Super-Resolution Characterization of Synaptic Neurexin-1 and Development of Highly Multiplexed RNA Imaging in the Brain

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th>Citation</th>
<th>Hao, Junjie. 2019. Super-Resolution Characterization of Synaptic Neurexin-1 and Development of Highly Multiplexed RNA Imaging in the Brain. Doctoral dissertation, Harvard University, Graduate School of Arts &amp; Sciences.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:42013099">http://nrs.harvard.edu/urn-3:HUL.InstRepos:42013099</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Super-resolution characterization of synaptic neurexin-1 and development of highly multiplexed RNA imaging in the brain

A dissertation presented
by
Junjie Hao
to
The Department of Chemistry and Chemical Biology
In partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Chemistry

Harvard University
Cambridge, Massachusetts
Aug 2019
Super-resolution characterization of synaptic neurexin-1 and development of highly multiplexed RNA imaging in the brain

Abstract

Imaging techniques, applied at different length scales, can shed light on the complex, interconnected molecular and cellular networks that comprise the brain – how are different classes or types of neurons connected, and why are they connected that way? Variations in synaptic connections reflect not only in the type of neurons joined by a chemical synapse, but also the properties of synaptic release and plasticity. One of the most attractive candidates for explaining the high degree of variability observed in synaptic junctions is the trans-synaptic interaction between neurexins and their ligands, owing to the high degree of diversity possible due to the combinatorial nature of their alternative splicing and pairing. Understanding the synaptic localization and organization of neurexins can help elucidate the mechanism by which they function. A spatially-resolved expression profile on the transcriptomic scale of intact brain tissue can then be used to define transcriptionally distinct neuron types within a neuronal network, which can be correlated with molecular information from their synapses to examine the relationship between neurexins and synaptic identity.
The first part of this thesis reports our efforts to probe the synaptic nanostructure and dynamics of neurexin-1 using multicolor 3D STORM. We found that synaptic neurexin-1 was organized into nanoclusters in a subset of Homer1(+) excitatory synapses, increasing in fraction over the course of development in cultured neurons and tissue from the mouse hypothalamus, up to ~40%. These nanoclusters also increased in size and neurexin content as the neurons mature. Furthermore, we found that the presence of synaptic neurexin-1 was correlated with an increase in markers of synaptic activity and strength. Neurexin-1 nanoclusters were dynamically regulated via ectodomain cleavage by ADAM10, and chemical inhibition of ADAM10 activity resulted in a doubling of the fraction of neurexin-1(+) synapses in cultured neurons. This cleavage is likely to be physiologically relevant, as neurexin-1 ectodomain fragments accounted for ~4-6% of total neurexin-1 levels in the mouse brain, varying across different brain regions.

The remainder of this thesis describes our improvements to MERFISH towards quantifying spatially-resolved transcriptomic-scale expression profiling in intact brain tissue. By switching our method of signal removal between hybridization rounds from photobleaching to chemical reduction, restricting our readout probe sequences to a 3-letter alphabet, scaling up the field of view area, and optimizing the acquisition and analysis software, we increased MERFISH throughput by a factor of ~100, and demonstrate the ability to measure ~40,000 cells in 18 hr. The experimental changes that enabled this significant increase in measurement throughput did not negatively affect the quality of MERFISH when measured against bulk sequencing and smFISH standards. We devised a method to remove background signal in MERFISH by casting a polyacrylamide gel over the sample, imprinting the RNA signal directly into the gel matrix, and removing proteins and lipids through clearing. In cleared cultured cell samples, we demonstrated the ability to perform MERFISH measurements in up to 4 colors, and detected no
loss in RNA counts despite removing cellular structural components. Additionally, we performed MERFISH measurements in tissue sections from the mouse hypothalamus medial preoptic area, and found good correlation in RNA counts measured by MERFISH and average expression levels measured by bulk sequencing of the same brain region.

Understanding the mechanisms and principles that govern organization of neuronal circuits in the brain requires probing biological interactions ranging in scale from nanometers between molecules to millimeters or more between cells. STORM and MERFISH are two powerful imaging techniques applicable to these length scales, and as both techniques continue to be improved, can perhaps together clarify complex trans-synaptic code of alternatively spliced isoforms of neurexins and their ligands, and its role in shaping synaptic identity and function.
# Table of Contents

Title ................................................................................................................................. i
Copyright........................................................................................................................... ii
Abstract............................................................................................................................. iii
Table of contents.............................................................................................................. vi
Citation to previously published work............................................................................ ix
Acknowledgements.......................................................................................................... x

Chapter 1: Introduction .................................................................................................... 1
  1.1 Neurexins are functionally important molecules responsible for synaptic function and maintenance ............................................................................................................. 1
  1.2 Disruption to neurexin function is implicated in neurological disorders .................. 3
  1.3 Applications of super-resolution imaging in neurobiology ...................................... 4
  1.4 Multicolor 3D super-resolution with STORM .......................................................... 5
  1.5 Transcriptome-level studies of brain tissue ................................................................ 6
  1.6 MERFISH for spatially-resolved transcriptomics .................................................... 7
  1.7 Outline of thesis ...................................................................................................... 9

Chapter 2: Synaptic Neurexin-1 Assembles into Dynamically Regulated Active Zone Nanoclusters ............................................................................................................. 11
  2.1 Introduction ............................................................................................................ 11
  2.2 Results .................................................................................................................. 14
    2.2.1 Generation of conditional knockin (cKI) mice expressing HA-tagged Nrxn1 ...... 14
    2.2.2 Nrxn1 localizes to unique nanoclusters at synaptic junctions ............................. 18
    2.2.4 Synaptic Nrxn1 nanoclusters are formed independent of HA-epitope tagging and include other neurexin isoforms ............................................................... 24
    2.2.5 Functional Differences in Synapses Containing Nrxn1 Clusters ......................... 29
    2.2.6 Dynamics of Nrxn1 expression and processing .................................................. 36
  2.3 Discussion ............................................................................................................. 45
  2.4 Materials and methods ......................................................................................... 49
    2.4.2 Cell Culture .................................................................................................... 52
    2.4.3 Quantitative real time PCR (qPCR) ................................................................. 53
2.4.4 Lentivirus Production and Infection ................................................................. 55
2.4.5 Electrophysiology ............................................................................................... 57
2.4.6 Immunocytochemistry ....................................................................................... 57
2.4.7 Excitatory Synapse Quantification .................................................................... 60
2.4.8 Tissue Preparation and Staining for STORM ................................................... 60
2.4.9 Generation of Expression vectors ....................................................................... 65
2.4.10 Immunoprecipitation ......................................................................................... 66
2.4.11 Cell lysis ............................................................................................................. 66
2.4.12 Immunoblotting ................................................................................................. 67
2.4.13 Studying Processing in Neurons ....................................................................... 68
2.4.14 Fractionation ....................................................................................................... 68
2.4.15 Quantification and Statistical Analysis ............................................................... 70

Chapter 3: High-throughput single-cell gene expression profiling with multiplexed error-robust fluorescence in situ hybridization (MERFISH) ................................................................. 71

3.1 Introduction ............................................................................................................. 71
3.2 Results ..................................................................................................................... 76
  3.2.1 Increasing the throughput of MERFISH measurements .................................... 76
3.3 Discussion ............................................................................................................... 92
3.4 Materials and methods ........................................................................................... 93
  3.4.1 Encoding probe design ....................................................................................... 93
  3.4.2 Encoding probe construction ............................................................................. 99
  3.4.3 Readout probe construction ............................................................................. 99
  3.4.4 High-throughput imaging platform .................................................................. 100
  3.4.5 Encoding probe staining .................................................................................. 101
  3.4.6 MERFISH imaging ............................................................................................ 103
  3.4.7 Single-Molecule fluorescence in situ hybridization (smFISH) ......................... 104
  3.4.8 Image registration and decoding ....................................................................... 104
  3.4.9 Cell segmentation ............................................................................................ 108
  3.4.10 RNA-seq ........................................................................................................ 110

Chapter 4: High performance multiplexed fluorescence in situ hybridization in culture and tissue with matrix imprinting and clearing ................................................................. 111
4.1 Introduction ........................................................................................................................................... 111
4.2 Results .................................................................................................................................................. 115
  4.2.1 A matrix imprinting and clearing approach to reduce the background for smFISH measurements ........................................................................................................................................... 115
  4.2.2 RNA is preserved during matrix imprinting and clearing ................................................................. 121
  4.2.3 Extending MERFISH to four-color imaging ....................................................................................... 124
  4.2.4 MERFISH measurements of brain tissue .......................................................................................... 127
4.3 Discussion ............................................................................................................................................ 129
4.4 Materials and methods ....................................................................................................................... 133
  4.4.1 MERFISH probe library design and construction .............................................................................. 133
  4.4.2 Silanization of coverslips .................................................................................................................. 134
  4.4.3 Cell culture and fixation .................................................................................................................... 135
  4.4.4 Encoding probe staining .................................................................................................................... 135
  4.4.5 Sample embedding and clearing ........................................................................................................ 136
  4.4.6 Imaging platforms ............................................................................................................................. 137
  4.4.7 MERFISH imaging ............................................................................................................................. 139
  4.4.8 Image registration and decoding ........................................................................................................ 141
  4.4.9 RNase treatment ................................................................................................................................. 143
  4.4.10 Protein and lipid staining .................................................................................................................. 143
  4.4.11 Tissue preparation ............................................................................................................................ 144
Chapter 5: Conclusion ............................................................................................................................... 146
  5.1 Characterization of synaptic neurexin-1 nanoclusters using super-resolution microscopy ................. 146
  5.2 Increasing throughput and performance of MERFISH towards achieving spatially resolved transcriptomics in brain tissue ...................................................................................................... 147

Appendix A ................................................................................................................................................. 150
Appendix B ................................................................................................................................................... 151
Bibliography .................................................................................................................................................. 152
Citations to previously published work

Chapter 2 has been published with minor changes as:

Chapter 3 has been published with minor changes as:

Chapter 4 has been published with minor changes as:

Other publications:

J. R. Moffitt*, D. Bambah-Mukku*, S. W. Eichorn†, E. Vaughn†, K. Shekhar, J. D. Perez, N. D. Rubinstein, J. Hao, A. Regev, C. Dulac, X. Zhuang, Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. Science. 362 (2018) pp. eaau5324 *These authors contributed equally to this work †These authors contributed equally to this work
Acknowledgements

First and foremost, I need to express my gratitude towards my advisor, Prof. Xiaowei Zhuang, for providing me with the opportunity to work on a number of interesting projects during my time in her lab. She is not only an excellent and passionate scientist, but also wonderful mentor who pushes us to achieve the best we can, and an inspiration to everyone in her lab. I am also thankful for my committee members Prof. Joshua Sanes and Prof. Adam Cohen, as well as Prof. Sunney Xie and Prof. Hongkun Park whose labs I rotated in.

It has been an honor and privilege to work with and learn from many talented scientists. First I would like to thank my collaborators Prof. Thomas Südhof, and Dr. Justin Trotter, as well as Dr. Stephen Maxeiner and Dr. Theodoros Tsetsenis with whom I worked closely on the STORM project presented in the thesis. Of course, I am indebted to Dr. Colenso Speer and Dr. Yaron Sigal, for mentoring me during my rotation, and providing guidance in everything from STORM imaging to working with mice and tissue. Next I need to thank Dr. Jeffrey Moffitt and Dr. Kok Hao Chen for all their advice and friendship, as well as patience in showing me the ropes of MERFISH. I would also like to thank Dr. Hazen Babcock for always being available to answer my questions regarding software and hardware.

There are many others in the Zhuang lab that I am thankful for. I have had the privilege to work with and have lively discussions with many people as part of Team MERFISH, and in particular Guiping Wang and Tian Lu whom I worked with on the throughput and clearing projects, as well as Chenglong Xia and Dr. Timothy Blosser on other projects not mentioned in this thesis. I must thank our wonderful lab manager, Alec Goodman, for always making sure that everything in the lab runs smoothly. I would like to thank my office mates, Dr. Ben Altheimer, Dr. Evan Heller, and Dr. Ruobo Zhou are always available for discussion, both scientific related
(and not). I would also like to thank the rest of the Zhuang lab members I have had the pleasure of interacting with for all of their help and support.

There are many others during my time at Harvard that have been there with me through the highs and lows of graduate student life, and who I feel very lucky to be friends with. I need to thank Dr. Chun Pong Tam, Dr. Tian-Ming Fu, Dr. Leon Liu, Dr. Xiabing Lou, and Dr. Ruihua Ding for all the memories of our times at CCB. I would like to especially thank Dr Maxim Wolf and Dr. MengMeng Xu, who are probably the closest I’ve come to having siblings. Talking with you never fails to brighten my mood, no matter how down I may be feeling. I also want to thank all the people I have had the pleasure to sing with in both the graduate student choir and the Collegium. Music has been and probably will always be my emotional pillar.

I have been very lucky to have a wonderful family, and probably no amount of words will be able to fully express my appreciation for you. My grandparents have been and probably will always be my greatest source of influence and inspiration, in science and in life. I have thanked you countless times, but I wish you could hear me say thank you countless more. I also want to thank my parents and in laws for their encouragement and support, my mother in particular for all the sacrifices I know you have made for me, though I probably do not express my gratitude enough. Finally, I need to tell my wife how truly grateful I am for her unconditional love, encouragement, and support.

For my grandparents and my wife
Chapter 1

Introduction

The brain is arguably the most difficult organ to fully understand due to the high degree of complexity resulting from the sheer number of neurons and synapses that connect them, manifesting in the emergent properties of behavior and thought. Estimates put the number of neurons in the adult human brain on the order of $10^{11}$, with $10^{15}$ connections between them (1), with roughly similar numbers of glial cells (2). One of the more fascinating questions that continues to drive the field is how exactly synaptic identity in the connections between neurons are formed, pruned, and maintained.

Initiation of synaptic formation seems deceptively simple, as contact from a single type of adhesion molecule, such as SynCAM (3) or Neuroligin-1 (4) expressed in nonneuronal cells was sufficient to drive synaptic assembly, yet in vivo, synapses exhibit a high level of specificity and are dynamically regulated based on activity (5, 6). Adhesion molecules, which by definition are likely found at points of neuronal contact, have been demonstrated to play many roles in directing synaptic formation and maintenance (7–10). One particular heterophilic trans-synaptic adhesion molecule, neurexin, has been a particularly attractive candidate for study since its discovery in 1992 (11), owing to the high number of splice isoforms for both neurexin itself and its post-synaptic binding partners (12).

1.1 Neurexins are functionally important molecules responsible for synaptic function and maintenance
The complexity of mammalian neurexin isoforms arises from 3 genes, each driven by independent promoters to produce longer α and shorter β forms with thousands of possible splice isoforms, as well as a third recently discovered γ form unique to neurexin-1 (11, 13–15). The three neurexin genes are transcribed at similar levels across the brain, but with much higher levels of α than β isoforms (16–18). Interestingly, the relative levels of different neurexin isoforms appear to vary with neuronal cell type, further supporting the possible role neurexins may play in determining synaptic identity (19, 20). Neurexins interact trans-synaptically with many ligands, including neuroligins, LRRTMs, cerebellins, and latrophillins, and their binding affinities has been demonstrated to depend on alternative splicing of neurexin (12). The wide array of neurexin and ligand isoforms and their interactions have been found to play a role in a plethora of synaptic functions.

While early functional studies reported that neurexin and its interaction with neuroligin were capable of inducing assembly of pre- and postsynaptic proteins (4, 21), their necessity for initiating synapse formation was called into question, when triple deletion of α-neurexins was found not inhibit formation of excitatory synapses in the mouse brain (22). More recently, it was reported that conditional deletion of all neurexin isoforms affected synapse formation in some, but not all regions of the mouse brain (23). However, this total knockout did result in defects in synaptic transmission and decreases in synapse number in a cell-type and brain region dependent manner, suggesting that neurexins may perform different functions in different types of neurons and synapses. Deletion of all α-neurexins have been demonstrated to affect synaptic vesicle release in a Ca\(^{2+}\)-channel dependent manner (22), while deletion of \(Nrxn3\) in the hippocampal CA1 region and olfactory bulb resulted in a contrasting decrease in either AMPA- or GABA-receptors, in the respective brain regions. Summarized in detail in a review by Südhof, the
postsynaptic ligands of neurexin also contain multiple genes and exhibit alternative splicing, and have been implicated in regulating a wide variety of synaptic properties including synaptic transmission, long-term potentiation (LTP), and long-term depression (LTD), as well as synaptic density and structure (19).

1.2 Disruption to neurexin function is implicated in neurological disorders

Given the pervasiveness of neurexins in synaptic function, it is natural to find disruptions to neurexin implicated in a range of phenotypes of neurological disorder. In rodent models, a variety of genetic techniques have allowed probing of specific deletion/mutation interactions of neurexin molecules with deficits in behavior. Homozygous deletion of neurexin-1α (24) or neurexin-2α (25), and to a lesser extent heterozygous deletion of either (26), results in autism-like behavior, generally typified by repetitive motion and decreased social cues. Triple KO of all alpha neurexin isoforms was found to be fatal in neonatal mice, and impaired neurotransmitter release at both excitatory and inhibitory synapses (22). Conversely, triple deletion of the shorter β-neurexin isoforms resulted in viable, but smaller and infertile mice, as well as impaired contextual fear memory (18).

Mutations in neurexin-1 are also implicated in clinical studies of neurological illnesses. Genetic manipulation tools are unavailable in clinical studies, and results are drawn from correlating phenotypes with genetic mutations found through genotyping at varying length scales using different techniques, including classical techniques like karyotyping, to more recently developed genome-scale techniques. Kasem et al. provided a broad summary of genetic studies of neurexin mutations and disorders to which they have been linked (27), including autism spectrum disorders, Pitt-Hopkins syndrome, schizophrenia, intellectual disability, developmental
disorders, and others. Given the significant health detriments of these neurological disorders, it is imperative that we fully understand the mechanisms of neurexin function, so that we may find the underlying causes of these phenotypes and properly target treatment.

1.3 Applications of super-resolution imaging in neurobiology

A multitude of complementary methods across a range of scientific disciplines have been applied in an effort to unravel its functions, of which imaging has offered an invaluable chance to study the structure of neurons and their connections in situ. Fluorescence-based imaging, in particular, has been used extensively because of its high specificity, ability to simultaneously image multiple molecules simultaneously in multiple color channels, and compatibility with live imaging. More recently, the introduction of super-resolution techniques (28–32) has significantly extended the scale of samples that can be imaged with fluorescence microscopy, allowing features below the diffraction limit of light to be resolved. The diffraction limit is a phenomenon that arises due to diffraction of light, causing images of point sources of light to be blurred to roughly less than half the wavelength their emitting wavelength, which is roughly a few hundred nanometers for visible light. Historically, shorter wavelength approaches, such as electron microscopy, or near-field techniques, like atomic force microscopy, have been applied for resolving minute features, yet the conditions required are often harsh and unsuitable for more delicate biological samples. Super-resolution techniques significantly extended the resolution of fluorescence-based microscopy imaging, and have found wide ranging applications across biology owing to their compatibility with wet and live samples (33–36).

Even within the field of neuroscience, super-resolution techniques have been used to characterize a range of molecular and cellular properties of neurons, including the organization
of synaptic scaffolding and release proteins (37, 38), the structure and dynamics of trans-synaptically organized active zone and receptor machinery nanocolumns (39–41), a novel membrane-associated periodic skeleton (MPS) structure in axons and dendrites (42), and distribution of synapses across the surface of entire cells in the mouse retina (43). We also chose to leverage the advantages of super-resolution imaging, and performed multicolor 3D stochastic optical reconstruction microscopy (STORM) to study the synaptic nanostructure and localization of neurexin-1.

1.4 Multicolor 3D super-resolution with STORM

Like other optical super-resolution techniques, STORM is able to resolve biological features smaller than the diffraction limit of light. STORM (29), and similar approaches photoactivated localization microscopy (PALM) (30) and fluorescence photoactivated localization microscopy (FPALM) (31) overcome the diffraction barrier by acquiring images of stochastically activated fluorophores labeled across a structure of interest over time, and fitting their centroid positions to achieve high-precision localizations. From multiple frames acquired over the same field of view, these high-precision localizations can be used to reconstruct and resolve the underlying biological structure at resolutions an order of magnitude or more below those achievable via conventional fluorescence microscopy.

After its initial demonstration, the capabilities of STORM were quickly augmented to include multicolor and 3D imaging. To achieve multicolor imaging, STORM experiments could either be performed using multiple fluorophores in different emission channels, or by activating the same emission fluorophore with different activation fluorophores in different frames of a STORM movie, and capturing the localizations in subsequent frames following each activation
To add 3D information to STORM localizations, a cylindrical lens can be added prior to the camera, which results in astigmatism, and elongation of circular point spread functions (PSFs) into differentially elongated ellipses in x and y directions as a function of positive or negative z-offset (45). Other advances to localization-based super-resolution techniques have since been reported, including live-cell imaging (46), ultrafast imaging with sub-second time resolution (47), and sub-10 nm resolution (48, 49), as well as combinations with complementary imaging techniques, such as adaptive optics (50) or light-sheet microscopy (51), for thick-sample and deep-tissue imaging.

1.5 Transcriptome-level studies of brain tissue

More recently, many -omics level approaches have also been applied to characterize the diverse range of cell types in the brain and dissect their interactions (52–55), with the hope that systematic classification and definition may help guide and focus efforts to understand how the brain works (56). Single-cell RNA sequencing (scRNAseq) enables quantification of the expression level of all genes in individual cells, providing a way to group neurons and glia into transcriptionally distinct cell types (57, 58). Indeed, scRNAseq has been used to classify cell types in a variety of brain regions to quantify numbers of transcriptionally distinct cell types and measure their degree of heterogeneity in brain tissue, as summarized by Cembrowski (55). However, only having transcriptional profiles of dissociated cells is insufficient to full map and dissect the neuronal networks that comprise brain tissue. Ever since Rámon y Cajal revealed the beauty and intricacy of neuronal arborizations traced using Golgi staining, there has been a strong focus on understanding the underlying mechanisms that give rise to the unique structures and shapes of different neurons and the computational circuits they form. Spatial information is
thus a key piece of information required to fully describe the transcriptional landscape inherent to the complex 3D structure of the brain.

A number of methods have been utilized to complement scRNAseq by providing spatial context to the transcriptomics data. Fluorescence in situ hybridization (FISH) has classically been a powerful technique to stain individual RNA species in tissue, and with the advent of single molecule fluorescence in situ hybridization (smFISH), it has become possible to visualize individual RNA molecules in their native context (59, 60). However, it would not be feasible to measure each RNA species individually to construct a spatially-resolved map at the transcriptome-scale over the course of tens of thousands of individual smFISH experiments. To address this issue of scaling, a number of technologies have been developed that combine in vivo spatial context to transcriptome-scale gene expression measurements, including in situ sequencing techniques (61, 62), as well as multiplexed smFISH techniques (63, 64). Other techniques for spatially-resolved transcriptome profiling have also been reported, but have not yet to be demonstrated with cellular resolution (65, 66). Our lab developed a multiplexed smFISH method called multiplexed error-robust FISH (MERFISH) that allows transcriptome-scale RNA imaging in single cells (63). Early demonstrations of MERFISH were in simpler cultured cell systems, and needed further adjustments before they could be applied in intact tissue. We aimed to improve MERFISH to enable spatially-resolved transcriptome measurements in brain tissue.

1.6 MERFISH for spatially-resolved transcriptomics

First reported in 2015, MERFISH was designed making for transcriptomic measurements in single cells by massively multiplexing smFISH measurements (63). As the acronym suggests,
MERFISH was designed with two key concepts. Multiplexity in MERFISH arises through the use of combinatorial barcoding, in which each RNA species is assigned a unique binary barcode of 0’s and 1’s, which in principle would allow encoding of $2^N$ RNA species with a barcode of length $N$. Although most of the experimental implementations of MERFISH use binary barcodes(63), ternary and higher-base barcodes are also possible (67, 68). When the RNA molecules are imaged sequentially in $N$ rounds, fluorescent smFISH spots are found in the rounds for which there is a corresponding “1” in $N^{th}$ digit in the binary barcode for each RNA. Practically, however, only a subset of all possible binary barcodes is used, fulfilling the namesake error robustness. First, the number of “1” bits is artificially kept low, due to the relative ease of 1 to 0 error compared to 0 to 1 error. Second, the minimum number of bit flips separating two different barcodes, termed Hamming distance (69), is set to 4 to enable error-correction, or reduced to 2 if only error-detection is desired.

Experimentally, in RNA molecules are labeled with two sets of oligo DNA probes. The first set, termed encoding probes, includes a targeting region that hybridizes to a particular RNA species, as well as readout regions that are designed to hybridize to the second set of oligos, termed readout probes. Encoding probes are all stained onto the sample simultaneously, before readout probes are then hybridized and imaged sequentially to read out the 1’s and 0’s corresponding to each RNA molecule. In the original demonstration, signal was removed between each round of readout hybridization via photobleaching, and as a result the measurement throughput, in terms of the number of cells imaged per experiment, is relatively low, ~ a few hundred per experiments (63). Nevertheless, MERFISH exhibited high quality performance, showing both high levels of correlation with expression levels from bulk RNA sequencing of the same IMR-90 cells, as well as significantly higher detection efficiencies in
comparison to scRNAseq techniques when compared to smFISH measurements genes spanning a representative expression range (63). We hoped to improve upon MERFISH, and increase its applicability towards intact brain tissue by significantly increasing throughput and decreasing background.

1.7 Outline of thesis

Biological imaging provides a unique opportunity to observe the native structure and distribution of biomolecules in intact cells and tissue. To fully dissect the functions of a tissue as complex as the brain, we must utilize all tools available, while continuously striving to improve upon existing techniques towards the goals of seeing things smaller, faster, and clearer.

Chapter 2 of this thesis describes our study on the synaptic distribution and dynamics of neurexin-1, utilizing the power of STORM to provide sub-diffraction-limit images of neurexin at synapses in both cultured neurons and brain tissue. We find that neurexin-1 is localized to nanoclusters that increase in size and neurexin content over the course of development. We further identify the metalloprotease ADAM10 as a regulator of neurexin-1 content at synapses through ectodomain cleavage.

Chapter 3 details our efforts to increase the throughput of MERFISH by two orders of magnitude. By adopting reductive-cleavage to efficiently remove dyes from labeled RNA, reducing readout probe sequences to a 3-letter alphabet, and improving imaging hardware and software, we were able to achieve imaging on the scale of multiple millimeters-squared in a single day, and on the order of tens of thousands of cells in a single overnight experiment. This allows MERFISH to overcome the significant hurdle of throughput hindering RNA profiling of whole tissues.
Chapter 4 further expands on our efforts to improve upon MERFISH by removing protein and lipids through clearing. After anchoring RNA to a polyacrylamide gel cast over the sample, a combination of proteinase and surfactant is used to digest proteins and lipids we found responsible for a significant fraction of observed background staining in MERFISH measurements. After sample clearing, we demonstrate the ability to perform MERFISH measurements in 4 color channels simultaneously, as well as in cryosections from the medial preoptic area of the mouse hypothalamus.
Chapter 2

Synaptic Neurexin-1 Assembles into Dynamically Regulated Active Zone Nanoclusters

2.1 Introduction

Synapses are specialized intercellular junctions that are dedicated to the transfer of information between neurons. Synaptic properties, such as release probability, postsynaptic receptor composition and short-term synaptic plasticity, vary widely and are subject to activity-dependent short- and long-term changes (70, 71). These properties are determined by interactions between pre- and postsynaptic neurons that may be largely shaped by trans-synaptic cell-adhesion molecules (72–76), including neurexins, which are arguably the best characterized (12).

Neurexins are encoded by three genes, each of which contains independent promoters that drive transcription of longer α-neurexin and shorter β-neurexin proteins (11, 13, 14) (Figure 2.1A). The large extracellular sequence of α-neurexins includes six LNS-domains with interspersed EGF-like repeats, while the extracellular sequence of -neurexins contains only a short β-specific N-terminal sequence that splices into the -neurexin sequence N-terminal to their sixth LNS-domain (Figure 2.1A). Following the sixth LNS-domain, α- and β-neurexins include a heavily glycosylated ‘stalk’ region that is interrupted by a cysteine-loop domain, a transmembrane region, and a cytoplasmic tail. The neurexin-1 gene (Nrxn1) also encodes a third shorter isoform called neurexin-1γ (Nrxn1γ) that is absent from the neurexin-2 (Nrxn2) and neurexin-3 (Nrxn3) genes and lacks LNS-domains (77, 78) (Figure 2.1A).
Due to alternative splicing neurexins are differentially expressed as thousands of isoforms throughout the brain (15, 17, 79). Neurexins are presynaptic proteins that interact with a myriad of postsynaptic ligands, often in a manner regulated by alternative splicing, to perform multiple functions at synapses (12). These functions likely depend on the specific isoforms expressed and on the ligands available, and include regulation of presynaptic Ca\textsuperscript{2+}-channels (22, 23), trans-synaptic recruitment of postsynaptic AMPA- (16) or NMDA-type glutamate receptors (80), and control of tonic postsynaptic endocannabinoid synthesis (18).

Figure 2.1: Schematic of the Nrxn1 cKI mice and impaired survival following constitutive truncation of Nrxn1. A. Diagram of wild-type, HA-tagged full-length, and HA-tagged truncated Nrxn1α, Nrxn1β, and Nrxn1γ. Top, wild-type Nrxn1 proteins; middle, Nrxn1 proteins containing inserted tandem-HA epitopes and loxP sequences in cKI mice; bottom, truncated HA-tagged Nrxn1 proteins after Cre-recombination (LNS1-6 = LNS1-6 domains; E = EGF-like domains; C = cysteine-loop domain). B. Amino-acid sequences of the juxtamembranous region of wild-type Nrxn1 (top), Nrxn1 with the HA-epitope/LoxP-site knockin (middle), and Cre-recombined truncated Nrxn1 (bottom; SS5 = alternative splice site #5). C. Constitutive Nrxn1 truncation impairs postnatal survival as analyzed in newborn (P1) and 21-day old mouse (P21) offspring from heterozygous matings. Statistical significance was assessed by the Chi Square test (**: p < 0.01).
Recent studies have uncovered that synapses contain nanocolumns (81–85). Specifically, presynaptic release sites and postsynaptic receptors were shown to cluster into nanocolumns that are aligned across the synaptic cleft (39, 40, 86–88). A different organization was observed for SynCams and N-cadherin, which are trans-synaptic cell-adhesion molecules that were localized to a ring surrounding the active zone (89, 90). How other synaptic cell-adhesion molecules, such as neurexins, are organized, however, remains unknown. Trans-synaptic signaling by synaptic cell-adhesion molecules is likely dynamic, as suggested by the observation that these molecules are often substrates of ectodomain proteases such as ADAM10 and BACE1 (91–93). Cleavage by ectodomain proteases could provide a mechanism for the rapid and regulated disassembly of trans-synaptic protein complexes and for fine-tuning of synaptic properties. Indeed, this mechanism has been suggested for the neurexin ligand neuroligin-1, which is proteolytically processed by ADAM10 (94) and/or MMP-9 (95).

In the present study, we analyzed Nrxn1 localization by 3D stochastic optical reconstruction microscopy (STORM) (29, 45). We found that at Homer1(+) excitatory synapses, Nrxn1 is localized to discrete nanoclusters that occupy a fraction of the synapse area and increase in neurexin content and physical size during synapse development. Moreover, we found that Nrxn1 is physiologically cleaved by ADAM10, thereby releasing a soluble fragment containing most of the extracellular sequences of Nrxn1. Inhibition of Nrxn1 cleavage by pharmacological blockage of ADAM10 dramatically enhances the Nrxn1 content of synaptic nanoclusters, and elevates the proportion of excitatory synapses containing Nrxn1 nanoclusters to nearly 100%. Taken together, our results reveal an unexpected nanoscale organization of synapses in which Nrxn1 is assembled into nanoclusters that are dynamically regulated by ADAM10 and may serve as a functional pivot for synapse specification.
2.2 Results

2.2.1 Generation of conditional knockin (cKI) mice expressing HA-tagged Nrxn1

To enable precise localization and monitoring of synaptic Nrxn1, we generated conditional knockin (cKI) mice that express HA-tagged Nrxn1 (Figure 2.1A, 2.1B, 2.2A). Specifically, we inserted a double HA-epitope tag and LoxP sequence into the extracellular stalk region of Nrxn1 between the Cys-loop domain and the transmembrane region (Figure 2.1A, 2.1B). In the Nrxn1 gene, this region is encoded by the last coding exon that is 3’ to alternatively spliced sequence #5 (SS5; Figure 2.1A, 2.2B, 2.2A). In addition, we introduced a second loxP sequence into the 3’ non-coding region of the targeted exon. With this design, the Nrxn1 cKI gene encodes full-length, HA-tagged Nrxn1 proteins in the absence of Cre-recombinase. Following Cre recombination, however, the Nrxn1 cKI gene synthesizes truncated Nrxn1 proteins that are still HA-epitope tagged but lack a transmembrane region (Figure 2.1A). We chose this cKI design for studying Nrxn1 because we could find no neurexin antibodies that are specific for a given neurexin isoform, target extracellular sequences of a neurexin, and recognize native neurexins in cells or tissue. Moreover, this design replicates a strategy we employed in human neurons to model a Nrxn1 mutation associated with schizophrenia (96, 97).

We crossed the Nrxn1 cKI mice with transgenic mice expressing Cre-recombinase in the germ line, and analyzed the survival of the resulting constitutive truncated Nrxn1 (‘Nrxn1-Tr’) mice. Homozygous Nrxn1-Tr mice were viable at birth, but only 1 in 40 mice survived until weaning (Figure 2.1C). mRNA measurements in brains from littermate wild-type and heterozygous Nrxn1-Tr mice suggest that the truncation mutation does not cause non-sense-mediated decay of the Nrxn1 mRNAs (Figure 2.2B-C). Thus, while truncation disables Nrxn1
functions that are essential for survival, the epitope-tag itself does not seem to have a deleterious effect.

Figure 2.2: Generation and characterization of HA-neurexin-1 (HA-Nrxn1) conditional knockin (cKI) mice and of constitutive HA-Nrxn1-truncation mice. A. Strategy for generation of HA-Nrxn1 cKI mice by homologous recombination in ES cells. A targeting vector containing exon-26 (which encodes the C-terminal Nrxn1 transmembrane region and cytoplasmic tail) was constructed. The targeting vector includes (i) homologous arms, (ii) sequences encoding tandem HA-epitope tags, a linker sequence, and a LoxP site with an open reading frame inserted into the 5’ end of exon-26, (iii) a second LoxP site into the 3’ UTR, (iv) a PGK-neo selection cassette flanked by frt sites, and (v) a diphtheria toxin expression cassette for negative selection. We performed homologous recombination experiments in ES cells to generate ES cell clones containing a targeted Nrxn1 allele. Mice were then produced from these ES cells by blastocyst injections, and offspring with the correctly targeted allele were crossed with mice expressing flp-recombinase to remove the PGK-neo selection cassette from the gene, resulting in the HA-Nrxn1 cKI allele in which full-length Nrxn1 is expressed with an inserted tandem HA-epitope. Crossing HA-Nrxn1 cKI mice with mice expressing Cre-recombinase or expressing Cre-recombinase in cells cultured from the HA-Nrxn1 cKI mice, finally, causes recombination of the two loxP sites in the HA-Nrxn1 cKI allele and thereby production of a truncated HA-Nrxn1 that lacks a transmembrane region and cytoplasmic sequence (see Figure 2.1A-B for the structures and sequences of the HA-Nrxn1 proteins). B. Measurements by quantitative RT-PCR of the mRNA levels of the indicated genes in the forebrains of adult...
littermate wild-type mice and mice heterozygous for the HA-Nrxn1 truncation mutation. The first bars depict the levels of Nrxn1 mRNAs lacking an insert sequence at the position in which the HA-epitope is inserted, exon-26; the Nrxn1α and Nrxn1β bars depict the levels of these mRNAs as determined with probes upstream of the HA-epitope insertion site. All other measurements assess mRNA levels for the indicated genes. Note that the normal levels of Nrxn1α and Nrxn1β mRNAs indicate that there is no nonsense-mediated decay induced by the deletion of exon-26. C. Same as B, except measurements reflect mRNAs containing the knockin and normalization is to mice carry a single HA-Nrxn1-Tr allele. D. Measurements by quantitative RT-PCR of the mRNA levels of the indicated genes in hippocampal neurons cultured from HA-Nrxn1 and infected with lentiviruses expressing inactive mutant Cre-recombinase (ΔCre) or active Cre recombinase. Neurons were infected at DIV4, and harvested at DIV12. Data show the levels observed in neurons after Cre recombination as percent of the levels in the matching control neurons expressing ΔCre. The first bars depict measurements of Nrxn1 mRNAs containing exon-26 sequences; the Nrxn1α and Nrxn1β bars depict the levels of these mRNAs as determined with probes upstream of the HA-epitope insertion site. Again, the normal levels of Nrxn1α and Nrxn1β mRNAs indicate that there is no nonsense-mediated decay induced by the deletion of exon-26. E. Representative images of DIV14 hippocampal neurons live surface-labeled for HA-Nrxn1 (red) and then co-labeled for MAP2 (green) and the various indicated presynaptic proteins (blue). Surface HA-Nrxn1 signals coincide with some (white asterisk) but not all signals of presynaptic proteins (yellow asterisk). Scale bars = 2 μm. Summary graphs show means ± SEMs. Statistical significance was determined by T-test to earliest time point or control (***, p < 0.001; ****; p < 0.0001). For B-C, n = 4 mice; for D, n = 4 cultures.

Hippocampal neurons cultured from HA-Nrxn1 cKI mice were infected with lentiviruses expressing mutant inactive (ΔCre; control) or active Cre-recombinase (Cre). After Cre expression, >95% of the Nrxn1 cKI gene was recombined (Figure 2.2D). Neurons were live surface labeled with an HA antibody (98), fixed and permeabilized, followed by labeling with a different HA antibody to visualize intracellular Nrxn1 and with an antibody to MAP2 to visualize dendrites (Figure 2.3A). We detected a punctate surface-staining pattern of HA-Nrxn1 in ΔCre-expressing neurons juxtaposed to dendrites, but found no HA-Nrxn1 puncta in Cre-expressing neurons. Intracellular HA-Nrxn1, conversely, was concentrated in the perinuclear region of neuronal somas and to a lesser extent in proximal primary dendrites (Figure 2.3A). Co-labeling of surface HA-Nrxn1 with synaptic markers revealed that some, but not all, synapses were associated HA-Nrxn1 (Figure 2.3A, 2.2E).
**Figure 2.3:** *Nrxn1* cKI mice express full-length HA-tagged Nrxn1 in the absence of Cre-recombination, but truncated HA-tagged Nrxn1 after Cre-recombination. A. HA-tagged Nrxn1 exhibits a punctate surface-staining pattern in cultured hippocampal neurons. Representative images of HA-Nrxn1 cKI neurons expressing ΔCre (control, left) or Cre (to conditionally truncate HA-Nrxn1, right), and stained at DIV14 for surface-exposed and internal HA-Nrxn1 as well as for MAP2 (top, overview; bottom, zoomed-in to showcase synapses). B. Immunoblot analysis of proteins from hippocampal neurons treated as described in A reveals HA-Nrxn1 truncation but normal expression of selected synaptic proteins (top, representative blots; bottom, summary graph of protein levels). Data are means ± SEM; statistical significance was determined by one sample T-test (**: p < 0.01, ***: p < 0.001; n = 3 cultures).

We next analyzed cellular and secreted proteins in hippocampal neurons following Cre recombination via quantitative immunoblotting (Figure 2.3B). Consistent with the alternative splicing of Nrxn1 mRNAs and glycosylation of Nrxn1 protein (14, 15, 99), we observed in ΔCre-expressing neurons multiple cellular HA-Nrxn1 species (Figure 2.3B). In contrast, we detected in Cre-recombinase expressing neurons a smaller Nrxn1 protein that was less abundant. No other cellular protein analyzed was changed. When we examined secreted proteins, we detected HA-tagged Nrxn1 in the medium of ΔCre-expressing neurons that was smaller than full-length cellular Nrxn1 (Figure 2.3B). In Cre-expressing neurons, truncated HA-tagged Nrxn1 was
slightly smaller than its wild-type counterpart in ΔCre-expressing neurons and was found at lower levels, suggesting that Nrxn1 may be cleaved at a significant rate to produce secreted extracellular fragments.

Overall, these findings demonstrate that HA-tagged Nrxn1 expressed in Nrxn1 cKI mice is functional, transported to the surface, and concentrated in puncta that are at least partly synaptic, but is also partly converted into a secreted protein. Conditional truncation of Nrxn1, conversely, produces only a secreted protein that is not functional and is present at lower levels, possibly because it is unstable (96).

2.2.2 Nrxn1 localizes to unique nanoclusters at synaptic junctions

To examine where in a synapse Nrxn1 is localized, we co-labeled HA-Nrxn1 and Homer1, an excitatory postsynaptic scaffolding protein (100), using direct and indirect immunofluorescence, respectively, on cultured hippocampal neurons and cryostat sections within the S. radiatum of the hippocampal CA1 region. We then imaged synapses using 3D STORM, a super-resolution imaging method (29, 45).

Instead of being evenly distributed across the synaptic junction, Nrxn1 was concentrated in the synaptic cleft in discrete nanoclusters (Figure 2.4A-B, 2.4G-H, 2.5A-B). Only a subset of Homer1(+) synapses contained Nrxn1 nanoclusters, and those with Nrxn1 nanoclusters mostly featured a single nanocluster (Figure 2.4C, 2.4I). Nrxn1 nanoclusters were similarly observed when we double-labeled synapses for Nrxn1 and the presynaptic markers Piccolo (Figure 2.5A) or Bassoon (Figure 2.5B). Supporting the specificity of the HA antibodies, we did not observe HA(+) signals in Nrxn1 cKI neurons expressing Cre (Figure 2.5C) or in neurons or brain sections from wild-type mice (Figure 2.5D-E).
Figure 2.4: Super-resolution STORM imaging identifies synaptic Nrxn1 nanoclusters. A. Representative image of cultured neurons at DIV18, showing multiple Homer1(+) synapses with a subset containing Nrxn1 nanoclusters. B. Representative images of individual synapses from DIV18 neurons showing Homer1(+) synaptic junctions containing Nrxn1 nanoclusters. C. Only a subset of Homer1(+) synapses contains Nrxn1 nanoclusters (39%), abbreviated as NC, in cultured neurons at DIV18. D. Homer1 ‘discs’ and Nrxn1 nanoclusters are separated by ~85 nm across the synapse. The trans-axial distribution of Homer1 and HA-Nrxn1 localizations was determined by STORM imaging of hippocampal cultures at DIV18. E. Nrxn1 nanoclusters
occupy on average ~19% area of the Homer1-defined synaptic junction. F. Nrxn1 nanoclusters are on average labeled by ~9 antibodies per nanocluster. G-K. Same as A-E, except that cryostat sections of the hippocampal CA1 region from HA-Nrxn1 cKI mice were analyzed at postnatal day 28 (P28). L. Quantification of the axial distribution of STORM localizations of surface HA-Nrxn1 in hippocampal neurons or cryosections relative to those of Homer1, Piccolo, and Bassoon. Short vertical bars = SEMs; horizontal bars = standard deviations. For C-F, n = 54 synapses / 3 cultures; for I-K, n = 77 synapses / 3 mice.

Because the synaptic images were acquired in 3D, we could rotate them to facilitate spatial analyses (Figure 2.4B, 2.4H). Quantifications showed that the distance between the Nrxn1 nanocluster and Homer1 disc along the trans-synaptic axis was similar in cultured hippocampal neurons and hippocampal sections (Figure 2.4D, 2.4J). We repeated these measurements using multiple synaptic marker proteins, and determined the position of Nrxn1 clusters relative to the synaptic cleft along the trans-synaptic axis based on the previously reported localizations of the same synaptic marker proteins (38). Using this approach, we localized the Nrxn1 HA-epitope to the presynaptic border of the synaptic cleft (Figure 2.4L), consistent with the juxtamembranous placement of the HA-epitope (Figure 2.1A).

Further quantifications revealed that each Nrxn1 nanocluster occupied ~15-20% of the area of the synaptic cleft (Figure 2.4E, 2.4K), with 8.6 ± 0.6 primary antibodies bound to each nanocluster in cultured hippocampal neurons (Figure 2.4F). Since maximally two antibodies can bind to an HA-tagged Nrxn1 molecule, each nanocluster contains at least ~4.3 Nrxn1 molecules, although the actual number is likely higher because not all epitopes will be saturated with antibodies.
Figure 2.5: Super-resolution microscopy by 3D STORM identifies mostly single Nrxn1 nanoclusters in excitatory synapses that are developmentally regulated. A. Representative 3D STORM images of synapses in hippocampal neurons that were cultured from HA-Nrxn1 cKI mice and double-labeled for surface HA-Nrxn1 and Piccolo (scale bar = 0.5 µm). B. Representative 3D STORM images of synapses in hippocampal sections from Nrxn1 cKI mice that were double-labeled for surface HA-Nrxn1 and Bassoon (scale bar = 0.5 µm). C-D. Control experiments documenting specificity of HA-immunolabeling in cultured neurons. Representative 3D STORM images show synapses in hippocampal neurons that were cultured from HA-Nrxn1 cKI mice and infected with Cre-recombinase expressing lentiviruses (C) or that were cultured from wild-type mice (D). Neurons were double-labeled for surface HA-Nrxn1 and Homer1 (scale bar = 0.5 µm). E. Control experiments documenting specificity of HA-immunolabeling in hippocampal sections from wild-type mice. Representative 3D STORM images show synapses in hippocampal sections from wild-type mice. F-G. Homer1 levels (F) and area (G) did not significantly change during synaptogenesis in cultured hippocampal neurons analyzed at the indicated days in vitro (DIVs; shown as % of DIV4 levels). H. Nrxn1 localizations that were not associated with Homer1(+) synapses significantly increased during synaptogenesis in cultured neurons. I. Nrxn1 nanoclusters shift towards the periphery during synapse maturation in cultured hippocampal neurons. J. Control for differences in synapse area, for each synapse, the average localization of HA-Nrxn1 and Homer1 signals were divided by a radius value defining the area that encompasses 95% of Homer1 localizations. This
normalized measurement of Nrxn1 radial displacement similarly shows that Nrxn1 nanoclusters move slightly towards the periphery during synapse maturation. K. The size of synaptic Nrxn1 nanoclusters was compared using three different methods at indicated culture days in vitro (DIVs; shown as % of DIV4 levels). L-M. Homer1 levels (L) and area (M) significantly increased during postnatal development in vivo. Analysis was performed as in F-G, except that cryostat sections of the hippocampal CA1 region from HA-Nrxn1 cKI mice were analyzed as a function of age. N-P. Same as H-J, but of Nrxn1 nanoclusters in cryosections of the hippocampal CA1 region from HA-Nrxn1 cKI mice of the indicated ages. Summary graphs show means ± SEMs. Statistical significance was determined by T-test to earliest time point or control, except for radial displacement in I-J and O-P, which was determined by a rank-sum test (*: p<0.05, **: p < 0.01, ***: p < 0.001). For For F-G and I-J, n as # of synapses = 6, DIV4; 12, DIV10; 78, DIV18; 85 DIV26 / 3 cultures; for H and N, n = 15 fields of view / 3 cultures or mice; for K, n as # of synapses = 6, DIV4; 12, DIV10; 78, DIV18; 85 DIV26 / 3 cultures; for L-M and O-P, n as # of synapses = 45, P1; 41, P7; 22, P14; 77, P28 / 3 mice.

2.2.3 Development of synaptic Nrxn1 nanoclusters

To study the development of Nrxn1 nanoclusters, we analyzed Homer1(+) synapses as a function of days in culture (DIV3-DIV26) or postnatal life (P1-P28). Generally, we found that Nrxn1 nanoclusters grow in size and content during synapse maturation and slightly shift towards the periphery of the synapse (Figure 2.6A, 1.6E). In immature neurons (<DIV5) and in mice younger than P10, less than 20% of Homer1(+) synapses contained Nrxn1 nanoclusters (Figure 2.6B, 2.6F). In more mature cultured neurons, the percentage of synapses with Nrxn1 nanoclusters more than doubled, leveling off at ~40% of synapses by DIV10 (Figure 2.6B). In the hippocampus tissues, however, the percentage of synapses with Nrxn1 nanoclusters continued to increase with age, reaching ~40% of synapses at P28 (Figure 2.6F).

Both the Nrxn1 content per nanocluster and the relative size of the nanoclusters increased substantially during neuronal maturation in cultured neurons (Figure 2.5K, Figure 2.6C-D) and in hippocampal tissue (Figure 2.6G-H). By comparison, Homer1 levels per synapse and size did not increase substantially during synapse maturation in cultured neurons (Figure 2.5F-G), and showed a major increase only at P28 in vivo (Figure 2.5L-M). Since Nrxn1 is also found at
inhibitory synapses and extrasynaptic sites (Figure 2.2E, 2.3), we quantified all Nrxn1 localizations not associated with Homer1 and found that maturation leads to increased Nrxn1 levels at sites other than Homer1(+) excitatory synapses (Figure 2.5H, 2.5N).

Figure 2.6: **Synaptic Nrxn1 nanoclusters are dynamic across development.** A. Example synapses illustrating the features of Nrxn1 nanoclusters in cultured neurons at DIV18 (top) and DIV26 (bottom), including central and peripheral localizations of Nrxn1 nanoclusters (scale bar = 0.2 µm). B. The percentage of Homer1(+) synapses with Nrxn1 nanoclusters in cultured neurons as a function of culture time. C-D. The relative Nrxn1 content (C) and size of synaptic Nrxn1 nanoclusters (D) increases during development in hippocampal neurons analyzed at the indicated culture days in vitro (DIVs; shown as % of DIV4 levels). E. Same as A, except of Nrxn1 nanoclusters in hippocampal CA1 region from HA-Nrxn1 cKI mice at P14 (top) and P28 (bottom) (scale bar = 0.2 µm). F. Same as B, except that cryostat sections of the hippocampal CA1 region from HA-Nrxn1 cKI mice were analyzed as a function of age. G-H. Same as C-D, but for cryostat sections of the hippocampal CA1 region of HA-Nrxn1 cKI mice analyzed at P1-P28 (values are expressed as % of P1). Numerical data are means ± SEM. Statistical significance was determined by two-sample T-test to earliest time point (*: p < 0.05, **: p < 0.01, ***: p < 0.001). For B/F, n = 3 cultures/mice, averaged per culture/mice; for C-D, n as # of synapses = 6, DIV4; 12, DIV10; 78, DIV18; 85 DIV26 / 3 cultures; for G-H, n as # of synapses = 45, P1; 41, P7; 22, P14; 77, P28 / 3 mice.
Finally, we measured the radial position of synaptic Nrxn1 nanoclusters relative to the synaptic cleft. We calculated both an absolute radial distance, as well as a normalized distance to account for variation in synapse size. We found that Nrxn1 nanoclusters were generally located closer to the center of the synapse during early stages of development in cultured neurons and tissue, but moved slightly peripherally at later stages (Figure 2.5I-J, 2.5O-P). Taken together, our data indicate that synaptic Nrxn1 nanoclusters substantially increase in abundance, size and content as a function of development and synapse maturation.

2.2.4 Synaptic Nrxn1 nanoclusters are formed independent of HA-epitope tagging and include other neurexin isoforms.

The observation of synaptic Nrxn1 nanoclusters was unexpected, as it suggests that synapses contain a hitherto unknown structural organization with a central neurexin pivot. However, our experimental approach may raise concerns about possible artifacts caused by the use of multivalent antibodies, the placement of the HA-tag and LoxP sequence, or the fixation method used. The discovery of Nrxn1 nanoclusters also raises an important question: are Nrxn1 nanoclusters specific for Nrxn1, or do they contain other neurexins?

To address these questions, we first tested whether the Nrxn1 nanoclusters are caused by antibody-induced aggregation during live-cell labeling. We compared Nrxn1 nanoclusters in cultured neurons live-labeled with HA antibody or stained after light and harsh chemical fixation, but observed no major differences (Figure 2.7A-J). Thus, the Nrxn1 nanoclusters are unlikely to be artifacts caused by live antibody labeling or fixation.
Figure 2.7: Additional control data to test the validity of Nrxn1 nanoclusters observed by 3D STORM microscopy. A. Synaptic Nrxn1 nanoclusters are equally detected in neurons immunolabeled when live or after fixation. Representative images show HA-Nrnx1 nanoclusters visualized in hippocampal neurons immunolabeled when live (left) or after fixation (right) at DIV18, and counterstained after permeabilization of fixed neurons for Homer1 (red). B-E. The neurexin content per nanocluster (B) and area occupied by a nanocluster (D) are only slightly reduced in neurons stained after fixation compared to live stained neurons, and the proportion of synapses with nanoclusters (C) and radial displacement (E) is indistinguishable. F. Synaptic Nrnx1 nanoclusters are independent of the fixation conditions. Representative images show Nrnx1 nanoclusters (HA-Nrnx1, green) visualized in hippocampal neurons fixed in 4% paraformaldehyde (PFA, left) or 3% paraformaldehyde-0.1% glutaraldehyde (GA, right) at DIV18, and counterstained after permeabilization for Homer1 (red) (scale bar = 0.5 µm). G-J.
The neurexin content per nanocluster (G), proportion of neurexin(+) synapses (H), area occupied by a nanocluster (I), and nanocluster radial displacement (J) are indistinguishable between neurons fixed in 4% PFA or 3% PFA and 0.1% GA. K-N. The absolute radial displacement (K), normalized radial displacement (L), levels of Homer1 (M), and Homer1 area (N) are indistinguishable between neurons from Nrxn1 cKO mice expressing mutant inactive Cre-recombinase (ΔCre) or active Cre-recombinase. O-P. The levels of Homer1 (O) and Homer1 area (P) was indistinguishable between neurons cultured from Nrxn123 triple cKO mice after expressing either mutant inactive Cre-recombinase (ΔCre) or active Cre. Q. Representative images of synapses stained for Homer1 (red) and for all neurexins using a pan-neurexin antibody (green; Frontier Institute, Af870) in hippocampal neurons cultured from HA-Nrxn1 cKI mice. Neurons were infected with lentiviruses encoding inactive mutant Cre-recombinase (ΔCre) or active Cre-recombinase at DIV4, and analyzed at DIV12. R-U. The levels of total neurexin signal per synapse (R), the proportion of neurexin(+) synapses (S), the size of the neurexin nanocluster (T), and nanocluster radial displacement (U) was indistinguishable between HA-Nrxn1 cKI neurons without or with Cre-recombination. V. Representative images of neurexin nanoclusters in synapses double-labeled in HA-Nrxn1 cKI neurons for surface HA-Nrxn1 and Homer1 (left) or in Nrxn1 cKO neurons for surface pan-neurexin and Homer1 (right). All neurons were infected with lentiviruses expressing inactive Cre-recombinase (ΔCre) to simulate viral infection state. Neurons were analyzed at DIV12. W-Z. Comparison of the levels of neurexin signal per synapse (W), the proportion of neurexin(+) synapses (X), the size of the neurexin nanocluster (Y) and nanocluster radial displacement (Z) between the two types of staining experiments described in panel Q. Summary graphs show means ± SEMs. Statistical significance of most plots was determined by T-test to control, except for radial displacement plots, which was determined by a rank-sum test (*: p<0.05). For B, D-E, n as # of synapses = 78, live; 55, fixed / 3 cultures; for C/H/S/X, n = 3 cultures, averaged per culture; for G, 1-J, n as # of synapses = 35, PFA; 23, GA / 3 cultures; for K-N, n as # of synapses = 17, ΔCre; 9, Cre / 3 culture; for O-P, n as # of synapses = 18, ΔCre; 14, Cre / 3 cultures; for R,T-U, n as # of synapses = 9, ΔCre; 11, Cre / 3 cultures; for W, Y-Z, n as # of synapses = 24, HA; 16, WT / 3 cultures.

Next, we examined whether the formation of the nanoclusters was a native property of neurexins that is independent of the HA-tagging and also includes other neurexins. We used a ‘pan-neurexin’ antibody that recognizes the conserved C-terminus of Nrxn1 and reacts with all neurexins and that was previously employed for immuno-EM analyses (101). Hippocampal neurons from Nrxn1 cKO mice (23) expressing ΔCre or Cre were labeled with the pan-neurexin antibody. We observed in the control neurons (ΔCre) the same synaptic nanoclusters as we had detected using HA-antibody in HA-Nrxn1 cKI neurons and found no difference in homer levels (Figure 2.7K-N, 2.8A-D).
Figure 2.8: Synaptic Nrxn1 nanoclusters are independent of HA-epitope tagging and contain other neurexins. A. Labeling with a pan-neurexin antibody shows neurexin nanoclusters in DIV12 hippocampal neurons from Nrxn1 cKO mice that are not abolished following Cre expression (scale bar = 0.5 µm). B-D. Neurexin content per nanocluster (B), proportion of neurexin(+) synapses (C) and area occupied by a nanocluster (D) are indistinguishable in Nrxn1 cKO mice expressing ΔCre or Cre. E. Expression of Cre in hippocampal cultures (DIV12) form Nrxn1/2/3 triple cKO mice abolishes nanoclusters labeled with a pan-neurexin antibody (scale bar = 0.5 µm). F-H. Pan-deletion of all neurexins abolishes neurexin content per nanocluster (F), proportion of neurexin(+) synapses (G), and area occupied by a nanocluster (H); “n.d.” indicates “not detectable.” I. Representative field of view (top) and a subset of exemplary Homer1(+) synapses containing HA-Nrxn1 (detected by anti-HA) and total neurexins (detected by pan-neurexin antibody) imaged using STORM. The Homer1 signal is imaged using conventional wide field microscopy (scale bar = 0.5 µm). J. Synaptic pan-Neurexin content of Homer1(+) synapses having one or more HA-Nrxn1 clusters compared to
those without clusters. **K.** Synaptic HA-Nrxn1 content of Homer1(+) synapses having one or more pan-Nrxn1 clusters compared to those without clusters. **L-M.** Fraction of HA-Nrxn1(+) (L) or pan-Nrxn(+) (M) synapses also containing pan-Nrxn and HA-Nrxn1 signal, respectively. **N.** Co-localization between neurexin nanoclusters defined by pan-Nrxn and HA-Nrxn1 using the coarse convex hull method (method 1) or convex hull method (method 2). Numerical data are means ± SEM. Statistical significance was determined by two-sample T-test to ΔCre or synapses having 0 HA-Nrxn1 clusters (*: p < 0.05, **: p < 0.01). For B,D, n as # of synapses = 17, ΔCre; 9, Cre / 3 cultures; for C/G, n = 3 cultures, averaged per culture; for F,H, n as # of synapses = 18, ΔCre; 14, Cre / 3 cultures; for J-M, n = 3 cultures; for N, n = 75 synapses).

Notably, the pan-neurexin antibody still recognized Nrxn nanoclusters in Nrxn1 cKO neurons (Figure 2.8A-D) and in HA-Nrxn1 cKI neurons (Figure 2.7Q-U) after expression of Cre-recombinase, whereas under the same conditions HA-antibodies no longer detected the nanoclusters (Figure 2.5C). However, we did not detect Nrxn nanoclusters in neurons cultured from Nrxn123 triple cKO mice (23) following expression of Cre (Figure 2.8E-H), confirming the specificity of the pan-neurexin antibody. No change in Homer1 levels or area were detected following Cre expression (Figure 2.7O-P). Thus, Nrxn2 and Nrxn3 likely substitute for Nrxn1’s absence in the synaptic nanocluster. Direct comparison of Nrxn1 nanoclusters observed using HA-antibodies in HA-Nrxn1 cKI neurons with neurexin nanoclusters observed using pan-neurexin antibodies in control neurons revealed that these nanoclusters had similar properties (Figure 2.7V-Z). Thus, Nrxn1 nanoclusters are independent of the HA-tag and contain other neurexins in addition to Nrxn1.

To confirm that pan-neurexin and HA antibodies recognize the same nanoclusters, we labeled hippocampal neurons from HA-Nrxn1 cKI mice with pan-neurexin and HA antibodies for two-color STORM imaging, and with Homer1 antibody for conventional wide-field microscopy (Figure 2.8I). We then measured the pan-neurexin signal at Homer1(+) synapses containing or lacking HA-Nrxn1 nanoclusters. We found that nearly all synapses containing HA-
Nrxn1 nanocluster(s) also featured pan-neurexin nanocluster(s) (Figure 2.8J, 2.8L), and that nearly all synapses containing a pan-neurexin nanocluster(s) also featured a HA-Nrxn1 nanocluster(s) (Figure 2.8K, 2.8M). Moreover, the HA-Nrxn1 and pan-neurexin clusters exhibited a high degree of co-localization (~80% volume overlap) (Figure 2.8N). Taken together, these findings confirm that HA and pan-neurexin antibodies recognize the same nanoclusters, which include Nrxn1 as well as other neurexin isoforms.

### 2.2.5 Functional Differences in Synapses Containing Nrxn1 Clusters.

What is the function of neurexin nanoclusters? Addressing this question is currently difficult, since there are no known neurexin mutants that selectively interfere with nanoclustering while preserving synaptic targeting. Extensive studies of Nrxn1 function using genetic manipulations *in vivo* followed by slice physiology have revealed that Nrxn1 has an important role in shaping synaptic properties, but not in the basic processes of synapse formation and synaptic transmission (24, 80). Using hippocampal neurons cultured from Nrxn1 cKI and cKO mice, we found that Nrxn1 is not essential for excitatory synapse formation or maintenance (Figure 2.9A-I, 2.9P-S). We also found no change in basic properties of synaptic transmission, including in spontaneous AMPA receptor currents (Figure 2.9J-O, 2.9T-Y).
Figure 2.9: Loss of synaptic Nrxn1 nanoclusters does not grossly affect excitatory synapse formation or function in cultured hippocampal neurons. A. Truncation of Nrxn1 does not alter excitatory synapse density in cultured neurons. Representative images show hippocampal neurons from HA-Nrxn1 cKI mice that were infected at DIV4 with lentiviruses expressing ΔCre (top) or Cre (to conditionally truncate HA-Nrxn1, bottom). Cultures were stained at DIV14 for
Another approach for ascertaining the function of neurexin nanoclusters is to compare the properties of synapses that contain or lack nanoclusters. We found that Homer1 levels were similar in synapses containing or lacking Nrxn1 or pan-neurexin nanoclusters (Figure 2.10A-F), suggesting that the overall size of synapses does not vary based on the presence of nanoclusters. We next asked whether nanoclusters were preferentially associated with active synapses. Since neurexin nanoclusters can only be identified using STORM, functional properties must be inferred optically. We used two optical approaches for these experiments, measurements of the relative amounts of postsynaptic surface GluA1 AMPA-receptors (16), and assessment of...
presynaptic vesicle exocytosis by live labeling of neurons with antibodies to the luminal domain of synaptotagmin-1 (Syt1) that becomes exposed during synaptic vesicle exocytosis (102). We observed that synapses containing Nrxn1 nanoclusters exhibited significantly more surface GluA1 receptors than synapses lacking Nrxn1 nanoclusters (Figure 2.10G-I), and displayed much higher levels of synaptic vesicle exocytosis (Figure 2.10J-L). These results suggest that excitatory synapses possessing Nrxn1 nanoclusters are functionally more active than those that don’t. These observations do not contradict our observation that conditional deletion of Nrxn1 had no effect on synapse density and activity for the following reasons: 1) Nrxn2 and Nrxn3 substitution at nanoclusters may compensate for the absence of Nrxn1 and 2) the function of Nrxn1 in cultured neurons may relate to additional synaptic properties not measured here, such as post-synaptic NMDA-receptor-mediated responses (80).
Figure 2.10: Excitatory synapses containing Nrxi1 nanoclusters exhibit higher surface GluA1 levels and more active presynaptic vesicle exocytosis. A. Representative synapses from HA-Nrxn1 cKi hippocampal neurons (DIV12) showing synapses without (top row) and with (bottom row) Nrxn1 nanoclusters (scale bar = 0.5 µm). B-C. Homer1 levels (B) and area (C) are indistinguishable at synapses having 0 or 1+ Nrxn1 nanoclusters. D. Representative images of individual synapses from wild-type hippocampal neurons (DIV12) showing synapses without (top row) and with (bottom row) pan-neurexin nanoclusters (scale bar = 0.5 µm). E-F. Homer1 levels (E) and area (F) are indistinguishable at synapses having 0 or 1+ pan-neurexin nanoclusters. G. Homer1(+) synapses containing Nrxn1 nanoclusters have higher levels of surface AMPA receptor subunit 1 (sGluA1) (bottom row) than synapses without nanoclusters.
Homer1(+) synapses containing one or more HA-Nrxn1 clusters relative to synapses without HA-Nrxn1 nanoclusters. J. Homer1(+) synapses containing Nrxn1 nanoclusters have higher levels of presynaptic vesicle exocytosis visualized following uptake of an antibody recognizing the luminal domain of Synaptotagmin-1 (Syt1). (scale bar = 0.5 µm). K-L. Luminal Syt1 antibody uptake levels (K) and area (L) are higher at Homer1(+) synapses containing one or more Nrxn1 nanoclusters relative to those that do not contain Nrxn1 nanoclusters. Numerical data are means ± SEM. Statistical significance was determined by a two-sample T-test to neurexin cluster # = 0 (*: p < 0.05, **: p < 0.01). For B-C, n as # of synapses = 27, 0 nanoclusters; 20, 1+ nanoclusters / 3 cultures; for E-F, n as # synapses = 24, 0 nanoclusters; 16, 1+ nanoclusters / 3 cultures; for H-I and K-L, n = 3 cultures, averaged per culture.

The fact that Nrxn1 nanoclusters are enriched in active synapses prompted us to ask whether synaptic activity regulates Nrxn1 nanoclusters. We found that Nrxn1 content, percentage of excitatory synapses with Nrxn1 nanoclusters, and the amount of Homer1 at excitatory synapses was not significantly influenced by manipulations of synaptic activity in mature (Figure 2.11A-H) and developing cultured neurons (Figure 2.11I-P). Instead, in a manner similar to what occurs during synapse maturation (Figures 2.5I-, 2.5O-P), increasing activity via synaptic disinhibition shifted HA-Nrxn1 nanoclusters from the center of the active zone towards the periphery in both developing and mature cultures (Figure 2.11E-F, 2.11M-N). Thus, although Nrxn1 nanoclusters are a marker of active synapses, their abundance is not controlled by synaptic activity.
Figure 2.11: Synaptic activity regulates the position of the Nrxn1 nanocluster, but not its abundance or susceptibility to proteolysis. A. Acute manipulation of synaptic activity does not alter the content of HA-Nrxn1 nanoclusters at excitatory synapses, but synaptic disinhibition shifts nanocluster position towards the periphery. Representative images show individual synapses labeled with antibodies recognizing Homer1 and HA-Nrxn1. Prior to labeling, cultured hippocampal neurons (DIV18) were treated with vehicle (0.1% DMSO), PTX (50 µM), APV (100 µM) + NBQX (10 µM), or APV (100 µM) + TTX (1 µM) for 2 hours to acutely manipulate synaptic activity (scale bar = 0.5 µm). B-D. The content of Nrxn1 per synapse (B), proportion of
synapses bearing Nrxn1 nanoclusters (C) and area of individual Nrxn1 nanoclusters (D) is unchanged following acute manipulation of synaptic activity. E-F. Treatment of neurons with PTX increased the absolute (E) and normalized (F) radial displacement of Nrxn1 nanoclusters from the center of the synapse, while other treatments did not alter nanocluster position relative to the control. G-H. Homer1 levels (G) and area (I) increased following APV + TTX treatment, but were unaffected by other treatments. I. Synaptic disinhibition of developing neurons accelerates movement of Nrxn1 nanoclusters towards the periphery of the cleft. Representative images of individual synapses labeled with antibodies recognizing Homer1 and HA-Nrxn1. Prior to labeling, hippocampal neurons were treated with vehicle (VEH, top) or PTX (bottom) for 48 hours starting on DIV10 (scale bar = 0.5 µm). J-L. Neurexin content per synapse (J), proportion of Homer1+ synapses bearing Nrxn1 nanoclusters (K) and area of individual nanoclusters (L) does not differ between neurons treated with either VEH or PTX. M-N. Treatment of neurons with PTX increases the absolute (M) and normalized (N) radial displacement of Nrxn1 nanoclusters away from the center of active zone. O-P. Homer1 content (O) and area (P) is reduced following treatment of neurons with PTX compared to VEH. Q. Immunoblot analysis of the HA-Nrxn1α fragment released into the medium and of cellular HA-Nrxn1α in hippocampal neurons cultured from HA-Nrxn1 cKI mice (Figure 2.12A). Direct comparison of Nrxn1 proteins in the cell lysates and medium from neurons expressing Cre or Cre-recombinase showed that without Cre-recombination, neurons release a Nrxn1 fragment into the medium that contains the HA-epitope and is ~20 kDa smaller than full-length Nrxn1 in cell lysates, suggesting that Nrxn1 is subject to site-specific proteolysis.

2.2.6 Dynamics of Nrxn1 expression and processing.

Cultured neurons release Nrxn1 fragments into the medium, suggesting that Nrxn1 is physiologically processed by ectodomain cleavage (Figure 2.3B). Does Nrxn1 ectodomain cleavage occur at a physiologically relevant rate? To begin to address this question, we examined secretion of Nrxn1 ectodomains in hippocampal neurons cultured from HA-Nrxn1 cKI mice (Figure 2.12A). Direct comparison of Nrxn1 proteins in the cell lysates and medium from neurons expressing Cre or Cre-recombinase showed that without Cre-recombination, neurons release a Nrxn1 fragment into the medium that contains the HA-epitope and is ~20 kDa smaller than full-length Nrxn1 in cell lysates, suggesting that Nrxn1 is subject to site-specific proteolysis.
in the sequence separating the HA-epitope from the membrane (Figure 2.12B). We measured the relative amount of ectodomain cleavage in hippocampal, cortical, and olfactory bulb neurons as a function of culture time (Figure 2.12C). Although hippocampal, cortical, and olfactory bulb neurons exhibited different Nrxn1 expression levels, they released comparable amounts of Nrxn1 fragments into the medium (Figure 2.12D-E). Surprisingly, Nrxn1 cleavage was not dependent on synaptic activity (Figure 2.11Q-S).

Figure 2.12: Nrxn1 is extensively cleaved across synaptic development in cultured neurons. A. Cultured hippocampal neurons from HA-Nrxn1 cKI mice for defined days in vitro (DIV5-18) were analyzed by quantitative immunoblotting directly (cell lysates) or after immunoprecipitation with HA antibodies (medium). B. Immunoblotting shows HA-Nrxn1α in cell lysates and in the medium of HA-Nrxn1 neurons expressing ΔCre, but detects only a faint amount of HA-Nrxn1α in either sample after Cre-mediated truncation of Nrxn1. C. Levels of cellular and cleaved Nrxn1α rise in parallel with culture time. Images show representative immunoblots of hippocampal, cortical and olfactory bulb neurons cultured from HA-Nrxn1 cKI mice and analyzed at DIV5, 7, 10 12, 15, and 18. D-E. The levels of cellular Nrxn1α and of Nrxn1α released by proteolysis into the medium increase coordinately with culture time. Summary plots depict protein levels determined by quantitative immunoblotting in the cells (D) and medium (E, monitored after immunoprecipitation) as a function of culture time. Data are means ± SEMs (n = 5, hippocampal; 6, cortical; 2, olfactory bulb cultures).
A potential concern of the experiments probing Nrxn1 ectodomain cleavage in cultured neurons is that neuronal cultures include activated astrocytes and microglia that may promote proteolysis of cell surface proteins. To address this concern, we examined the relative amount of the released soluble Nrxn1 fragment in mouse brain. We fractionated total brain homogenates (H) from HA-Nrxn1 cKI mice and from wild-type control mice into a membranous pellet (P) and a soluble supernatant (S), and analyzed these fractions by immunoblotting with a series of antibodies (Figure 2.13A). These analyses showed that the protein levels in the fractions from HA-Nrxn1 cKI and wild-type mice were similar overall, ruling out major changes induced in brain composition by the HA-Nrxn1 cKI (Figure 2.13B-C).

Figure 2.13: Nrxn1 is cleaved physiologically by proteolysis at approximately twice the rate of neuroligin-1. A. Brain subcellular fractionation protocol to obtain the total homogenate (H) and to separate soluble proteins (S) from membrane-associated proteins in the particulate fraction (P). B. Validation of subcellular fractions isolated from the hippocampus of adult HA-Nrxn1 cKI (left) and WT mice (right) at P56. Lanes were loaded with equivalent amounts of H and P fractions, but 10 times more of the S fraction. C. Nrxn1α and Nrxn1β are detected in the particulate and soluble brain fractions characterized in panel C. Note that soluble Nrxn1α and Nrxn1β proteins are smaller than cellular Nrxn1α and Nrxn1β, as would be expected for
fragments released by proteolytic cleavage. Note also that the HA-antibody only detects HA-Nrxn1 from cKI mice, whereas the pan-neurexin antibody detects Nrxn1 in both Nrxn1 cKI and in wild-type brain lysates. D-E. Soluble Nrxn1α fragments are produced in all brain regions, and are present at approximately twice the levels of soluble Nlgn1 fragments. Summary graphs show the amount of soluble Nrxn1α (left) and Nlgn1 (right) as percent of the total as determined by quantitative immunoblotting in subcellular fractions from the indicated brain regions of HA-Nrxn1 cKI mice analyzed at P56 (hippocampus, Hp; Cortex, Cx; Cerebellum, Ce; Olfactory Bulb, OB; Brain Stem, BS). Lanes were loaded with equivalent amounts of H and P fractions, but 20 times more of the S fraction. Data are means ± SEM; statistical significance was determined by a one-way ANOVA and Tukey’s post-hoc test (*: p < 0.05; ***: p < 0.001; n = 6 mice for the Nrxn1α and 3 mice for the Nlgn1 quantification).

Next, we quantified the relative amounts of ectodomain-cleaved vs. membrane-associated Nrxn1 in adult brain at P56 using the fractionation scheme outlined above. In these experiments, we also examined the levels of ectodomain-cleaved vs. membrane-associated neuroligin-1 (Nlgn1), a neurexin ligand that was previously shown to be proteolytically processed (94, 95). We observed the highest levels of Nrxn1 ectodomain cleavage in cortex and hippocampus, in which soluble Nrxn1 proteins represented ~6% of the total full-length Nrxn1 (Figure 2.13D-E). Nrxn1 ectodomain cleavage in the cerebellum was significantly lower than in other regions, suggesting that the tissue context of Nrxn1 expression may control ectodomain cleavage (Figure 2.13D-E). Importantly, the relative Nlgn1 ectodomain concentration was two-fold lower than that of the Nrxn1 ectodomain, suggesting that Nrxn1 ectodomain cleavage exceeds Nlgn1 cleavage (Figure 2.13D-E).
2.2.7 Nrxn1 ectodomain cleavage is largely mediated by ADAM10.

To determine what proteases are responsible for Nrxn1 ectodomain cleavage, we first surveyed ectodomain protease inhibitors for their ability to decrease Nrxn1 cleavage cultured neurons from HA-Nrxn1 cKI mice (Figure 2.14A-B). After application of various inhibitors at DIV10 for 48 h, we found inhibitors of matrix metalloproteases in general (GM6001) or ADAM-type sheddases in particular (TAPI-2) suppressed the release of the cleaved Nrxn1 fragment into the medium by ~60% in hippocampal and cortical neurons (Figure 2.14C-H). Because ADAM10...
is a major ectodomain metalloprotease in brain that has been suggested to cleave Nlgn1 (92, 95, 103), we tested more specific ADAM10/17 and ADAM10 inhibitors (GW280264X or GI254023X, respectively). In hippocampal neurons, we found that the ADAM10-specific inhibitor was as effective as the broad-spectrum metalloproteinase inhibitors and the ADAM10/17 inhibitor in blocking Nrnx1 ectodomain cleavage, but was less effective in cortical neurons, indicating that ADAM10/17 activity may differ between brain regions (Figure 2.14C-H). Blocking MMP9 (SB-3CT), which has been implicated in synaptic function and reported to mediate Nlgn1 cleavage (95, 104), had no effect on Nrnx1 ectodomain cleavage (Figure 2.14C-H). Similarly, blocking BACE1 (C3), another abundant ectodomain protease in the brain, did not significantly affect Nrnx1 ectodomain cleavage (Figure 2.14C-H). These data suggest that Nrnx1 ectodomain cleavage is primarily mediated by ADAM10. This conclusion is further supported by our findings that both a partial knockdown of ADAM10 and overexpression of a dominant-negative ADAM10 mutant decrease ectodomain cleavage of Nrnx1 (Figure 2.15A-F).

Figure 2.15: Effect of ADAM10 knockdowns or overexpression in cultured hippocampal HA-Nrxn1 cKI neurons. A. Strategy for manipulating HA-Nrxn1 processing in hippocampal
neurons via RNAi-mediated ADAM10 knockdown or overexpression of wild-type or mutant dominant-negative ADAM10. Neurons cultured from HA-Nrxn1 cKI were infected on DIV3-4 with lentiviruses encoding shRNAs targeting ADAM10 or a control shRNA, or with lentiviruses expressing wild-type or dominant-negative (DN) ADAM10, using two different dominant-negative ADAM10 proteins (E384A or ΔPro/MP). On DIV10, the HA-Nrxn1 cKI neuron culture medium was replaced with age-matched WT medium, and on DIV12 the culture medium and cell lysates of the HA-Nrxn1 cKI neurons were harvested and subjected to immunoprecipitation (for medium) and immunoblotting. In this manner, the HA-Nrxn1 cKI neuronal culture medium contains only material released during the last 48 hours. B. ADAM10 mRNA measurements of neurons expressing the control shRNAs or one of three different ADAM10 shRNAs. In controls, ADAM10 mRNA expression in cultured hippocampal neurons is approximately 50-fold lower than that of the housekeeping gene GAPDH. All three shRNAs caused a similar partial suppression of ADAM10 expression. C-D. Knockdown of ADAM10 in hippocampal neurons cultured from HA-Nrxn1 cKI mice decreases the release of proteolytic HA-Nrxn1α fragments into the medium. All three ADAM10 shRNAs reduced HA-Nrxn1 processing by approximately 50%. Only one of the three ADAM10 shRNAs caused a decrease in cellular Nrxn1α levels, possibly by an off-target effect. E-F. Overexpression of dominant-negative ADAM10 in hippocampal neurons cultured from HA-Nrxn1 cKI mice reduces the release of a proteolytic fragment from HA-Nrxn1 into the medium. Rather than being infected with shRNA, cultures were infected with lentiviruses to express ADAM10-WT or dominant-negative forms ADAM10-E384A or ADAM10-ΔPro/MP. Experiments were performed similar as A. Both dominant-negative ADAM10 variants decreased processing significantly, albeit, not to the same extent as the ADAM10-specific inhibitor GI254023X (GI). Overexpression of wildtype or dominant-negative ADAM10 in hippocampal neurons did not significantly alter the levels of cellular Nrxn1-HA. Summary graphs show means ± SEMs. Statistical significance for qPCR was determined with a one-sample T-test and quantitative immunoblotting was determined with a one-way ANOVA followed by a Dunnett’s multiple comparison post-hoc test (D, F). (*: p < 0.05, **: p < 0.01, ***: p < 0.001). For B, n = 3 cultures; for C-D, n as # of cultures = medium, 5; cellular, 3; for E-F, n as # of cultures = medium, 4 for all groups, except 3 for GI; cellular, 3
Figure 2.16: ADAM10 inhibition dramatically enhances Nrxn1 nanocluster presence and content at excitatory synapses. A. Representative 3D STORM images of Homer1(+) synapses with surface HA-Nrxn1 clusters visualized in hippocampal neurons treated with the indicated protease inhibitors from DIV10-12 (scale bar = 0.5 μm). Drugs are detailed in Figure 2.14. B. Pie charts of the number of nanoclusters per Nrxn1(+) synapse show that MMP/ADAM10 blockage increases the proportion of synapses containing Nrxn1 nanoclusters. C-G. Summary graphs showing that ADAM10 inhibition substantially increases the content of Nrxn1 nanoclusters (C) and the fraction of synapses containing Nrxn1 nanoclusters (C). The area of Nrxn1 nanoclusters (E), Homer1 levels (F), and Homer1 disc area (G) were not consistently and significantly affected by all metalloproteinase inhibitors. Data are means ± SEM. Statistical significance was determined by two-sample T-test to treatment with vehicle (*: p < 0.05; **: p < 0.01). For BC, D-G, n as # of synapses = 20, VEH; 49, GM; 45, GW; 39, GI; 20, C3 / 3 cultures; for D, n = 3 cultures, averaged per culture.

2.2.8 Blocking proteolytic Nrxn1 cleavage substantially enhances Nrxn1 nanoclusters.

We next asked whether inhibition of Nrxn1 ectodomain cleavage alters synaptic Nrxn1 nanoclusters. We imaged synapses by STORM at DIV12 after a 48 hour application of protease
inhibitors (Figure 2.16A). Strikingly, protease inhibitors that impaired Nrxi1 ectodomain cleavage caused a large enhancement of synaptic Nrxi1 nanoclusters. Specifically, the broad-spectrum matrix metalloprotease inhibitor (GM6001) or protease inhibitors targeting ADAM10 (GW280264X or GI254023X) increased the percentage of synapses containing Nrxi1 nanoclusters almost two-fold, enhanced the number of synapses with two or even three Nrxi1 nanoclusters, and augmented the Nrxi1 content of these nanoclusters (Figure 2.16B-D). The protease inhibitors additionally increased the physical size of synaptic Nrxi1 nanoclusters, although this effect was statistically significant only for the ADAM10 inhibitor GI254023X (Figure 2.16E). The inhibitors did not consistently affect the Homer1 content of synapses or the radial position of Nrxi1 nanoclusters (Figure 2.16F-G, 2.17A-B). In contrast to metalloprotease and ADAM10 inhibitors, BACE inhibition had no effect on synaptic Nrxi1 nanoclusters (Figure 2.16A-G, 2.17A-B). The substantial increase in the content of Nrxi1 nanoclusters and the doubling of synapses associated with Nrxi1 nanoclusters observed following ADAM10 inhibition suggests that nanoclusters are dynamically regulated by ADAM10, and that all (or nearly all) synapses are capable of acquiring such nanoclusters. Inhibition of ADAM10 did not significantly alter excitatory synapse density (Figure 2.17D-L), suggesting that the increase in synapses associated with Nrxi1 nanoclusters following ADAM10 inhibition is not due to a selective loss of synapses lacking Nrxi1 nanoclusters. Moreover, the actions of ADAM10 on constraining surface Nrxi1 levels appears to be a major feature of excitatory synapses, since none of the MMP inhibitors significantly altered the levels of Nrxi1 not associated with Homer1(+) synapses (Figure 1.17C). Altogether, these results demonstrate that Nrxi1 nanoclusters likely exist in a steady state of formation vs. destruction, partially or entirely.
Figure 2.17: Effect of MMP Inhibition on Nrxn1 nanocluster dynamics and synapse formation. A-B. Absolute (A) and normalized (B) radial displacement of HA-Nrxn1 with respect to Homer1 ‘discs’ of indicated protease inhibitor treatments. C. Summary graph showing that blocking MMPs does not affect the levels of HA-Nrxn1 not associated with Homer1(+) synapses, whereas C3 modestly decreases HA-Nrxn1 levels. D-E. Inhibition of ADAM10 does not prevent normal excitatory synapse formation in cultured hippocampal neurons. Representative images of HA-Nrxn1 cKI neurons following treatment with vehicle control or the ADAM10 inhibitor GI254023X (GI) for 48 hours starting on DIV10. Neurons were fixed on DIV12 and labeled with antibodies recognizing vGluT1, Homer1, and MAP2 (scale bar = 20 μm, left panels; 5 μm, right panels). F. Compared to control neurons, treatment with GI did not change the density of MAP2-associated excitatory synapses, which were positive for both Homer1 and vGluT1 puncta. G-L. The density, intensity, and area of MAP2-associated vGLUT1 puncta (G-I) and Homer1 Puncta (J-L) were not changed following GI treatment. Summary graphs show means ± SEMs. Radial distance was tested for significance using a rank-sum test (A-B) and remaining staining analysis with a T-test to VEH (*: p < 0.05). For A-B, n as # of synapses = 20, VEH; 49, GM; 45, GW; 39, GI; 20, C3 / 3 cultures; for C, n = 15 fields of view / 3 cultures; for F-L, n = 2 ROI per neuron / 9-10 neurons per culture / 3 cultures, avg. per culture).

2.3 DISCUSSION
Here, we identify by super-resolution imaging a previously unknown nanoscale organization of excitatory synapses characterized by discrete neurexin nanoclusters. We found that Nrxn1 is co-assembled with other neurexins in the synaptic cleft into nanoclusters, with a given Nrxn1(+) synapse generally containing only a single Nrxn1 nanocluster that is on average comprised of at least 4 Nrxn1 molecules and that is dynamically regulated by ectodomain cleavage but not by synaptic activity. Thus, the spatial distribution of Nrxn1 is neither similar to the proposed nanocolumns that align the presynaptic release machinery with postsynaptic receptors (83–85, 88), nor to the perisynaptic localization displayed by N-cadherin or SynCAM (89, 90). Given that neurexins function in regulating synapse properties (12), the Nrxn1 nanoclusters identified here may serve as pivots for the synapse-organizing functions of neurexins. In support of this hypothesis, we found that synapses containing Nrxn1 nanoclusters exhibit higher levels of postsynaptic GluA1 receptors and a greater rate of presynaptic vesicle exocytosis than synapses lacking Nrxn1 nanoclusters. In contrast to the Nrxn1 nanoclusters, the pre- and postsynaptic proteins that are aligned in nanocolumns were proposed to coordinate neurotransmitter release and reception (83), as recently revealed by localization of AMPA receptor nanocolumns as a function of long-term plasticity (85). Thus, the distinct structural organizations of Nrxn1 nanoclusters and repeated nanocolumns are consistent with the difference in their proposed roles in synaptic function.

Our study focused on excitatory synapses marked by the presence of Homer1, with Nrxn1 visualized both via a knocked-in HA-epitope tag and by application of a pan-neurexin antibody. Our conclusion that Nrxn1 assembles into a novel, dynamically regulated nanocluster in excitatory synapses thus is based on four major lines of evidence. First, using 3D STORM imaging, we found that Nrxn1 could be detected in synaptic nanoclusters that occupy 15-20% of
the synaptic cleft area, with most Nrxn1(+)-synapses having only a single nanocluster (Figures 2.4-2.8, 2.10, 2.11, 2.16, 2.17). Second, we found that Nrxn1 is present at more active synapses, which contain higher levels of synaptic vesicle exocytosis and surface GluA1, than synapses without Nrxn1 nanoclusters (Figure 2.10). Third, we found that Nrxn1 is physiologically processed by ectodomain proteases, with the dominant contribution of proteolytic cleavage coming from ADAM10 that cleaves Nrxn1 at a site between the Cys-loop domain and transmembrane region (Figure 2.3, 2.12-2.17). As a result of ectodomain cleavage, ~6% of Nrxn1 in adult mouse cortex is present as the released extracellular fragment; in contrast, the previously characterized ectodomain cleavage of neuroligin-1 (94, 95) causes a release of only ~2% of the extracellular neuroligin-1 fragment in adult cortex (Figure 2.13E). Fourth, pharmacological inhibition of ADAM10 proteolysis in cultured hippocampal neurons caused a substantial increase in both the number of synapses containing Nrxn1 nanoclusters and the Nrxn1 content of Nrxn1 nanoclusters (Figure 2.16), while not significantly affecting excitatory synapse number (Figure 2.17D-L). Remarkably, ADAM10 inhibition did not cause a significant change in total Nrxn1 protein, although it did decrease the release of the extracellular Nrxn1 fragment by more than 60% (Figure 2.14). These findings show that ADAM10 inhibition not only caused a decrease in Nrxn1 proteolytic turnover, but also a redistribution of Nrxn1 from a ‘free’ state into synaptic nanoclusters. Since ADAM10 is a versatile multifunctional protease that is tightly regulated (103, 105), these data suggest that Nrxn1 nanoclusters can be controlled via regulation of ADAM10. Based on these data, a plausible mechanism would be that Nrxn1 molecules are continuously recruited to the nanoclusters and released into the extracellular space via ectodomain cleavage by ADAM10, thus maintaining Nrxn1 nanoclusters in a dynamic equilibrium.
Is it possible that the nanoclusters constitute experimental artifacts? We raise this question particularly because we first identified the Nrxn1 nanoclusters using knockin mice in which endogenous Nrxn1 was tagged, which may have changed its properties. To address these concerns, we showed that Nrxn1 nanoclusters were similarly observed, with identical properties and similar developmental dynamics, in cultured hippocampal neurons and in hippocampal sections (Figure 2.4-2.6), and that the same nanoclusters were detected independent of labeling and fixation conditions (Figure 2.7, 2.8). Most importantly, we identified the same nanoclusters using a pan-neurexin antibody and neurons expressing unmodified neurexins (Figure 2.7, 2.8), and that HA-Nrxn1 and pan-neurexin localizations mostly co-localized (Figure 2.8I-N). Finally, we documented that the nanoclusters were abolished upon genetic deletion of neurexins 1, 2 and 3 (Figure 2.8E-H). We believe that these observations together rule out conceivable artifacts. A recent immunoelectron microscopy study using the same pan-neurexin antibodies that we applied for control labeling obtained results consistent with a non-random distribution of neurexins (106).

The discovery of Nrxn1 nanoclusters in excitatory synapses raises new questions. For example, do Nrxn2 and Nrxn3 also assemble into these nanoclusters, how do the nanoclusters form, and does alternative splicing of neurexins regulate the nanoclusters? Partial answers to the first question are already provided by our experiments demonstrating that upon deletion of only Nrxn1, the pan-neurexin antibody still detects synaptic nanoclusters, whereas deletion of all neurexins abolishes synaptic neurexin nanoclusters (Figure 2.8). This result reveals that Nrxn2 and/or Nrxn3 detected by the pan-neurexin antibody must also be present in the nanoclusters. As regards the mechanism of neurexin nanocluster formation, it seems likely that interactions of their cytoplasmic tail with the cytoskeleton (107, 108) or specific features of their glycosylation
(99) may be important, although testing this will be difficult since overexpression of mutant proteins is likely to introduce unintended side effects. Finally, the issue of alternative splicing seems particularly important to us given the close proximity of splice site #5 (SS5) to the region of ADAM10 cleavage in Nrxn1 (Figure 2.12-2.15), providing a possible mechanism for alternative splicing to regulate both neurexin cleavage and nanocluster dynamics. However, these three questions are only the proverbial tip of the iceberg with regards to synapse organization and neurexin function, an overall question that will occupy the field for years to come.

2.4 Materials and Methods

2.4.1 Generation of HA-Nrxn1 cKI Mice.

In generating HA-tagged Neurexin-1 conditional knockin (cKI) mice we aimed to (1) introduce a double hemagglutinin tag (2xHA-tag) to allow the detection of the HA-epitope tagged Nrxn1 including alpha-, beta-, and –gamma isoforms, and (2) to introduce a single loxP-site into the reading frame to allow for Cre-mediated in-frame recombination resulting in truncated Nrxn1 proteins (Figure 2.1A). Transgenic mice were generated by homologous recombination in R1 embryonic stem cells (109) using the strategy outlined in Figure 2.2A that targets the last exon of Nrxn1, exon-26 (79, 110). Briefly, immediately after the first two bases of Nrxn1 exon-26, in maintaining frame order, we introduced the coding sequence of two hemagglutinin motifs (amino acid sequence: YPYDVPDYA) as well as the 34 bp sequence of a ‘wild-type’ loxP-site complemented by two additional bases to avoid a frame shift and a SpeI restriction site for cloning purposes. The amino acid sequence of the translated loxP-site including the two additional bases and the SpeI site within the mutated full exon 26, reads as follows: ITSYSIHYTKLSTS. The second loxP-site was introduced immediately downstream of
the endogenous stop codon. Upon Cre/loxP-mediated recombination the coding region now terminates in a stop codon shortening the translated loxP-site into the sequence ITSYSIHYTEKL. Since the transmembrane region is encoded by the sequence downstream of the introduced loxP-site, Cre/loxP-mediated recombination will result in a truncated Nrxn1 protein lacking its membrane anchoring sequence and intracellular C-terminus.

The 3 prime loxP site is followed by a frt-site flanked neomycin resistance gene under control of the phosphoglycerate-kinase promoter, PGK-neoR, to allow for the selection of homologously recombined ES cells. The mutated Nrxn1 exon 26 and the selection cassette were both flanked by an approx. 6 kb homology regions, respectively. Additionally, upstream of the 5 prime homology region, a diphtheria toxin minigene was inserted for negative selection. Electroporation into (129X1/SvJ x 129S1/Sv)F1-Kitl++-derived R1 embryonic stem cells, drug selection and clone isolation was performed as previously described (111) (UT Southwestern Transgenic Facility, Dallas, TX, USA). Positive ES clones were identified by Southern blotting and PCR genotyping. Blastocyst injection was performed by the Stanford Transgenic Facility, Stanford, CA, USA. Chimeric offspring were bred to C57BL/6 mice to select for the recombined allele. Mutated mice were subsequently bred to Flp-deleter mice (112) for the removal of the frt-flanked selection cassette, Nrxn1KI allele, and Cre-deleter mice (113) generating the Nrxn1 truncated allele, Nrxn1tr. Both deleter mouse lines were maintained on a C57BL/6 background (JAX stock number #003800 (Flp-deleter), #006054 (Cre-deleter)) and Nrxn1 cKI mice have been bred several rounds to C57BL/6 mice with the intention to remove the transgenic Flp- and Cre-alleles so that all experiments were performed on congenic backgrounds of at least >95% C57BL/6 unless noted differently. The HA-Nrxn1 cKI mouse line was deposited with the
Jackson Laboratory Mouse Repository for distribution (B6.129-Nrxn1<sup>tm3Sud</sup>/J, stock number #021777).

All mouse work was performed as prescribed by approved protocols at Stanford University. Mice were weaned at 20 days of age and housed in groups of up to five on a 12h light/dark cycle with food and water ad libidum. Animals were kept in the Stanford Animal Housing Facility with all procedures conforming to the standards set by the National Institutes of Health Guidelines for the Care and Use of Laboratory Mice and approved by the Stanford University Administrative Panel on Laboratory Animal Care.

The following primers were used for genotyping to discriminate between different alleles:

MX10434 (5’-GCCTTGAGAGGGTGAACACTTTATTTGG-3’)

MX10435 (5’-CCATTGGACTGTGCTGAGTTACTGATG-3’)

MX10472 (5’-TTTGCTTTATGAATGCGTGCGGTCCTCACC-3’).

Oligonucleotide combinations MX10434/MX10435 were used to identify the presence of Nrxn1<sup>wt</sup> (382 bp) and Nrxn1<sup>Ki</sup> (472 bp) alleles, MX10434/MX10472 identified the Nrxn1<sup>tr</sup> (682 bp) allele.

Survival analyses were performed by comparing the genotype distribution observed in litters resulting from Nrxn1<sup>wt/tr</sup> x Nrxn1<sup>wt/tr</sup> (het x het) crosses at either P1 or P21. Observed genotype distribution was compared to expected distribution based on Mendelian inheritance (114). The Chi-Square test was used to determine whether the observed offspring ratio differed significantly from the expected ratio.
2.4.2 Cell Culture.

2.4.2.1 Primary Neuronal Culture.

Hippocampal, olfactory bulb and cortical neurons were cultured from newborn mice as described previously (115) with some modifications. Dissected hippocampi, olfactory bulbs, or cortices were digested for 20 minutes with 10 U/ml papain in HBS buffer in an incubator, washed with HBS buffer, dissociated in plating media (MEM supplemented with 0.5% glucose, 0.02% NaHCO₃, 0.1 mg/ml transferrin, 10% FBS, 2 mm l-glutamine, and 0.025 mg/ml insulin), and seeded on Matrigel (BD Biosciences) pre-coated coverslips placed inside 24-well dishes. The day of plating was considered as 0 days \textit{in vitro} (DIV0). After 24 hours (DIV1), 95% of the plating media was replaced with neuronal growth media (Neurobasal-A media supplemented with 2% B27 supplement and 0.5 mm l-glutamine). At DIV2-3 (for hippocampal and olfactory bulb cultures) or DIV3-4 (for cortical cultures), 50% of the medium was exchanged with fresh growth medium additionally supplemented with 4 μm Ara-C (Sigma-Aldrich) to restrict glial overgrowth. When applicable, lentivirus supernatants were always added on DIV3-4 following exchange of media. For long-term culture of neurons, 25% fresh media was added every 4–5 days starting from DIV7. Most imaging experiments utilized P0 pups from HA-Nrxn1 cKI mice. For biochemistry experiments, HA-Nrxn1 cKI cultures were prepared simultaneously from newborn HA-Nrxn1 cKI mice and wild-type cultures were prepared from newborn CD1 mice. For select control experiments, cultures were also prepared from Nrxn1 cKO mice and Nrxn1/2/3 cKO mice (23). For generation of cultures form P0 pups, gender was not taken into account. In general, pooling tissue from 3-6 mice in a given preparation was used to generate cultures.

2.4.2.2 HEK293T Cells.
Wild-type Human embryonic kidney 293 T (HEK293T) cells were maintained in Dulbecco’s Modified Eagle’s medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and penicillin and streptomycin. For producing lentivirus, cells were transfected using the calcium phosphate method (as described below).

2.4.3 Quantitative real time PCR (qPCR).

2.4.3.1 Brain Tissue.

For quantitative RT-PCR analysis, age-matched, adult (10 week old) males from the same litters were used and were either homozygous for the \textit{Nrxn1} wild-type locus (wild-type mice) or carried a single truncated allele, \textit{Nrxn1}^{wt/tr} (n = 4 per genotype). Mice were euthanized using isoflurane, the entire was brain dissected and then collected in TRIZol reagent. RNA was isolated according to the manufacturer’s protocol (Invitrogen, Carlsbad, CA, USA). RNA concentration was measured using a ND-1000 spectrophotometer (Thermo Scientific, Waltham, MA). Briefly, reactions were loaded onto an ABI7900 fast RT-PCR machine (Applied Biosystems, Foster City, CA) and run in triplicates with 100 ng RNA per run using VeriQuest Probe one-step qRT-PCR master mix with ROX (Affymetrix Inc., Cleveland, OH, USA). Gene-specific qPCR assays were purchased from Integrated DNA Technologies (Coralville, IA, USA), mouse ACTB assay (4352933E) was purchased from Applied Biosystems. The following PrimeTime qPCR Assays (IDT) were used (shown as gene, primer1, probe, primer2): Nrxn1\textalpha, TCCTCTTAGACATGGGATCAGG, CAACGGGATGGACCGGTCTAGGT, GTGTAGGGAGTGGGTAGTG; Nrxn1\textbeta, TGGCCCTGATCTGGATAGTC, ACCACATCCACCATTTCAT, AATCTGTCCACCACCTTTGC, Nrxn2\textalpha/\textbeta, ATCATCATTGGACACTCAGC, CAGGAGGTCATCTGTTTCTGTTGT, ACAATGAGGGACAGCAC, Nrxn2\textalpha, GTCAGCAACAACTTTCATGG,
CTTCATCTTCCGGGTCCCCCTTCCCT, AGCCACATCCTCACAACG; Nrxn2β,
CCACCACTTCCACAGCAAG, GGACCACATACATCTTCGGG,
CTGGTGTGCTGCTGAAGCCTA; Nrxn3α/β, CTTTTGCTCTTTCCTCCGATG,
TTTTCCTGCAGCACTCCTCTCTACG, CACTGATAATGAACGCCTCCA; Nrxn3α,
GGGAGAACCCTGCAGGAAGAG, CTGCCGTCATAGCTCAGGATAGATGC,
ATGAAGCGGAAGGACACATC; Nrxn3β, CACCACTCTGTGCCTATTTC,
TCTATCGCTCCCTCTTTTCC, GGCCAGGTATAGAGGATGA; Nlgn1,
GGTTGGGTCTTCATGGATAGA, TGAGGAACTGGTGTGATTTGGGTCACC,
GATGTTGAGTGCAGTAGTAATGAC; Nlgn2, CCGTGTAGAAACACGATCGACC,
TCAATCCGCCAGACACAGATATCCG, TGCCCTGTACCTCAACCTCTA; Nlgn3,
CAGTCTCGGATGTCTTTCA, CCTGTTTCTTAGCGCCGGATCCAT,
CCTCTATCTGAATGTGTATGTGC; LRRTM2, GGCCACTTGAATGTAAGCC,
TGCAGCCTCAATGTGCTCAGAA, CACTGCGTTGAGTCTGACAA; LRRTM3,
CATATGCCAGAAAGGTGACAC, AGGCTCCAGGAATGTGAGATACCT,
GAGATGCTGCTGAACGGGA; LRRTM4, GAAATAGCACCAGAAACACACTC,
ACGGAACCATCCTTTTGCTCAATCCA, GACCAATAAGAAGAAAGCTGAGAG; Lphn1,
GACTGATGCTCGACTCATGT, TGGGCACACAGAAGTGAAGGGGAC,
CTGGAAACCTACAAAATACCTGGGA; Lphn2, CTCGTGGTGAGTATTTGTTGGT,
TGACCCTGCCAAGTGCTCCTAC, TTACGGGTATTCCTGAGTGGTG; Lphn3,
AGAAACATCCAGGTGAAGGC, AAGAAATGCAAAAAGGAGCCCGGAACA,
GAATCAAAGAACCAGCAACAC. The Nrxn1-HA assay was designed to identify the presence
of the mutated alleles (Nrxn1\textsuperscript{KI} and Nrxn1\textsuperscript{tr}), its sense oligo (5'-'ACAGATGACATCCTTTTGCC-3') and the probe (5'-'ACATTGACCCCTGTGAGCCGAG-
3’) are both located in the proceeding exon, the antisense oligo (5’GTTATAGCATAGTCAGGTACGTCG-3’) is located in the mutated sequence of exon 26 including parts of the HA coding sequence and the first 5 bases of the loxP-site. The relative transcript level of the target mRNA in each sample was calculated by normalization of Ct values to the reference mRNA (β-actin) using the following equation: \( V = 2^{\text{CT}[\text{reference}]/\text{CT}[\text{target}]} \), \( V \) is the relative value of target gene normalized to the reference.

### 2.4.3.2 Primary Neurons

RNA was isolated from DIV12 neurons that had been previously infected with lentivirus on DIV3-4 to allow Cre-mediated gene deletion or knockdown of ADAM10 using shRNA (see below). RNA was isolated using the PrepEase® RNA spin kit according to the manufacturer’s instructions (Affymetrix Inc., Cleveland, OH, USA). Nrxn1, Nrxn3, Lphn1, and Nlgn1 probes are described above. A pre-validated mouse ADAM10 assay (Assay ID: Mm.PT.58.14225600) was obtained from Integrated DNA Technologies (Coralville, IA, USA). For each reaction, 1 microliter of RNA was run in triplicate using VeriQuest Probe one-step qRT-PCR master mix with ROX (Affymetrix Inc., Cleveland, OH, USA). Reactions were loaded onto a QuantStudio3 Real-Time PCR machine (ThermoFisher, Waltham, MA). The relative transcript level of the target mRNA in each sample was calculated as described above.

### 2.4.4 Lentivirus Production and Infection

Recombinant lentiviral particles were produced in HEK293T cells by co-transfecting cells with LTR-containing vector and helper plasmids (pRSV-REV, pMDLg/gRRE, and pVSVG) using calcium phosphate. Media was exchanged 1 hour prior to transfection and included 25 μM of chloroquine diphosphate. Per 75 cm² of cells, 0.5 mL 250 mM CaCl₂ containing molar equivalents of DNA (12 μg of LTR-containing vector, 3.9 μg pRSV-
REV, 8.1 μg pMDLg/gRRE, and 6.0 μg pVSVG) was added dropwise to an equal volume of 2X-HBS (0.4 M NaCl, 10 mM KCl, 1.5 mM Na₂HPO₄, 0.2% glucose, 38.4 mM Hepes, pH 7.05) under vigorous mixing, incubated for 20 minutes at room temperature and added drop-wise to the cells. 16-20 hours following transfection, cells were washed with plain DMEM and replaced with neuronal growth media lacking AraC. After 24 hours, media containing lentiviral particles was cleared by centrifugation (1,500 x g, 10 minutes), aliquoted, and snap-frozen. Neuronal cultures were infected with lentivirus on DIV3 or 4 by adding 25-30 μL of viral supernatant per well of a 24-well plate.

For generating conditional knockout cultures, matched sets of lentiviruses were produced using LTR-containing vectors that included Cre or mutant ΔCre (inactive) fused to EGFP and driven by the human synapsin-1 promoter (116). For gene knockdown, lentiviruses were produced using lentiviral constructs containing shRNAs against mouse ADAM10 purchased from MISSION shRNA Library (Sigma-Aldrich). Three different shRNA’s were used for ADAM10 knockdown in the pLKO.1 lentivirus expression vector, including ADAM10 shRNA 1 (TRC Number: TRCN0000031844; Sequence: CCGGGCAGAGAGATACATTAAATGTATCTCTCTCTGCTTTT G), ADAM10 shRNA 2 (TRC Number: TRCN0000031847; Sequence: CCGGCAGCTCTATATCCAGACAGATCTCGAGATCTGTCTGGATATAGAGCTGTTTTT G), and ADAM10 shRNA 3 (TRC Number: TRCN0000031848; Sequence: CCGGCCAGGAGTCTAAGAATCTTACTCGAGAAGTTCTTTAGACTCTCTCTGTTTTT G). All other viruses were generated using a similar lentiviral backbone as Cre (i.e. FSW), which contains the human synapsin-1 promoter to drive gene expression in neurons (described further below). For ADAM10 wild-type and mutant overexpression, FSW viruses were prepared either
containing no insert (control), wild-type ADAM10 (A10-wt), ADAM10-E384A (A10-E384A), or ADAM10 lacking its prodomain and metalloproteinase domain (A10-ΔPro/MP).

2.4.5 Electrophysiology

For electrophysiology recordings, hippocampal neurons were prepared from newborn Nrxn1 cKO or HA-Nrxn1 cKI mice. Cultures were infected with lentiviruses carrying Cre or mutant ΔCre (inactive) fused to EGFP on DIV3. Culture coverslips were superfused with ACSF solution (in mM): 120 NaCl, 2.5 KCl, 1 NaH2PO4, 26.2 NaHCO3, 2.5 CaCl2, 1.3 MgSO4-7 H2O, 11 D-Glucose, ~290 mOsm. mEPSCs were recorded with an internal solution containing (in mM): 117 Cs-methanesulfonate, 15 CsCl, 8 NaCl, 10 TEA-Cl, 0.2 EGTA, 4 Na2-ATP, 0.3 Na2-GTP, 10 HEPES, pH 7.3 with CsOH (~300 mOsm). mEPSCs were recorded in ACSF containing 0.5 μM tetrodotoxin (TTX) and 100 μM picrotoxin. Miniature events were handpicked and analyzed in Clampfit 10 (Molecular Devices) using template matching. Synaptic currents were monitored with a Multiclamp 700B amplifier (Molecular Devices). Data were collected at 10 kHz and filtered with a lowpass filter at 2 kHz. For all experiments, the experimenter was blind to the recording condition.

2.4.6 Immunocytochemistry

For live surface labeling experiments, primary neurons were first washed at room temperature once with 1X HEPES bath solution, which contained the following (in mM): 140–150 NaCl, 4–5 KCl, 2 CaCl2, 1 MgCl2, 10 glucose, and 10 HEPES, with pH adjusted to 7.4 with NaOH, and osmolarity of 300 mOsm. For conventional fluorescence imaging, cultures were then incubated at RT for 20 minutes with purified mouse anti-HA monoclonal antibody (1:250; BioLegend, anti-HA.11) diluted in 1X HEPES bath solution. For STORM imaging experiments, cultures were incubated with the same antibody dye-labeled with 405-647 (1:250) in 1X HEPES
bath solution at RT for 20 minutes. For double-labeling Nrxn1 nanoclusters with markers for synapse activity, dye-labelled anti-HA 405-647 was co-incubated with either anti-GluA1 (Calbiochem, PC246) at 1:10 or an antibody recognizing the luminal domain of Synaptotagmin-1 (Synaptic Systems, 105103) at 1:100 for 20 minutes at RT. Cultures were then gently washed 3 times with 1X HEPES bath solution, followed by fixation for 20 minutes at RT with 4% (wt/vol) paraformaldehyde (PFA). Following fixation, cultures were washed 3X with DPBS. For surface labeling experiments to be used for conventional imaging, cultures were blocked for 1 hour at RT with antibody dilution buffer ADB(-), which contains 5% normal goat serum diluted in DPBS. Cells were then labelled with Alexa-conjugated secondary antibodies (1:1000, Invitrogen) diluted in ADB(-) for 2 hours at RT. Cells were washed and then permeabilized and blocked for 1 hour with ADB(+), which contains 0.2% triton X-100 and 5% normal goat serum diluted in DPBS. Non-surface primary antibody were diluted in ADB(+) and cells were incubated O/N at 4 °C. Cultures were washed 3X and then incubated with either Alexa-conjugated secondary antibodies (1:1000, Invitrogen) or STORM secondary antibodies (1:250; described below) in ABD(+) for 2 hours. Following 3 washes, coverslips for conventional imaging were inverted onto glass microscope slides with Fluoromount-G mounting media (Southern Biotech). For STORM imaging, samples were stored in the dark at 4 °C until further processing, for no longer than one week until downstream processing.

For non-surface labeled stains, cultures were first washed 1X with DPBS followed by a 20 minute fixation at RT with 4% (wt/vol) paraformaldehyde (PFA). Following fixation, cultures were washed 3X with DPBS. For stronger fixation (as used for Figure 2.7F-J), cultures were incubated with 3% PFA and 0.1% glutaraldehyde for 20 minutes, followed by treatment with sodium borohydride (1 mg/ml, 12 minute incubation) to quench autofluorescence. Cultures were
incubated for 1 hour with ADB(+), followed by overnight incubation at 4 °C with antibodies diluted in ADB(+). Coverslips were washed 3X and then incubated for 2 hours with either Alexa-conjugated secondary antibodies (1:1000, Invitrogen) or STORM secondary antibodies (1:250) in ABD(+). Coverslips were further processed as described above for surface labelled specimens.

The antibodies used for conventional imaging included: Rabbit anti-HA (1:250, Cell Signaling, #3724), mouse anti-HA monoclonal antibody (1:1000 for pre-fixed; BioLegend, anti-HA.11), rabbit anti-Homer1 (1:500; Synaptic Systems, 160003), guinea pig anti-Homer1 (160004), rabbit anti-VGLUT1 (1:1000; TCS, YZ6089), rabbit anti-Synapsin (1:500; TCS, E028), guinea pig anti-VGLUT1 (1:500; Millipore, AB5905), rabbit anti-VGAT (1:500, Millipore; AB5062P), chicken anti-MAP2 (1:1000; Encor, CPCA-MAP2), rabbit anti-Munc13-1 (1:500; Synaptic Systems, 126103), and rabbit anti-RIM1/2 (1:500; Synaptic Systems, 140203). The antibodies used for STORM imaging included: rabbit anti-pan-Neurexin (1:250; Frontier Institute, Af870), rabbit anti-Piccolo (1:200; Synaptic Systems, 142003), rabbit anti-Bassoon (1:200; Synaptic Systems, 141003), mouse anti-Homer1 (1:500; Synaptic Systems; 160011), and rabbit anti-Homer1 (1:500; Synaptic Systems, 160003). For STORM imaging, Neurexin-1-HA or pan-Neurexins were directly labeled with mouse HA and rabbit anti-pan-Neurexin (1:250, Frontier Institute, Af870) that had been pre-conjugated with 405-647 dye pairs (see below). Remaining antibodies used for STORM were recognized by dye-labelled secondary antibodies (described below).

For conventional microscopy (Figures 2.2, 2.3, 2.9, 2.17), confocal images were acquired at room temperature using an inverted Nikon A1 Eclipse Ti confocal microscope (Nikon) equipped with a 60× objective (Apo, NA1.4) and operated by NIS-Elements AR acquisition
software. Images were taken at 1,024 × 1,024 pixels with z-stack distance of 0.5 µm. Images were acquired sequentially in order to avoid bleed-through between channels. Imaging parameters (i.e. laser power, photomultiplier gain, scan speed, etc.) were optimized to prevent pixel saturation and kept constant for all conditions within the same experiment. Images were analyzed using NIS-Elements Advanced Research software (Nikon).

2.4.7 Excitatory Synapse Quantification

For Figure 2.9, 3-4 independent cultures were prepared and separate wells were infected with lentivirus carrying Cre or ΔCre. At DIV14, cultures were fixed and stained as described above for either vGluT1, total HA, Homer1, and MAP2 (for Nrxi1 cKI) or just MAP2 and vGluT1 (for Nrxi1 cKO). For Figure 2.17J-L, DIV10 HA-Nrxi1 cKI hippocampal neurons were incubated for 48 hours with GI250423X followed by fixation and staining of vGluT1, MAP2, and Homer1 on DIV12. For all experiments, 9-10 neurons were blindly selected for both Cre and ΔCre groups per culture. Neurons were always confirmed to be healthy (i.e. continuous processes) and have pyramidal morphology by viewing the MAP2 staining. General Analysis was performed with NIS-Elements and binaries were generated following background subtraction applied equally to groups being compared and imaged on the same day. Binary operations were used to distinguish Homer1 and vGluT1 puncta associated with MAP2. Two regions of interest were blindly drawn on secondary and tertiary dendrites of imaged pyramidal neurons. Quantification of puncta density, area and intensity was averaged per culture before statistical analysis.

2.4.8 Tissue Preparation and Staining for STORM

2.4.8.1 Tissue Preparation
Hippocampal tissue slices were prepared for STORM imaging following a modified protocol described previously (38). Specifically, mouse brains that have been perfusion fixed with 4% PFA (Electron Microscopy Sciences) were cryo-protected by immersion in 30% sucrose (Sigma-Aldrich) O/N and embedded in a 2:1 mixture of 30% sucrose:TissueTek® OCT Compound (Sakura). Frozen tissue blocks were cut into 10 µm sections on a cryostat and collected on glass slides (Thermo Fisher Scientific). Sections were blocked with 5% bovine serum albumin (BSA, Jackson ImmunoResearch) in 1X DPBS and incubated with anti-hemagglutinin tag (HA, BioLegend) antibody labeled with Alexa405 and Alexa647 (Thermo Fisher Scientific) dyes following previously described protocols (44) for 30 minutes at room temperature (RT) in 5% BSA in 1X DPBS and washed 3X for 15 minutes each with 1% BSA in DPBS. Sections were then permeabilized with 0.1% w/v saponin (Sigma-Aldrich) in 5% BSA in 1X DPBS for 30 min, and stained O/N at 4 °C in 5% BSA in 1X DPBS with antibodies against Homer1 (1:400; Synaptic Systems), Piccolo (1:400; Synaptic Systems), or Bassoon (1:400 Enzo Life Sciences). After washing 3X with 1% BSA in PBS, samples were stained for 2 hours at RT with appropriate secondary antibodies (Jackson ImmunoResearch) labeled with Alexa647 and Cy3 (Invitrogen) in 5% BSA in 1X DPBS, and washed 3X with 1% BSA in 1X DPBS. Sections were then immediately imaged.

2.4.8.2 Imaging Buffer and Sample Mounting

The STORM imaging buffer was prepared with 100 mM cysteamine (Sigma-Aldrich), 5% w/v D-glucose (Sigma-Aldrich), 0.8 mg/mL glucose oxidase (Sigma-Aldrich), and 40 µg/mL catalase (Sigma-Aldrich) in 1X DPBS. For tissue sections, 100 µL imaging buffer was dropped onto glass slides holding the cryosection, and covered with a #1.5 rectangular glass coverslip (22 mm × 30 mm, Thermo Fisher Scientific). Excess imaging buffer was wicked away and the
sample was sealed using nail polish. For cultured neuron sections, 150 µL imaging buffer was first deposited into an imaging chamber before the sample coverslip was inverted over the chamber, and pressed down firmly before sealing with nail polish.

2.4.8.3 **STORM setup**

STORM imaging experiments were conducted on a custom-built Nikon Eclipse-Ti inverted microscope with illumination channels at 405-nm (OBIS 405-LX, Coherent), 460-nm (Sapphire 460-10, Coherent), 488-nm (Genesis MX488-1000 STM, Coherent), 561-nm (Sapphire 561–200 CW CDRH, Coherent), and 647-nm (F-04306–113; MPB Communications). Lasers were introduced through the backport of the microscope, passing through a ZT405/488/561/647/752RPC (Chroma) dichroic mirror and ZET405/488/561/647–656/752 penta-band notch filter (Chroma) emission filter. A translation stage allowed the incident beams to be directed at low incidence angles to illuminate only fluorophores near the coverglass (i.e. within ~1-2 µm of the coverglass). For 3D imaging, a Roper Scientific Dual View system was inserted between the microscope body and the EMCCD camera (iXon, DU-897E-CSO-#BV, Andor), where a 1 m focal-length cylindrical lens was inserted in place of a beam-splitter. Focus was maintained via a custom-built feedback system in which an 850-nm infrared laser was directed into the objective back aperture, and its reflection from the glass/liquid interface was imaged onto a quadrant photodiode, driving a piezo objective nanopositioner (Nano F-100, Madcity Labs).

2.4.8.4 **STORM Imaging and analysis**

For 2-color STORM imaging, the 405- and 561-nm lasers were used to activate the corresponding Alexa405-Alexa647 or Cy3-Alexa647 dye pair, and the 647-nm laser was both used to excite Alexa647 fluorophores and switch them into the dark state. STORM movies were
acquired at 60 Hz until most dyes were photobleached within the field of view. During acquisition, the power of the activation lasers (405-nm and 561-nm) were gradually adjusted so that only a small subset of fluorophores were activated at in any frame, allowing individual activated molecules to be imaged and localized, and the two color channels were roughly even in their activation. The 647-nm laser was maintained at ~80 mW throughout the experiment, while the 405-nm and 561-nm lasers were ramped up to a maximum of 2 mW and 1 mW, respectively.

The illumination lasers were either mechanically or digitally shuttered in 4-frame events consisting of one frame of activation laser (405- or 561-nm) and 3 frames of imaging laser (647-nm), and alternating in the activation channels (i.e. between 405 nm and 561 nm) for four frames in each channel. STORM movies were automatically analyzed using custom-written software implementing 3D-DAOSTORM (117), generating molecule lists for each STORM image that could be used for subsequent analyses. In brief, the x and y positions of the fluorophores are determined from the centroid positions of their images, and the z position of the fluorophores are determined from the ellipticity of their images, as previously described (29, 45). Cross-talk between color channels was subtracted using methods previously described for activator-based multi-color STORM (38). For 3-color imaging, a conventional image was acquired using the 488-nm laser to excite Alexa488, labeling Homer1, in addition to the 2-color STORM image.

2.4.8.5 Synapse selection

To select synapses in 2-color STORM images, homer localizations were first clustered through an implementation of DBSCAN with a distance of 80 nm and a minimum of 5 points per assigned cluster. To be designated as a synapse, a homer cluster must contain at least 100 localizations, and the spatial distribution of localizations in the cluster must have two dimensions
as least twice as large as the third dimension. Synapses were then defined as the homer cluster, as well as all neurexin localizations within 500 nm of the edge of the homer cluster.

STORM localizations from synapses were plotted based on their coordinates to create STORM images of individual synapses, which were rotated such that the trans-synaptic axis aligned along the x-axis, with the synaptic protein signal, e.g. homer, towards the negative side. By defining the origin as the centroid of the homer signal, both trans-axial (defined as distance along x-axis) and radial distances (defined as $R = \sqrt{y^2 + z^2}$) could be calculated. To determine the normalized radial distance $R_{Norm}$, a radius was determined for each synapse, $R_{95}$, which defined a ring encompassing 95% of homer localizations. $R_{Norm}$ could then be calculated as a simple ratio of the 2 values: $R_{Norm} = R/R_{95}$. Neurexin and Homer levels were calculated from cross-talk-subtracted localization counts for each synapse. The number of neurexin clusters per synapse was determined by applying an implementation of DBSCAN to identify and count the number of neurexin clusters in each synapses. HA antibodies per cluster was estimated by first determining the average number of localizations per isolated antibody, $L_{AB}$, and then dividing the number of localizations per neurexin cluster, $L_{NxN}$, by the value of $L_{AB}$. Neurexin and Homer areas were calculated via a coarse convex hull method, where 2D bins of 35 x 35 nm were first defined on the geometric plane parallel to the synaptic face, and centered around the centroid of the homer cluster. The cluster area was then defined as the sum of the area of bins containing at least two STORM localizations (to eliminate spurious outliers) after crosstalk subtraction. Two alternative methods of area measurements (i.e. convex hull area and radius of gyration based on the localizations) were also compared, and produced results with similar trends to those obtained from the coarse convex hull quantification. Although the same trends were observed from all three methods, quantitative numbers differed (Figure 2.5K), likely because these methods have
different sensitivities to outliers and holes (convex hull is more sensitive to outliers than coarse convex hull and radius gyration, and convex hull and radius gyration do not detect holes unlike coarse convex hull).

For 3-color imaging experiments, synapses were defined by the area of homer puncta in the conventional image. The degree of cluster volume overlap was calculated for the two neurexin antibody labels using two methods. The first method is based on the coarse convex hull method described above but extended to 3D to determine cluster volume. The overlap fraction is calculated as the number voxels containing STORM localizations in both antibody channels divided by the totally number of voxels that contain STORM localizations regardless of antibody channels. The second method is the convex hull method. The overlap fraction is calculated as the volume of overlapping region covered by the convex hulls of both antibody channels divided by the volume of the entire region covered by two convex hulls.

The STORM resolution was determined by measuring x/y/z localization SD for 100 isolated antibodies adhered to the surface of a coverslip, and multiplying the average SD by 2.35 to obtain FWHM. We obtained FWHM values of 30 nm and 28 nm for x and y dimensions, respectively, and 67 nm for z (axial resolution).

### 2.4.9 Generation of Expression vectors

Lentiviral vectors for expression of Cre and ΔCre (truncated, inactive) recombinase driven by the human synapsin-1 promoter have been described previously (116). For all other experiments using the Lentiviral backbone with a human synapsin-1 vector (FSW), an empty vector was used as a control. ADAM10 shRNA expression constructs were in the pLKO.1 lentivirus expression vector (Sigma-Aldrich), including ADAM10 shRNA 1 (TRC Number: 65
TRCN0000031844), ADAM10 shRNA 2 (TRC Number: TRCN0000031847), and ADAM10 shRNA 3 (TRC Number: TRCN0000031848). To generate ADAM10 expression constructs, the mouse ADAM10 ORF was purchased from GE Dharmacon (OMM5895-202525673). In-Fusion cloning (Clonetech) was used to clone wild-type ADAM10 and mutants into FSW (according to Manufacturer’s instructions). A catalytically-inactive version of ADAM10 was generated by mutating Glu → Ala at position 384 (118). Another dominant-negative version of ADAM10 was generated by deleting the prodomain and metalloproteinase domain as previously reported (119). A Flag epitope tag (DYKDDDDK) was included at the 3’ end of all ADAM10 constructs. Due to toxicity of the ADAM10 constructs when expressed in normal copy number E. coli, we used Copy Cutter (Lucigen) low-copy number, inducible E. coli for successful cloning and generation of ADAM10 plasmids.

2.4.10 Immunoprecipitation

To remove debris, media from neurons was centrifuged at 3,000 x RPM for 5 minutes. Media was carefully transferred to a new tube and a final concentration of 0.5X cOmplete ULTRA protease inhibitor cocktail (Sigma-Aldrich) and 2 mM EDTA was added. Then 30 µl of pre-washed 1:1 HA agarose beads (Sigma-Aldrich) was added and samples were rotated overnight at 4 °C. Beads were washed 1-2 times with ice-cold DPBS before elution with 2X Laemmle’s buffer containing fresh DTT (100 mM) at 60 °C for 10 minutes.

2.4.11 Cell lysis

Primary neurons plated in 24-well plates were isolated with 50-80 µl of cRIPA per well with 350 µl modified complete RIPA lysis buffer (cRIPA) containing 150 mM NaCl, 5 mM EDTA, 1% Triton X-100, 0.1% SDS, and 25 mM Tris-HCl (pH 7.6). cRIPA also contained freshly added 1X cOmplete ULTRA protease inhibitor cocktail (Sigma-Aldrich). Lysates were
incubated on ice for 20 minutes and then clarified by centrifugation for 20 minutes at 13,000 x
RPM at 4 °C. The same volume of lysis buffer was always used for wells being compared within
a given experiment. Lysates were stored at -80 °C until further processing.

2.4.12 Immunoblotting

Proteins were separated by SDS-PAGE using 4-20% MIDI Criterion TGX precast gels
(Bio-Rad). Proteins were transferred onto nitrocellulose membranes for 10 minutes at 2.5 V
using the Trans-blot turbo transfer system (Bio-Rad). Membranes were blocked in 5% milk
diluted in PBS or TBS for 1 hour at RT. Membranes were then incubated with primary
antibodies diluted in TBST (containing 0.1% Tween-20) overnight at 4C. The following
antibodies were used: HA monoclonal antibody (1:1000; BioLegend, anti-HA.11), CASK
(1:1000, Neuromab, 75-000), PSD-95 (1:500, Neuromab, 73-028), synapsin (1:500, TCS, E028),
CamKII (1:1000, Millipore, MAB8699), synaptophysin (1:1000, Millipore, MAB5258), Mint1
(1:500, TCS, P730), SNAP-25 (1:500, TCS, P913), Nlgn1 (1:1000, Synaptic Systems, 4C12),
GluA1 (1:1000, Millipore, Ab1504), Syntaxin-1 (1:500, TCS, 438B), Tau (1:1000, Millipore,
MAB361), ApoE (1:1000, ThermoFisher, #701241), HSC70 (1:1000, TCS 903A), Grp78
(1:1000, Abcam), alpha-synuclein (1:1000, TCS, Q698), pan-Neurexin (1:500, A473, Sudhof),
SynCAM4 (1:500, Neuromab, 75-247), GDI (1:2000, Synaptic Systems, 130001), CASPR2
(1:500, Neuromab, 73-075), actin (1:2000, Sigma-Aldrich, A1978), and Munc18 (1:1000, BD,
610337). Actin was used as a loading control for protein quantifications.

Combinations of the following IRDye secondary antibodies were used (1:10,000 in TBST
with 5% milk). IRDye 800CW donkey anti mouse (926-32212), IRDye 680LT donkey anti
mouse (926-68022), IRDye 800CW donkey anti rabbit (926-32213), IRDye 680LT donkey anti
rabbit (926-68023), from LI-COR. Detection of the signal was obtained by Odyssey CLx
imaging systems (LI-COR). Pseudo colors were applied to the signals and quantification was performed using Image Studio 5.2 free software.

2.4.13 Studying Processing in Neurons

At DIV10, the media of hippocampal neurons or cortical neurons from HA-Nrxn1 cKI cultures was swapped with media from age-matched wild-type cultures supplemented with DMSO (0.1%), GM6001 (10 µM), TAPI-2 (10 µM), GW280264X (10 µM), GI254023X (20 µM), SB-3CT (30 µM), or C3 (1 µM). At 48 hours post-swap, media was clarified by centrifugation at 3500 RPM for 5 minutes, supplemented with protease inhibitors, and incubated overnight with 30 µl of 1:1 HA-agarose beads. Cells were lysed using RIPA containing protease inhibitor cocktail. Following 2 washes with cold DPBS, IP’s were eluted with 50 µl of 2X Laemmli buffer containing DTT.

For ADAM knockdown and overexpression experiments, the media of HA-Nrxn1 cKI hippocampal neurons (DIV10) was replaced with media from age-matched wild-type cultures. Media and cells were harvested on DIV12 and processed similar to other experiments.

For analysis of activity-dependent changes in processing, wild-type and HA-Nrxn1 cKI hippocampal neurons were prepared. On DIV10, media was swapped from wild-type neurons onto Nrxn1 cKI neurons and included 0.1% DMSO, picrotoxin (50 µM), or tetrodotoxin (1 µM). Media and cell lysate was harvested 48 hours later.

2.4.14 Fractionation

To maximize protein isolation, the average mass of dissected brain tissue (i.e. hippocampus, cortex, cerebellum, olfactory bulb, and brainstem) was pre-determined for 3, male HA-Nrxn1 cKI mice (P56). Based on these masses, the volume used for tissue homogenization
was at 40 mg/ml of buffer. For fractionation, 6 male HA-Nrxn1 cKI mice were euthanized using isoflurane and fresh brain tissue was dissected on ice including the hippocampus, cortex, cerebellum, olfactory bulb, and brainstem. Until homogenization, tissue was stored in cold homogenization buffer containing 1X cOmplete ULTRA protease inhibitor cocktail (Sigma-Aldrich) and 2 mM EDTA. Tissue from one hemisphere was dounce homogenized in RIPA containing protease inhibitor cocktail (complete RIPA, cRIPA). This tissue was incubated on ice for at least 20 minutes and then clarified at 14,000 x RPM for 20 minutes at 4 °C. The supernatant was saved as the total particular fraction (H).

Tissue from the other hemisphere was dounce homogenized in the same volume of ice-cold ACSF (containing in mM: 126 NaCl, 2.5 KCl, 1 NaH2PO4, 26.2 NaHCO3, 2.5 CaCl2, 1.3 MgSO4-7 H2O, 11 D-Glucose, ~290 mOsm.) with protease inhibitors (including 2 mM EDTA) for 10 strokes. Lysates were initially cleared at 13,000 x RPM for 20 minutes (spin 1). The crude supernatant was saved and pellet was re-suspended in equal volume as starting material of cRIPA and incubated on ice for at least 20 minutes. Lysates were clarified at 13,000 x RPM for 20 minutes (spin 2) and the supernatant was saved as total particular fraction (P). The supernatant from spin 1 was further subjected to clarification for 2 hours at 165,000 x g (spin 3). The supernatant from spin 3 was saved as total soluble fraction (S).

For comparing the amount of Nrxn1α and Nlgn1 in particulate and soluble fractions, a BCA assay was performed (ThermoFisher) according to manufacturer’s instructions. The particulate fractions were adjusted to 2 µg/µl and soluble fractions from paired samples were adjusted with similar volumes. Proteins detected in the S fraction were much less abundant than H/P fractions, and were therefore loaded at 10 (Figure 2.13C) or 20 times (Figure 2.13D-E) more than their corresponding H or P fractions. Therefore, final measurements for Nrxn1α and Nlgn1
in the soluble fraction were divided by 20 to estimate the percentage of soluble protein relative to particular protein.

2.4.15 Quantification and Statistical Analysis

Quantification has been described in the respective method details sections and statistical details are provided in the figure legends. Statistical significance between various conditions was assessed by determining P values (95% confidence interval). For animal survival analysis and quantitative immunoblotting experiments, statistical analysis was performed using GraphPad Prism 6 software. Blot intensity quantification was performed using Image Studio Lite. Graphs depict mean ± SEM. Statistical analyses for STORM images, including rank-sum tests for radial displacement and 2-sample T-tests for other quantifications, were performed in MATLAB. Values were plotted as mean ± SEM.
Chapter 3

High-throughput single-cell gene expression profiling with multiplexed error-robust fluorescence in situ hybridization (MERFISH)

3.1 Introduction

Image-based approaches to single-cell transcriptomics, in which RNA species are identified and counted in situ via imaging, have emerged as a powerful complement to single-cell methods based on RNA sequencing of dissociated cells. These image-based approaches naturally preserve the native spatial context of RNAs within a cell and the organization of cells within tissue, which are important for addressing many biological questions. However, the throughput of these image-based approaches, in terms of the number of cells measured per experiment, was relatively low. Here we report advances that leads to a drastic increase in the measurement throughput of multiplexed error-robust fluorescence in situ hybridization (MERFISH), an image-based approach to single-cell transcriptomics. In MERFISH, RNAs are identified via a combinatorial labeling approach that encodes RNA species with error-robust barcodes followed by sequential rounds of single-molecule fluorescence in situ hybridization (smFISH) to read out these barcodes. Here, we increase the throughput of MERFISH by two orders of magnitude through a combination of improvements, including using chemical cleavage, as oppose to photobleaching, to remove fluorescent signals between consecutive rounds of smFISH imaging, increasing the imaging field of view, and utilizing multi-color imaging. With these improvements, we imaged 130 RNA species in more than 100,000 human cells, with as
many as 40,000 cells measured in a single 18-hour measurement. This throughput should substantially extend the range of biological questions that can be addressed by MERFISH.

Single-cell transcriptomics, powered by next-generation RNA-sequencing, has transformed many aspects of cellular and tissue-scale biology (120–122). This new capability has allowed researchers to address exciting questions ranging from the response of single immune cells to antigen (123–125) to the number of transcriptionally distinct cell types and the cellular heterogeneity within complex tissues (58, 126–131). Recent advances in the automated handling of individual cells and the sequencing library preparation for these cells have substantially increased the number of cells that can be routinely characterized with these approaches — notably, state-of-the-art droplet-based RNA sequencing approaches provide the ability to quantify the transcriptome of tens of thousands or more cells (57, 132). This throughput allows rare populations of cells to be characterized and transcriptionally distinct cell types within sizable tissue blocks to be mapped.

However, in most approaches to single-cell transcriptomics, cells are dissociated from tissues and RNAs are extracted from cells, and, as a result, the native spatial context of these RNAs is lost. However, this spatial information is important for a complete understanding of many biological behaviors (133). For example, the spatial organization of individual cell types within most tissues is crucial to how tissue function or dysfunction arises from the behavior of individual cells. Likewise, the intracellular spatial organization of RNAs is a powerful form of post-transcriptional regulation; thus, it is often important to know not only how many copies of an RNA are present within a cell but also where they are located within that cell (134). Addressing questions such as these requires spatially resolved approaches to single-cell transcriptomics (133).
Figure 3.1: **Approaches to improve the measurement throughput of MERFISH.** A. Simplified schematic of a MERFISH readout protocol. Target RNAs are stained with encoding probes that contain a barcode comprised of a combination of readout sequences unique to each RNA species. The barcode is then identified through successive rounds of smFISH, each with a readout probe complementary to one readout sequence. A registered stack of smFISH images for each sample produces an ensemble of fluorescence spots with on/off patterns that define binary barcodes (‘1’ represents fluorescent signal on and ‘0’ represents fluorescent signal off), which allow individual RNA species to be identified. A more detailed hybridization and imaging procedure is shown in Figure 3.2. B. The time required to perform a MERFISH experiment for a given sample area for the published protocol (63, 135) which utilizes photobleaching to remove smFISH signal (red), a modified protocol without photobleaching (purple), a modified protocol without photobleaching and a larger field of view (green), and a modified protocol without photobleaching, a large field of view, and two-color imaging (blue).

Recently we introduced an image-based approach to spatially resolved, single-cell transcriptomics — multiplexed error-robust fluorescence in situ hybridization (MERFISH) (63). In this approach, RNAs are identified via single-molecule fluorescence in situ hybridization (smFISH) (59, 60), as opposed to alternative in situ methods utilizing sequencing (136, 137). MERFISH uses error-robust barcoding schemes to encode RNA species and reads out these barcodes with sequential rounds of smFISH measurements (Figure 3.1A). In our previous implementation of MERFISH (63), RNAs inside cells are first hybridized with complex sets of oligonucleotide probes, termed encoding probes (Figure 3.2). Each encoding probe contains a targeting sequence that directs its binding to a given cellular RNA and multiple readout sequences. The collection of readout sequences associated with a cellular RNA forms a barcode that is unique to that RNA species. These barcodes are then read out through a series of smFISH
measurements; in each round, the sample is stained with a readout probe complementary to one of the readout sequences, the sample is imaged, and the fluorescence signal is extinguished via photobleaching. This process is then repeated with a different readout probe, and the specific on/off pattern of fluorescence observed across multiple smFISH rounds defines the binary barcode used to identify each RNA. We use error-robust barcodes, which allow measurement errors to be identified and, in some cases, corrected, to ensure high-accuracy MERFISH measurements (63). Using this approach we have previously demonstrated the ability to image 140 RNA species with an 80% detection efficiency using 16 rounds of smFISH imaging with an encoding scheme capable of detecting and correcting errors and to image 1,000 RNA species with a 30% detection efficiency with an encoding scheme capable of detecting but not correcting errors (63). In both cases, we were able to quantify the copy number and spatial distribution of these RNAs within ~100 human fibroblast cells in a single ~18-hour measurement. However, for many biological questions, such as the study of rare populations of cells or the survey of sizeable volumes of tissues, it is highly desirable to increase the throughput of MERFISH such that many more cells can be measured.
Figure 3.2: **Diagram of the hybridization and imaging procedure with encoding and readout probes.** Encoding probes are first hybridized to each cellular RNA. Each encoding probe contains a 30-nt-long target region (black) that binds to the target RNA and three 20-nt-long readout sequences (purple, green, blue, or orange). The specific choice of readout sequences for a given RNA determines the barcode that will be used to identify it. During each readout hybridization, one readout probe complementary to a given readout sequence (depicted in orange for the first hybridization round) conjugated to a dye (red circle) is hybridized to the sample. The sample is imaged, and the fluorescence signal is eliminated (as indicated by the gray circles). This process is repeated, with a different readout probe hybridized in each of the $N$ rounds of readout hybridization.

Here, we present an improved MERFISH method that drastically increases the throughput of this technique, simplifies several aspects of this protocol, and increases the measurement accuracy. With these improvements, we demonstrated the ability to perform spatially resolved gene expression profiling of ~40,000 cultured human osteosarcoma (U-2 OS) cells in a single 18-hour experiment. As a simple illustration of the benefits of this increased throughput, we characterized 130 genes in ~100,000 cells, identified a sub-population of cells
undergoing DNA replication or cell division, and characterized both the expression profile and spatial distribution of cells that comprised this subpopulation.

3.2 Results

3.2.1 Increasing the throughput of MERFISH measurements

The total time required for a MERFISH measurement can be divided into an area-dependent time that scales with the total imaged area and an area-independent time that does not. The area-dependent time includes the time required to position, focus, and image each field-of-view (FOV). In addition, because of the high illumination intensity required to photobleach the fluorescence signals between consecutive rounds of smFISH, each FOV must be photobleached individually; thus, this time is also a part of this area-dependent time. The area-independent time includes buffer exchange times and incubation times required for sample staining, thus, scales with the number of rounds of smFISH that must be performed. Figure 3.1B illustrates the scaling of the duration of a MERFISH measurement with the imaged area (red line). For 16 rounds of hybridization and imaging, the total area-independent time amounts to several hours; however, this area-independent time is dominated by the area-dependent time when the imaged sample area is larger than ~1 mm$^2$.

To improve the throughput of MERFISH, we first sought to decrease the area-dependent time. In our previously published MERFISH protocols (63, 135), imaging a FOV of $\sim40 \mu m \times 40 \mu m$ required only 0.1 s, yet photobleaching of this same FOV required a significantly longer exposure, ~3 s. Thus, we devised a scheme in which the smFISH signal from the entire sample could be extinguished simultaneously by chemical reaction instead of photobleaching. Specifically, we reasoned that fluorescent dyes conjugated to readout probes via a disulfide
linkage could be rapidly cleaved from these probes with a mild reducing agent such as Tris(2-carboxyethyl)phosphine (TCEP) (Figure 3.3A).

Figure 3.3: Reductive cleavage of disulfide-linked fluorophores efficiently removes fluorescent signal. A. Schematic diagram of the use of Tris(2-carboxyethyl)phosphine (TCEP) to extinguish fluorescence signal via cleavage of a disulfide bond linking a fluorescent dye to a readout probe. B. Images of a region of a human fibroblast (IMR-90) stained with an encoding probe for the FLNA RNA and a readout probe linked to Cy5 via a disulfide bond as a function of time exposed to 50 mM TCEP. Each panel represents the same portion of a field of view. Scale bars: 2 µm. The contrast of all but the upper left panel has been increased 5-fold to better illustrate the remaining fluorescent signal in the sample after TCEP treatment. C. The average brightness of readout probe 1 bound to encoding probes targeting FLNA (normalized to the brightness prior to TCEP exposure) as a function of total exposed time to 50 mM TCEP. Error bars represent SEM (N provided in Figure 3.4B), and the blue region represents the 95% confidence range for a fit to an exponential decay. D. The measured half-life for the average brightness when exposed to 50 mM TCEP for four readout probes (1-4) each with a different sequence and linked to one of two different dyes, Cy5 (green) and Alexa750 (red). Error bars represent the 95% confidence interval for the fit to an exponential decay shown in (C) for readout probe 1 and Figure 3.4A for readout probes 2-4.

To test this approach, we hybridized encoding probes containing readout sequences to the FLNA mRNA in human IMR-90 cells and then stained this sample with a readout probe that was conjugated to a Cy5 dye via a disulfide bond. As expected, the sample exhibited bright fluorescent spots representing individual molecules of the FLNA mRNA, and these fluorescent
spots reduced in brightness and eventually disappeared upon treatment with 50 mM TCEP (Figure 3.3B). When averaged across thousands of RNAs, the brightness of these spots decayed exponentially (Figure 3.3C) with a half-life of 1.17 ± 0.07 min (95% confidence interval). This half-life did not depend on the sequence of the readout probe or the dye to which it was conjugated (Figure 3.3D). After ~15 minutes of TCEP treatment, the average brightness of each RNA spot and the number of detected RNA spots were reduced by 10⁵-fold and 10⁴-fold, respectively (Figure 3.3C, and Figure 3.4A-B). Furthermore, the TCEP treatment did not inhibit the ability of the next round of readout probes to bind to the sample (Figure 3.4C). Our calculation shows that the use of this chemical approach to remove fluorescence signals between successive rounds of smFISH should substantially reduce measurement time and increase throughput (Figure 3.1B, purple line).

Figure 3.4: TCEP cleavage efficiently extinguishes fluorescence signal from readout probes for different readout sequences and fluorophores. A. The average brightness of all smFISH spots observed for labeled FLNA mRNAs in human fibroblast cells (IMR-90) as a function of the total exposure time to cleavage buffer (50 mM TCEP in 2xSSC) for four different readout sequences (blue, green, cyan, and red) and two different fluorophores (Cy5 was conjugated to readouts 1 and 4 and Alexa750 to readouts 2 and 3). The readout sequences are provided in Appendix 1. The brightness values are normalized to the values observed before TCEP treatment (time zero). B. The fraction of smFISH spots that have brightness greater than half of the brightness determined for a single dye (either Cy5 or Alexa750) as a function of the total exposure time to TCEP cleavage buffer. The colors indicate the same readout and dye
combinations depicted in (A). C. Representative images of the FLNA mRNA stained with a readout probe corresponding to the first bit (top), treated with TCEP cleavage buffer (middle), and restained with a readout probe corresponding to the second bit (bottom). The error bars in (A) represent SEM based on the number of RNA spots observed at each time point. The number of RNA spots observed prior to TCEP treatment (time zero) were 19,696, 17,644, 20,156, 17,415 for readout probes 1, 2, 3, and 4, respectively. The numbers of spots determined at all other time points are specified by the survival fraction in (B). Missing data points indicate times at which no spots were visible in the sample. Scale bars: 2 µm.

Next, we reasoned that, without the requirement for high illumination intensities for efficient photobleaching, it should be possible to further decrease the area-dependent time by expanding the size of the imaging FOV. To explore this idea, we designed and constructed a microscope that utilizes a 2048-pixel × 2048-pixel, scientific CMOS (sCMOS) camera in combination with a high numerical aperture (NA = 1.3), high magnification (60X), silicone oil objective (Materials and Methods). We utilized a silicone oil objective because we found that it had less field curvature than comparable oil immersion 60X objectives. With this optical configuration, we could image a FOV of 223 µm × 223 µm, which is ~25-fold larger than our previously reported FOV, with an exposure time of 0.5 s. This increase in the size of the FOV should further increase imaging speed and hence measurement throughput substantially (Figure 3.1B, green line).

As a third step to reduce measurement time, we utilized multi-color imaging. Specifically, we simultaneously stained the sample with two readout probes per hybridization round, each probe conjugated to one of two spectrally distinct dyes, and used two-color imaging to reduce the number of imaging rounds in half, thereby cutting the area-independent time required to stain, wash, and extinguish signals accordingly (Figure 3.1B, blue line). We used Cy5 and Alexa750 dyes because of the low cellular autofluorescence observed in the red and near
infra red spectral ranges. In total, the use of reductive cleavage to extinguish fluorescence signal between successive imaging rounds in combination with the increase in the area of the FOV and the use of two-color imaging should dramatically reduce both the time required to perform MERFISH for a given area and increase the measurable area size for a given time (Figure 3.1B, blue line versus red line).

3.2.2 Improving the robustness of MERFISH measurements

We also made a series of protocol changes aimed at simplifying measurement procedures and improving the robustness of the measurement. First, we found that readout probes can bind to encoding probes with similar rates at room temperature to those observed at 37 °C (Figure 3.5A). Room temperature hybridization avoids any variation in measurement results associated with non-uniform sample heating. Second, we shortened readout probes from 30 to 20 nt, which allowed us to include more readout sequences on each encoding probe without increasing the total length of the encoding probe. This modification allows us to either increase the brightness of signals from single mRNA molecules by preserving the number of encoding probes per RNA or to achieve the same signal brightness with fewer encoding probes per RNA, allowing shorter RNAs to be targeted. Third, we created readout probes that bind to readout sequences with comparable rates to our previous probes but at 10-fold lower concentrations. Specifically, we exploited the published observation (138) that oligonucleotide sequences that contain only three of the four nucleotides have significantly less secondary structure than sequences that use all four nucleotides and, thus, have faster hybridization rates (Figure 3.5B). Fourth, we replaced the toxic RNA denaturing agent formamide used in the readout hybridization and wash buffers with nontoxic ethylene carbonate (139), which we found also moderately increased the rate of readout hybridization (Figure 3.5C).
Figure 3.5: Characterization of the hybridization properties of different readout probes and different hybridization conditions. A. The average normalized smFISH spot brightness for FLNA molecules labeled first with encoding probes then with readout probes versus the total time the sample is exposed to 10 nM readout probes at 37 °C (green crosses) or at room temperature (25 °C; purple stars). The sequence of the readout probe is CGCAACGCTTGGGACGGTTCCAATCGGATC, which is one of our previously published readout probe sequences. The hybridization buffer is our previously published, formamide-based hybridization buffer (63, 135). B. The average normalized smFISH spot brightness as in (A) but when the sample is stained with either 10 nM of a previously published 30-nt 4-letter readout probe (purple stars; reproduced from (A)), 10 nM of a new 20-nt 3-letter readout probe which does not contain G (red circles), 1 nM of the previously published 30-nt 4-letter readout probe (orange circles), or 1 nM of the new 20-nt 3-letter readout probe (blue crosses). The sequence of the new readout probe is ATCCTCCTTTCAATACATCCC. Hybridization was conducted at room temperature in the formamide-based buffer. C. The average normalized smFISH spot brightness as in (A) for 1 nM of the new 20-nt 3-letter readout probe hybridized at room temperature but using different buffers: a hybridization buffer containing 10% formamide as described previously (63, 135) (blue crosses, reproduced from (B)), a hybridization buffer in which formamide was replaced with 1% v/v ethylene carbonate (red stars), or a hybridization buffer with 10% v/v ethylene carbonate (green circles). D. The coefficient of variation (the standard deviation divided by the mean) for the average brightness of smFISH spots across all rounds of imaging in the 16-bit MERFISH experiment conducted with the previously published 30-nt readout probes and formamide-based hybridization protocol (63, 135) (old protocols) as well as with the new readout protocols published here (20-nt 3-letter readout sequence and an ethylene-carbonate-based hybridization protocol). Error bars in (A)-(C) represent SEM across all measured RNA spots, and more than 10,000 RNA spots were measured for each data point.
We also found that these modified readout probes and readout hybridization protocols reduced the variance in staining quality between different rounds of readout hybridization as compared to our previous protocols (Figure 3.5D). Of the multiple changes made above, the new readout sequences likely account for the majority of this improvement since we have previously observed that some of the variability across different readout staining rounds (Figure 3.5D) can be attributed to sequence variations, presumably, due to unanticipated secondary structures. By design, such secondary structures should be far less likely with the new readout sequences that utilize only three of the four nucleotides (138). We anticipate these improvements to increase the accuracy of our MERFISH measurements because lower-quality (or varying-quality) readout hybridizations can result in dim fluorescence signals in some imaging rounds and increase the rate at which read out errors are made.

3.2.3 A new image analysis algorithm to handle high-throughput MERFISH data

In parallel, we anticipated that our previous computational methods for MERFISH data analysis (63, 135), which typically required several hours to a day to analyze a single MERFISH data set, would not be adequate for analyzing the two orders of magnitude higher data volume generated per experiment. Thus, we developed a new analysis pipeline capable of handling this drastic increase in imaging throughput (Materials and Methods). The major advance in this pipeline is the adoption of a pixel-based decoding approach, as opposed to a spot-finding approach, to reduce computation time. Briefly, images of the same FOV from different imaging rounds are registered using images of fiducial beads collected in each round. These images are high-pass filtered to remove background and deconvolved to sharpen and better resolve closely positioned spots. Previously we observed that signals from the same RNA often varied in position from round to round by ~100 nm (63). Thus, to better connect signals from one round to
another, we applied a low-pass filter with a kernel size of 100-nm radius. The intensities of each pixel across all 16 rounds of images were then used to form a 16-dimensional vector, which we normalized to unit amplitude. This vector was then compared to the set of unit vectors defined by all valid barcodes. The pixel was assigned to a given barcode if the Euclidean distance between its normalized intensity vector and the closest barcode vector was less than the distance defined by a single-bit error. Contiguous sets of pixels that matched to the same barcode were combined to form a single detected RNA. Background pixels mistakenly matched to a barcode were identified and removed based on their low brightness and small number of contiguous pixels matched to the same barcode (Figure 3.6). With this new pipeline, analysis of large MERFISH data sets (~40 mm$^2$ with ~40,000 human cells) can be completed in 2-3 days utilizing multiple cores on a computer cluster.

**Figure 3.6:** **Thresholding of RNA signals based on area and brightness.**

A. Histogram of the log$_{10}$ brightness for all observed single-RNA-molecule signals from the data presented in Figure 3.7. The gray dashed line defines the brightness threshold used to discard dim single-molecule signals that are likely due to background rather than real RNA signals. B. Scatter plot of the observed log$_{10}$ brightness for single-molecule signals with a given area (gray markers), i.e. the number of contiguous pixels assigned to the same RNA molecule, with the associated probability distributions (cyan). Only 1,000 randomly selected single-molecule signals are plotted for each area for clarity. Note that single-molecule signals with smaller areas also tend to be low brightness. The gray dashed lines represent the cuts applied to separate spurious background signals from foreground RNA signals, i.e. a brightness greater than $10^{0.75}$ and an area of 4 pixels or larger.

### 3.2.4 High-throughput MERFISH measurements of tens of thousands of cells
To demonstrate the substantial increase in imaging throughput made possible by the above advances, we measured 130 RNAs in cultured human osteosarcoma (U-2 OS) cells with a previously published 16-bit, modified Hamming-Distance-4 (MHD4) encoding scheme (63). In this encoding scheme, all utilized barcodes are separated at least by a Hamming Distance of 4, and hence at least 4 bits must be read incorrectly to change one valid barcode into another. Therefore, every single-bit error produces a barcode uniquely close to a single valid barcode, allowing such errors to be detected and corrected. Two-bit errors can also be detected but are not correctable since the resulting barcode is no longer uniquely close to a single valid barcode. To further account for the fact that it is more likely to miss a hybridization event (1-to-0 error) than to misidentify a background spot as an RNA (0-to-1 error) in smFISH measurements, our MHD4 code contains a constant and relatively low number (four) of ‘1’ bits. This 16-bit MHD4 encoding scheme includes 140 distinct barcodes in total (63). We assigned 130 of these barcodes to different RNA species, leaving 10 barcodes unused to serve as ‘blanks’ (not corresponding to any RNA) for misidentification controls.

Figure 3.7A illustrates one such measurement over an area of 3.2 mm × 6.2 mm. The cells were fixed, permeabilized, and labeled with encoding probes to 130 RNA species. We then performed 8 rounds of hybridization, imaging, and TCEP cleavage with 16 different readout probes; each round of imaging employed two readout probes conjugated to Cy5 and Alexa750, respectively. Single-molecule spots were clearly observed across the entire imaged area in both Cy5 and Alexa750 channels in each round of smFISH staining and imaging (Figure 3.7B-C). The identities of individual RNA molecules were then decoded via the algorithm discussed above (Figure 3.7D). To assign RNAs to individual cells, we utilized DAPI to identify cell nuclei and the local density of RNAs to define cellular boundaries (Materials and Methods). In total, Figure
3.7 contains 15,181 cells. Among these, 12,607 segmented cells satisfied our conservative criteria on cell morphology designed to eliminate segmentation errors (Materials and Methods), and these properly segmented cells contained 9.7 million identified RNA molecules.

**Figure 3.7:** A MERFISH measurement of a ~20 mm² sample area (~15,000 cells). **A.** Mosaic image of a 3.2-mm × 6.2-mm region of cultured U-2 OS cells stained with DAPI (purple), encoding probes for 130 RNAs, and a Cy5-labeled readout probe (green). Scale bar: 1 mm. **B.** Image of the Cy5 channel in the first round of readout hybridization for the small portion of the field in (A) marked by the gray square. Scale bar: 20 µm. **C.** Two color images of the smFISH stains for all 8 rounds of hybridization and imaging, for the small portion of the field in (B) marked by the gray square after the application of a high-pass filter to remove background, deconvolution to tighten spots, and a low-pass filter to better connect spots in different images (Materials and Methods). Green, red, and orange represent the Cy5 channel, the Alexa750 channel, and the overlay between the two, respectively. Scale bars: 500 nm. **D.** The decoded barcodes for the region shown in (B). Spots represent individual molecules color-coded based on their RNA species identities (barcodes). Both the nuclear boundaries and the boundaries used to assign RNAs to individual cells are depicted (gray). Scale bar: 20 µm. Inset: An image of the barcode assignment (indicated by color) for each pixel in the images shown in (C). Scale bar: 500 nm.
Figure 3.8: Additional metrics to evaluate the performance of MERFISH measurements. A. The total number of RNAs decoded without (Exact) and with (Corrected) error correction. B. The confidence ratio for all barcodes representing real RNAs (blue) and the blank controls (red) sorted from largest to smallest value. The confidence ratio for any given gene (or barcode) is defined as the ratio between the number of exact matches to this barcode and the total number of exact matches plus matches with single-bit errors to this barcode. 123 of the 130 barcodes encoding real RNAs have a confidence ratio larger than that of the largest confidence ratio of the blank barcodes. C. The error rate—the fraction of measured barcodes that contains a given bit flip—for each bit. Both 1-to-0 error rates and (blue) and 0-to-1 error rates (red) are show for each bit. The data presented in this figure represent the error properties of the data set presented in Figure 3.7 and are representative of that observed for all other data sets.

To determine the RNA decoding quality, we considered two types of errors for each RNA species. First, some RNAs can be misidentified as the wrong species, leading to a non-zero misidentification rate. Second, some RNAs can be missed, leading to a non-100% calling rate. To assess these errors, we first examined the fraction of decoded RNAs that required error correction (Figure 3.8A). In our previous published MERFISH experiments using the same 16-bit MHD4 code, we observed that ~60% of all decoded RNAs required error correction (63). By contrast, with the new protocols described here, only ~20% of RNAs required correction. Lower levels of error correction would suggest a lower level of misidentification and a higher calling rate. To test the level of misidentification, we examined the number of times that the blank barcodes were counted. Indeed, these barcodes were counted relatively infrequently with 120 of the 130 (92%) RNA species counted more frequently than the most abundant blank barcode
In addition, we used an alternative metric, the confidence ratio, to further assess the misidentification rate. As previously defined (63), the confidence ratio for each measured barcode was determined as the number of RNA molecules exactly matching this barcode over the total number of exact matches and matches with single-bit errors for this barcode. We have previously shown that blank barcodes tend to have lower values of the confidence ratio relative to RNA-encoding barcodes (63). Indeed, here we found that 95% of the 130 RNA species had a confidence ratio higher than the maximum confidence ratio observed for the blank barcodes (Figure 3.8B). Next, to examine the calling rate of these measurements, we first used the frequency with which errors were corrected at each bit to determine the average per-bit error rate, as described previously (63). Previously we observed an average 1-to-0 error rate of ~10% and an average 0-to-1 error rate of ~4% (63). By contrast, our new MERFISH protocol produced substantially lower per-bit error rates: a 1-to-0 error rate of ~1% and a 0-to-1 error rate of ~0.5% (Figure 3.8C). With these per-bit error rates we would predict a very high calling rate of ~99%.

To assess the calling rate experimentally, we determined the copy numbers of 10 different RNAs using conventional smFISH and compared them with our MERFISH results. We found that the average copy number per cell for these 10 RNAs determined with MERFISH correlated strongly with the values determined via smFISH (Figure 3.9B). Moreover, the average ratio of copy numbers between the MERFISH and smFISH measurements was 0.94 ± 0.06 (SEM; n =10), consistent with the high calling rate estimated from our observed per-bit error rates. Together, these metrics indicate a slightly lower misidentification error rate and higher calling rate as compared to those of our previous lower-throughput MERFISH measurements (63).
Figure 3.9: **Performance of the high-throughput MERFISH measurements.** A. The average RNA copy numbers per cell measured in Figure 3.7 and sorted from largest to smallest abundance. Barcodes assigned to real RNAs are marked in blue and those not assigned to RNAs, i.e. blank controls, are marked in red. B. The average RNA copy numbers per cell determined via MERFISH versus that determined via conventional smFISH for 10 of the 130 RNAs. The dashed line represents equality. The average ratio of counts determined by MERFISH to that determined by smFISH indicates a calling rate of 94 ± 6% (SEM, n = 10). Plotted error bars represent SEM across the number of measured cells (> 300 cells) for each gene measured via smFISH. C. The average RNA copy number per cell determined by MERFISH versus the abundance as determined by bulk sequencing (FPKM, fragments per kilobase per million reads). The Pearson correlation coefficient (ρ_{10}) between the log_{10} values is 0.86 with a P-value of 6×10^{-39}.

We further compared the average copy number per cell determined by MERFISH to that determined from published bulk RNA-seq for U-2 OS (140). The values determined by MERFISH correlated with those determined from RNA-seq with a high Pearson correlation coefficient for the logarithmic abundances (ρ_{10} = 0.86) (Figure 3.9C).

Finally, to demonstrate the reproducibility of these high-throughput measurements, we performed MERFISH measurements for a range of confluencies of cells and for two different sample areas, ~20 mm^2 and ~40 mm^2. Figure 3.10 shows that the average RNA copy number determined by each of these measurements correlated strongly with those determined by the measurement presented in Figure 3.7 (ρ_{10} >= 0.95). Across all seven measurements we observed an average calling rate of 90% ± 10% (SEM across 7 replicate measurements) by comparison to smFISH results. In total, we measured 105,966 cells with 87,632 cells segmented. Of these data
sets, the largest contained 39,523 cells (35,873 segmented) in an area of 40 mm² measured in less than 18 hours. This throughput represents a 250-fold increase in the sample area imaged in a single 18-hour measurement relative to that published (63) and, due to the smaller size of U-2 OS cells (used here) as compared to IMR-90 cells (used previously), a nearly 400-fold increase in the number of measured cells.

Figure 3.10: Reproducibility of high-throughput MERFISH measurements. A. The average RNA copy number per cell for a replicate MERFISH data set versus that shown in Figure 3.7. $\rho_{10}$ represents the Pearson correlation coefficient between the log₁₀ copy numbers. B-F. As in (A) but for five additional MERFISH data sets. The strong correlation between these values shows the high reproducibility of MERFISH measurements. The number of segmented cells and the total imaged area are also listed for datasets 2-7. The number of segmented cells and the total imaged area for dataset 1 is described in the main text. The P-values for all Pearson correlation coefficients are less than $1 \times 10^{-71}$.

3.2.5 Characterization of a sub-population of cells

One advantage of the significantly enhanced throughput is the ability to image potentially rare or transient subpopulations of cells with sufficient statistics to characterize the properties of such subpopulations. As a simple illustration of this ability, we identified a subpopulation of cells undergoing DNA replication or cell division in the three datasets collected at the highest
confluency (total 78,815 cells). To identify this subpopulation, we determined the distribution of DAPI signal intensity observed in individual cells (Figure 3.11A). A local minimum in this distribution divided the cells into two groups: Group 1 cells contained lower DAPI levels whereas Group 2 cells contained roughly twice the DAPI signal of Group 1 cells, suggesting that Group 2 contained cells undergoing DNA replication or cell division. Group 2 represented a relatively small population of ~20% of the measured cells; nonetheless, due to the large number of cells measured, this population contained 16,036 cells. To identify how the transcriptional profile of these 130 genes differed between Group 2 and Group 1, we determined for each gene a fractional expression level defined as the copy number of this RNA divided by the total copy number of all 130 RNAs detected in the cell. Figure 3.11B displays the ratio of this fractional expression level between Group 1 and Group 2 cells for each gene, showing that some genes were up-regulated and some were down-regulated in Group 2 cells. The large number of cells measured here allowed us to distinguish even small changes in expression levels with confidence. Figure 3.11C plots the observed distribution of expression levels for both groups for the 10 most up-regulated (top) and 10 most down-regulated (bottom) genes. The most up-regulated genes included CENPF, CKAP5, POLQ, and BUB3, which encode a centromere-binding protein, a spindle-binding protein, a DNA polymerase, and a protein involved in the mitotic checkpoint, respectively, supporting the association of Group 2 with cells undergoing DNA replication or cell division. Interestingly, the expression of these genes, in particular CKAP5 and CENPF, could also be used to identify this sub-population of cells without the DAPI signal information (data not shown). The set of the most down-regulated genes included THBS1, FBN2, and TSPAN3, which encode thrombosin, fibrillin, and tetraspanin, respectively, as well as other genes involved in cell-cell interactions and adhesion. We speculate that the differential
The regulation of these proteins might facilitate the disruption and reformation of cell-cell interactions that must occur during cell division.

**Figure 3.11: Characterizing the expression differences of a sub-population of cells undergoing DNA replication or cell division.**

**A.** Violin plot of the distribution of total DAPI intensity for individual cells. The dashed line defines the intensity threshold (based on a local minimum) used to group cells into two groups: Group 1, low-DAPI-signal cells, and Group 2, high-DAPI-signal cells. Gray dots indicate the values for individual cells and the blue shaded area represent the probability distributions. Only 1,000 randomly selected cells are displayed for clarity. **B.** The log₂ ratio of the mean fractional expression level of each RNA species in Group 2 relative to that of Group 1. The fractional expression level for an RNA species is defined as the copy number of that RNA divided by the total copy number of all 130 RNAs detected in the cell. The mean and SEM are computed across three biologic replicates. Green and red markers indicate genes further examined in (C). **C.** Violin plots of the distribution of expression levels for individual genes within Group 1 (blue) or Group 2 (red) for the 10 genes with the largest magnitude of up-regulation (top; marked green in (B)) or 10 genes with the largest magnitude of down-regulation (bottom; marked red in (B)) in Group 2 relative to Group 1. The solid black lines represent the mean, and the colored curves represent the probability distributions. The gray dots represent the expression level for 1,000 randomly selected cells. **D.** A small region of one data set showing the location of MALAT1 (gray), CENPF (red), and CKAP5 (green). The gray lines represent the boundaries of cells (segmented based on the density profile of all 130 measured RNAs). Note that MALAT1 clearly defines the nucleus. **E.** The Pearson correlation coefficient for the relative expression of CENPF (red) or CKAP5 (green) observed between pairs of cells separated by various distances. The cell-cell separation index is defined to be 1 for any given cell and its nearest neighbor, 2 for any given cell and its second closest neighbor, etc. These correlations were calculated for all cells within each of the three data sets, and then averaged across these data sets.

Finally, to illustrate the power of a spatially resolved measurement, we investigated the spatial distribution of the Group 2 cells. To probe this organization, we examined the copy numbers of CKAP5 and CENPF, the two RNAs most up-regulated in Group 2 cells (Figure...
As expected, we found that the expression levels of these RNAs were highly correlated and varied significantly between cells. Moreover, Figure 3.11D-E reveals that neighboring cells tended to express a similar level of these RNAs. Such spatial correlations could have been caused by a variety of potential mechanisms, including the fact that neighboring cells likely shared a common progenitor, which may have resulted in an apparent synchronization of their cell cycles, or that there may have been local cues that promoted or repressed cell division, such as direct or indirect communication between cells. The ability to directly reveal these cellular-scale spatial organizations is one of the benefits of an image-based approach to single-cell transcriptomics.

3.3 Discussion

Image-based approaches to single-cell RNA profiling, which identify RNAs via multiplexed smFISH (63, 141–145) or in situ sequencing (136, 137), can directly provide the native spatial context of individual RNAs both within cells and within the context of the culture or tissue. Recently we introduced MERFISH, which uses massively multiplexed smFISH to perform spatially resolved RNA profiling of single cells at the transcriptomic scale (63). However, the measurement throughput of these image-based approaches (the number of measured cells) has been relatively limited. Here we developed several advances to the MERFISH method that increased the throughput of this approach by two orders of magnitude: we profiled 130 RNAs across 40 mm² of sample containing as many as 39,000 human cells in just 18 hours. In total, we performed such measurements in ~100,000 cells, generating a dataset comparable in size to those published utilizing droplet-based single-cell sequencing approaches (57, 132). Previously, we have shown that MERFISH can be used to measure ~1,000 RNA species in individual cells using a very similar experimental procedure but different encoding
schemes (63). Thus, we anticipate that this increase in throughput could be applied to the measurement of thousands of RNAs with MERFISH.

This substantial increase in throughput should extend the range of questions that can be addressed via MERFISH. For example, we demonstrated here the ability to identify a subpopulation of cells and to utilize the sizeable number of cells within this subpopulation to quantify the, potentially small, differences in their gene expression profiles with statistical significance. We also envision that the increase in imaging throughput reported here will be instrumental in applying MERFISH to the de novo identification of cell types in sizable volumes of tissues. Finally, we anticipate that with further optimization of the hybridization and dye-cleavage protocols, utilization of more colors per imaging round, and additional improvements in camera, optics, and light sources to further increase the FOV area and reduce the imaging time, it will be possible to further increase the throughput of MERFISH and to characterize millions of individual cells in their native culture and tissue contexts. Given that the MERFISH experimental setup is, at its core, a simple epi-fluorescence microscope with a sensitive camera in combination with an automated fluid handling system composed of commercially available components and controlled by open-source software (63, 135), we anticipate that this technique can be readily adopted by many laboratories.

### 3.4 Materials and Methods

#### 3.4.1 Encoding Probe Design.

Each encoding probe is comprised of the following components: two priming regions, multiple readout sequences, and a target region. To design the target regions, we developed a computational pipeline that requires significantly less computational cost than the previous
methods that we employed (63, 135, 146). The central challenge to target region design was determining the optimal specificity of the probes, e.g. how narrow or broad a range of local GC content or melting temperature (T_M). The specificity should be as high as possible to promote specific binding while maintaining permissive enough conditions to produce ample target regions for all desired genes. Our previous computational approach was inefficient, in part, because basic calculations of specificity properties, e.g. GC content or T_M, were re-performed whenever a new set of stringency conditions were probed. To address this issue, we developed an alternative approach in which we pre-calculated properties of the transcriptome that allow the rapid calculation of specificity parameters, namely, local GC content, T_M, relative abundance of potential off-target binding partners, and relative specificity of a given probe to a given isoform of a gene. This approach reduced the time required to design probes for a given set of stringency conditions from days to minutes.

Specifically, using the human transcriptome from the genome build hg38 (http://useast.ensembl.org/Homo_sapiens/Info/Index), we calculated the local GC content and the local nearest neighbor thermodynamic properties (entropy and enthalpy) using the parameters defined previously by SantaLucia et al. (147). From the pre-determined local nearest neighbor thermodynamic properties, the T_M for any length of probe could be rapidly computed, assuming a monovalent salt concentration of 300 mM (the concentration of NaCl in the encoding and hybridization readout buffers) and a probe concentration of 5 nM. In addition, we created a series of look-up tables that allowed us to rapidly calculate a penalty for off-target binding for each potential target region. First, we created a look-up table for each gene comprised of all of the unique 17-nt-long sequences present in all of the isoforms of that gene with the penalty associated with each sequence defined as the sum of the abundance of each isoform determined
from RNA-seq (see the “RNA-seq” section below) in which that sequence appeared. Multiple appearances of the same 17-nt-long sequence in the same isoform each contributed to this penalty. We termed these tables the isoform penalty tables. Second, we created a look-up table comprising the same penalty terms but for the entire transcriptome. We termed this table the transcriptome penalty table. For both types of penalty tables, we chose 17-nt-long homology sequences to balance the desire to eliminate short regions of homology with the dramatic increase in the frequency of such regions in the transcriptome as this length is decreased. Finally, because some non-coding RNAs (ncRNAs) are far more abundant than coding RNAs and could, thus, contribute more significantly to background, we calculated an additional penalty table corresponding to the number of times all unique 15-nt-long sequences appear in the set of human rRNA and tRNAs as well as the human mitochondrial rRNAs and tRNAs (GRCh38, ncRNA: ftp://ftp.ensembl.org/pub/release-84/fasta/homo_sapiens/ncrna/), and termed this table the ncRNA penalty table. We decreased the homology length for off-target binding to these ncRNAs to increase the stringency of selection against partial homology against these highly abundant RNAs.

Using the isoform and transcriptome penalty tables, we calculated two quantities for each 17-nt-long region in each transcript—an isoform-specificity index and a gene-specificity index. The isoform-specificity index for each 17-nt-long region within every transcript was calculated by dividing the measured abundance of the given isoform (i.e. the abundance of the correct target) by the isoform-specific penalty for that sequence as determined by the isoform penalty table for that gene (i.e. the sum of the abundance of the correct target plus all potential off-targets in other isoforms). This value varies between 0 and 1, and can be roughly thought of as the fraction of probes that contain the given 17-nt sequence that would bind to this given isoform out
of those that could bind to any isoform derived from that gene. This quantity is most likely an underestimate of that fraction since it is unlikely that a probe would bind to a 17-nt-long region of homology with the same affinity as the full length (30-nt) target of the probe. The gene-specificity index for each 17-nt-long sequence in a given transcript was calculated by dividing the penalty associated with this 17-nt-long sequence derived from the specific isoform penalty table for that gene (i.e. the abundance of that 17-nt sequence in all isoforms of that gene) by the penalty for that sequence derived from the transcriptome penalty table (i.e. the abundance of that sequence in all transcripts, which includes all isoforms of the target gene as well as all other transcripts). Again, this quantity varies between 0 and 1 and can be roughly thought of as the fraction of a potential probe containing a given 17-nt-long region that would bind to any of the isoforms of a given gene as opposed to any other member of the transcriptome. The specificity indices for individual target regions, which were typically 30-nt long, were derived by averaging the isoform and gene specificity indices for all 17-nt-long sequences within each potential target region.

Using the GC and thermodynamic annotations in conjunction with these specificity indices, we calculated the GC content, $T_M$, isoform- and gene-specificity indices for all possible 30-nt target regions, as well as the frequency of homology regions in ncRNAs, and then chose a subset of target regions based on the desired ranges for each of these quantities. From these chosen target regions, we identified non-overlapping regions starting with the first valid target region at the 3’ end of each isoform. Computationally, construction of the penalty tables and GC and thermodynamic annotations for the transcriptome was slow, requiring a few hours on a desktop computer running in parallel on multiple cores. However, once these annotations were computed, construction of target regions for a given range of stringencies, e.g. $T_M$, GC,
specificity index ranges, etc., required only ~5-10 minutes. Thus, we were able to screen a wide range of stringency ranges and identify the set of parameters that provided the narrowest, most stringent conditions on the target regions while still producing enough target regions for the desired set of transcripts. For the library reported here, the target regions used were designed with a GC range of 43 to 63%, a $T_M$ range of 66 to 76 °C, an isoform-specificity index range of 75 to 100%, an gene-specificity index range of 75 to 100%, and no regions of homology longer than 15 nt to human rRNAs, tRNAs, or mitochondrial rRNAs and tRNAs (calculated using the ncRNA penalty table). The used target regions are provided as part of the encoding probes in Appendix A. All calculations were performed in MATLAB with custom, open-source functions and scripts (http://zhuang.harvard.edu/merfish/).

20-nt-long, 3-letter readout sequences were designed by generating a random set of sequences with the per-base probability of 25% for A, 25% for T, and 50% for G. Sequences generated in this fashion can vary in their nucleotide content. To eliminate outlier sequences, only sequences with a GC content between 40 and 50% were kept. In addition, sequences with internal stretches of G longer than 3 nt were removed to eliminate the presence of G-quadruplets, which can form secondary structures that inhibit synthesis and binding. To remove the possibility of significant cross-binding between these readout sequences, we used a published algorithm (148) to identify a subset of these sequences with no cross-homology regions longer than 11 contiguous bases. We then used BLAST (149) to identify and eliminate sequences with contiguous homology regions longer than 11 nt to the human transcriptome. From the readout sequences satisfying the above requirements, 16 were selected. The corresponding readout probes, i.e. the reverse complement of these readout sequences, are provided in Appendix A.
To construct the library of encoding probes, we first selected a set of target RNAs (130 genes) drawn from the human transcriptome. We chose 85 of the genes used in our previous 140-gene MERFISH library (63) and then selected the remaining 45 genes at random from those expressed in the range of $10^{-1}$ to $10^3$ FPKM. We assigned to each RNA a unique barcode drawn from the same 16-bit Modified Hamming-Distance-4 (MHD4) code that we utilized previously (63). This code, which has a Hamming Distance of 4 and a constant Hamming Weight, i.e. the number of ‘1’ bits per barcode, of 4, contains 140 barcodes. We randomly assigned one of the 140 barcodes to each of the 130 RNA species and left the remaining 10 as blank controls. 92 putative encoding probes were created for each gene; each of the encoding probes contained a target region that was randomly selected from the target regions of the gene and three readout sequences that were randomly selected from the four readout sequences associated with the gene. These readout sequences were concatenated with the target regions in one of two randomly selected configurations: either one or two readout sequences at the 5’ end of the target region with the remaining sequences at the 3’ end of the target region. Additional adenosine nucleotide spacers were added between readout sequences and target regions to prevent terminal G triplets in the readout sequences or target regions to combine with Gs from adjacent sequences to form G quadruplets. Priming regions for the amplification of these probes were designed by randomly generating a set of 20-nt-long sequences, selecting those with a $T_M$ in the range of 70 to 72 °C, a GC content in the range of 50 to 65%, no contiguous region of four or more of the same base, and no region of self-complementarity longer than 6 nt. A final set of orthogonal primers were then designed again as described previously (148) with the requirement that there be no region of cross homology longer than 8 nt. Two primers were drawn at random from these sequences and added to the 5’ and 3’ of each encoding probe. Finally, this set of putative probes was screened
for any additional homology to human rRNA, tRNA, and mitochondrial rRNA and tRNAs, using
the same approach and parameters as in the design of the target regions. These sequences
(Appendix A) were then ordered from CustomArray. The code for the construction of this library
was written in MATLAB and is open source (http://zhuang.harvard.edu/merfish/).

3.4.2 Encoding Probe Construction

Encoding probes were constructed via a high-yield, enzymatic amplification protocol
published previously (63, 135), with a few notable differences to account for the use of the new,
3-letter readout sequences. Briefly, we created in vitro transcription templates from the complex
oligopool using limited cycle PCR, and then amplified RNA from these templates using a high-yield
in vitro transcription kit (NEB, E2050S). To account for the disproportionate use of C in
these sequences, we added additional CTP (ThermoFisher, R0451) to bring the final
concentration to 16.7 mM while the concentrations of ATP, GTP, and UTP were each 10 mM.
We then transcribed ssDNA probes from these RNA templates via reverse transcription (Maxima
RT H-, ThermoFisher, EP0752). To address the additional requirement for G in this reaction, we
doubled the concentration of all dTNPs to 3 mM. The RNA template was removed via alkaline
hydrolysis, the sample neutralized with 1 N HCl, and the DNA probes were purified through
phenol-chloroform extraction and two rounds of ethanol precipitation with ammonium acetate.
The final probes were resuspended in RNase-free water and stored at -20 °C.

3.4.3 Readout Probe Construction.

Readout probes that are complementary to the readout sequences on the encoding probes
and conjugated to the desired dye via a disulfide linkage were synthesized and purified by Bio-
synthesis, Inc. Lyophilized probes were immediately resuspended in Tris-EDTA buffer pH 8
(TE; ThermoFisher, AM9849) to a concentration of 100 µM to prevent degradation of the
fluorophore linkage (only observed for the Alexa750-linked readout probes) and stored at -20 °C. To reduce the number of freeze-thaw cycles experienced by these probes, 1 µM aliquots were made in TE buffer and stored at -20 °C.

3.4.4 High-throughput Imaging Platform

Samples were imaged on a custom-built, high-throughput, imaging platform. Briefly, the system was constructed around an Olympus IX71 microscopy body. Illumination was provided at 754, 647, 561, and 405 nm with solid-state lasers (Toptica, DL100/BooSTa; MBP Communications, F-04306-113; Crystalaser GCL-150-561; Coherent, Cube 405). These laser lines were used to excite Alexa750 and Cy5-labeled readout probes, orange fiducial beads, and DAPI, respectively. The 647-, 561-, and 405-nm lasers were collimated with custom 3-lens, 0.4X-to-3X-zoom systems, combined via a series of long pass filters (Chroma, z561bcm-xr; Semrock, LM01-503; Semrock, BLP01-405R), and then coupled into a single-mode fiber (Thorlabs, S405-XP) to purify each mode. The output of this fiber and that of the single-mode-fiber-coupled, 754-nm laser were collimated each with a 60-mm achromat and combined with a long-pass filter (Semrock, FF735-Di02). The sizes of these collimated beams were adjusted to 6.2 mm using a pair of custom, 3-lens, 0.5X-to-2X-zoom systems (one for the output of the 754-nm-laser fiber and the other for all other beams). The Gaussian distribution of these beams was then converted to a round, flat-top distribution using a refractive beam shaper (AdlOptica, piShaper 6_6). This distribution was focused onto a pair of galvanometer mirrors (Thorlabs, GVS201) and then relayed to the back-focal plane of a 300 mm achromat, which focused this illumination onto the back-focal plane of a 60X, Planapo, 1.3 NA, silicone oil objective (Olympus, UPLSAPO 60XS2). The fluorescence emission from the sample was separated from the laser illumination using a penta-band dichroic (Chroma, zy405/488/561/647/752RP-UF1).
Stray laser light was further removed with two copies of a custom notch filter (Chroma, ZET405/488/561/647-656/752m), and the fluorescent signal was imaged with a scientific CMOS camera (sCMOS; Andor, Zyla 4.2). The sample was positioned with a motorized microscope stage (Marzhauser, SCAN IM 112x74), and the focus was maintained via a custom-built autofocus system, which utilizes an objective nanopositioner (Mad City Labs, NanoF200) to maintain the position of a reflected IR laser (Thorlabs, LP980-SF15) on an inexpensive CMOS camera (Thorlabs, uc480). The sCMOS camera pixel size, 109.2 nm, was calibrated by imaging fields of fluorescent beads moved in defined increments with the motorized stage.

The sample coverslip was housed in a flow chamber (Bioptechs, FCS2), and the flow through this chamber was controlled via a home-built fluidics system composed of three computer-controlled eight-way valves (Hamilton, MVP and HVXM 8-5) and a computer-controlled peristaltic pump (Gilison, Minipuls 3) as described previously (63, 135). The entire system was fully automated, so that imaging and fluid handling were performed for the entire experiment without user intervention, using home-built software that is open source (https://github.com/ZhuangLab/storm-control).

### 3.4.5 Encoding Probe Staining

Human osteosarcoma cells (American Type Culture Collection, U-2 OS) were cultured with Eagle’s Minimum Essential Medium (American Type Culture Collection, 30-2003) containing 10% v/v fetal bovine serum (ThermoFisher, 10437). Cells were plated on 40-mm-diameter, #1.5 coverslips (Bioptechs, 0420-0323-2) at 300,000 cells per coverslip and incubated at 37 °C with 5% CO₂ for 48 to 72 hours within petri dishes. Cells were fixed, permeabilized, and stained with encoding probes as described previously (63, 135). Briefly, cells were fixed for 20 minutes in 4% paraformaldehyde (Electron Microscopy Sciences, 15714) in 1× PBS at room
temperature, washed three times with 1× PBS, permeabilized for 10 minutes with 0.5% v/v Triton (Sigma, T8787) in 1× PBS at room temperature, and washed three times with 1× PBS. Permeabilized cells were incubated for 5 minutes in encoding wash buffer comprising 2× saline-sodium citrate buffer (2× SSC; Ambion, AM9763), 30% v/v formamide (Ambion, AM9342), and 2 mM vanadyl ribonucleoside complex (VRC; NEB, S1402S). 30 μL of ~200 μM encoding probes (final concentration titrated for each probe batch) in encoding hybridization buffer was added to a glass microscope slide (Fisher Scientific, 22-265446), over which a cell-containing coverslip was placed. Samples were then incubated in a humid chamber inside a 37°C-hybridization oven for 36 to 48 hours. Encoding hybridization buffer is composed of encoding wash buffer supplemented with 0.1% w/v yeast tRNA (Life technologies, 15401-011), 1% v/v murine RNAse inhibitor (NEB, M0314L), and 10% w/v dextran sulfate (Sigma, D8906-50G). Cells were then washed with encoding wash buffer, incubated at 47 °C for 30 minutes, and this wash was repeated once. The cells were stained with 4', 6-diamidino-2-phenylindole, dihydrochloride (DAPI; ThermoFisher, D1306) during the second wash by adding 10 μg/mL DAPI to the encoding wash buffer. The sample was then incubated for 10 minutes with a 1:200,000 dilution of 0.1-μm-diameter carboxylate-modified orange fluorescent beads (Life Technologies, F-8800). The bead solution was aspirated, and the sample post-fixed with 4% v/v paraformaldehyde in 1× PBS at room temperature for 10 minutes. The sample was washed three times with 2× SSC and either imaged immediately or stored for no longer than 24 hours at 4 °C in 2× SSC containing 0.1% v/v murine RNAse inhibitor. All solutions were prepared as RNase-free. The beads were used as fiducial markers to align images obtained from successive rounds of hybridization.
Human lung fibroblast cells (American Type Culture Collection, IMR-90) were prepared using the same protocols as described above for the U-2 OS cells but without the DAPI staining.

3.4.6 MERFISH Imaging

Each readout hybridization mixture contained 1 nM each of two different readout probes, one conjugated to Cy5 and the other to Alexa750 via a disulfide bond, in readout hybridization buffer comprised of 2×SSC, 1% v/v ethylene carbonate (Sigma-Aldrich, E26258), 10% w/v dextran sulfate, and 2 mM VRC. The sample chamber was initially flushed with 2 mL of this readout hybridization mixture over the span of 5 minutes to fully exchange buffers. Then an additional 2 mL of this readout mixture was continuously flown across the sample for 6 more minutes. This total incubation time was several times that required to saturate binding (Figure 3.5) to reduce round-to-round and experiment-to-experiment variation in hybridization. The sample was then washed by flowing 2 mL of readout wash buffer (2×SSC, 10% v/v ethylene carbonate, and 2 mM VRC complex) for 9 minutes. Then 2 mL of imaging buffer comprising 2×SSC, 50 mM Tris-HCl pH 8, 10% w/v glucose, 2 mM Trolox (Sigma-Aldrich, 238813), 0.5 mg/mL glucose oxidase (Sigma-Aldrich, G2133), 40 μg/mL catalase (Sigma-Aldrich, C30) and 50 units/mL murine RNase inhibitor was flown across the sample for 4 minutes. Flow was stopped, and ~500 to ~1000 FOV were imaged. Because the imaging buffer is sensitive to oxygen, it was stored under a layer of mineral oil (Sigma-Aldrich, 330779) throughout the measurement (63, 135). 50% v/v stock solutions of ethylene carbonate were made by melting solid ethylene carbonate in a 65-°C water bath and then diluting to 50% v/v with RNase-free water.

After imaging, the fluorescence of the readout probes was extinguished via reductive cleavage using Tris(2-carboxyethyl)phosphine (TCEP). 2 mL of cleavage buffer comprising
2×SSC and 50 mM TCEP hydrochloride (Sigma, 646547) was flown across the sample for 4 minutes, the flow speed was reduced to 0.1 mL/min, and the sample incubated in this continuous flow for 15 minutes. After cleavage, the chamber was flushed with 2 mL of 2xSSC for 4 minutes to remove the risk of premature cleavage of the probes within the subsequent hybridization buffer. All buffers were prepared fresh for each experiment.

The above hybridization, imaging, and chemical cleavage process was repeated 8 times with the 488-nm and 405-nm channels imaged in conjunction with the first round of readout imaging. A complete MERFISH measurement of ~500 FOV covering 19.8 mm² required 12 hours while ~1000 FOV covering 40.8 mm² required 18 hours.

3.4.7 Single-Molecule Fluorescence In Situ Hybridization (smFISH)

smFISH stains were prepared at 1 µM probe concentrations following the same procedures described above in the “Encoding Probe Staining” section. Probes for smFISH on U-2 OS cells were designed using the same target regions selected for MERFISH measurements and were synthesized conjugated to Quasar760 (Biosearch, Stellaris). Probes for smFISH on human fibroblast cells (American Type Culture Collection, IMR-90) cells were generated using the target regions published previously for the FLNA mRNA (63) conjugated to multiple readout sequences, drawn either from the previously published sequences (63, 135) or those provided in Appendix A. These probes were synthesized by Biosearch.

3.4.8 Image Registration and Decoding

Registration of images of the same FOV in different rounds of hybridization was performed as described previously (63, 135). Briefly, the centroids of individual beads within the fiducial bead images collected for each FOV in each round of imaging were found using the
single-emitter fitting routine daoSTORM (117) and used to align images of the different rounds of hybridization using an affine transformation (non-reflective similarity) which corrected for translation, rotation, and a uniform coordinate scaling. In practice, we found that only an X and Y translation was required and that calculated transforms that contained a non-zero rotation or a non-uniform scaling were indicative of a rare registration failure. These affine transformations were then used to warp each image to the same coordinate system using linear interpolation. We found no systematic offsets between the centroid of spots in the Cy5 and Alexa750 channels, and, thus, did not perform any additional chromatic warping.

Warped images were saved as tiff stacks, one per FOV, with the frames in the order of the bits in the barcodes that they represented. These stacks were then pre-processed to remove background and better resolve overlapping fluorescent spots. Specifically, we used a high-pass filter comprised of a Gaussian filter with a kernel size of 3 pixels to remove background. This kernel was slightly larger than the PSF of the system so as not to remove regions of partially overlapping RNA signals. These high-pass filtered images were then deconvolved using Lucy-Richardson deconvolution and the estimated PSF of the system (2 pixels). In practice, we found it unnecessary to modify this kernel for the slight difference in PSF size between the Cy5 and Alexa750 channels. We then low-pass filtered these images using a Gaussian kernel with a width of 1 pixel (~100 nm). We found that this low-pass filter improved the quality of decoding, a result consistent with our previous observation that the spot centroids for the same RNA varied in position by ~100 nm in different imaging rounds, potentially due to the finite cellular volume occupied by each RNA (63).

To decode these images, we first recognized a few geometric properties of this problem. First, each set of 16 normalized intensity values observed for each pixel in each FOV represent a
vector in 16-dimensional space—we term this a pixel vector; second, the 140 barcodes of our 16-bit MHD4 code represent preferred directions in this space; and, finally, the set of all 16 single-bit errors generated from any of the 140 barcodes define a unique volume within this space, containing the set of all possible deviations from these barcodes that corresponds to a single-bit error or (in the case of an analog signal) less. The central premise of our decoding approach was that pixel vectors that fell within the volume defined by all single-bit errors from a given barcode should be associated with that barcode (we term this volume Hamming-Sphere 1, HS1). To identify all pixel vectors that fell within one of the 140 different HS1, we first mapped each pixel vector to the 16-dimensional unit sphere by dividing it by its magnitude, i.e. the $L^2$-norm. All barcodes were mapped to the unit sphere in a similar fashion. Because all barcodes shared the same Hamming Weight, the HS1 for each barcode was defined by the same distance, in this case, the maximum Euclidean distance between a barcode and all single-bit errors was 0.5176. Thus, occupancy in this volume could be calculated simply by determining the nearest barcode for each pixel vector and thresholding on the distance to that barcode. Any pixel vector with a distance to the nearest barcode larger than 0.5176 was left unassigned.

Because this decoding approach was conducted on individual pixels and because we observed that the signal from RNAs spreads across multiple pixels, we combined adjacent pixels assigned to the same barcode into a single putative RNA. We then calculated various properties of this RNA, including its magnitude-weighted centroid, the area (in pixels) it covered, the average magnitude across all pixels, as well as the average pixel vector across all combined pixels. Because this approach assigned barcodes within HS1 to a given barcode, it can be thought of as applying error correction. To determine whether or not error correction was applied to a given RNA, we computed the distance between the average 16-dimensional pixel vector for the
set of pixels associated with each RNA and a set of normalized barcodes including the barcode to which it was assigned and all barcodes generated from single-bit errors. If the nearest barcode to the average pixel vector for an RNA was one of the single-bit error barcodes, we then considered the RNA decoded with error correction applied. The specific single-bit error barcode to which it was closest defined the bit at which the error occurred. These quantities were then used to calculate the total number of RNAs decoded with or without error correction (Figure 3.8A), the confidence ratio associated with the counts of each barcode (Figure 3.8B), and the error rate for each type of error (1-to-0 or 0-to-1) at each bit (Figure 3.8C).

In this decoding approach, differences in the brightness between different imaging rounds and color channels will lead to different weights for ‘1’ values in each of the different bits, which, in turn, can lead to increased per-bit error rates. To remove this source of error we developed a two-step approach to removing these brightness differences between imaging rounds. First, we applied a crude normalization by setting the 90% quantile of pixel intensities for each imaging round to 1. Second, we used this normalization to decode RNAs from 100 randomly selected FOV as described above, and then used the observed pixel vectors from these decoded RNAs to refine this normalization. Specifically, we selected all RNAs of which the barcodes read ‘1’ in the first bit, and averaged the first component of their pixel vectors (calculated as described above). This produced an average intensity for a ‘1’ in the first bit. In a similarly way, we then calculated this average intensity quantity for all other bits. We then calculated an overall average value by taking the average of this average intensity quantity for each bit and we calculated the deviation from this average for each bit. We then renormalized the brightness of each imaging round based on the deviation observed for the average intensity of the corresponding bit to the overall average. Because this renormalization step will change the
quality of the decoding, we iterated this second step. In practice, we find that 10 iterations were sufficient to remove any substantial variation in the intensity of ‘1’s in each bit.

The majority of computations were run on the Odyssey cluster supported by the FAS Division of Science, Research Computing Group at Harvard University. By utilizing 32 cores and 64 GB of RAM, the complete analysis of a MERFISH data set could be completed in 3 days. Some datasets were analyzed on a desktop server, which contained two 10-core Intel Xeon E5-2680 2.8 GHz CPUs and 256 GB of RAM. Data were stored on a Synology DiskStation connected directly to this server via a high-speed switch. With this configuration, it was possible to analyze a single MERFISH dataset in 2 days using only 10 of the available cores and no more than 64 GB of RAM. This increase in speed relative to that of the Odyssey computing cluster was due to the increased read/write speed provided by the high-speed connection between the data storage system and the server.

All barcodes were stored in a custom binary format accessible with open source MATLAB functions (http://zhuang.harvard.edu/merfish/). Barcode files can be downloaded for all seven data sets from http://zhuang.harvard.edu/merfish/.

3.4.9 Cell Segmentation

Our cell segmentation approach to associate individual barcodes with individual cells exploited the observation that the density of RNAs dropped significantly at the edges of cells. Specifically, to generate the cell boundaries, we first created composite mosaic images in which a single FOV was flanked by its 8 surrounding neighbors. We then normalized the DAPI signal within this composite to the maximum observed value, thresholded this signal, and defined the cell nuclei as contiguous sets of pixels above this threshold. We then calculated the local density
of barcodes throughout each composite image by binning decoded barcodes into 2-µm × 2-µm bins. These binned images were then smoothed with a Gaussian filter of width equal to the bin size and resized to the original pixel size via bicubic interpolation. Segmentation boundaries for individual cells were calculated using the watershed algorithm, utilizing an inverted image of the barcode density (so that regions of low density formed natural watershed boundaries) and with the cell nuclear regions set to zero to insure that each watershed region contained a cell nucleus. Even though we included the 8 flanking FOVs for the computation, only cells with the centroid of their nuclei within the center FOV were kept. This approach produced some segmentation errors on a small fraction of cells, which were identified and removed via a series of thresholds. First, improperly segmented cells were identified by a threshold on the total effective cytoplasm area of 3000 µm² and by removing cells with segmented boundaries that shared more than 10 µm of that boundary with the edge of a FOV. Second, multiple overlapping nuclei were identified via a combination of thresholds on the total nucleus size and on a ratio of area covered by the nuclei boundary to that of a convex hull defined by the same boundary. We set the threshold of the nucleus size to 1000 µm² (nuclei with sizes larger than this value are considered arising from multiple overlapping nuclei). We set the threshold of the ratio to 1.06 (nuclei with the ratio larger than this value are considered arising from multiple overlapping nuclei because the boundary of overlapping nuclei have regions of concavity which tend to increase this ratio). Finally, this algorithm occasionally produced two boundaries for the same cell if the nucleus happened to be shared between two FOV. These cells were identified by finding cells for which 97% of the boundary was contained within the boundary of another cell. On average, roughly 98% of the identified cells passed the cellular boundary thresholds and ~90% of the resulting cells passed thresholds associated with the nuclear segmentation thresholds.
3.4.10 RNA-seq

All three replicates of RNA-seq data for U-2 OS (140) (GEO accession GSM1231610) were downloaded as sra files, converted to fastq files, and analyzed using the human transcriptome (hg38), the indices provided by Illumina’s iGenomes (http://support.illumina.com/sequencing/sequencing_software/igenome.html), Bowtie 2.2.1, TopHat 2.0.11, and Cufflinks 2.2.1 (150). The reported FPKM represents the average of that derived from these three replicates.
Chapter 4

High performance multiplexed fluorescence in situ hybridization in culture and tissue with matrix imprinting and clearing

4.1 Introduction

Highly multiplexed single-molecule fluorescence in situ hybridization (smFISH) has emerged as a promising approach to spatially resolved single-cell transcriptomics due to its ability to directly image and profile numerous RNA species in their native cellular context. However, background—from off-target binding of FISH probes and cellular autofluorescence—can become limiting in a number of important applications, such as increasing the degree of multiplexing, imaging shorter RNAs, and imaging tissue samples. Here, we developed a sample clearing approach for FISH measurements. We identified off-target binding of FISH probes to cellular components other than RNA, such as proteins, as a major source of background. To remove this source of background, we embedded samples in polyacrylamide (PA), anchored RNAs to this PA matrix, and cleared cellular proteins and lipids, which are also sources of autofluorescence. To demonstrate the efficacy of this approach, we measured the copy number of 130 RNAs in cleared samples using multiplexed error-robust fluorescence in situ hybridization (MERFISH). We observed a reduction both in the background due to off-target probe binding and in the cellular autofluorescence without detectable loss in RNA. This led to an improved detection efficiency and detection limit of MERFISH, and an increased measurement throughput via extension of MERFISH into four color channels. We further demonstrated MERFISH measurements of complex tissue samples from the mouse brain using this matrix imprinting and
clearing approach. We envision that this method will improve the performance of a wide range of in-situ-hybridization-based techniques in both cell culture and tissues.

Single-molecule fluorescence in situ hybridization (smFISH) is a powerful technique that allows the direct imaging of individual RNAs within single cells \( (59, 60) \). In this approach, individual copies of a specific RNA species are labeled via the hybridization of fluorescently labeled oligonucleotide probes, producing bright fluorescent spots for single RNA molecules, which reveal both the abundance and the spatial distribution of these RNAs inside cells \( (59, 60) \). The ability of smFISH to image gene expression at the single-cell level in both cell culture and tissue has led to exciting advances in our understanding of the natural noise in gene expression and its role in cellular response \( (151, 152) \), the intracellular spatial organization of RNAs and its role in post-transcriptional regulation \( (134, 153) \), and the spatial variation in gene expression within complex tissues and its role in the molecular definition of cell types and tissue functions \( (133) \).

In order to extend the benefits of this technique to systems-level questions and high-throughput gene expression profiling, approaches to increase the multiplexing of smFISH — i.e. the number of different RNA species that can be simultaneously quantified within the same cell — have been developed \( (63, 141–145) \). Most of these approaches take advantage of color multiplexing, which has allowed a few tens of RNA species to be imaged simultaneously. We have recently introduced multiplexed error robust fluorescence in situ hybridization (MERFISH), a massively multiplexed form of smFISH that allows RNA imaging and profiling at the transcriptomic scale \( (63, 154) \). MERFISH achieves this level of multiplexing by assigning error-robust barcodes to individual RNA species, labeling RNAs combinatorically with oligonucleotide probes that contain a representation of these barcodes, and then reading out these
barcodes through sequential rounds of single-color (63) or multi-color (154) smFISH imaging. Using this approach, we have imaged 140 and 1000 RNA species in individual cells with two different error-detecting and/or error-correcting encoding schemes (63), and the number of addressable RNA species can be further changed by using different encoding schemes. Recently, we have increased the measurement throughput of MERFISH and demonstrated the ability to profile gene expression in tens of thousands of cells in a single-day-long measurement (154). A different multiplexed smFISH method using color-based barcodes and sequential imaging (seqFISH) has been independently proposed and initially demonstrated with measurements of 12 RNA species in individual cells (145). While this current paper was in review, a paper reporting an extension of seqFISH that incorporates error correction and demonstrates the capability of imaging 125 or 250 RNA species was published (155).

smFISH measurements typically benefit from high signal-to-background ratios, resulting in the detection of individual RNA molecules with high accuracy and detection efficiency (59, 60): In many cases, the bright fluorescent signals that arise from the tens of fluorescently labeled probes bound to each copy of an RNA substantially exceed the background that arises from probes binding off target or from cellular autofluorescence. However, we have observed that as the degree of multiplexing is increased, the background level also tends to increase. The resulting decrease in the signal-to-background ratio makes a number of important applications and extensions of multiplexed smFISH challenging. For example, efforts to further increase the degree of multiplexing, to thousands or potentially tens of thousands of RNAs, will likely be limited by increased background. In addition, many RNAs are not long enough to accommodate tens of oligonucleotide probes, limiting the ability to measure relatively short RNAs and to
discriminate many different RNA isoforms. Finally, background is typically more pronounced in complex tissues, making multiplexed smFISH measurements in tissues more challenging.

Here we report a sample clearing approach aimed at improving the signal-to-background ratio in RNA FISH measurements by substantially reducing background fluorescence signal. Many modern tissue clearing approaches are designed to preserve the protein content of the sample while reducing scattering and autofluorescence background by extracting lipids and matching refractive index \[156-163\]. For example, embedding and crosslinking tissues to hydrogels provides a powerful approach to tissue clearing, minimizing sample distortion during lipid removal and index matching while maintaining the protein content of the sample \[159, 160\]. These approaches have also been made compatible with RNA FISH by stabilizing RNA molecules, for example through crosslinking of RNAs to proteins, without removing the protein content of the cell \[160, 164\]. However, we report here that a major source of background in RNA FISH measurements is the non-specific binding of FISH probes to cellular components other than RNAs, such as proteins. For this reason, a clearing method that preserves RNAs while removing proteins and lipids is desired for RNA FISH imaging. In the recently developed expansion microscopy method, proteins \[165\] and, more recently, RNAs \[166, 167\] are physically anchored to a solvent-expandable and clearable poly-electrolyte matrix, effectively imprinting signals of these components on this matrix and allowing these molecular signals to be expanded along with the matrix for super-resolution imaging. Inspired by this approach, we anchored RNA molecules to a non-swellable polyacrylamide (PA) matrix and then removed unwanted, non-RNA components, such as proteins and lipids, with the aim to remove their contribution to background fluorescence. We demonstrated that this matrix imprinting and clearing approach substantially reduced the background due to off-target binding of FISH probes.
and cellular autofluorescence. By comparing the copy number of 130 RNAs measured via MERFISH in uncleared and cleared cultures of human cells, we demonstrated that this matrix-imprinting-based clearing approach improves the detection efficiency and detection limit of MERFISH with no detectable loss in RNAs. Moreover, the reduction in autofluorescence, in particular in the blue-green spectral range, allowed us to extend MERFISH imaging from two to four distinct color channels with no reduction in performance. This improvement substantially reduced the number of hybridization rounds needed for MERFISH measurements, which should further increase the MERFISH measurement speed and throughput. Finally, we demonstrated that this clearing approach substantially reduces the background in tissue, allowing us to perform MERFISH measurements in cryosections of adult mouse brain tissues. Given the simplicity and efficacy of this matrix-imprinting-based clearing method, we envision that this approach could be used to substantially improve the performance of a wide range of in-situ-hybridization methods for both RNA and DNA in cell cultures and tissues.

4.2 Results

4.2.1 A matrix imprinting and clearing approach to reduce the background for smFISH measurements

Our first step in the development of a sample clearing method for smFISH was to determine the physical origin of off-target binding of oligonucleotide probes: are these probes binding to the incorrect RNA, or other cellular components such as proteins or lipids? To address this question, we stained human lung fibroblast (IMR-90) cells using FISH probes targeting the Filamin A (FLNA) mRNA. As expected, we observed both bright fluorescence spots marking individual molecules of FLNA mRNA (Figure 1A, left) and a diffuse background due to off-target probe binding (Figure 4.1A, middle) that was not present in samples not stained with FISH
probes (Figure 4.2). We then measured the RNase sensitivity of both the foreground RNA spots and the diffuse background, reasoning that if the background arose from off-target binding to incorrect RNAs, both the foreground spots and the background should be RNase sensitive. We found that a brief RNase A treatment completely removed the bright foreground spots, but produced little if any reduction in the background (Figure 4.1A, right). Thus, we conclude that the vast majority of off-targeting binding of smFISH probes arose from binding to cellular components other than RNA, such as proteins and lipids.

Figure 4.1: Matrix imprinting and clearing reduces background in smFISH measurements. A. A human fibroblast cell (IMR-90) stained with smFISH probes targeting the FLNA mRNA before (left and middle) and after (right) treatment with RNase A. The contrast of the middle and right panels has been increased 5-fold from that of the left panel to better visualize the background from probes bound off-target. Scale bars: 10 µm. B. Schematic diagram of a matrix-imprinting and clearing approach to reduce background in smFISH measurements. Cells are stained with smFISH probes or encoding probes for MERFISH measurements, and a poly-dT anchor probe which targets the polyA tail of mRNAs. Cells are then embedded in a polyacrylamide (PA) matrix, to which the poly-dT anchor probes are covalently linked via a terminal acrydite moiety. Proteins and lipids are then digested and extracted, freeing off-target bound smFISH probes to diffuse out of the PA matrix and removing cellular components that contribute to autofluorescence. C. U-2 OS cells labeled with MERFISH encoding probes targeting 130 RNAs followed by staining with a readout probe conjugated to Cy5 that binds to the encoding probes in an uncleared sample (upper panel) and a sample treated with the matrix imprinting and clearing protocol (lower panel). Scale bars: 20 µm.
Figure 4.2: Off-target binding of FISH probes is largely insensitive to RNase treatment. Images of different background sources in IMR-90 cells: Cells stained with encoding probes but no fluorescently labeled readout probe (left), cells stained with a fluorescently labeled readout probe but no encoding probes (middle), and cells stained with encoding probes, a fluorescently labeled readout probe that can bind to a readout sequence on these encoding probes, and then treated with RNase A in order to remove all specific RNA signals (right). All three images are displayed at the same contrast to illustrate the relative intensity of the signal from the autofluorescence background of the cell (left), the very low level (if any) of non-specific binding of readout probes, and the signal from the off-target (RNase-insensitive) binding of encoding probes followed by binding of the readout probes to the encoding probes. The encoding probes used here target the FLNA mRNA only, and the readout probe used here is the Bit-1 readout probe conjugated to Cy5 (Appendix B). Scale bars: 5 μm.

Since this background arose from binding of FISH probes to cellular components other than RNA, we reasoned that one way to reduce it would be to remove unwanted components, such as proteins and lipids, from the sample. Moreover, since these components are also a major source of autofluorescence, the autofluorescence background might be reduced by such an approach as well. To this end, we fixed the sample and hybridized it with oligonucleotide probes as in standard smFISH (59, 60) or MERFISH measurements (63, 135), and then embedded the sample in a inert, non-fluorescence matrix to which RNA molecules were anchored, effectively imprinting the desired RNA signal onto this matrix. Once RNAs were anchored, cellular proteins and lipids were removed without, in principle, affecting the number and localization of RNAs.
within the sample. smFISH probes bound off-target to these components should then be free to diffuse from the matrix. We utilized polyacrylamide (PA) as the inert matrix and a 15-nt poly-dT oligonucleotide to bind and anchor polyadenylated (polyA) RNAs to the PA matrix. This anchor probe was comprised of 50% locked-nucleic acid bases to stabilize the hybridization to polyA tails of the RNAs (168) and additionally contained a terminal acrydite moiety which can be covalently incorporated into the PA matrix as it polymerizes.

To test whether this clearing approach led to a reduction in off-target binding, we first measured the efficacy of protein and lipid removal and observed that this protocol efficiently removed cellular proteins and lipids from embedded cultured human osteosarcoma (U-2 OS) cells (Figure 4.3). Next, we performed labeling as in MERFISH experiments and tested whether off-target probe binding was indeed reduced by clearing. In a MERFISH measurement, we typically stain cells first with a complex library of “encoding” oligonucleotide probes (63, 135). These encoding probes are not themselves fluorescently labeled. Instead, each encoding probe contains a targeting sequence that directs its binding to a cellular RNA and multiple readout sequences. Multiple encoding probes are targeted to each RNA, and the set of readout sequences contained within these encoding probes form a specific barcode that is unique to that RNA. These barcodes are then measured in a series of hybridizations, each round of hybridization using either one fluorescently labeled ‘readout’ probe complementary to a specific readout sequence, reading out one bit, in the single-color imaging mode (63) or multiple ‘readout’ probes labeled with spectrally distinct dyes, reading out multiple bits simultaneously, in the multi-color imaging mode (154). One advantage of this two-step labeling approach with encoding hybridization followed by readout hybridization is that it substantially reduces the time required for each hybridization round because hybridization of the readout probes to encoding probes
(including all fluid handling and sample washing) requires <30 minutes (63, 135) as compared to the overnight hybridization typically required for direct hybridization of FISH probes to cellular RNAs, because the readout sequences on the encoding probes do not form secondary structure and are not occluded by cellular proteins.

Figure 4.3: **Protease digestion and detergent treatment efficiently remove protein and lipid from poly-acrylamide embedded cells.** A. Images of U-2 OS cells stained with Krypton, a non-specific protein dye, before (Uncleared) or after matrix imprinting and clearing (Cleared). The contrast at which the right image is displayed has been increased 10× relative to the middle image to better illustrate the reduction in fluorescence signal. B. The average fluorescence signal observed from the samples in (A). The average fluorescence has been normalized to the fluorescence observed in the uncleared sample. The error bar represents SEM (n = 3 replicates). C. As in (A) but for DiD, a non-specific lipid stain. D. As in (B) but for the samples stained with DiD. Scale bars: 20 µm.

To demonstrate the clearing efficacy, we stained U-2 OS cells with encoding probes used for a MERFISH measurement of 130 RNAs at a total concentration of 300 µM, which is 3-fold higher than typically used in our MERFISH experiments, in order to generate high background. We then embedded and cleared the sample in the PA matrix as described above, and stained the RNA-imprinted matrix with a readout probe labeled with a Cy5 dye. Figure 4.1C shows that the
cleared samples contained visible smFISH spots but substantially lower background than uncleared samples, demonstrating that this approach indeed reduced the background due to off-target probe binding.

As an aside, MERFISH measurements require repeated sample staining with a series of readout probes and, in cases where the FISH signal is removed by chemical cleavage of the fluorophores (154), the efficient removal of cleaved fluorophores. To facilitate the rapid penetration of readout probes as well as the rapid removal of cleaved dyes, we embedded samples in 50–100-µm thick PA films. These films were thick enough to cover cultured cells or moderately sized tissue slices, yet thin enough that the rate of readout probe hybridization and the rate of dye cleavage/removal were not substantially changed from those observed in uncleared samples (Figure 4.4).

Figure 4.4: Matrix imprinting and clearing in polyacrylamide films does not reduce the rate of readout probe binding or reductive cleavage of fluorescent dyes. A. The average brightness of individual RNA spots as a function of time exposed to a readout probe conjugated to Cy5 in uncleared samples (orange) or matrix imprinted and cleared samples (blue). The average brightness was normalized to the average of the brightness observed in the final two time points. B. The average brightness of individual RNA spots as a function of time exposed to cleavage buffer (Materials and Methods). The average brightness has been normalized to that observed prior to exposure to cleavage buffer. Both measurements were conducted on IMR-90 cells stained with encoding probes targeting the FLNA mRNA and the first readout probe (Bit 1; Appendix B). The readout hybridization buffer utilized in (A) differed slightly from that
described previously (154) in that it contained 3 nM of the readout probe and no dextran sulfate. All error bars represent SEM (n = 3 replicates).

4.2.2 RNA is preserved during matrix imprinting and clearing

To determine if any RNAs were lost during matrix imprinting and clearing, we used MERFISH to determine the copy number of 130 RNAs in a cleared sample of U-2 OS cells and compared these numbers to that derived previously from a uncleared sample (154). We utilized our previously published 16-bit, modified Hamming-distance-4 (MHD4) code to encode RNAs (63). In this encoding scheme, all valid binary barcodes used to encode RNAs are separated by a Hamming distance of at least 4, which means that at least four bits must be read incorrectly to change one valid barcode to another, drastically reducing the probability of mis-identifying RNAs. Furthermore, this scheme also allows us to correct single-bit errors because every single-bit error produces a barcode uniquely close to a single valid barcode. This specific MHD4 code contains 140 valid barcodes (63), and we only utilized 130 of them to encode RNAs, leaving the remaining 10 barcodes to serve as ‘blank’ controls to determine the rate of spurious RNA detection and estimate misidentification rates.

We performed this MERFISH measurement of these 130 RNA species as described previously (154), using two-color imaging to read out 16 bits in 8 rounds of hybridization and imaging (two bits per round), as well as reductive cleavage of disulfide bonds to remove the fluorophores linked to the readout probes between consecutive rounds of smFISH imaging. Figure 4.5A shows that individual RNA molecules could be clearly detected in each of the 8 hybridization and imaging rounds, allowing their identity to be decoded. As described previously (154), we used the depletion of RNAs near cell boundaries to perform cell segmentation. Figure
4.5B shows that the copy number per cell observed for these RNAs measured in the cleared sample correlated strongly with those measured in a uncleared sample with a Pearson correlation coefficient of 0.94 between the log\(_{10}\) copy numbers (\(\rho_{10} = 0.94\) for the 116 RNA species whose measured copy numbers were larger than that observed for the largest 'blank' barcode). On average, the ratio between the copy numbers measured in the cleared sample to those measured in the uncleared sample was 1.12 ± 0.04 (SEM, n = 116 RNAs), and this ratio was largely independent of the length of the RNAs (Figure 4.5C). Although we conservatively used only those RNAs with copy numbers greater than the largest 'blank' count for analysis here, the results were similar when all 130 RNAs were used.

**Figure 4.5:** Matrix imprinting and clearing improves MERFISH performance with no loss in RNA. **A.** Left: two-color smFISH images from each of the 8 rounds of hybridization and imaging in a MERFISH measurement of 130 RNA species in matrix imprinted and cleared U-2 OS cells utilizing readout probes labeled with Cy5 (green) or Alexa750 (red). Yellow represents the overlay between the two dyes. Only a small portion of the MERFISH imaging field of view is shown. Scale bars: 2 \(\mu\)m. Right: All identified RNAs (colored markers) detected in a single field-of-view with the barcodes of the RNAs represented by the colors of the markers. The white box represents the portion of this field of view displayed in the left panels. Scale bar: 25 \(\mu\)m. **B.**
The average copy numbers per cell observed for these RNA species in matrix imprinted and cleared U-2 OS cells versus the copy numbers obtained from previously published measurements in an uncleared sample (154). Copy numbers were corrected by subtracting the average copy number observed for the ‘blank’ barcodes. Uncorrected copy numbers are displayed in Figure 4.6B. The log_{10} counts correlate with a Pearson correlation coefficient of 0.94 (p-value: 10^{-54}). The dashed line represents equality. C. The average ratio of the copy number per cell for a sample that was matrix imprinted and cleared to that observed for an uncleared sample for RNAs within the specified RNA length range. Error bars represent SEM (n = 26 genes for each bin). D. Average copy number per cell of the ‘blank’ barcodes, i.e. barcodes not assigned to an RNA, in an uncleared sample and in a matrix imprinted and cleared sample. Error bars represent SEM (n = 10 ‘blank’ barcodes).

These measurements showed that several aspects of MERFISH performance were improved with matrix imprinting and clearing. Previously, we observed a MERFISH detection efficiency of ~90% (154); thus, a copy number ratio of ~1.1 between the cleared and uncleared samples suggested that clearing increased the detection efficiency to near 100%. Second, we observed that the average frequency at which the ‘blank’ barcodes were observed per cell in the cleared samples dropped substantially relative to that observed in the uncleared samples (Figure 4.5D). The average level of ‘blank’ barcode counts observed in the uncleared sample (Figure 4.5D) was comparable to the observed copy number for the lowest abundance RNAs measured here, leading to the possibility that the copy number observed for these low abundance RNAs might have been biased by a background rate of spurious RNA counts in uncleared samples. Indeed, we observed an excess of these low abundance RNAs in uncleared samples relative to that expected from bulk RNA-seq (Figure 4.6A), whereas this bias was substantially reduced in cleared samples (Figure 4.6A-B). Thus, we conclude that the increased signal-to-background in cleared samples results in an improvement in both the detection efficiency and the detection limit in MERFISH measurements.
Figure 4.6: Matrix imprinting and clearing reduces bias in the detection of low abundance RNAs. A. The ratio of the copy number per cell determined via MERFISH to the abundance determined via RNA-seq (140) as measured in FPKM for uncleared samples (blue) and for matrix imprinted and cleared samples (red). Error bars represent SEM (n = 26 RNAs in each abundance range). B. The copy number per cell determined via MERFISH in a matrix imprinted and cleared sample as compared to that determined for an uncleared sample. These copy numbers have not been corrected for the average rate of ‘blank’ barcode detection as in Figure 4.5B. The dashed line represents equality. The deviation from equality in (B) and the excess MERFISH counts relative to those estimated from bulk-seq at the low abundance range are consistent with the increased rate of ‘blank’ barcode detection observed for untreated samples (Figure 4.5).

4.2.3 Extending MERFISH to four-color imaging

In addition to providing a substantial decrease in the background due to off-target binding of FISH probes, the removal of proteins and lipids from the sample may also reduce the level of autofluorescence. To quantify this decrease, we measured the fluorescence of unlabeled U-2 OS cells in uncleared and cleared samples with four excitation wavelengths — 750 nm, 647 nm, 561 nm, and 488 nm. Consistent with the expectation that cell autofluorescence is substantially higher in the blue-green spectral range than in the red range, the clearing protocol had little effect on the already low autofluorescence background in the 750-nm and 647-nm channels but produced a substantial reduction in the autofluorescence observed in the 561-nm and 488-nm channels (Figure 4.7A).
Figure 4.7 Autofluorescence reduction by matrix imprinting and clearing facilitates four-color MERFISH. A. The average autofluorescence observed for unstained U-2 OS cells before (blue) and after (red) matrix imprinting and clearing when excited with 750-nm, 647-nm, 561-nm, or 488-nm light. Error bars represent SEM (n = 3 replicates). B. Images of cleared U-2 OS cells stained with MERFISH encoding probes targeting 130 RNAs and the first four readout probes each conjugated to one of the following dyes: Alexa750, Cy5, ATTO565, or Alexa488. Samples were imaged with excitation light listed in (A). Scale bars: 10 µm. C. Average copy number per cell determined via four-color MERFISH to that determined with two-color MERFISH, both in cleared samples. The copy numbers have been corrected by subtracting the average rate of ‘blank’ barcode detection as in Figure 4.5B. The dashed line represents equality. The Pearson correlation coefficient between the log10 abundances is 0.99 (p-value: 10^-98). D. The average rate of observing a ‘1’ to ‘0’ error (blue) or a ‘0’ to ‘1’ error (red) per bit for bits that are read out with each of the four color channels, as indicated by the excitation wavelength. Each error rate (‘1’ to ‘0’ or ‘0’ to ‘1’) was calculated for each individual bit using the frequency at which errors were corrected at that bit, as described previously (63), and then these per-bit error rates were averaged over the bits that were detected in the same color channel (Appendix B). Error bars represent SEM (n = 4 bits read out with each color channel).

With this reduction in the autofluorescence, we explored the possibility of using all four excitation channels to read out four different bits of the 16-bit code simultaneously in each round of imaging during MERFISH measurements. We again stained U-2 OS cells with the same MERFISH encoding probe set as described above and performed MERFISH measurements in
which each round of hybridization utilized four different readout probes, conjugated respectively to Alexa750, Cy5, ATTO565, or Alexa488 via a disulfide bond (Figure 4.7B, Appendix B). With such four-color imaging, the 16-bit MERFISH measurement only required four rounds of hybridization and imaging.

We then compared the measured copy numbers derived from this four-color measurement to those determined with a two-color (750 nm and 647 nm) measurement in the cleared sample. Figure 4.7C demonstrates that these copy numbers correlated strongly with a $\rho_{10}$ of 0.99 and had an average ratio of $1.01 \pm 0.02$ (SEM, $n=109$ RNAs with copy numbers greater than that observed for the largest 'blank' barcode). To confirm that imaging in the new color channels did not introduce additional error, we determined the ‘1’ to ‘0’ or ‘0’ to ‘1’ error rates per bit and found that these error rates did not vary substantially with the color channel (Figure 4.7D).

Finally, to confirm that the improved performance that we observed with cleared samples was reproducible, we performed additional two-color and four-color MERFISH measurements in cleared samples. Figure 4.8 shows that the copy numbers derived from all of these measurements correlated strongly (all $\rho_{10}$ are 0.94 or greater). By comparing each of these data sets to the previously determined detection efficiency of MERFISH measurements in uncleared samples (154), we estimated an average MERFISH detection efficiency of $96 \pm 7\%$ (SEM, $n=4$ replicates) for cleared samples. Furthermore, we observed a $\sim$4-fold reduction in the average rate of ‘blank’ barcode detection ($0.08 \pm 0.03$ counts per cell [SEM, $n=4$ replicates] for cleared samples versus $0.30 \pm 0.07$ counts per cell [SEM, $n=7$ replicates (154)] for uncleared samples).
Figure 4.8: Two- and four-color MERFISH measurements in matrix imprinted and cleared samples are reproducible. Comparison of the average copy number per cell measured in different two-color or four-color MERFISH measurements in matrix imprinted and cleared U-2 OS cells. $\rho_{10}$ represents the Pearson correlation coefficient between the log$_{10}$ copy numbers for all RNAs. The p-values associated with all $\rho_{10}$ are less than $10^{-44}$.

4.2.4 MERFISH measurements of brain tissue

To explore whether clearing can overcome the increased background that we have observed in tissues, we performed MERFISH measurements of 130 RNA species on four cryosections taken from adult mouse hypothalamus, each ~2-mm×2-mm wide and 10-µm thick (Figure 4.9A-B). We performed 3D imaging with seven ~1.5-µm thick optical sections measured per field of view. These RNAs were again encoded with the 16-bit MHD4 code and read out with 8 rounds of hybridization using two-color imaging per round. These samples were matrix imprinted and cleared as described above but with the addition of a brief treatment with 4% w/v
sodium dodecyl sulfate (SDS) prior to PA embedding, which further improved tissue clearing. Figure 4.9C and D illustrate that this clearing approach substantially reduced the background observed in these tissue slices. As a result, smFISH spots representing individual RNA molecules were observable in the cleared sample in each round of imaging with similarly low ‘1’ to ‘0’ or ‘0’ to ‘1’ error rates as detected in cultured cells, allowing individual RNAs to be decoded (Figure 4.9E-G).

**Figure 4.9: MERFISH measurements of adult mouse brain tissue.** A. Nissl-stained images of coronal and sagittal slices of an adult mouse brain taken from the Allen brain atlas (169). The black box and dashed line represent the region of the mouse hypothalamus studied. Scale bar: 2 mm. B. Image of a single, 10-µm-thick cryosection of the mouse hypothalamus stained with DAPI (4’, 6-Diamidino-2-phenylindole, dihydrochloride). The complete volume of the central 2-mm × 2-mm region of this slice was imaged with MERFISH using seven 1.5-µm thick optical sections per field-of-view. Scale bar: 1 mm. C-D. Images of a small portion of a mouse hypothalamus slice stained with an encoding probe set used for a MERFISH measurement of 130 RNAs and a readout probe conjugated to Cy5. (C) Single optical section image of an uncleared sample; (D) Single optical section image of a matrix imprinted and cleared sample. Scale bar: 50 µm. E. Zoom-in of the region of (D) marked with the white dashed box. F. Decoded RNAs (different colors represent different barcodes) from all seven optical sections of the region shown in (E). Not all RNA molecules shown in (F) are observed in (E) because (E) represents only one of the seven optical sections and one of the 16 bits. G. The average rate of observing a ‘1’ to
‘0’ error (blue) or a ‘0’ to ‘1’ error (red) per bit for bits that are read out with each of the two color channels, as indicated by the excitation wavelength. Error rates were calculated as in Figure 4.7. Error bars represent SEM (n = 8 bits read out with each color channel). H. The density of 130 RNA species as determined via MERFISH versus the abundance as determined via bulk RNA-seq for the region of the mouse hypothalamus shown in (A). The Pearson correlation coefficient between the log_{10} abundances is 0.84 (p-value: 10^{-35}).

We compared the average RNA density determined via MERFISH from these four tissue slices with the abundance determined via bulk RNA-seq data derived from the same region of the hypothalamus (170), and observed a strong correlation (\( \rho_{10} = 0.84 \)) between our MERFISH results and the bulk RNA-seq results (Figure 4.9H). At the very low abundance range corresponding to those RNAs that are expressed poorly in the hypothalamus (< 0.5-1 FPKM), the correlation between MERFISH and bulk RNA-seq results reduced substantially, suggesting that the abundance of these RNAs was near or below our current detection limit, a conclusion supported by the similarity between the copy numbers of these RNAs and the average copy number observed for the blank barcodes (\( 6 \times 10^6 \pm 2 \times 10^6 / \text{mm}^3 \) [SEM, n =10 ‘blank’ barcodes]).

### 4.3 Discussion

Massively multiplexed smFISH allows spatially resolved gene expression profiling within single cells. However, a number of important applications of and advances to this approach are limited by the fluorescence background encountered in these experiments. Here we described a clearing approach that substantially reduced several background sources in RNA FISH measurements by effectively imprinting the desired RNA signal onto an inert, non-fluorescent, PA matrix and then removing unwanted cellular components that give rise to background due to off-target probe binding and autofluorescence. The reduction in fluorescence background provided by this approach led to improvement in both the detection efficiency and
the detection limit in MERFISH measurements. Moreover, this matrix imprinting and clearing approach produced a substantial reduction in the background observed for measurements in tissue samples, allowing us to perform MERFISH measurements in tissue sections of the mouse hypothalamus.

This matrix imprinting and clearing approach complements several existing methods that improve the signal or reduces non-specific binding background for RNA detection using in situ hybridization (155, 164, 166, 171, 172) or in situ sequencing (136, 137). For example, signal amplification techniques, such as branched DNA (171, 173), rolling-circle amplification (136, 137, 174, 175), and hybridization chain reaction (155, 164, 166, 172, 176) have been utilized to increase the signal associated with each RNA molecule, thereby increasing the ratio between RNA signals and autofluorescence background, although background due to off-target probe binding may also be amplified concurrently by these approaches. Alternatively, proximity-dependent approaches, in which a fluorescent signal is only produced when two separate probes or two ends of the same probe are in close proximity, have been used to reduce the effect of non-specific binding and thereby increase the ratio between signals and background due to off-target probe binding (136, 171, 175, 177, 178). Because matrix imprinting and clearing is compatible with each of these approaches, it may be combined with these techniques to further improve the performance of both conventional and highly multiplexed smFISH measurements.

In addition, the reduction in autofluorescence in the blue and green color channels provided by matrix imprinting and clearing allowed us to extend MERFISH measurements from two colors to four colors with no loss in performance, which should substantially increase the measurement speed and throughput of MERFISH. We previously reported a high-throughput MERFISH imaging platform and demonstrated the ability to profile 140 RNA species in 40,000
human cells in ~18 hours using a 16-bit MHD4 code, two-color imaging, and 8 hybridization rounds (154). This platform also allowed us to profile a similar number of RNAs across a ~16-mm² x 10-μm tissue volume in ~18 hours here. This throughput was facilitated by our ability to rapidly bind FISH probes (through hybridization of readout probes to encoding probes, which requires <30 minutes) and then rapidly remove FISH signals (though the use of chemical cleavage to remove fluorophores from readout probes, which requires <15 minutes) in each round of hybridization. With the increase to four color channels, we can now read out the 16-bit code in just four rounds of hybridization with four bits detected per round, further reducing the measurement durations and allowing tens of thousands of cells to be profiled in <10 hours with MERFISH. Similarly, we also anticipate that our previously demonstrated MERFISH measurement of 1001 RNA species using a 14-bit MHD2 encoding scheme (63) can also be reduced to just four hybridization rounds. A recent paper demonstrated the profiling of >100 RNA species in tissue using a different multiplexed smFISH method (seqFISH) with a similar number of hybridization rounds, for example, 125 RNA species were measured with a 4-letter, 5-color barcoding scheme using four rounds of hybridization, though the approach used for probe hybridization (overnight hybridization of FISH probes directly to cellular RNA) and signal removal (4 hours of DNase I digestion of FISH probes) in each round made the overall measurement time longer (155).

Finally, we envision that the substantial reduction in background provided by this matrix imprinting and clearing approach will facilitate several additional extensions of MERFISH. First, an increase in the degree of multiplexing — to the simultaneous measurement of thousands or tens of thousands of RNAs — would likely require substantially higher encoding probe concentrations than are currently used and, thus, would benefit from the much lower off-target
probe binding achieved here in cleared samples. Second, we have performed MERFISH on RNAs that are 3 kb or longer with 92 encoding probes per RNA (63, 135). With the dramatic decrease in background enabled by this clearing approach, it should be possible to detect RNAs that are much shorter, potentially with as few as <10 encoding probes per RNA. This advance would facilitate the detection of relatively short messenger and long-non-coding RNAs, and possibly some small RNAs. The ability to detect RNA molecules with relatively few FISH probes will also substantially improve the ability to distinguish RNA isoforms. Third, the combination of expansion microscopy (165) with MERFISH may be facilitated by a common matrix imprinting approach, which may help RNA profiling in RNA-dense regions of cells and further increases in the degree of multiplexing. We also anticipate that the alternative RNA anchoring approach reported for expansion microscopy (166)—in which RNAs are alkylated with a crosslinker that is covalently incorporated into the PA gel—may be used for background reduction by matrix imprinting and clearing as well. Finally, while our current implementation of the matrix imprinting and clearing approach removes cellular proteins and, thus, information regarding the protein content of the sample, we envision that this information could be restored by labeling samples with antibodies conjugated to oligonucleotides prior to embedding. These oligonucleotides could then be anchored to the matrix, followed by digestion of the antibodies along with cellular proteins, allowing the original location of the antibodies to be determined via FISH imaging of these oligonucleotides (165). Such labeling approaches could be used to label cell boundary markers to facilitate cell segmentation in cases where a depletion of RNA at the edge of the cell is not sufficient to identify cell boundaries. Moreover, such approaches can convert protein identities into oligonucleotide signals, which may be used to perform highly multiplexed proteomic and transcriptomic measurements simultaneously in single cells.
4.4 Materials and Methods

4.4.1 MERFISH probe library design and construction

MERFISH measurements in human osteosarcoma cells (American Type Culture Collection, U-2 OS) were performed with the same MERFISH encoding probe set as previously described (154). Briefly, this encoding scheme utilized a 16-bit modified Hamming-distance-4 code (MHD4) to encode the RNAs (63). In this encoding scheme, each of the 140 possible barcodes required at least four errors to accumulate to be converted into another barcode. This property permitted the detection of errors at up to any two bits, and the correction of errors to any single bit. In addition, this encoding scheme utilized a constant Hamming weight, i.e. the number of ‘1’ bits in each barcode, of 4, in order to avoid potential bias in the measurement of different barcodes due to a differential rate of ‘1’ to ‘0’ and ‘0’ to ‘1’ errors, as described previously (63). We used 130 of the 140 possible barcodes to encode cellular RNAs and the remaining 10 barcodes were left unassigned to serve as ‘blank’ controls. The encoding probe set that we used contained 92 encoding probes per RNA, with each encoding probe containing three of the four readout sequences assigned to each RNA.

The MERFISH encoding probes for measurements in the mouse hypothalamus were designed using the same 16-bit MHD4 code as above. Again, 130 of the 140 possible barcodes were assigned to RNAs that were selected to cover roughly three orders of magnitude in average expression in the hypothalamus with expression levels estimated from previously published RNA-seq (170). The remaining 10 barcodes were left unassigned to serve as ‘blank’ controls. Encoding probes were designed using a previously published pipeline with the same stringency conditions and design criteria as described previously (154). Transcript sequences were derived.
from the mouse genome (mm9) downloaded from ensembl (http://ensembl.org/Mus_musculus/Info/Index).

Construction of the encoding probe sets was conducted from complex oligonucleotide pools, as described previously (135, 154). Briefly, we amplified the oligopools (CustomArray) via limited-cycle PCR to make in vitro transcription templates, converted these templates into RNA via in vitro transcription, converted the RNA back to DNA via reverse transcription, and then purified the DNA via alkaline hydrolysis (to remove RNA templates), phenol-chloroform extraction (to remove proteins), and ethanol precipitation (to remove nucleotides and concentrate probes). To improve probe purity and reaction yield, we modified the previous protocol (135, 154) in the following ways. First, excess NTPs or dNTPs were removed via desalting columns (40K MWCO Zeba; ThermoFisher, 89894) after both the in vitro transcription and the phenol-chloroform extraction. We found that removal of stray NTPs improved the performance of the reverse transcription and that removal of excess dNTPs aided in quantification of the final yield of the protocol. In addition, to further improve yield, we switched the salt in our ethanol purification from 2.5 M ammonium acetate (which allows nucleotides to be removed, but decreases DNA recovery in our hands) to 300 mM sodium acetate.

4.4.2 Silanization of coverslips

To stabilize the polyacrylamide (PA) film, we coated our coverslips with a silane layer containing an allyl moiety which could be actively incorporated into PA gels during polymerization, covalently crosslinking the PA film to the coverslip. Coverslips were silanized using a modified version of a published protocol (179). Briefly, 40-mm-diameter, #1.5 coverslips (Bioptechs, 0420-0323-2) were washed for 30 min via immersion in a 1:1 mixture of 37% HCl and methanol at room temperature (RT). Coverslips were then rinsed three times in deionized
water and once in 70% ethanol. Coverslips were dried in a 70-°C oven and then immersed in 0.1% v/v triethylamine (Millipore, TX1200) and 0.2% v/v allyltrichlorosilane (Sigma, 107778) in chloroform for 30 min at RT. Coverslips were washed once each with chloroform and ethanol and then baked in a 70-°C oven for 1 hour to dehydrate the silane layer. Silanized coverslips could then be stored at room temperature (RT) in a desiccated chamber for weeks with no obvious reduction in the quality of the silane layer.

4.4.3 Cell culture and fixation

To promote cell adhesion, silanized coverslips were coated with 0.1 mg/mL poly-D-lysine (PDL) (molecular weight 30,000-70,000 Da; Sigma, P7886) diluted in nuclease-free water for 1 hour at RT. Coverslips were washed three times with nuclease-free water, incubated in water at RT overnight, and then dried and UV sterilized prior to plating cells.

U-2 OS cells were cultured, fixed, and permeabilized using protocols previously described (135, 154), before staining with encoding probes. Briefly, cells cultured with Eagle’s minimum essential medium (American Type Culture Collection, 30-2003) containing 10% v/v fetal bovine serum (ThermoFisher, 10437) were plated on PDL-coated, silanized coverslips at a density of 300,000 cells per coverslip and incubated at 37 °C with 5% CO₂ for 48 to 72 hours before fixing with 4% paraformaldehyde (PFA; Electron Microscopy Sciences, 15714) in 1×phosphate buffered solution (PBS; ThermoFisher, AM9625) for 20 min. Cells were washed three times with 1×PBS, and permeabilized using 0.5% v/v Triton X-100 (Sigma, T8787) in 1×PBS for 10 min at RT. Cells were then washed three times with 1×PBS.

4.4.4 Encoding probe staining
Encoding probe staining was performed as described previously (63, 135, 154). Briefly, cells were incubated for 5 min in a 30% formamide wash buffer, containing 2×saline-sodium citrate (SSC; ThermoFisher, AM9763) and 30% v/v formamide (ThermoFisher, AM9342) and then stained with encoding probes in encoding hybridization buffer, containing 2×SSC, 30% v/v formamide, 0.1% w/v yeast tRNA (Life technologies, 15401-011), 1% v/v murine RNase inhibitor (NEB, M0314L), and 10% w/v dextran sulfate (Sigma, D8906), in a humidity-controlled 37°C incubator for 36 to 48 h. Encoding probes were stained at a concentration of 100 µM unless otherwise specified. Where appropriate, the encoding probes were supplemented with 1 µM of anchor probe—a 15-nt sequence of alternating dT and thymidine-locked nucleic acid (dT+) with a 5′-acrydite modification (Integrated DNA Technologies). After staining, cells were washed two times for 30 min each with 30% formamide wash buffer at 47 °C.

Human lung fibroblast cells (American Type Culture Collection, IMR-90) were cultured, fixed, and stained following the same protocols described above for U-2 OS cells using 1 µM of a smFISH probe set targeting Filamin A (FLNA, Biosearch) described previously (154).

4.4.5 Sample embedding and clearing

In order to anchor RNAs in place, the encoding-probe-stained samples were embedded in thin, 4% PA gels. Briefly, stained samples on coverslips were first washed for 2 min with a degassed PA solution, consisting of 4% v/v of 19:1 acrylamide/bis-acrylamide (BioRad, 1610144), 60 mM Tris-HCl pH 8 (ThermoFisher, AM9856), 0.3 M NaCl (ThermoFisher, AM9759), and either a 1:500 dilution of 0.1-µm-diameter light-yellow beads (Spherotech, FP-0245-2) when samples were used for four-color MERFISH measurements or a 1:200,000 dilution of 0.1-µm-diameter carboxylate-modified orange fluorescent beads (Life Technologies, F-8800) when samples were used for two-color MERFISH measurements. The beads served as fiducial markers.
for the alignment of images taken across multiple rounds of smFISH imaging. Cells were then washed again for 2 min with the same PA gel solution supplemented with the polymerizing agents ammonium persulfate (Sigma, A3678) and N,N,N′,N′-Tetramethylethylenediamine (TEMED; Sigma, T9281) at final concentrations of 0.3% w/v and 0.15% v/v, respectively.

To cast a thin PA film, 50 µL of this gel solution was added to the surface of a glass plate (TED Pella, 26005) that had been pre-treated for 5 min with 1 mL GelSlick (Lonza, 50640) so as not to stick to PA. The sample on the coverslip treated as described above was aspirated to remove excess PA gel solution, then gently inverted onto this 50-µL droplet to form a thin layer of PA between the coverslip and the glass plate. The volume of this gel droplet could be used to control the thickness of this PA film. The gel was then allowed to cast for 1 ½ hours at RT. The coverslip and the glass plate were then gently separated, and the PA film washed twice with a digestion buffer consisting of 0.8 M guanidine-HCl (Sigma, G3272), 50 mM Tris-HCl pH 8, 1 mM EDTA, and 0.5% v/v Triton X-100 in nuclease-free water. After the final wash, the gel was covered with digestion buffer supplemented with 1% v/v proteinase K (NEB, P8107S). The sample was digested in this buffer for 16 to 20 hours in a humidified, 37-°C incubator and then washed with 2×SSC three times. MERFISH measurements were either performed immediately or the sample was stored in 2×SSC supplemented with 0.1% v/v murine RNase inhibitor at 4 °C for no longer than 24 h.

4.4.6 Imaging platforms

Cultured-cell samples were imaged on a home-built imaging platform with minor modifications from that described previously (63, 135). Briefly, this microscope was built using an Olympus IX-71 body and a 1.45 NA, 100× oil-immersion objective. Illumination in 750 nm, 641 nm, 561 nm, and 488 nm were provided using solid-state lasers (MPB communications,
VFL-P500-751; MPB communications, VFL-P500-642; Coherent, 561-200CWCDRH; and Coherent, 1069413/AT) for excitation of readout probes labeled with Alexa750, Cy5, ATTO565 and Alexa488, respectively. For two-color MERFISH measurements using Alexa750-labeled and Cy5-labeled readout probes, the 561-nm laser was used to excite the orange fiducial beads. A 405-nm solid-state laser (Coherent, Cube) was used to excite the nuclear stain 4’, 6-diamidino-2-phenylindole, dihydrochloride (DAPI), where appropriate. For four-color MERFISH measurements using Alexa750-labeled, Cy5-labeled, ATTO565-labeled and Alexa488-labeled readout probes, the 405 nm light was also used to excite the light-yellow fiducial beads. All laser lines were combined with a custom dichroic (Chroma, zy405/488/561/647/752RP-UF1), and the emission was filtered with a custom dichroic (Chroma, ZET405/488/561/647-656/752m). Fluorescence was separated with a custom penta-notch filter and imaged with an EMCCD camera (Andor, iXon-897). The pixel size for the EMCCD camera was determined to correspond to 167 nm in the sample plane. Each FOV was imaged at a single z plane with a 100 ms exposure for each color channel.

Tissue slices were imaged on a second home-built high-throughput imaging platform as described previously (154). Briefly, this microscope was constructed around an Olympus IX-71 microscope body and a PlanApo, 1.3 NA, 60× silicone-oil-immersion objective (Olympus, UPLSAPO 60×S2). Illumination in 754 nm, 647 nm, 561 nm, and 405 nm was provided using solid-state lasers (Toptica, DL100/BoosTA; MBP Communications, F-04306-113; Crystalaser GCL-150-561; Coherent, Cube 405). These laser lines were used to excite readout probes labeled with Alexa750 and Cy5, orange fiducial beads, and DAPI, respectively. The illumination profile was flattened with a square multi-mode fiber (Andor, Borealis). The fluorescence emission from the sample was separated from the laser illumination using a penta-band dichroic (Chroma,
zy405/488/561/647/752RP-UF1) and imaged using a scientific CMOS camera (sCMOS; Andor, Zyla 4.2) after passing through two duplicate custom penta-notch filters (Chroma, ZET405/488/561/647-656/752m) to remove stray excitation light. The pixel size for the sCMOS camera was determined to correspond to 109.2 nm in the sample plane. During the imaging of tissue, z-stacks consisting of seven, 1.5-µm-thick optical sections were collected in each color channel at each field-of-view (FOV) so as to image the entire volume of the tissue. Each exposure was 500 ms. The z-steps were controlled via an objective nanopositioner (Mad City Labs, NanoF200).

On both setups, sample position was controlled via a motorized microscope stage (Marzhauser, SCAN IM 112×74) and focus was maintained via a custom focus-lock system, realized through a feedback system between an objective nanopositioner (Mad City Labs, NanoF200) and the reflection of an IR laser (Thorlabs, LP980-SF15) onto an inexpensive CMOS camera (Thorlabs, uc480). The sample coverslip was held inside a flow chamber (Bioptechs, FCS2), and buffer exchange within this chamber was directed using a custom-built automated fluidics system described previously (63, 135, 154), controlling three eight-way valves (Hamilton, MVP and HVXM 8-5) and a peristaltic pump (Gilison, Minipuls 3).

4.4.7 MERFISH imaging

Samples were hybridized with readout probes and imaged following protocols similar to those previously described (154), with slight adjustments to readout hybridization buffer composition and flow times. Readout hybridization buffer was composed of 2×SSC, 10% v/v ethylene carbonate (EC; Sigma-Aldrich, E26258), 0.1% v/v murine RNase inhibitor in nuclease-free water, and 3 nM of the appropriate readout probes. Previously we utilized dextran sulfate in this buffer to increase the rate of readout probe hybridization; however, we have found that the
same hybridization kinetics can be achieved without dextran sulfate by increasing the readout probe concentrations from 1 nM to 3 nM. Removing dextran sulfate from the readout buffer dramatically reduced the viscosity of this buffer, and this reduction in buffer viscosity, in turn, effectively eliminated the occasional flow failures that arose from the high pressures required to pull high viscosity buffers through the fluidics system.

Two different configurations of readout probes were utilized: for two-color MERFISH measurements, two readout probes, one conjugated to Cy5 and the other to Alexa750 via a disulfide bond were used in each round of hybridization; and for four-color MERFISH four different readout probes each conjugated to one of Alexa750, Cy5, ATTO565, or Alexa488 via a disulfide bond were used in each round of hybridization. Appendix B contains the readout probe sequences and dye combinations used for both two- and four-color measurements. All readout probes were purchased from Biosynthesis, Inc.

The sample was stained with readout probes by first flushing the sample chamber with 2 mL of readout hybridization buffer over the span of 5 min to fully exchange buffers. Then an additional 2 mL of buffer was flown across the sample for 6 min. The sample was then washed by flowing 2 mL of readout wash buffer, containing 2×SSC and 10% v/v EC, over a span of 9 min. Finally, 2 mL of imaging buffer, containing 2×SSC, 50 mM Tris-HCl pH 8, 10% w/v glucose, 2 mM Trolox (Sigma-Aldrich, 238813), 0.5 mg/mL glucose oxidase (Sigma-Aldrich, G2133), 40 μg/mL catalase (Sigma-Aldrich, C30) and 0.1% v/v murine RNase inhibitor, was flown across the sample for 6 min, after which the flow was halted and ~400 FOVs were imaged. This imaging buffer was used to decrease the effect of photobleaching (180). Imaging buffer was stored under a layer of mineral oil (Sigma-Aldrich, 330779) throughout the measurement as a barrier against oxygen. Because glucose oxidase was determined to contain trace amounts of
RNase, the imaging buffer also contained 0.1% v/v murine RNase inhibitor. We replaced ribonucleoside vanadyl complex (VRC; NEB, S1402S), which was used previously (63, 135), with Murine RNase inhibitor because the citrate in 2×SSC was found to significantly reduce the effective lifetime of VRC.

After each round of imaging, the fluorescent dyes were removed from readout probes by reductive cleavage of the disulfide bond conjugating these dyes to the probes. 3 mL of cleavage buffer comprising 2×SSC and 50 mM of the reducing agent Tris(2-carboxyethyl)phosphine (TCEP; Sigma, 646547) was flown across the sample over the course of 15 min. After cleavage, the chamber was flushed with 2 mL of 2×SSC for 4 min to flush any residual cleavage buffer from the sample prior to the introduction of the subsequent hybridization buffer. All buffers were freshly prepared before each experiment using nuclease-free water.

After the final round of hybridization and imaging, the sample was stained with DAPI at a concentration of 1 μg/mL in 2×SSC for 10 min to mark nuclei, and then imaged at 405 nm.

4.4.8 Image registration and decoding

Registration of images of the same FOV across imaging rounds as well as decoding of the RNA barcodes was conducted using a previously described analysis pipeline (154). Briefly, the locations of the fiducial beads in each round of imaging were found via a Gaussian fitting routine (117), and these locations were used to create affine transformations that correct offsets between images in each imaging round. Additional corrections to account for minor chromatic aberrations were not applied because the offsets in the centroid of RNAs labeled simultaneously with Alexa750, Cy5, ATTO565, and Alexa488 were not substantial. Images were then high-pass filtered to remove background, deconvolved to tighten RNA spots, and then low-pass filtered so
as to connect RNA centroids that differ slightly in location between images, a property that we have previously observed (63). Individual pixels were then assigned to barcodes by comparing the intensity of each pixel across the 16 images collected across all color channels and all hybridization rounds (2 color channels in 8 hybridization rounds or 4 color channels in 4 hybridization rounds) to each of the different barcodes. Specifically, the set of 16 intensities for each pixel derived from each of the 16 images were used to define a vector that was normalized to unitary magnitude, i.e. by dividing by the L₂ norm. A unit vector was similarly defined for each of the 140 barcodes. The Euclidean distance was then calculated between each pixel vector and each of the barcode vectors. A pixel was assigned to a barcode if the Euclidean distance separating it from a barcode was smaller than a given threshold. This distance threshold was determined from the largest Euclidean distance between each normalized barcode and the set of normalized barcodes formed from all single-bit errors to that barcode. Conceptually, this distance defines a 16-dimensional sphere that contains all possible modifications to a barcode that correspond to a single-bit error to that barcode, and this decoding approach can be thought of as assigning pixels to a given barcode based on whether they fall within this sphere for a given barcode. Pixels with vectors that do not fall within one of these 140 spheres are left unassigned. Contiguous pixels assigned to the same barcode were combined to form a single RNA. Each RNA was then identified as requiring error correction (or not) by comparing the average pixel vector across all pixels assigned to that RNA to the set of unitary vectors defined by all single-bit errors to the assigned barcode. If the average pixel vector was closer to a vector corresponding to a single-bit error than it was to the correct barcode, the RNA was marked as requiring error correction.
This decoding approach assumes that the brightness of each RNA spot is identical between imaging rounds. To correct for differences in the brightness between color channels, images were initially normalized by equalizing their intensity histograms, as described previously (154). This normalization was then refined via an iterative process, again as described previously (154). A background of spurious RNAs were removed with thresholds on the brightness of the RNA, i.e. the $L_2$ norm of the pixel vector, and the number of pixels combined to form that RNA, i.e. its area, as described previously (154). For tissue imaging, this pipeline was modified to accommodate z-stacks. Because each z-stack was separated by a distance larger than the axial extent of the point-spread-function, each stack was decoded independently of the others. Nuclei were identified and counted via intensity thresholding of the DAPI images as described previously (154).

Computations were split between the Odyssey cluster supported by the FAS Division of Science, Research Computing Group at Harvard University and a desktop server that contained two 10-core Intel Xeon E5-2680 2.8 GHz CPUs and 256 GB of RAM.

### 4.4.9 RNase treatment

U-2 OS or IMR-90 samples were stained with smFISH probes or MERFISH encoding probes as described above in the sections “Cell culture and fixation” and “Encoding probe staining.” Samples were then stained with readout probe 1 (Appendix B) as in the “MERFISH imaging” section above. Samples were imaged, and then treated for 30 min with 1% v/v RNase A (Qiagen, 19101) in 2×SSC, and then reimaged.

### 4.4.10 Protein and lipid staining
U-2 OS cells were cultured, fixed, and labeled with smFISH probes or MERFISH encoding probes as described in the “Cell culture and fixation” and “Encoding probe staining” sections above. Samples were then either matrix imprinted and cleared as described in the “Sample embedding and clearing” section above or stored at 4 °C in 2×SSC. Cells were stained with a 1:10 dilution of Krypton Fluorescent Protein Stain (ThermoFisher, 46629) in 2×SSC at RT for 15 min and washed once in 2×SSC at RT for 15 min. 100 FOVs were imaged with the 561-nm laser. Samples for lipid staining were prepared in the same fashion but stained with a 1:200 dilution of Vybrant® DiD Cell-Labeling Solution (ThermoFisher, V22887) in 2×SSC at RT for 15 min and washed once briefly with 2×SSC. 100 FOVs were imaged with the 641-nm laser. Imaged samples were quantified by averaging the observed fluorescence across all FOVs, and this value was then averaged across three biological replicates.

4.4.11 Tissue preparation

MERFISH imaging in tissue was performed on 10-μm-thick cryosectioned mouse hypothalamus slices. Whole brain tissue was removed from mice (C57BI6/J) euthanized using CO₂, and immediately frozen in optimum cutting temperature compound (Tissue-Tek O.C.T.; VWR, 25608-930). Frozen blocks were coarsely sectioned to the hypothalamus region, trimmed to an area of roughly 3 mm × 3 mm, and sectioned at a thickness of 10 μm at −18 °C on a cryostat (MICROM, HM550). Sections were collected on silanized coverslips coated with PDL prepared following protocols described in the “Silanization of the coverslips” and “Cell culture and fixation” sections above. These sections were then immediately fixed in 4% PFA in 1×PBS for 12 min at RT and washed with 1×PBS for 5 min three times. The samples were then partially cleared by treating them with 4% w/v SDS in 1×PBS for 2 min with gentle agitation at RT. After
this treatment, samples were washed three times with 1×PBS for 5 min, and then immersed in 70% ethanol and stored at 4 °C for at least 18 h.

Tissue samples were then stained and cleared following the protocols described in the “Encoding probe staining” and “Sample embedding and clearing” sections above. Tissue samples were measured using 2-color MERFISH as described in the “MERFISH imaging” section above. Four cryosections were imaged in a single MERFISH experiment on the high-throughput imaging platform described in the “Imaging platforms” section above.
Chapter 5

Conclusion

Many complementary techniques are required to dissect complex biological functions driven by molecular interactions between DNA, RNA, and protein. Of all the tools available, imaging techniques provide a unique opportunity to observe these interactions occur in their native context. In particular, ever-higher resolution and degree of multiplexity has been the focus of technical advancement to enable understanding dense and complex tissues such as the brain.

This thesis is comprised of three projects, which separately focus on: 1) visualizing and characterizing the structure and dynamics of synaptic neurexin-1 nanoclusters using STORM, 2) increasing the throughput of MERFISH by two orders of magnitude to allow quantification of ~40,000 cells in a single 18-hour measurement, and 3) creating a sample-clearing protocol for MERFISH to significantly reduce background by removing protein and lipid, improve the quality of MERFISH measurements, and extend the technique into tissue, specifically in the mouse brain. Through these works, we aimed to simultaneously leverage the power of existing super-resolution imaging in addressing questions in fundamental neurobiology, as well as improve upon the recently introduced technique of multiplexed RNA imaging towards the eventual goal of spatially-resolved transcriptome profiling at the whole tissue scale in brains.

5.1 Characterization of synaptic neurexin-1 nanoclusters using super-resolution microscopy

Neurexins and related synaptic adhesion molecules have been widely studied for their role in directing and maintaining synaptic organization and function. Disruptions to the function
of neurexins have also been implicated in many neurological diseases. To characterize the organization and localization of synaptic neurexin-1, we applied STORM to visualize natively-expressed neurexin at the active zone in mouse neurons and brain tissue.

We reported the creation of conditional knock-in mice that express HA-tagged neurexin-1, which we then used to characterize endogenous neurexin-1 with high specificity. After labeling with HA-antibody, we imaged synaptic neurexin-1 in both cultured neurons and tissue sections, and found a novel organization of discrete nanoclusters at presynaptic terminals. We further find that neurexin-1 content in these nanoclusters to be dynamic, and partly regulated via cleavage of its extracellular domain by ADAM10. While only ~40% of Homer1(+) excitatory synapses contain neurexin-1 nanoclusters, pharmacologically inhibiting ADAM10 activity nearly doubles the fraction of synapses containing neurexin-1.

Neurexins purportedly play many roles in organizing synaptic function, and has many other interacting partners, both syn- and trans-synaptically. With newer methods of multiplexing super-resolution imaging, it would be interesting to view neurexin-1 nanoclusters and their dynamics in the context of other important players in synaptic terminal and cleft. It would also be of interest to further explore the mechanistic basis of nanocluster formation and regulation, as well as the relationship between diseases linked to neurexin-1 dysfunction and the organization of neurexin-1 at synaptic terminals.

5.2 Increasing throughput and performance of MERFISH towards achieving spatially resolved transcriptomics in brain tissue

Spatially-resolved transcriptomics and single-cell RNA sequencing are complementary techniques that together can provide profound insight into the complex network of cell states that
give rise to the emergent properties observed in whole tissues. This is particularly true in the brain, which is arguably one of the most difficult tissues to unravel, due to the sheer number of inputs and outputs from neurons and their synapses. Understanding the transcriptome of such a complex system requires both high numbers of cells, and given the complex spatial patterns of neuron types already known within the brain and its substructures, the preservation of spatial information is also especially important. Being able to describe the spatially-resolved transcriptome of the brain via MERFISH requires the ability to image volumes of tissue on the scale of multiple millimeters, a challenge that is further complicated by the difficulty of imaging in tissues arising from background signal.

Chapters three and four described our efforts to improve MERFISH to tackle the above problems hindering large-scale tissue imaging. First by using reductive cleavage as an alternative to photobleaching and significantly expanding our imaging field of view, we were able to dramatically cut the imaging time of MERFISH experiments and increase throughput by two orders of magnitude. To address the increase in background staining observed in brain tissue, we developed a clearing protocol tailored for MERFISH measurements. We demonstrate that by removing the majority of cellular proteins and lipids, we can improve MERFISH performance and significantly reduce fluorescence background due to off-target binding of RNA probes.

Spatially resolved transcriptomics offered the possibility of characterizing the expression profile of intact tissue supplemented by the contextual information of neighboring cells. By increasing throughput and decreasing background signal found in MERFISH, we extend its capabilities towards achieving the ability to performing measurements across entire tissue blocks, in particular in the brain. We also bring the throughput of MERFISH in line with, and in
some cases, beyond those achievable with single-cell RNA sequencing, allowing easier integration between these two complementary techniques.
Appendix A: Readout probe sequences.

The dye was attached to each readout probe via a disulfide bond at the 3’ end of the listed probe sequences.

<table>
<thead>
<tr>
<th>Bit</th>
<th>Readout probe name</th>
<th>Sequence</th>
<th>Dye</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RS0015</td>
<td>ATCCTCCTTCAATACATCCC</td>
<td>Cy5</td>
</tr>
<tr>
<td>2</td>
<td>RS0083</td>
<td>ACACTACCACCATTTCTAT</td>
<td>Alexa750</td>
</tr>
<tr>
<td>3</td>
<td>RS0095</td>
<td>ACTCCACTACTACTCACTCT</td>
<td>Alexa750</td>
</tr>
<tr>
<td>4</td>
<td>RS0109</td>
<td>ACCCTCTAACTTCCATACA</td>
<td>Cy5</td>
</tr>
<tr>
<td>5</td>
<td>RS0175</td>
<td>ACCACAACCCATTCTTTCA</td>
<td>Cy5</td>
</tr>
<tr>
<td>6</td>
<td>RS0237</td>
<td>TTCTACCACTAATCAACCC</td>
<td>Alexa750</td>
</tr>
<tr>
<td>7</td>
<td>RS0247</td>
<td>ACCCTTTACAACACACACCC</td>
<td>Cy5</td>
</tr>
<tr>
<td>8</td>
<td>RS0255</td>
<td>TCCTATTCTCAACCTAACC</td>
<td>Alexa750</td>
</tr>
<tr>
<td>9</td>
<td>RS0307</td>
<td>TATCTTTCAATCCCTCCAC</td>
<td>Alexa750</td>
</tr>
<tr>
<td>10</td>
<td>RS0332</td>
<td>ACATTACACCTCATTCTCC</td>
<td>Cy5</td>
</tr>
<tr>
<td>11</td>
<td>RS0343</td>
<td>TTACTCCCTACACCTCAA</td>
<td>Cy5</td>
</tr>
<tr>
<td>12</td>
<td>RS0384</td>
<td>TTCTCCCTCTATCAACTCTA</td>
<td>Alexa750</td>
</tr>
<tr>
<td>13</td>
<td>RS0406</td>
<td>ACCCTTACTACTACATCTC</td>
<td>Cy5</td>
</tr>
<tr>
<td>14</td>
<td>RS0451</td>
<td>TCCTAACAACCAACTACTCC</td>
<td>Alexa750</td>
</tr>
<tr>
<td>15</td>
<td>RS0468</td>
<td>TCTATCATTACCCTCTCTCT</td>
<td>Alexa750</td>
</tr>
<tr>
<td>16</td>
<td>RS0548</td>
<td>TATTCACCTCTAAACCCTC</td>
<td>Cy5</td>
</tr>
</tbody>
</table>
Appendix B: Readout dye specification for 2- and 4-color measurements

The dye was attached to each readout probe via a disulfide bond at the 3’ end of the listed probe sequence.

<table>
<thead>
<tr>
<th>Bit</th>
<th>Readout probe name</th>
<th>Sequence</th>
<th>Dye (2-color MERFISH)</th>
<th>Dye (4-color MERFISH)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RS0015</td>
<td>ATCCTCCTTCAATACATCCC</td>
<td>Cy5</td>
<td>Cy5</td>
</tr>
<tr>
<td>2</td>
<td>RS0083</td>
<td>ACACTACCACCATTTCTAT</td>
<td>Alexa750</td>
<td>Alexa750</td>
</tr>
<tr>
<td>3</td>
<td>RS0095</td>
<td>ACTCCACTACTACTCACTCT</td>
<td>Alexa750</td>
<td>ATTO565</td>
</tr>
<tr>
<td>4</td>
<td>RS0109</td>
<td>ACCCTCTAACTTCCATCACA</td>
<td>Cy5</td>
<td>Alexa488</td>
</tr>
<tr>
<td>5</td>
<td>RS0175</td>
<td>ACCACAACCCATTCCTTTC</td>
<td>Cy5</td>
<td>Cy5</td>
</tr>
<tr>
<td>6</td>
<td>RS0237</td>
<td>TTTCTACCTAATACACCC</td>
<td>Alexa750</td>
<td>Alexa750</td>
</tr>
<tr>
<td>7</td>
<td>RS0247</td>
<td>ACCCTTTCAAAACACACCT</td>
<td>Cy5</td>
<td>Alexa488</td>
</tr>
<tr>
<td>8</td>
<td>RS0255</td>
<td>TCCTATTCTCAACCTAACC</td>
<td>Alexa750</td>
<td>ATTO565</td>
</tr>
<tr>
<td>9</td>
<td>RS0307</td>
<td>TATCCTTCAATCCCTCCACA</td>
<td>Alexa750</td>
<td>Alexa750</td>
</tr>
<tr>
<td>10</td>
<td>RS0332</td>
<td>ACATTACACCTCATTCTCC</td>
<td>Cy5</td>
<td>Cy5</td>
</tr>
<tr>
<td>11</td>
<td>RS0343</td>
<td>TTTACTCCCTACACCTCCA</td>
<td>Cy5</td>
<td>ATTO565</td>
</tr>
<tr>
<td>12</td>
<td>RS0384</td>
<td>TTCTCCCTCTATCAACTTA</td>
<td>Alexa750</td>
<td>Alexa488</td>
</tr>
<tr>
<td>13</td>
<td>RS0406</td>
<td>ACCCTTACTACTACATCATC</td>
<td>Cy5</td>
<td>Cy5</td>
</tr>
<tr>
<td>14</td>
<td>RS0451</td>
<td>TCCTAACAACCAACTACTCC</td>
<td>Alexa750</td>
<td>Alexa750</td>
</tr>
<tr>
<td>15</td>
<td>RS0468</td>
<td>TCTATCATTACCTCCTCCTC</td>
<td>Alexa750</td>
<td>ATTO565</td>
</tr>
<tr>
<td>16</td>
<td>RS0548</td>
<td>TATTCACCTTACAAACCCTC</td>
<td>Cy5</td>
<td>Alexa488</td>
</tr>
</tbody>
</table>
Bibliography


156


121. A. K. Shalek et al., Single-cell transcriptomics reveals bimodality in expression and


160. B. Yang et al., Single-Cell Phenotyping within Transparent Intact Tissue through Whole-


179. A. Buxboim, K. Rajagopal, A. E. X. Brown, D. E. Discher, How deeply cells feel: