Genomic Data Visualization

Monday Jan 31, 2022

Reflection Card Review

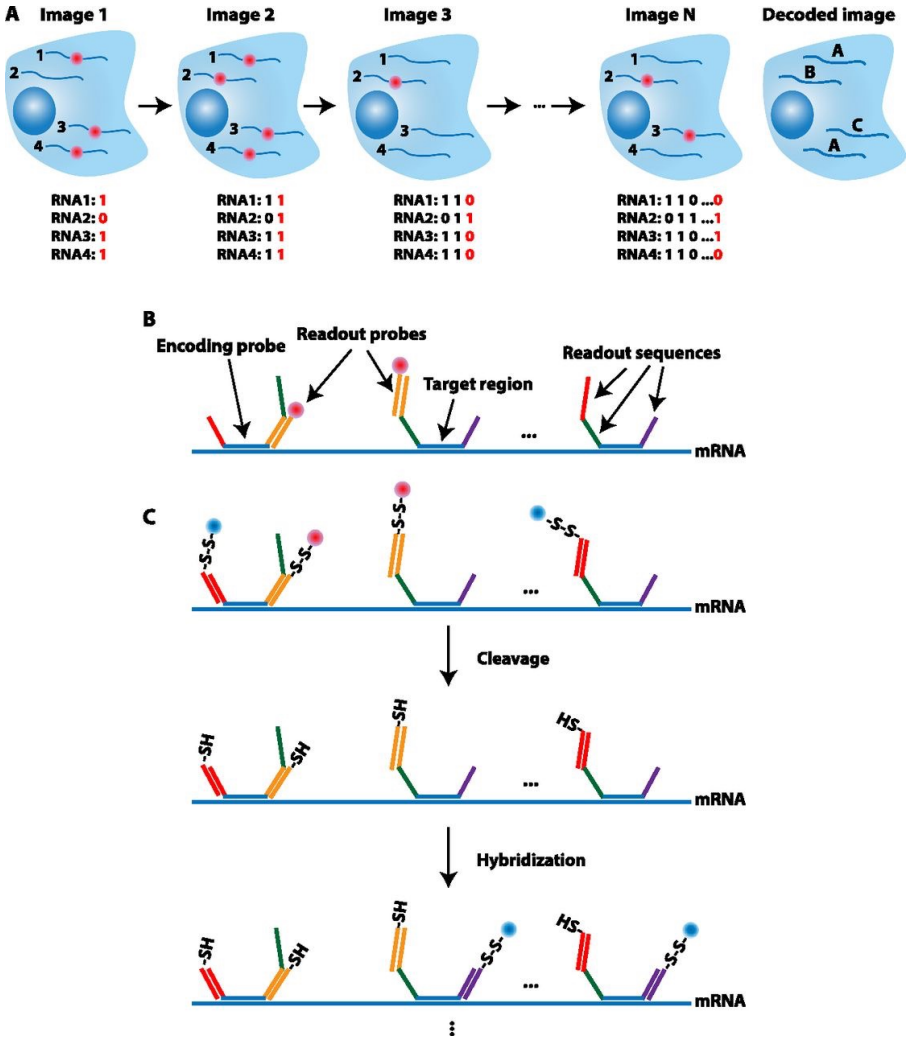
MERFISH read out different mRNAs as binary barcodes of length N -> reading out these binary barcodes through N rounds of sequential smFISH

- Imaging based

Q: Do we need to know the sequence of the RNAs we're targeting

-> design probes with target regions that are complementary to our RNAs of interest -> YES

-> MERFISH is targeted and limited to genes where we know what the sequence is



Q: How long does the target sequence need to be?

-> what do you think would happen if the target sequence was 1 nt long?

- bind everywhere; too short, non-specific binding
- Annealing and clearing chemistry

-> what do you think would happen if the target sequence was like 1000 nt long?

- Secondary structure interactions
- Annealing temperature -> time

Q: How many flourescnece probes do we need?

- 10 to 50

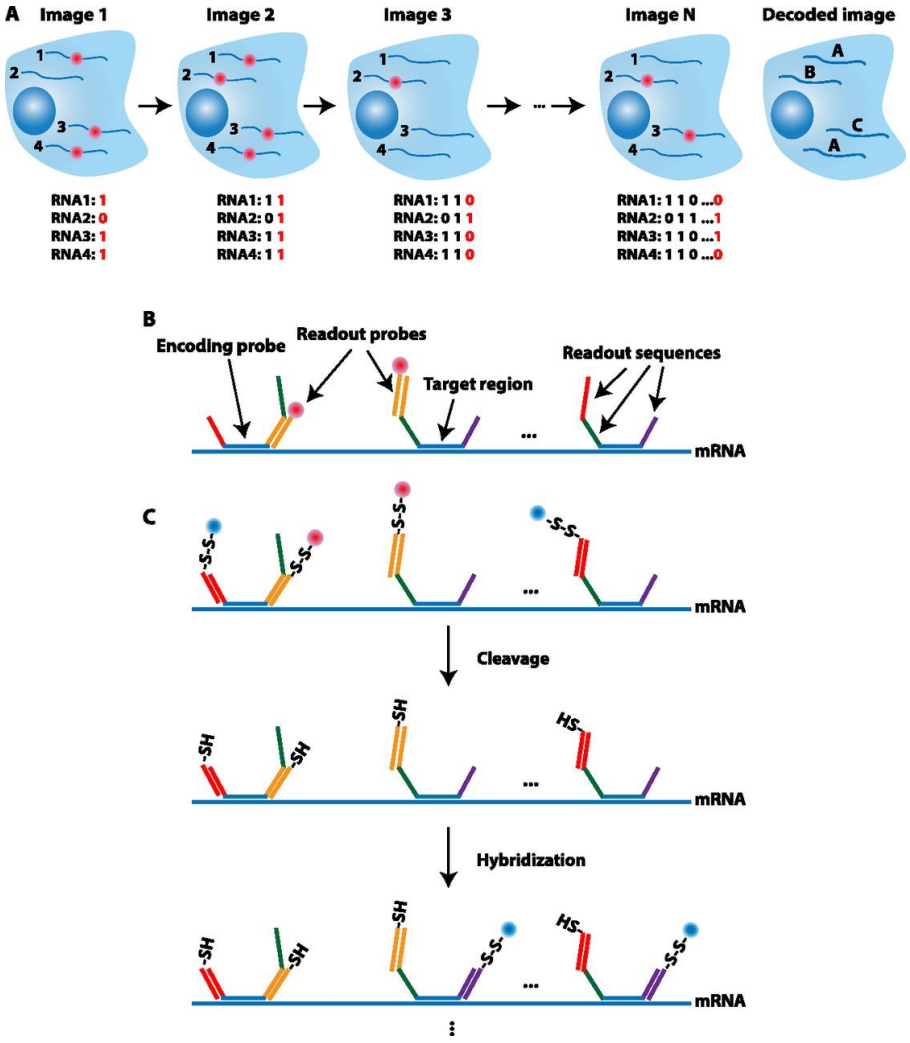
-> What does this mean for targeting short mRNAs?

Length 5 binary barcode, 10 probes per bit, each non-overlapping probe is 30 nt long -> what is the shortest RNA we can target?

- 10*5 probes, 50*30 -> 1500

-> microRNAs are small non-coding RNAs average length of 20 nt

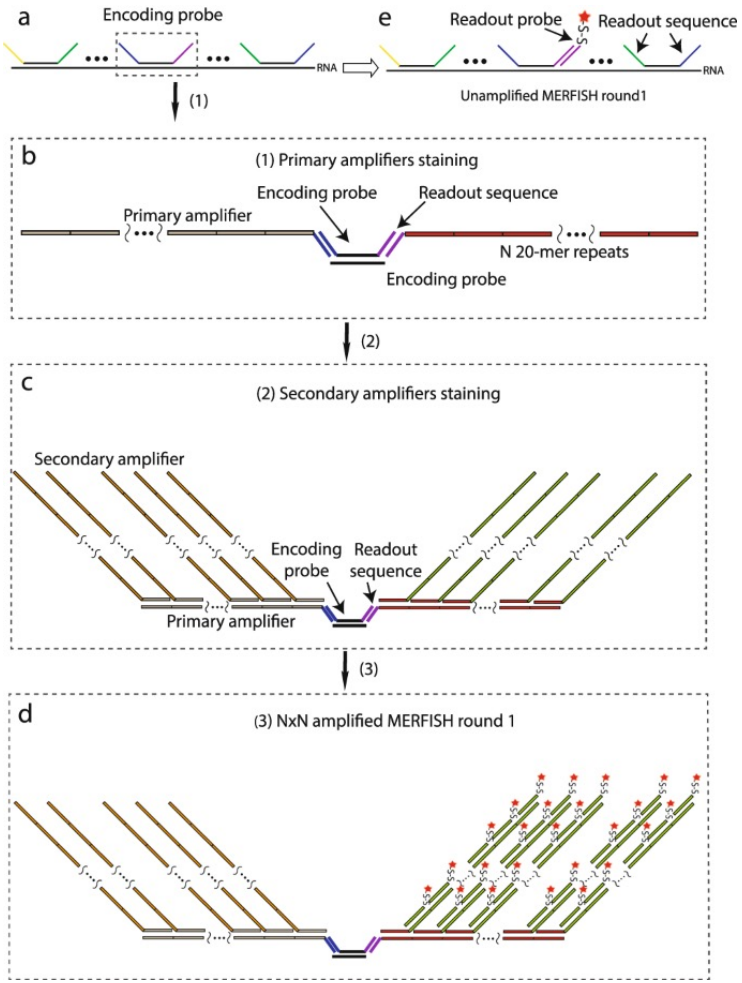
-> can MERFISH be used to profile microRNAs?

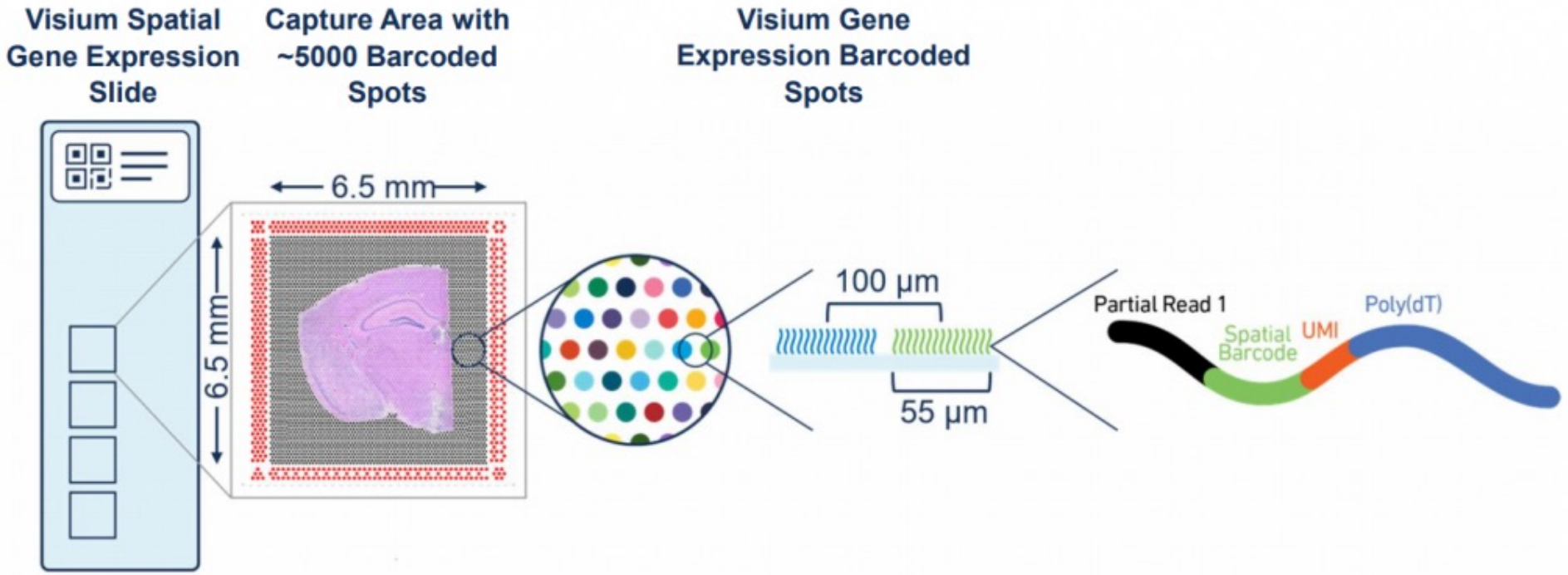


We can tile probes along an RNA in a non-overlapping manner

-> could also do overlapping

- There are many ways to amplify a fluorescent signal



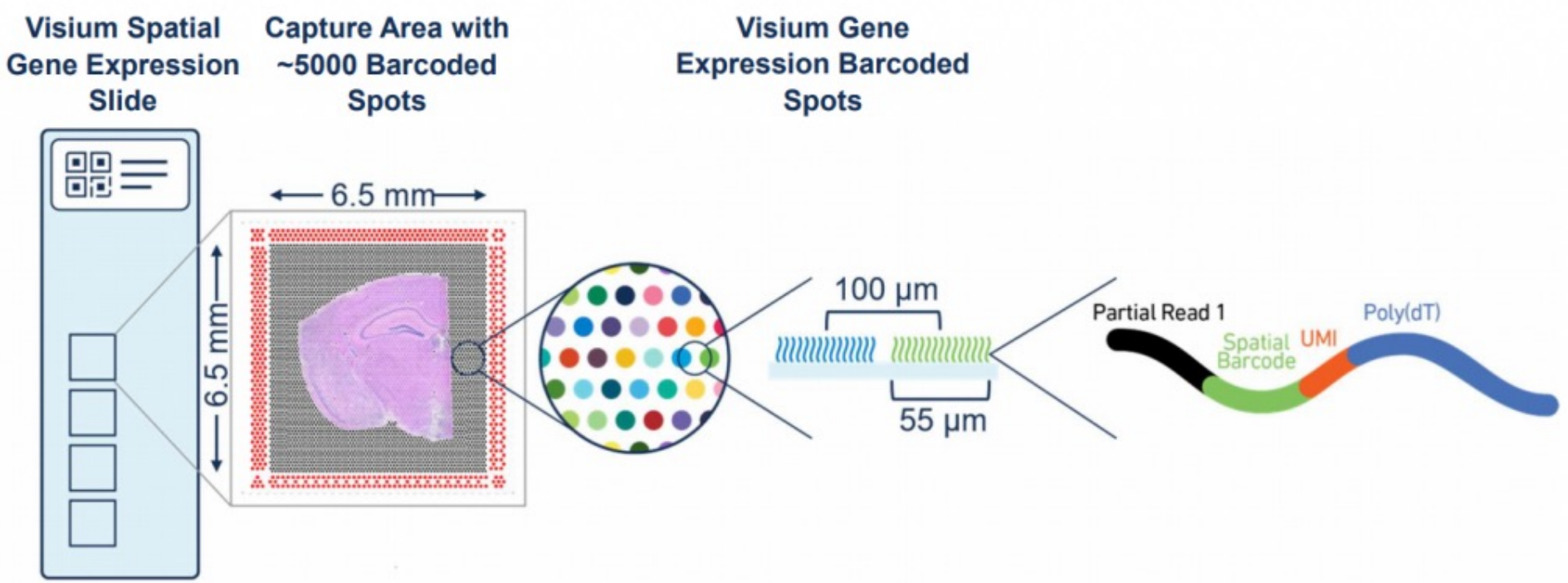


Visium (Spatial Transcriptomics) -> sequencing based approach

- Figure out RNA identity and also which spatial spot it came from

-> How many cells can fit into each of these spots?

- How large is a cell? 10um in diameter -> up to 5 cells per spot
- These measurements are not necessarily single-cell resolution



- How many mRNA molecules are there in a cell? -> 100k transcripts (1k unique transcripts)
- 100k transcripts per cell, 5 cells per spot -> 500k capture oligos

Q: how many probes are in these spots?

-> millions of probes per spot (more probes than RNA molecules)

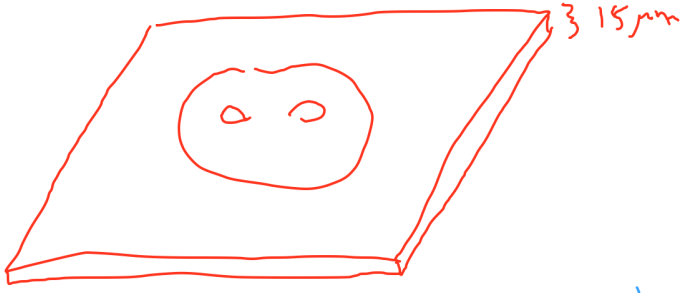
-> Do you think we're capturing everything? (MERFISH and Visium) -> "detection efficiency"

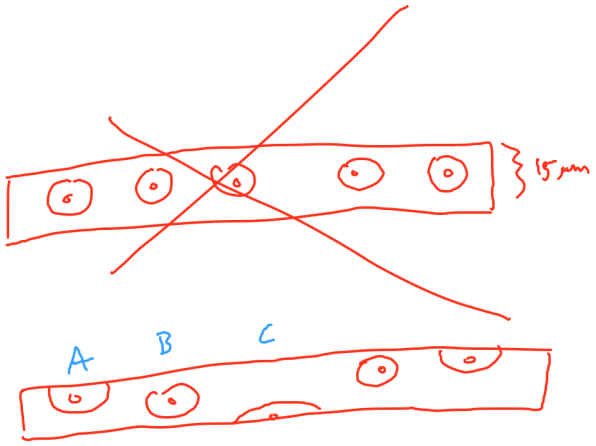
- Visium: 10% (capture of polyA, RT, sampling for sequencing) -> see this as you play with the data
- MERFISH: 90% (smFISH)

Q: if a cell is expressing 1 copy of a transcription factor, will the technology allow us to measure it?

Genomic Data Visualization

Learning objectives:
- normalization, quanlity control metrics (how data visualization can help us)





cell	total genes	gene X
A	1000	100
B	1500	150
C	500	50

→

10%
10%
10%

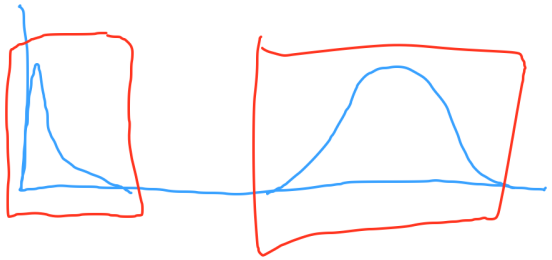
× 1e6

~~0.001%~~

"Counts per million" -> if I had sequenced or imaged 1e6 genes in every cell, what would be the proportion levels for each gene

Quality Control
-> data you guys are using are already cleaned up for you

Q: # genes detected per cell?



RNAs are sensitive to RNAase
- enzymes that degrade RNAs

MERFISH: what does it mean if we have cells with no genes?
- cell is dead
- it's not really cell

- because we are designing probes to target specific genes, maybe we just didn't pick genes that were expressed

Visium: what does it mean if we a spot with no genes?
- all the cells in that spot are dead

Sometimes experiments just fail

Q: # cells that express each gene?

- genes that are not expressed by any cell -> throw them out
Visium: data has already been cleaned for you (genes that are not expressed have been removed)

Class website -> jef.works/genomic-data-visualization
- notes, code, etc get updated on there

- homework assignments
-> no homework assigned today
-> will be a homework on creating a data visualization using datasets will be assigned next class
-> do a demonstration of submission process (Github pulls)

- if you are still less comfortable with R or ggplot, reading in the MERFISH or Visium datasets, interacting them with -> playing around or tinkering

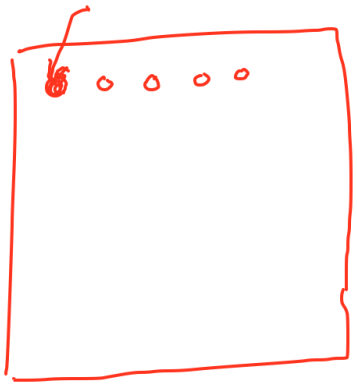
Reflection cards
-> identify areas in need of further clarification

Next class
-> diving into dimensionality reduction

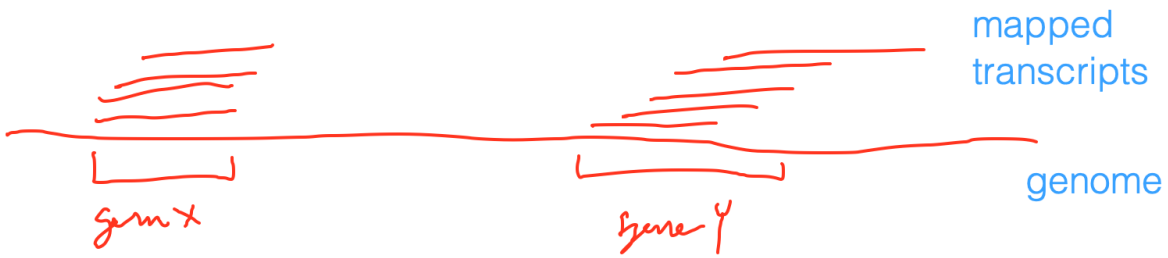
ATTATCTTGGTCTTATATATGCGGTCTA

Spatial

mRNA



-> align to a transcriptome



counting the number of
transcrpts mapped to each gene

quantify counts of genes per spot