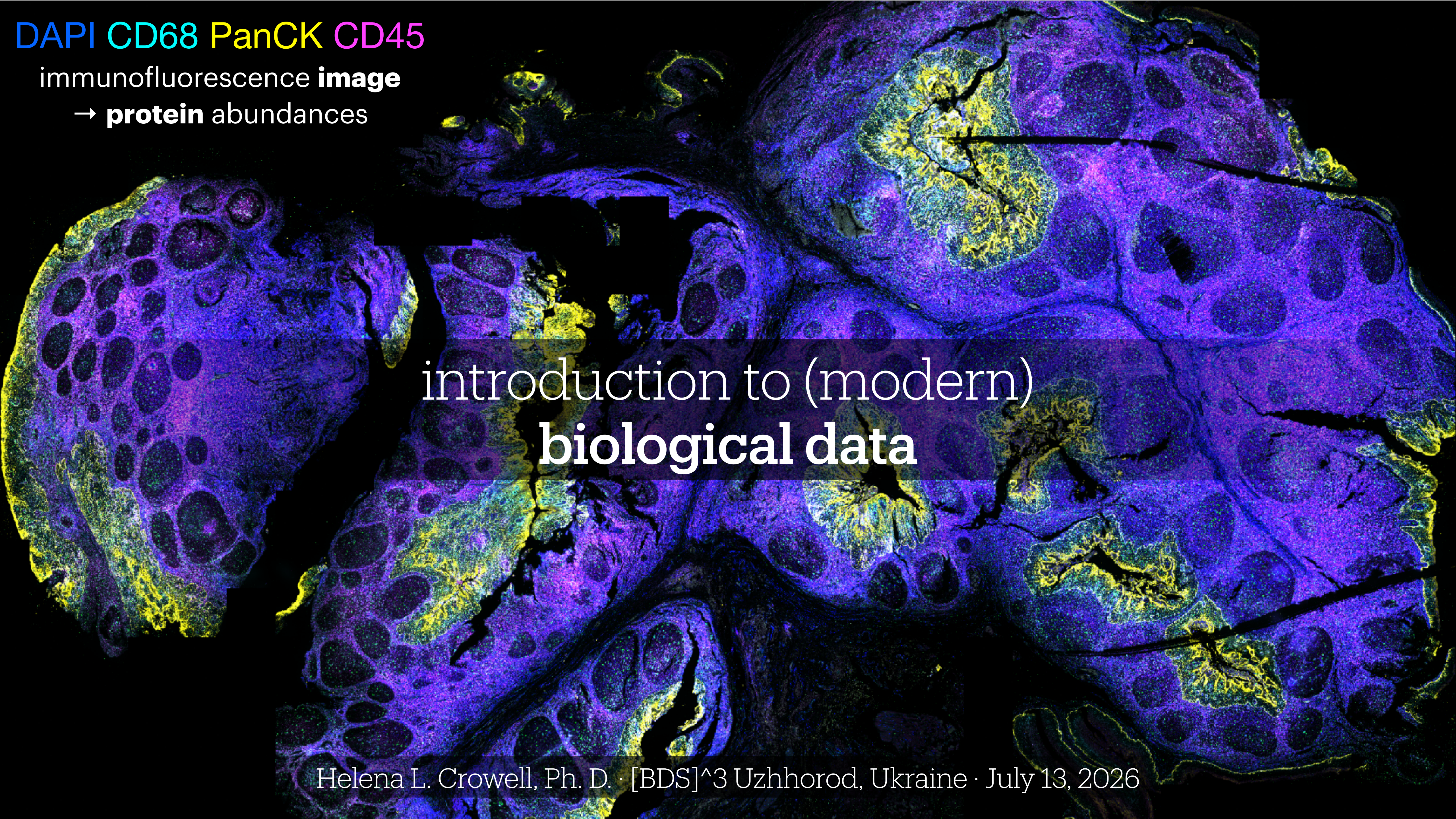




DAPI CD68 PanCK CD45

immunofluorescence **image**

→ **protein** abundances

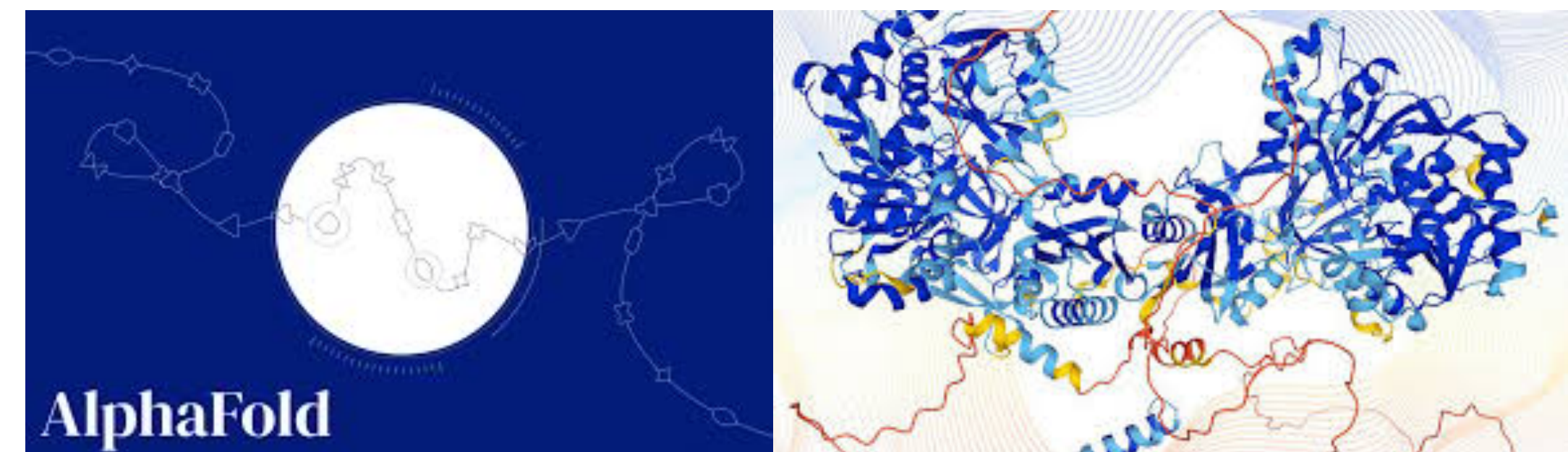
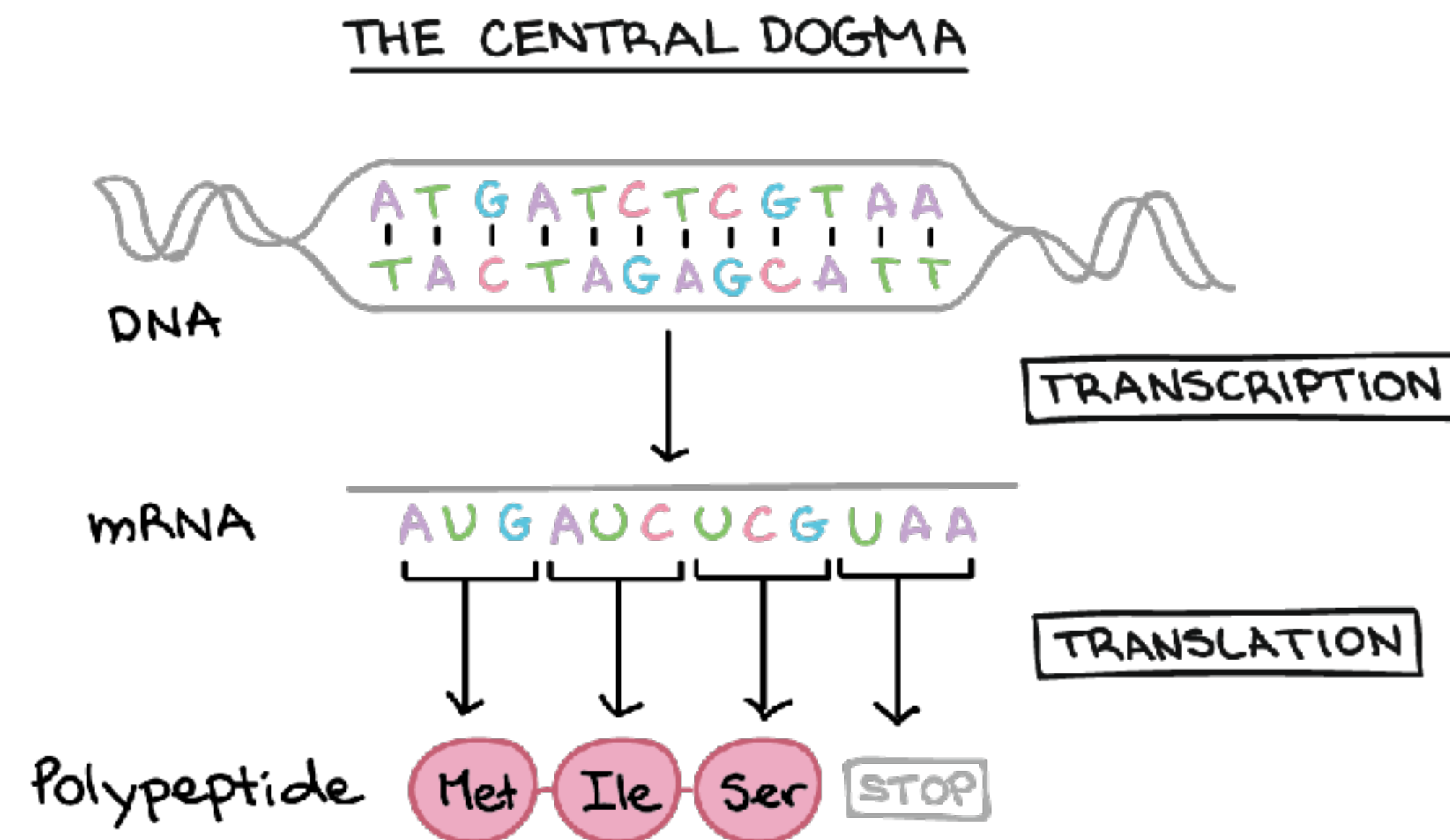
introduction to (modern) **biological data**

Helena L. Crowell, Ph. D. · [BDS]^3 Uzhhorod, Ukraine · July 13, 2026

sequences

sequence alignment

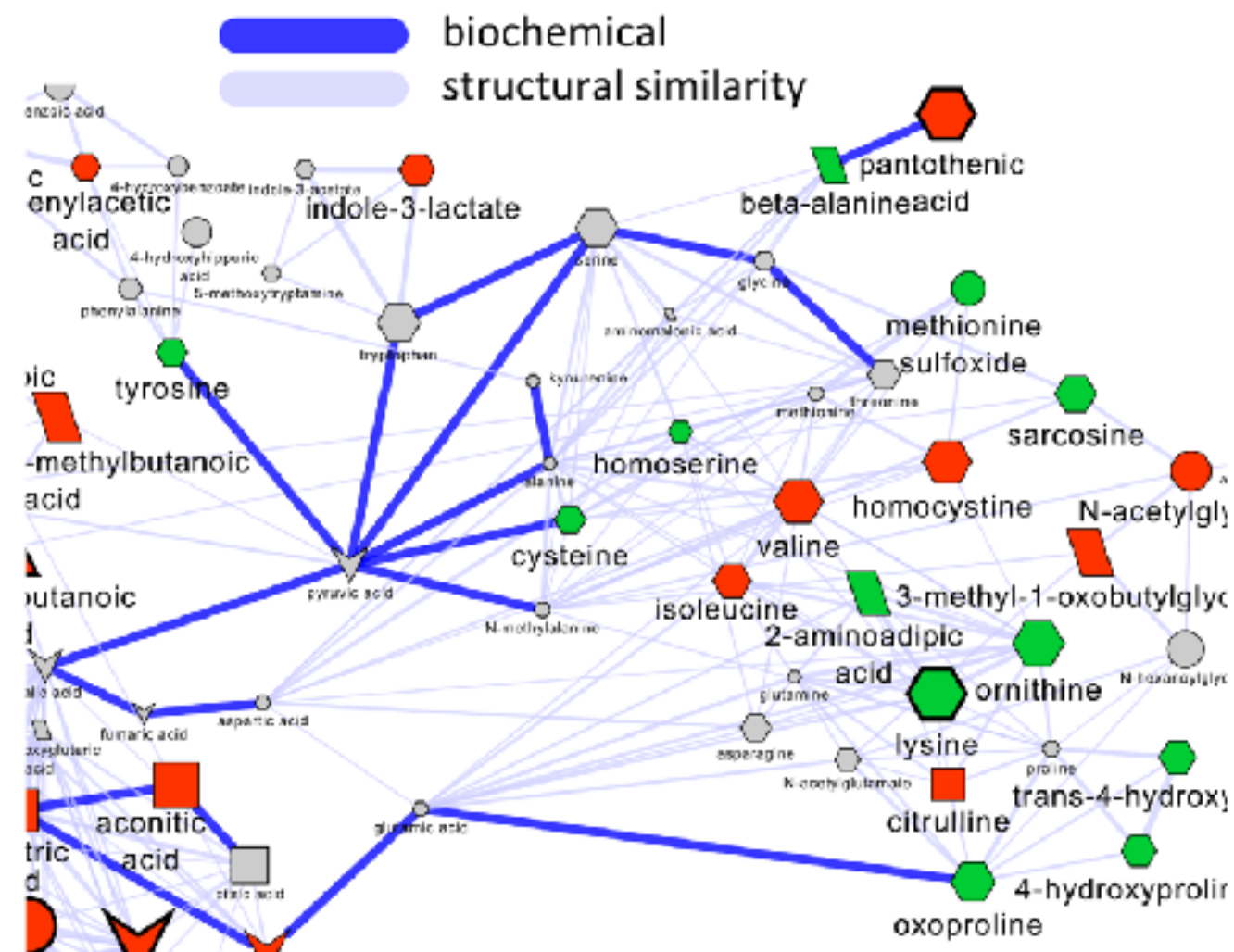
Q5E940 BOVIN	-----MREDRATWKSNYFLKIIQLDDYPKCFIVGADNVGSKOMQIRMSLRGK-AVVLGKNTMMRKAIRGHLENN--PALE	76
RLA0_HUMAN	-----MREDRATWKSNYFLKIIQLDDYPKCFIVGADNVGSKOMQIRMSLRGK-AVVLGKNTMMRKAIRGHLENN--PALE	76
RLA0_MOUSE	-----MREDRATWKSNYFLKIIQLDDYPKCFIVGADNVGSKOMQIRMSLRGK-AVVLGKNTMMRKAIRGHLENN--PALE	76
RLA0_RAT	-----MREDRATWKSNYFLKIIQLDDYPKCFIVGADNVGSKOMQIRMSLRGK-AVVLGKNTMMRKAIRGHLENN--PALE	76
RLA0_CHICK	-----MREDRATWKSNYFLKIIQLDDYPKCFIVGADNVGSKOMQIRMSLRGK-AVVLGKNTMMRKAIRGHLENN--PALE	76
RLA0_RANBY	-----MREDRATWKSNYFLKIIQLDDYPKCFIVGADNVGSKOMQIRMSLRGK-AVVLGKNTMMRKAIRGHLENN--PALE	76
Q7ZUG3 BRAKE	-----MREDRATWKSNYFLKIIQLDDYPKCFIVGADNVGSKOMQIRMSLRGK-AVVLGKNTMMRKAIRGHLENN--PALE	76
RLA0 ICTPU	-----MREDRATWKSNYFLKIIQLDDYPKCFIVGADNVGSKOMQIRMSLRGK-AVVLGKNTMMRKAIRGHLENN--PALE	76
RLA0 DROME	-----MREDRATWKSNYFLKIIQLDDYPKCFIVGADNVGSKOMQIRMSLRGK-AVVLGKNTMMRKAIRGHLENN--PALE	76
RLA0 DICDI	-----MSGAG-SKRKKLFTEKATKLETTIDKMIYAEADFGVSSQLQIRKSIRGI-GAVLMGKNTMIRKIVIRDLADSK--PELD	75
Q54LPO DTCDI	-----MSGAG-SKRKNVFTKKATKLETTIDKMIYAEADFGVSSQLQIRKSIRGI-GAVLMGKNTMIRKIVIRDLADSK--PELD	75
RLA0 PLAFH	-----MAKLSKQQRKQMYTEKLSLITQQYKSLIVHVDNVGSKOMASVRRSLRGK-ATIMGKNTMIRKIVIRDLADSK--PELD	76
RLA0 SULAC	-----MGLAVTTTKKIAKKVDEVAELTEKLETKITITIANIEGFPADKLEIRKKLRGK-ADIKVTEKNTLPHIALEHAG--DQLE	79
RLA0 SULTO	-----MRIMAVITQERKIAKKIIEKVELEKLEKREHTITIANIEGFPADKLEIRKKLRGK-ADIKVTEKNTLPHIALEHAG--DQLE	80
RLA0 SULSO	-----MKRIALALQQRKVASWLEEVKELTELKNSNTILIGNLEGGPADKLEIRKKLRGK-ATIKVTEKNTLPHIALEHAG--DQLE	80
RLA0 AERPE	MSVVSIVGQMYKREKIPKWTLMLELEELFSKIRVVFADDTGTFVYVQVRRKLLWKK-YPMVAKKRIILRAMKAAGLE--LQDN	86
RLA0 PYRAE	MMLAIGKRRYVTRQYPAKVKIVSEATELLQKYDYVFLDHLGSLRILHEVRYLLARY-GVTKIKPILFKIAFTKVYGG--LEAE	85
RLA0 METAC	MAEERHHTETPQKKRDEIENIKELIQSKVFGMVGIDGILATKMKIRADLDKV-AVLEKVRNTLLEALNQIG--CTIP	78
RLA0 METMA	MAEERHHTETPQKKRDEIENIKELIQSKVFGMVGIDGILATKMKIRADLDKV-AVLEKVRNTLLEALNQIG--CTIP	78
RLA0 ARCFU	MAAVRGS--PDEYKVRVVEIKRMISSEKDVVATVSPFNVPAGOMQIRMSLRGK-AVLEKVRNTLLEALNQIG--CTIP	75
RLA0 METKA	MAVKAQKQEPSEYEPKVAEKKRREVKELKELMDEYENVGLYDLGIPAPOLQIRAKLRERDTIIRSRNTLLEALNQIG--CTIP	88
RLA0 METTH	MAHVAEWKKKEVQELHDLIKSYEVVGLANLADIPAROLQMKRQTLRDS-ALIKSRNTLLEALNQIG--CTIP	74
RLA0 METTL	MITAENSEKIAFWKIEEVNKLKLLKNGQIVALVDMEVPAAROLQIRAKLRERDTIIRSRNTLLEALNQIG--CTIP	82
RLA0 METVA	MIDAKSEHKIAFWKIEEVNKLKLLKNGQIVALVDMEVPAAROLQIRAKLRERDTIIRSRNTLLEALNQIG--CTIP	82
RLA0 METJA	METKVKARVAPKIEEVNKLKLLKNGQIVALVDMEVPAAROLQIRAKLRERDTIIRSRNTLLEALNQIG--CTIP	81
RLA0 PYRAD	MAHVAEWKKKEVEELANLIRSYDVIALVDVSSMAYPLSQMRRLIRENGLLRVSRNTLLEALNQIG--CTIP	77
RLA0 PYRHO	MAHVAEWKKKEVEELANLIRSYDVIALVDVSSMAYPLSQMRRLIRENGLLRVSRNTLLEALNQIG--CTIP	77
RLA0 PYRFU	MAHVAEWKKKEVEELANLIRSYDVIALVDVSSMAYPLSQMRRLIRENGLLRVSRNTLLEALNQIG--CTIP	77
RLA0 PYRKO	MAHVAEWKKKEVEELANLIRSYDVIALVDVSSMAYPLSQMRRLIRENGLLRVSRNTLLEALNQIG--CTIP	76
RLA0 HALMA	MSAESEKKTETIPKWKQEEVDALIVEHIESYTESVGVVNLGIEPQKOLQMKRDLHGT-AELKVRNTLLEALNQIG--CTIP	79
RLA0 HALVO	MSAESEKKTETIPKWKQEEVDALIVEHIESYTESVGVVNLGIEPQKOLQMKRDLHGT-AELKVRNTLLEALNQIG--CTIP	79
RLA0 HALSA	MSAESEKKTETIPKWKQEEVDALIVEHIESYTESVGVVNLGIEPQKOLQMKRDLHGT-AELKVRNTLLEALNQIG--CTIP	79
RLA0 THEAO	MKEVYQQKKELVNEITIRIKAKSKVALVDLAGIKRQIQQIRGKNRGK-INLEKVIKLLLEKALENLGQ--EKLS	72
RLA0 THEVO	MKEVYQQKKELVNEITIRIKAKSKVALVDLAGIKRQIQQIRGKNRGK-INLEKVIKLLLEKALENLGQ--EKLS	72
RLA0 PICTO	MTEPAQKKIDFVKNLENEINSRKKVAATVSKIRNNFQKIRNSIRGK-ARIKVSRARLLRLAIENTGK--NNIV	72
ruler	1.....10.....20.....30.....40.....50.....60.....70.....80.....90	



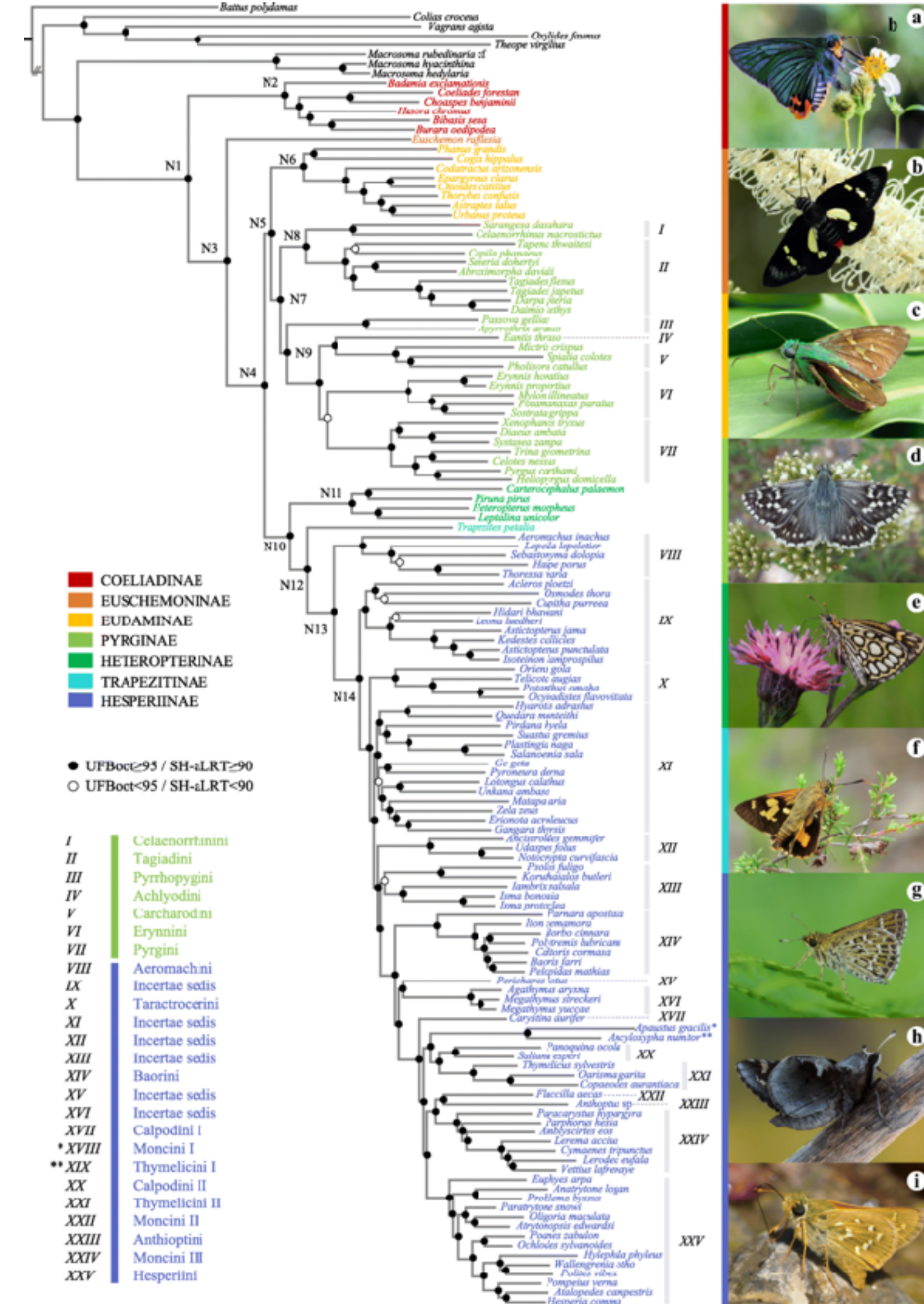
structural prediction

graphs & trees

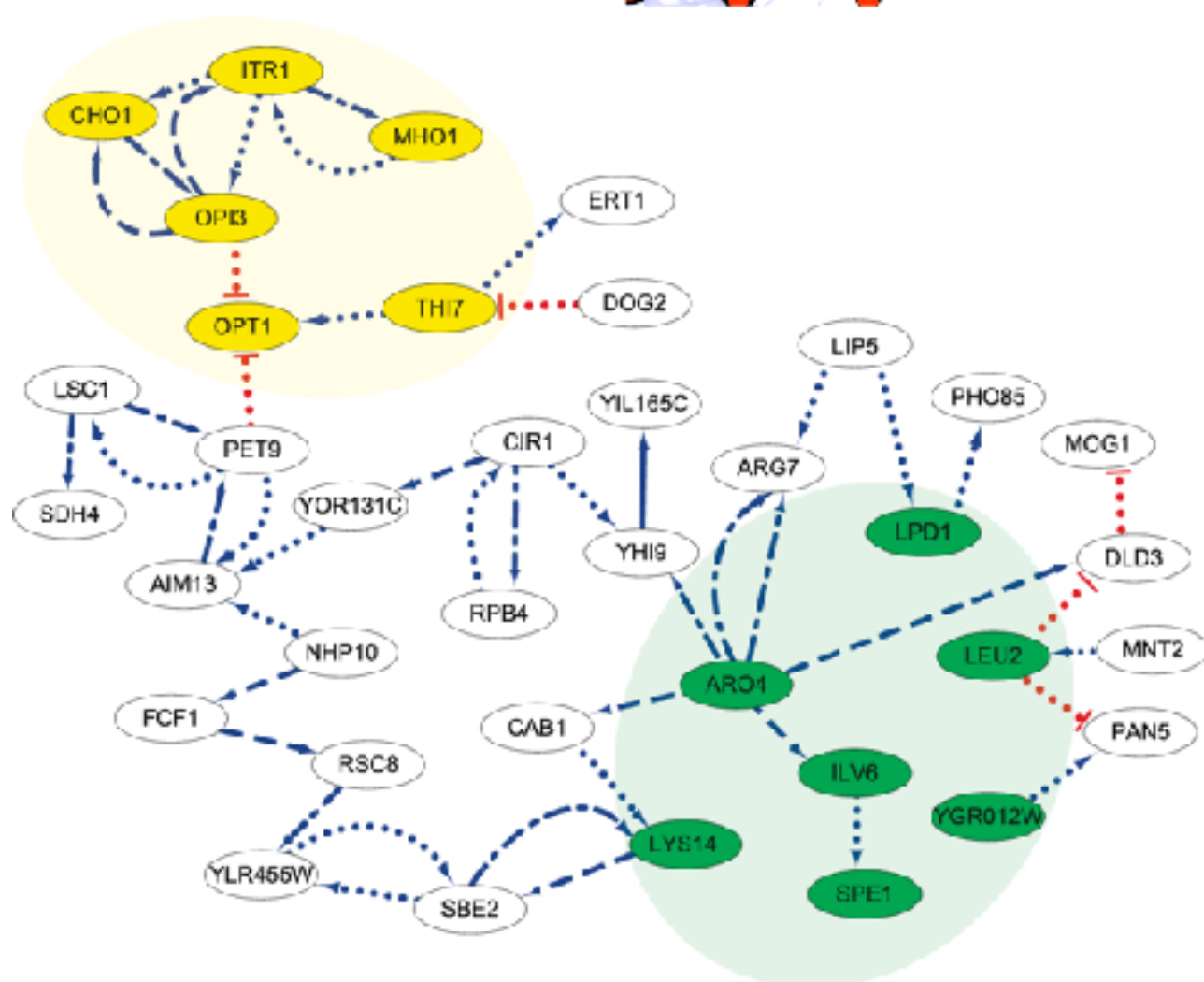
metabolites



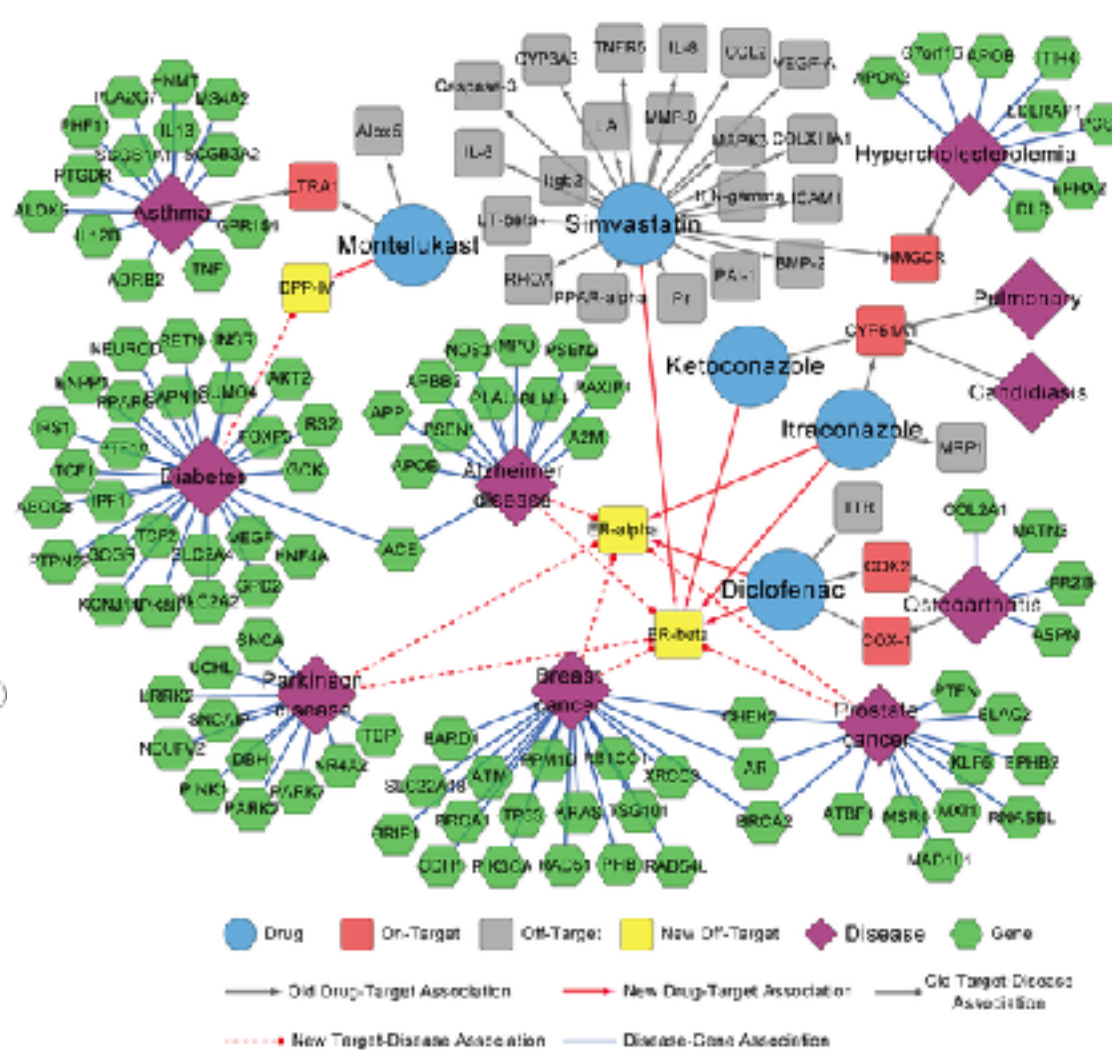
phylogenies



gene regulation

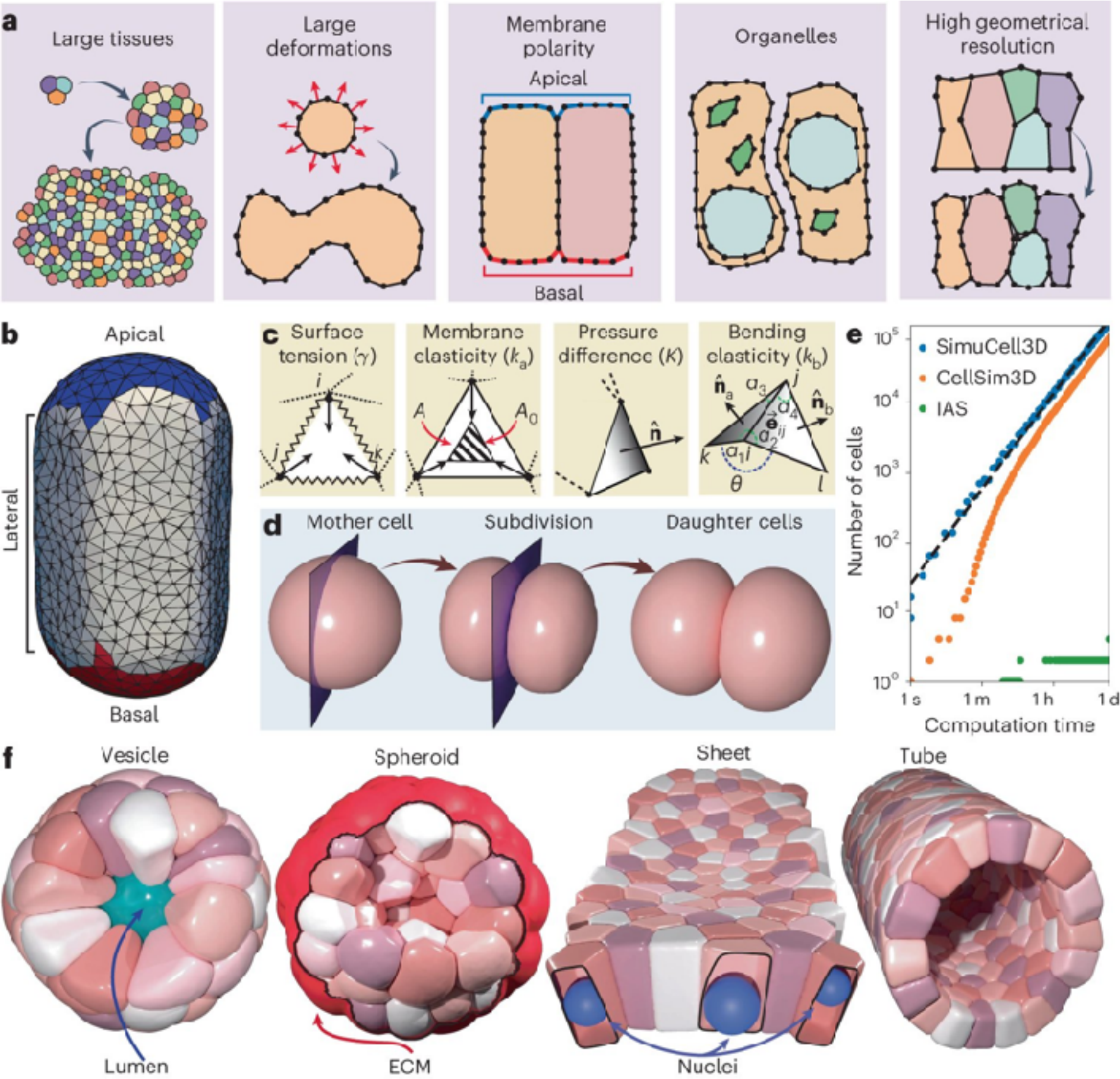


drug targets

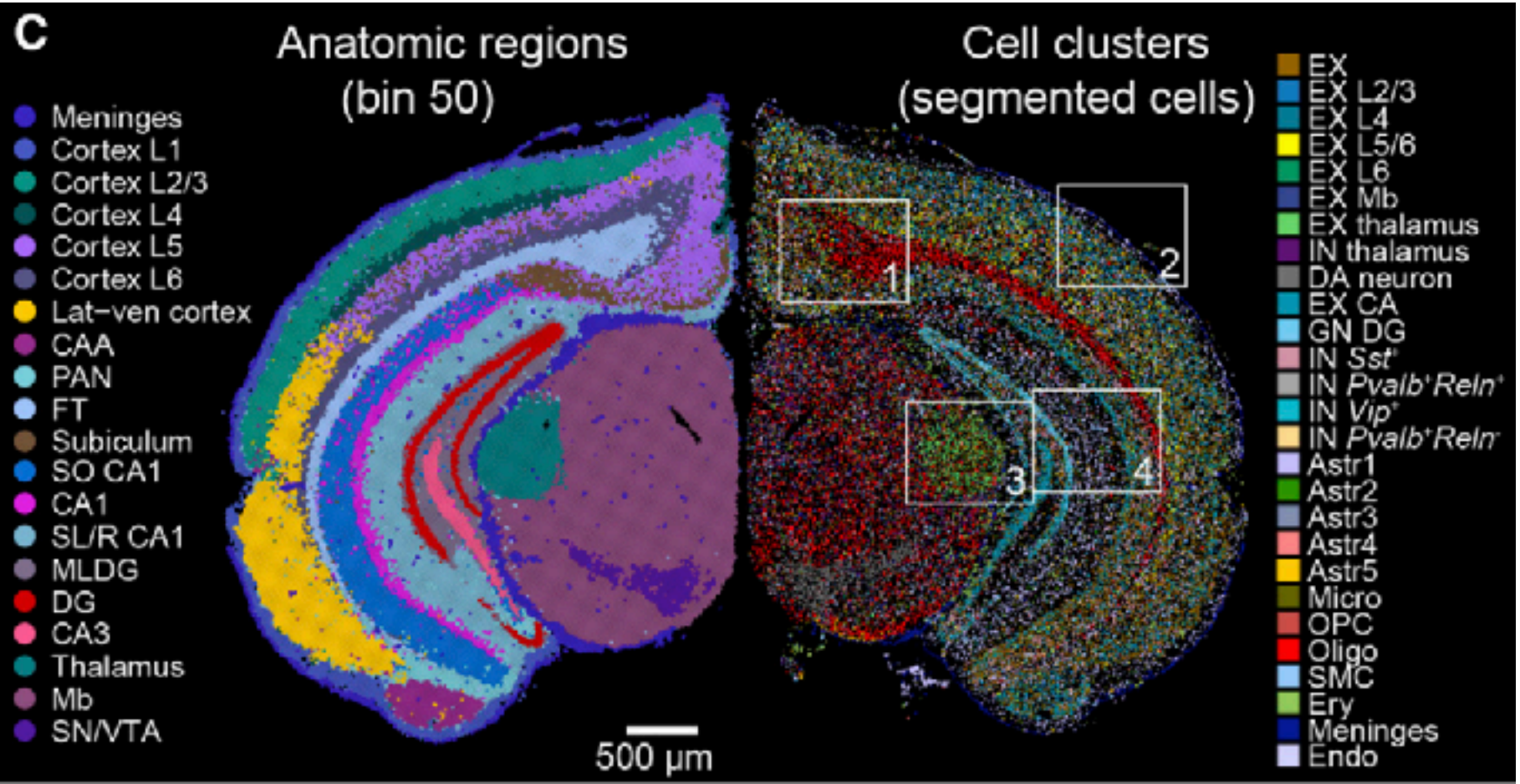


geometries, patterns & shapes

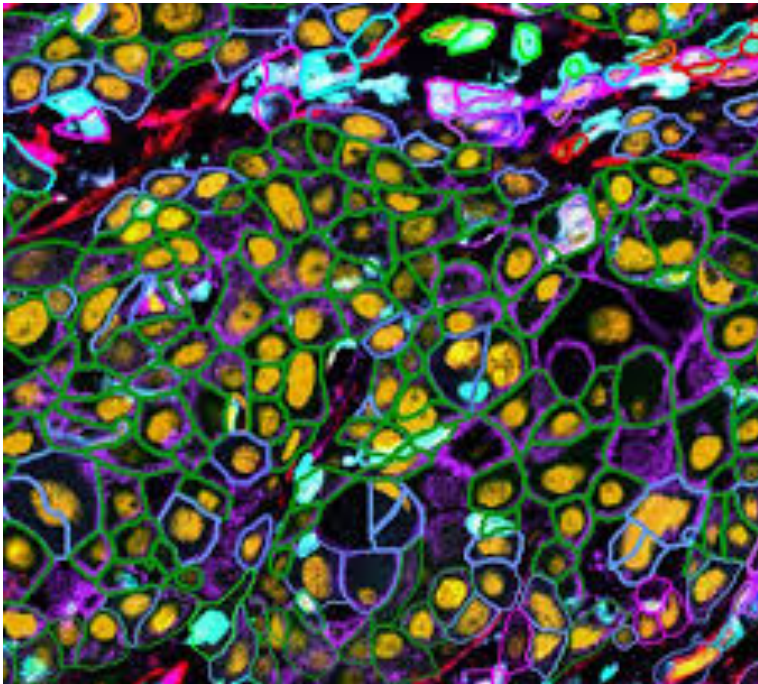
mechanistic modeling



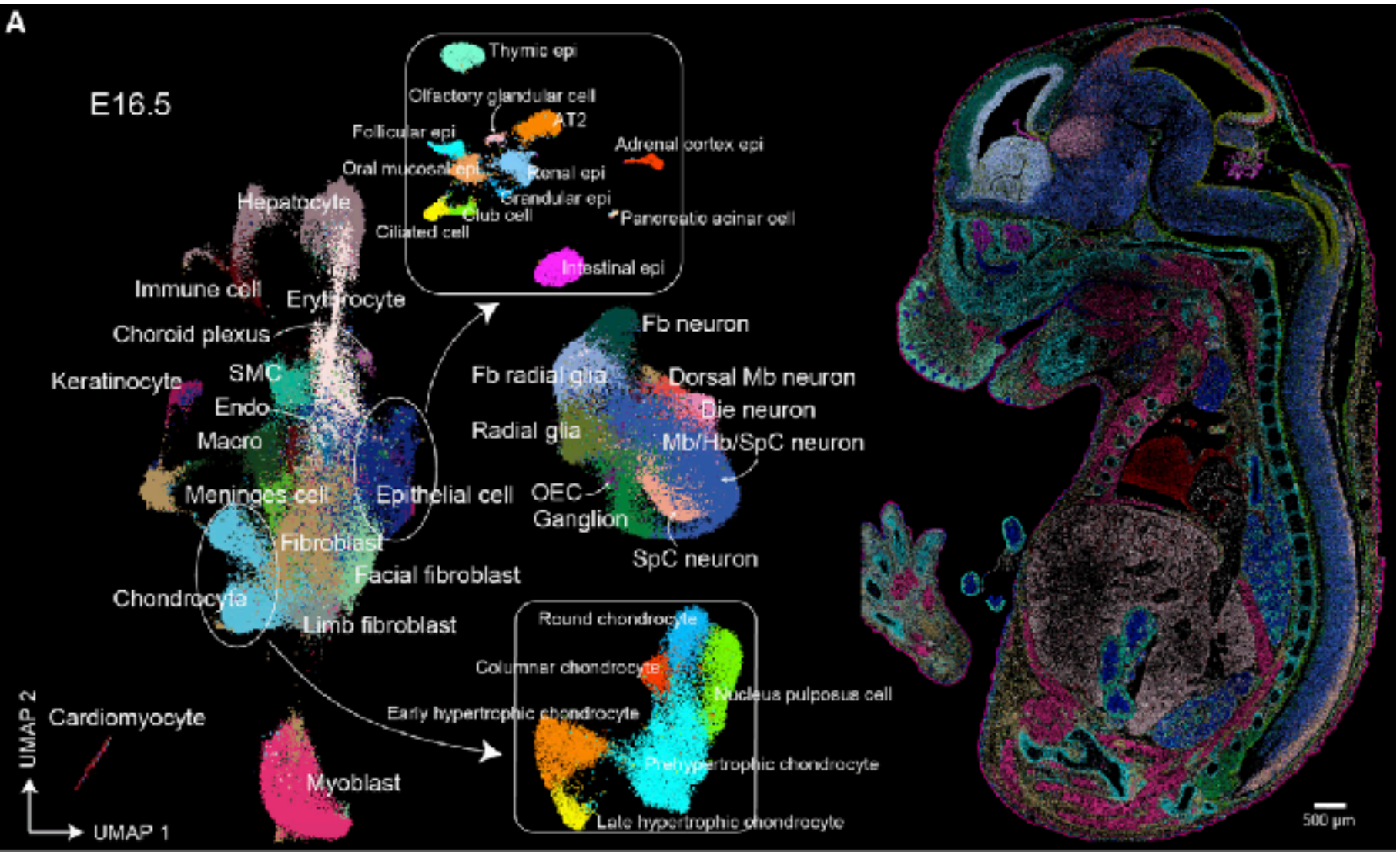
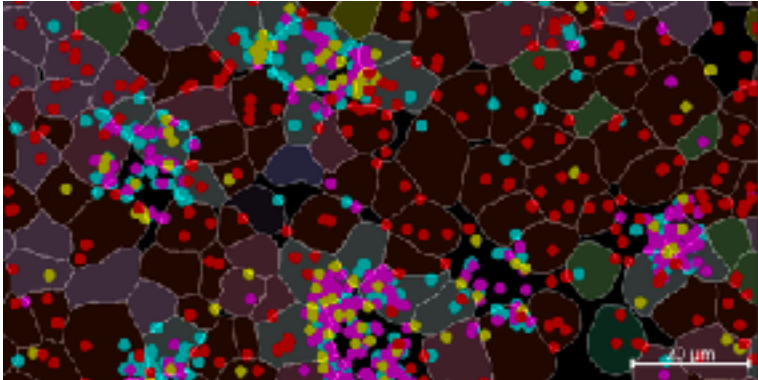
anatomical

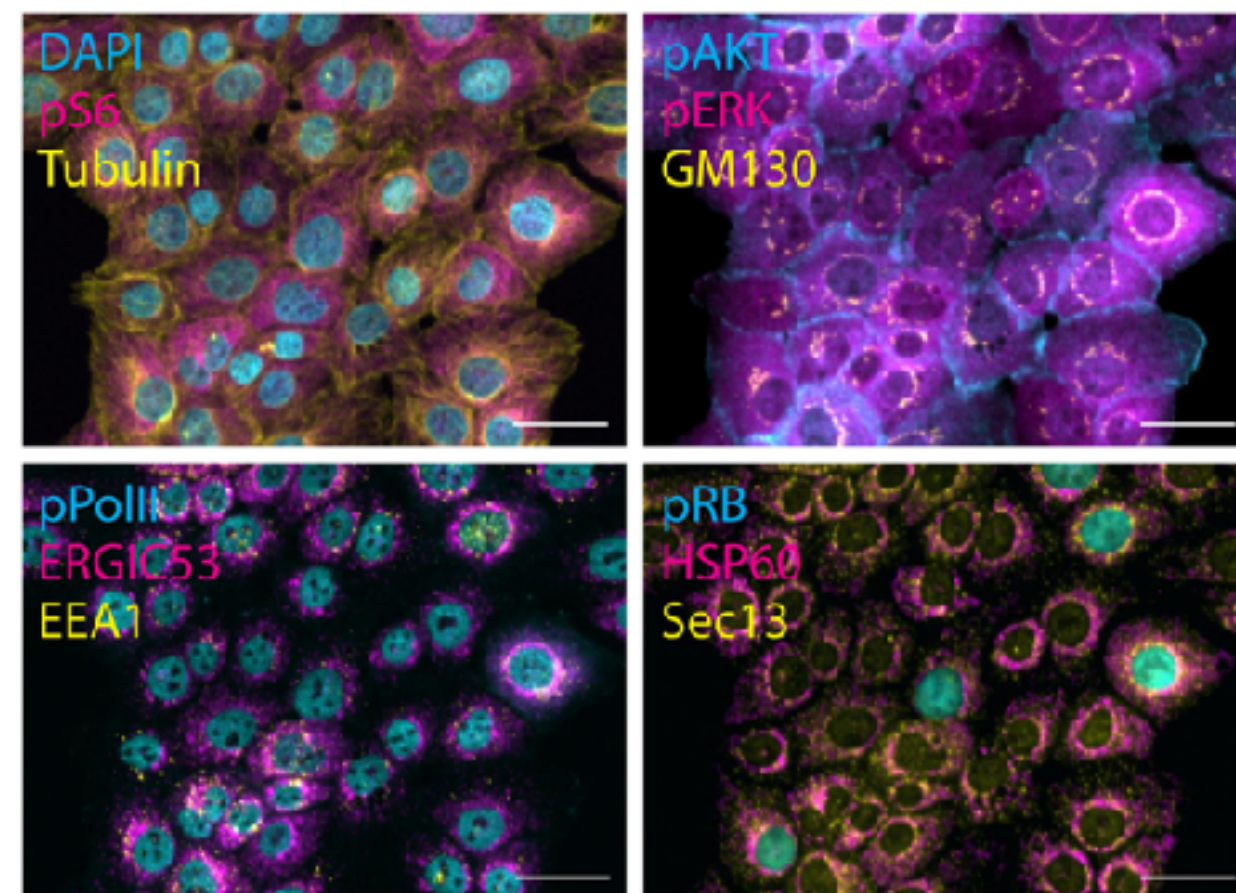


cellular

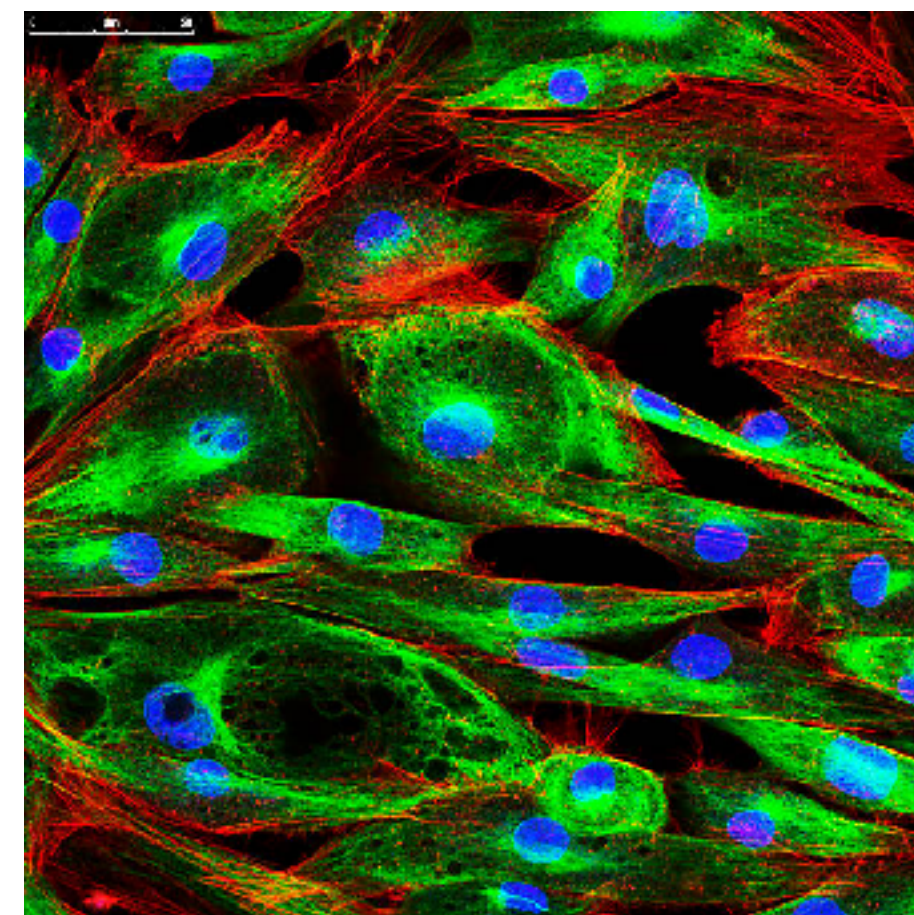


subcellular

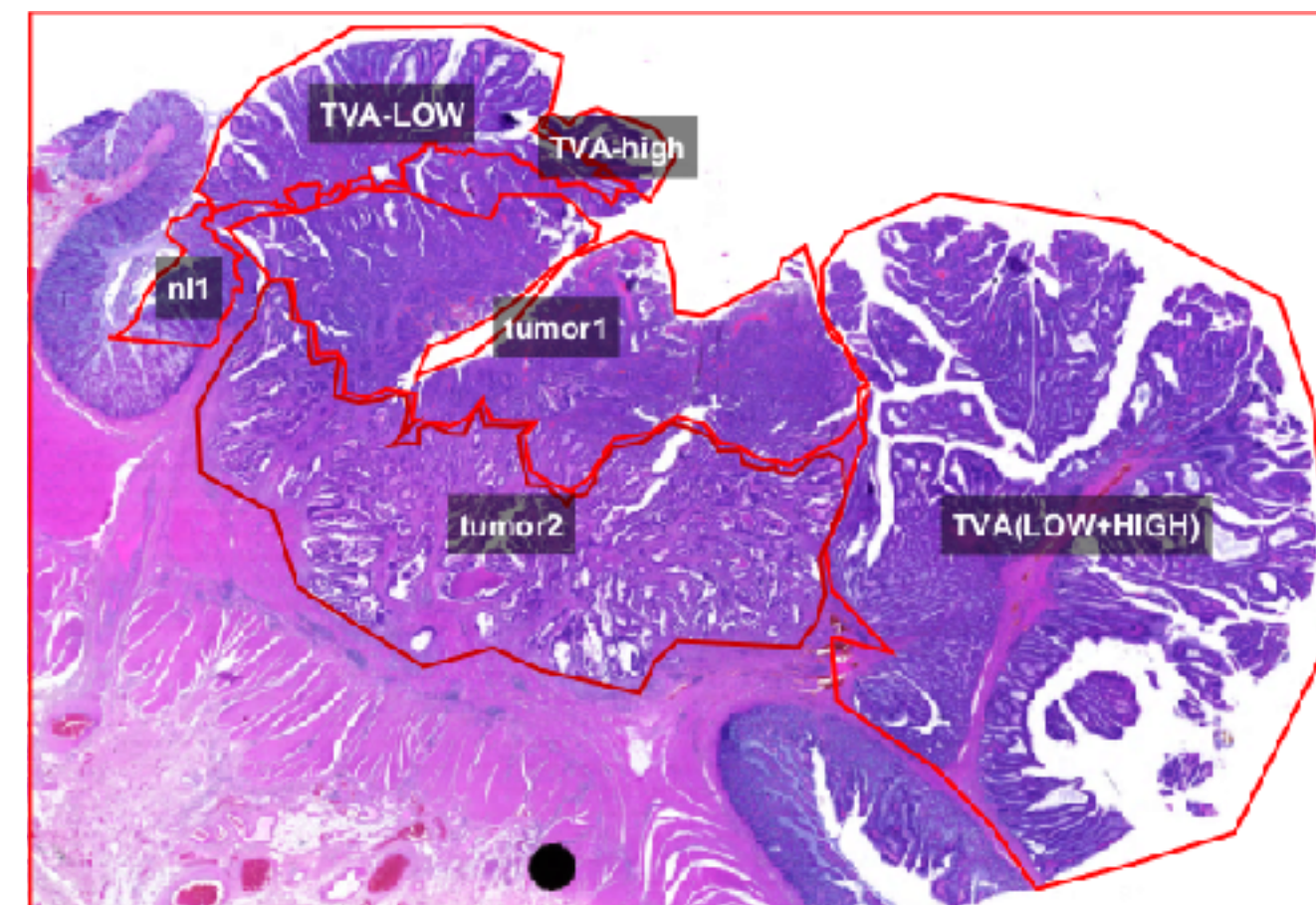




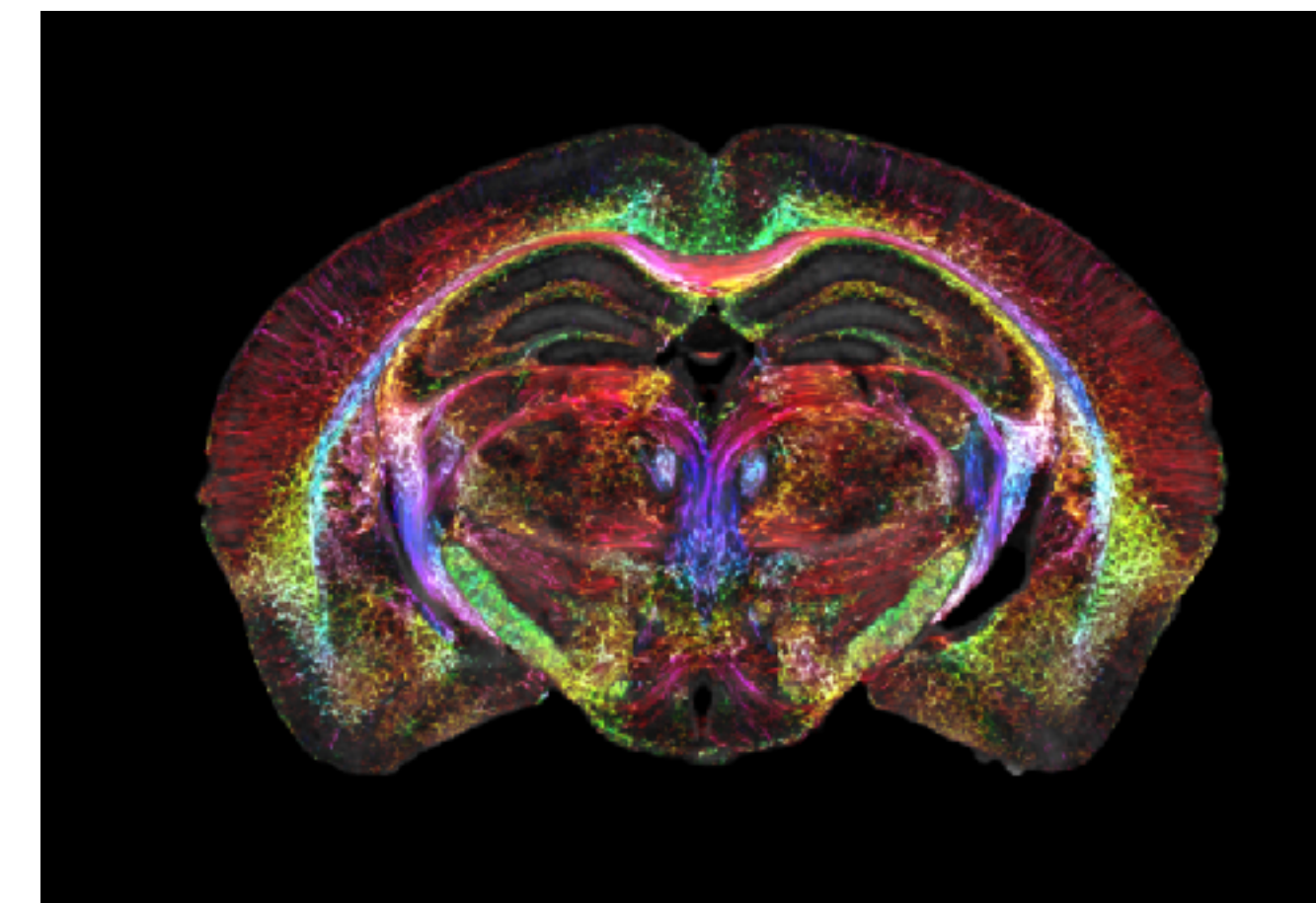
multiplexed imaging
(single cells)



confocal microscopy
(fibroblast fillaments)



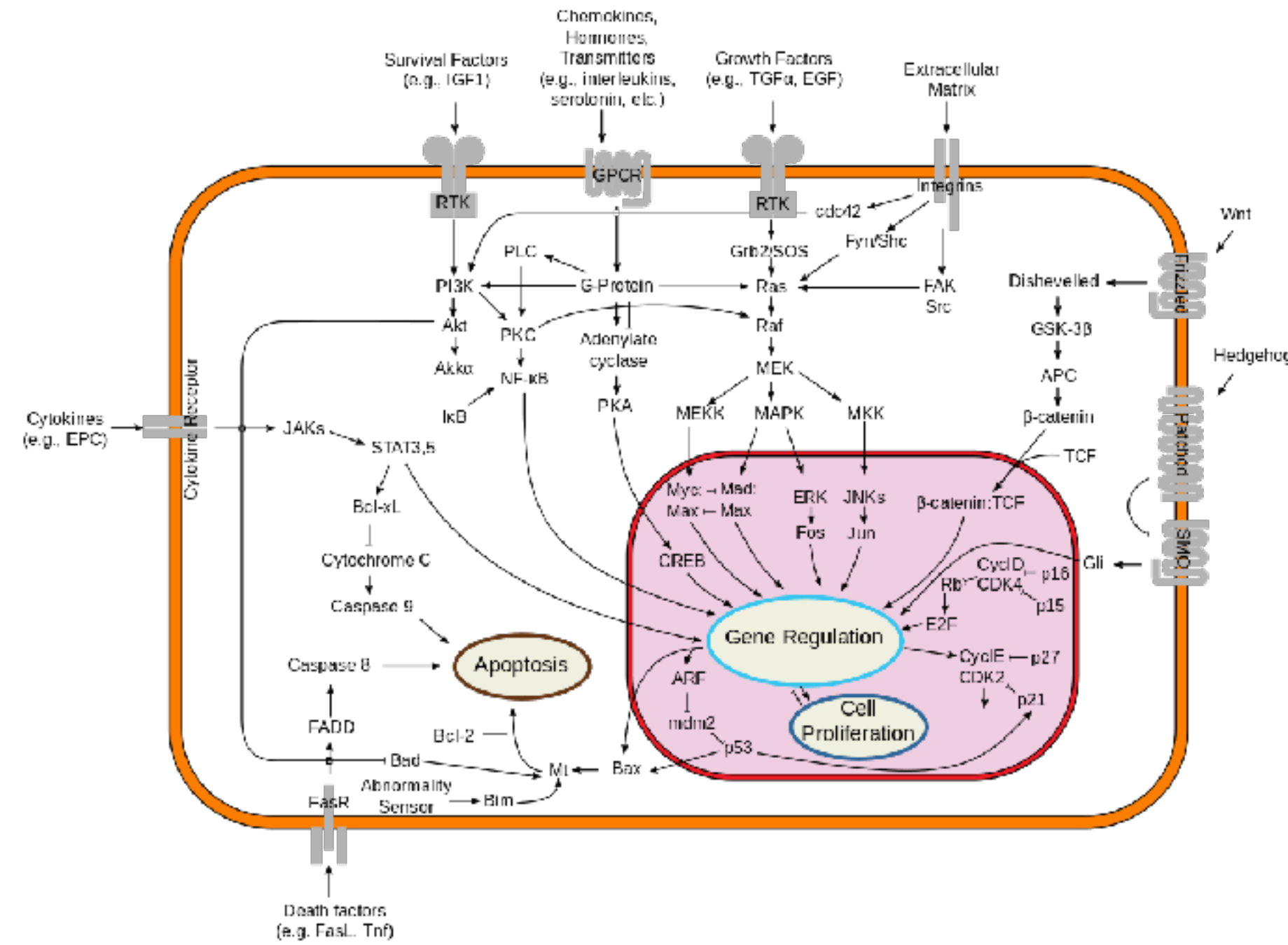
H&E staining
(colorectal cancer)



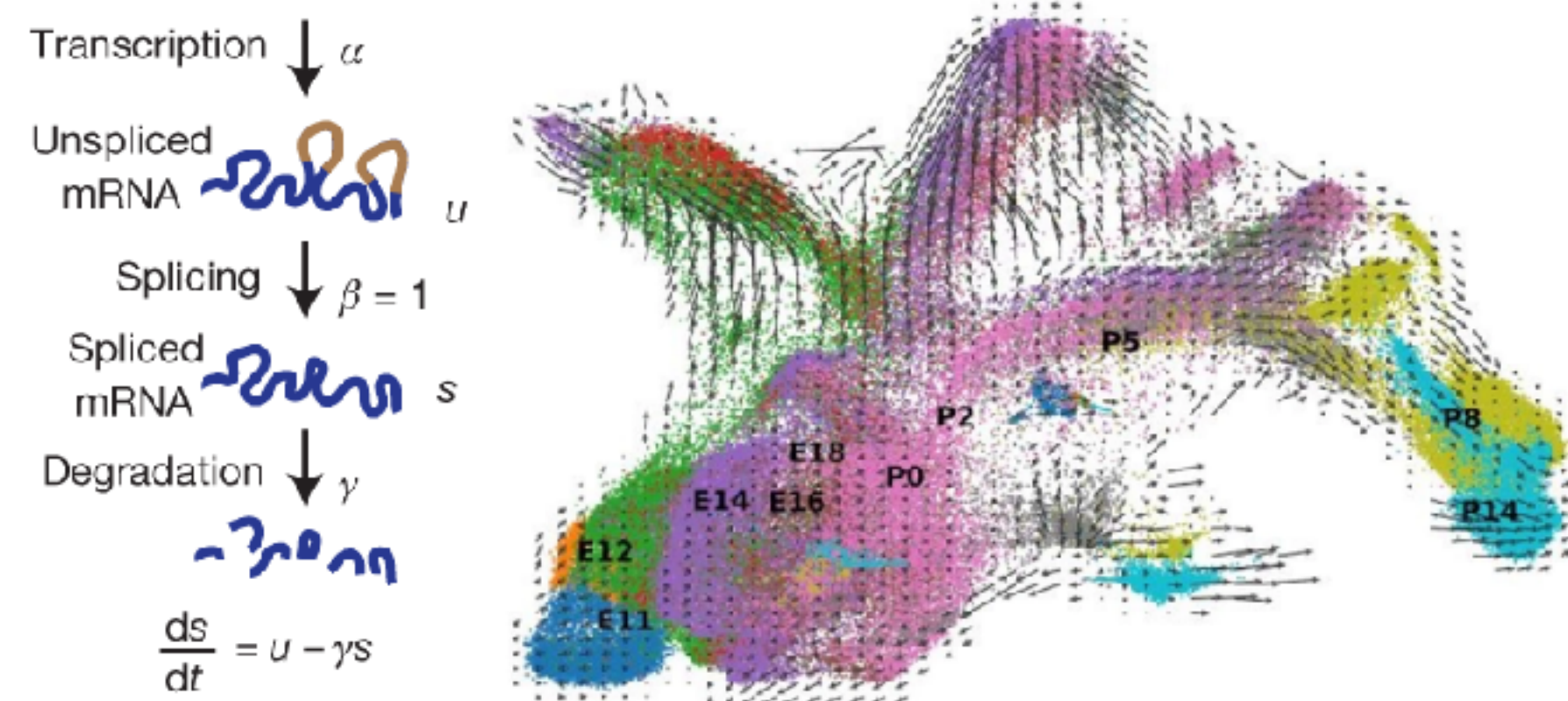
light sheet microscopy
(mouse brain)

models

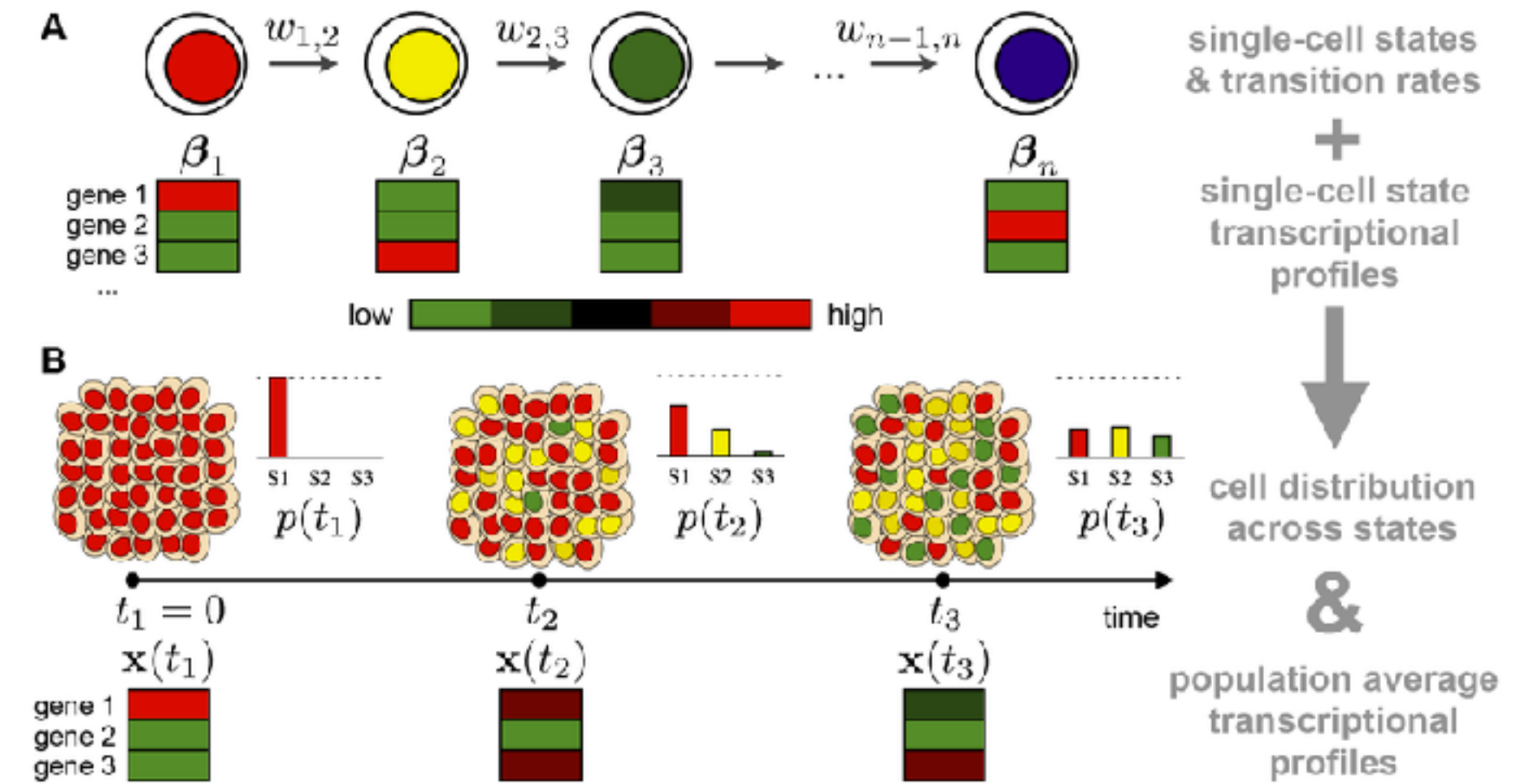
cellular models



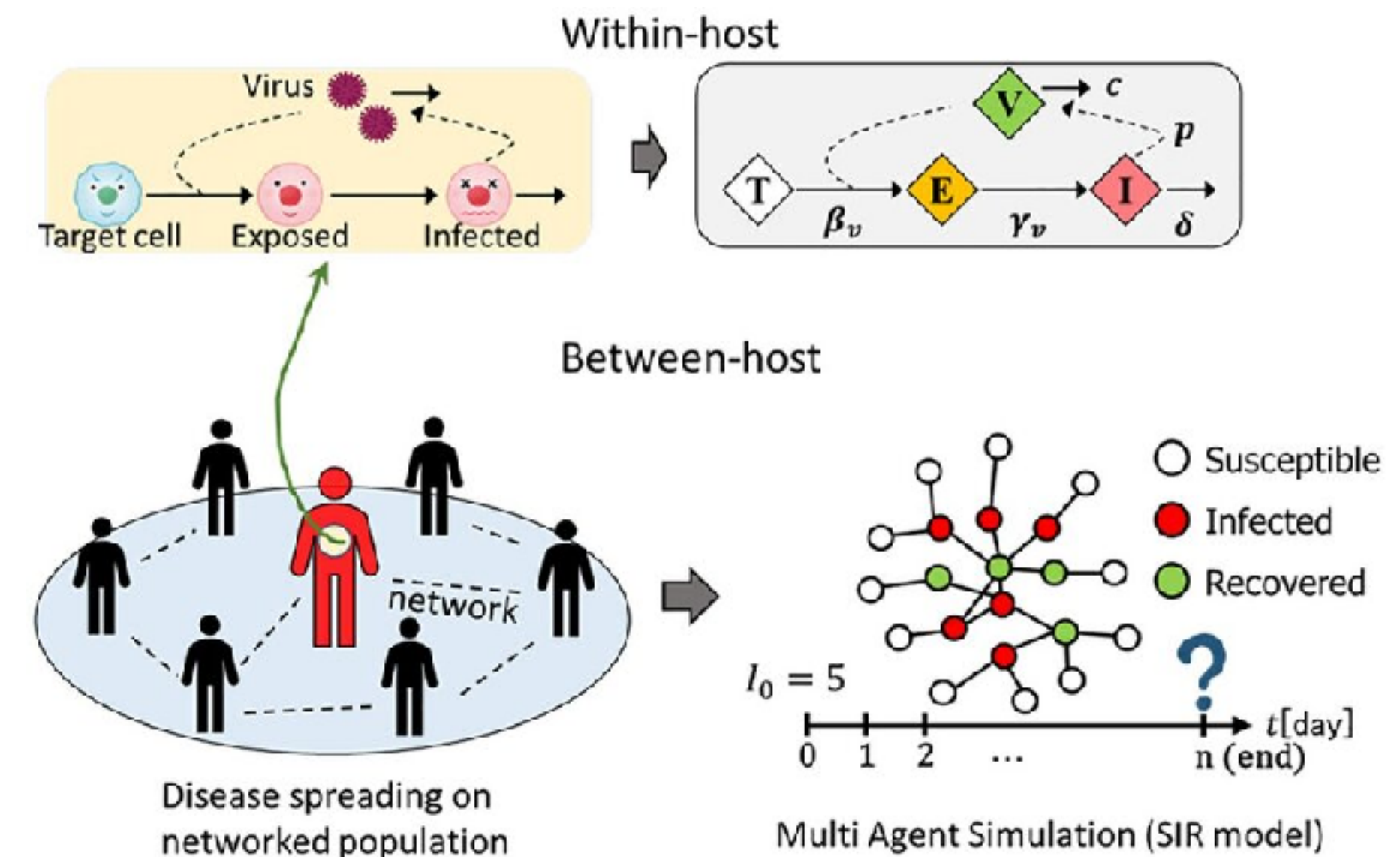
RNA velocity



cell-level Markov models



viral dynamics



data needs biotechnology

Fluidigm CyTOF

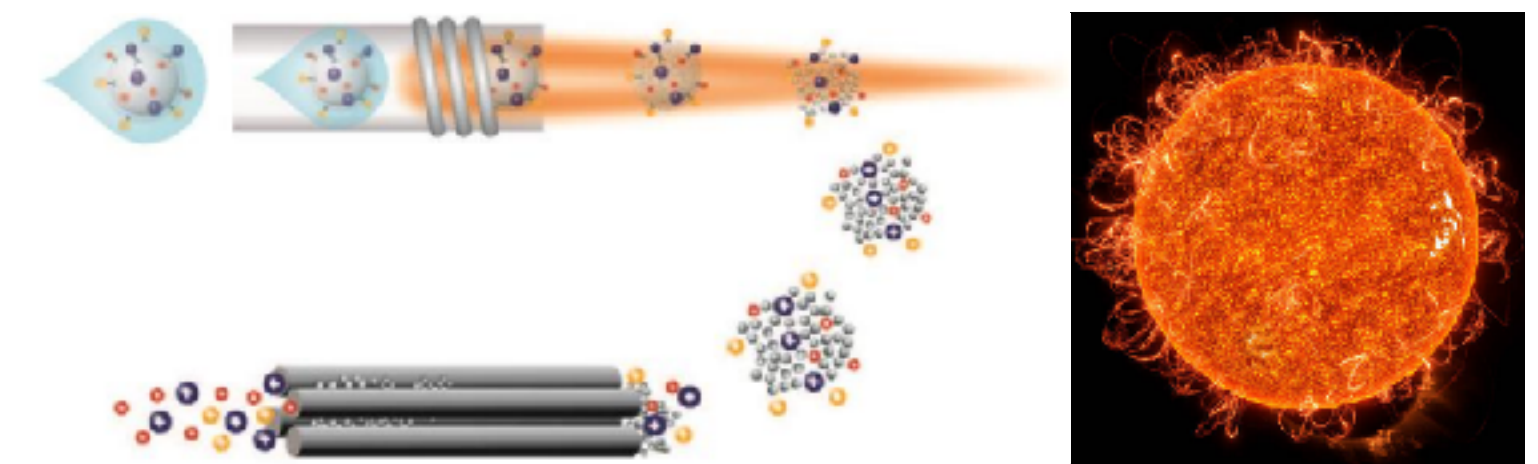
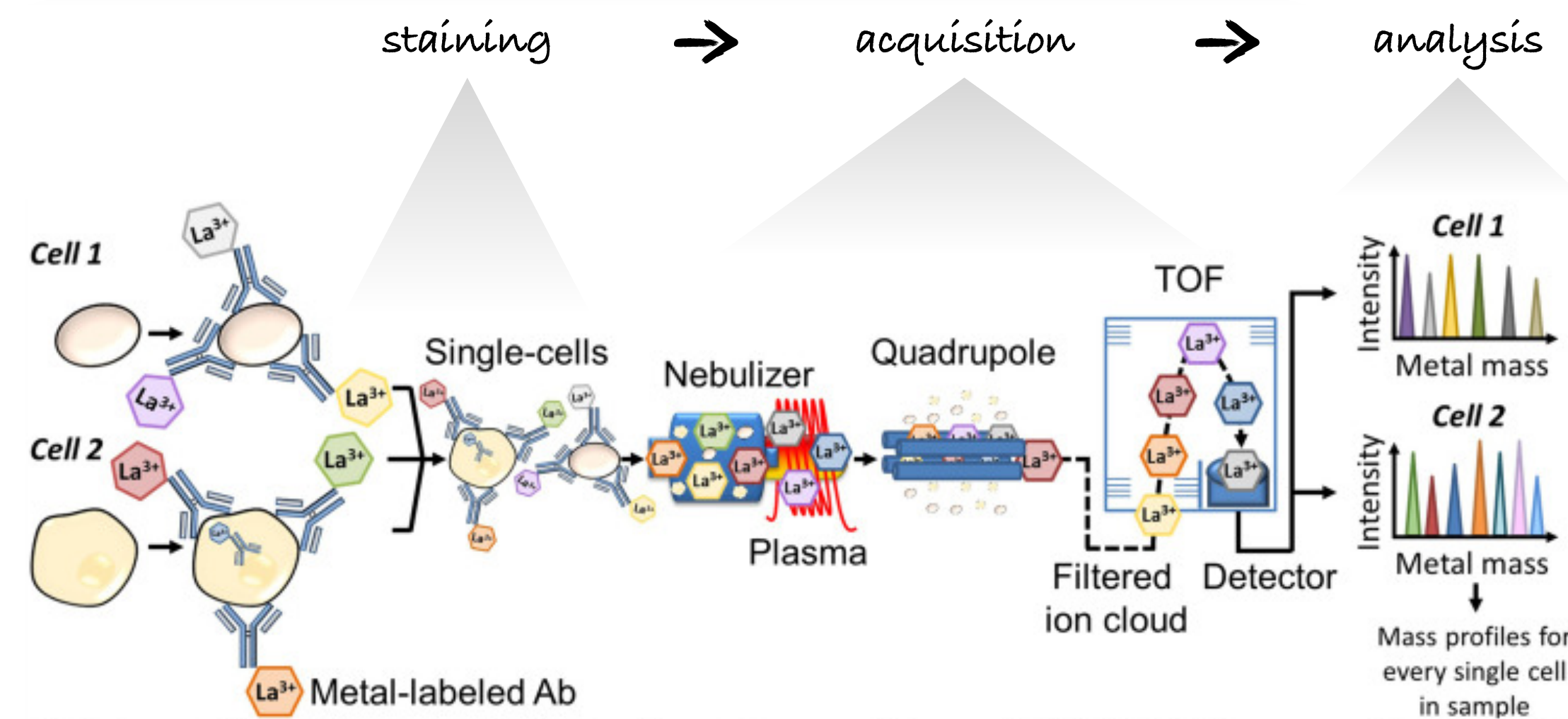


10x Chromium



Burker CosMx

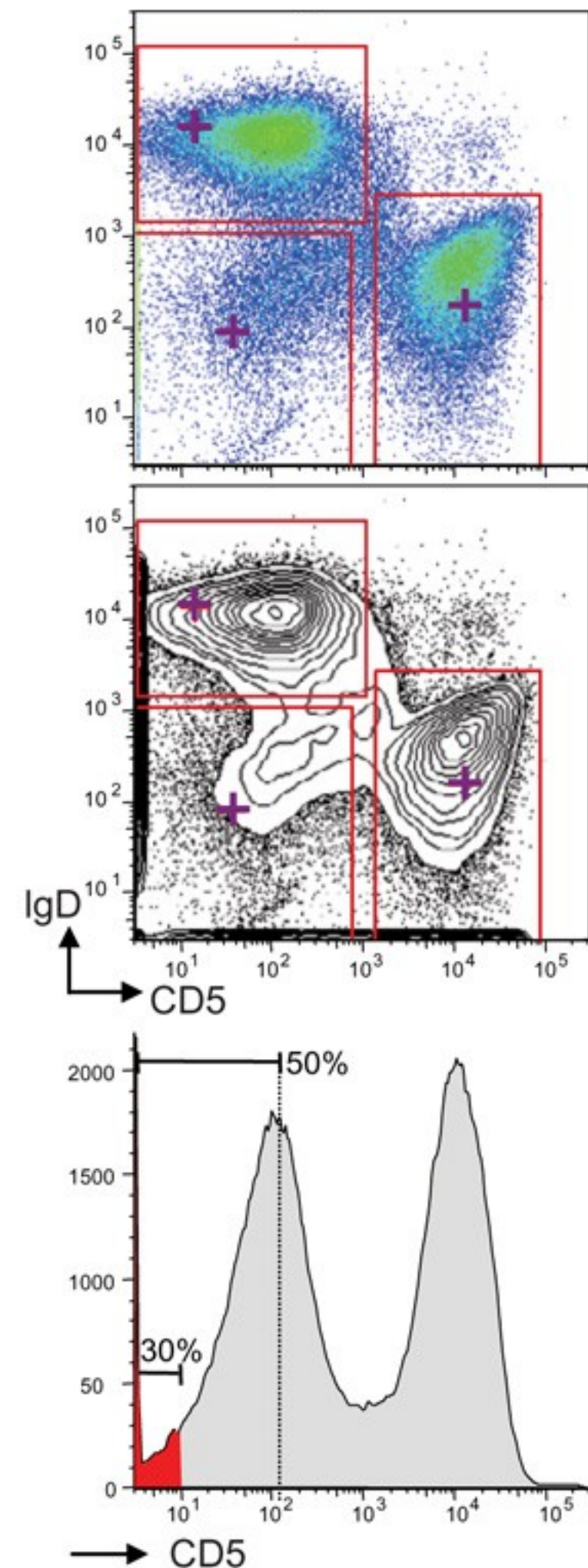
10x Xenium



CyTOF plasma torch: 7,200°C
> surface of the sun: 5,600°C

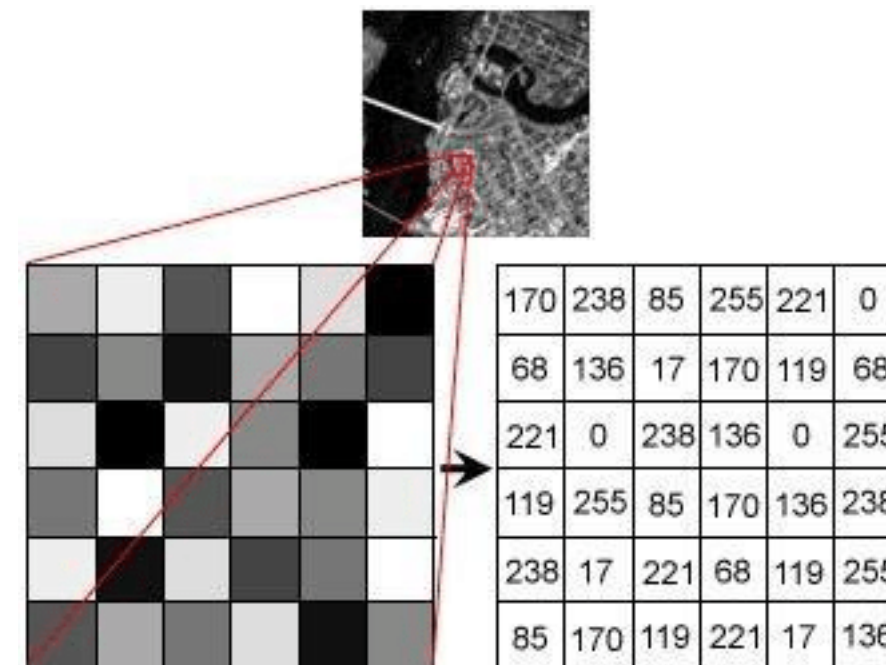
biology by the numbers

fluorescence intensities (FACS)



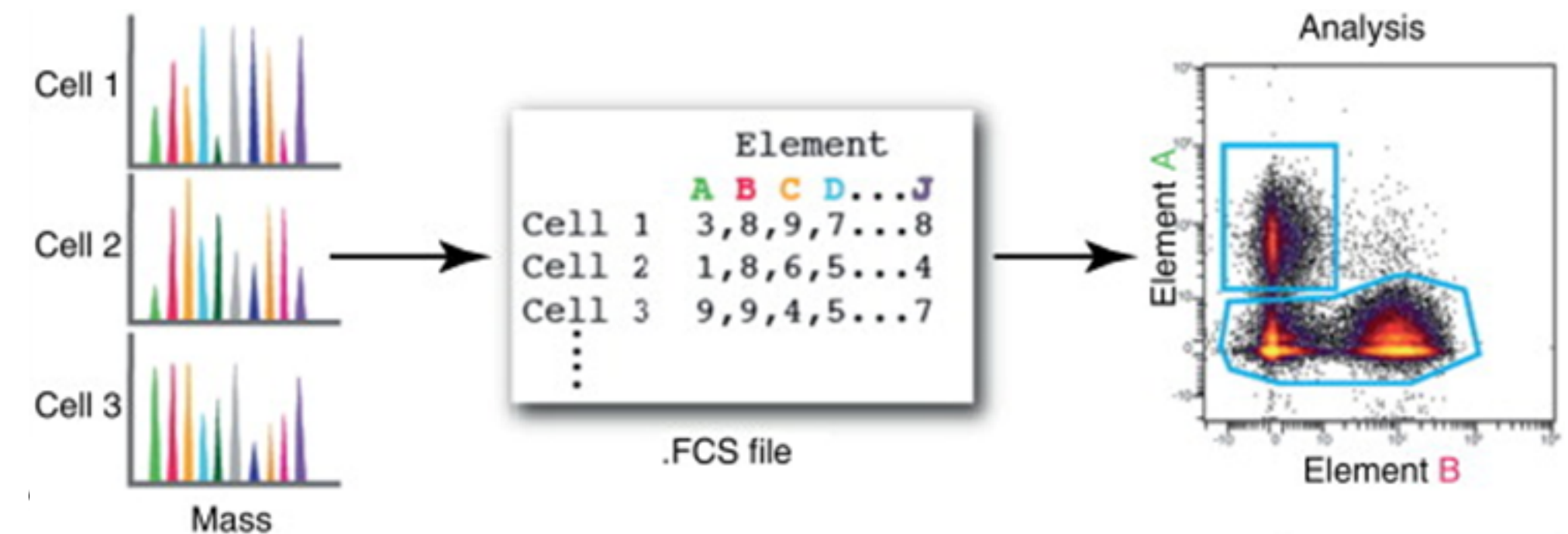
gene counts (sc/RNA-seq)

gene	ctrl_1	ctrl_2	exp_1	exp_1
geneA	10	11	56	45
geneB	0	0	128	54
geneC	42	41	59	41
geneD	103	122	1	23
geneE	10	23	14	56
geneF	0	1	2	0
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...

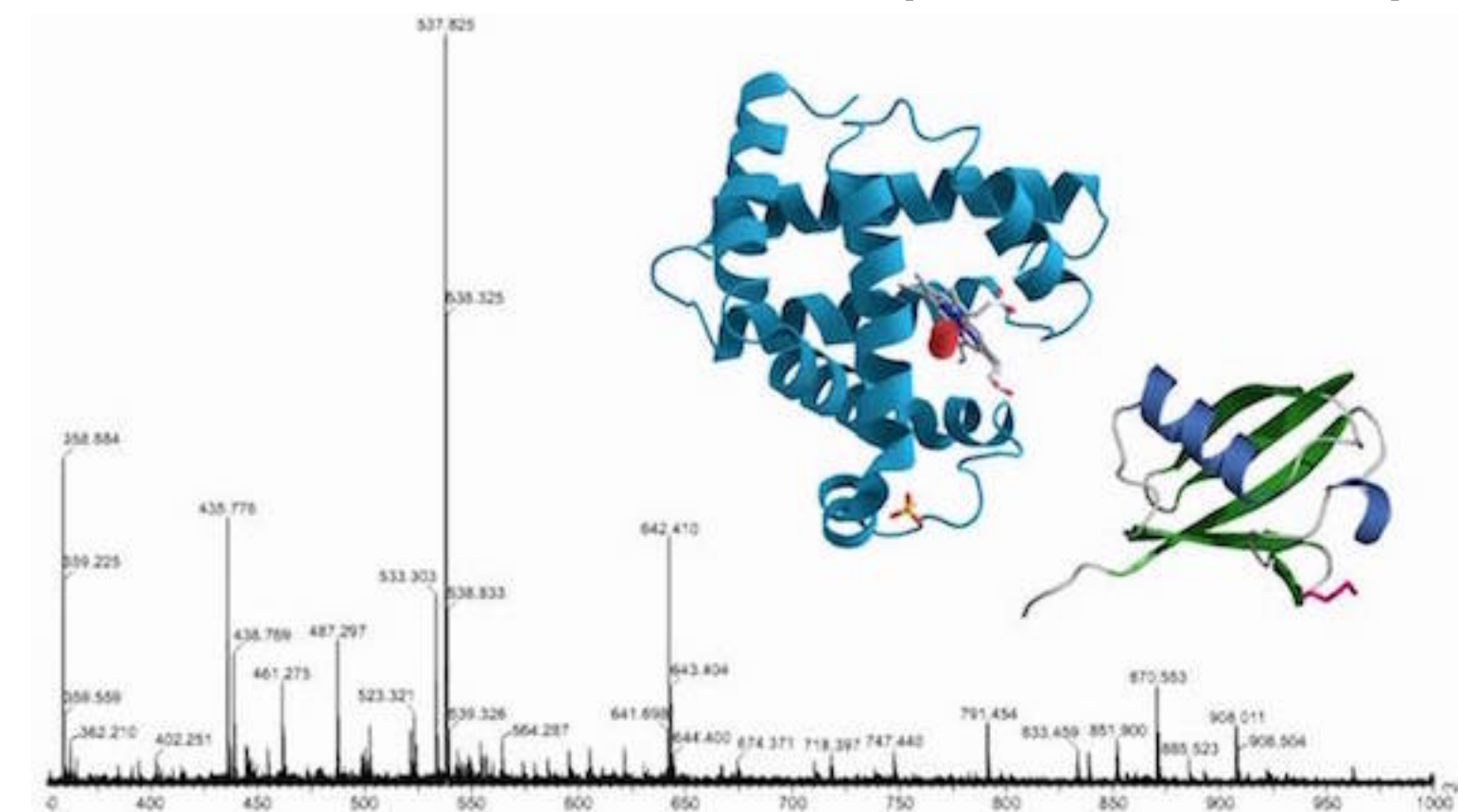


digital representation of images

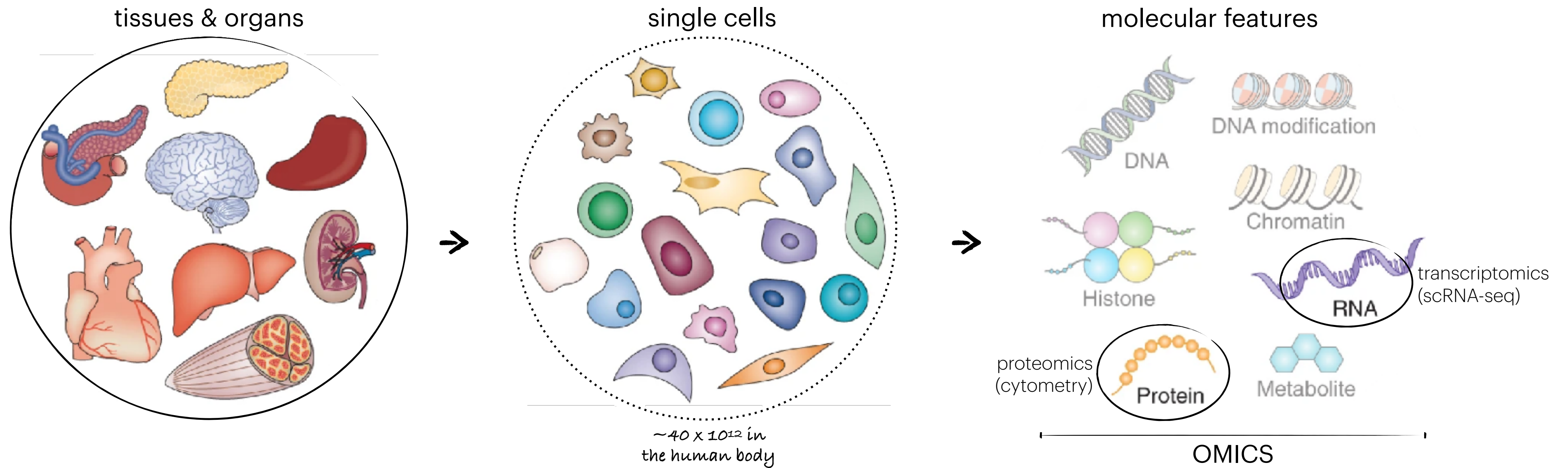
ion counts (mass cytometry)



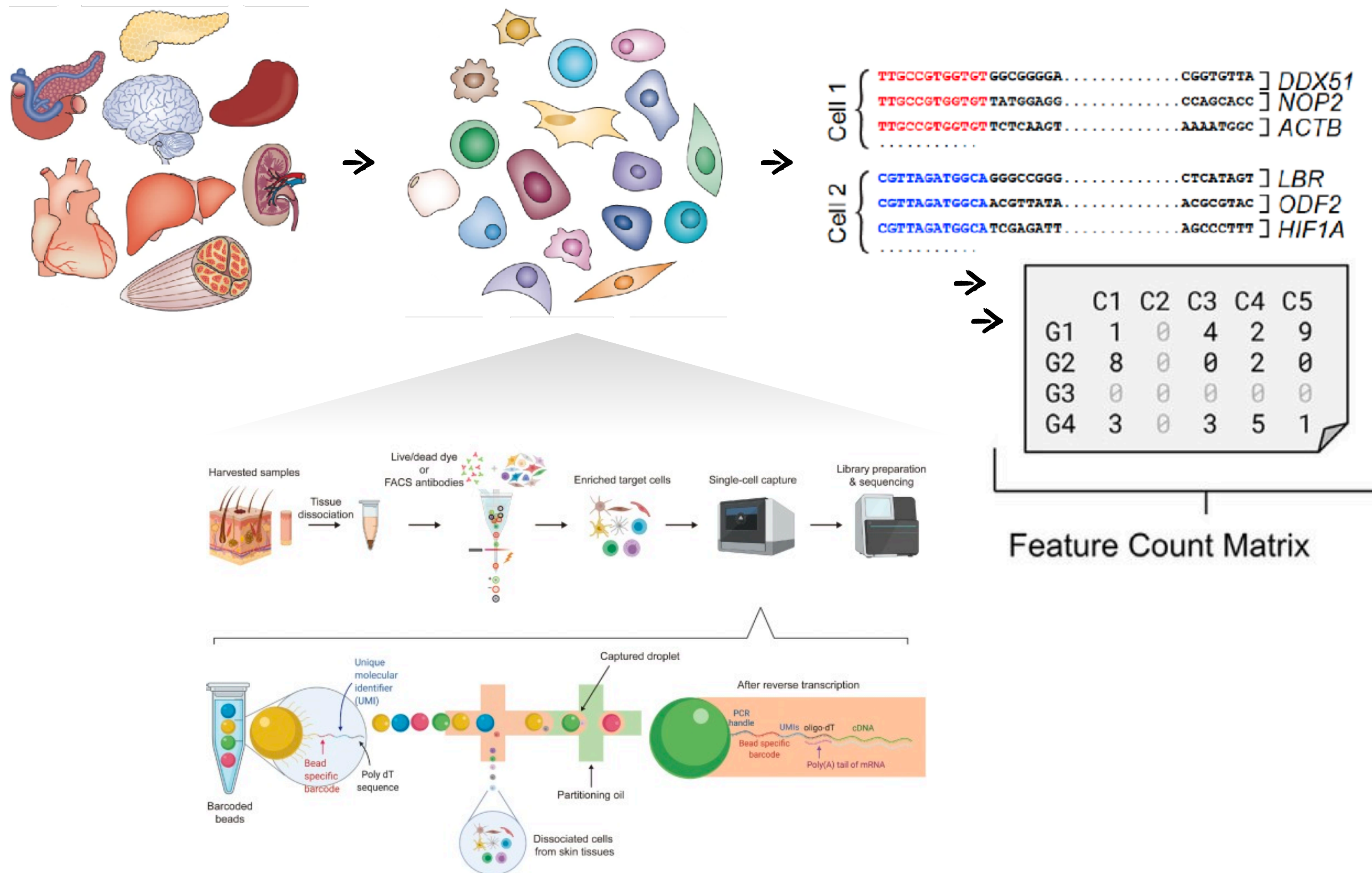
m/z spectra (mass spectrometry)



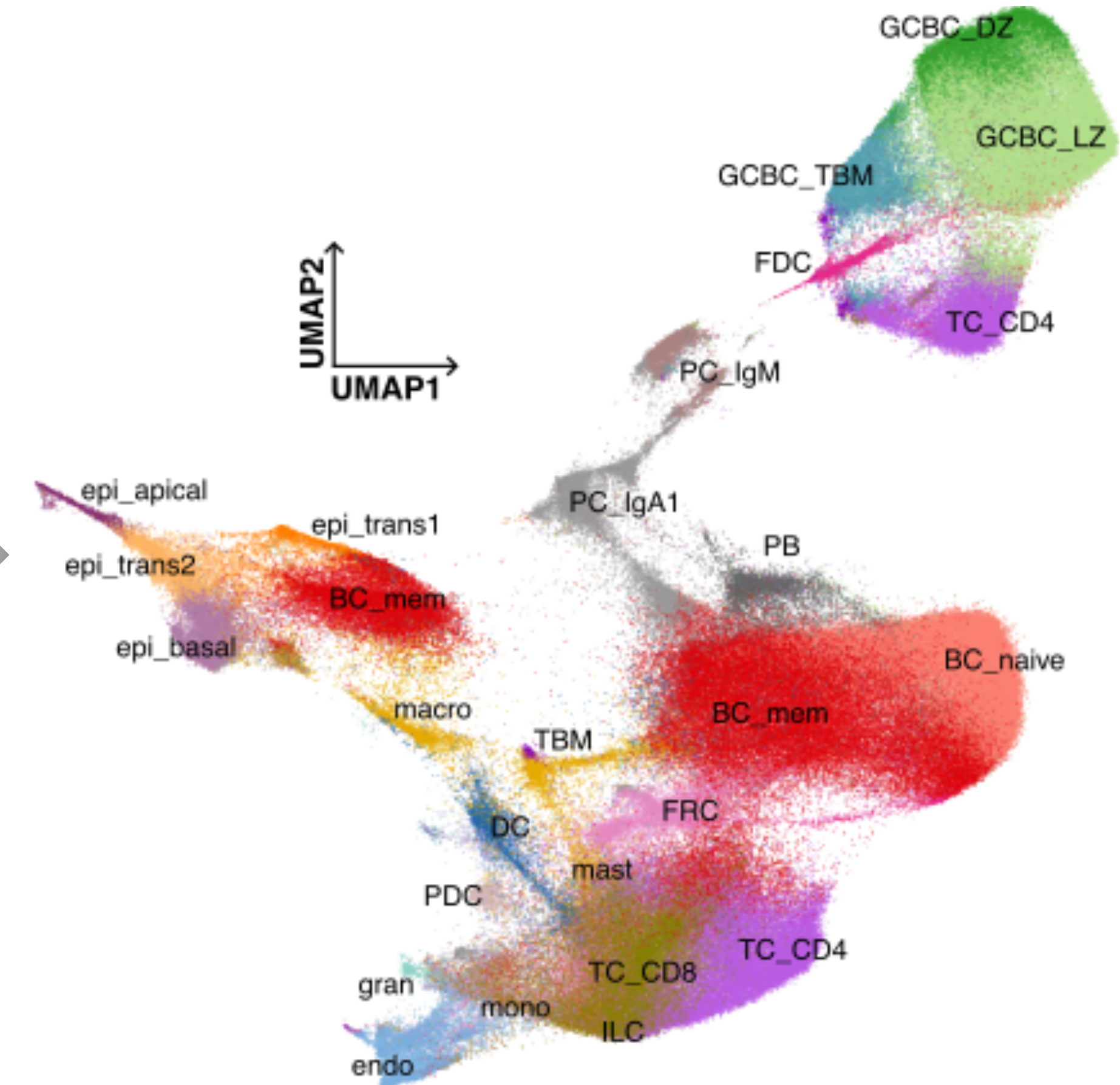
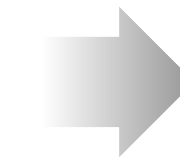
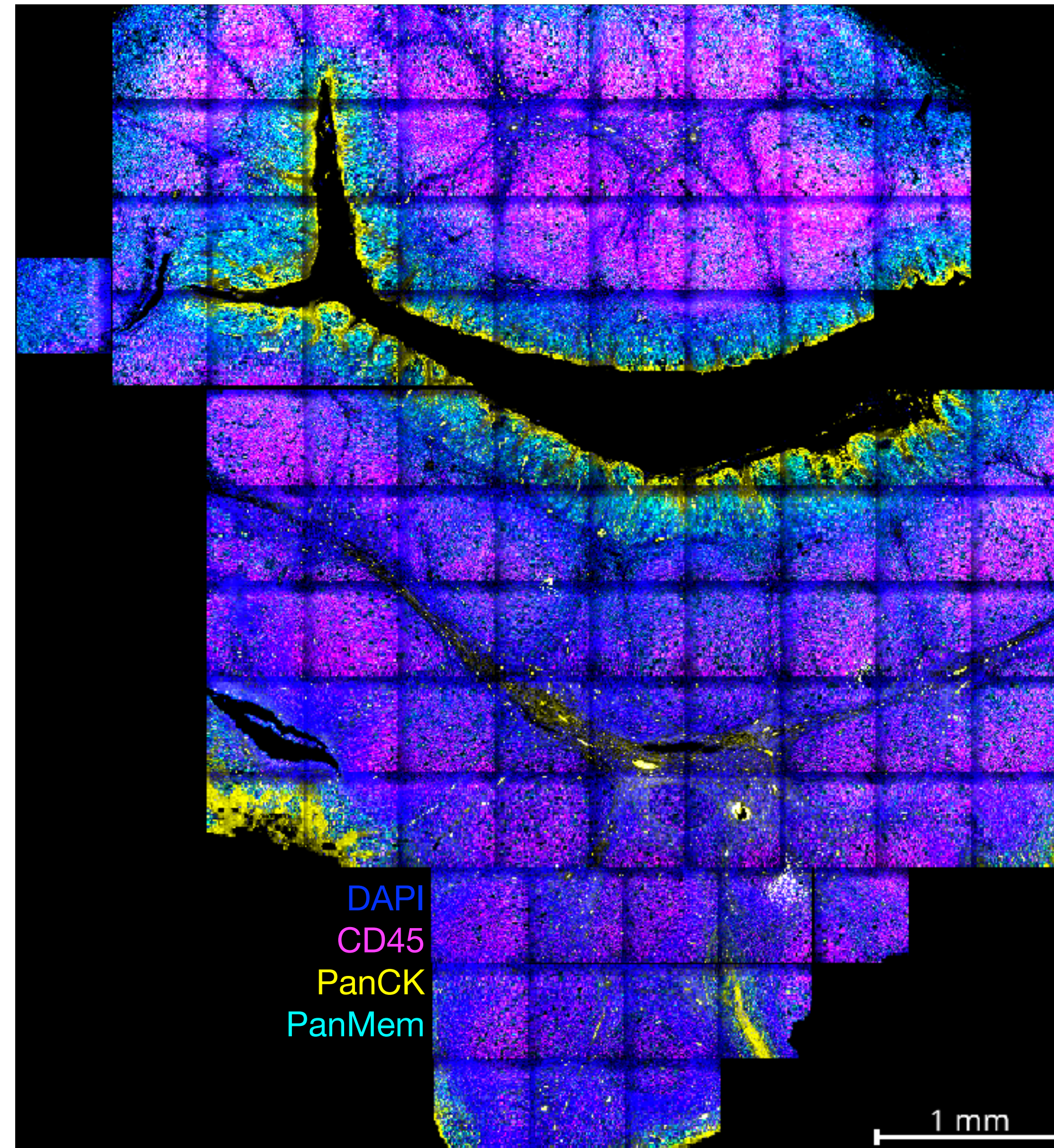
multi-cellular systems & the omics era



from bulk to single-cell omics

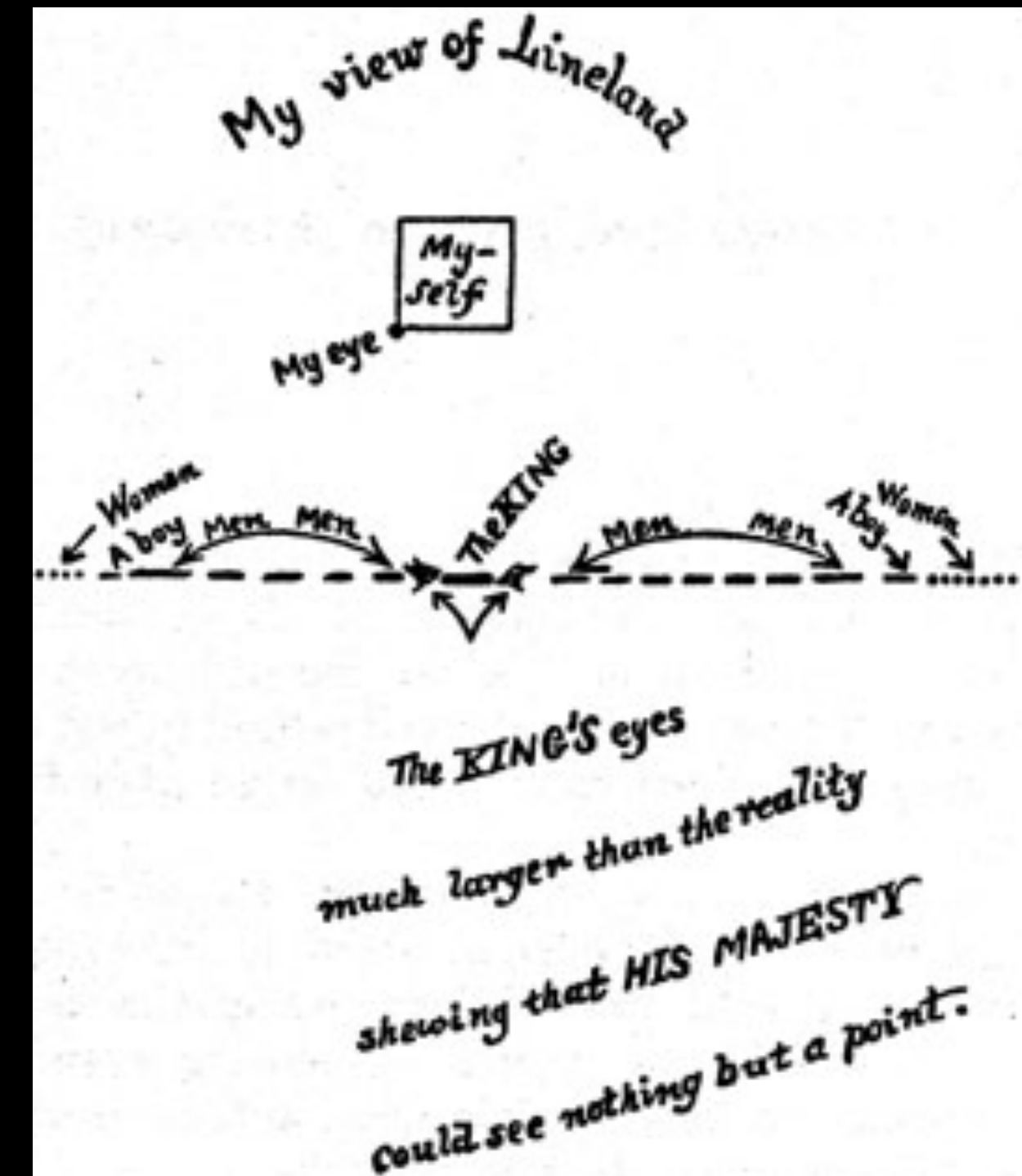


single-cell “maps”



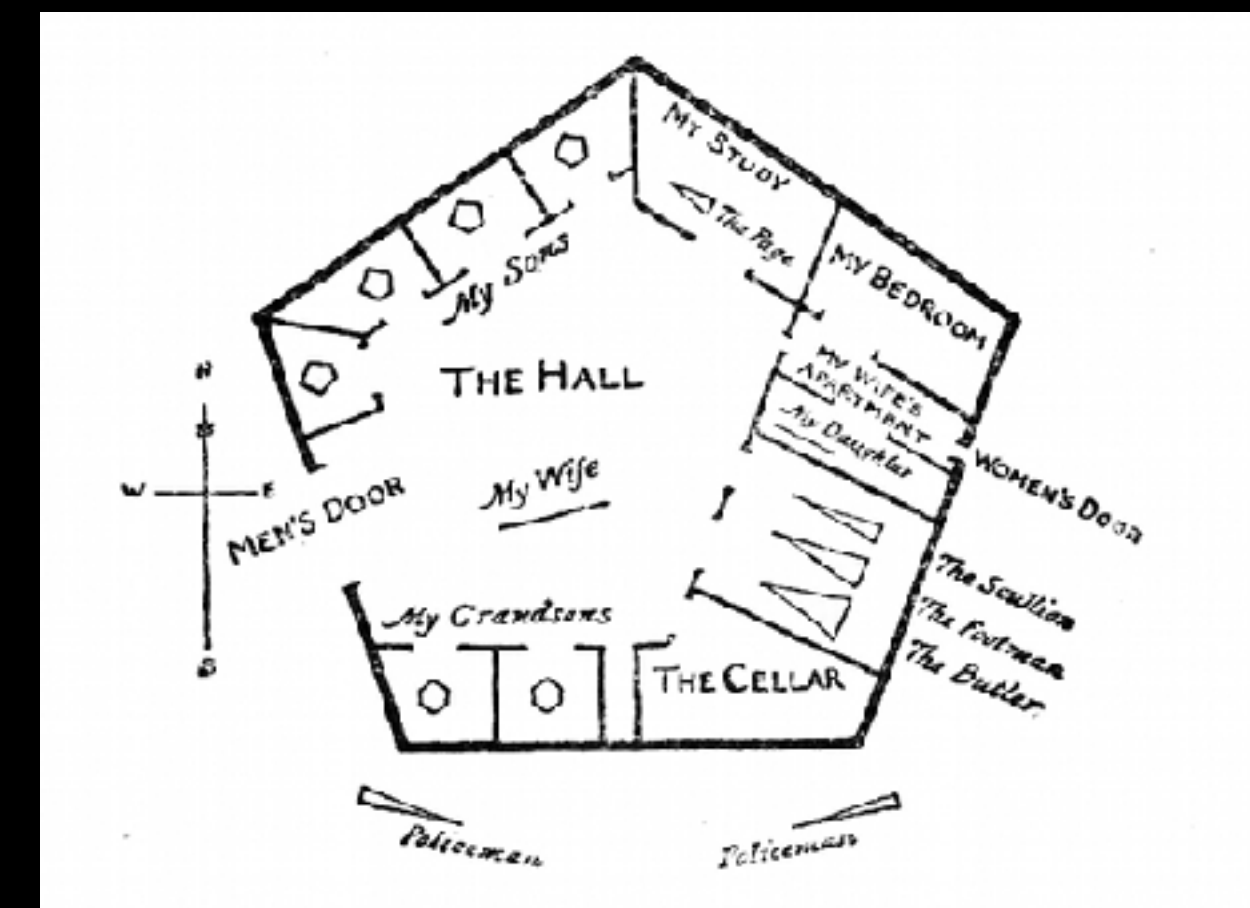
Lineland

[...] The small Lines were men and the Points Women, [...] all alike confined in motion and eye-sight to that single Straight Line, which was their World. It need scarcely be added that the whole of their horizon was limited to a Point; nor could any one ever see anything but a Point. [...] It followed that no Linelanders could ever pass another. Once neighbours, always neighbours.

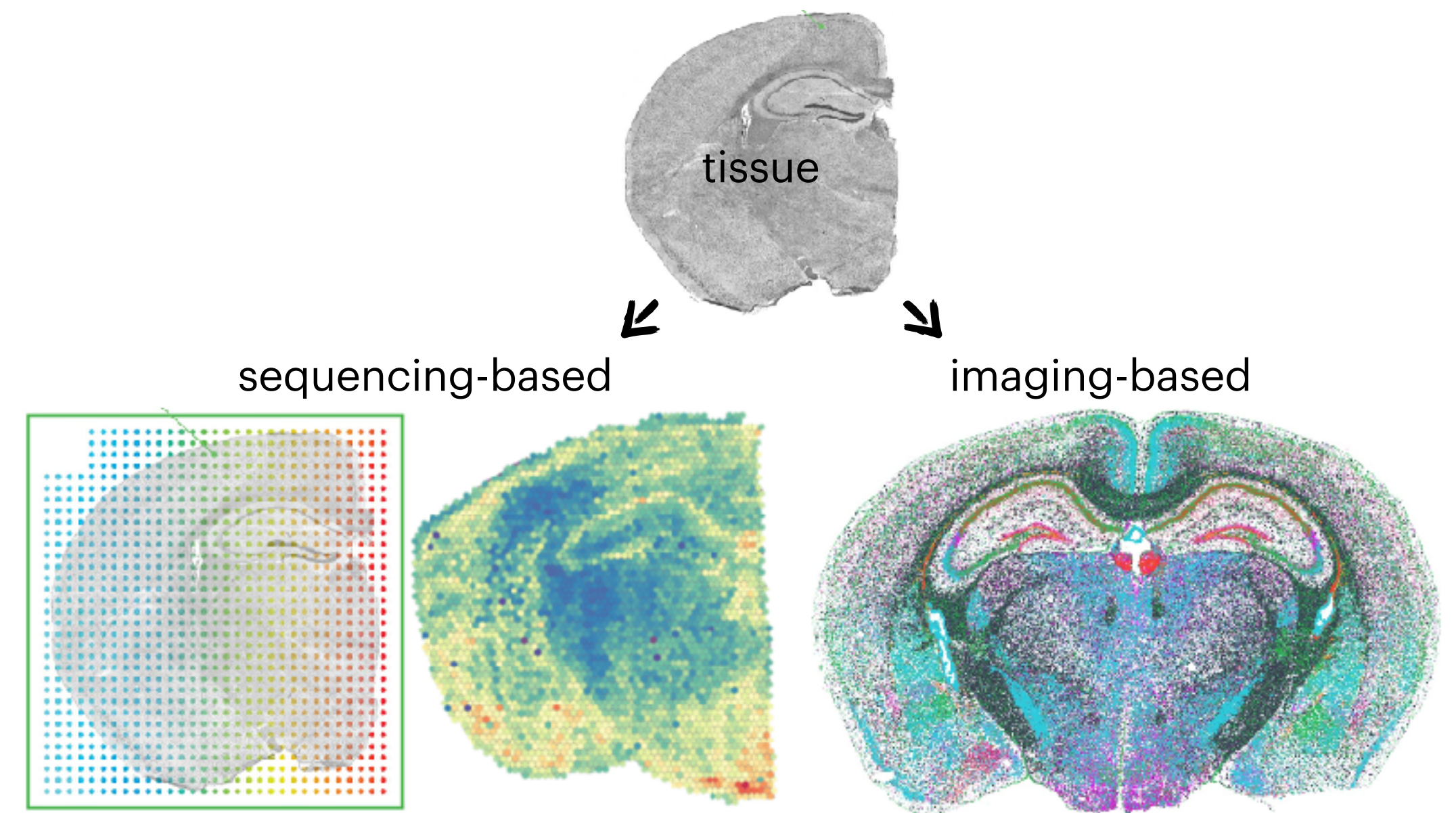


Flatland

Imagine a vast sheet of paper on which straight Lines, Triangles, Squares, Pentagons, Hexagons, and other figures, instead of remaining fixed in their places, move freely about, on or in the surface, but without the power of rising above or sinking below it [...].



from bulk to single-cell to spatial omics



...but if we can **proteomics**, why **transcriptomics**?



2'000

How many...
**protein-coding
genes** in human



20'000★

including
splice variants

5'000

including
**post-translational
modifications**

80'000

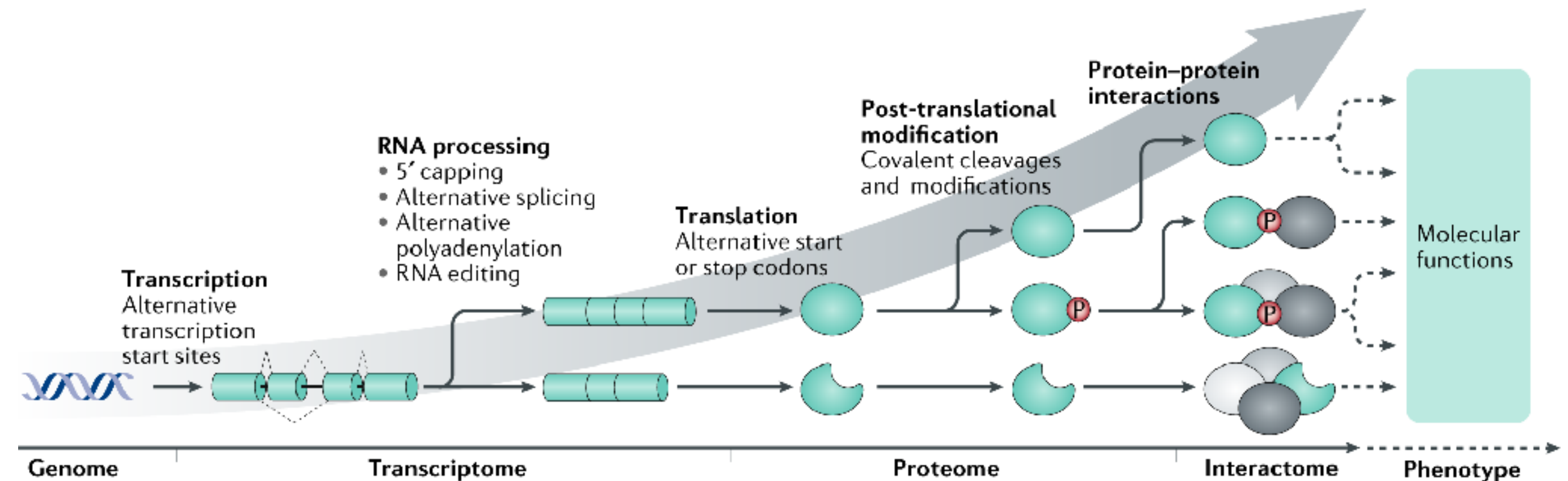
including
**somatic
recombination**

billions★

millions

The size of the human proteome is a matter of debate, and numbers in the literature range from as few as 20,000 to several million^{1,2}. The huge discrepancy between these numbers is not a scientific controversy, but more a matter of definition. Thanks to the human genome project, we can now estimate the number of protein-coding genes to be in the range of 19,587–20,245 (refs. 1,3,4). Thus, if a single representative protein from every gene is used as the definition of the proteome, the estimated size is just ~20,000. This number may decrease somewhat, as it has been difficult to find an expressed protein encoded by some of these putative protein-coding genes^{5,6}. However, if one considers that many genes

are transcribed with splice variants, the number of human proteins increases to ~70,000 (per Ensembl³). In addition, many human proteins undergo PTMs that can strongly influence their function or activity. These PTMs include glycosylation, phosphorylation and acetylation, among a few hundred others (Fig. 1a), giving rise to many hundreds of thousands of additional protein variants⁵; furthermore, though many proteins are unmodified, some fraction of proteins are already annotated with multiple modifications (Fig. 1b). Finally, selected genes for proteins like immunoglobulins and T-cell receptors undergo somatic recombination to increase the number of potential protein variants into the billions in certain cell types across one's lifetime^{7,8}.

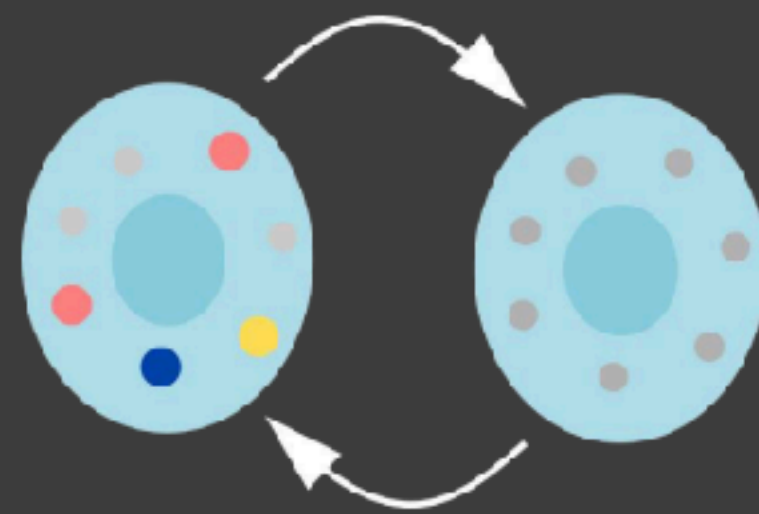


spatial RNA detection

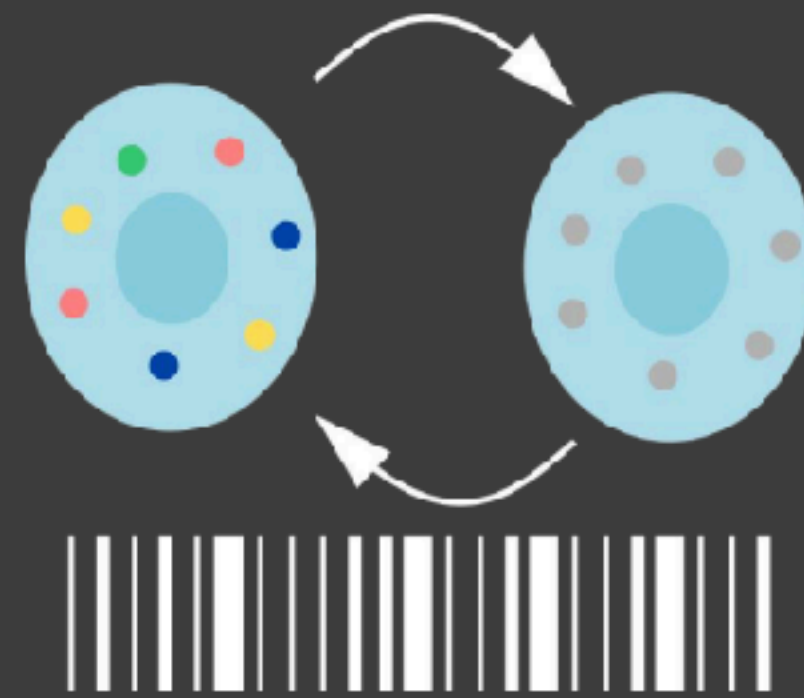
imaging/
microscopy

sequencing

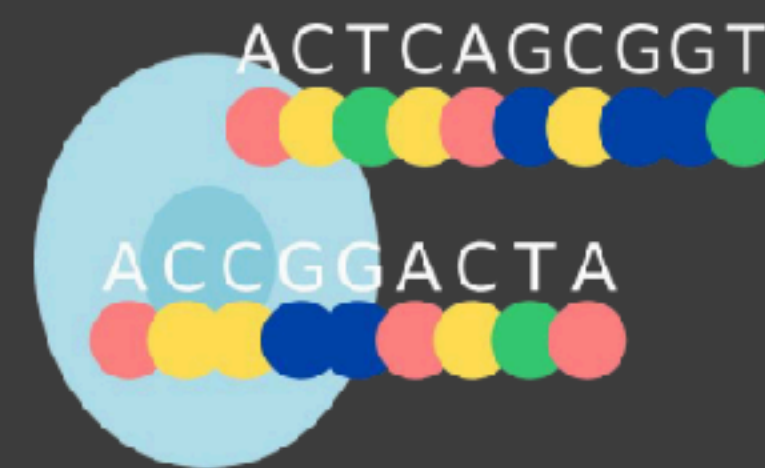
Cyclic FISH



Barcoded FISH



in situ Sequencing



Spatial Sequencing

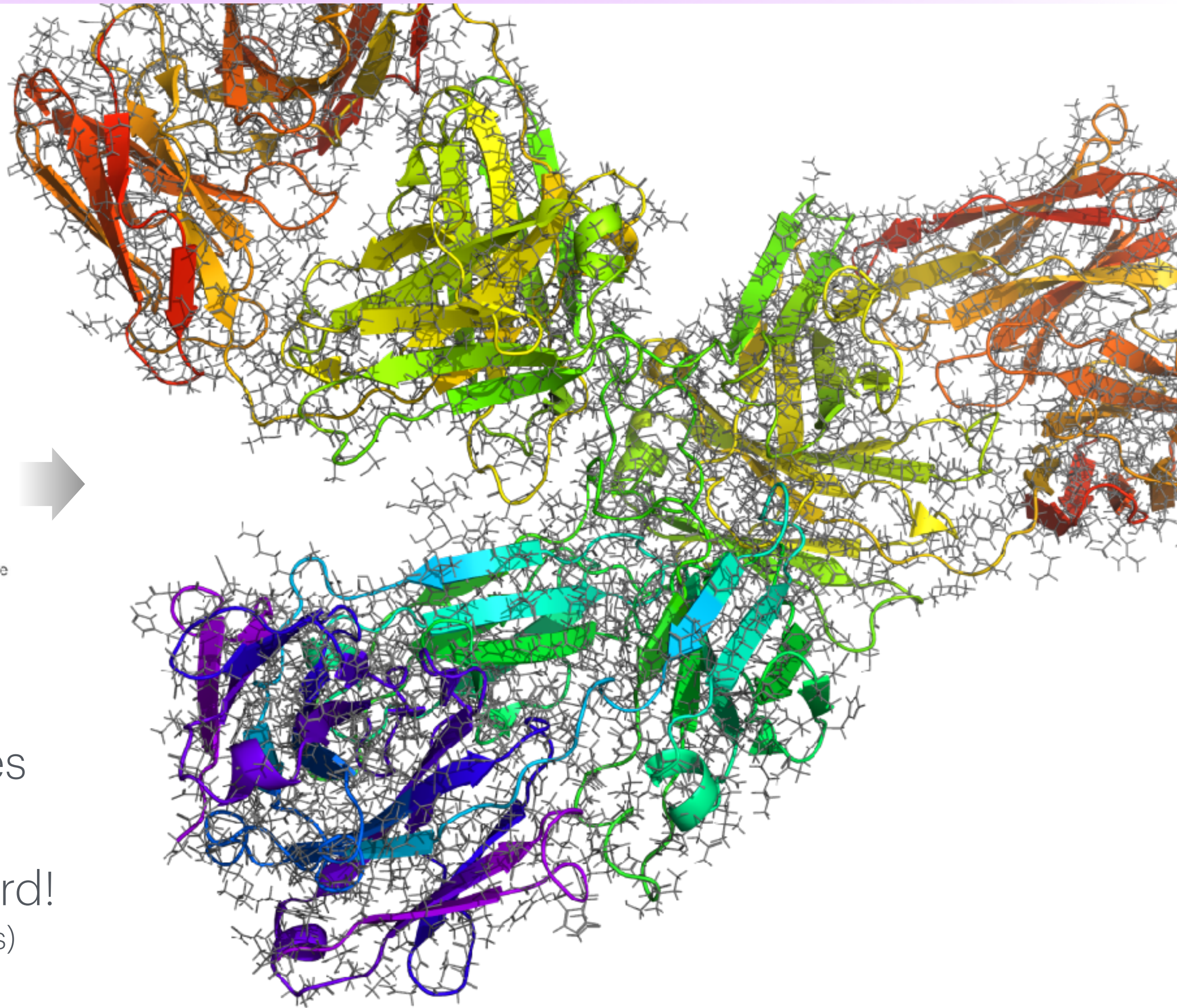
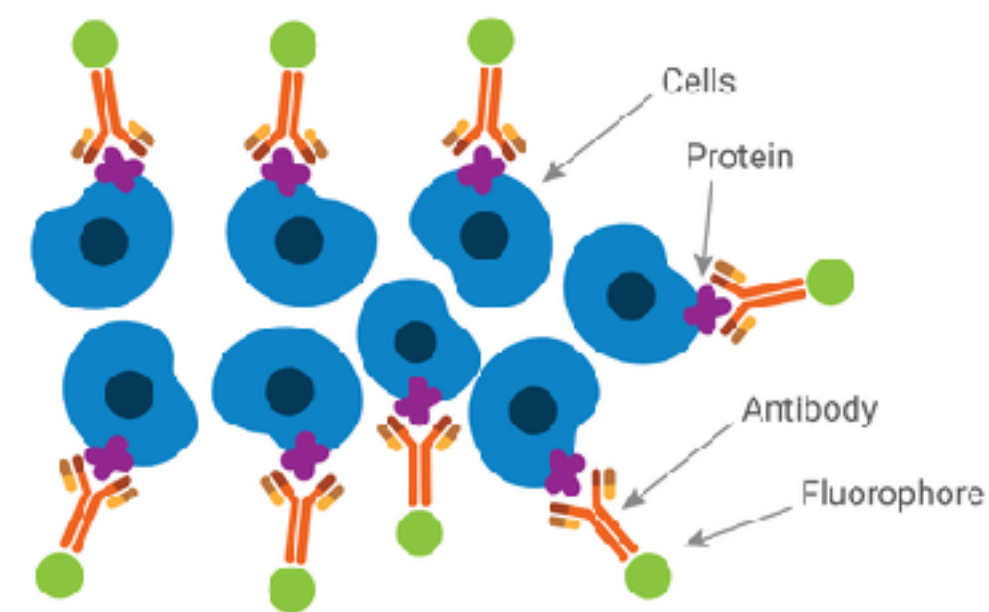
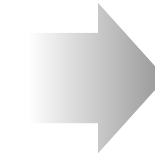
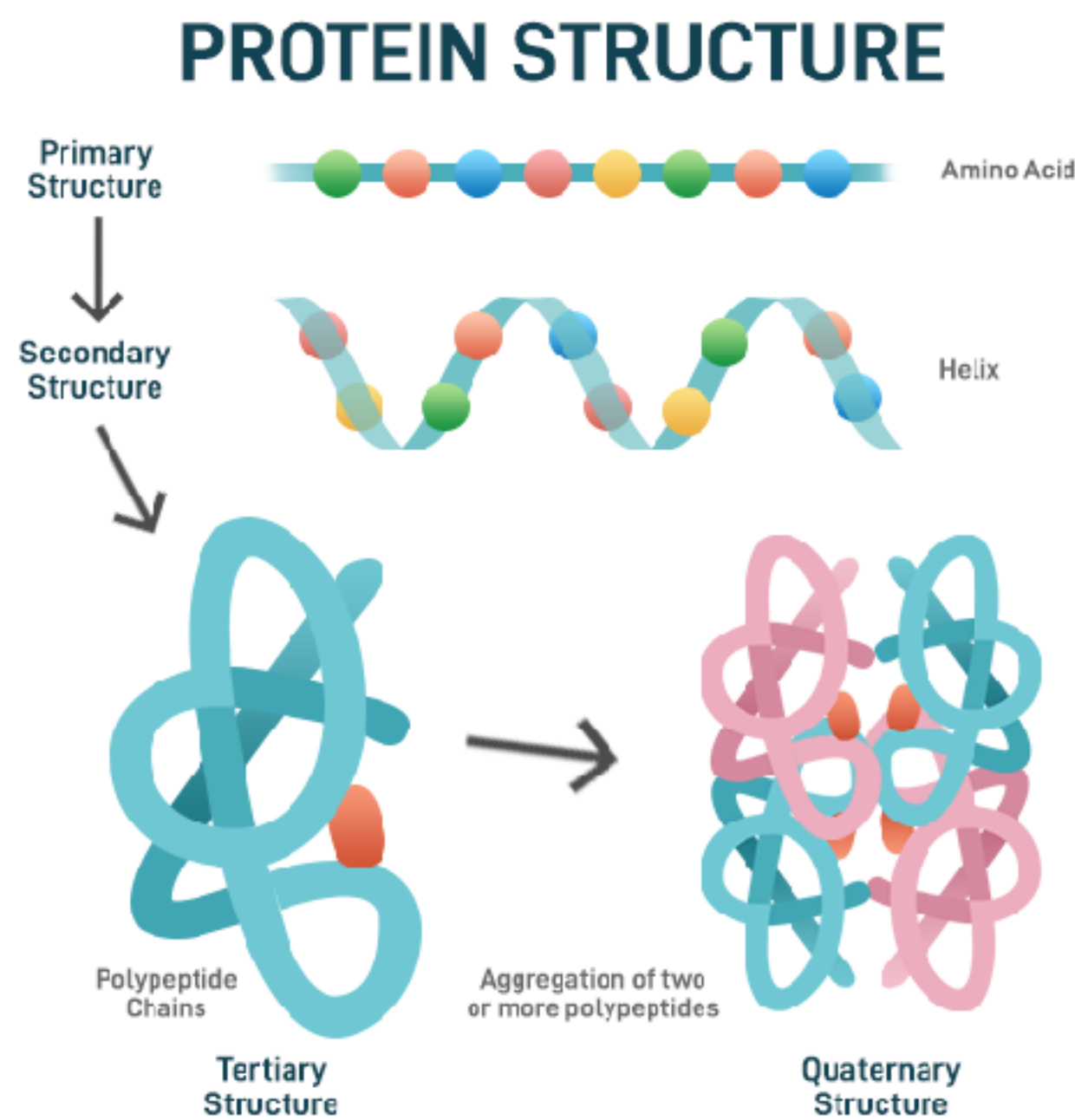


- HybISS
- Xenium
- CosMx
- EEL-FISH
- Mol. Cartography
- MERFISH/MERSCOPE

- ISS
- G4X
- STARmap
- StellarOomics

- ST
- Open-ST
- Slide-seq
- Stereo-seq
- Visium/HD
- Curio Seeker

RNA vs. protein detection

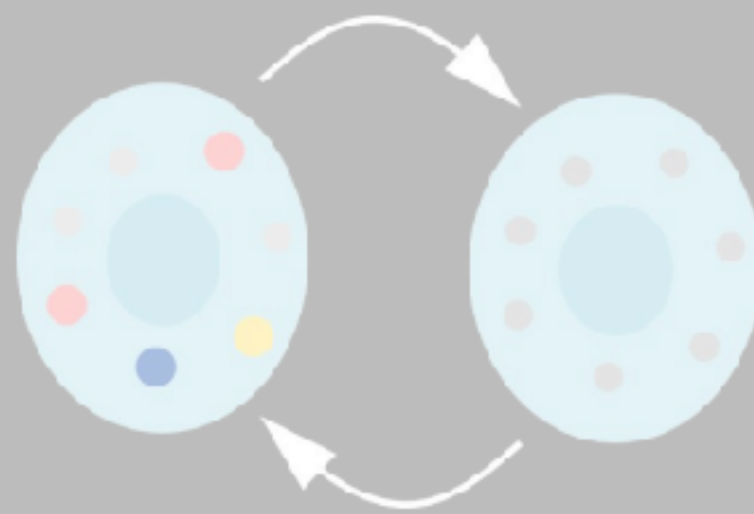


- we do not resolve amino acid sequences (MVYWTDRPAGSQHN...)
- antibody design is hard! (they may fail & specificity varies)

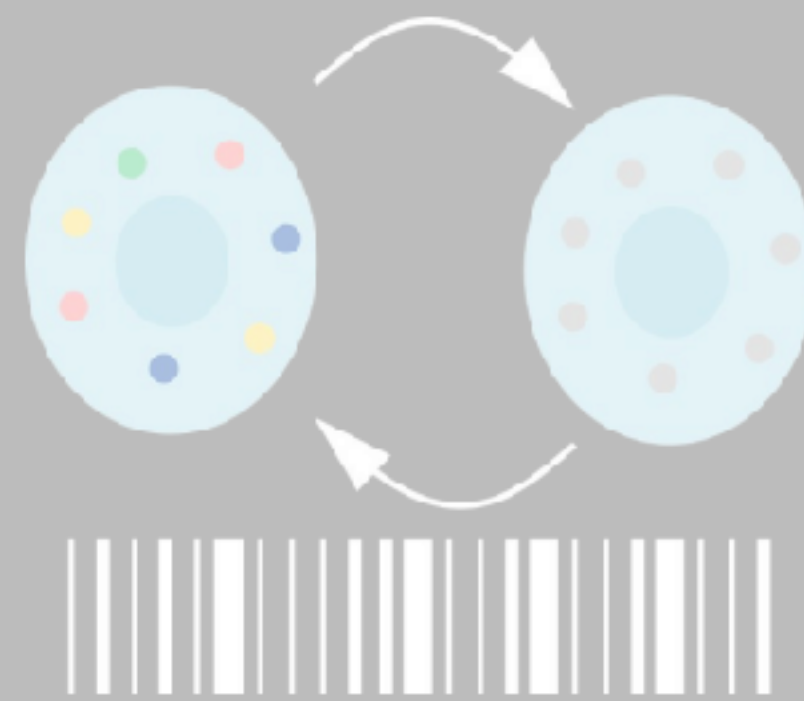
spatial RNA detection

imaging/ microscopy

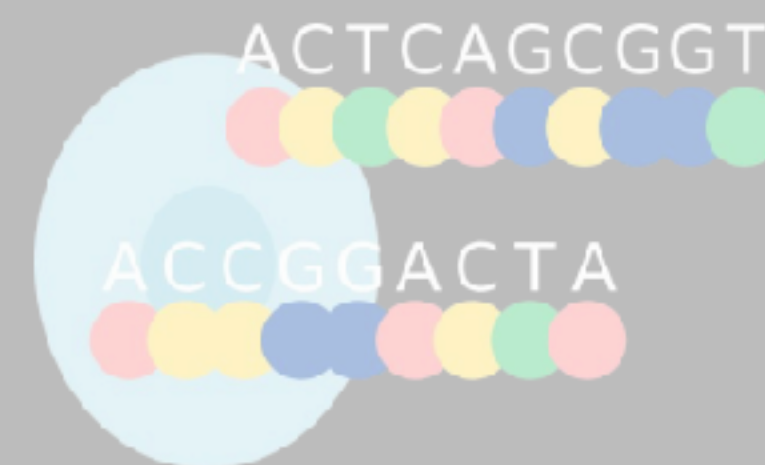
Cyclic FISH



Barcoded FISH



in situ Sequencing



- HybISS
- Xenium
- CosMx
- EEL-FISH
- Mol. Cartography
- MERFISH/MERSCOPE

- ISS
- G4X
- STARmap
- StellarOomics

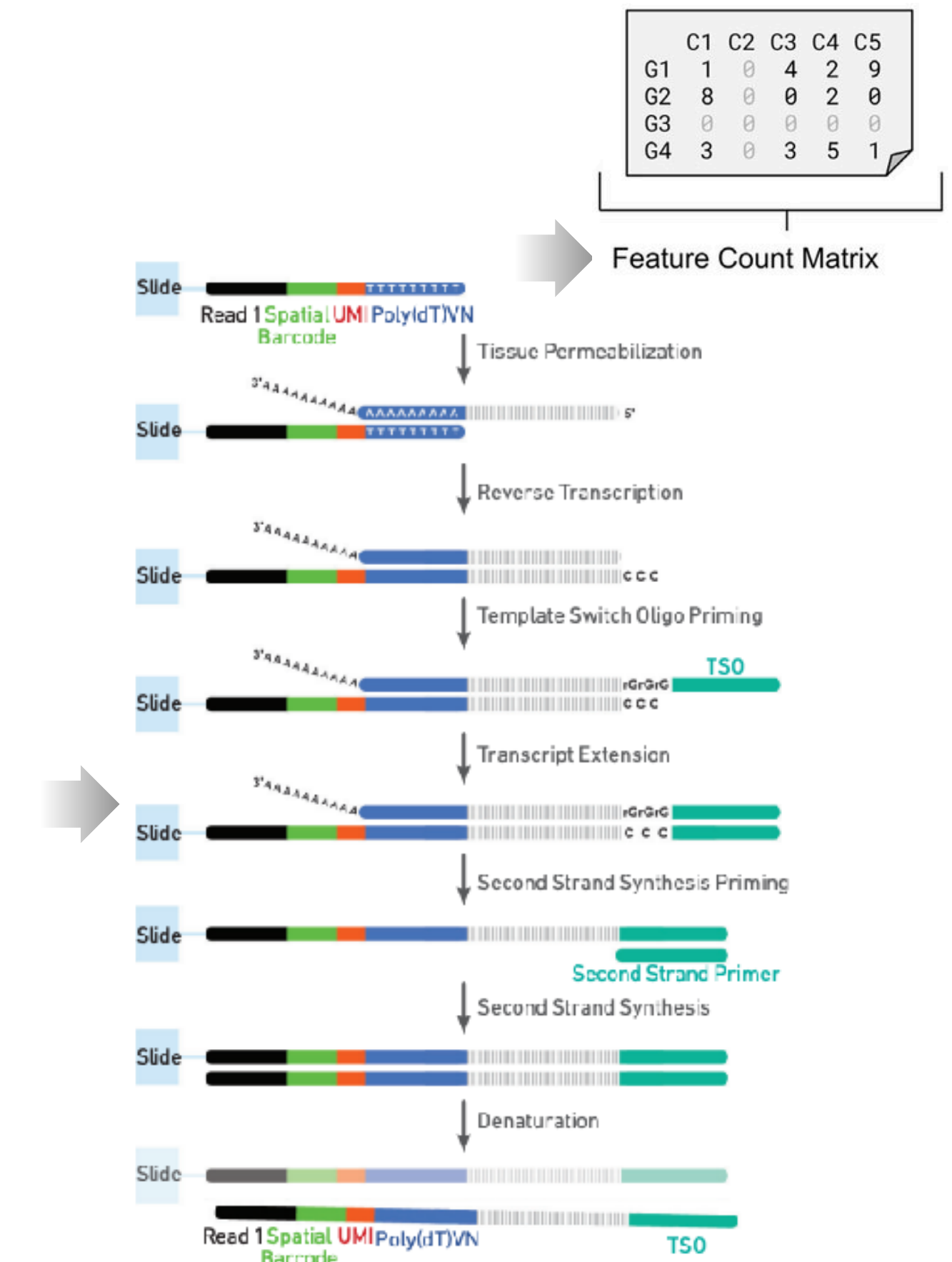
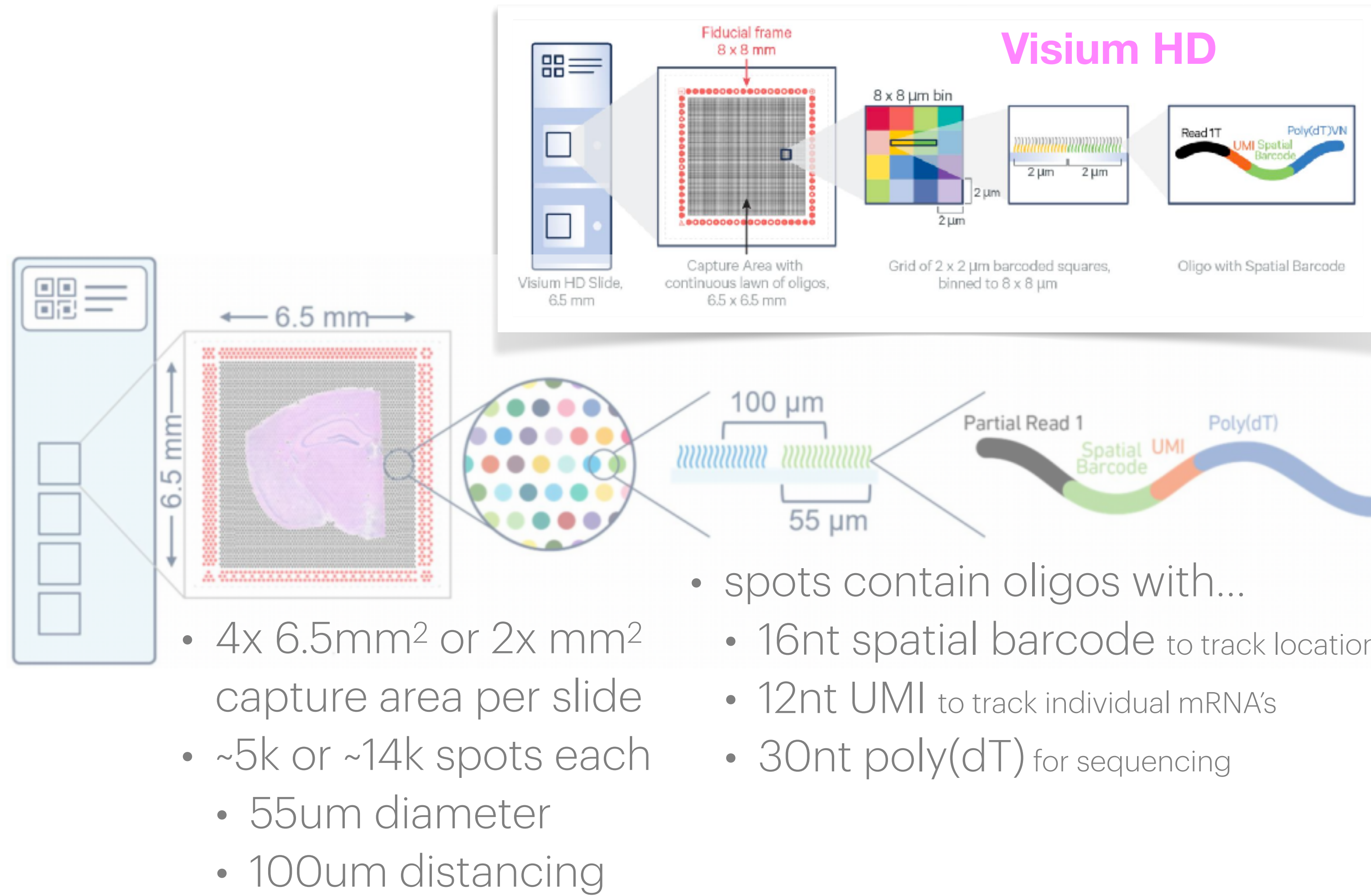
sequencing

Spatial Sequencing

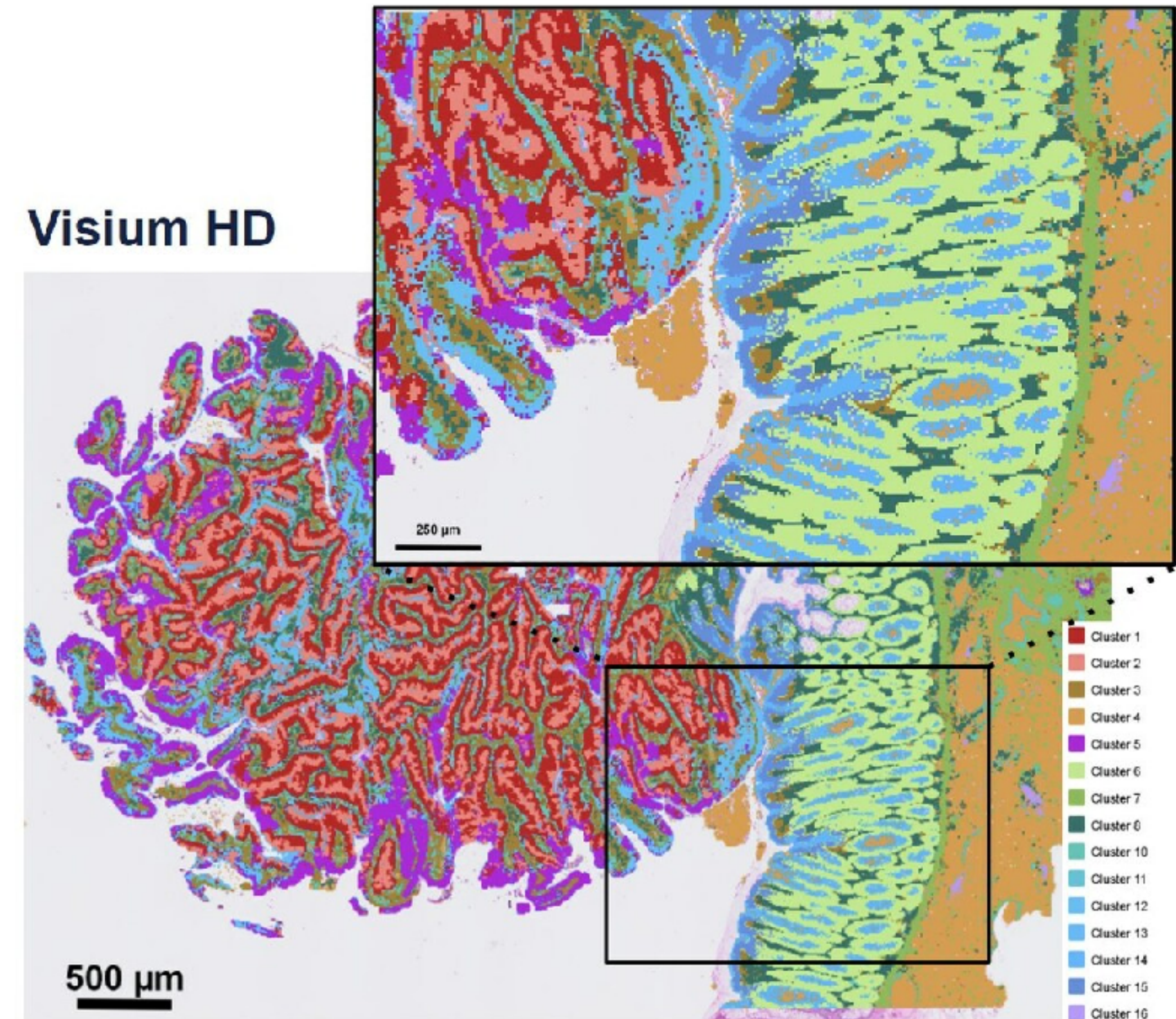
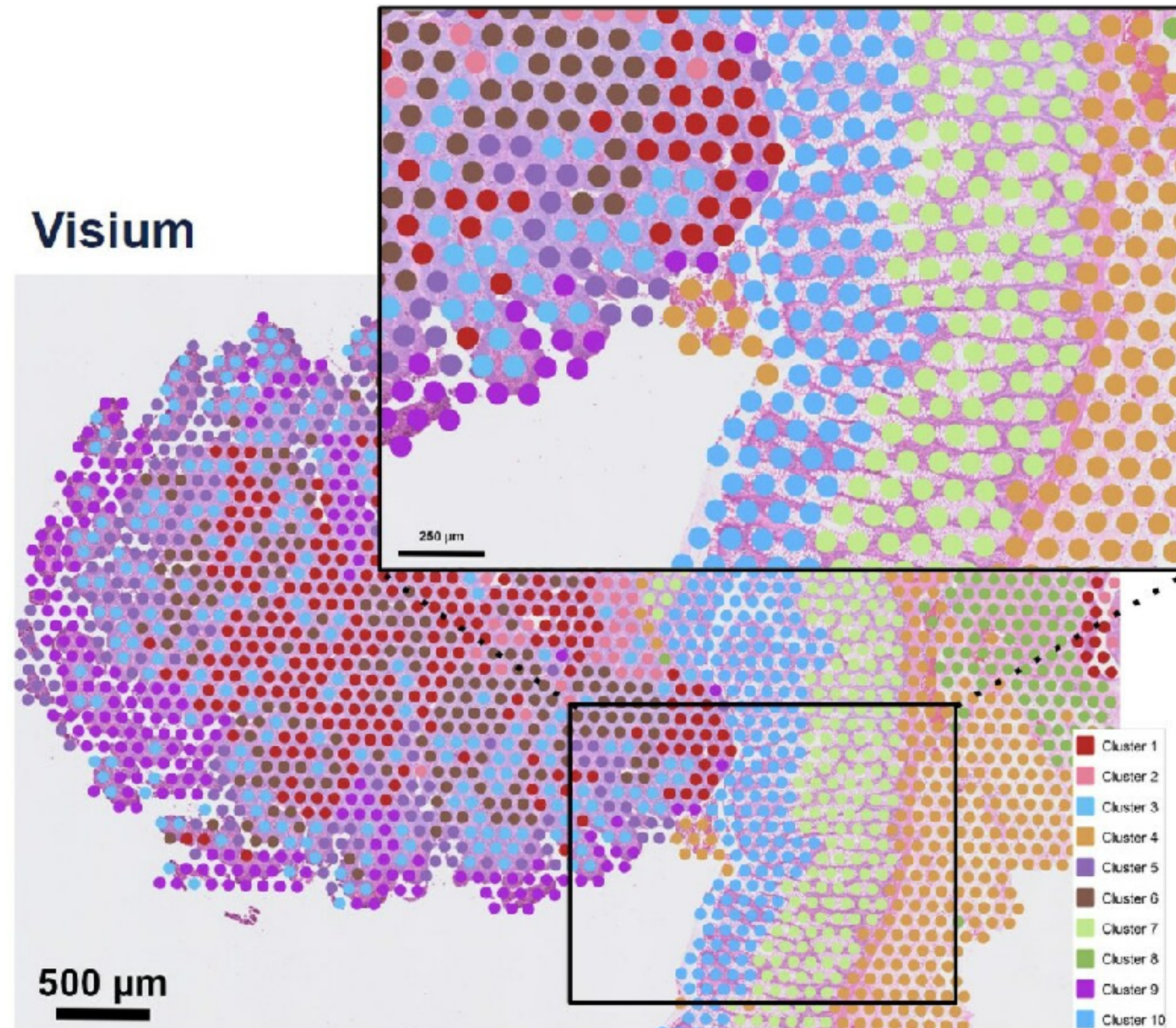


- ST
- Open-ST
- Slide-seq
- Stereo-seq
- Visium/HD
- Curio Seeker

10x Genomics Visium platform



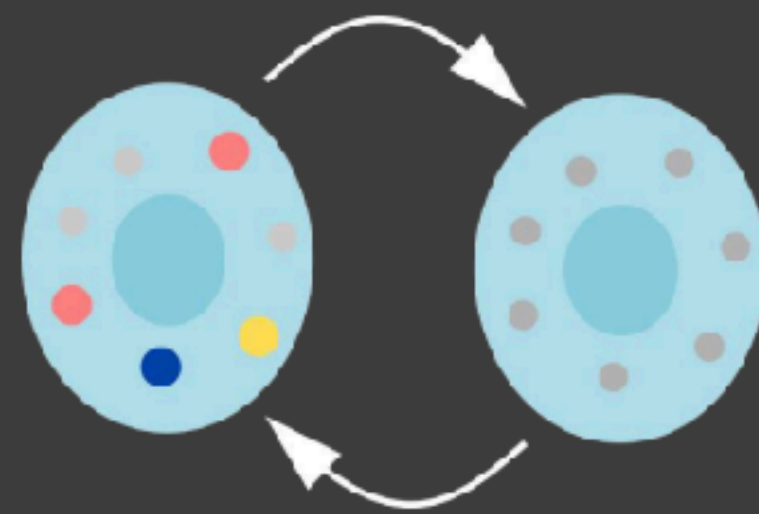
from super- to **sub-cellular** resolution



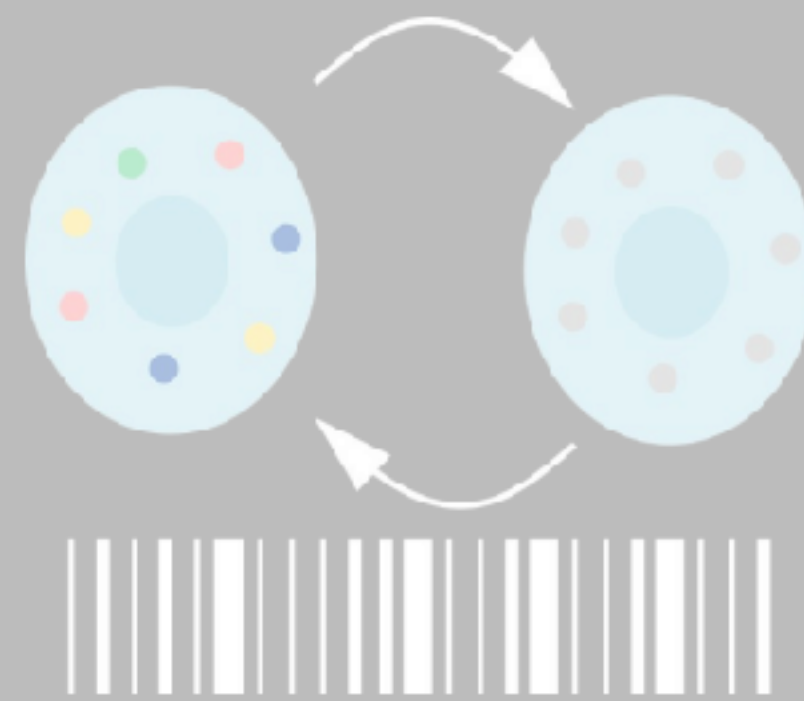
spatial RNA detection

imaging/ microscopy

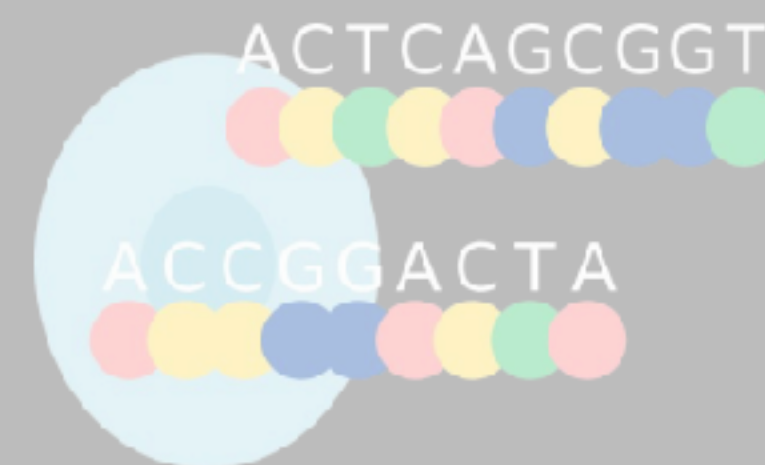
Cyclic FISH



Barcoded FISH



in situ Sequencing



sequencing

Spatial Sequencing



- HybISS
- Xenium
- CosMx
- EEL-FISH
- Mol. Cartography
- MERFISH/MERSCOPE

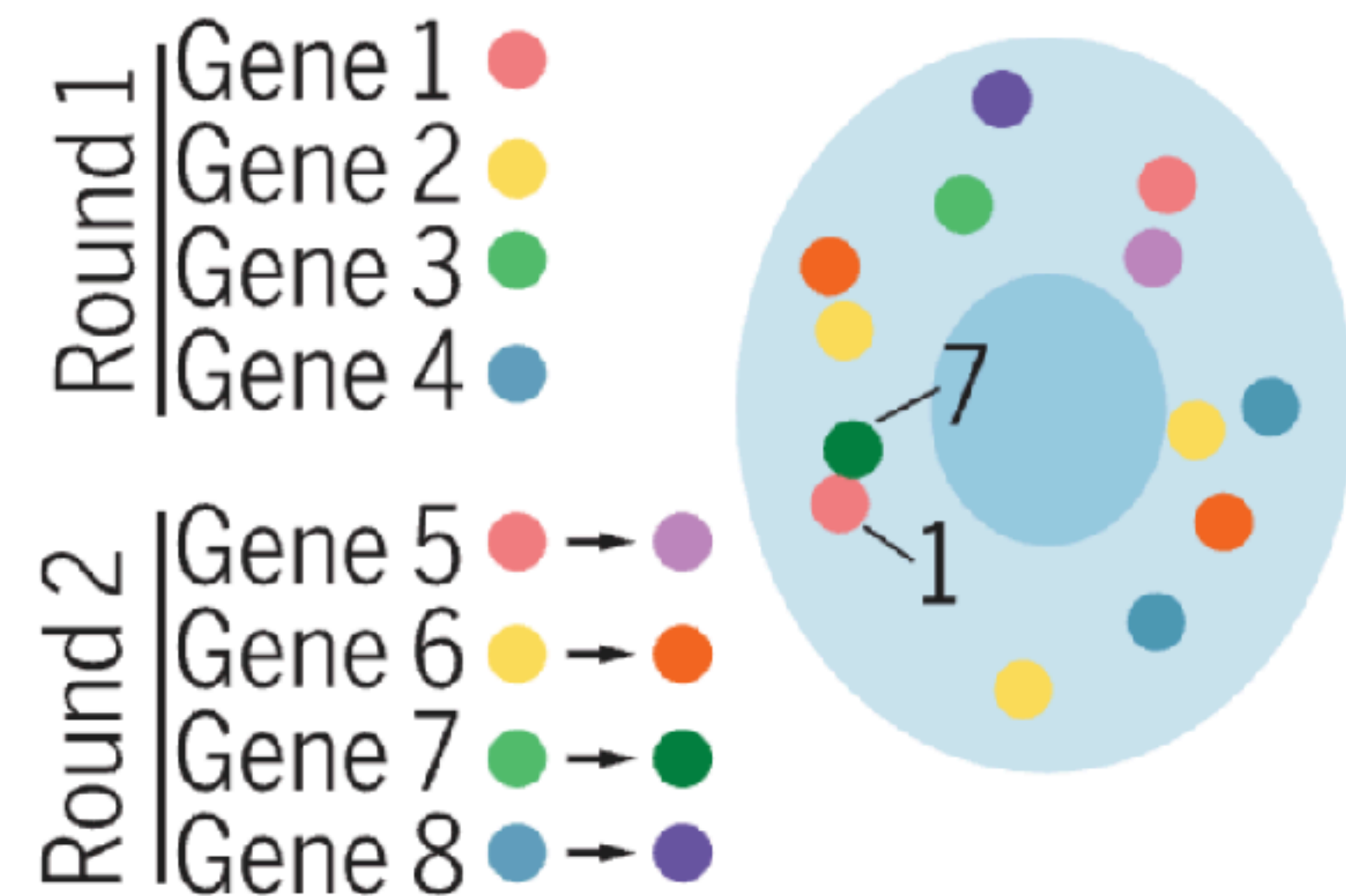
- ISS
- G4X
- STARmap
- StellarOomics

- ST
- Open-ST
- Slide-seq
- Stereo-seq
- Visium/HD
- Curio Seeker

cyclic FISH (= fluorescence in situ hybridization)



Raj et al. (Nature Methods, 2008)



scaling:

#rounds x
#channels

scalability:

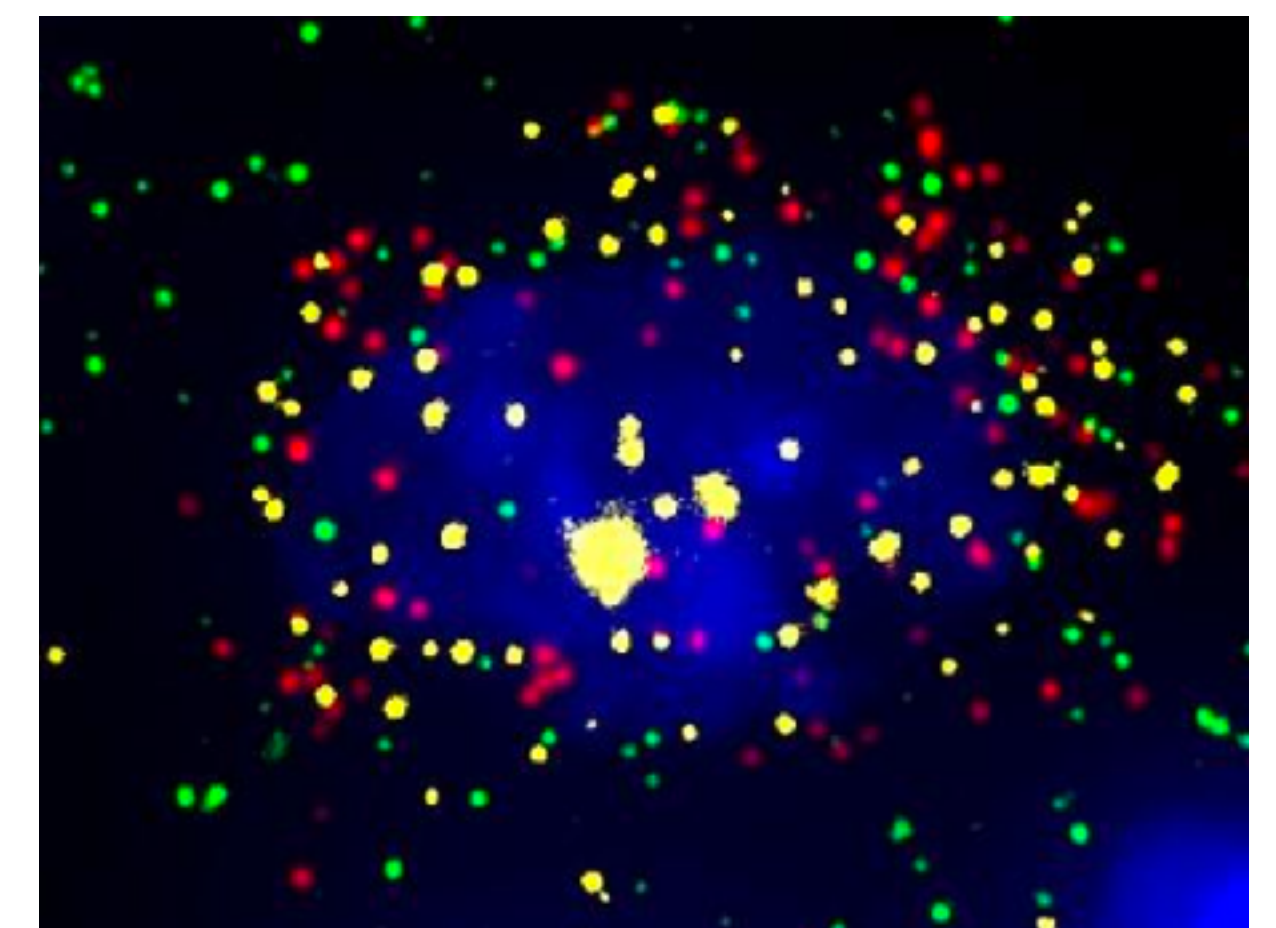
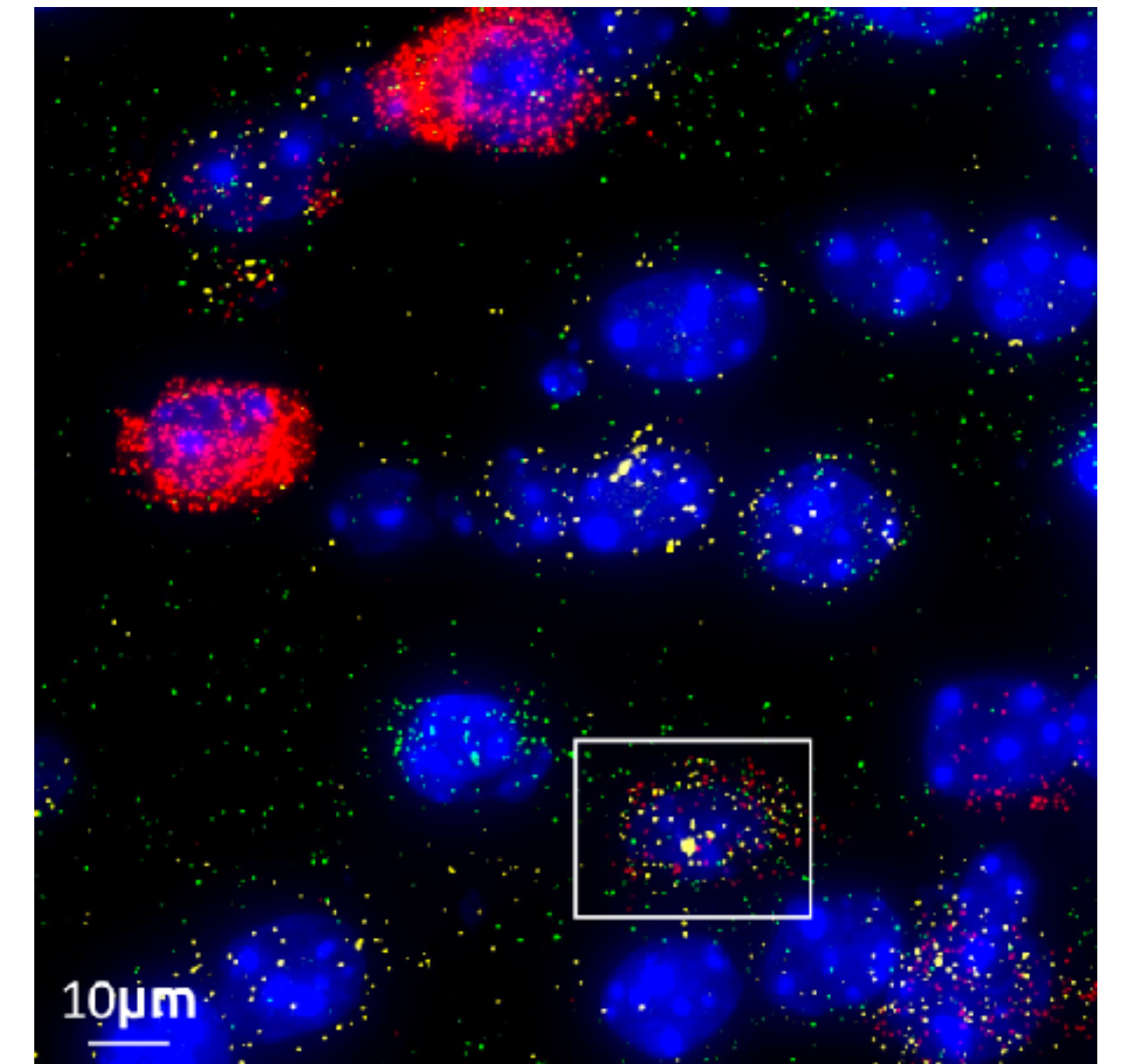
~100 targets

limitation:

fluorophore
availability,
sample
deterioration



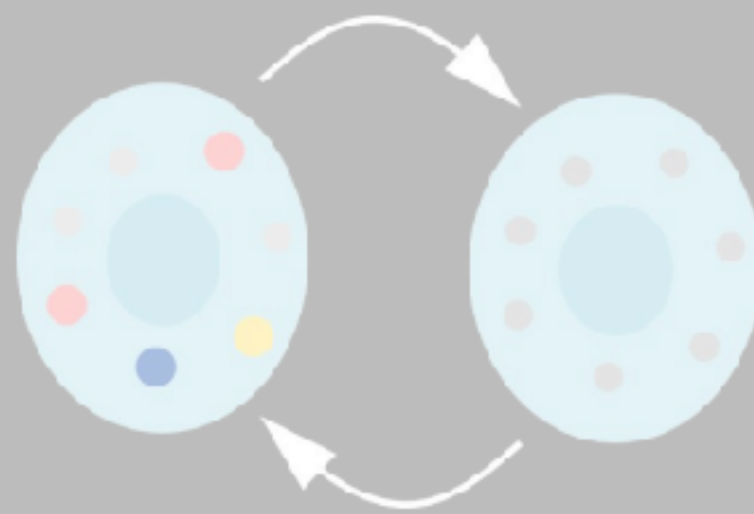
gene 1 gene 2 gene 3 DNA



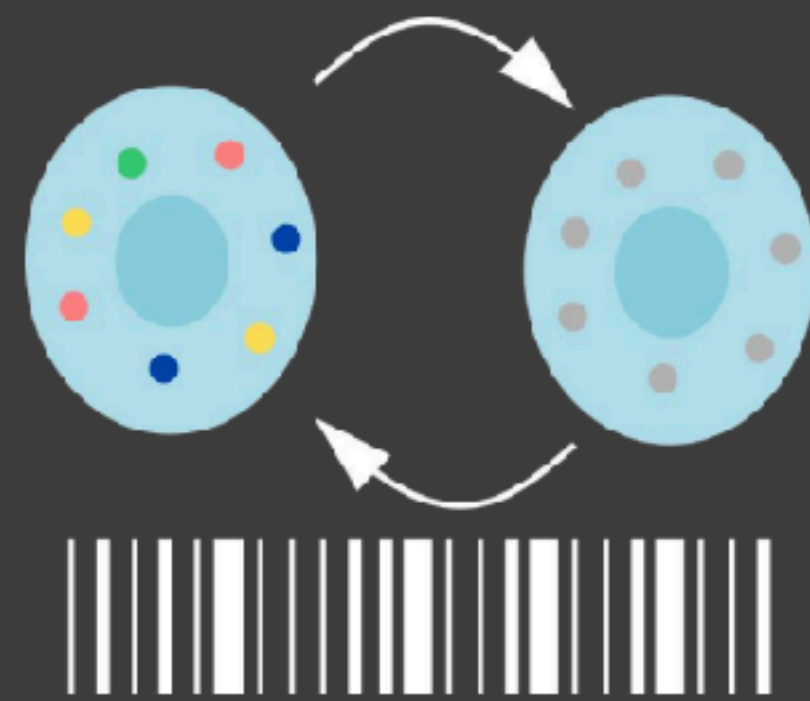
spatial RNA detection

imaging/ microscopy

Cyclic FISH

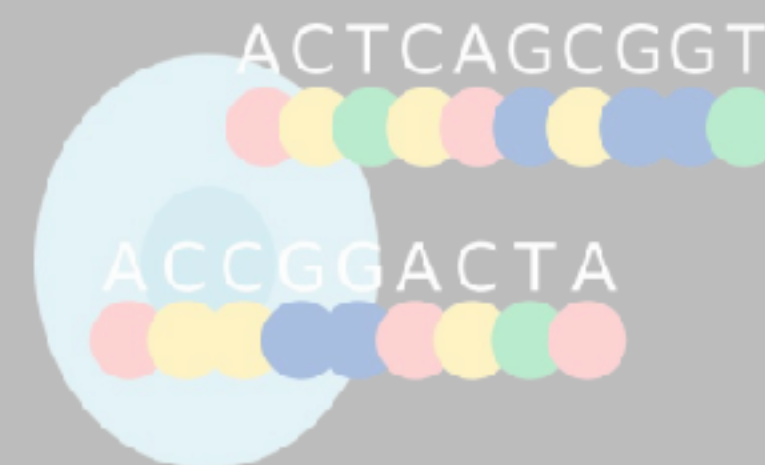


Barcoded FISH



- HybISS
- Xenium
- CosMx
- EEL-FISH
- Mol. Cartography
- MERFISH/MERSCOPE

in situ Sequencing



- ISS
- G4X
- STARmap
- StellarOomics

sequencing

Spatial Sequencing



- ST
- Open-ST
- Slide-seq
- Stereo-seq
- Visium/HD
- Curio Seeker

from cyclic to **barcoded** FISH

scaling:

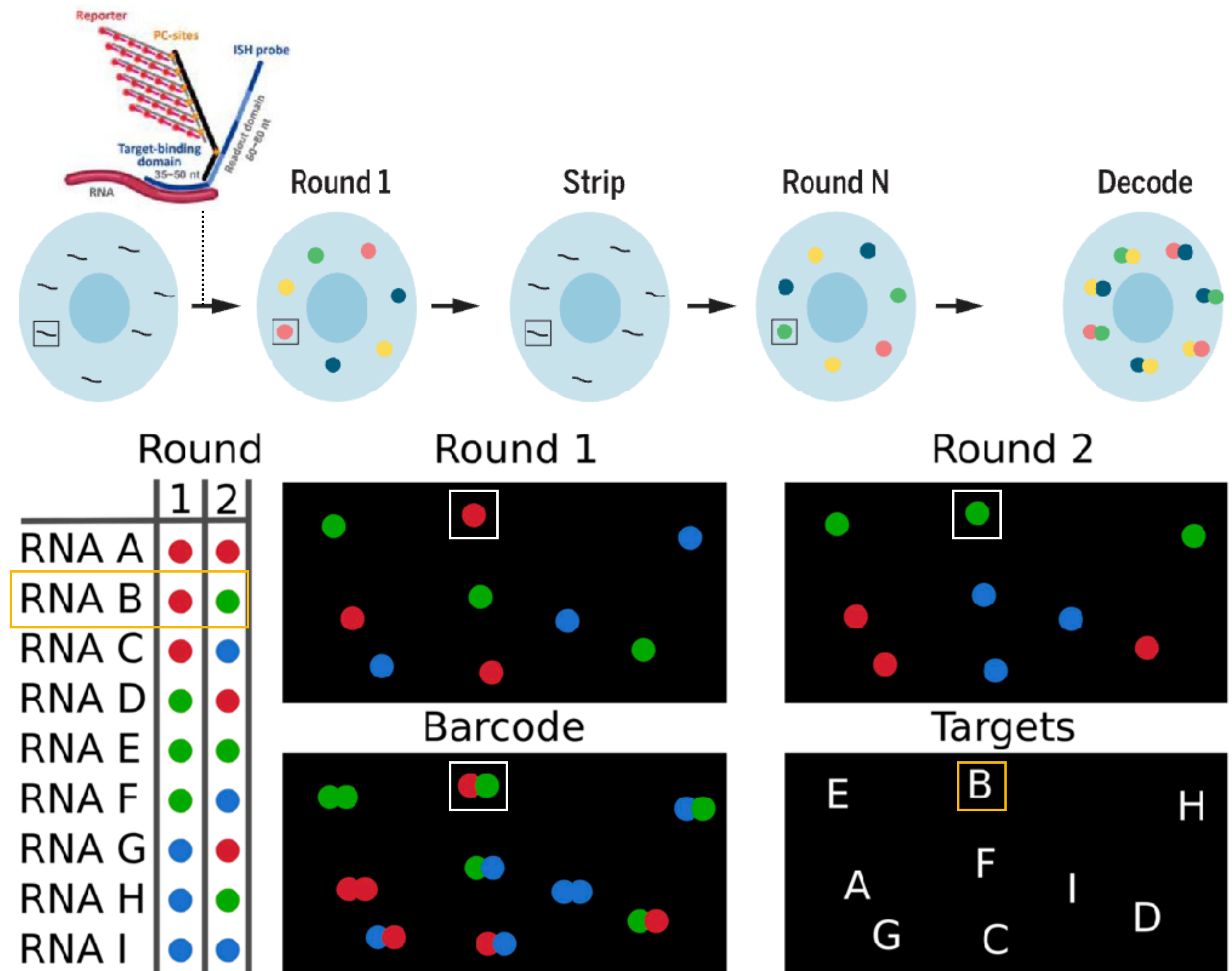
#channels $\wedge$ #rounds
($4^7 = 16'384$, $4^8 = 65'536$)

scalability:

1000s of targets

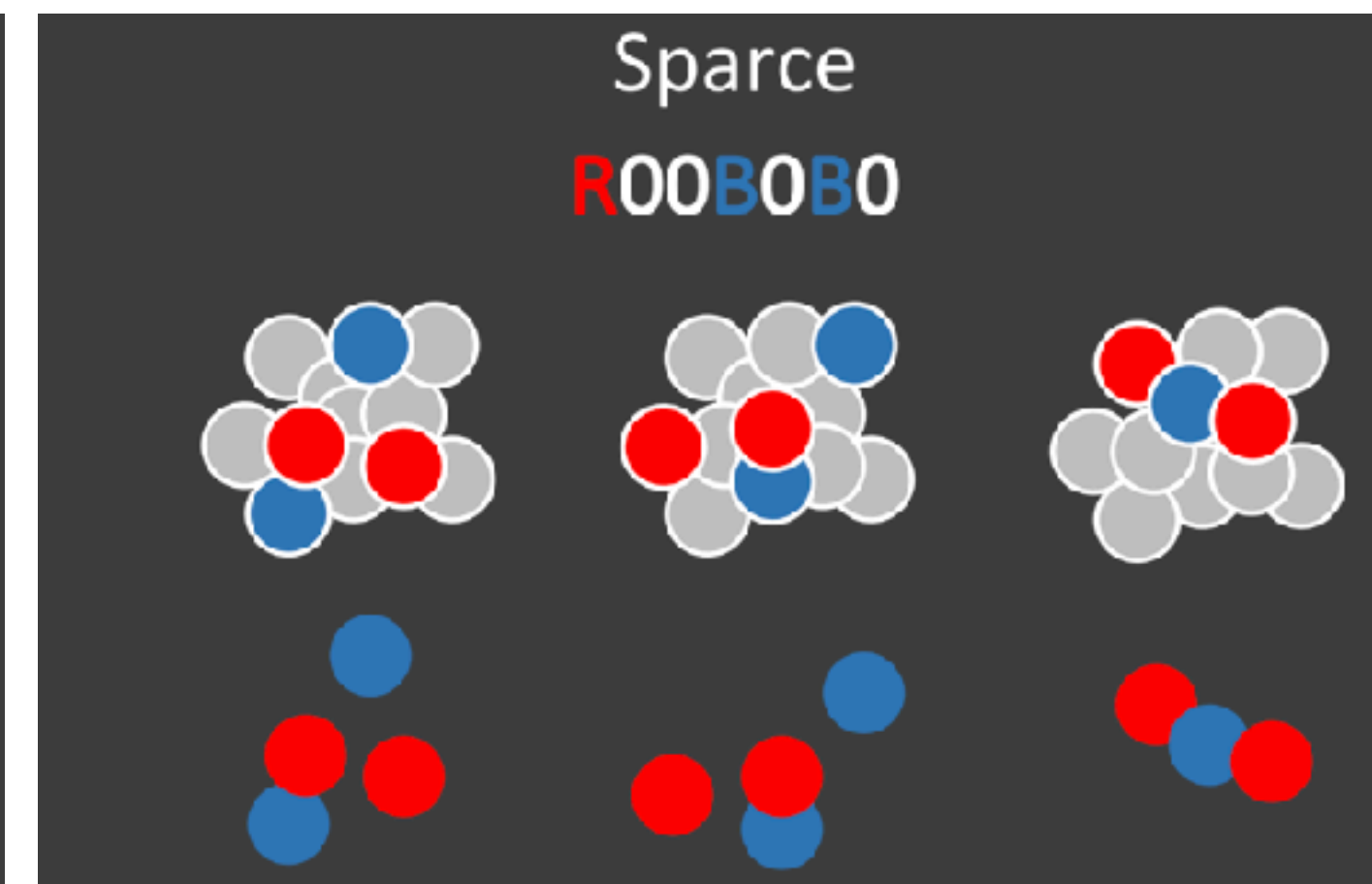
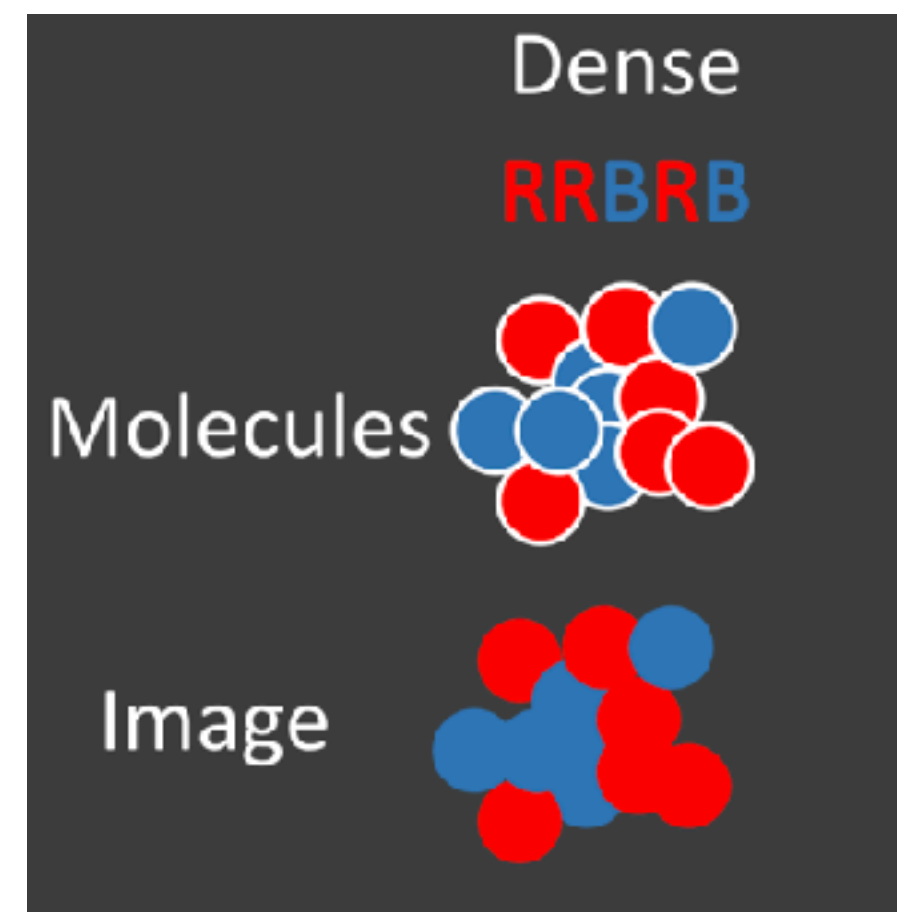
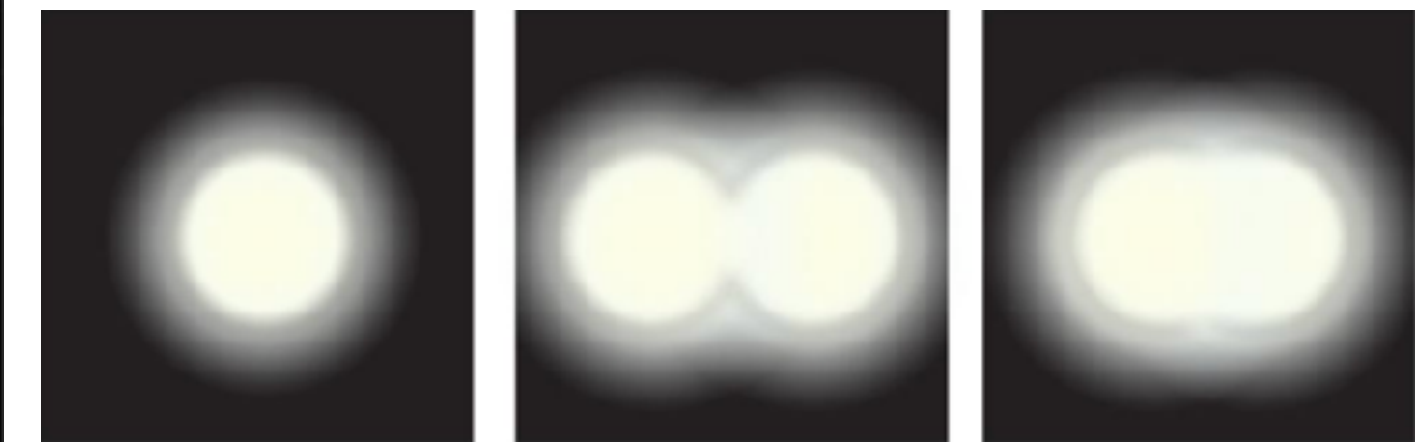
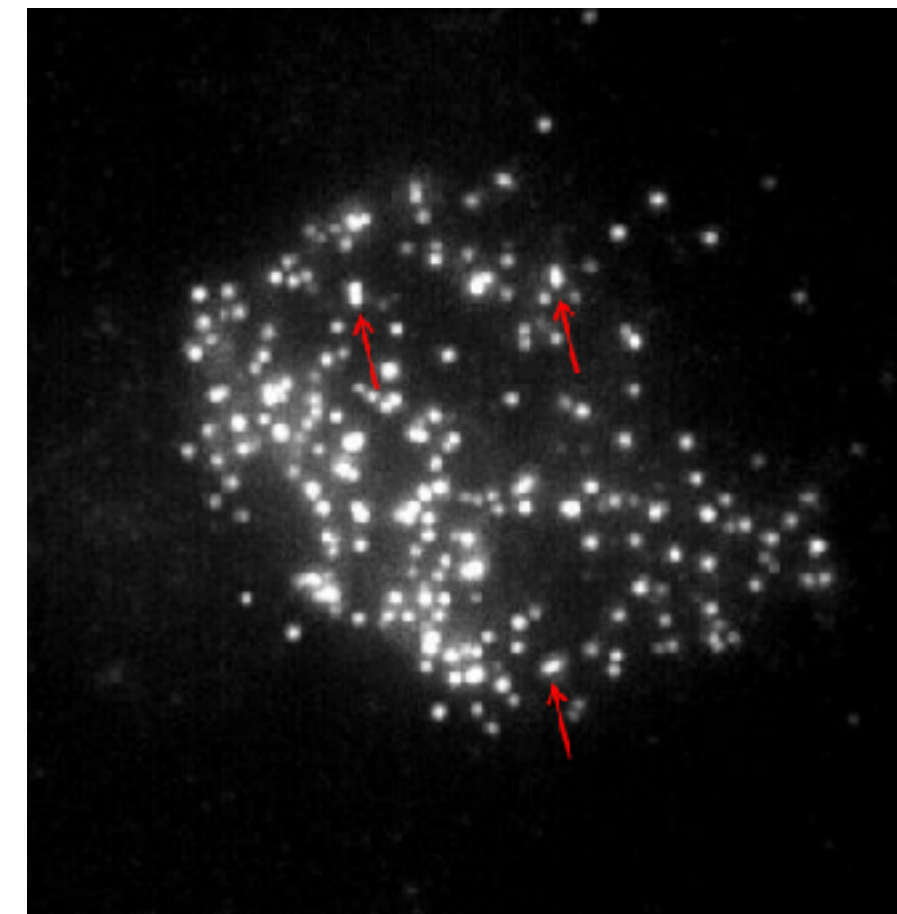
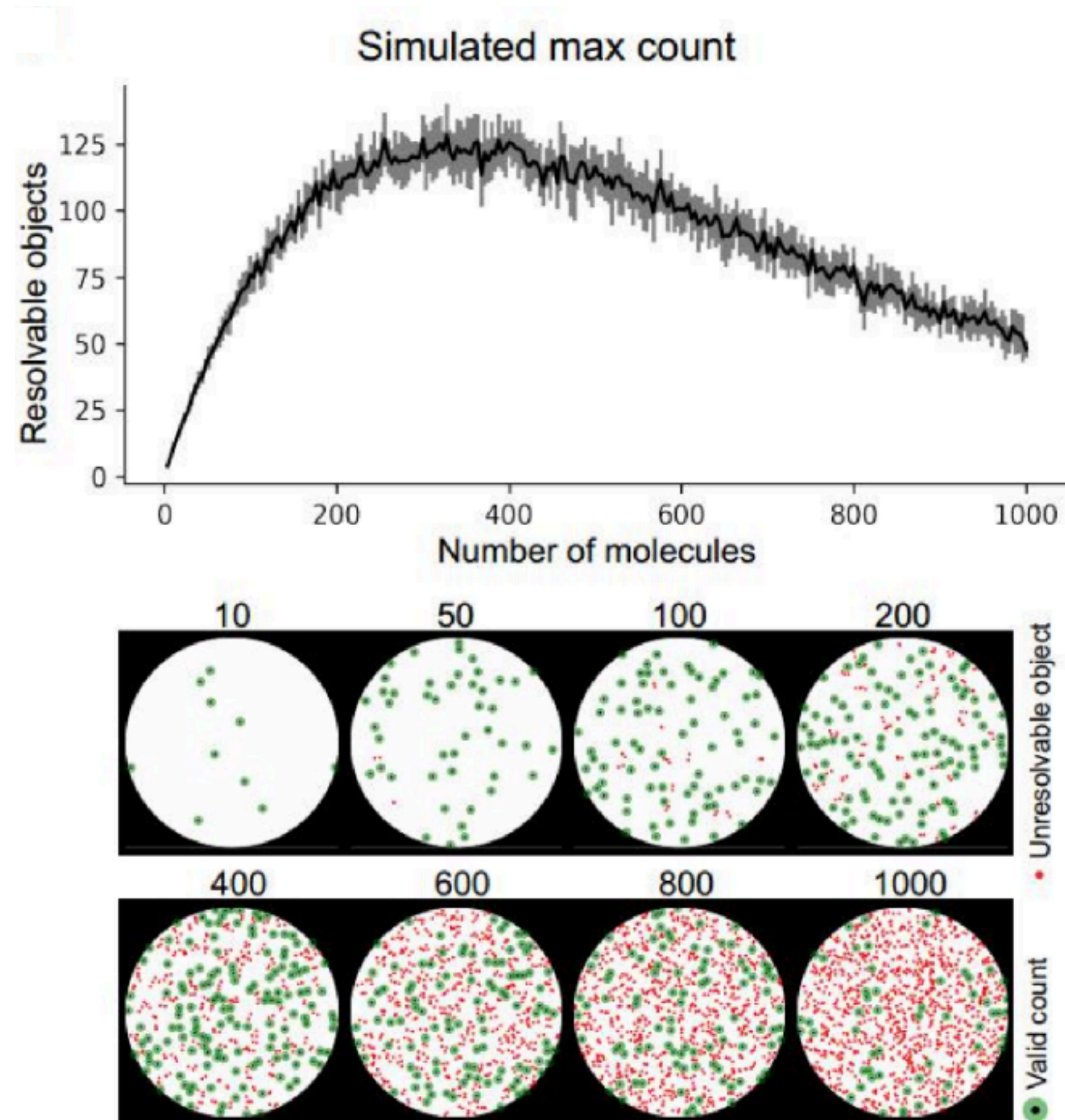
limitation:

false discoveries,
molecular crowding



Lein et al. *Science* **358**, 64-69 (2017)

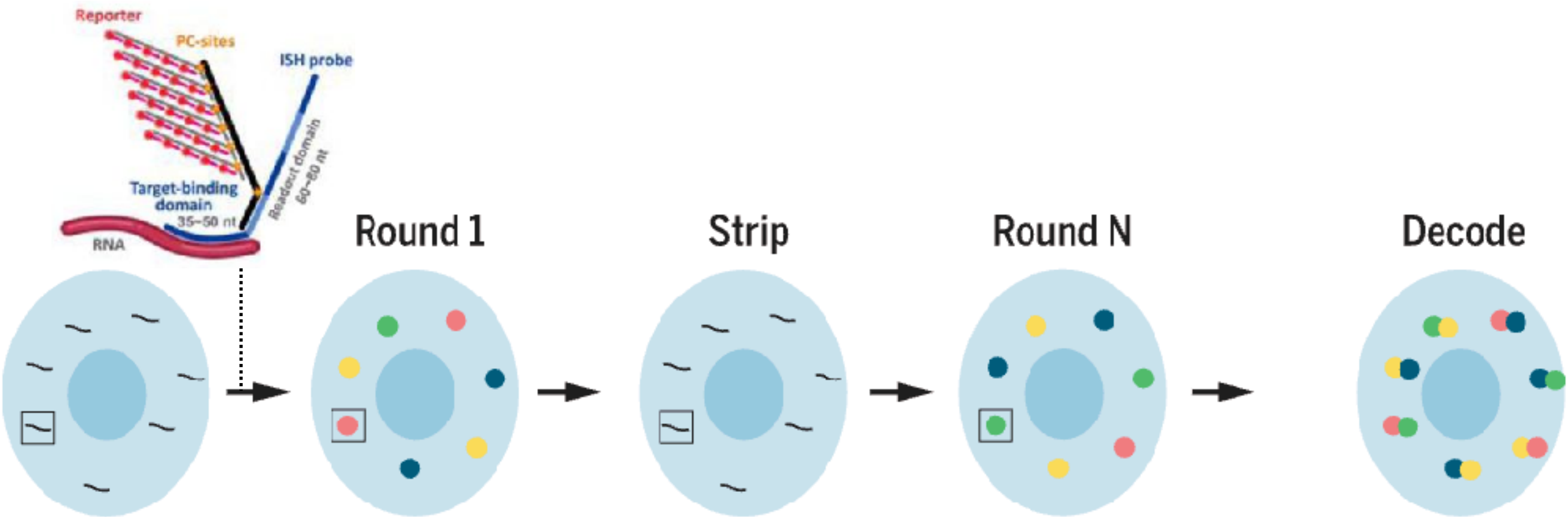
sparsity alleviates molecular crowding



error-robust barcoded FISH

same thing, but
barcodes are...

- 1. **sparse**
(we don't probe for every target, every time)
- 2. **disimilar**
(we can make mistakes during probing & decoding)



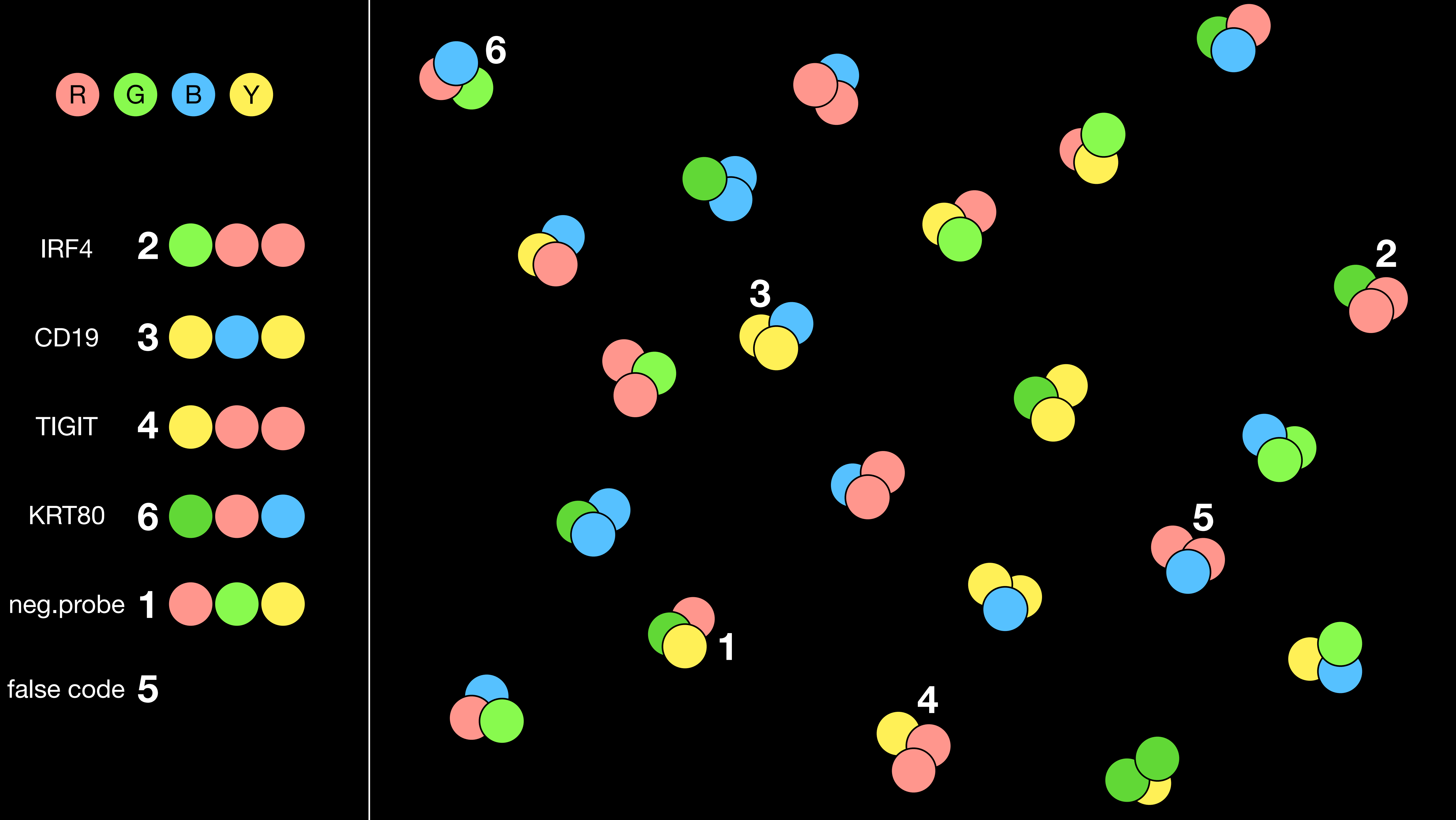
(predefined barcode catalogue)

	Round 1				Round 2			
Bit	1	2	3	4	5	6	7	8
Gene 1	1	0	0	0	0	0	1	0
Gene 2	0	0	1	0	1	0	0	0
Gene 3	1	0	0	0	0	1	0	0
Gene n	0	0	0	1	0	0	0	1

(observation)

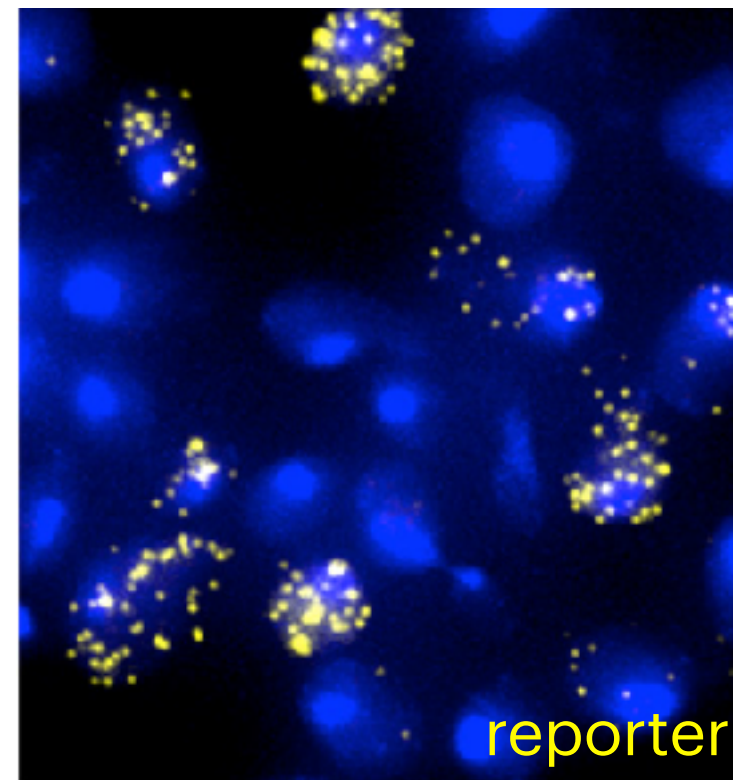
barcode

locations	R	.	.	.	.	.	G	.
	.	.	G	.	R	.	.	.
	R	.	.	.	.	Y	.	.
	.	.	B	.	.	.	B	.

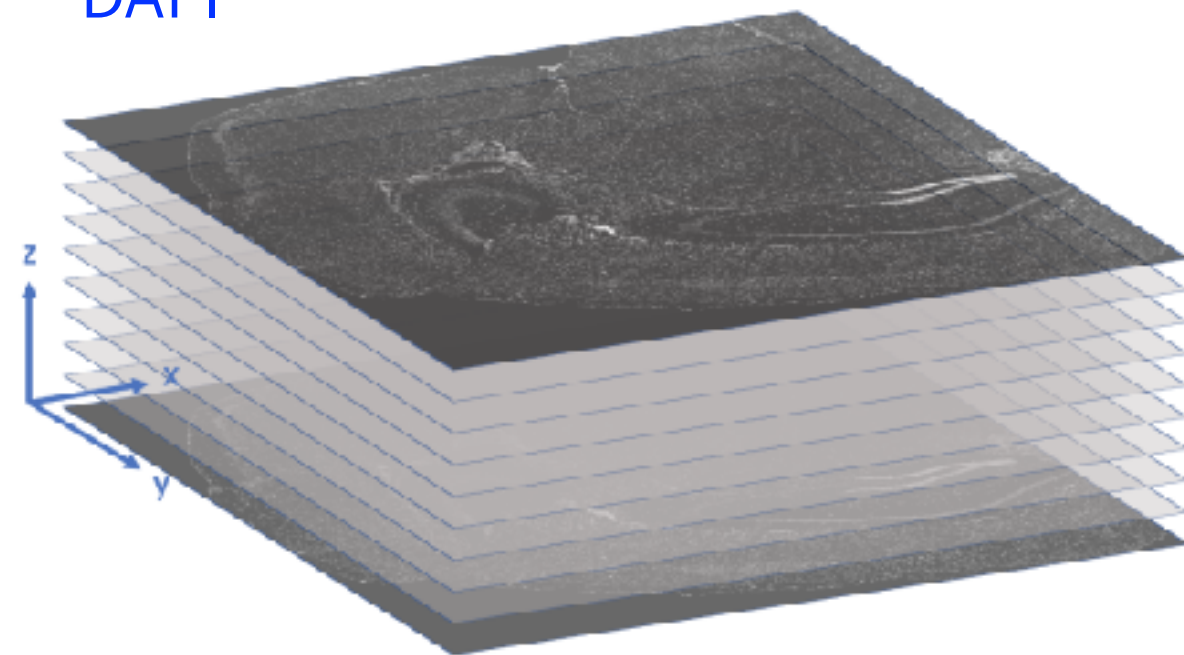


from images to gene x cell counts

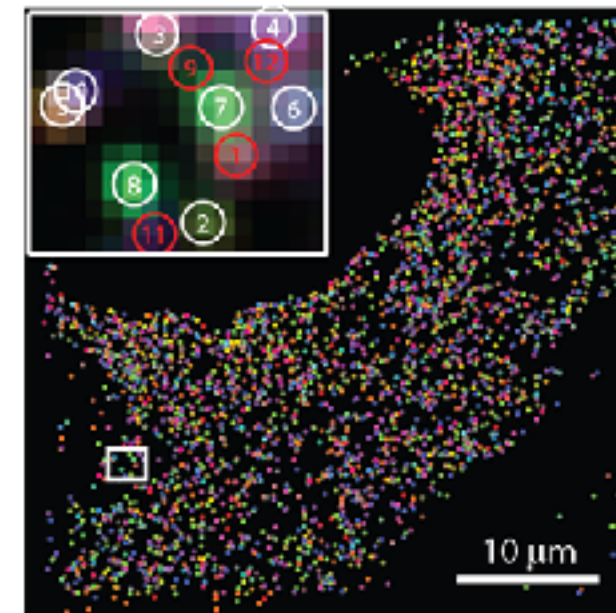
imaging



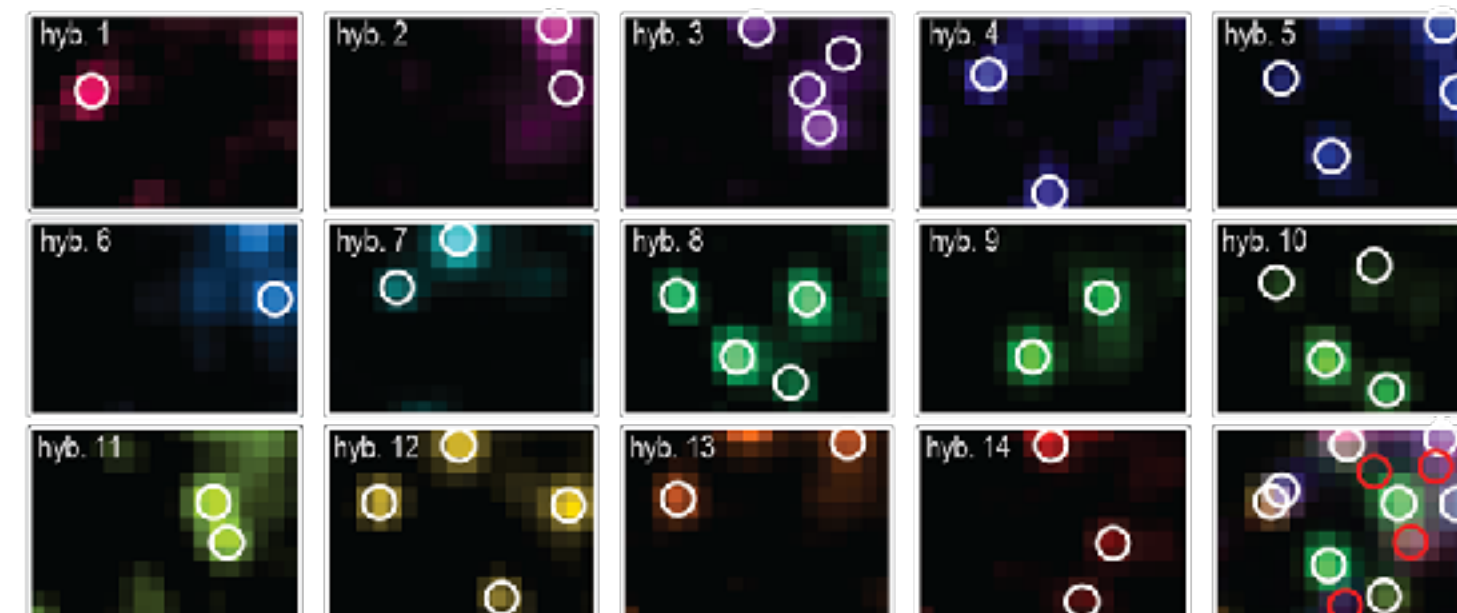
DAPI



spot-calling



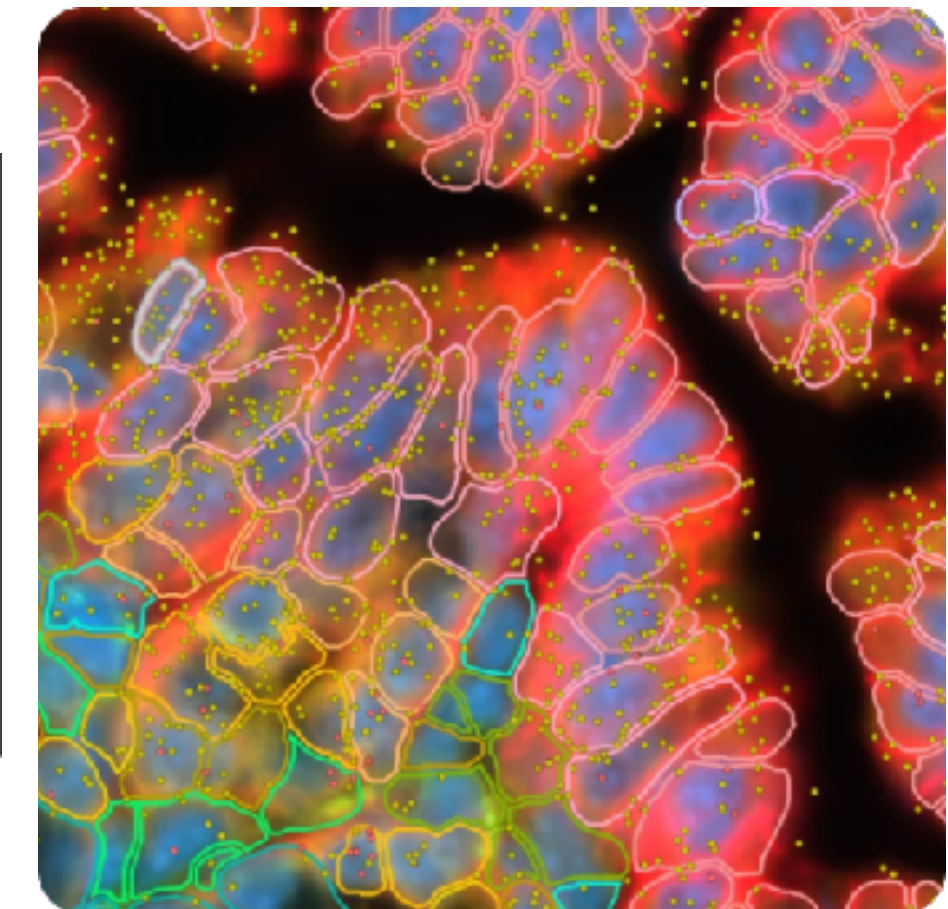
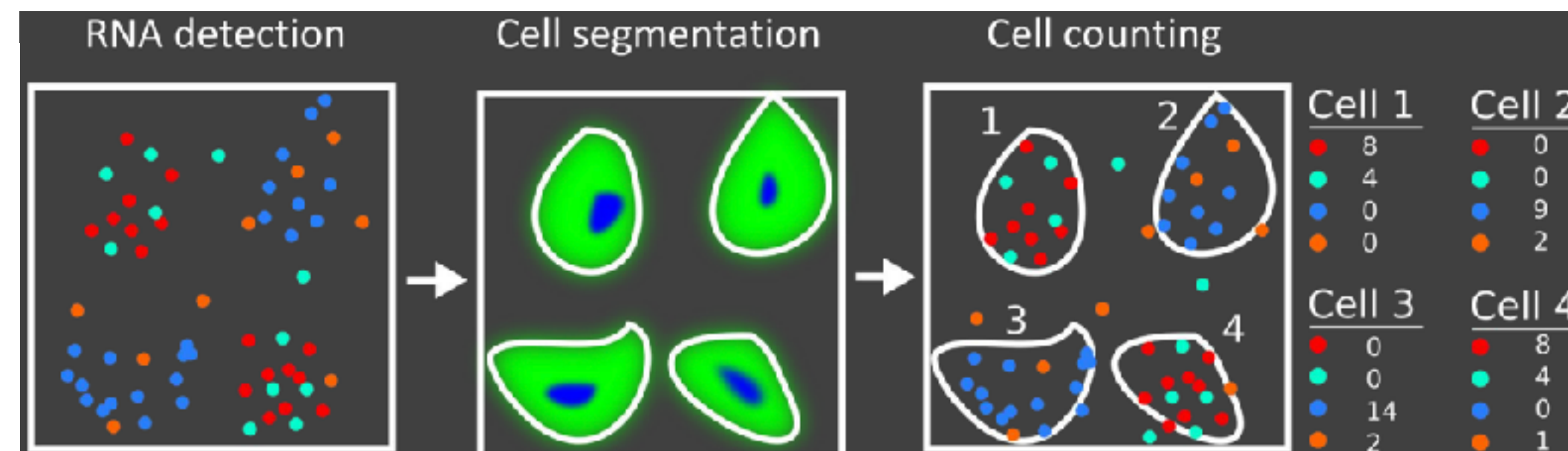
decoding



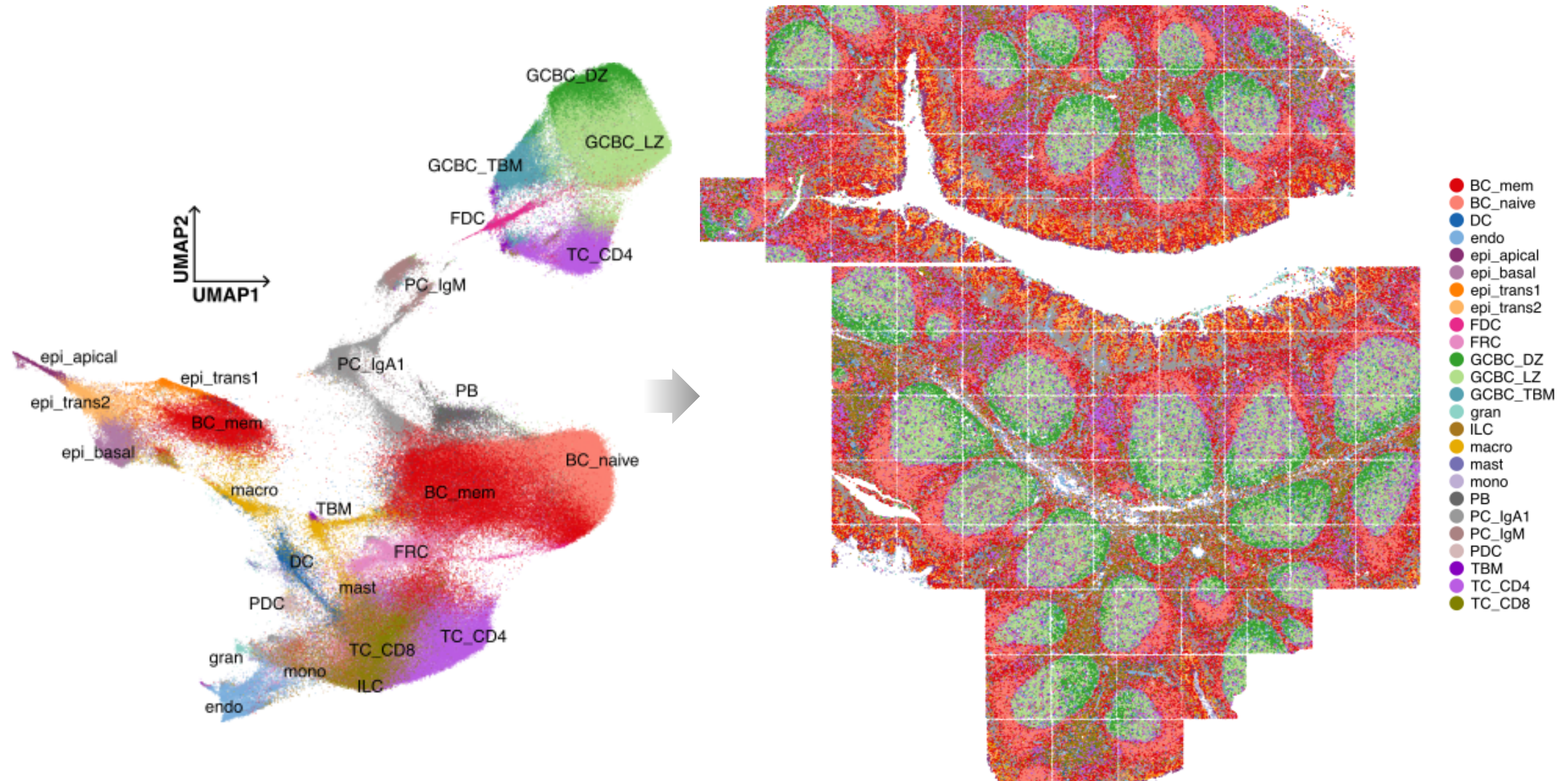
segmentation

	C1	C2	C3	C4	C5
G1	1	0	4	2	9
G2	8	0	0	2	0
G3	0	0	0	0	0
G4	3	0	3	5	1

Feature Count Matrix



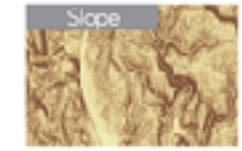
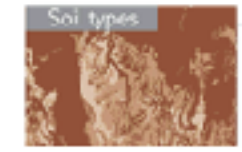
from single-cell to tissue maps



ST data is **multifaceted** & finds resemblance in many fields...

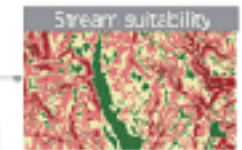
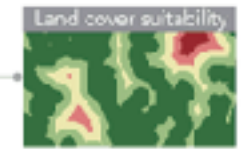
Collect source layers

Data is first digitized into either polygon or raster layers. This housing suitability data is raster.



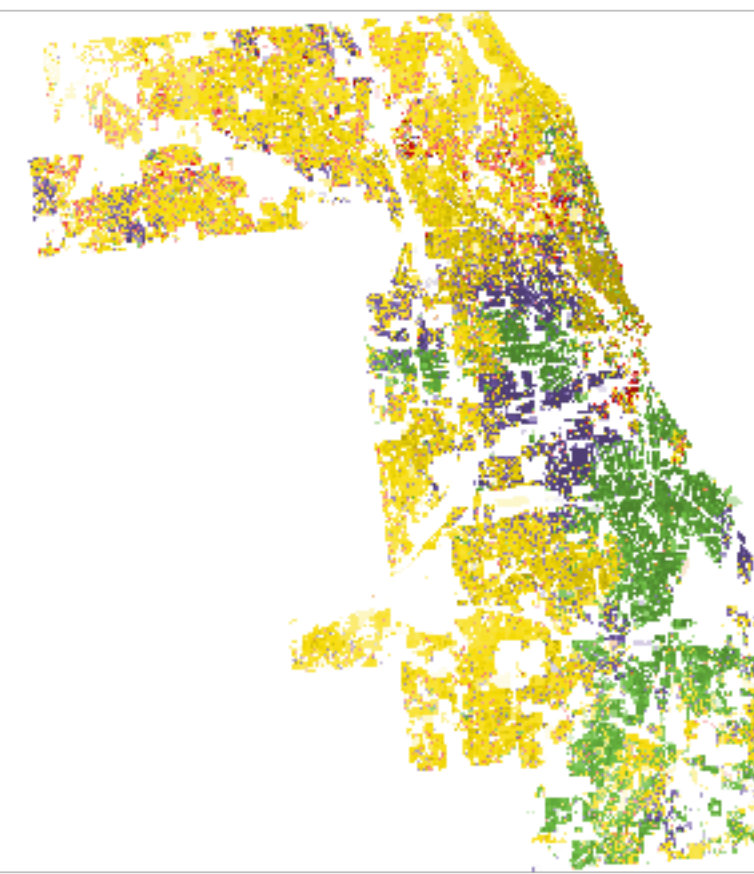
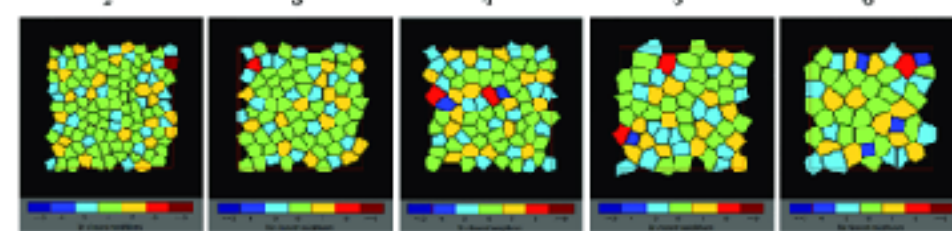
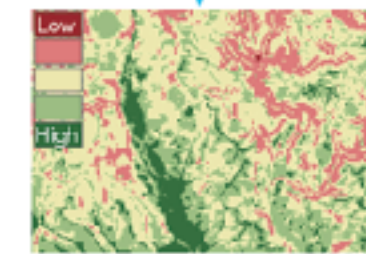
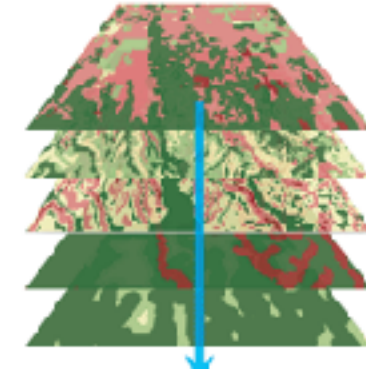
Create suitability layers

Each layer is now classified to use a common suitability scale: for example, low suitability could be assigned a value of 1 (dark red) and high suitability a value of 5 (dark green).

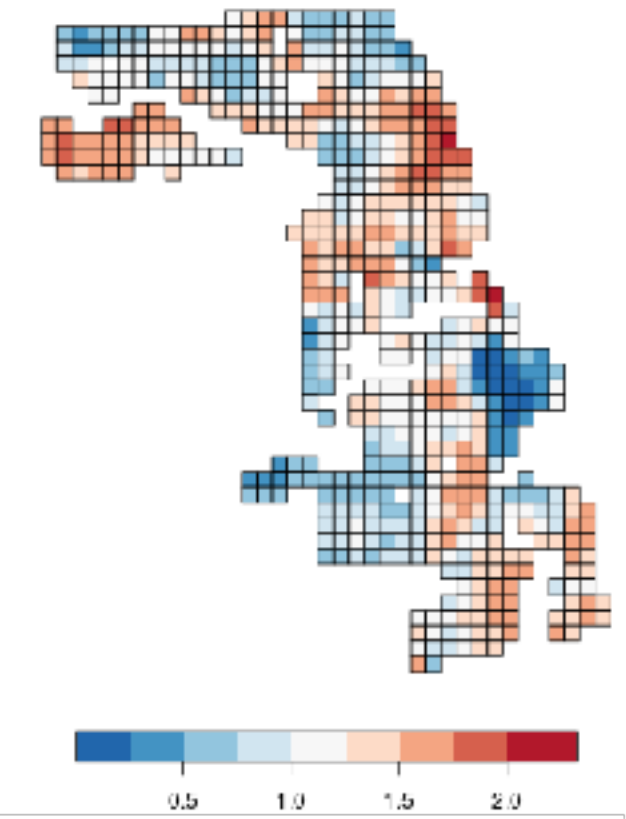


Calculate weighted overlay

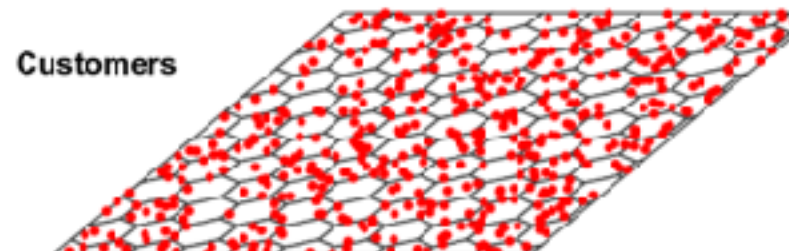
Suitability layers are overlaid so that each cell gets an overall suitability rating. Weights of relative importance are assigned to each layer.



Racial diversity (Entropy) at the scale of 1.8 km



Customers



Roads



Population density

