MERFISH Workshop Day 1

Robert Foreman
UCLA Collaboratory
Feb 2020

What is MERFISH and why is everyone here?

1. Way to measure 100s of gene expression values with spatial preservation and single-cell resolution
2. Why is everyone here interested in MERFISH
    a. Have you done smFISH before?
        i. Was it easy? Did you face any particular challenges?
3. How would you rate your preparedness for each of the skills of MERFISH?
    a. Bioinformatics
        i. Gene selection
        ii. Oligo design
    b. Molecular Biology
        i. Oligo Amplification
    c. Fluorescent Microscopy
    d. Data Analysis
        i. Image analysis
        ii. Single-cell anaylsis

Goals of workshop

Provide overview of all the steps that go into MERFISH experiments

- After the workshop everyone should at least know what the challenges will be, what they can do in house in their own labs, and what areas they're need to find collaboration
- Practical workshop in the design of MERFISH experiments
  a. If you have a project in mind you should be able to select genes and design oligo sets for the genes
- Detailed overview of the steps and challenge points of each step

Outline of workshop schedule

Day 1

1. Detailed Overview of MERFISH
2. Gene Selection
3. Oligo Design
4. Oligo Amplification

Day 2

1. Sample Preparation
2. Sample Imaging
3. Automated Serial Hybridization
4. Microscope Requirement

Day 3

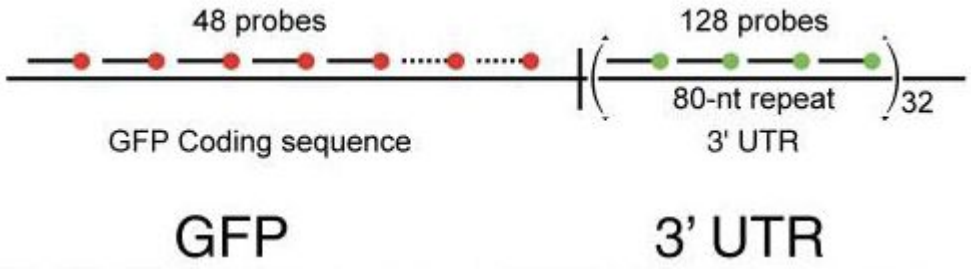
1. Image Processing
2. Spot Calling
3. Cell Segmentation
4. Counts Per Cell Analysis
    a. Single-cell gene expression analysis

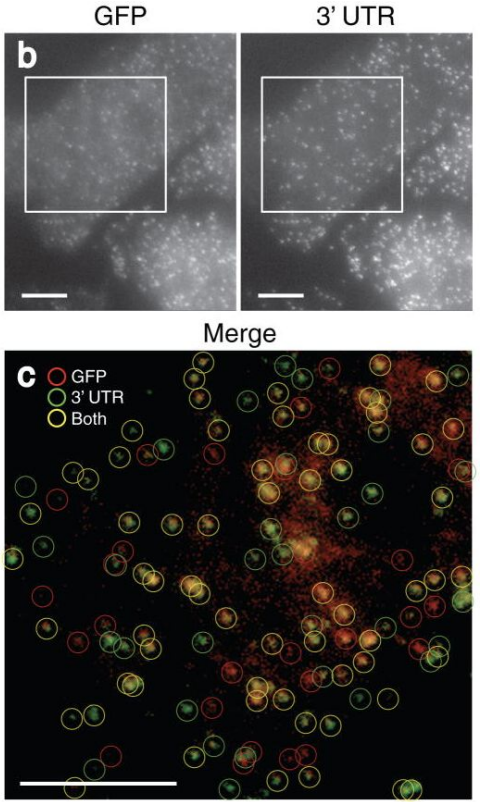
History of smFISH and development of multiplexed smFISH technologies

For a decade or so gene expression measurements using smFISH (single-molecule fluorescent in-situ hybridization) has been limited to the measurement of 3-5 genes per experiment until new approaches were used to barcode genes.

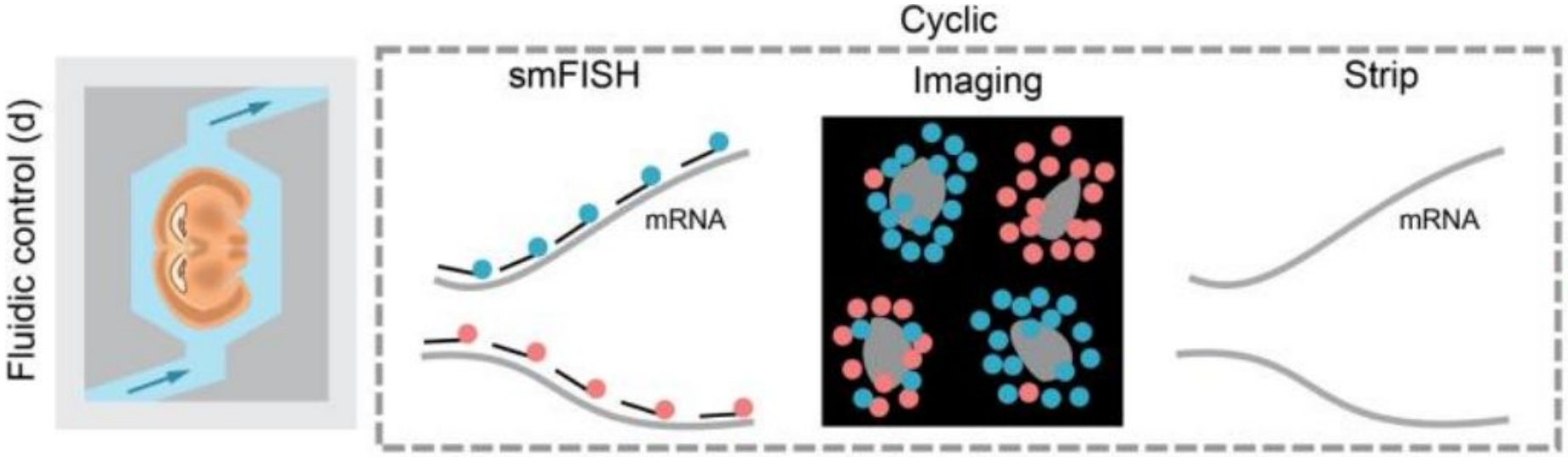
1. Standard smFISH
2. Multiple Hybridizations
3. Spectral Barcoding
4. MERFISH/seqFISH

smFISH - Raj *et al* 2008

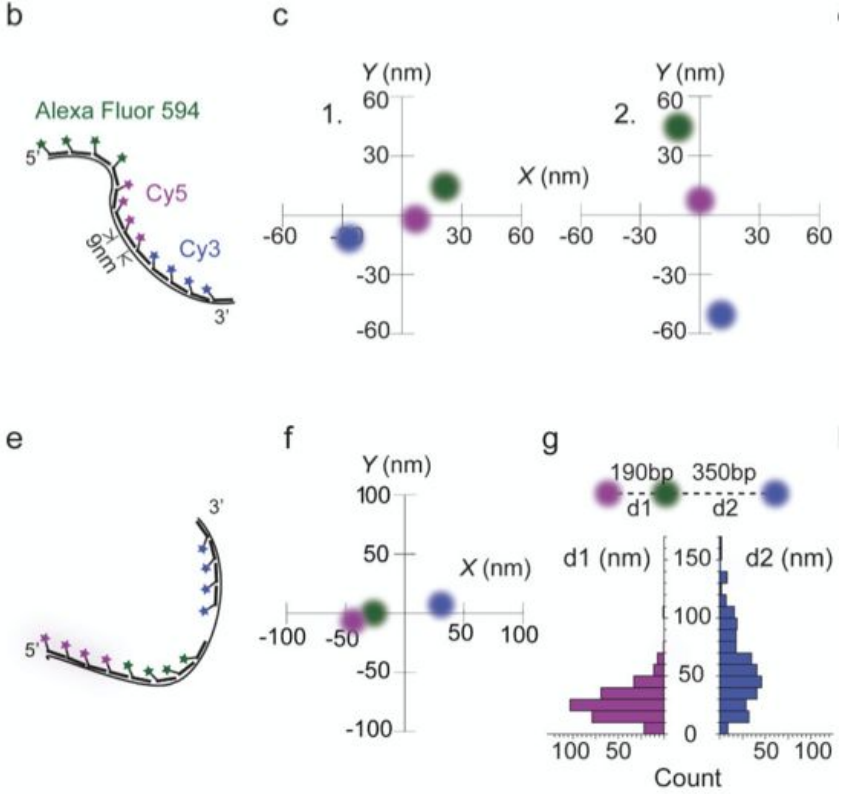


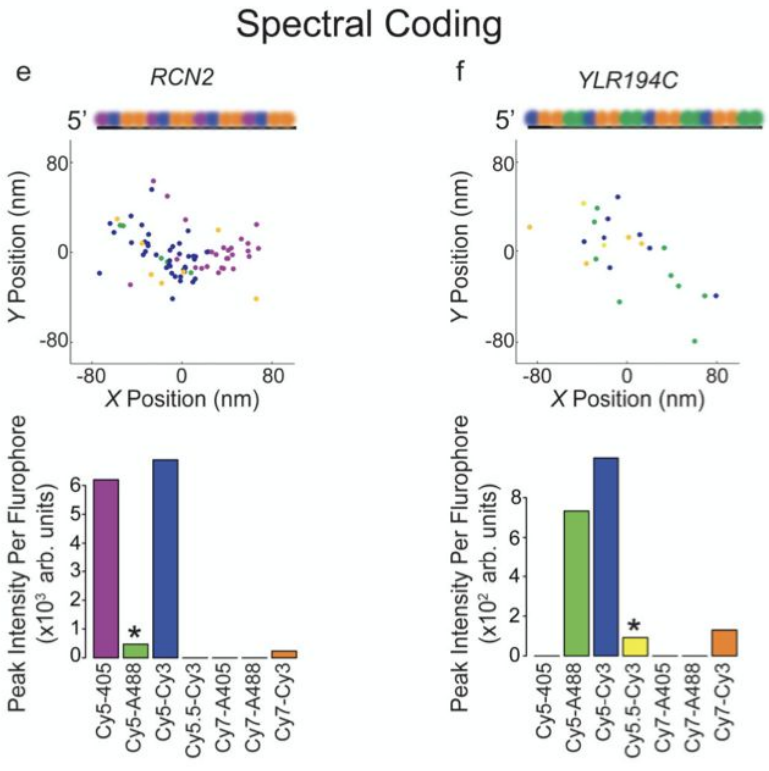


Multiple Hybridization smFISH



Spatial and spectral barcoding

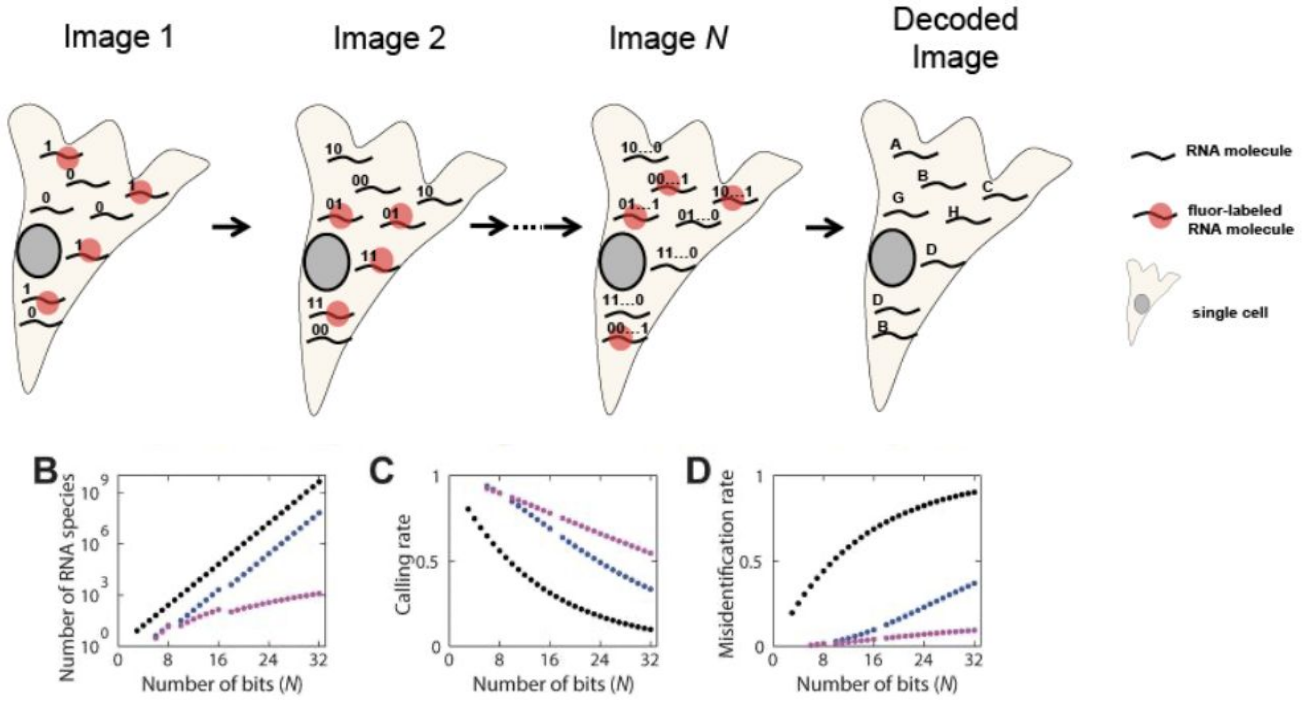




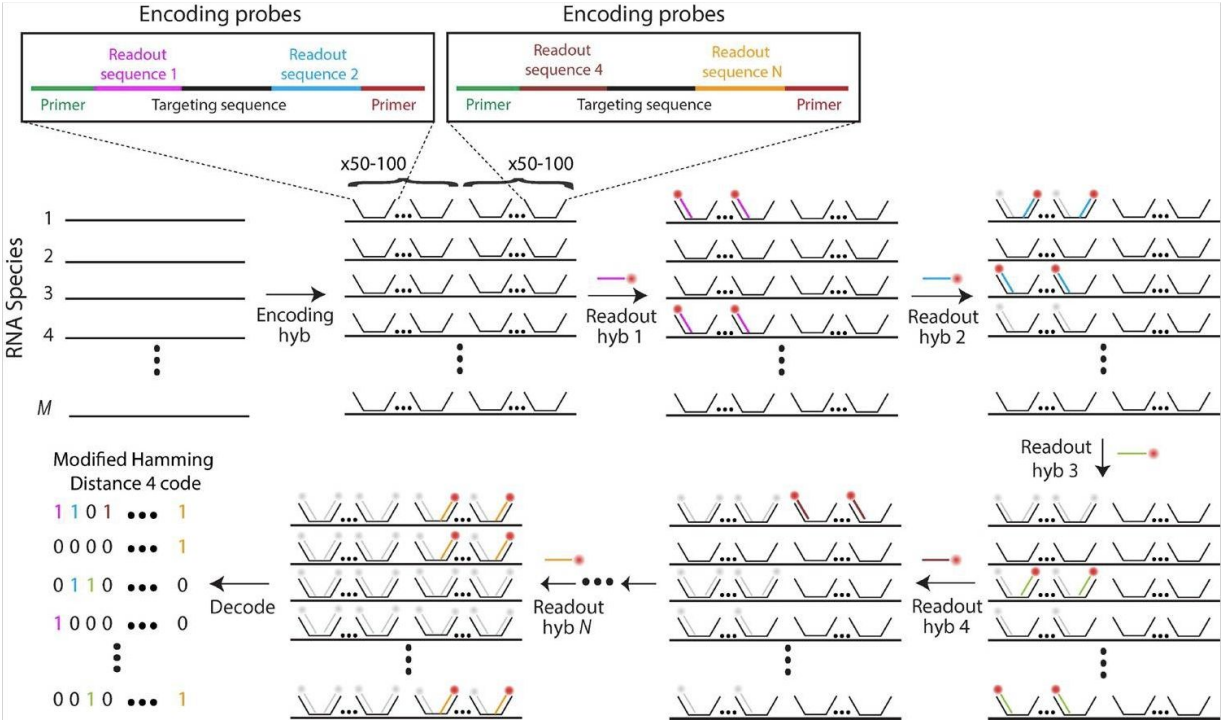
Long cai's spatial/spectral scaling

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

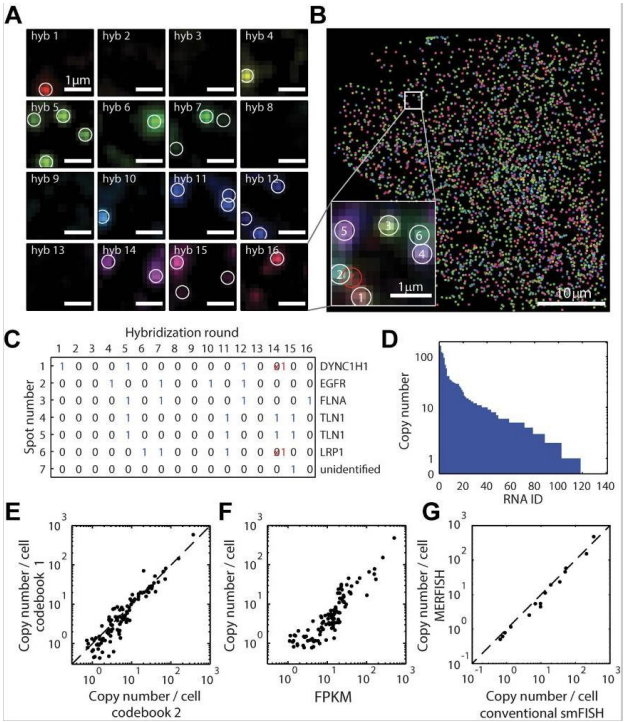
MERFISH and seqFISH - substituting time for space in RNA identity encoding scheme



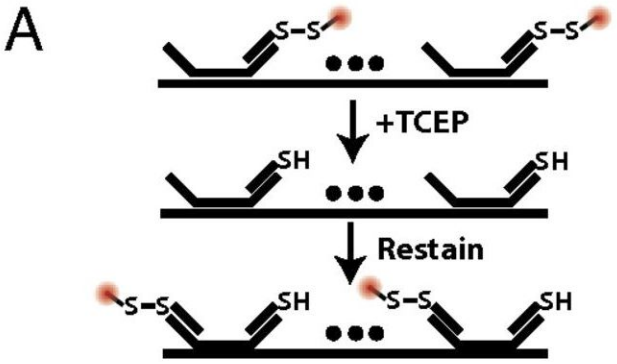
MERFISH and seqFISH - substituting time for space in RNA identity encoding scheme

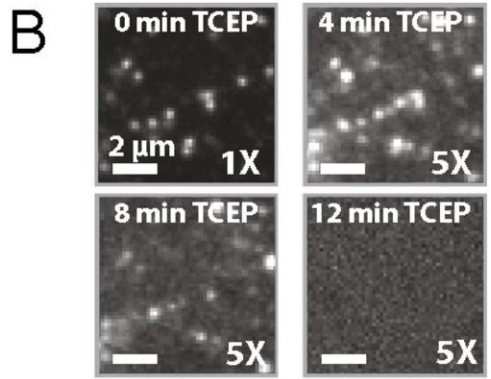


MERFISH and seqFISH - substituting time for space in RNA identity encoding scheme

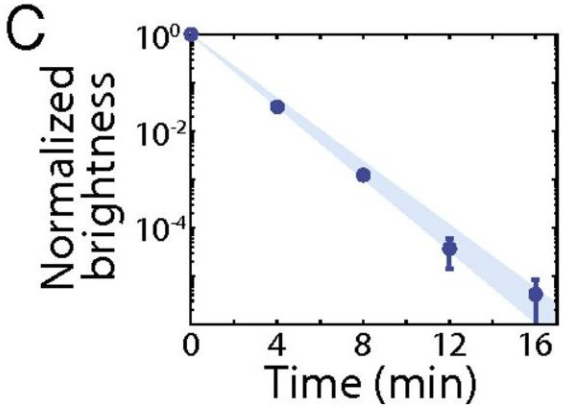


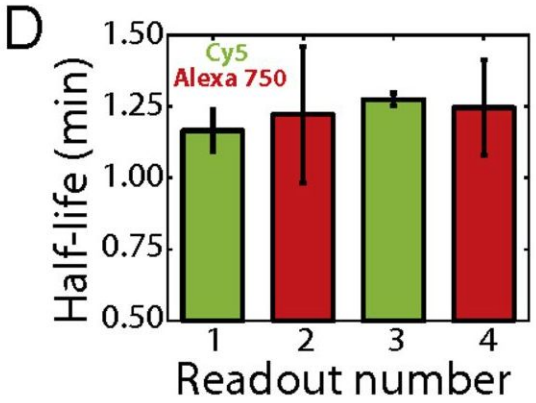
High performance MERFISH (V2)



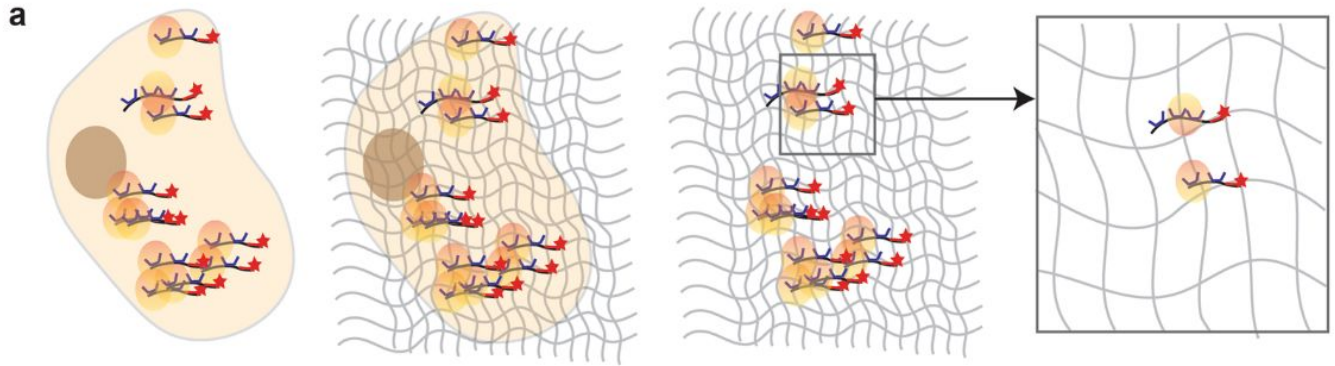


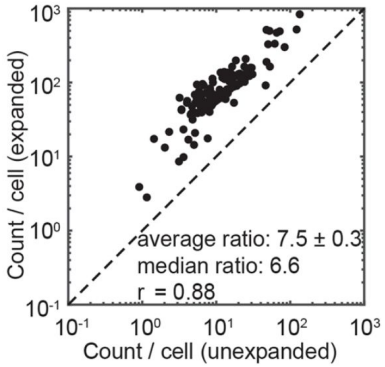
And sample clearing that we will talk about in more detail tomorrow

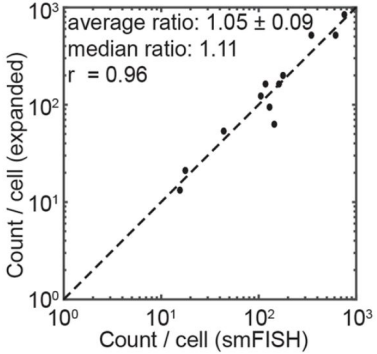




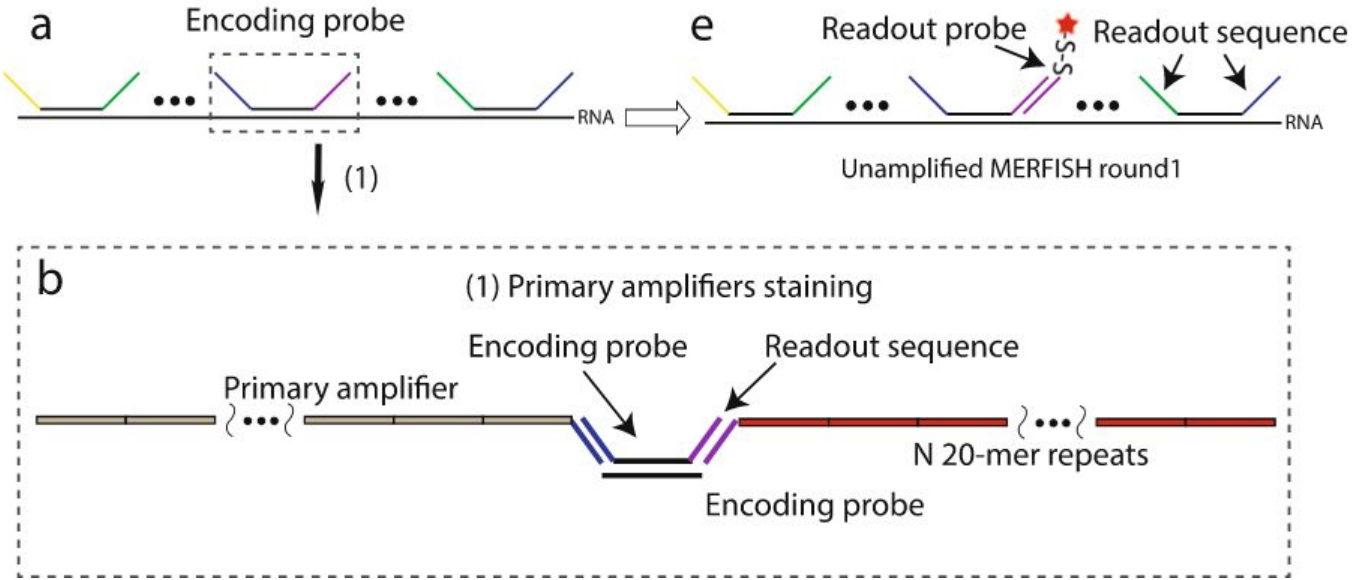
MERFISH V3 - expansion microscopy



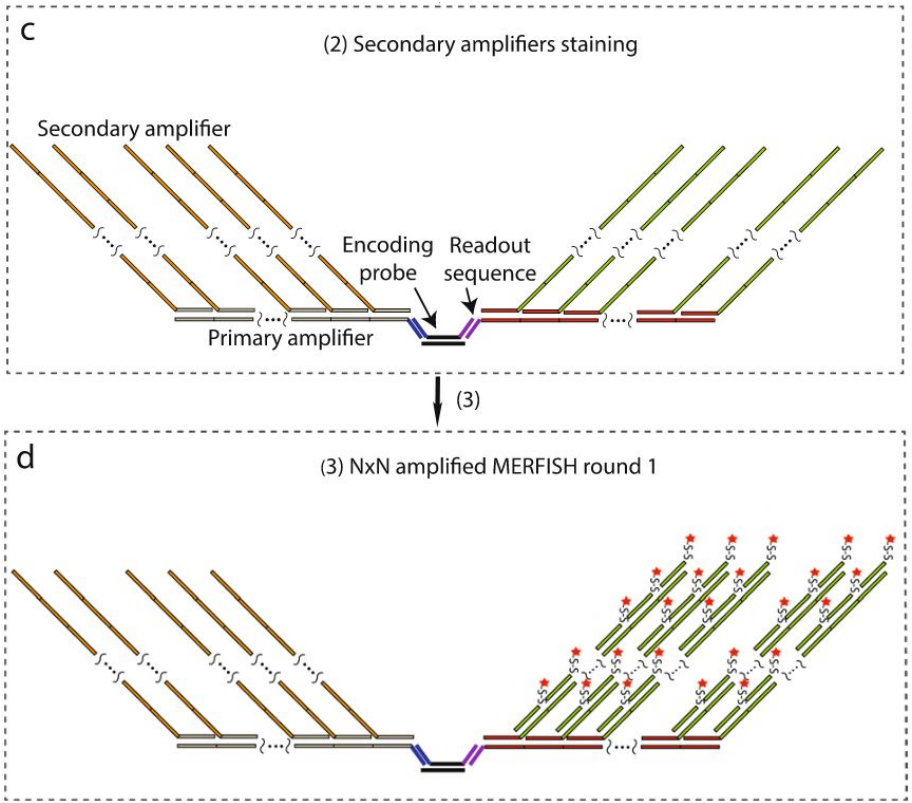




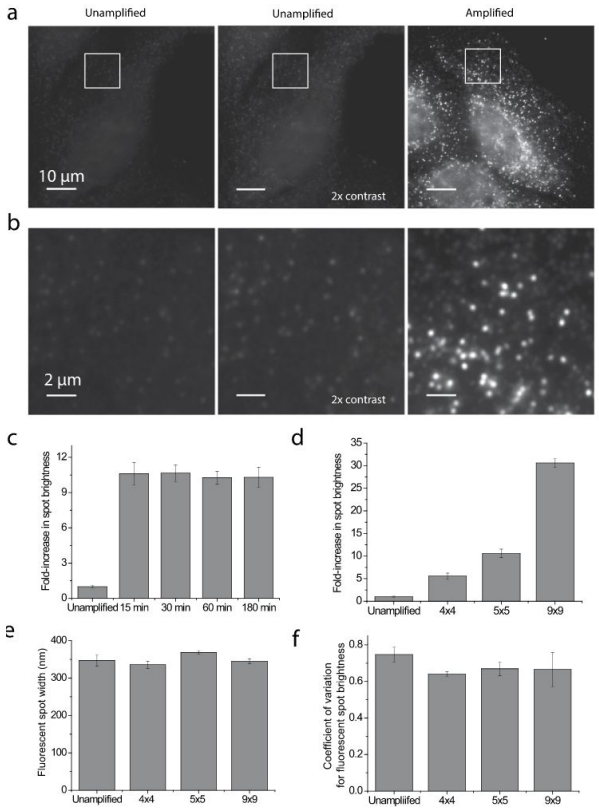
MERFISH V4 - amplified signal



MERFISH V4 - amplified signal



MERFISH V4 - amplified signal



Comparison to droplet based RNA-Seq

Droplet Based	MERFISH
<0.05 to 0.2 detection probability (dropout)	0.75-0.9+ detection sensitivity
No prior knowledge necessary	Prior knowledge of genes of interest
Scales to 100k to 1M+ cells	Scales to 10,000s to 100,000s cells per sample
Loss of spatial information	Preserves spatial information

Why are you interested in MERFISH over scRNA-Seq??

Part 2

Designing a MERFISH experiment - selecting a set of genes to measure

Getting a set of genes

Informed by RNA-Seq - know certain genes are DE and want to see what cell types or area of a tissue

scRNA-Seq - know single-cell expression patterns and want more accurate measurement and/or where the cells with single-cell DE are localizing

Targeted Biology - interested in calcium signaling, or immune response

Targeted biology resources

KEGG

Reactome

Literature Mining

scRNA-Seq data

Differential tissue expression

What do ya’ll know about?

Steps of oligo design

1. Generate gene list
    a. Select isoforms to target
    b. Find RNA expression (not necessary but helps)
2. Tile across the gene creating candidate oligos
3. Screen candidate oligos
    a. GC Content
    b. Forbidden sequences (GGGG, TTTTT+, AAAAA+, CCCCC+)
    c. Secondary Structure
    d. Off target binding (oligos not specific to target)
        i. Gene level non-specificity
        ii. Isoform level non-specificity
        iii. Ribosomal and tRNA non-specificity
4. Create oligo fusion
5. Screen final oligo for off-targets and homology to other oligos in the oligo pool

Generate gene and transcript list

NCBI (RefSeq) and Ensembl

Ensembl is more comprehensive in terms of gene annotations (not always a good thing if there is low support for the alternative isoforms)

Basic strategy - go with the ‘top refseq’ or use expression data (your own or public from GEO) to determine the dominant isoform

- Helpful for selecting isoforms MANE transcripts and APPRIS Annotation

Alternative to isoform selection

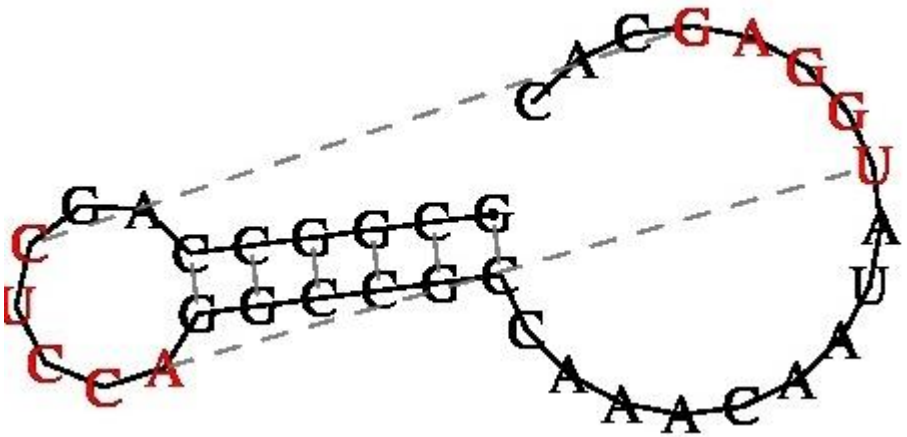
Use conservation/shared exons between alternative isoforms to design oligos against these

- Not currently implemented in the oligo selection software we use, but it wouldn’t be a super crazy thing to implement
- Biggest challenge would be figuring out how to parse/munge input files into a form that the current software can use (discuss more in future slides)

DNA Secondary structure screening

Limits effective concentration of Oligo

Affects the specificity of the oligo



Forbidden sequences

Homopolymers tend to be repeated in the genome and long strings of homopolymers often create issues with synthesize/polymerases

GGGG forms a tertiary structure called a G-quadruplex that can strongly inhibit amplification

Bioinformatics of oligo design

Let’s say you have 100 genes each ~3000bp and your candidate oligos are 30bp of RNA complementarity

3000bp ~ 3000 candidate oligos (sliding window of 30bp accross the gene) * 100 genes = 300k candidate oligos being screened for specificity in ~30k to 200k annotated transcripts in the human genome

There are very good sequence tools and it is actually not crazy to consider doing this in ‘brute force way’, but the current software from Zhuang lab uses k-mer approach to increase speed of oligo design

Offtarget binding screening (k-mer approach)

Parse the transcriptome into a set of 17bp k-mer's (substrings of length k)

For each 17bp kmer store which genes and isoforms it is contained by

For each candidate_oligo check all it's 17bp kmers against the database and enumerate the number of off target hits

If RNA-seq data available transform offtarget hits to expected number of off-target RNA molecules and estimate the fraction of ontarget to total off target

Offtarget k-mer approach

Two parameters

Gene Specificity Fraction (0-1) Default 0.75

Isoform Specificity Fraction (0-1) Default 0.6

Intuition is fraction of ‘spots’ that would be from the true target

- I.e. TP/(TP+FP)

Probably tends to overestimate (is more stringent) the problem of off target binding because a 17bp substring is not likely to hybridize with the same probability as a 30bp ontarget region of homology between oligo and target gene

Zhuang Lab MERFISH oligo design software (MATLAB)

Input Files

1. Transcriptome FASTA file
2. Non-coding RNA FASTA
3. Ribosomal and tRNA FASTA
4. Sequences of Readout Probes
5. Codebook csv

Parameters

1. Non-coding cutoff length (15)
2. Low-abundance cutoff (0.01)
3. Probes per gene (92)
4. Gene specificity (0.75-1)
5. Isoform specificty (0.5-1)
6. Encoding Olio Overlap (-15bp)
7. Encoding Length (30bp)
8. GC (43-63%)
9. T_m (66-75C)

Input Files

1. Transcriptome FASTA file
   a. Ensembl or NCBI RefSeq
2. Non-coding RNA FASTA
3. Ribosomal and tRNA FASTA
4. Sequences of Readout Probes
   a. We have 18bit readout sequences that are defined sequences with S-S linked fluorophores
5. Codebook csv

```
version    1
codebook_name    MHD4_18bit_187cwords
bit_names    RS0109    RS0175
name    id    barcode
CAMK4    ENST00000282356    1010011
BEST1    ENST00000378043    0011011
CABIN1    ENST00000263119    0001111
CALM1    ENST00000356978    1110001
```

What makes a good MERFISH target gene??

1. Not too highly expressed
    a. There is a limit to the number of spots in a cell before they overlap and become hard to count
2. Relatively long
    a. 1500bp+ can usually target
    b. Too small and it’s hard to get enough oligos
3. Not to similar in sequence to other genes
    a. Receptor families with high sequence homology might be hard to target

Part 3

Going through the oligo design tool code

Part 4

Life post-oligo design

Let’s say you are interested in 150 genes each gene has 92 oligos targeting it to tile along it’s length = 150*92 = 13,800

Oligo length is >(20bp Readout1 + 30bp Encoding Probe + 20bp Readout2)

Or > 70bp @ 10 cents/bp = 13800*70*0.1 ~ $100k

Oligo pools to the rescue

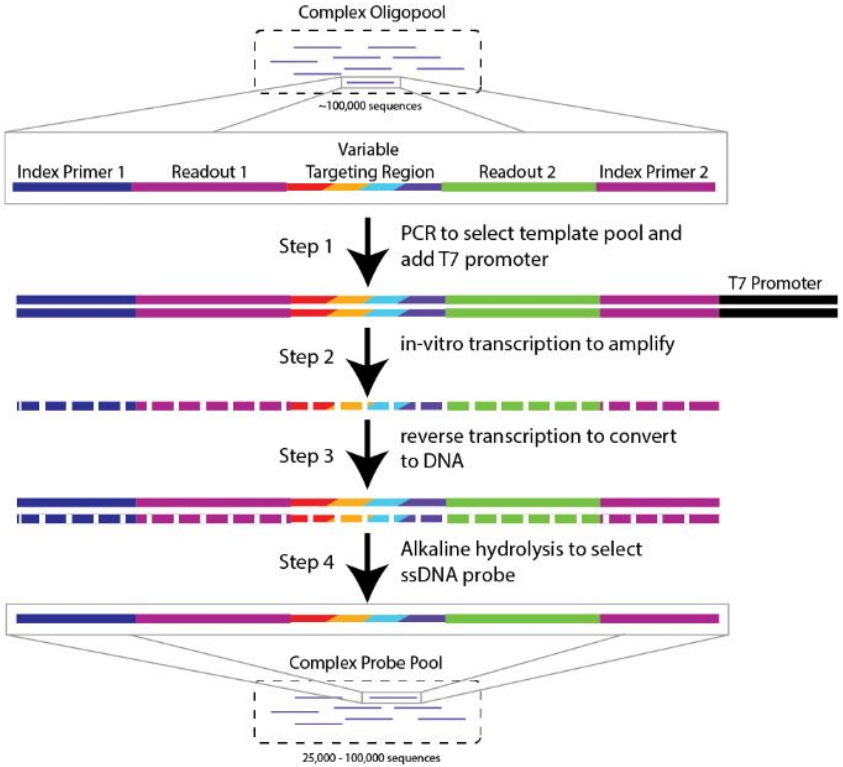
CustomArray, Agilent, and Twist biosciences all cell oligo pools for << $/bp

- As low at 0.04 cents or 0.1/0.0004 = 50-250X cheaper

However, the product has much lower full length product so amplification is necessary (femtomoles yield vs nano-micro moles yield from IDT)

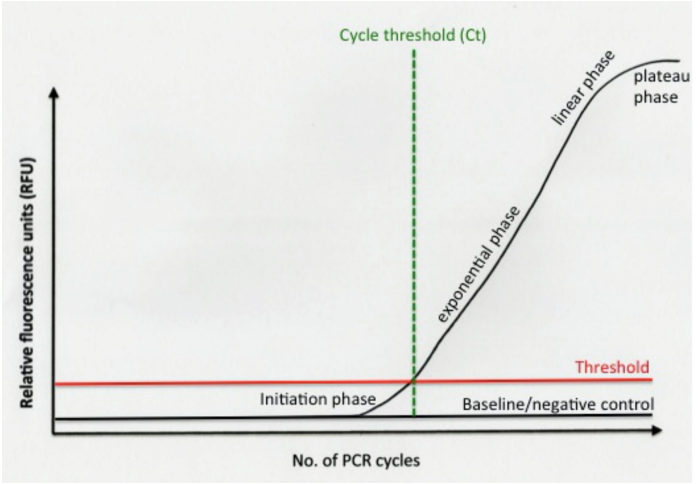
Oligo pool amplification steps

1. PCR
   a. Initial enrichment of full length product
   b. Increase in concentration to nM range
2. In-vitro Transcription
   a. 100-200 fold increase in concentration
   b. Creates single-stranded RNA product
3. Reverse transcription
   a. Go from RNA to RNA/DNA hybrid
4. Degradation of ssRNA
   a. Strong base hydrolyses RNA allowing ssDNA purification



PCR Considerations

Monitor the amplification live to observe transition from log-linear growth to saturating/diminishing sub-log-linear growth



Stop the amplification at the extension phase (~72C before the melting step so that when cooled to room temperature there isn’t hairball of oligos)

In vitro transcription

Must have clean input

Use NEB high yield in vitro transcription kit

Cost prospectus

GenScript (formerly CustomArray)

12k Array

~140 genes with 90 oligos/gene

$2,000 + amplification

92k Array

~1000 genes with 90 oligos/gene

$5k + amplification

Can include multiple gene sets with different primers

oPools from IDT
20-100 experiments with no amplification

Bases	Genes	Cost
10*90*48=43k	10	$1300
86k	20	$2200
129k	30	$2900
172k	40	$3300

Oligo Amplification Cost

PCR Kit - $500 and use ~$50 dollars/amplification

In-vitro transcription kit - $250/kit and use whole kit

Reverse transcription - $600 enough for 4 amplifications

Desalting columns - $25

Total = 50+250+150+25 ~ $500 dollars for 10-20 experiments typically

How to design oligos that target all isoforms of a gene?

1. Download GTF/GFF annotation file
2. Download genomic sequence
3. Remove GTF/GFF transcript features that are lowly expressed or noncanonical transcripts (retained introns etc)
4. Count frequency of exons in transcripts
    a. Extract sequences of most frequent exons and concatenate
5. Remove the transcript sequences of target genes from the off-target fasta
6. Run software as usual
7. Remove oligos that span alternatively spliced junctions

MERFISH Workshop Day 2

Agenda

Day 2

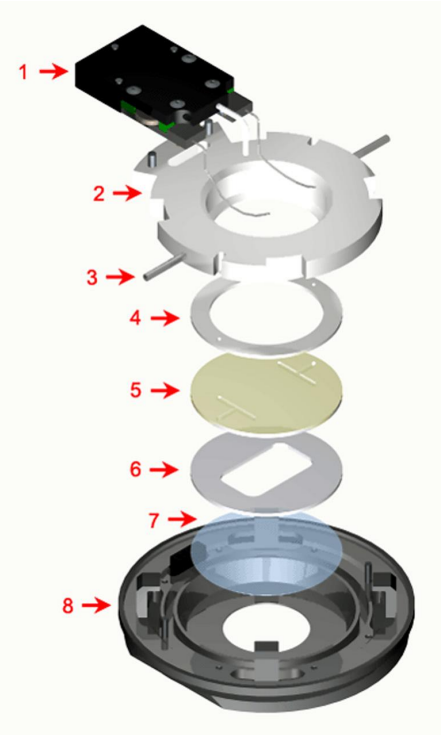
1. Sample Preparation
2. Sample Imaging
3. Automated Serial Hybridization
4. Microscope Requirement

Sample Preparation

1. Modify Coverslips
2. Coat coverslips
3. Culture Cells and perform experiment
4. Fix Cells
5. Permeabilize Cells
6. Hybridize Encoding Oligos and poly-T anchor probes
7. Wash Encoding Oligos
8. Embed cells in 4% acrylamide gel with fiduciary beads
9. Clear cells with detergent and proteinase K

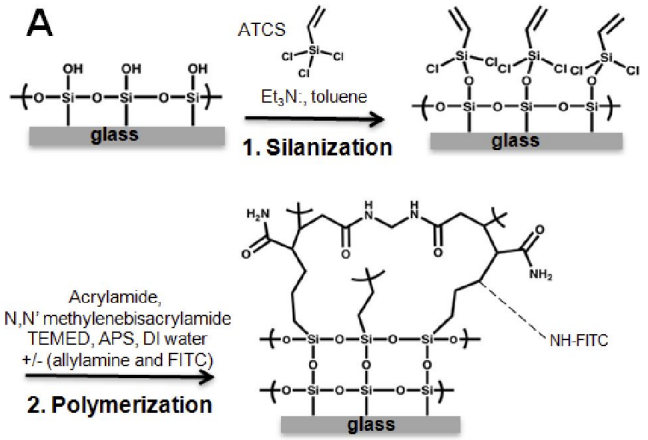
MERFISH Flow System (more details later)

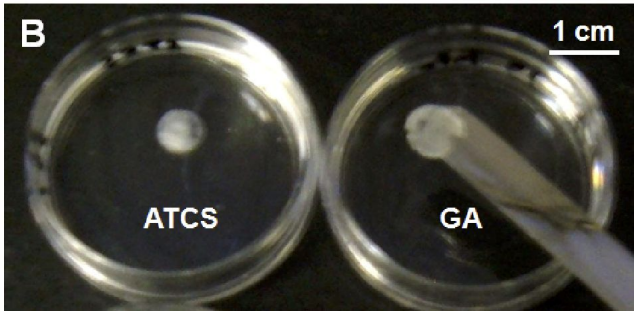
40mm diameter #1.5 coverslips from bioptech



Coverslip modification

1. Strong acid and organic solvent wash (HCl and Methanol)
    a. Removes debri and cleans surface
2. Wash and equilibrate to aqeuous free solution
3. Perform allyl-chloro silane modification in chloroform solvent
    a. Functionalization of glass surface with moeity that will covalently link PA gel to the coverslip surface
4. Wash and dry
5. Store dessicated





Coverslip coating

After modification with allyl silane the coverslips will be very hydrophobic and coating will generally help cell attachment/health

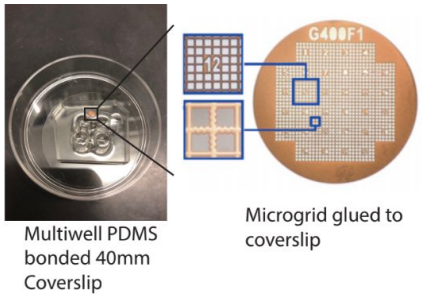
Your choice and the best choice can depend on the cell/tissue type we haven't had any issues with the following methods in terms of compatibility with other steps

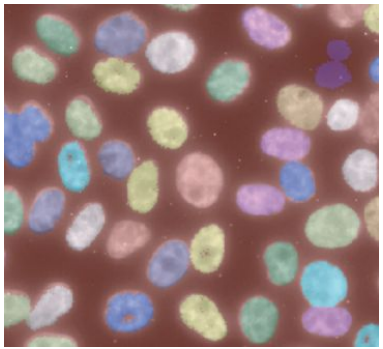
1. Poly-L Lysine
2. Collagen, Fibronectin, and BSA
3. Gelatin

Culture cells and experiment

Big question is how do I do multiple treatments and adapt my experimental system to the use of coverslips compatible with our flow chamber

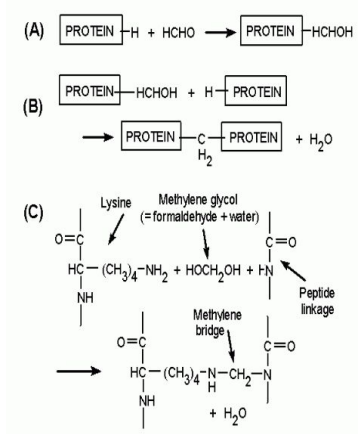
You can use the whole middle of the coverslip for 1 treatment, or we have used PDMS wells to do 4-5 5mm diameter wells on single coverslip then peel off the PDMS after cells are fixed



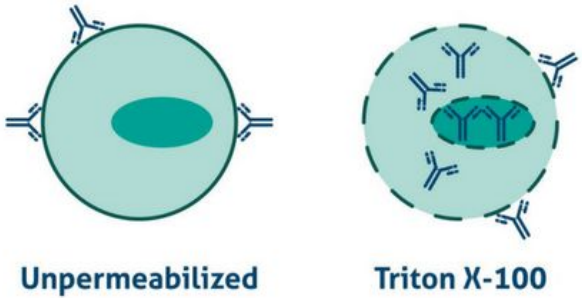


Fix and permeabilize cells

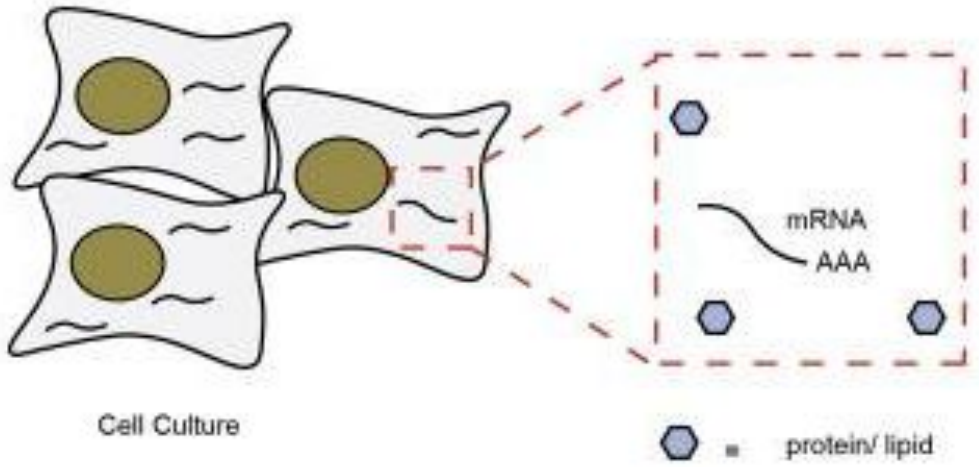
Standard fixation is 4% paraformaldehyde for 20 minutes at room temperature (may consider other protocols for thick tissues or special applications)



Permeabilization is done with 0.5% v/v Triton X-100 in PBS



Encoding and anchor probe hybridization

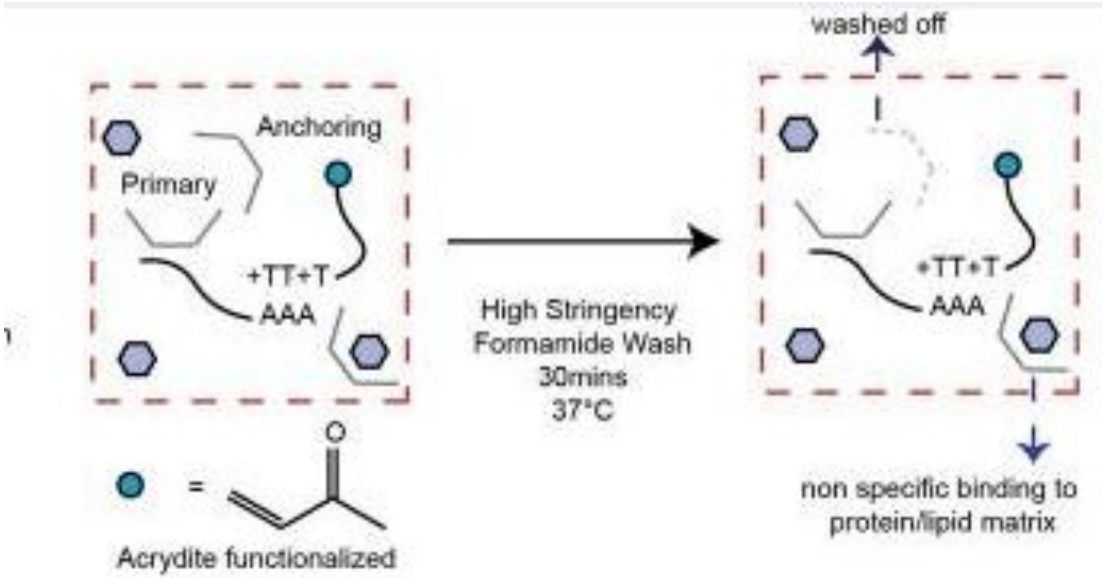


https://www.pnas.org/content/113/50/14456
https://www.pnas.org/content/113/39/11046

Encoding and anchor probe hybridization

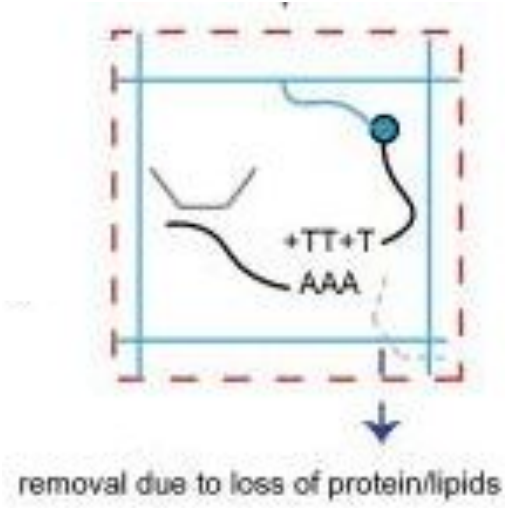
24-48 hours

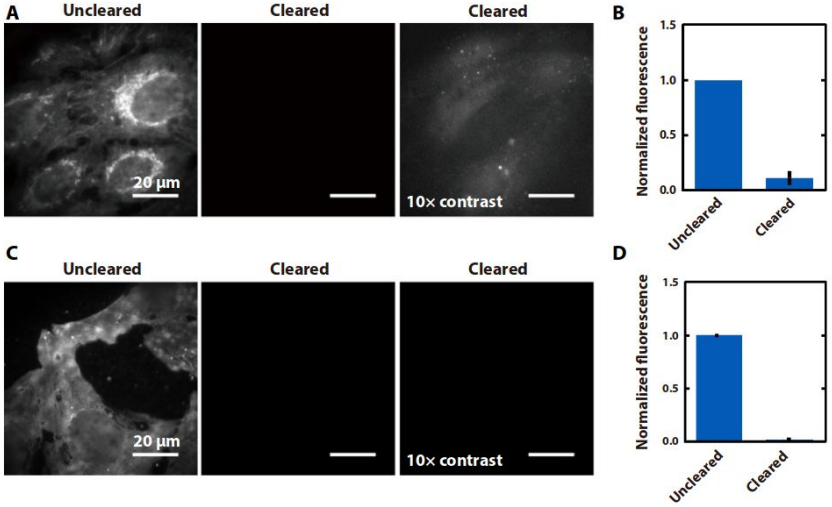
2x 30 minute washes



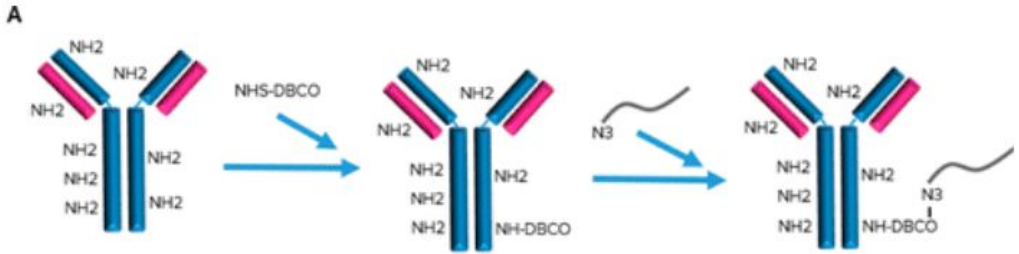
Protein and lipid clearing

+ Proteinase K
+ Trion X-100

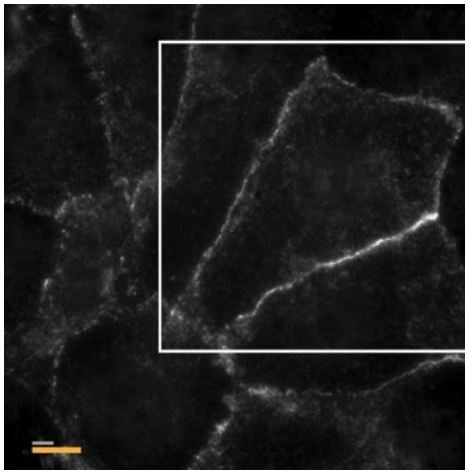




Protein and mRNA measurement in the same cells

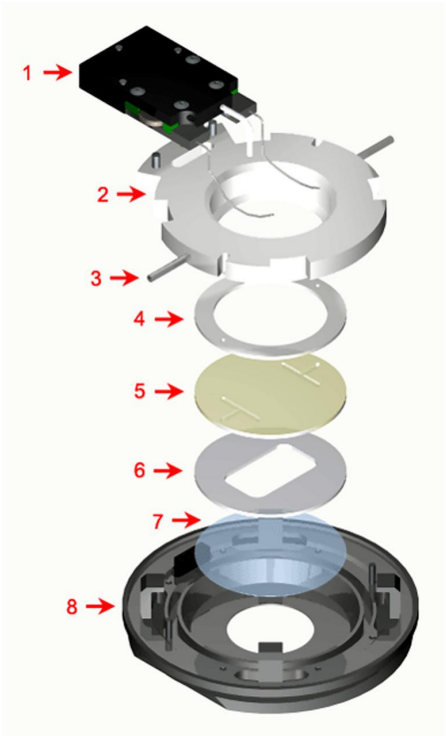


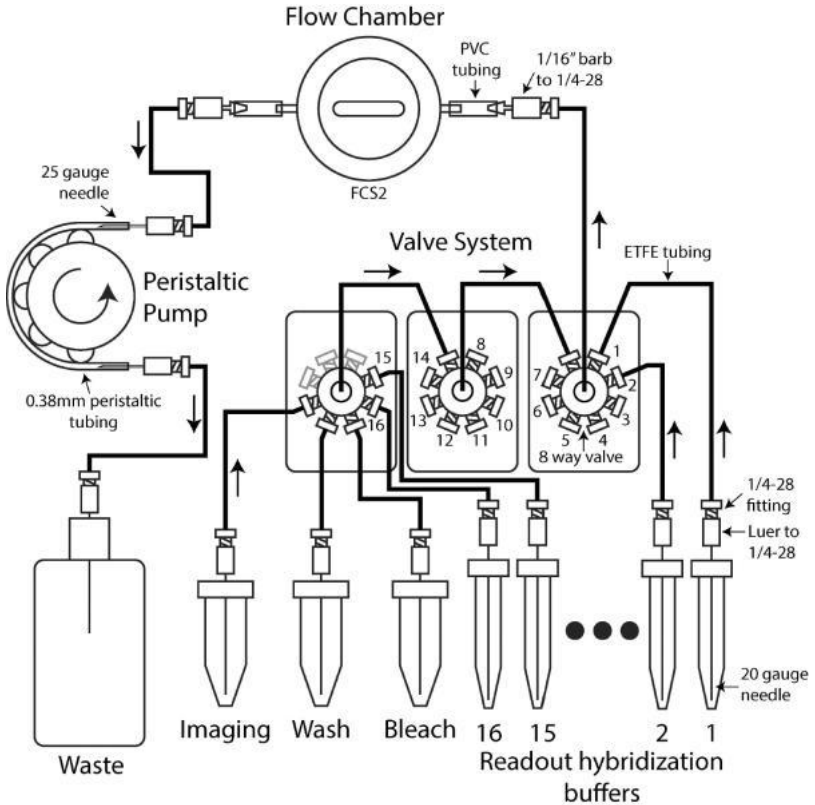
Cadherin oligo-antibody staining



https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5859009/
https://pubs.acs.org/doi/full/10.1021/acs.bioconjchem.5b00613

Fluidics system for automated imaging





https://www.ncbi.nlm.nih.gov/pubmed/27241748

Imaging

1. Set regions of interest
    a. Tissue culture we image at low magnfiication while cells are alive and use the glued grid to find the same positions again at higher magnification
2. Set the autofocus system
    a. We want to image the exactly same physical coordinates with <100nm precision after registration
    b. Across the coverslip and between rounds of hybridization the XYZ physical coordinates of cells are shifting
3. Acquire Z-stack in all channels at each position
4. Move to new FOV/position and repeat
5. After imaging all Z of all FOV during Hybridization *N* wash the sample and repeat

Multiple rounds of readout hybridization and imaging

The entire method depends on the ability to image RNA stained with dye-labeled readout oligos then strip all the signal and re-hybridize with a different dye-labeled oligo

Steps of Readout Hybridization

1. Rinse imaging buffer
2. Add TCEP to reduce disulfide bond
3. Wash away decoupled fluorophores
4. Equilibrate chamber to readout hybridization buffer
5. Hybridize readouts
6. Wash away unbound readouts
7. Flow in imaging buffer

Imaging Experiment Overview

2-3 days microscope time per experiment

15um imaging volume for ~300 FOV at 63X

Each FOV is ~200x200um so ~12mm^2 area imaged over 9 rounds of hybridization

Infastructure costs for MERFISH

1. Standard epifluorescent microscope body (20-100k)
    a. Motorized stage
2. High end apochromatic corrected ~60X objective (2-5k)
3. Quality sCMOS Camera (5-20k)
4. High powered illumination system
    a. 4 Channel Led (5k)
    b. Lasers (buncha money?)
5. Fluidics system (13k)
    a. Pump (1k)
    b. FCS2 system (7k)
    c. Valves (5k)
6. Dye labeled readout oligos (3k to 12k)
    a. 200-800 each * 16

Total cost = 60k to 250k

MERFISH Workshop
Day 3

Outline

1. Image Processing
2. Spot Calling
3. Cell Segmentation
4. Counts Per Cell Analysis
    a. Single-cell gene expression analysis

The data is collected now what?

Sense of scale of the data

9 rounds of imaging in 3 colors = 27 images per Z section

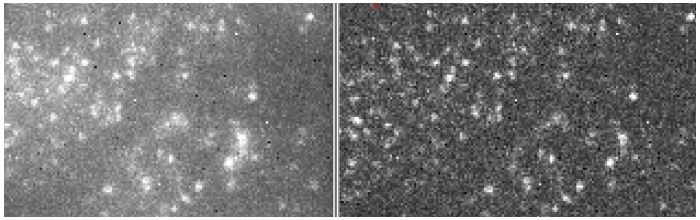
15um in Z every 0.5um = 30 Z positions

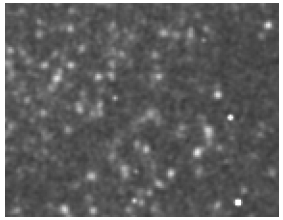
~300 FOV

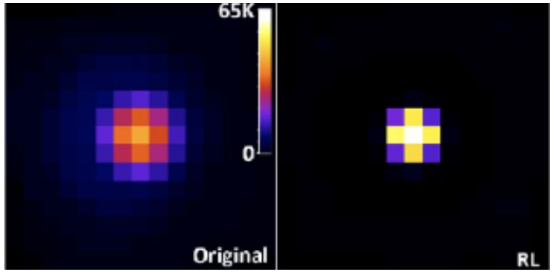
Total images = 27*30*300 ~ 250k 2048x2048 pixel images ~ 2.5TB data

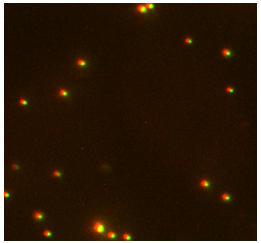
Image processing workflow

1. High pass filter
    a. We know spots are ~diffraction limited spots
    b. Subtract background that is bigger than these spots
2. Deconvolve images
    a. Richardson-Lucy
    b. Tightens the spatial distribution of spots
3. Sub-diffraction blur images
4. Correct Chromatic Aberration
5. Register images







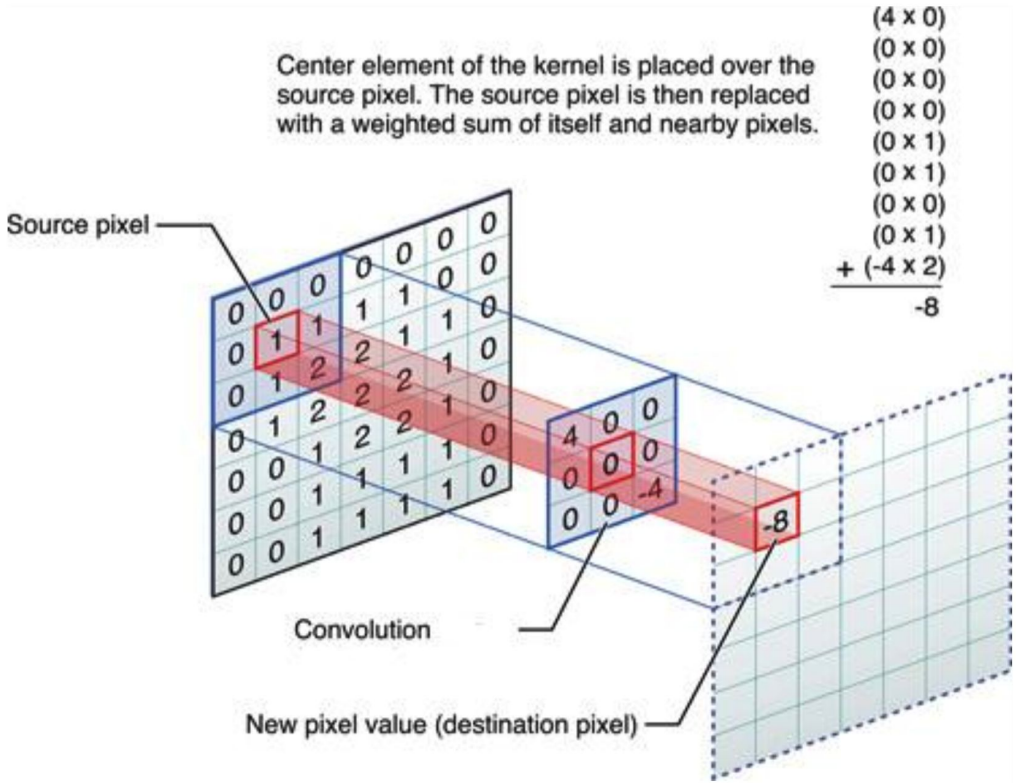


Embedded Bead Images

Red = Hybe 3
Green = Hybe 7

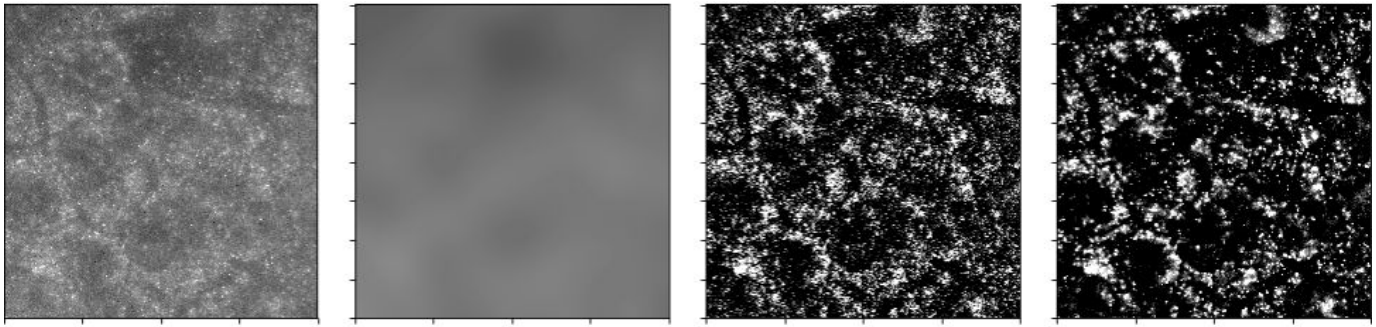
~250nm shift that will ruin experiment if not corrected well

Convolution

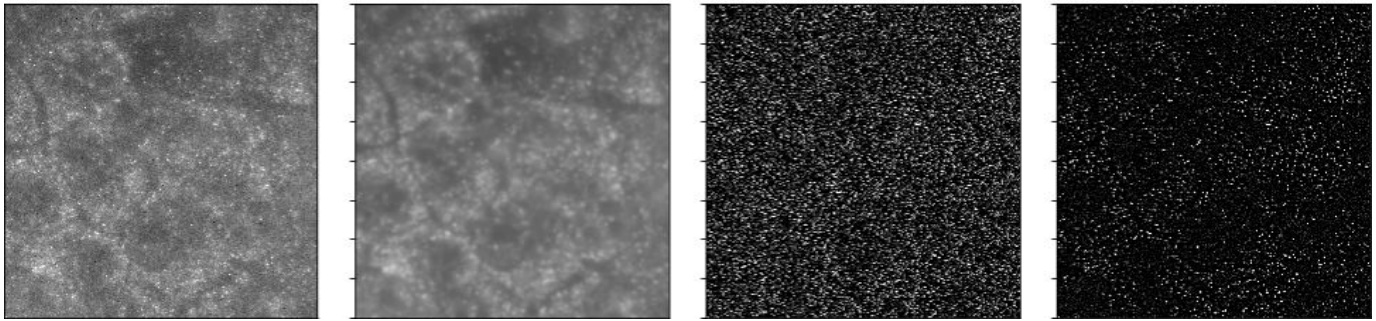


Combination of high pass filter and low pass filter is a band pass filter

1-50 pixel band pass

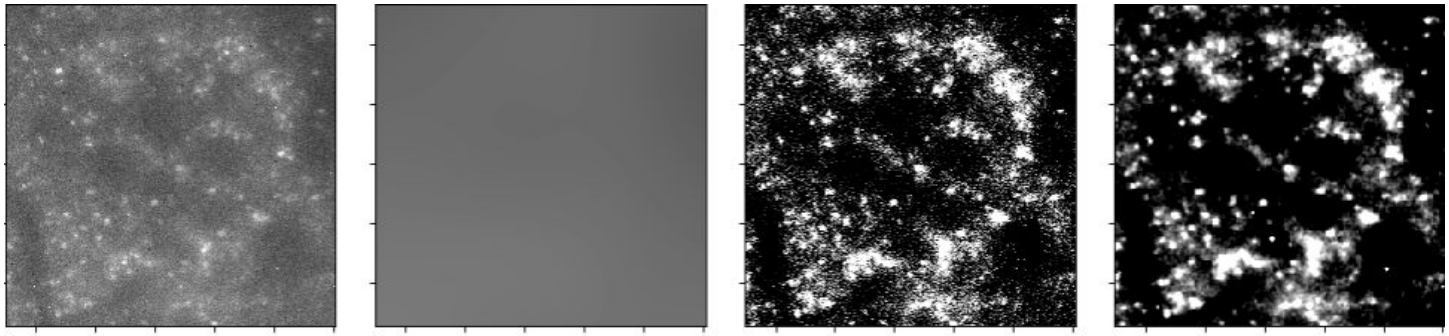


1-3 pixel band pass



Combination of high pass filter and low pass filter is a band pass filter

1-50 pixel band pass



1-3 pixel band pass

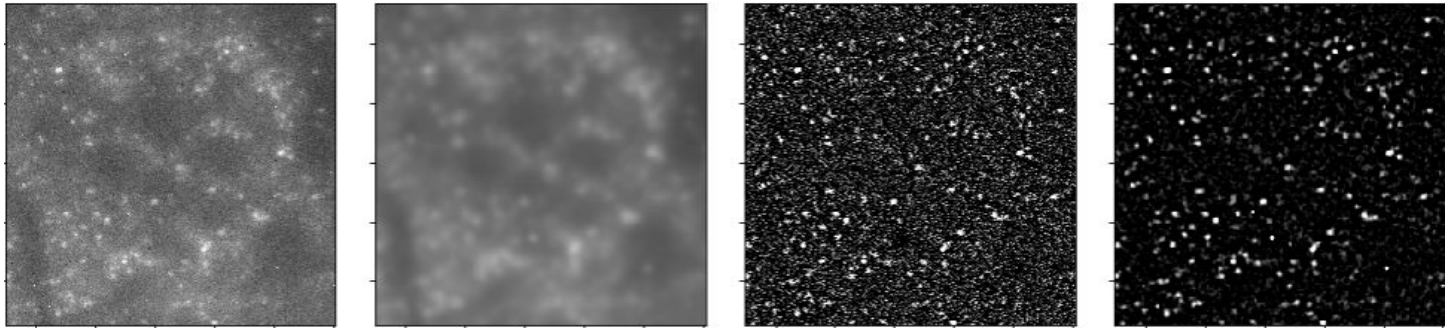


Image registration approaches - cross correlation

Cross correlation between two images (any problems)

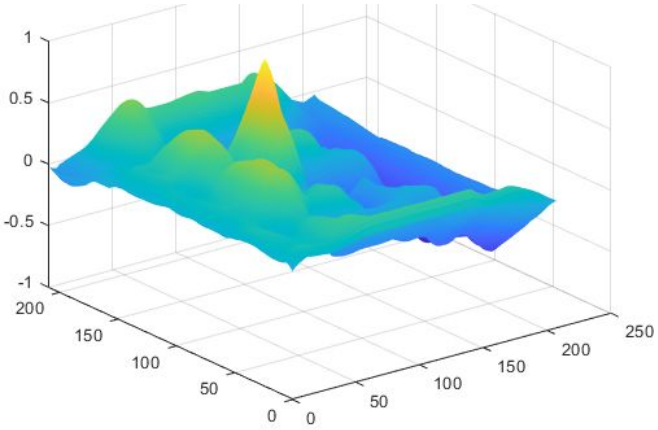
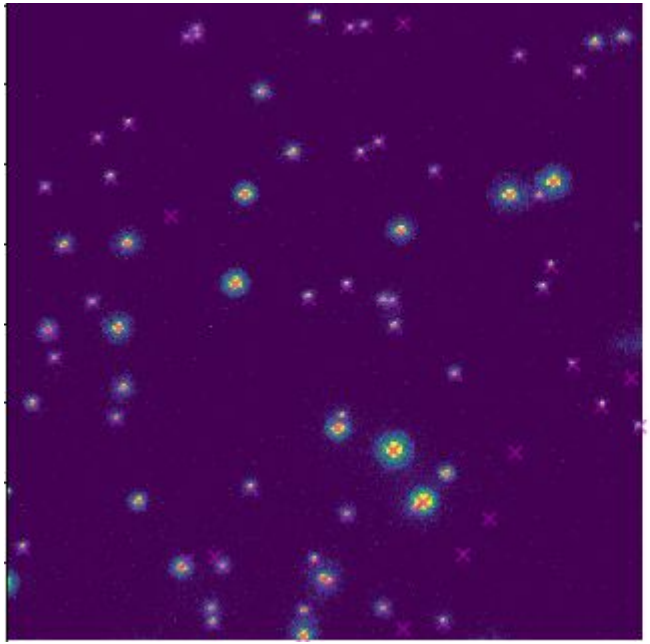


Image registration - paired fiduciary marks

Hybe 1



Hybe 6

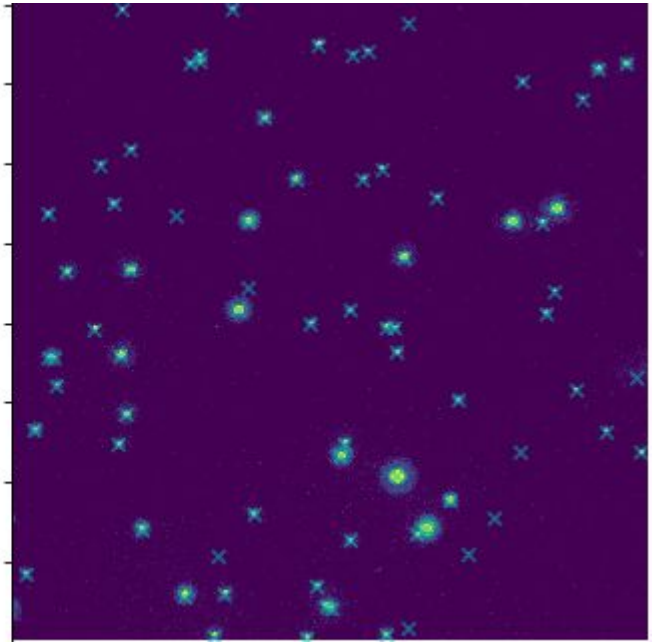


Image registration - paired fiduciary marks - rejecting bad pairs

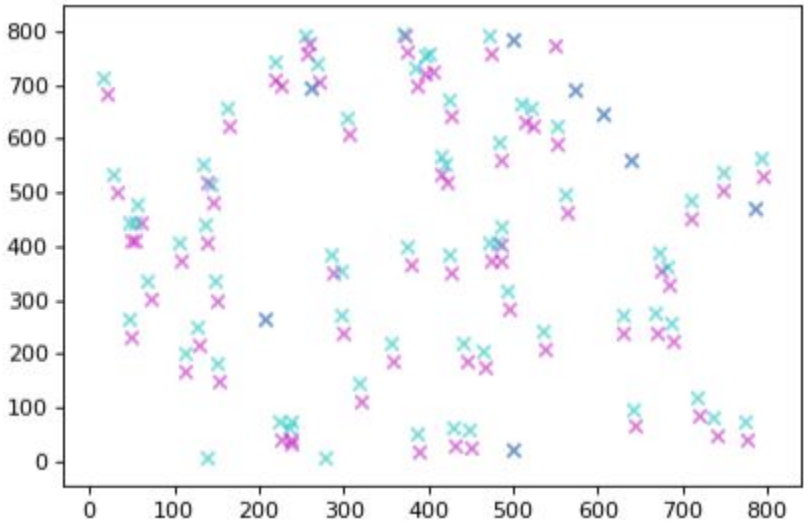
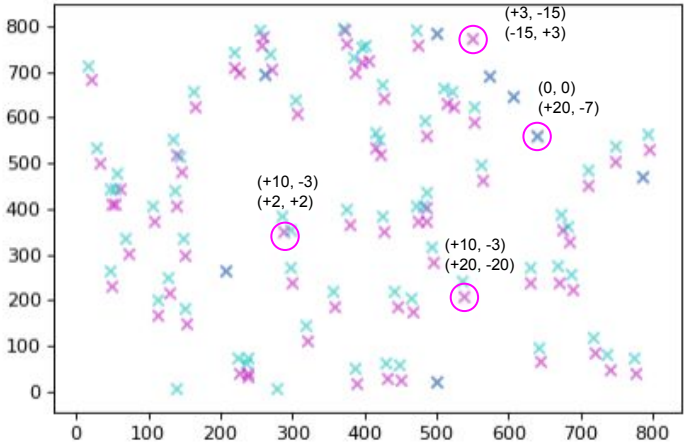


Image Registration

Translational registration from Hybe to Hybe

a. Use template matching in upscaled image
    i. Template is a bead profile (either gaussian estimate or empircal average of beads)
    ii. Normalized cross-correlation at every coordinate in image
b. Find local maxima in template matching image
c. Local maxima are coordinates of candidate beads
    i. Issue is that there can be signficant number of false positive matches and false negative matches
d. Find pairing of candidate beads beads Source Hybe and Destination Hybe
e. Reject pairs that are unlikely to be the same bead
    i. Example on the board



Chromatic aberration introduction

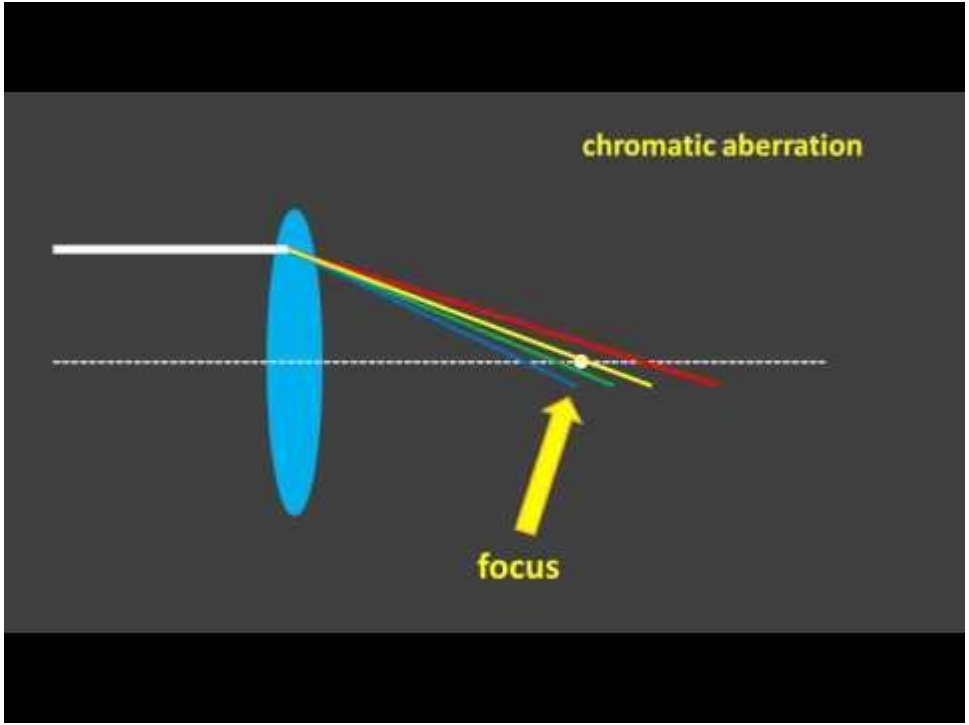
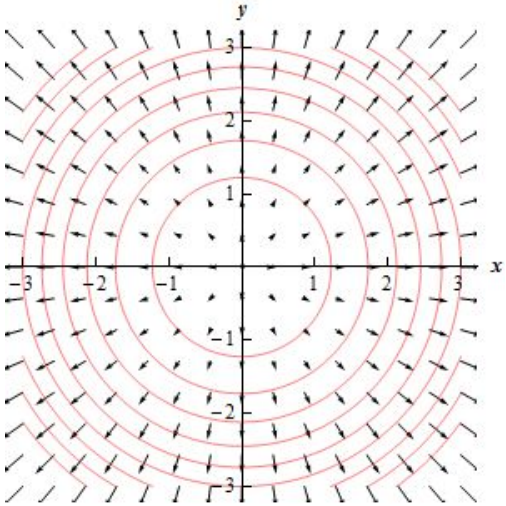


Image Registration

Chromatic Aberration Correction Calibration

1. Embed tetraspeck beads in gel
2. Image in all the channels of a MERFISH experiment
3. As function of XY calculate shift between each bead
4. Reject outliers
5. d(X, Y) for all pairs of channels
6. Apply d(X, Y) and then interpolate back onto the same coordinate space for all images



How to decide if a ‘spot’ or coordinate in a codestack is a gene or is background?

Two main approaches

1. Call Spots and Pair Spots (Long Cai and original MERFISH)
    a. Classify each spot based on it’s spatial distribution (gaussian fit parameters) for each image in the codestack independently
    b. Group ‘called spots’ based on proximity
    c. Check codebook to see if the called sequence of 0’s and 1’s matching a gene
2. Pixel Based Classification
    a. Decode each pixel based on it’s vector distance to each codebook entry
    b. Group pixels that are neighboring with same codeword
    c. Filter grouped pixels based on the number of pixels and their average intensity

Spot calling -- some reminders

After image processing we a ‘codestack’ for each Z plane of each FOV imaged

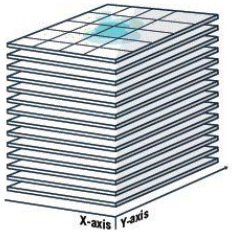
Codestacks are 2048x2048x18 (YXB) where B is the number of ‘bits’ or pseudocolors in the codewords assigned to each gene

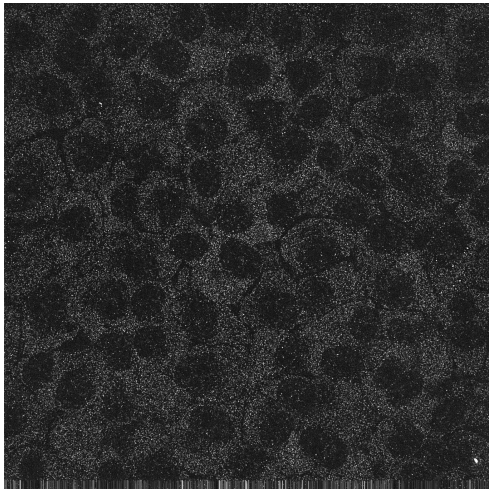
Every gene has exactly 4 1-valued bits in its codeword

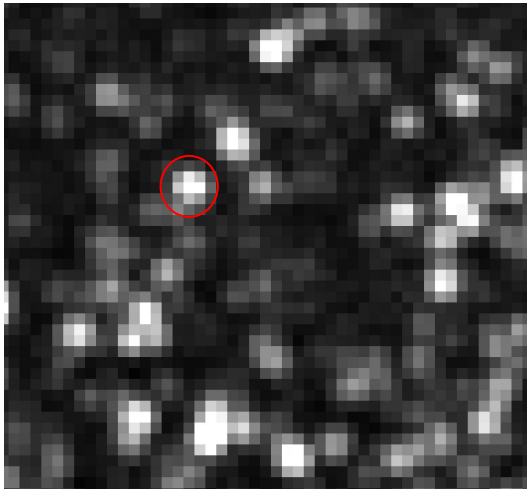
Expectation a real gene will have exactly 4 bright spots in a specific sequence of the B dimension of a codestack

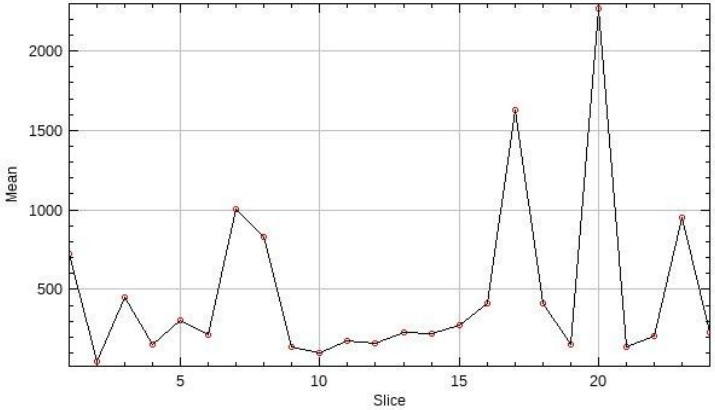
What does the data look like?

MCF10A cells



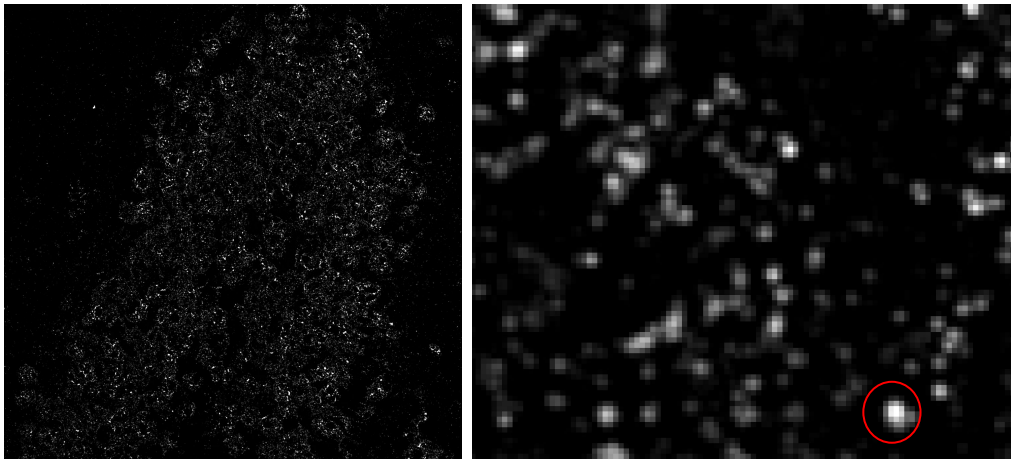


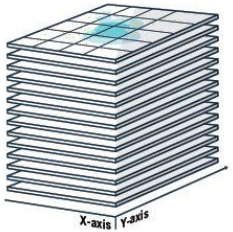


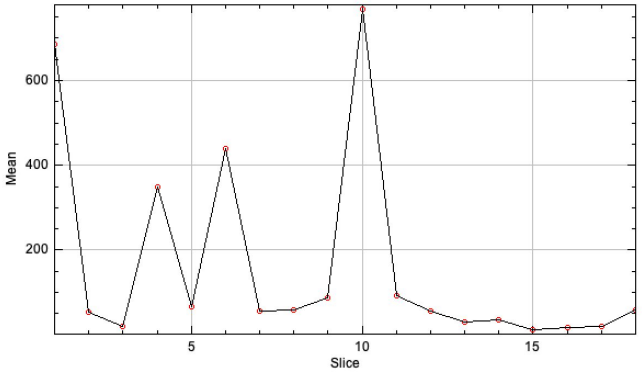


What does the data look like?

Mouse hippocampus







Call Spots and Pair Spots (Long Cai and original MERFISH)

What are some potential problems with this method?

1. What if spots vary a little bit in their position image to image?
    a. mRNA can be +/- 1 pixel in terms of their center
    b. Registration not perfect
    c. Chromatic Abberation

    If the radius of pairing becomes too big then optical crowding is a big problem and often have to consider all combinations of 4 for each spot which leads to more false positives

2. What if a spot is just slightly below the threshold for calling it a spot?
    a. Signficant but sub threshold spots are completely disregarded
    b. Setting the threshold always subject to FN/FP tradeoff

Issues with pixel based method

1. Maybe slow? Lots of pixels
2. Implicitly assumes that all hybes/dyes have the same intensity in the images?
    a. Must normalize spot intensity accross hybes
3. Also subject to issues if spots from different images don’t overlap well
    a. Not possible to iterate over combinations if the spot overlap isn’t high enough
4. More subject to issues from dense spots
    a. Spots can be dense in a single image as well as dense accross the codestack (Example)

Comparison of the two methods

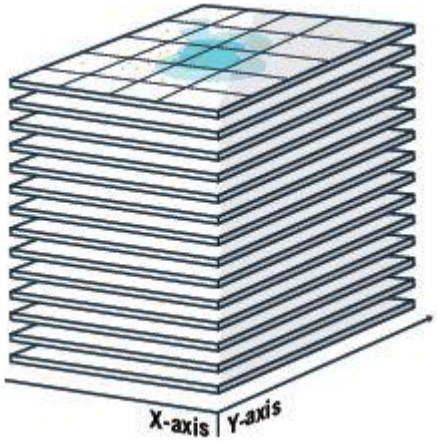
Long Cai’s group consistently reports lower sensivity and specificity than the Zhuang lab

Generally both methods are highly specific and they differ in sensitivity with Long Cai’s group reporting 50-80% sensitivity and Zhuang lab reporting 99%+

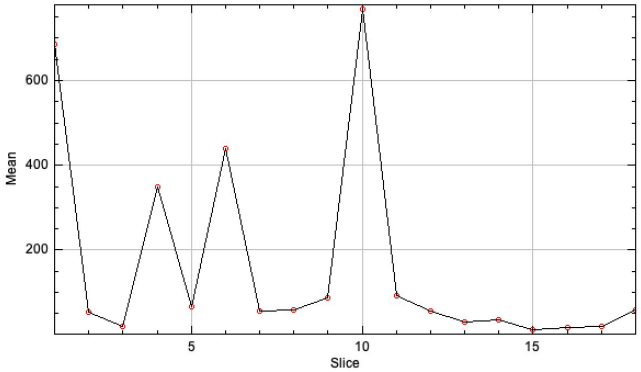
The pixel based method can be tricky and subject to high quality data

Spot calling

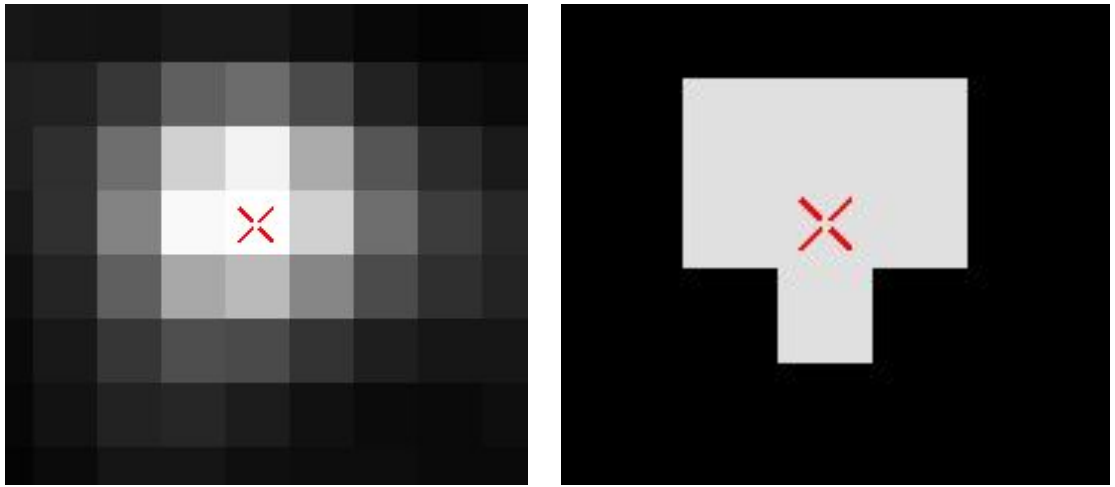
Therefore if we consider every X,Y coordinate of the codestack along the B dimension we consider a vector of image intensities



In 18 dimensional space this vector has some distance from each 18 dimensional codeword assigned to a gene



Spot calling output is ‘classification image’



Classification image processing and filtering

1. Adjacent pixels of same label/class are grouped and segmented
2. Intensity of the spot is extract from all 1 bit channels and the average is calculated
3. Spots are filtered by their intensity and the number of pixels grouped to filter out false positives
    a. Intensity depends on experiment
    b. Typically >2 pixels must be classified (can vary experiment ot experiment)
    c. Thresholds can be calibrated using the blank codewords and their FPR at different thresholds

Part 2 - iterative estimation of normalization factor

Estimating normalization factor through iterative spot classification

1. Normalize images
    a. Divide by a bitwise normalization factor
2. Iterate over X,Y coordinates of codestack
    a. For each X,Y
        i. Extract intensity vector along B and normalize by vector magnitude
        ii. Calculate the distance between normalized vector and each codeword from the codebook
        iii. If minimum distance < half the distanceness to the next nearest codeword classify that pixel as an instance of the gene with minimal distance
3. Calculate the average intensity of each bit using only the subset of pixels that were classified as genes
4. Repeat with the newly estimated normalization factor

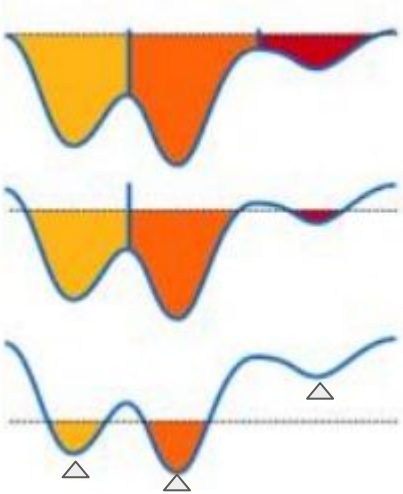
General factors affecting the quality of gene decoding

1. High signal to background (that's why we use 92 oligos per gene)
2. Equal brightness of each 'bit' of the codevectors
    a. Perhaps after normalization
3. Large spacing between codewords assigned to genes (minimum hamming distance of 4 between any two codewords used)

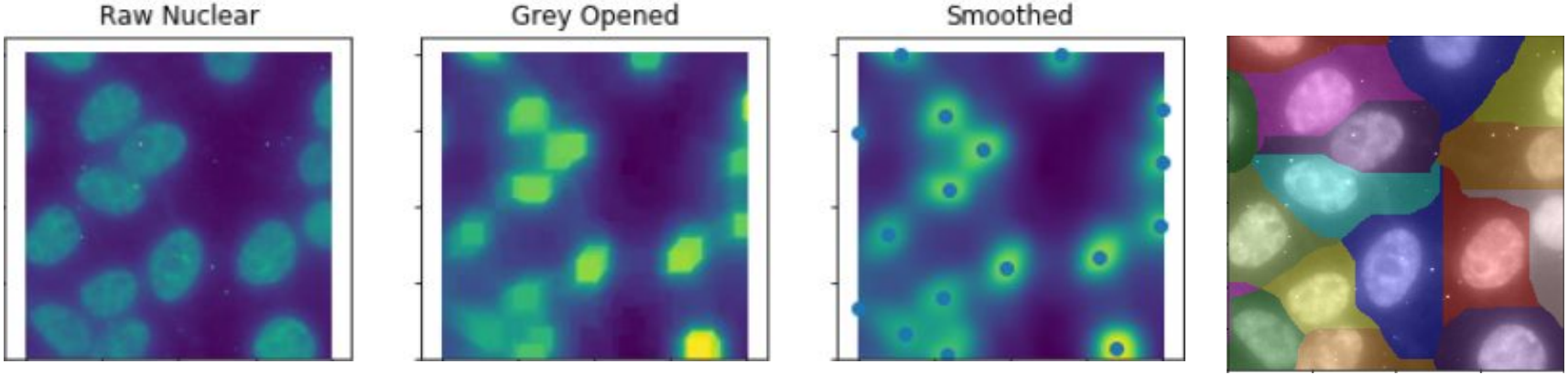
Call spots with final estimated normalization factor

Watershed segmentation introduction

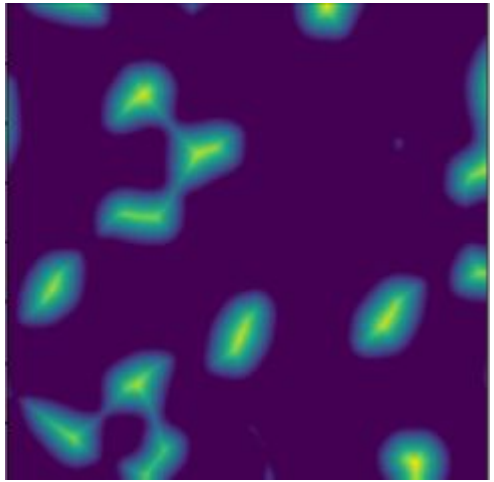
Initialize with some seeds (triangles) and flood the ‘basin’ until seed regions collide at ridges in the landscape. These ridges form segmentation boundaries

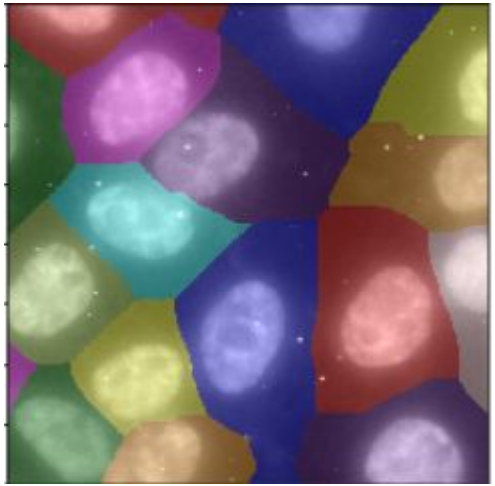


Cell segmentation - watershed seeds intensity based



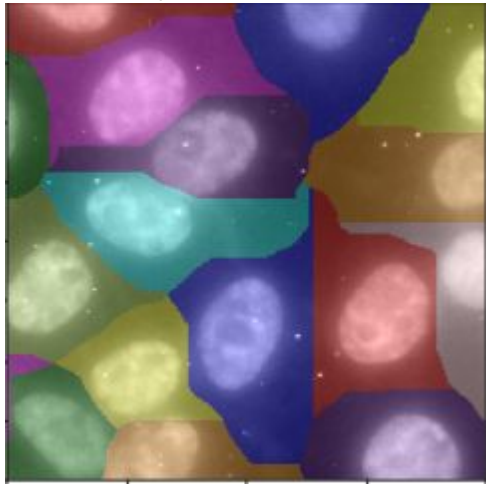
Cell segmentation - watershed distance based



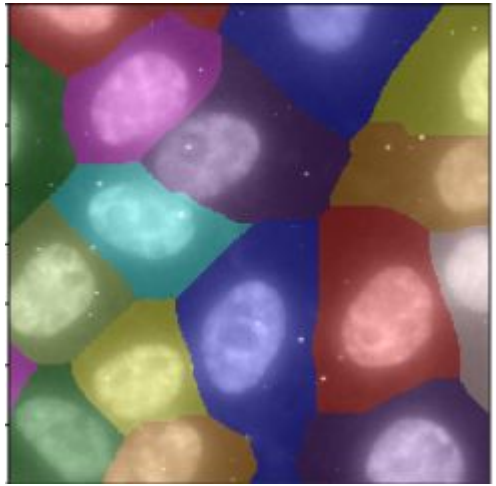


Comparison

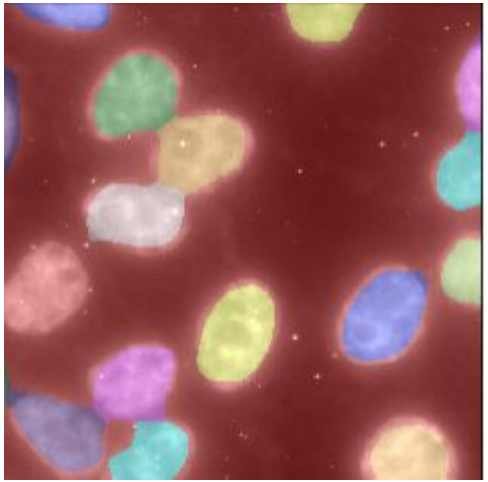
Intensity



Distance

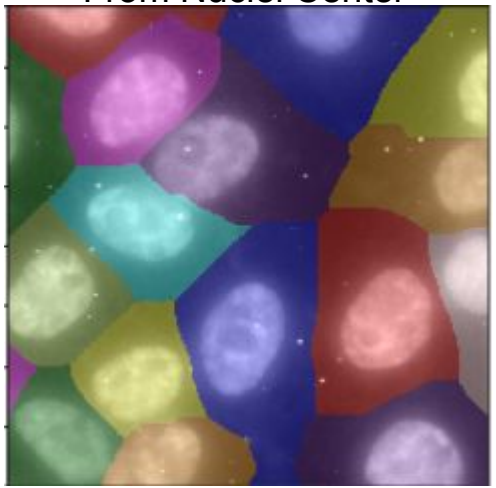


Watershed plus threshold mask to segment nuclei

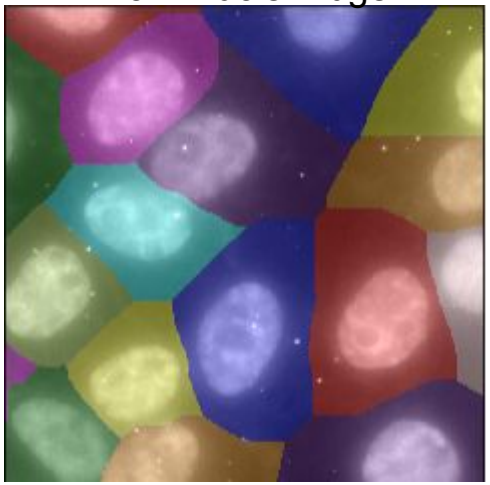


Repeat but now using segmented nuclei as seeds

From Nuclei Center



From Nuclei Edge

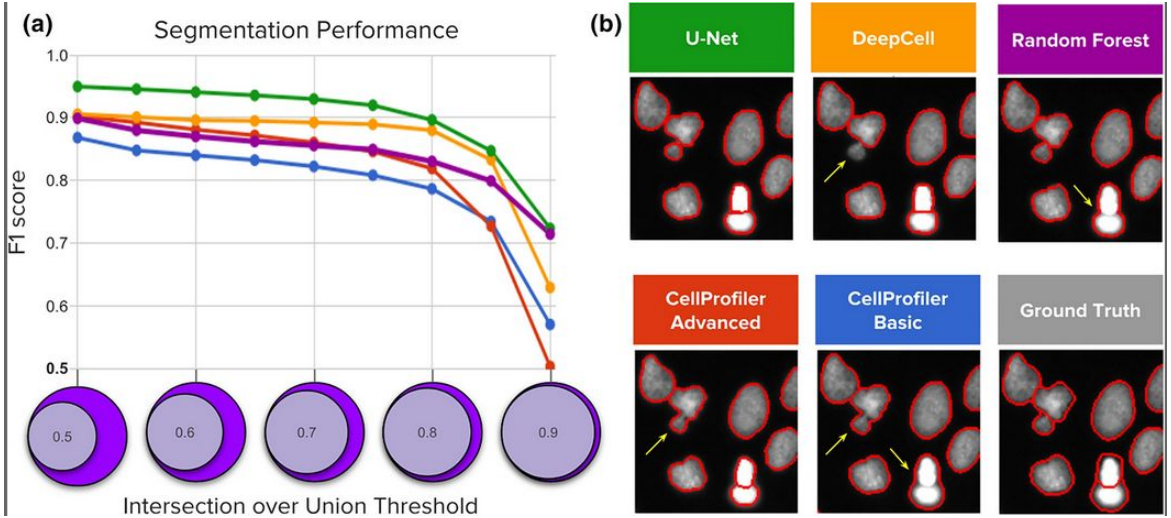


Filtering of segmented cells

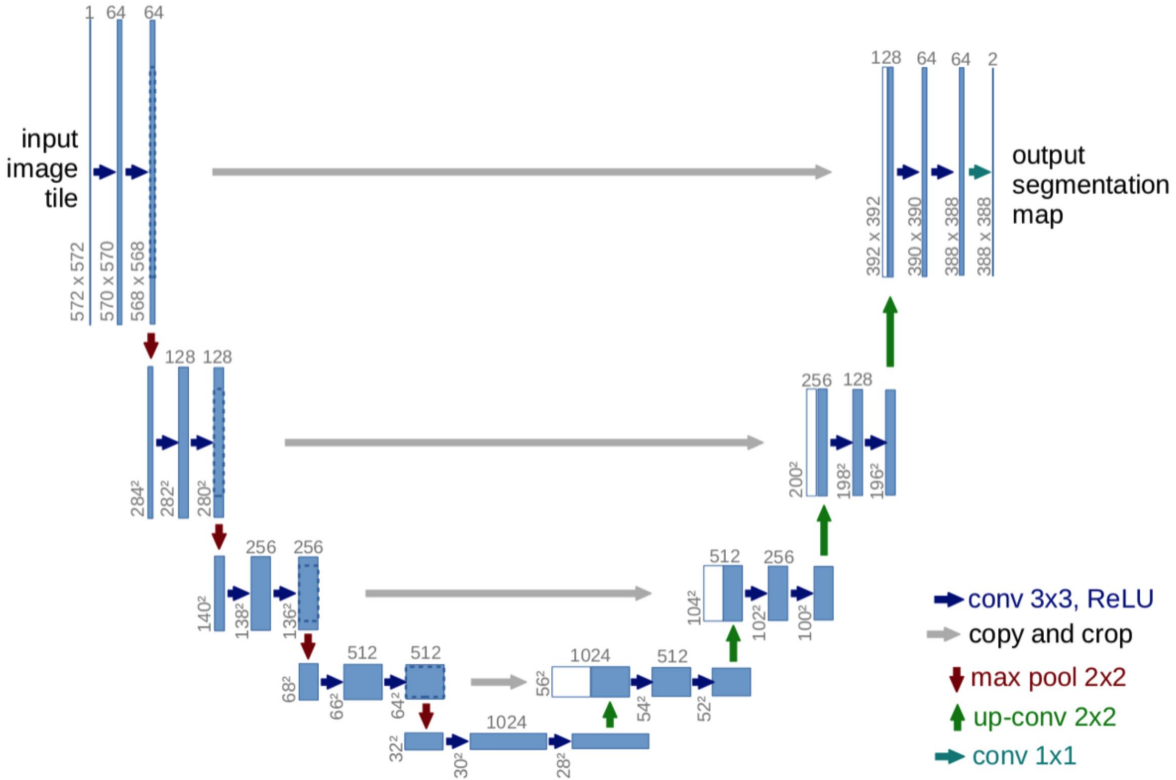
1. Cells touching the edge of the image
2. Cells with too big/small nuclei
3. Cells with too few RNA molecules assigned to them

Deep learning segmentation

Evaluation of Deep Learning Strategies for Nucleus Segmentation in Fluorescence Images



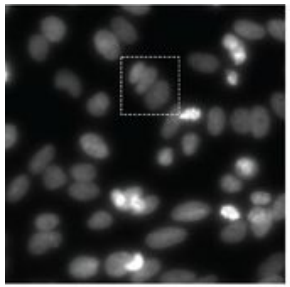
U-Net Architecture for image based deep learning

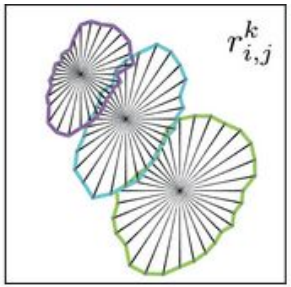


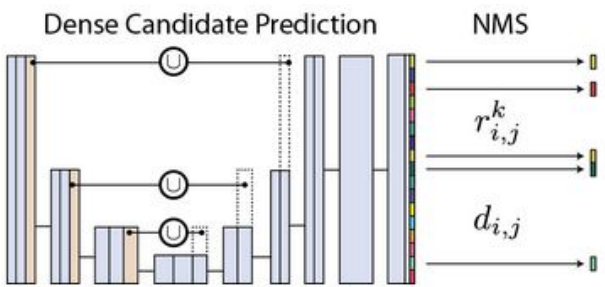
[U-Net Further Reading]

[transposed convolution]

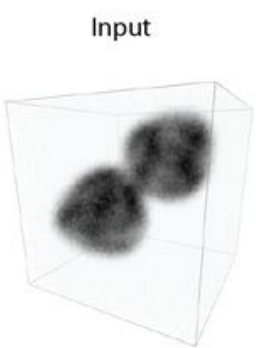
Stardist deep learning based segmentation

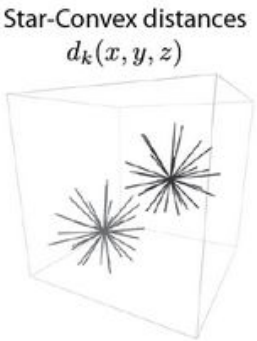


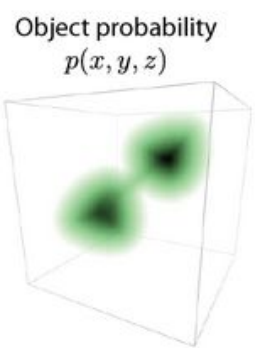


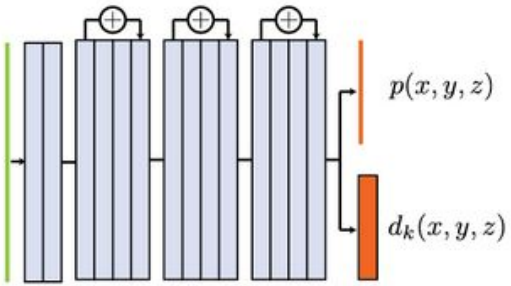


The approach for 3D volumes is similar to the one described for 2D, using pairs of input and fully annotated label volumes as training data.



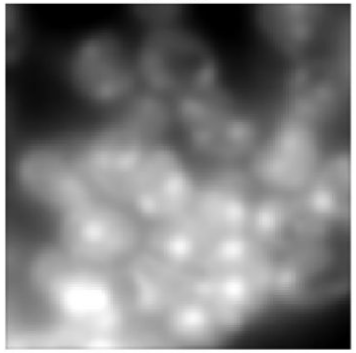






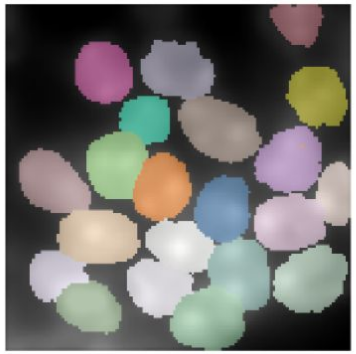
Stardist results in 3D

XY slice



XZ slice

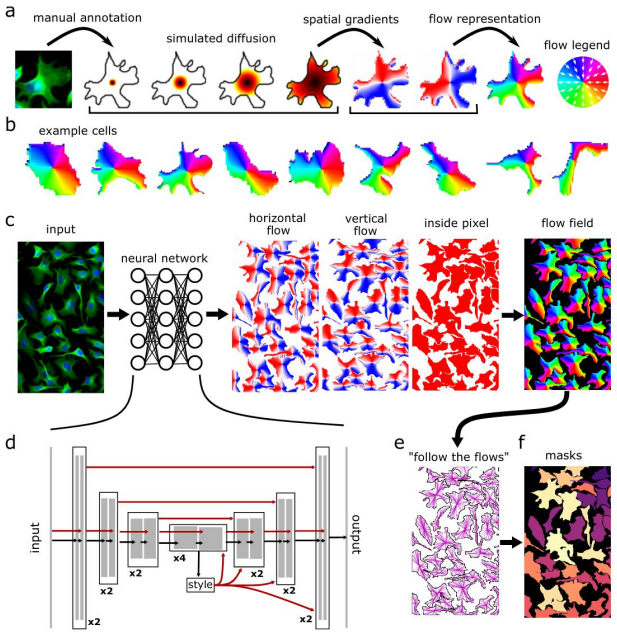
XY slice

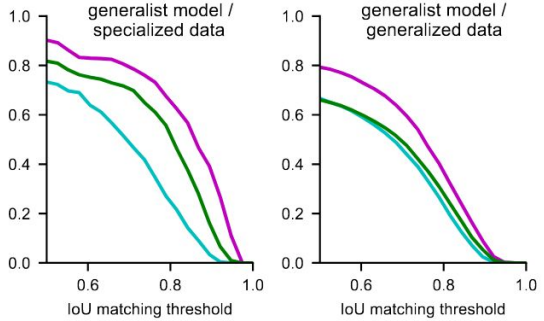


XZ slice

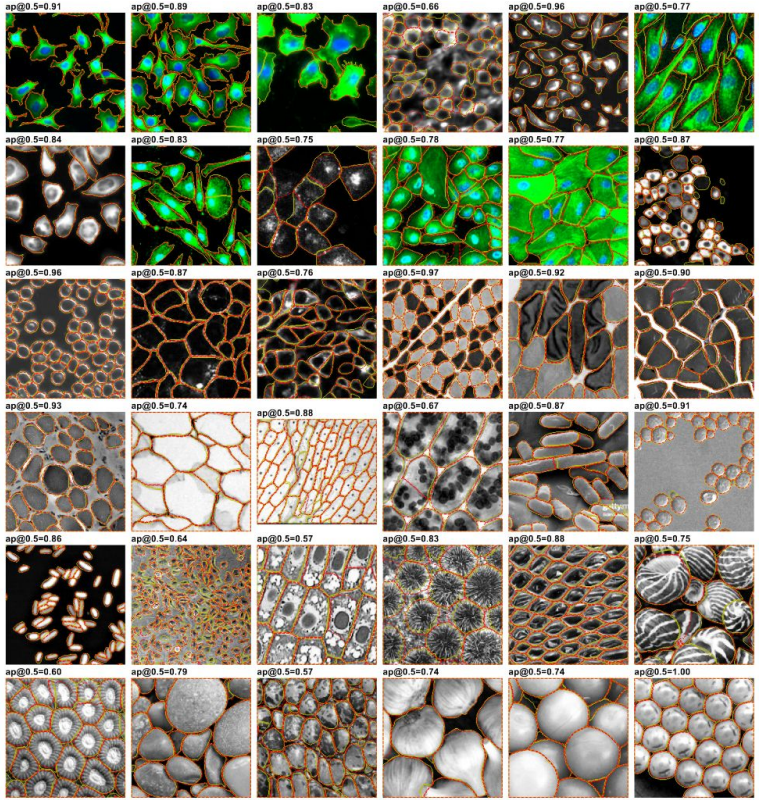
Cellpose

Does not support 3D currently but they’re working on it

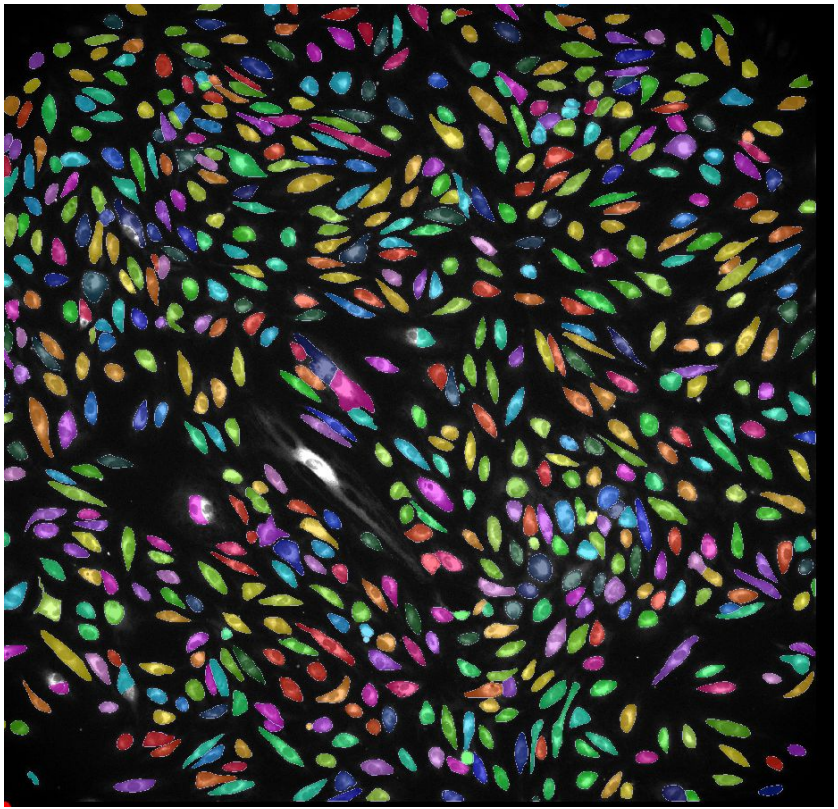


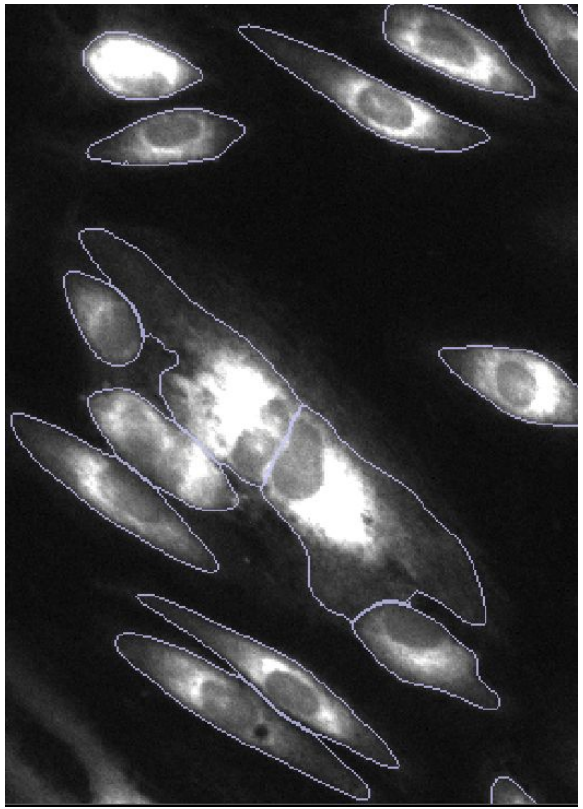


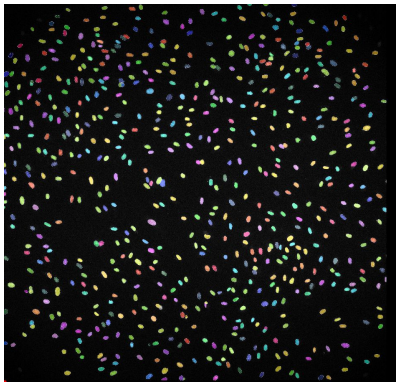
Cellpose - generalist test set

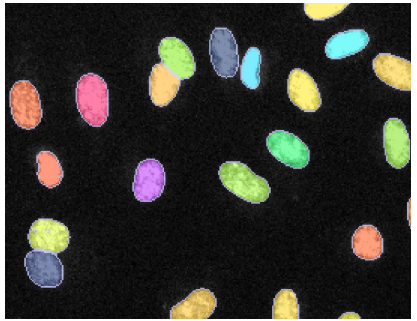


Cellpose performance on different images: HAEC



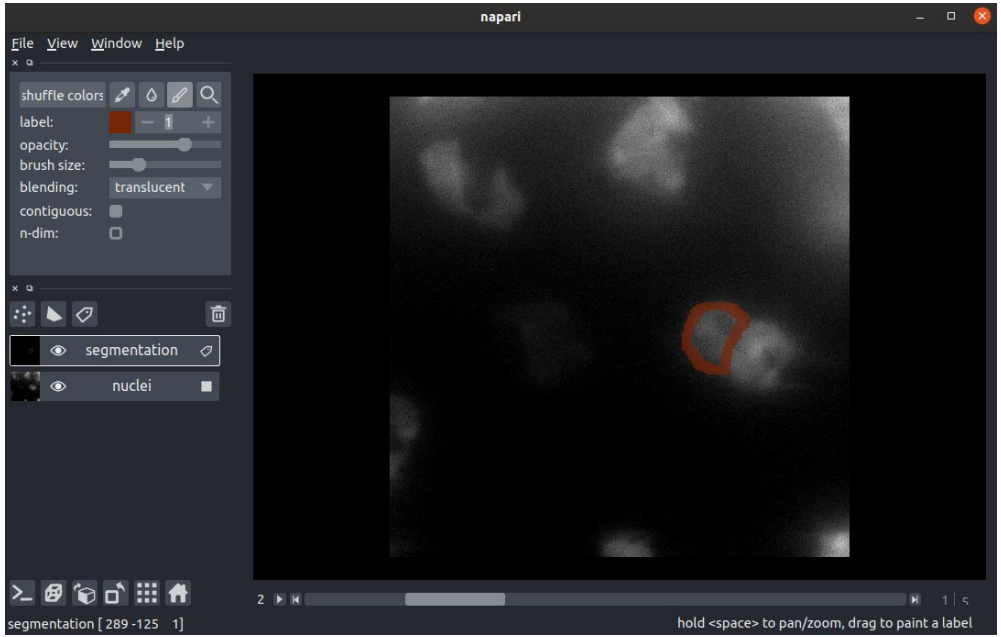






Napari

Tool with development supported by CZI for fast native python manipulation of 3D/4D imaging volumes



Notes for deep learning methods

Often need tweaking for specific datasets

Most will assume that you downsample your images so that the objects you're interested in have pixel diameter of 20-30 pixels

3D manual annotation is labor intensive

Spatial transcriptomics

Differential expression as a function of space is new and still in the rapid expansion of methods phase driven by new data.

MERFISH, seqFISH

10X Genomics Visium platform

Tools for spatial gene expression analysis

1. Space Ranger (10X Genomics)
   a. Tools from 10X for visium data
2. Seurat
   a. Implementing visualization and spatial DE tools
3. Spaniel
   a. R package for visualization and sharing spatial transcriptomic data
4. Giotto
   a. Implments trendcreek and spatialDE tests for differential expression

Approaches to differential expression accross space

1. Cluster cells according to space and test differential expression between clusters
2. SpatialDE
    a. Gaussian process regression to decompose variance into spatial and nonspatial components
3. Trendcreek
    a. Marked point processes - statistical framework borrowed from astronomy/geo fields

Alternative segmentation approaches

Use expression level of different genes to segment cells based on transitions in cell type specific expression patterns

1. Baysor - Bayesian segmentation of spatial transcriptomics data (julia)
2. Ssam - Spot-based spatial cell-type analysis by multidimensional mRNA density estimation (Python/R)
3. Cell_call (javascript)