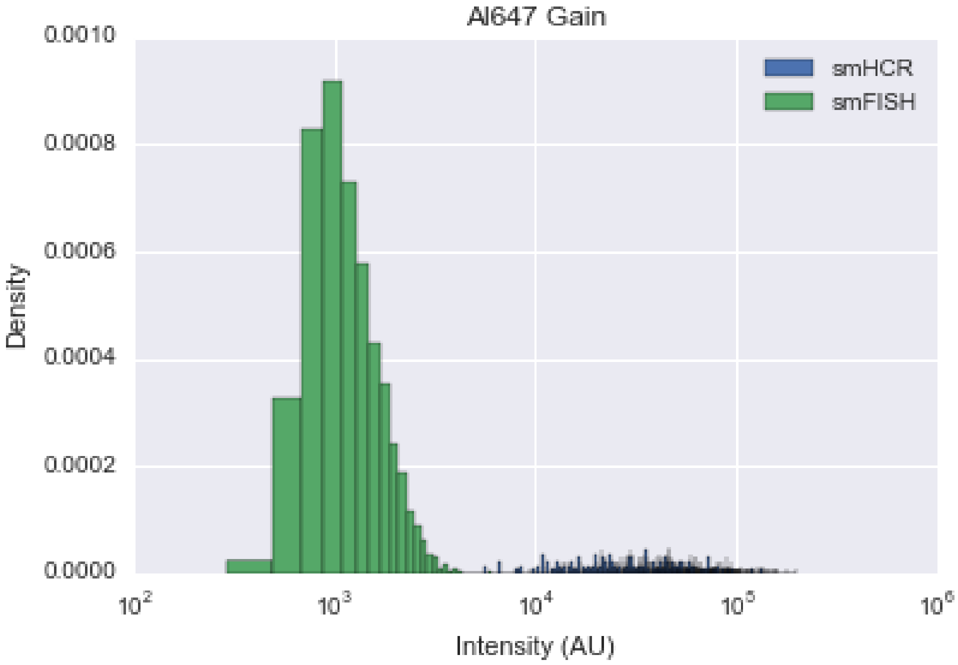
4.5.2 Supplementary Figures



Supplementary Figure 1. Thresholding for the 3-channel colocalization studies of Figure 4.1cd. Channel1 = smHCR B1 Alexa647, Channel2 = smHCR B3 Alexa594, Channel3 = smFISH Cy3B. Dashed lines depict thresholds selected for Figure 4.1cd. Target mRNA: Pcdh. Sample: cultured CAD cells.

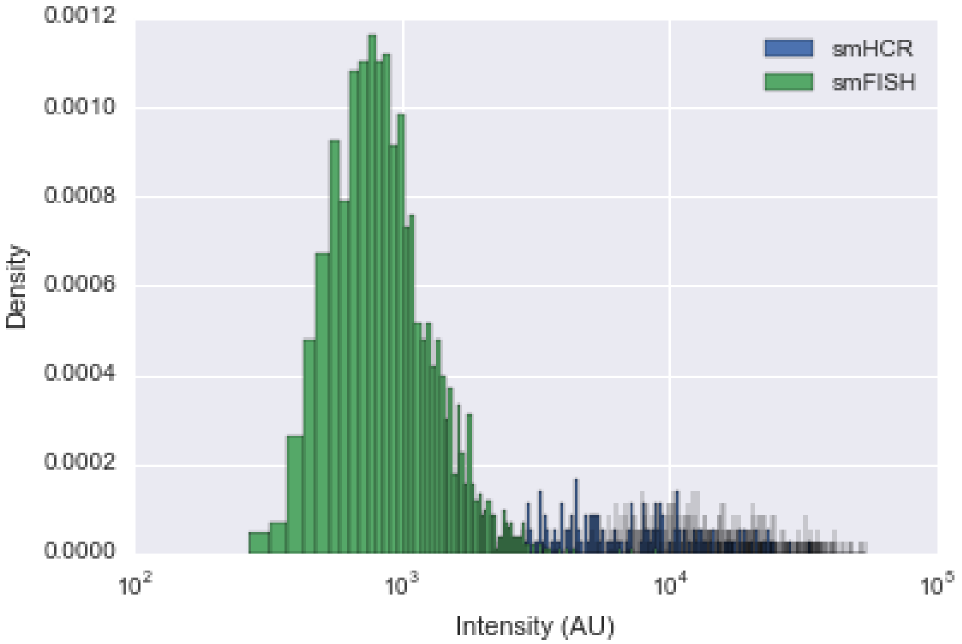


Supplementary Figure 2. Quantification of false positives for the three-channel colocalization studies of Figure 4.1cd. A dot that is present in only one channel is interpreted as a false positive in that channel and a true negative is present in the other two channels. Target mRNA: Pcdh. Sample: cultured CAD cells.

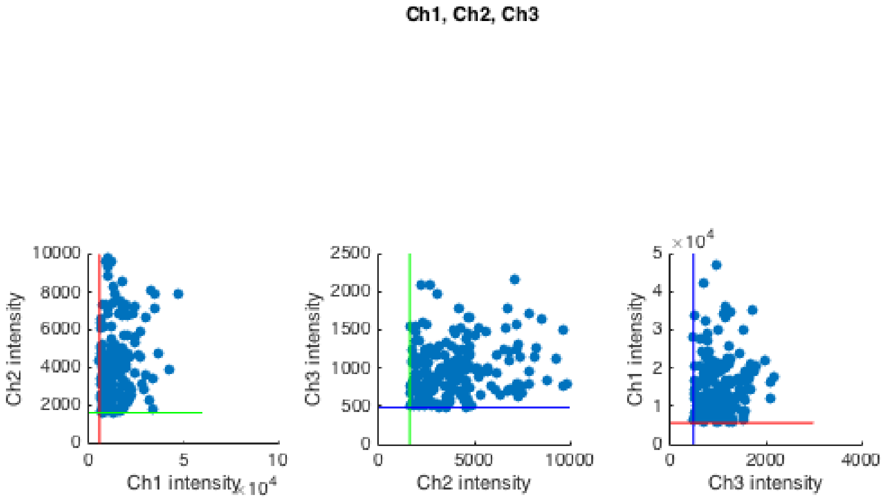


Supplementary Figure 3. Comparison of smFISH (Alexa 647) and smHCR (DNA HCR B1-Alexa 647) true-positive dot intensities. The ratio of median signal intensities is 36. Target mRNA:

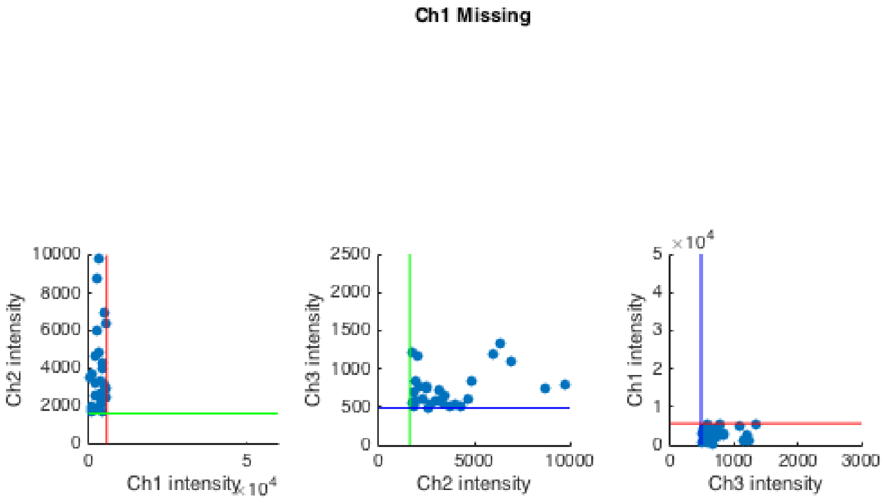
Pcdh. Sample: cultured CAD cells.



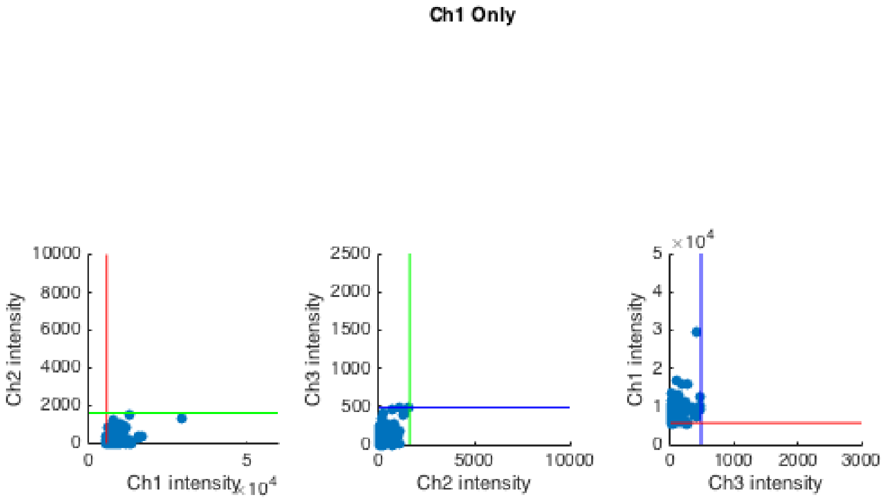
Supplementary Figure 4. Comparison of smFISH (Alexa 594) and smHCR (DNA HCR B3-Alexa 594) true-positive dot intensities. The ratio of median signal intensities is 13. Target mRNA: Pcdh. Sample: cultured CAD cells.



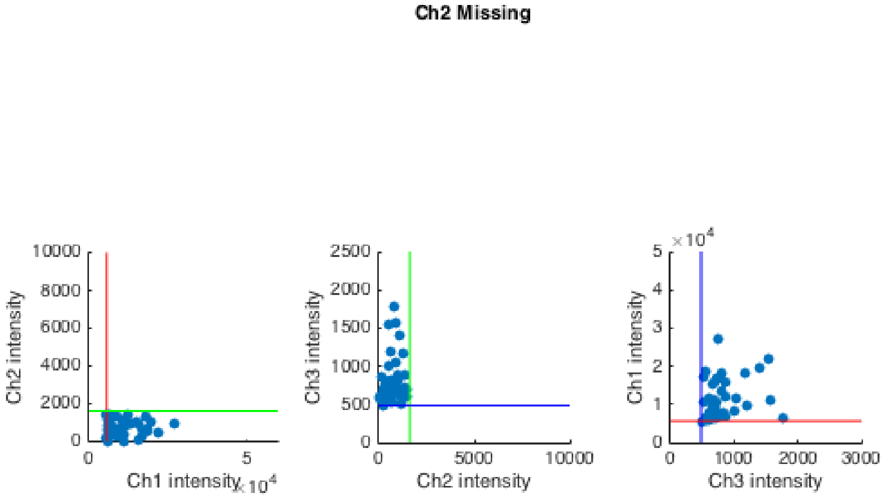
Supplementary Figure 5. Pairwise intensity of true-positive dots colocalized in all three channels in the study of Figure 4.1cd. Dot intensities depicted as the maximum pixel intensity within then dot. Pixel thresholds depicted as colored lines. Ch1 = Alexa 647 smHCR, Ch2 = Alexa 594 smHCR, Ch3 = Cy3B smFISH. Target mRNA: Pcdh. Sample: cultured CAD cells.



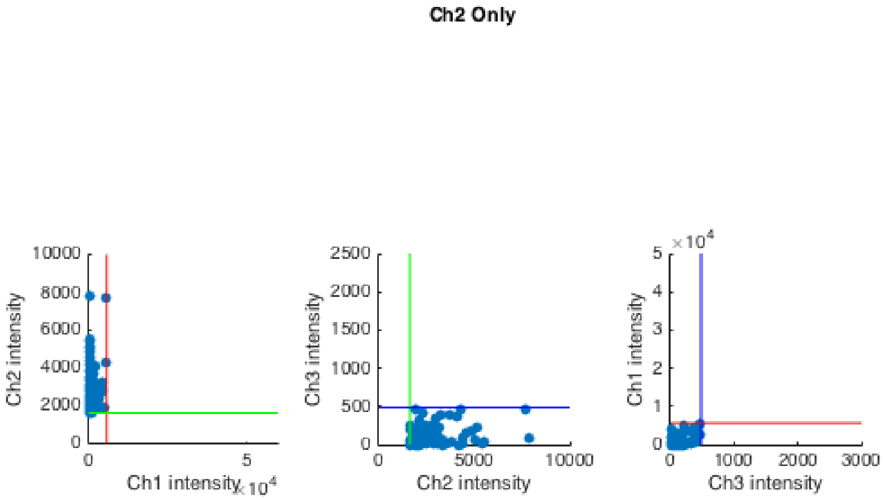
Supplementary Figure 6. Pairwise intensity of false-negative dots in Channel 1 for colocalization study of Figure 4.1cd. A dot that is absent in only one channel is interpreted to be a false negative in that channel. Dot intensities depicted as the maximum pixel intensity within the dot. Pixel thresholds depicted as colored lines. Ch1 = Alexa 647 smHCR, Ch2 = Alexa 594 smHCR, Ch3 = Cy3B smFISH. Target mRNA: Pcdh. Sample: cultured CAD cells.



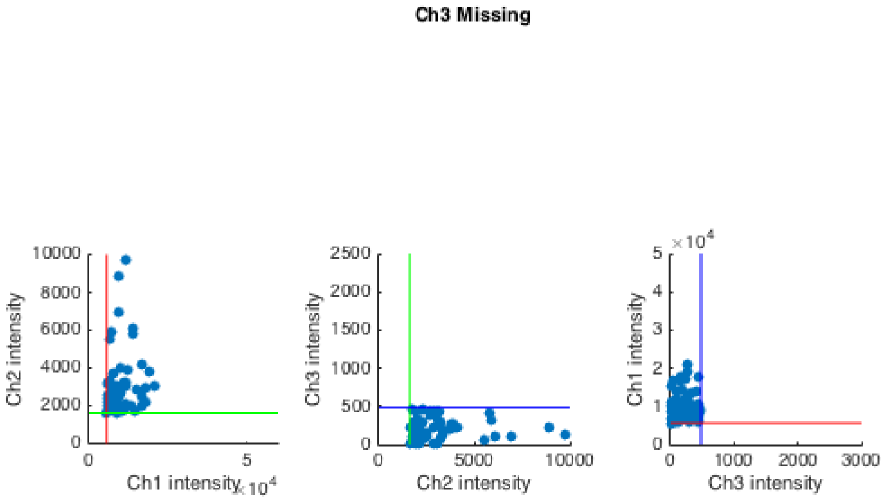
Supplementary Figure 7. Pairwise intensity of false-positive dots in Channel 1 for colocalization study of Figure 4.1cd. A dot that is present in only one channel is interpreted to be a false positive in that channel. Dot intensities depicted as the maximum pixel intensity within the dot. Pixel thresholds depicted as colored lines. Ch1 = Alexa647 smHCR, Ch2 = Alexa594 smHCR, Ch3 = Cy3B smFISH. Target mRNA: Pcdh. Sample: cultured CAD cells.



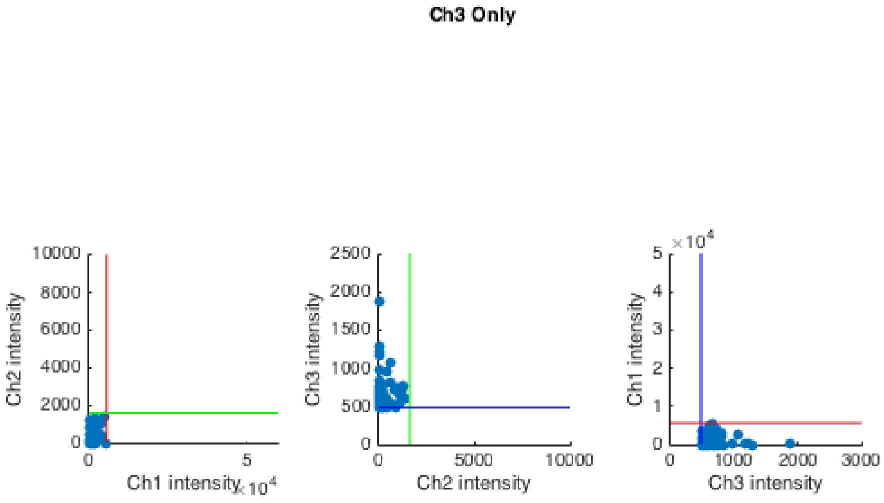
Supplementary Figure 8. Pairwise intensity of false-negative dots in Channel 2 for colocalization study of Figure 4.1cd. A dot that is absent in only one channel is interpreted to be a false negative in that channel. Dot intensities depicted as the maximum pixel intensity within the dot. Pixel thresholds depicted as colored lines. Ch1 = Alexa 647 smHCR, Ch2 = Alexa 594 smHCR, Ch3 = Cy3B smFISH. Target mRNA: Pcdh. Sample: cultured CAD cells.



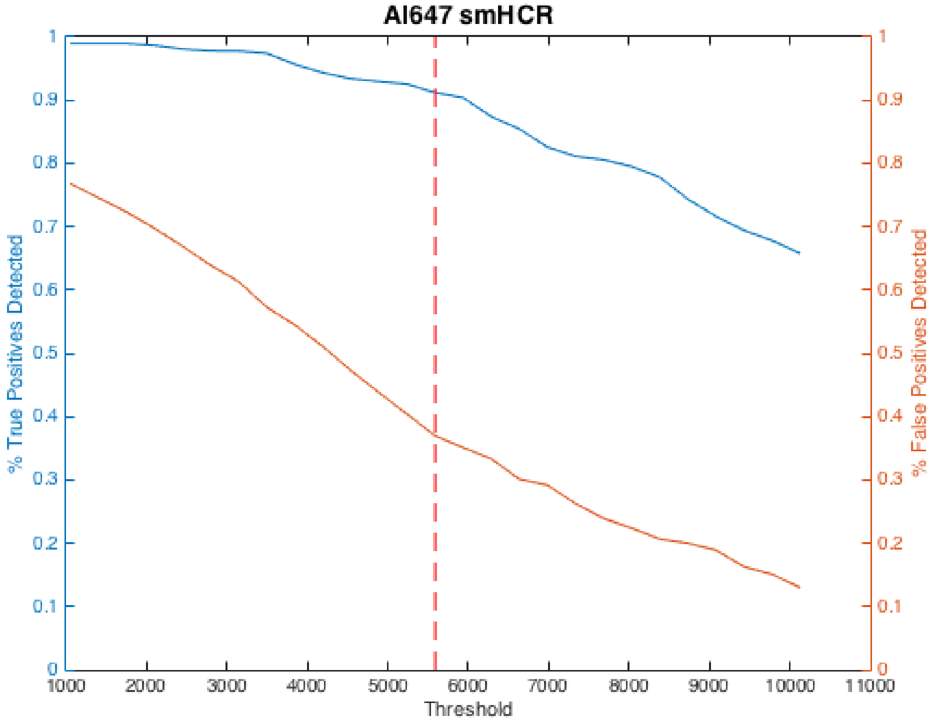
Supplementary Figure 9. Pairwise intensity of false-positive dots in Channel 2 for colocalization study of Figure 4.1cd. A dot that is present in only one channel is interpreted to be a false positive in that channel. Dot intensities depicted as the maximum pixel intensity within the dot. Pixel thresholds depicted as colored lines. Ch1 = Alexa 647 smHCR, Ch2 = Alexa 594 smHCR, Ch3 = Cy3B smFISH. Target mRNA: Pcdh. Sample: cultured CAD cells.



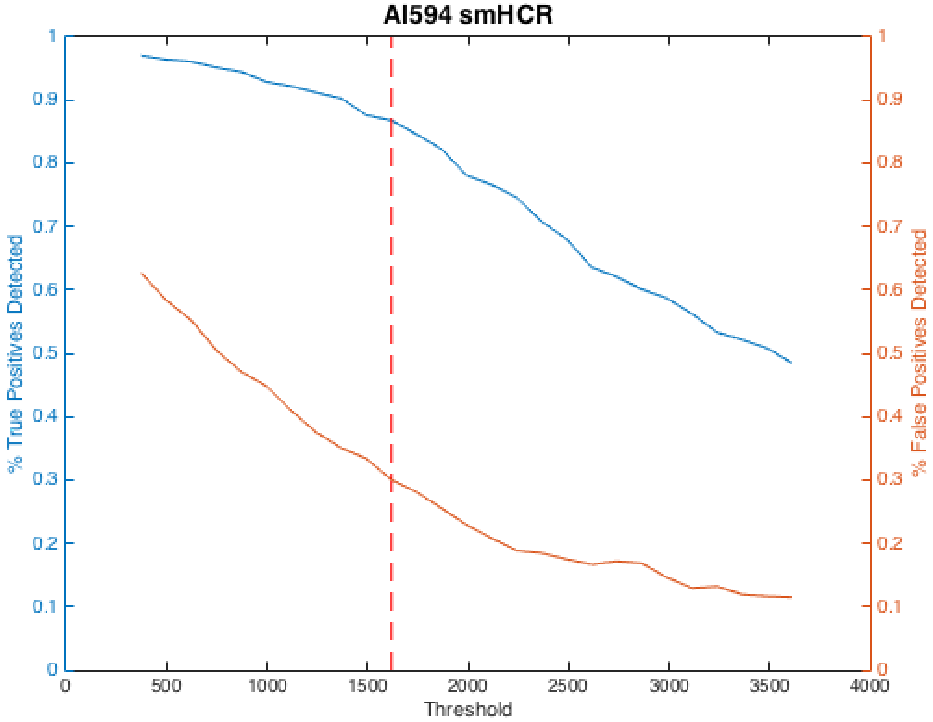
Supplementary Figure 10. Pairwise intensity of false-negative dots in Channel 3 for colocalization study of Figure 4.1cd. A dot that is absent in only one channel is interpreted to be a false negative in that channel. Dot intensities depicted as the maximum pixel intensity within the dot. Pixel thresholds depicted as colored lines. Ch1 = Alexa 647 smHCR, Ch2 = Alexa 594 smHCR, Ch3 = Cy3B smFISH. Target mRNA: Pcdh. Sample: cultured CAD cells.



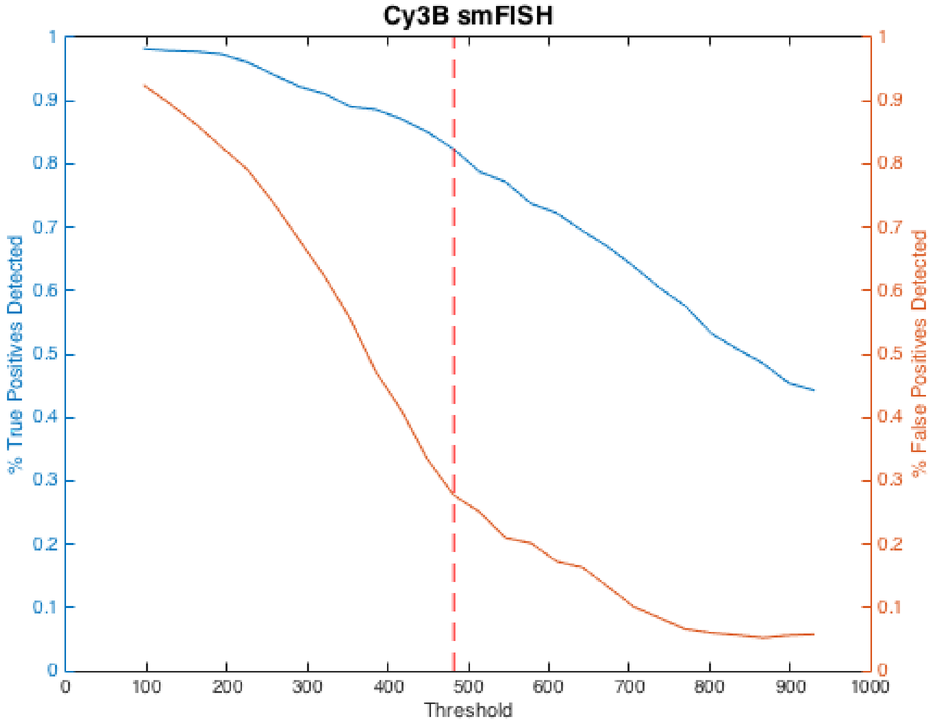
Supplementary Figure 11. Pairwise intensity of false-positive dots in Channel 3 for colocalization study of Figure 4.1cd. A dot that is present in only one channel is interpreted to be a false positive in that channel. Dot intensities depicted as the maximum pixel intensity within the dot. Pixel thresholds depicted as colored lines. Ch1 = Alexa647 smHCR, Ch2 = 594 smHCR, Ch3 = Cy3B smFISH. Target mRNA: Pcdh. Sample: cultured CAD cells.



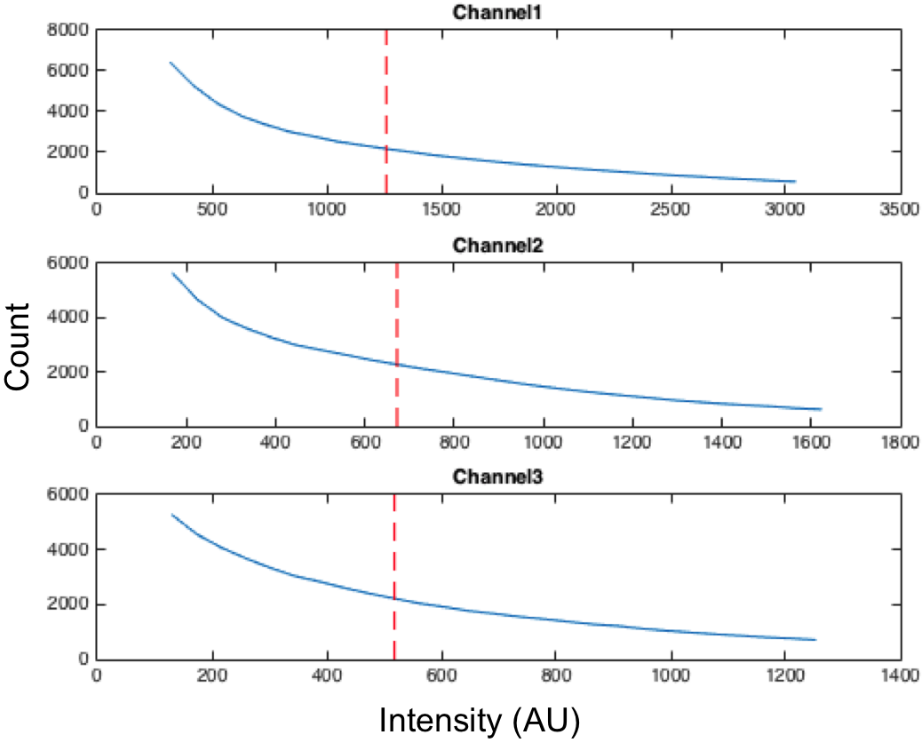
Supplementary Figure 12. Tradeoff between sensitivity and selectivity (Channel 1; Alexa 647 smHCR) as a function of pixel intensity threshold for the three-channel colocalization study of Figure 4.1cd. Dots that are present in at least two of the three channels are classified as true positives. A dot that is absent in only one channel is interpreted as a false negative in that channel; a dot that is present in only one channel is interpreted as a false positive in that channel. Threshold used for Figure 4.1cd depicted as a dashed line. Decreasing the threshold increases true positives (improving sensitivity) at the cost of increasing false positives (damaging selectivity). If each target is detected in two channels, false positives can be discarded based on colocalization information, enabling near-quantitative sensitivity. Target mRNA: Pcdh. Sample: cultured CAD cells.



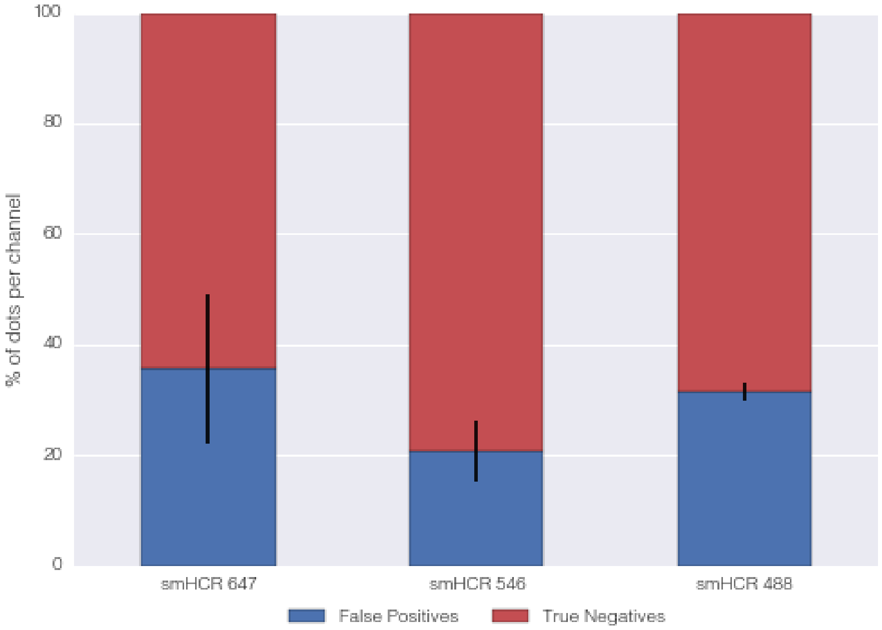
Supplementary Figure 13. Tradeoff between sensitivity and selectivity (Channel 2; Alexa 594 smHCR) as a function of pixel intensity threshold for the three-channel colocalization study of Figure 4.1cd. Dots that are present in at least two of the three channels are classified as true positives. A dot that is absent in only one channel is interpreted as a false negative in that channel; a dot that is present in only one channel is interpreted as a false positive in that channel. Threshold used for Figure 4.1cd depicted as a dashed line. Decreasing the threshold increases true positives (improving sensitivity) at the cost of increasing false positives (damaging selectivity). If each target is detected in two channels, false positives can be discarded based on colocalization information, enabling near-quantitative sensitivity. Target mRNA: Pcdh. Sample: cultured CAD cells.



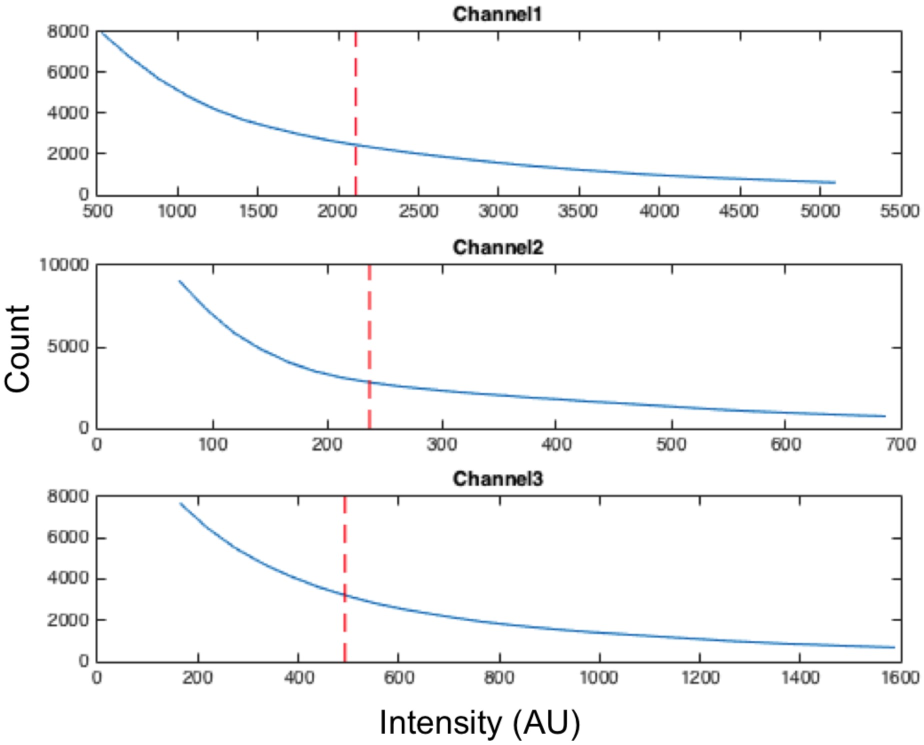
Supplementary Figure 14. Tradeoff between sensitivity and selectivity (Channel 3; Cy3B smFISH) as a function of pixel intensity threshold for the three-channel colocalization study of Figure 4.1cd. Dots that are present in at least two of the three channels are classified as true positives. A dot that is absent in only one channel is interpreted as a false negative in that channel; a dot that is present in only one channel is interpreted as a false positive in that channel. Threshold used for Figure 4.1cd depicted as a dashed line. Decreasing the threshold increases true positives (improving sensitivity) at the cost of increasing false positives (damaging selectivity). If each target is detected in two channels, false positives can be discarded based on colocalization information, enabling near-quantitative sensitivity. Target mRNA: Pcdh. Sample: cultured CAD cells.



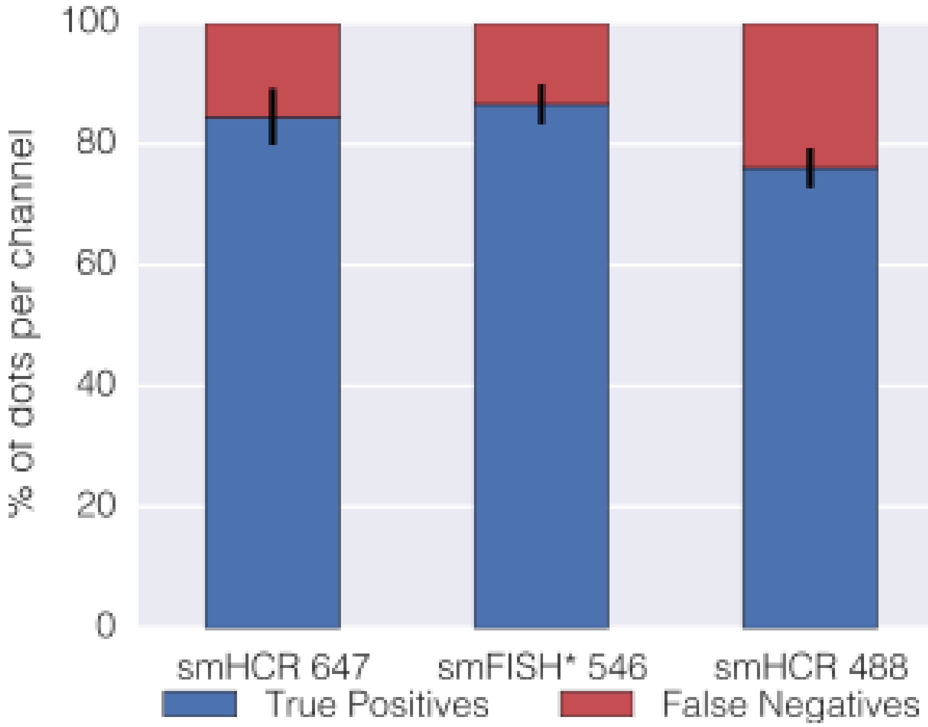
Supplementary Figure 15. Thresholding for the 3-channel colocalization studies of Figure 2a. Channel1 = smHCR Alexa647, Channel2 = smHCR Alexa546, Channel3 = smHCR Alexa488. Dashed lines depict thresholds selected for Figure 2a. Target mRNA: *kdrl*. Sample: whole-mount zebrafish embryo.



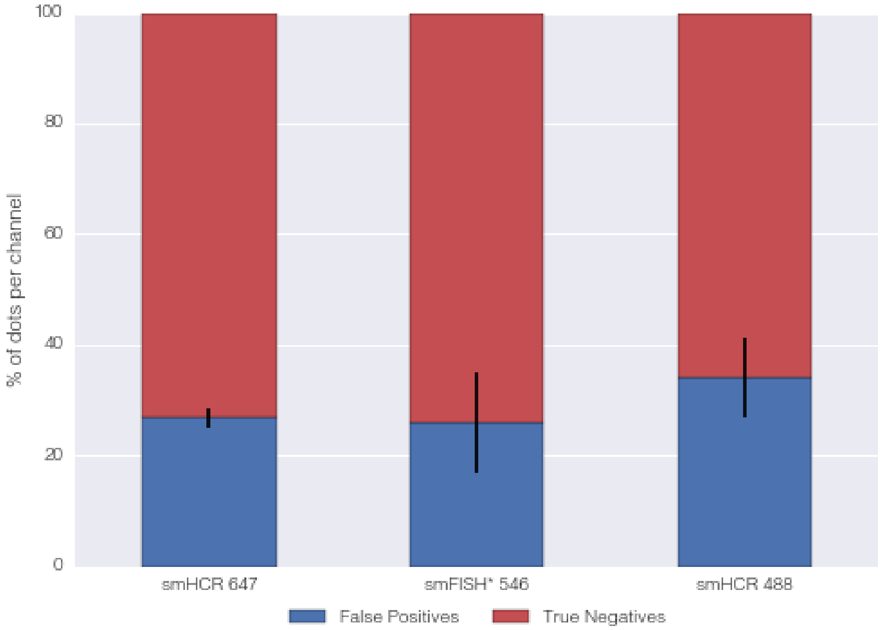
Supplementary Figure 16. Quantification of false positives for the three-channel colocalization studies of Figure 2a. A dot that is present in only one channel is interpreted as a false positive in that channel and a true negative in the other two channels. Target mRNA: *kdrl*. Sample: whole-mount zebrafish embryo.



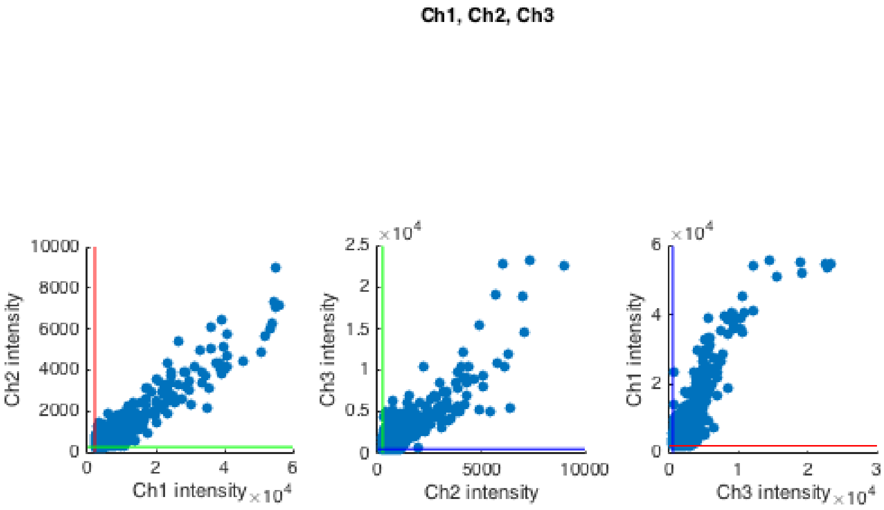
Supplementary Figure 17. Thresholding for the 3-channel colocalization studies of Figure 2c. Channel1 = smHCR Alexa647, Channel2 = smFISH* Alexa546, Channel3 = smHCR Alexa488. Dashed lines depict thresholds selected for Figure 2c. Target mRNA: *kdrl*. Sample: whole-mount zebrafish embryo.



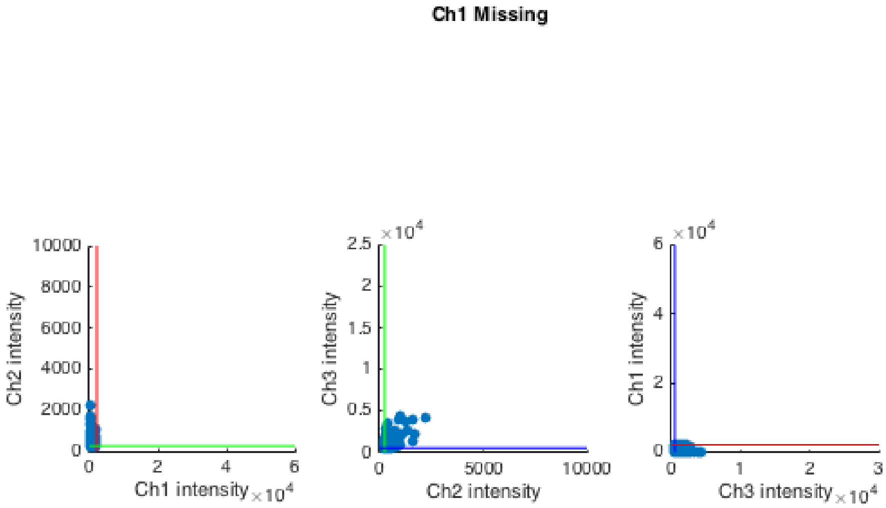
Supplementary Figure 18. Quantification of true positives for the three-channel colocalization studies of Figure 2c. Dots that are present in at least two of the three channels are classified as true positives. A dot that is absent in only one channel is interpreted as a false negative in that channel and as a true positive in the other two channels. Target mRNA: *kdrl*. Sample: whole-mount zebrafish embryo.



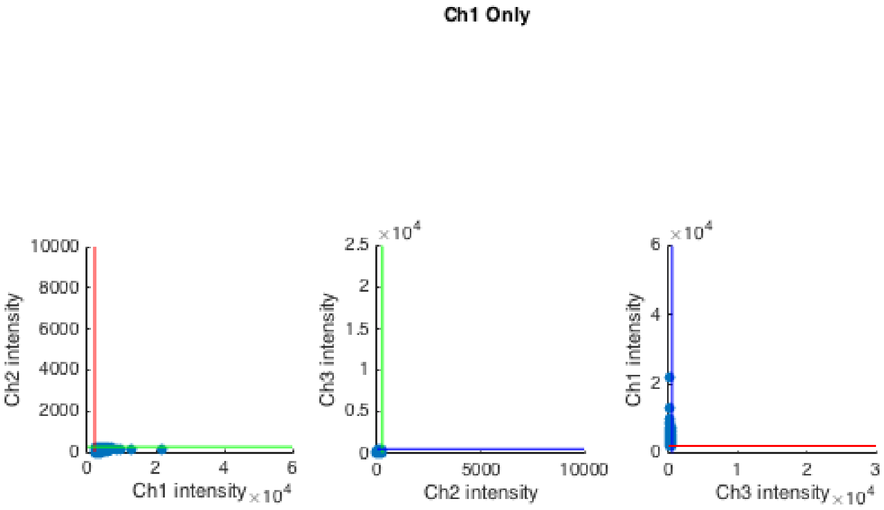
Supplementary Figure 19. Quantification of false positives for the three-channel colocalization studies of Figure 2c. A dot that is present in only one channel is interpreted as a false positive in that channel and as a true negative in the other two channels. Target mRNA: *kdrl*. Sample: whole-mount zebrafish embryo.



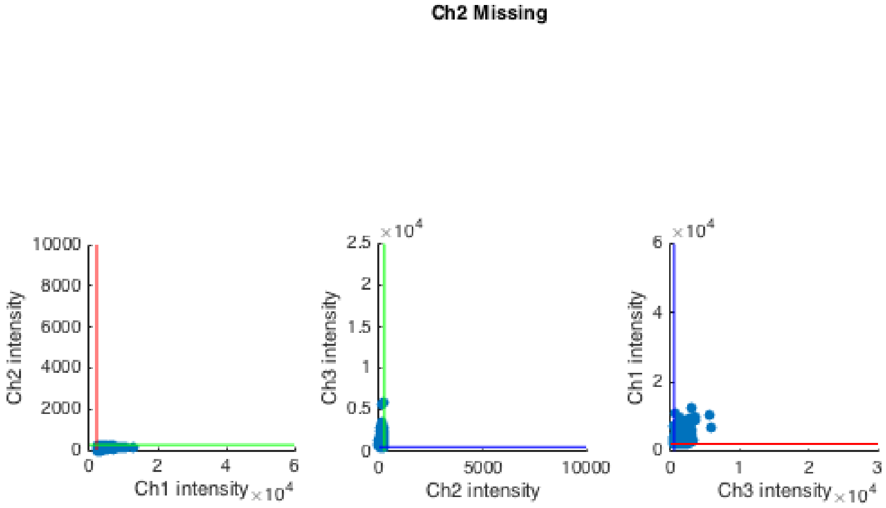
Supplementary Figure 20. Pairwise intensity of true-positive dots colocalized in all three channels in the study of Figure 2c. Dot intensities depicted as the maximum pixel intensity within then dot. Pixel thresholds depicted as colored lines. Ch1 = Alexa647 smHCR, Ch2 = Alexa546 smFISH*, Ch3 = Alexa488 smHCR. Target mRNA: *kdrl*. Sample: whole-mount zebrafish embryo.



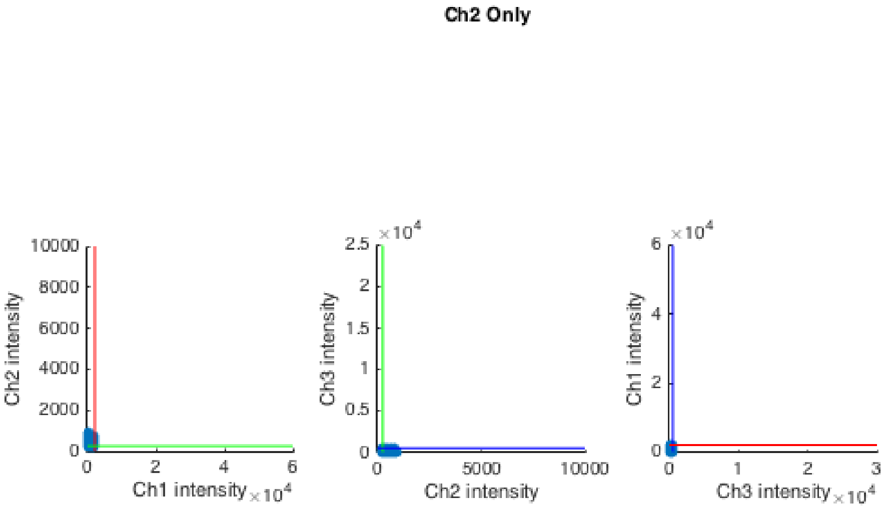
Supplementary Figure 21. Pairwise intensity of false-negative dots in Channel 1 for colocalization study of Figure 2c. A dot that is absent in only one channel is interpreted to be a false negative in that channel. Dot intensities depicted as the maximum pixel intensity within the dot. Pixel thresholds depicted as colored lines. Ch1 = Alexa647 smHCR, Ch2 = Alexa546 smFISH*, Ch3 = Alexa488 smHCR. Target mRNA: *kdrl*. Sample: whole-mount zebrafish embryo.



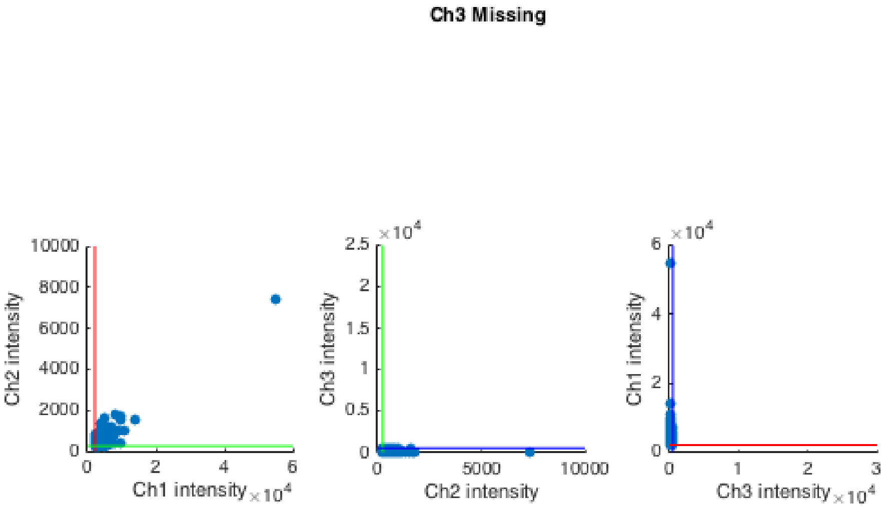
Supplementary Figure 22. Pairwise intensity of false-positive dots in Channel 1 for colocalization study of Figure 2c. A dot that is present in only one channel is interpreted to be a false positive in that channel. Dot intensities depicted as the maximum pixel intensity within the dot. Pixel thresholds depicted as colored lines. Ch1 = Alexa647 smHCR, Ch2 = Alexa546 smFISH*, Ch3 = Alexa488 smHCR. Target mRNA: *kdrl*. Sample: whole-mount zebrafish embryo.



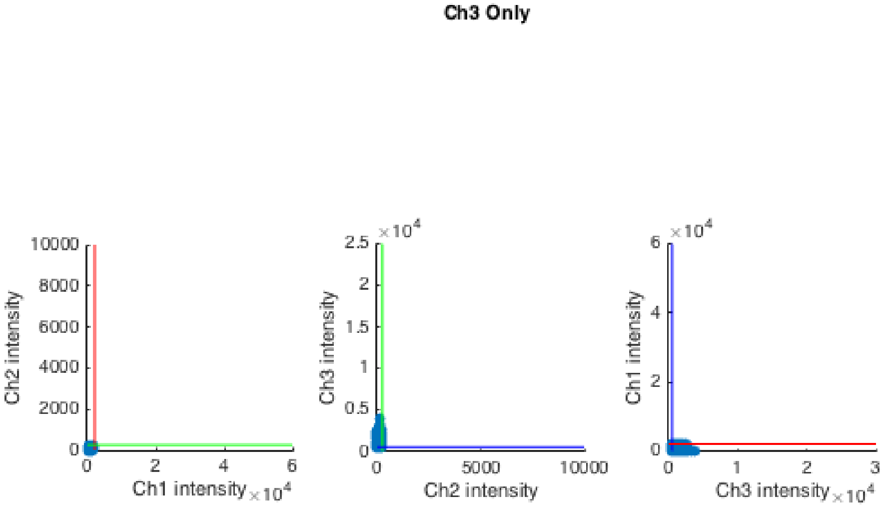
Supplementary Figure 23. Pairwise intensity of false-negative dots in Channel 2 for colocalization study of Figure 2c. A dot that is absent in only one channel is interpreted to be a false negative in that channel. Dot intensities depicted as the maximum pixel intensity within the dot. Pixel thresholds depicted as colored lines. Ch1 = Alexa647 smHCR, Ch2 = Alexa546 smFISH*, Ch3 = Alexa488 smHCR. Target mRNA: *kdrl*. Sample: whole-mount zebrafish embryo.



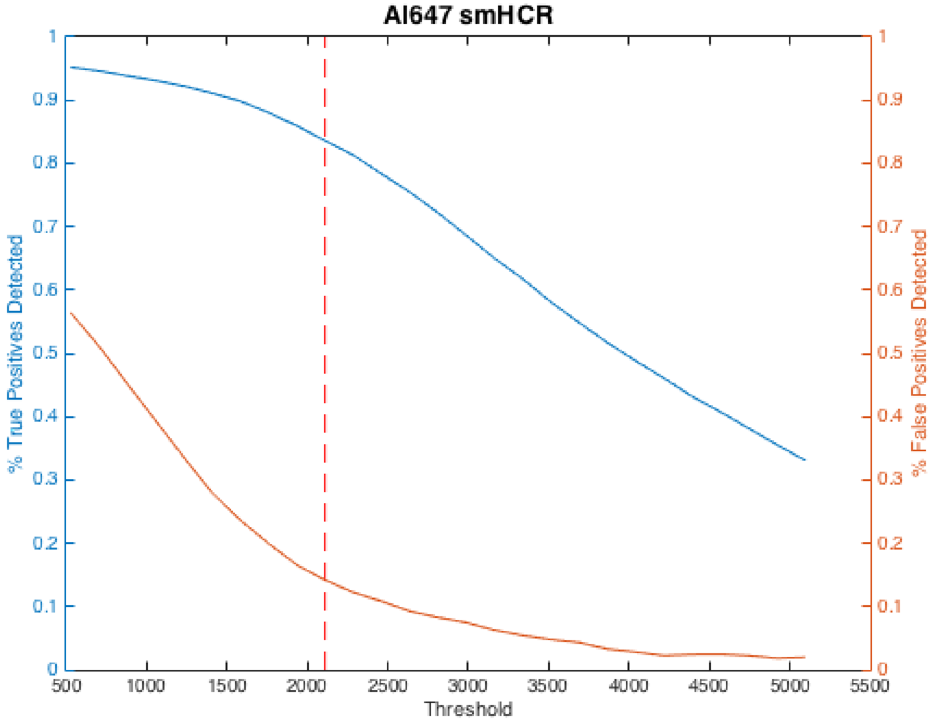
Supplementary Figure 24. Pairwise intensity of false-positive dots in Channel 2 for colocalization study of Figure 2c. A dot that is present in only one channel is interpreted to be a false positive in that channel. Dot intensities depicted as the maximum pixel intensity within the dot. Pixel thresholds depicted as colored lines. Ch1 = Alexa647 smHCR, Ch2 = Alexa546 smFISH*, Ch3 = Alexa488 smHCR. Target mRNA: *kdrl*. Sample: whole-mount zebrafish embryo.



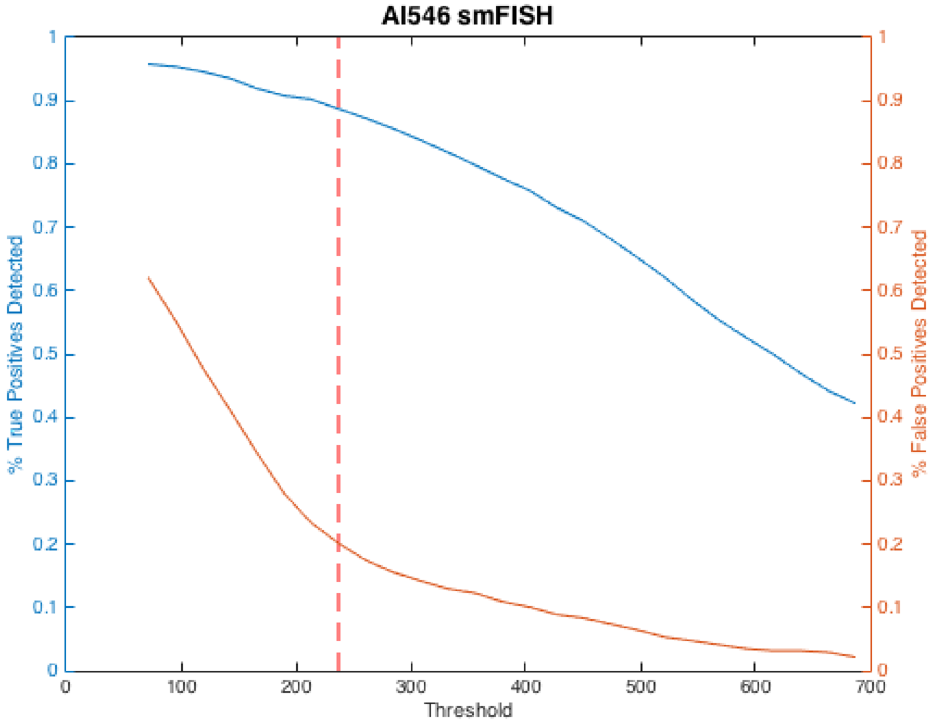
Supplementary Figure 25. Pairwise intensity of false-negative dots in Channel 3 for colocalization study of Figure 2c. A dot that is absent in only one channel is interpreted to be a false negative in that channel. Dot intensities depicted as the maximum pixel intensity within the dot. Pixel thresholds depicted as colored lines. Ch1 = Alexa647 smHCR, Ch2 = Alexa546 smFISH*, Ch3 = Alexa488 smHCR. Target mRNA: *kdrl*. Sample: whole-mount zebrafish embryo.



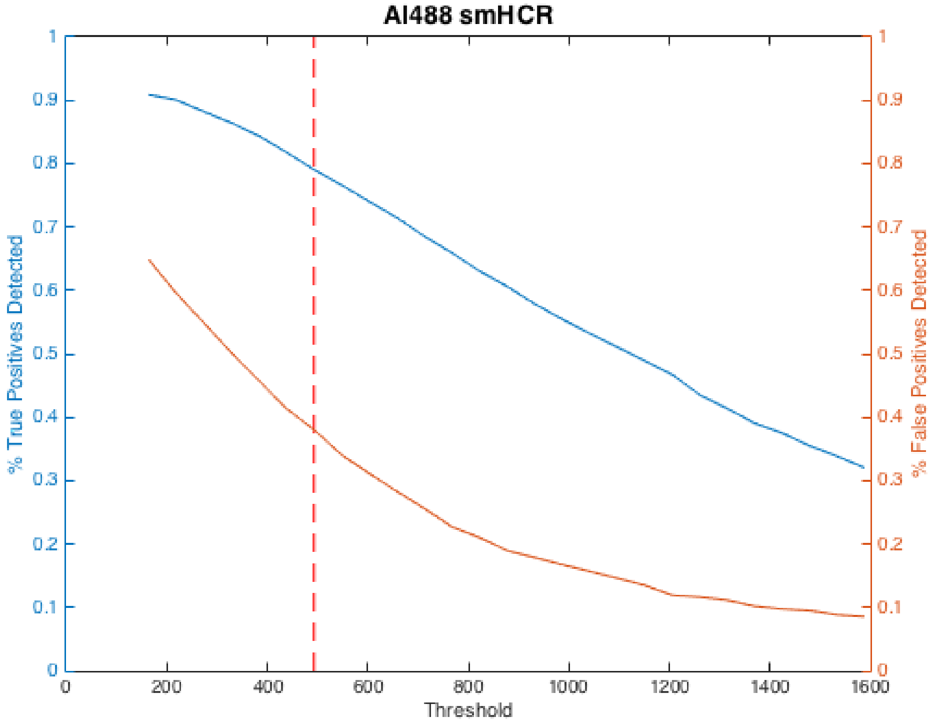
Supplementary Figure 26. Pairwise intensity of false-positive dots in Channel 3 for colocalization study of Figure 2c. A dot that is present in only one channel is interpreted to be a false positive in that channel. Dot intensities depicted as the maximum pixel intensity within the dot. Pixel thresholds depicted as colored lines. Ch1 = Alexa647 smHCR, Ch2 = Alexa546 smFISH*, Ch3 = Alexa488 smHCR. Target mRNA: *kdrl*. Sample: whole-mount zebrafish embryo.



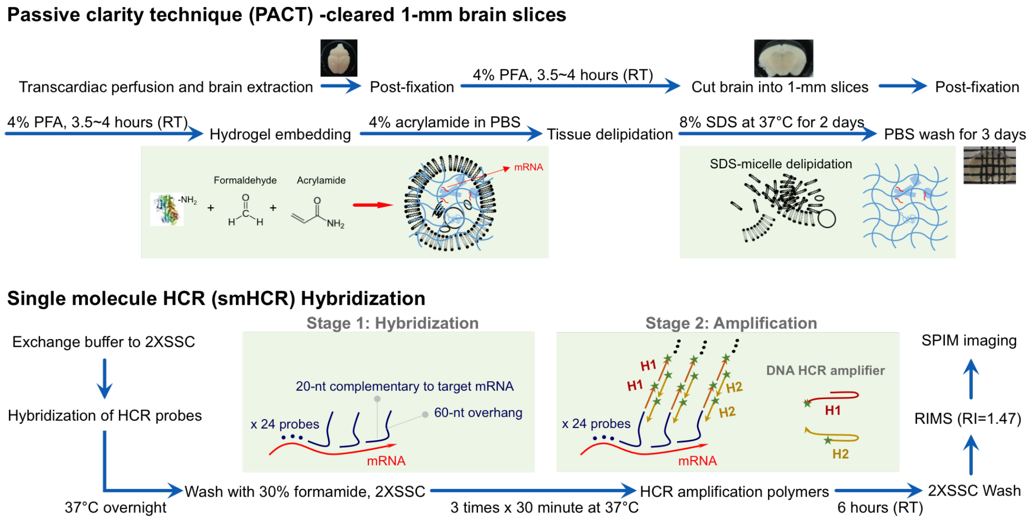
Supplementary Figure 27. Tradeoff between sensitivity and selectivity (Channel 1; Alexa 647 smHCR) as a function of pixel intensity threshold for the three-channel colocalization study of Figure 2c. Dots that are present in at least two of the three channels are classified as true positives. A dot that is absent in only one channel is interpreted as a false negative in that channel; a dot that is present in only one channel is interpreted as a false positive in that channel. Threshold used for Figure 2c depicted as a dashed line. Decreasing the threshold increases true positives (improving sensitivity) at the cost of increasing false positives (damaging selectivity). If each target is detected in two channels, false positives can be discarded based on colocalization information, enabling near-quantitative sensitivity. Target mRNA: *kdrl*. Sample: whole-mount zebrafish embryo.



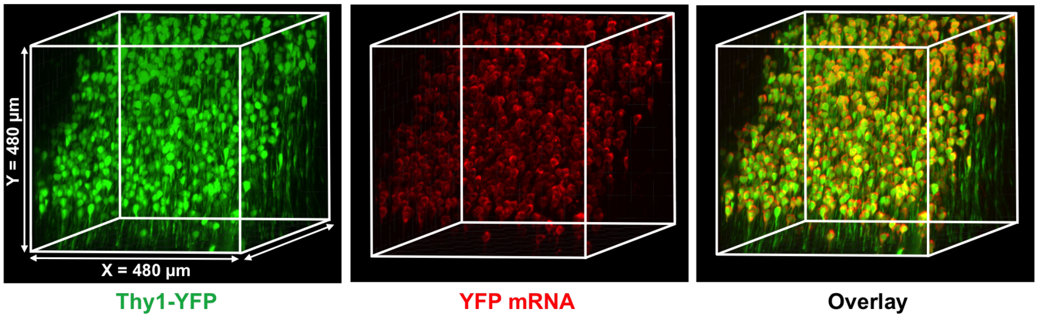
Supplementary Figure 28. Tradeoff between sensitivity and selectivity (Channel 2; Alexa 546 smFISH*) as a function of pixel intensity threshold for the three-channel colocalization study of Figure 2c. Dots that are present in at least two of the three channels are classified as true positives. A dot that is absent in only one channel is interpreted as a false negative in that channel; a dot that is present in only one channel is interpreted as a false positive in that channel. Threshold used for Figure 2c depicted as a dashed line. Decreasing the threshold increases true positives (improving sensitivity) at the cost of increasing false positives (damaging selectivity). If each target is detected in two channels, false positives can be discarded based on colocalization information, enabling near-quantitative sensitivity. Target mRNA: *kdrl*. Sample: whole-mount zebrafish embryo.



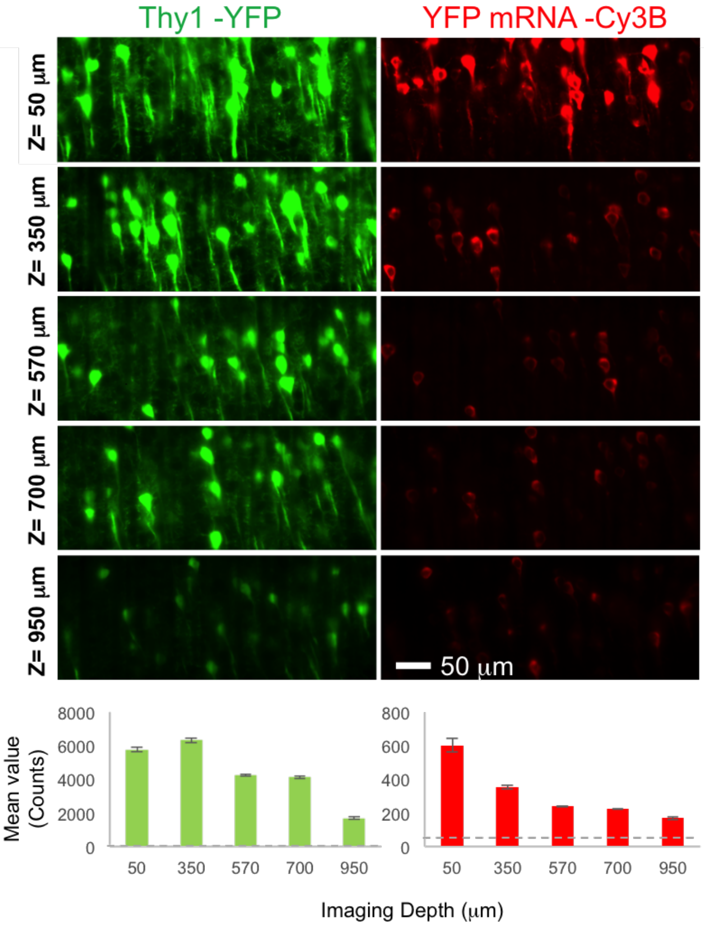
Supplementary Figure 29. Tradeoff between sensitivity and selectivity (Channel 3; Alexa 488 smHCR) as a function of pixel intensity threshold for the three-channel colocalization study of Figure 2c. Dots that are present in at least two of the three channels are classified as true positives. A dot that is absent in only one channel is interpreted as a false negative in that channel; a dot that is present in only one channel is interpreted as a false positive in that channel. Threshold used for Figure 2c depicted as a dashed line. Decreasing the threshold increases true positives (improving sensitivity) at the cost of increasing false positives (damaging selectivity). If each target is detected in two channels, false positives can be discarded based on colocalization information, enabling near-quantitative sensitivity. Target mRNA: *kdrl*. Sample: whole-mount zebrafish embryo.



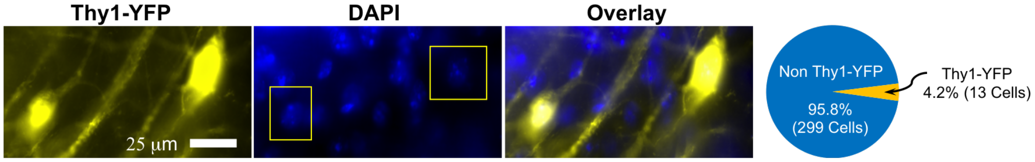
Supplementary Figure 30. PACT-clearing and smHCR protocol overview. (Top) PACT flow chart. The main deviation from the standard PACT protocol (Yang et al, Cell, 2014) is an additional fixation step after brain slicing, in order to better preserve mRNA transcripts. (Bottom) smHCR flow chart. The PACT-cleared brain slices were first hybridized with DNA HCR probes (probe concentration: 10 nM) at 37 °C overnight, and washed in high-stringency conditions (30% formamide, 2× SSC at 37 °C for 1.5 hours) to remove the free and nonspecifically-bound probes. During HCR amplification, hairpins (120 nM) were added to the sample and incubated for 6 hours at RT. The HCR brain slices were then imaged in RIMS (refractive index 1.47; Yang et al, 2014) using SPIM microscope (Treweek, in press).



Supplementary Figure 31. 3D reconstruction of 1-mm PACT-cleared Thy1-YFP mouse brain slice. (Left) Endogenous YFP Protein. (Middle) Cy3B labeling of YFP mRNA. (Right) Overlay shows the specificity of the HCR for YFP transcripts throughout the thick 1-mm slice.



Supplementary Figure 32. Fluorescence intensity decreases as a function of tissue depth. An attenuation in the fluorescence signal of the YFP and smHCR-Cy3B channels was observed as a function of tissue depth. However, even at 1-mm depth, and without light power compensation, the HCR hybridization still shows a detectable signal that co-localizes with the YFP signal. The dashed line in the lower graph indicates the camera dark count.



Supplementary Figure 33. Across a 1-mm cleared Thy1-YFP mouse brain slice, only a small subset of brain cells show YFP expression. (Left) Widefield microscope image (YFP, DAPI, and overlay); yellow boxes in the DAPI image indicate the location of YFP expressing cells. (Right) YFP expressing cells account for less than 5% (13 cells) of the total number of cells (312 cells).

4.5.3 References

Baumgart, E., and Kubitscheck, U. (2012). Scanned light sheet microscopy with confocal slit detection. *Optics Express* 20, 21805- 21814.

Choi, H.M.T., Beck, V.A. and Pierce, N.A. (2014). Next-generation in situ hybridization chain reaction: higher gain, lower cost, greater durability. *ACS Nano* 8(5):4284-4294.

Keller, P.J., Schmidt, A.D., Santella, A., Khairy, K., Bao, Z., Wittbrodt, J., and Stelzer, E.H. (2010). Fast, high-contrast imaging of animal development with scanned light sheet-based structured-illumination microscopy. *Nature Methods* 7, 637-642.

Lubeck, E. & Cai, L. Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nature Methods* 9, 743–748 (2012).

Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. & Cai, L. Single-cell in situ RNA profiling by sequential hybridization. *Nature Methods* 11, 360–361 (2014).

Tomer, R., Ye, L., Hsueh B., and Deisseroth K. (2014). Advanced CLARITY for rapid and high-resolution imaging of intact tissues. *Nature Protocols* 9, 1682-1697.

Treweek, J.B., Chan, K.Y., Flytzanis, N.C., Yang, B., Deverman, B.E., Greenbaum, A., Lignell, A., Xiao, C., Cai, L., Ladinsky, M.S., Bjorkman, P.J., Fowlkes, C.C. and Gradinaru, V (2015). Whole-Body Tissue Stabilization and Selective Extractions via Tissue-Hydrogel Hybrids for High Resolution Intact Circuit Mapping and Phenotyping. /Nature Protocol/s, dx.doi.org/10.1038/nprot.2015.122).