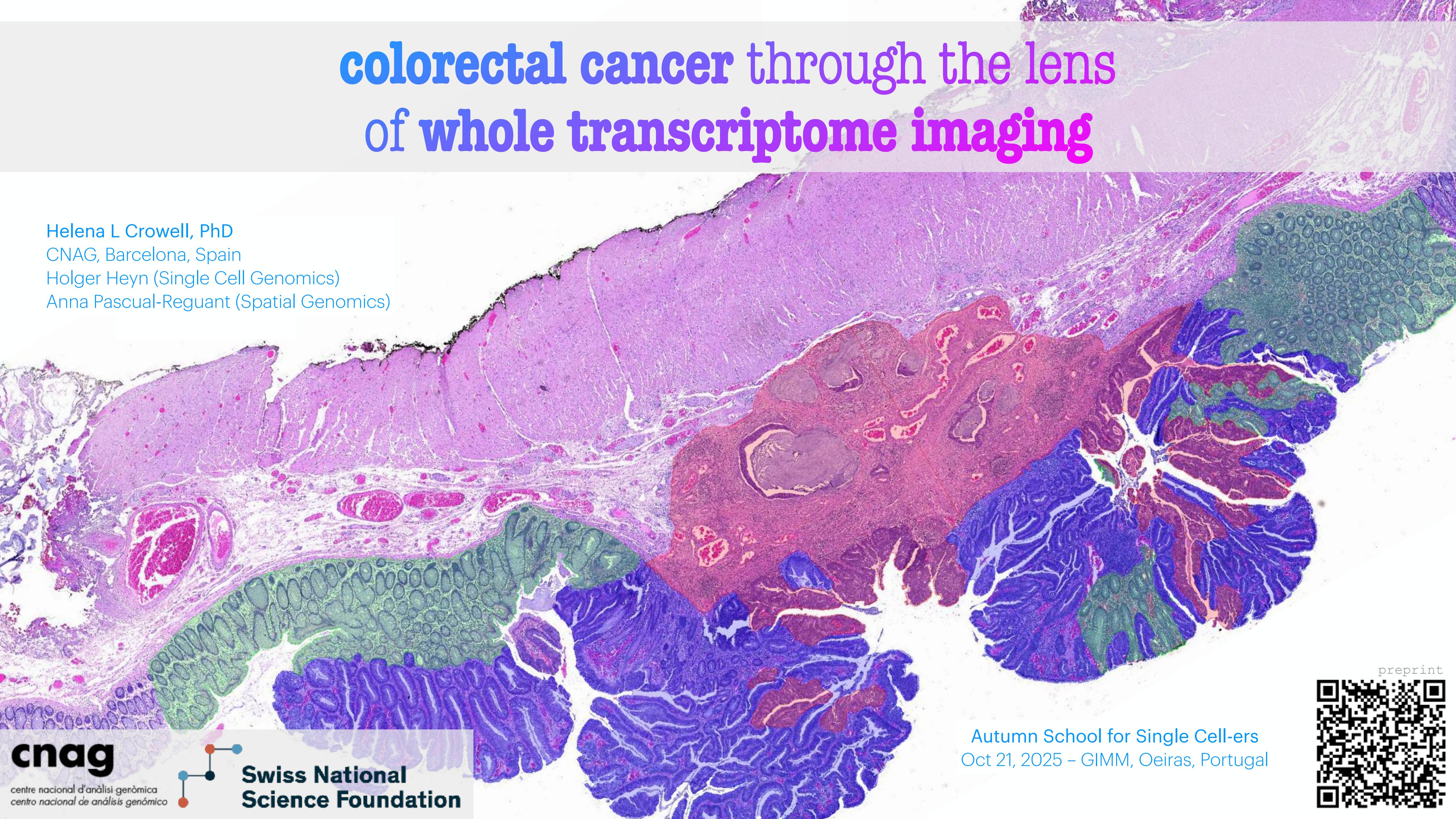


colorectal cancer through the lens of whole transcriptome imaging

Helena L Crowell, PhD
CNAG, Barcelona, Spain
Holger Heyn (Single Cell Genomics)
Anna Pascual-Reguant (Spatial Genomics)

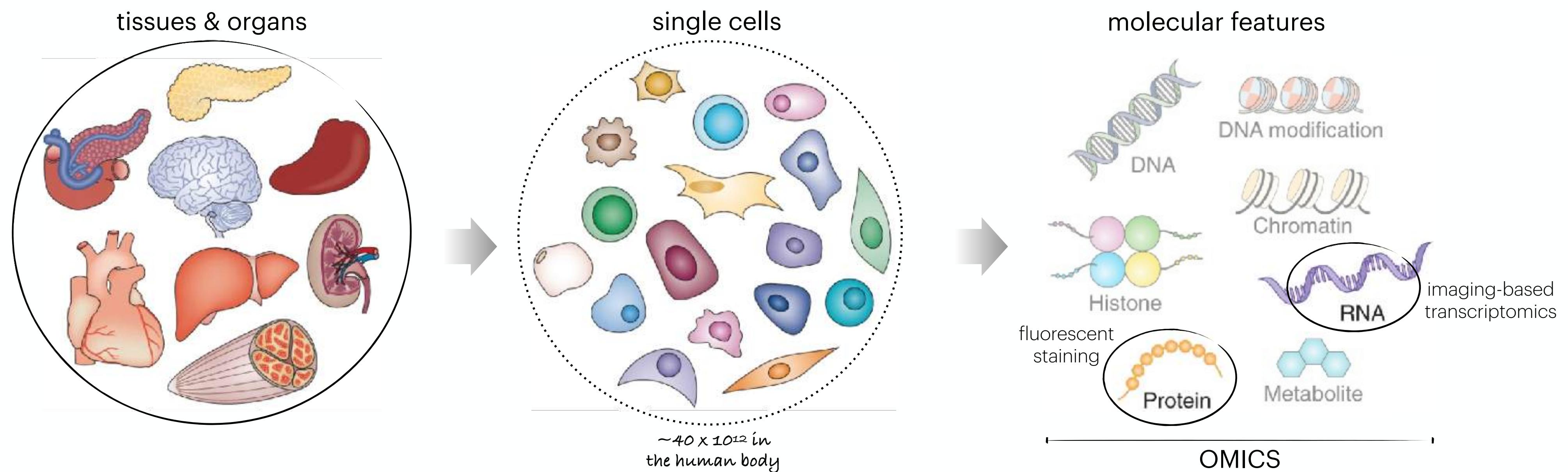


preprint

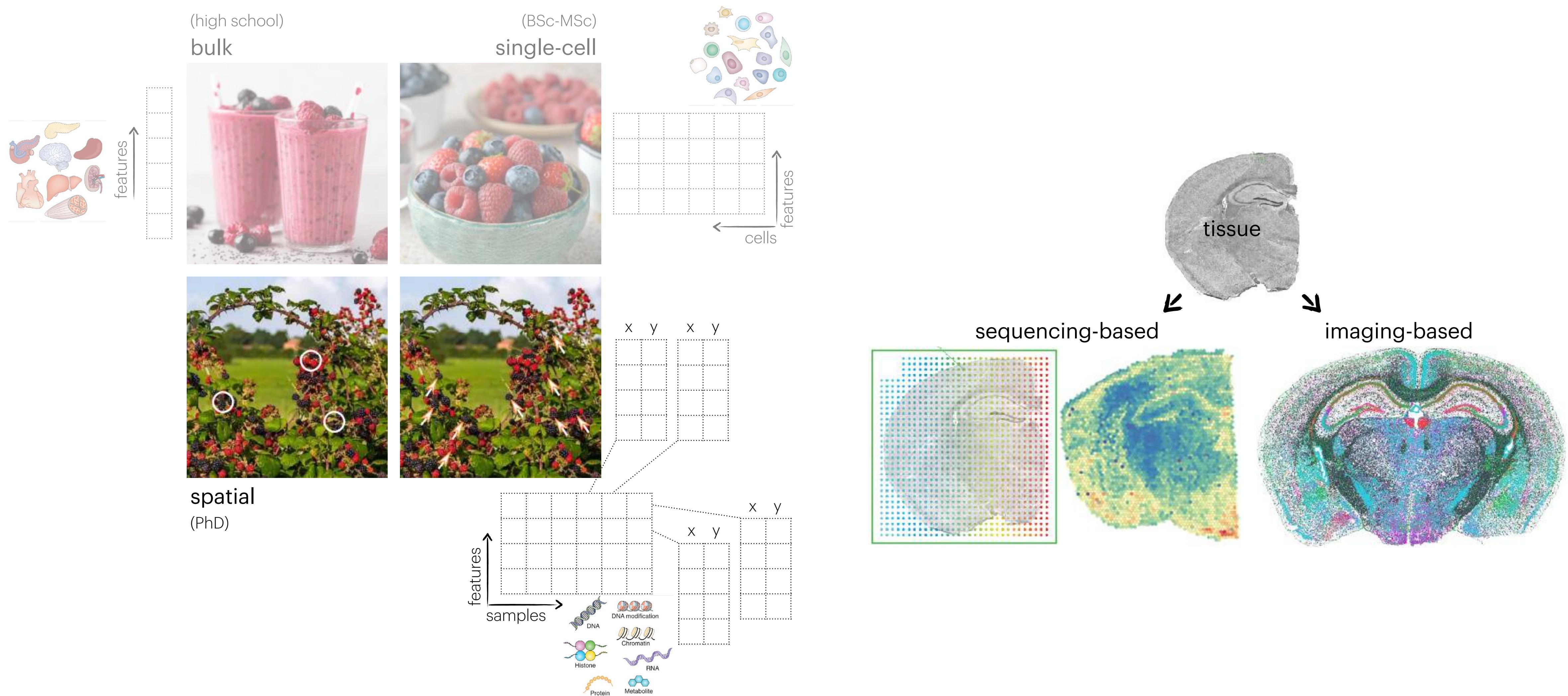
Autumn School for Single Cell-ers
Oct 21, 2025 – GIMM, Oeiras, Portugal



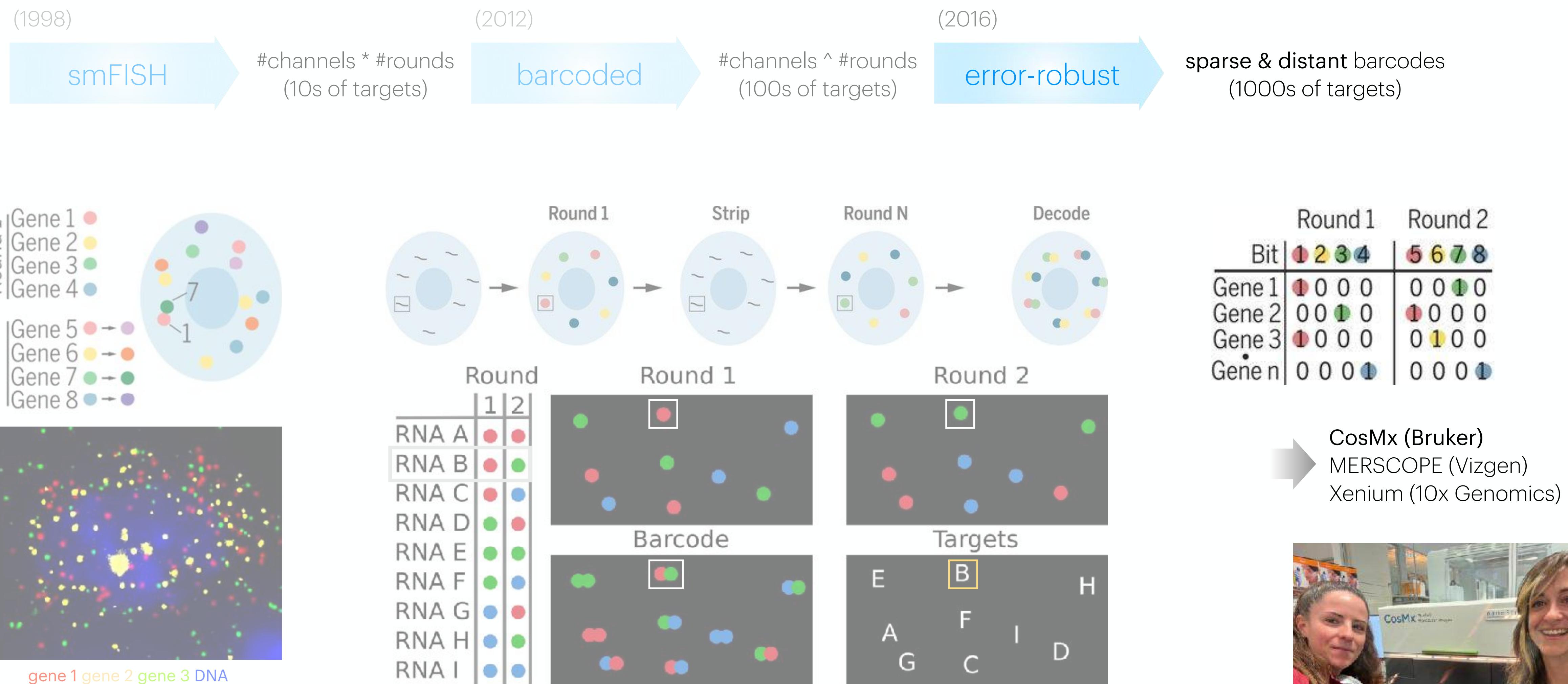
multi-cellular system & the **omics** era



from bulk to single cell to **spatial** omics



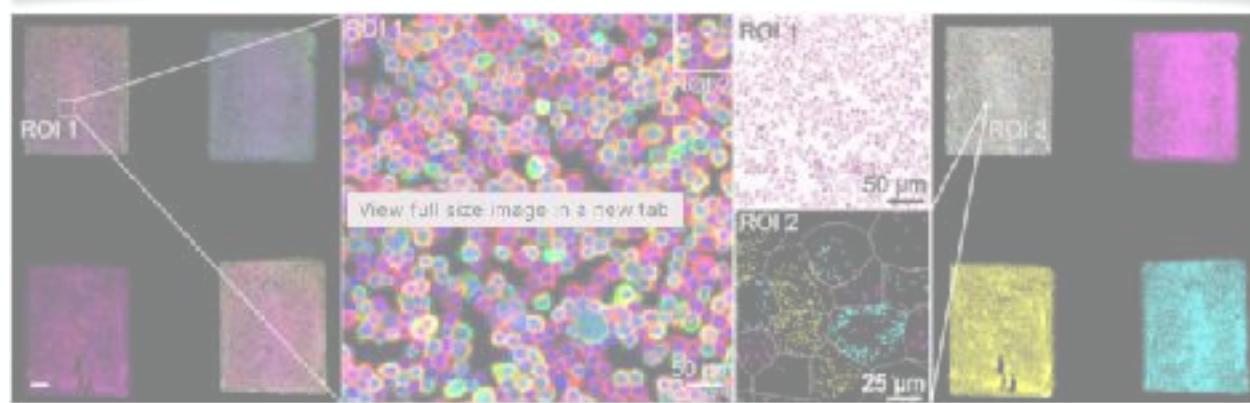
from smFISH to whole transcriptome imaging



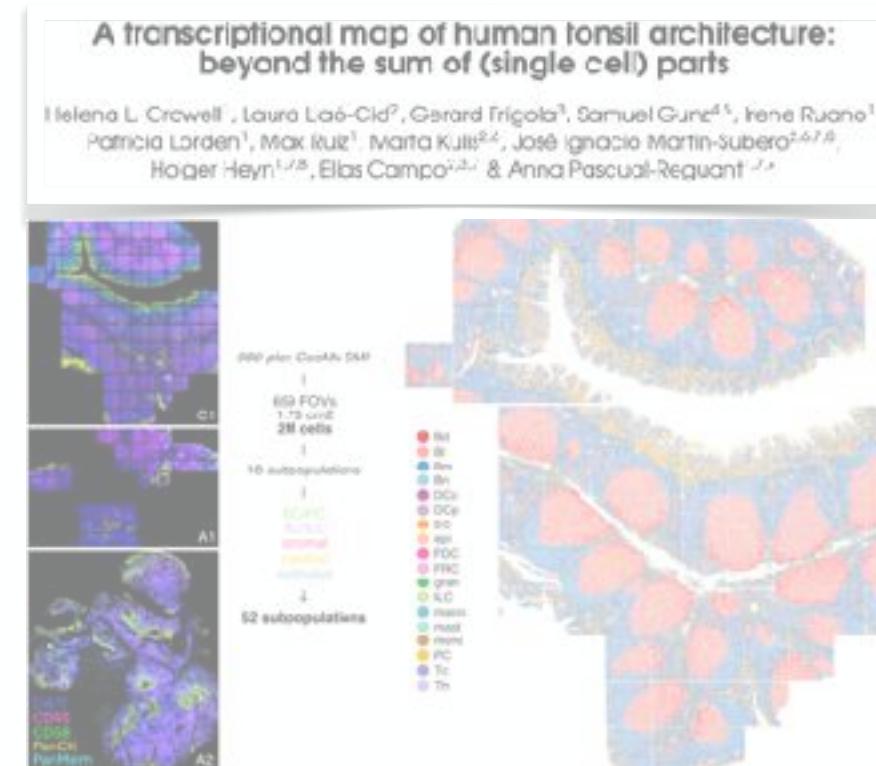
CosMX serial #1

research at CNAG: lots of tech, lots of data

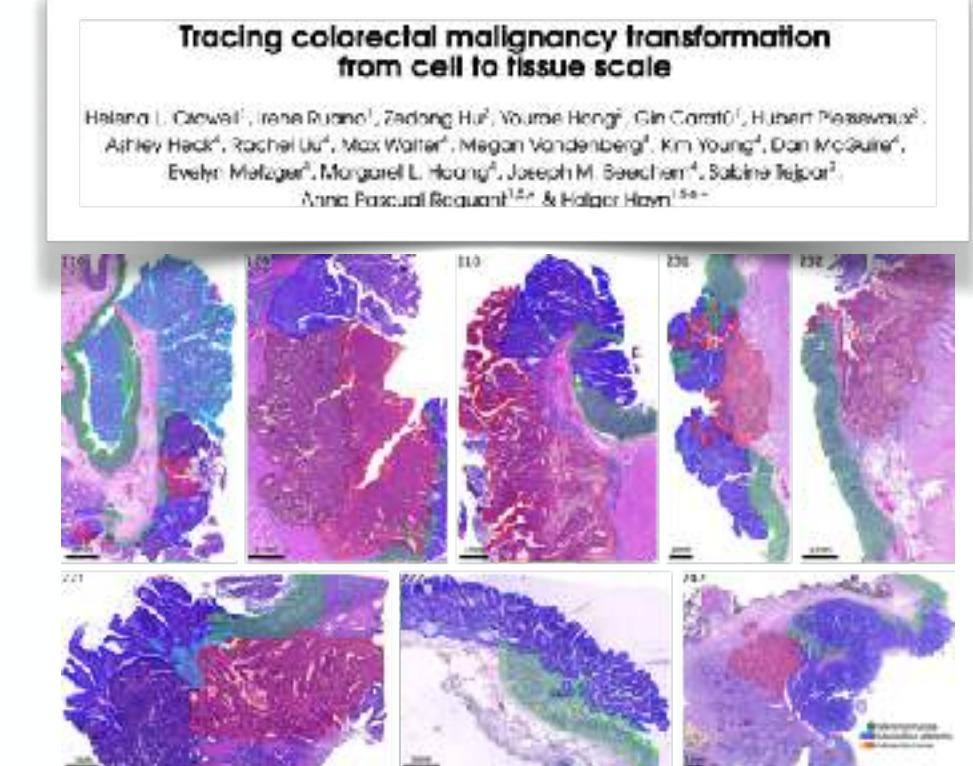
STAMP: Single-cell transcriptomics analysis and multimodal profiling through imaging



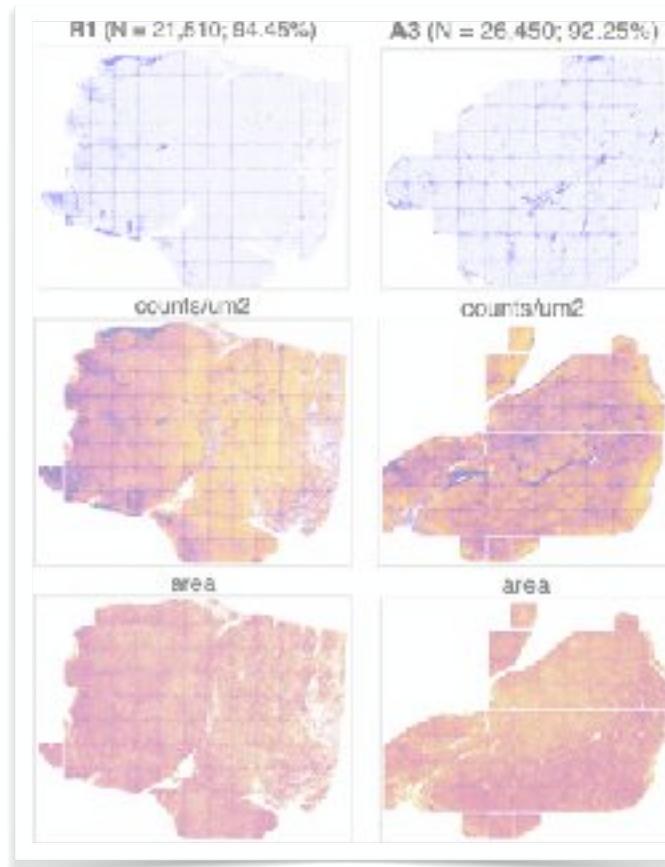
imaging instead
of sequencing



1k-plex CosMx on human tonsil



19k-plex CosMx on colorectal cancer



6k-plex CosMx on TMA of lymphoma

single-cell
FACS
scRNA-seq
scATAC-seq
snRNA-seq
snPATHO-seq
10x Multiome
BCR- & TCR-seq
Chromium Connect

spatial
CosMx & Xenium
Visium & HDST
Stereo-seq



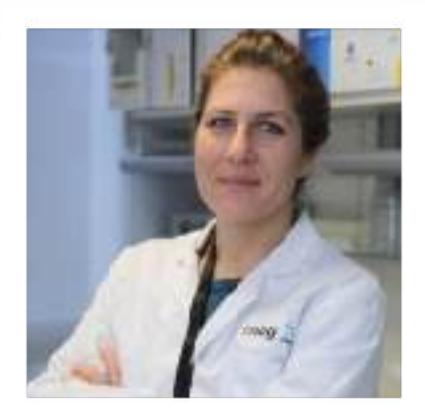
acknowledgements



Centro Nacional de Análisis Genómico (CNAG), Barcelona



Irene Ruano
in-house pathologist



Ginevra Caratù
lab technician



Anna Pascual-Reguant
Spatial group lead



Holger Heyn
Single Cell group lead



dry lab '23



dry lab '24



dry + wet lab '24



CosMx serial #1



Dept. of Oncology/Hepato-Gastroenterology, KU Leuven



Sabine Tejpar
clinician



Zhedong Hu
predoc



Yourae Hong
postdoc



Huber Piessevaux
clinician



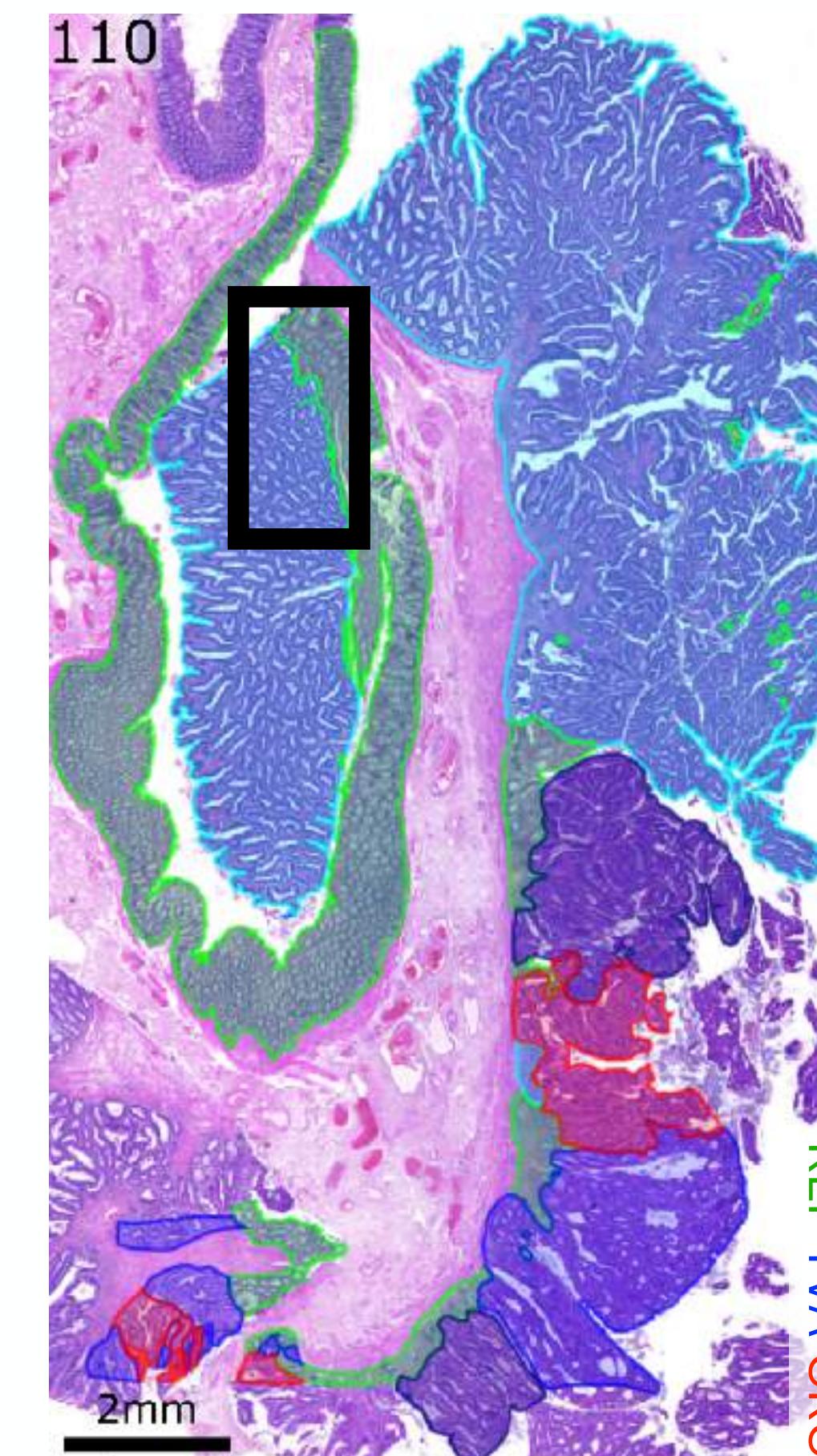
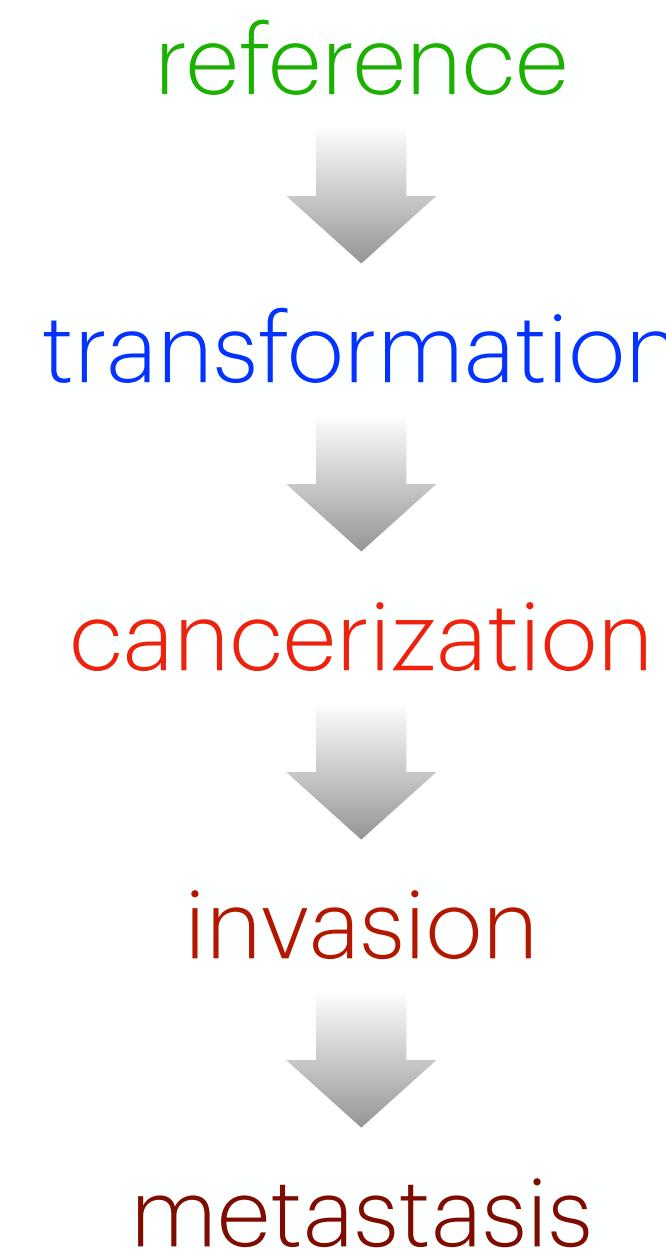
Bruker Spatial Biology, Seattle

Ashley Heck
Rachel Liu
Maximilan Walter

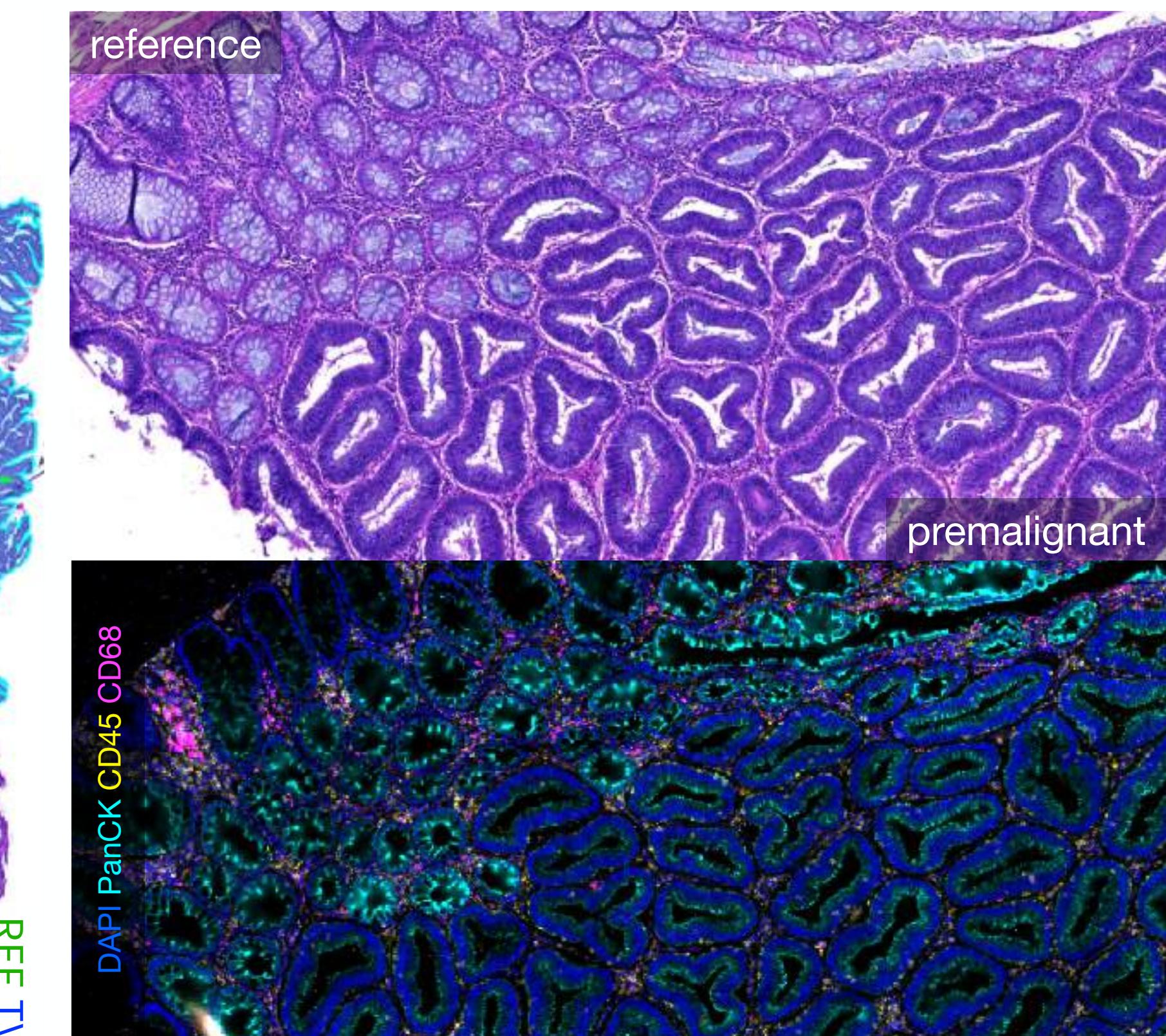
Megan Candenbergh
Kimberley Young
Dan McGuire

Evelyn Metzger
Margaret Hoang
Joseph Beechem

a brief intro to colorectal cancer



TVA = tubulovillous adenoma
CRC = colorectal cancer



? REF → TVA → CRC ?
healthy | pre- | malignant

bookkeeping: the ‘starting material’

WTx CosMx SMI data

IF staining

- DAPI (nuclei)
- CD45 (immune)
- PanCK (epithelia)
- CD68 (macrophages)
- B2M/CD298 (membrane)

~19k RNA targets

- 2,745 false codes
- 50 negative probes

~1,6k FOVs = 4.3cm² area

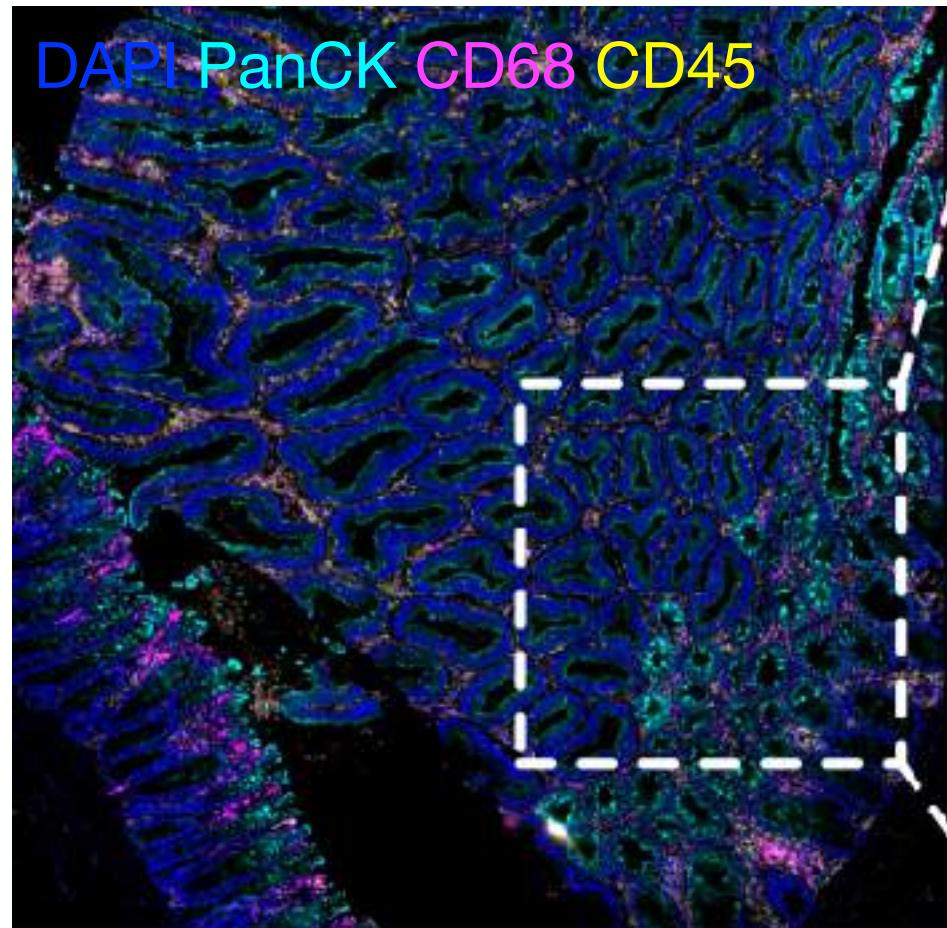
~3.5M cells from 9 sections

snPATHO-seq data

- ~125k cells on adjacent sections

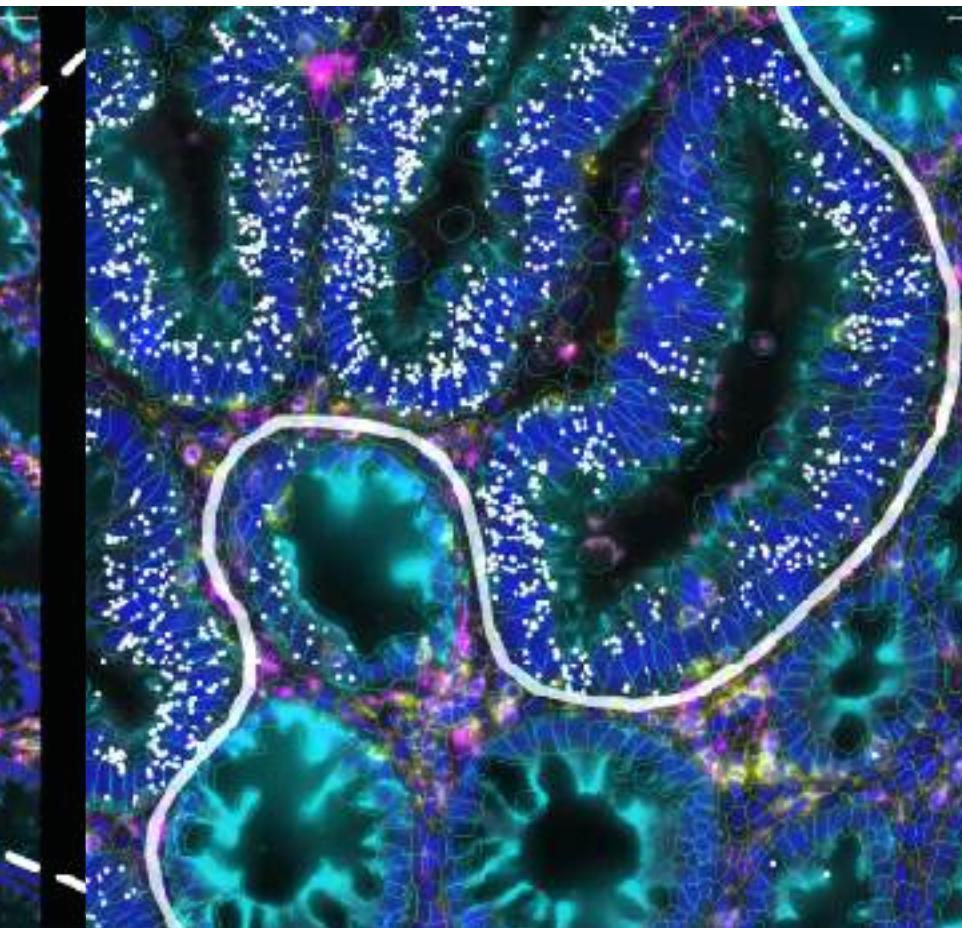
→ ~300GB

immunofluorescence

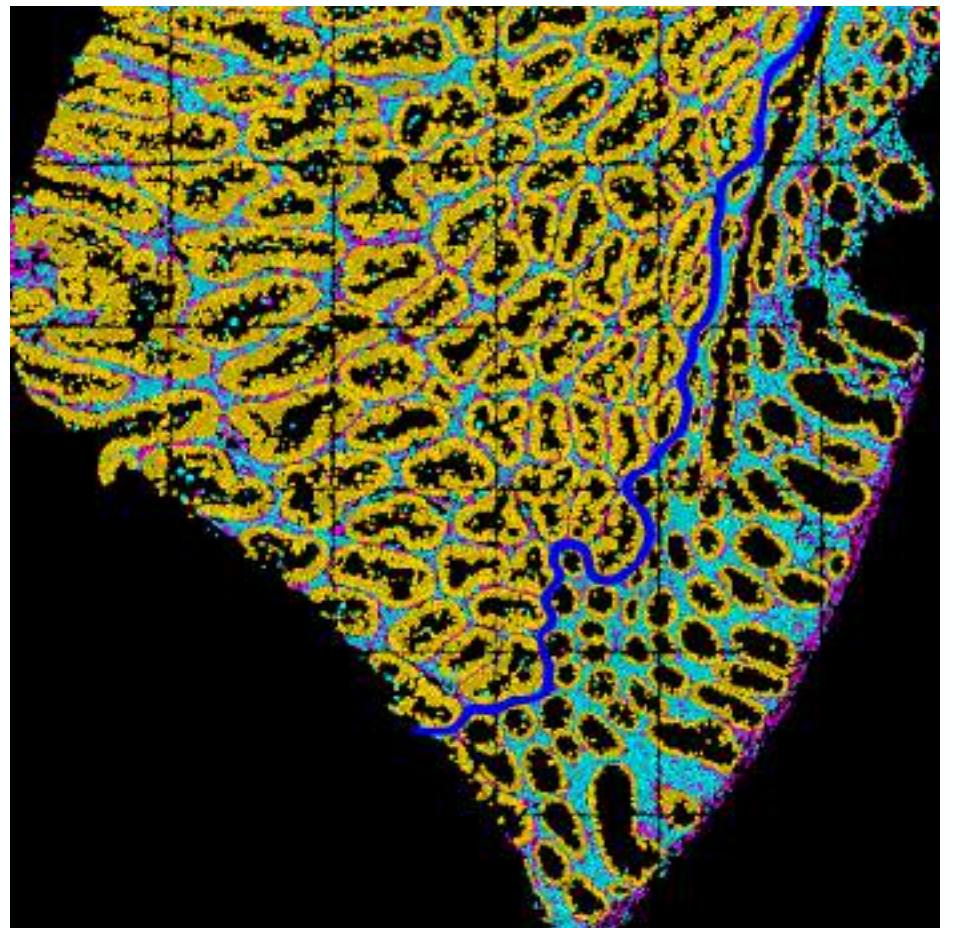
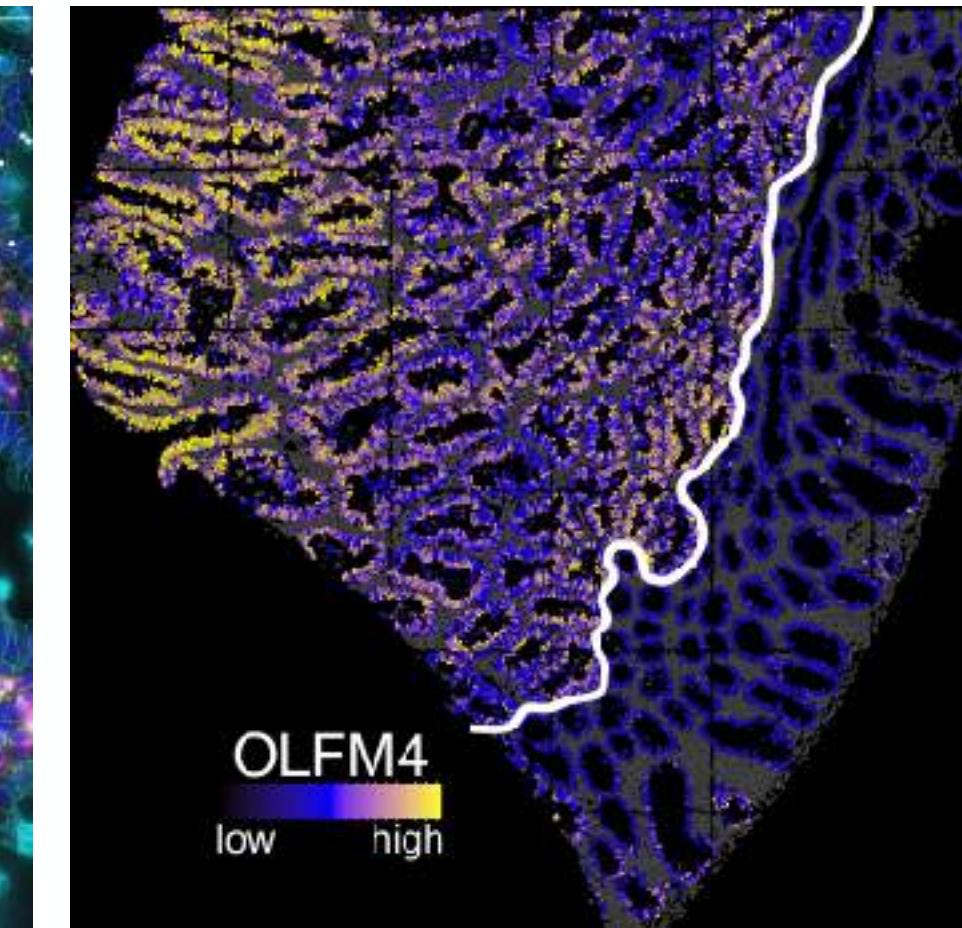


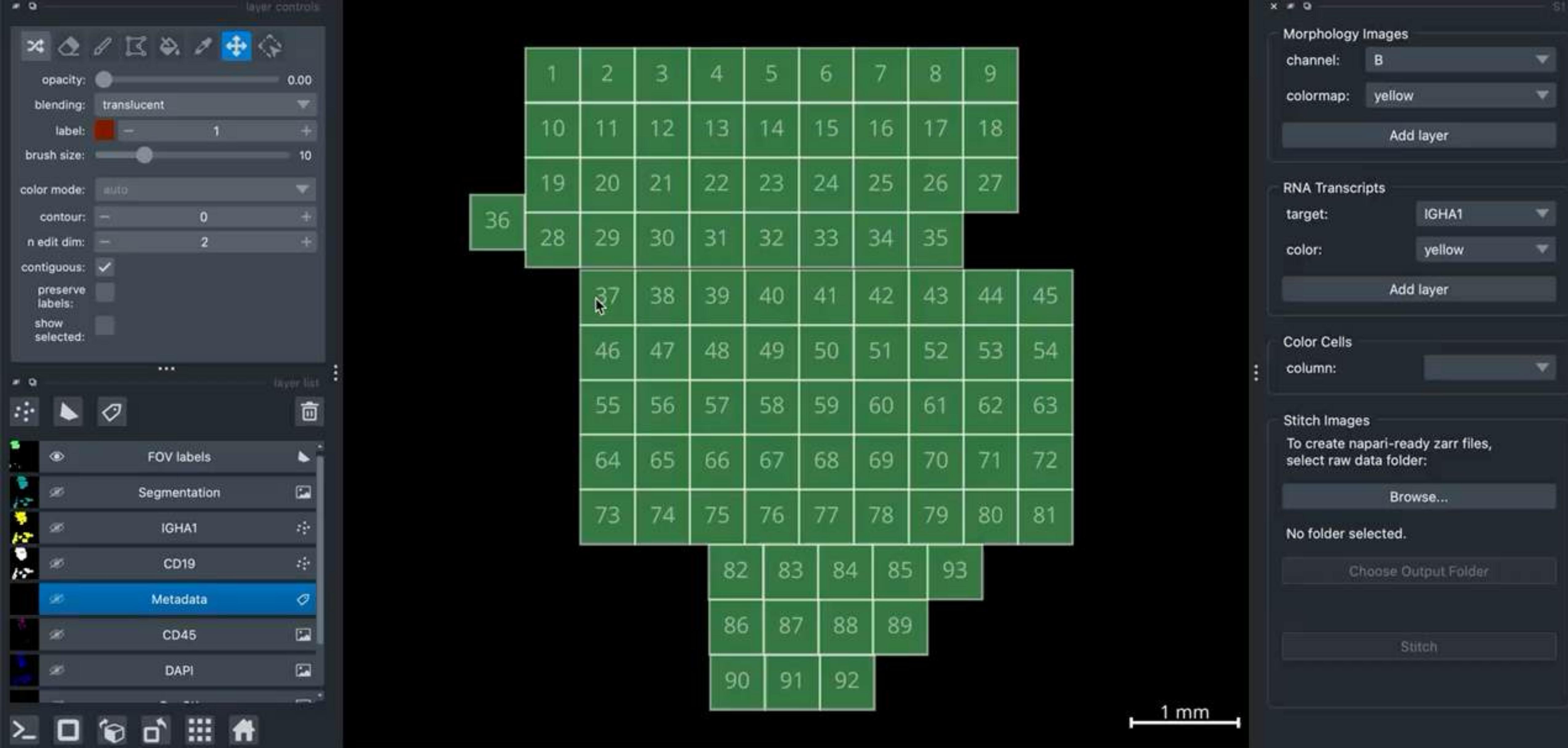
TVA | REF

transcript molecules



gene expression



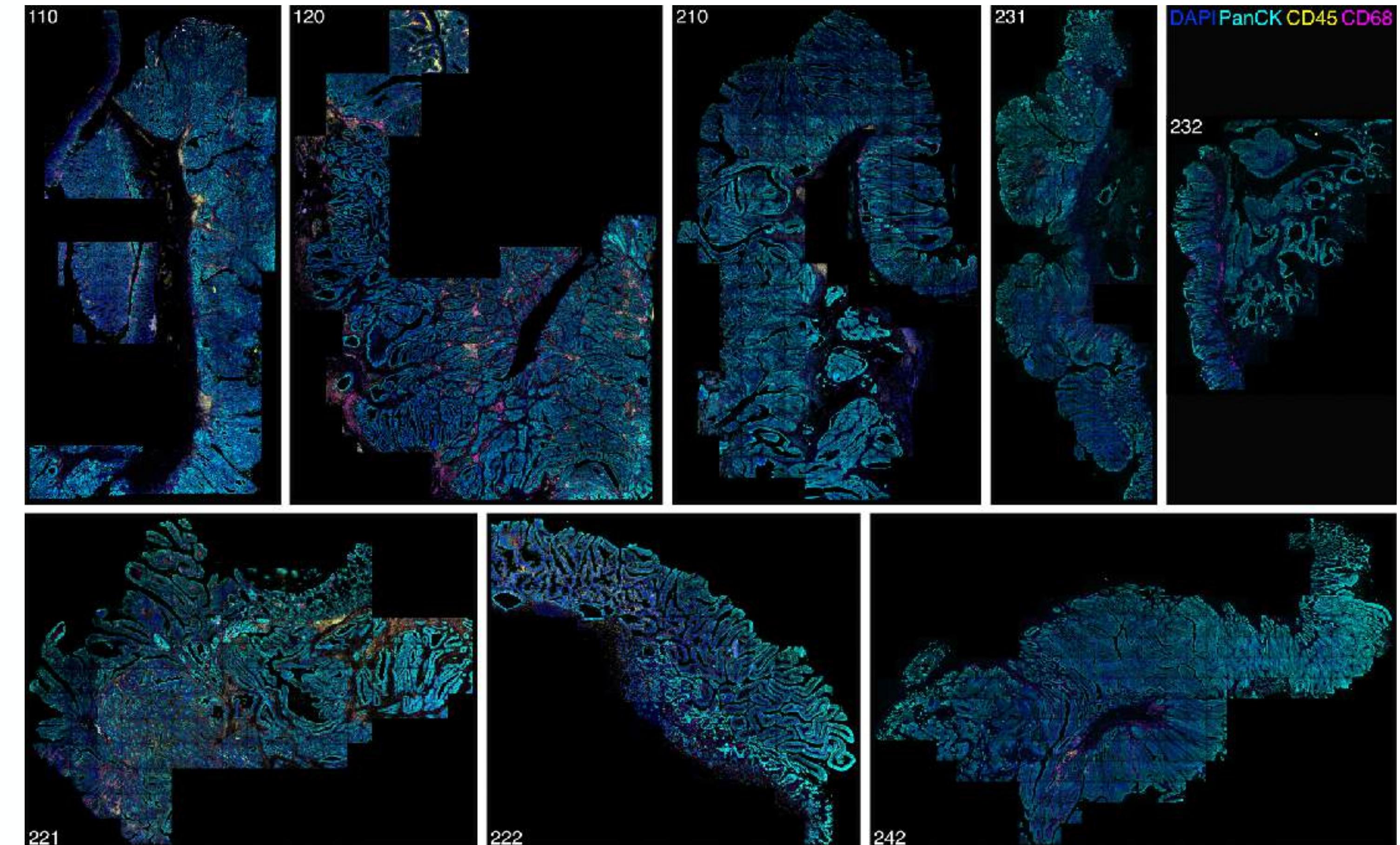


```
...: viewer.layers["DAPI"].contrast_limits = [50, 200]
...: viewer.layers["CD45"].contrast_limits = [220, 580]
...: viewer.layers["PanCK"].contrast_limits = [250, 1000]
...: viewer.layers["PanMem"].contrast_limits = [150, 700]
```

use <7> for transform, use <1> for activate the label eraser, use <2> for activate the paint brush, use <3> for activate the polygon tool, use <4> for activate the fill bucket, use <5> for pick mode

^activity

histopathology: REF/TVA/CRC annotation

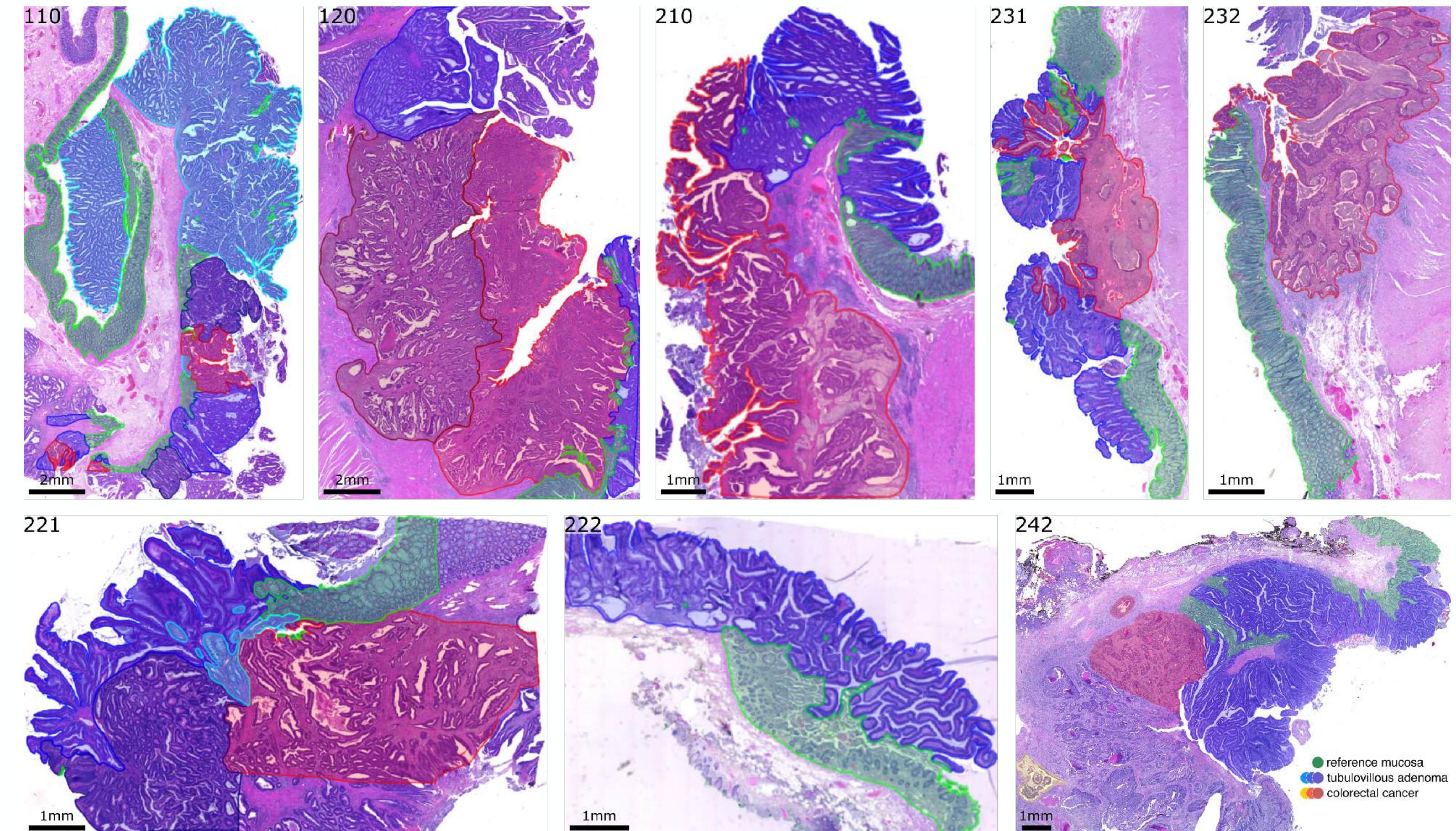


histopathology: REF/TVA/CRC annotation

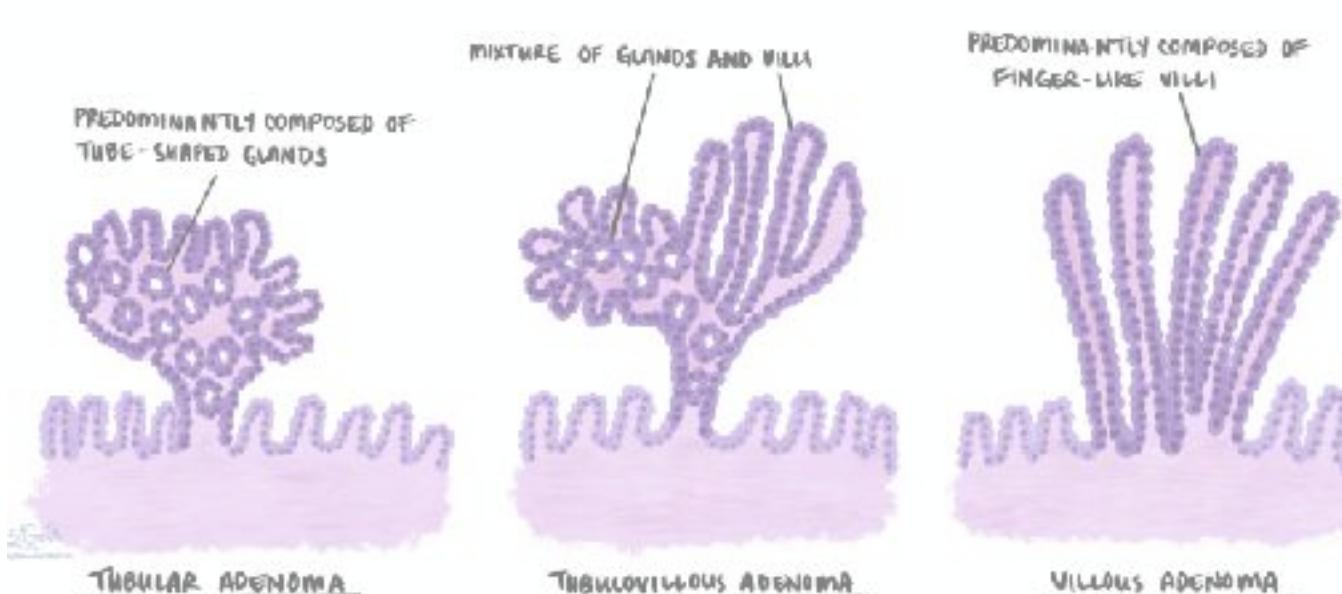
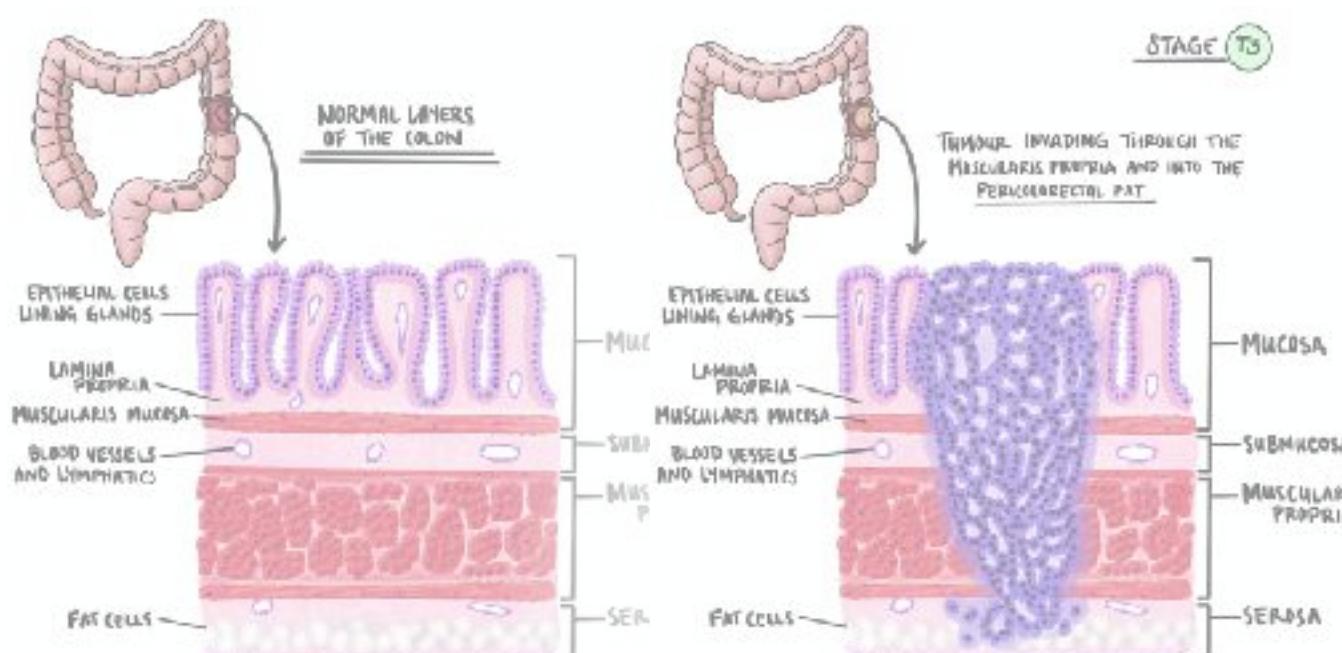
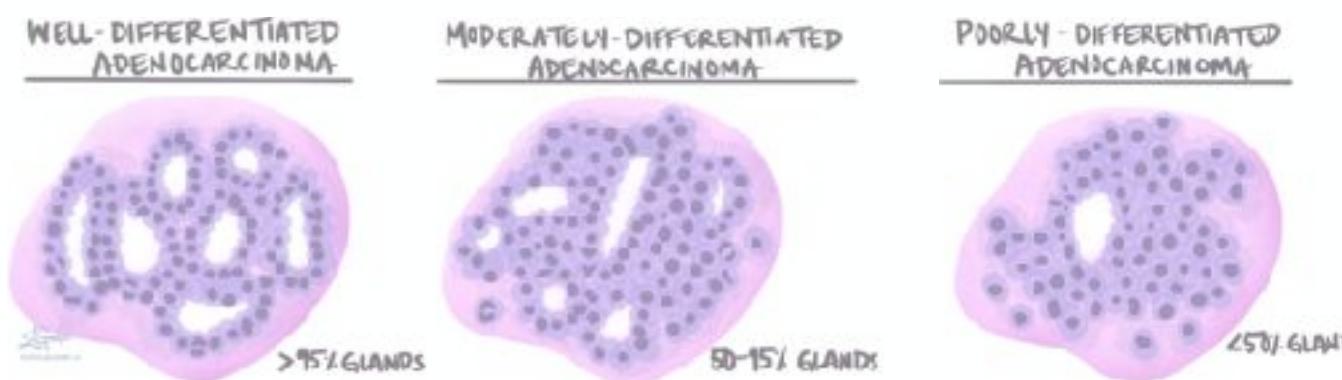
healthy reference mucosa (REF)
pre- tubulovillous adenoma (TVA)
malignant colorectal cancer (CRC)



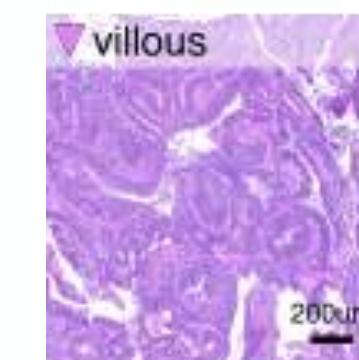
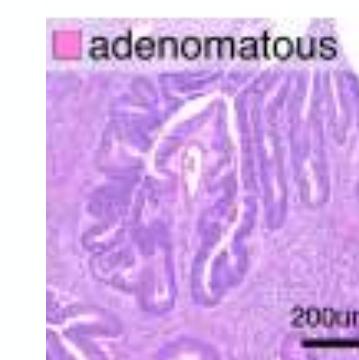
27 regions across sections
(1 REF, 0-3 TVA, 0-2 CRC each)



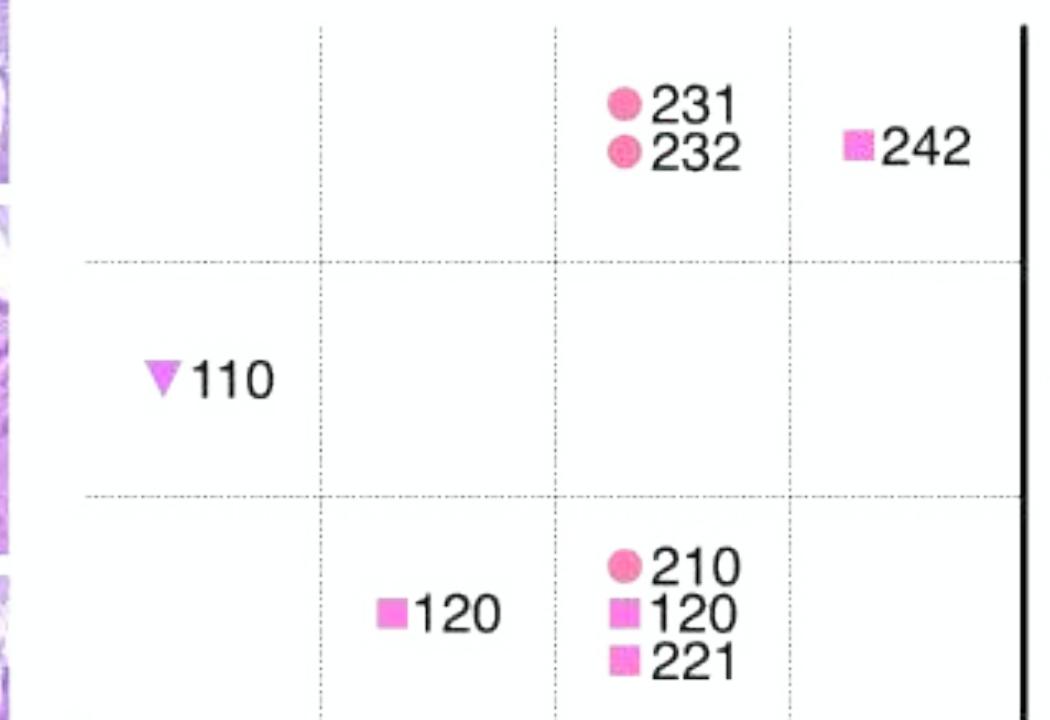
histopathology: CRC classification



grading



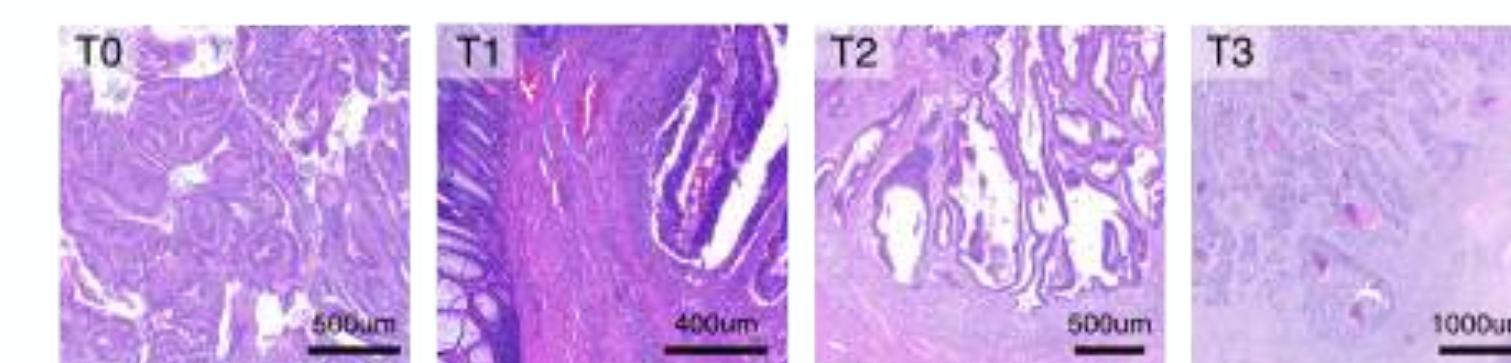
staging



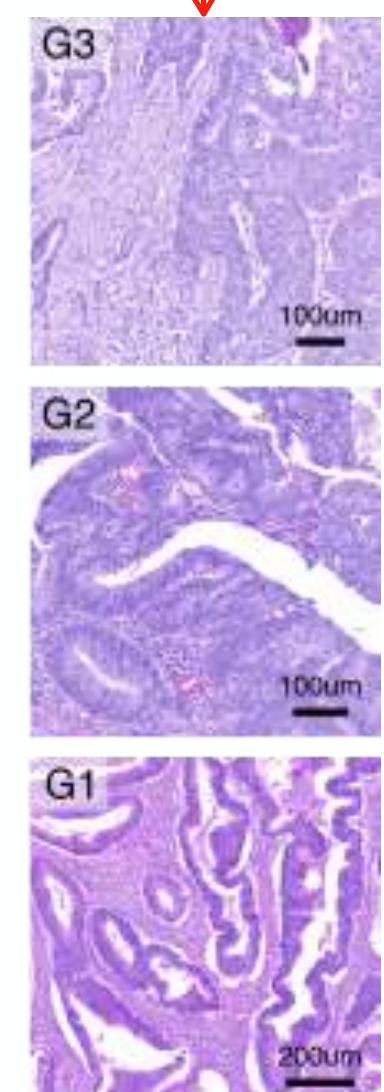
staging-invasion

features

- ◆ inflammation
- ◆ angiogenesis
- ◆ growth pattern
- ◆ lymphovascular invasion
- ◆ tertiary lymphoid structure
- ...

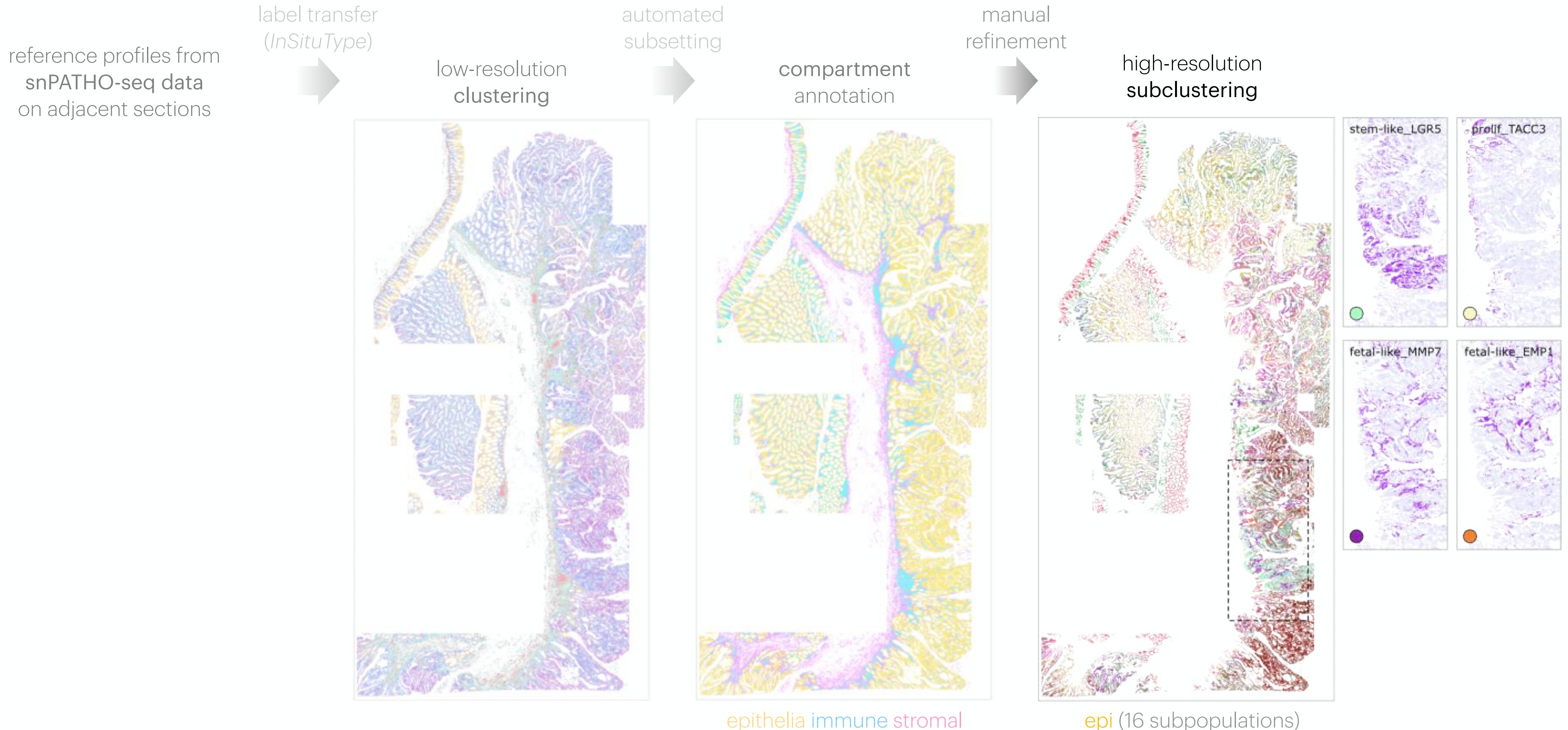


section	region	dysplasia/growth pattern	inflammation severity	cell type	grading/staging	fibrosis	angio-desmoplasia	angiogenesis	necrosis	LI	TIS
110	REF	low	3	P							
	TVA1	low	3	P							
	TVA2	low	3	M							
	TVA3	high, F	2	M							
	CRC	V	1	P	G2/T0	-	-	+	-	+	
120	REF		2	P							
	TVA	low	2	P							
	CRC1	A	4	P	G1/T1	+	-	+	-	-	
	CRC2	A	3	P	G1/T2	-	+	-	+	+	
210	REF		3	M							
	TVA	low, F	2	M							
	CRC	M	4	P	G1/T2	-	+	-	-	+	
221	REF		2	M							
	TVA1	low	3	P							
	TVA2	low	2	P							
	TVA3	low	3	P							
	CRC	A	4	P	G1,T2*	+	-	-	-	+	-
222	REF		4	P							
	TVA	low	3	P							
231	REF		4	M							
	TVA	low	4	P							
	CRC	M	4	P	G3/T2	+	+	-	+	+	
232	REF		3	M							
	CRC	M	4	P	G3/T2	+	-	-	+	+	
242	REF		3	P							
	TVA	high, F	4	P							
	CRC	A, N	3	P	G3/T3	+	+	+	+	+	-



27 regions
8 CRCs

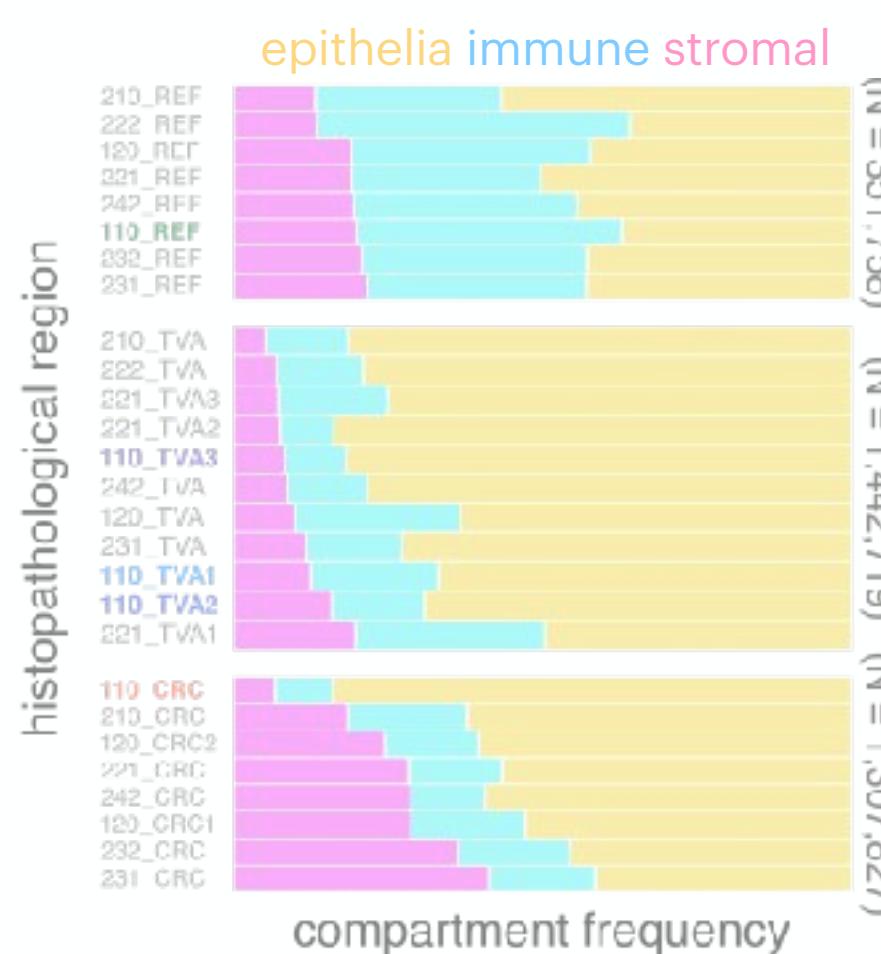
hierarchical clustering workflow



comparing regions: same data, different aes()

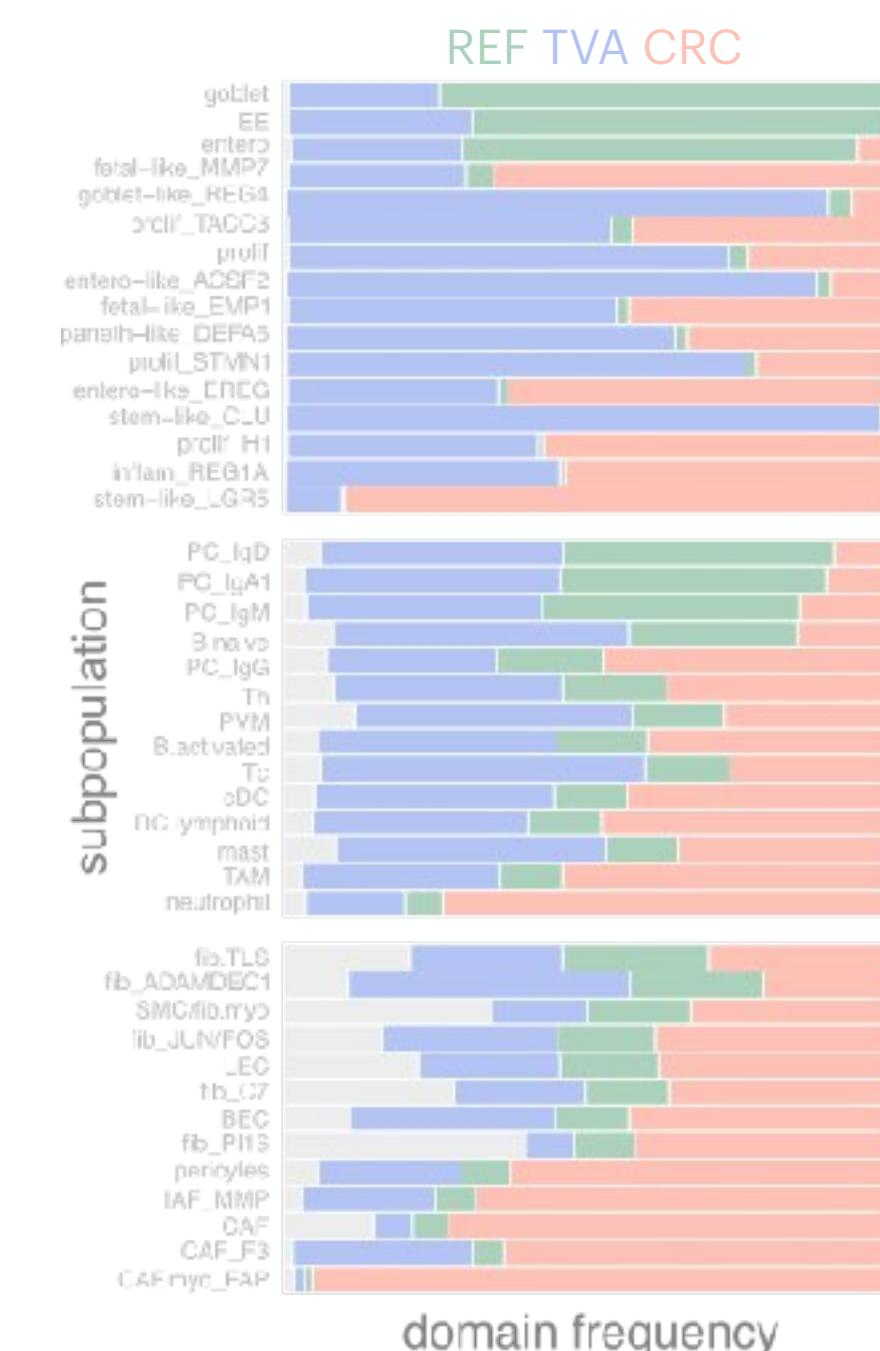
compartments
across regions

- ▶ along REF-TVA-CRC,
epi/str increase &
imm decrease



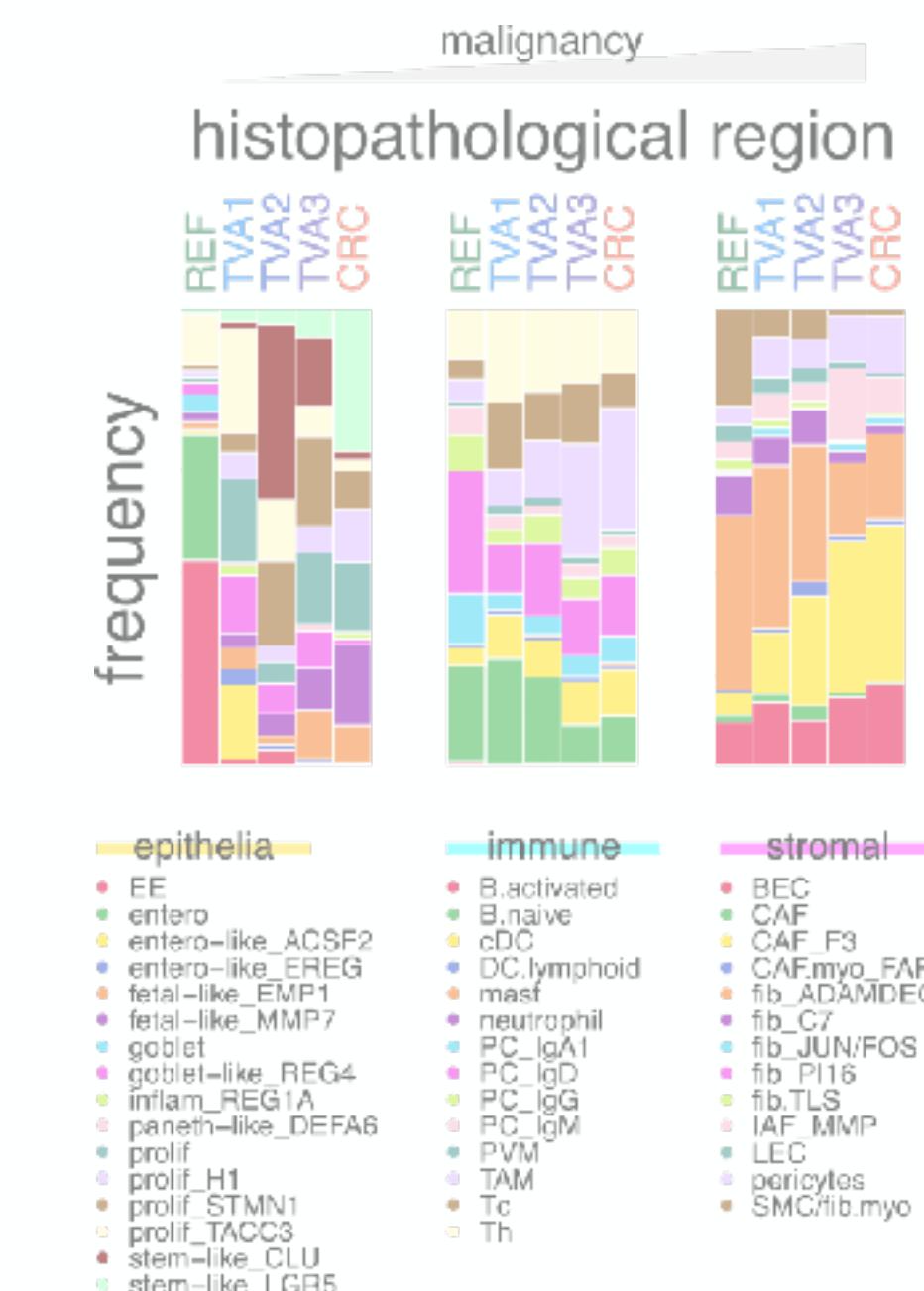
regions across subpopulations

some clusters
are more or less
domain-specific



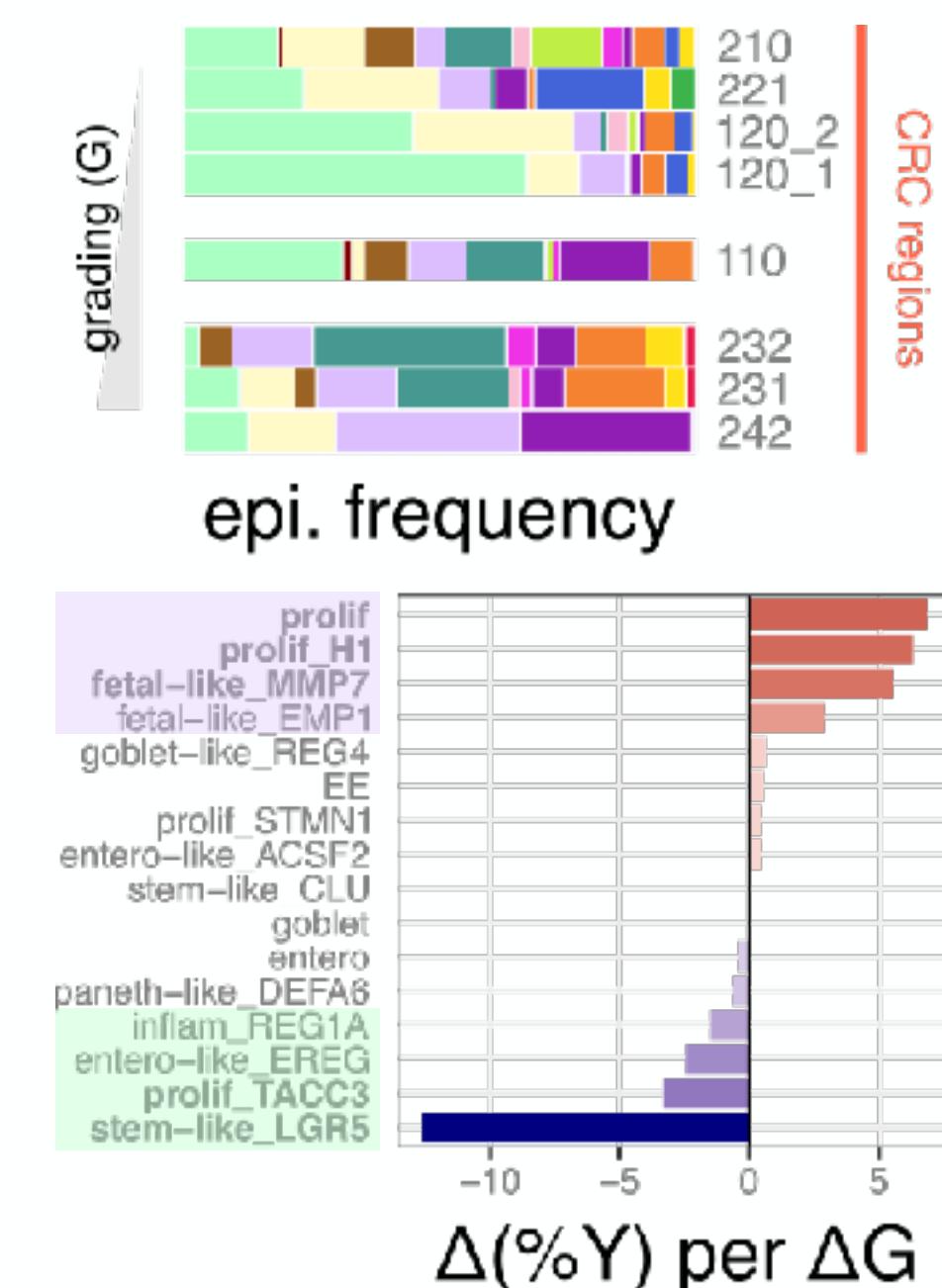
subpopulation
across regions

frequency shift
align with histo
pathology label

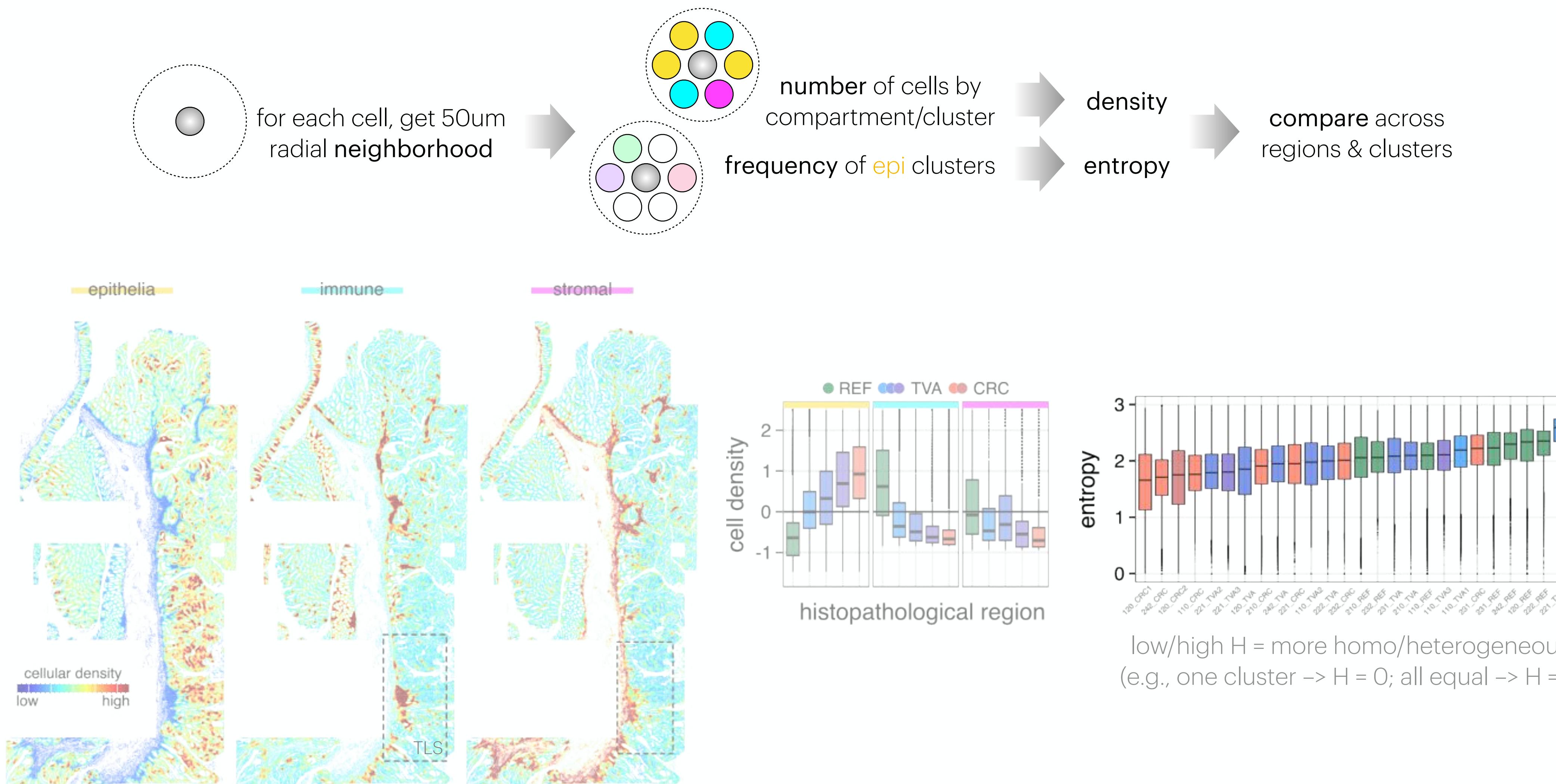


CRC across sections

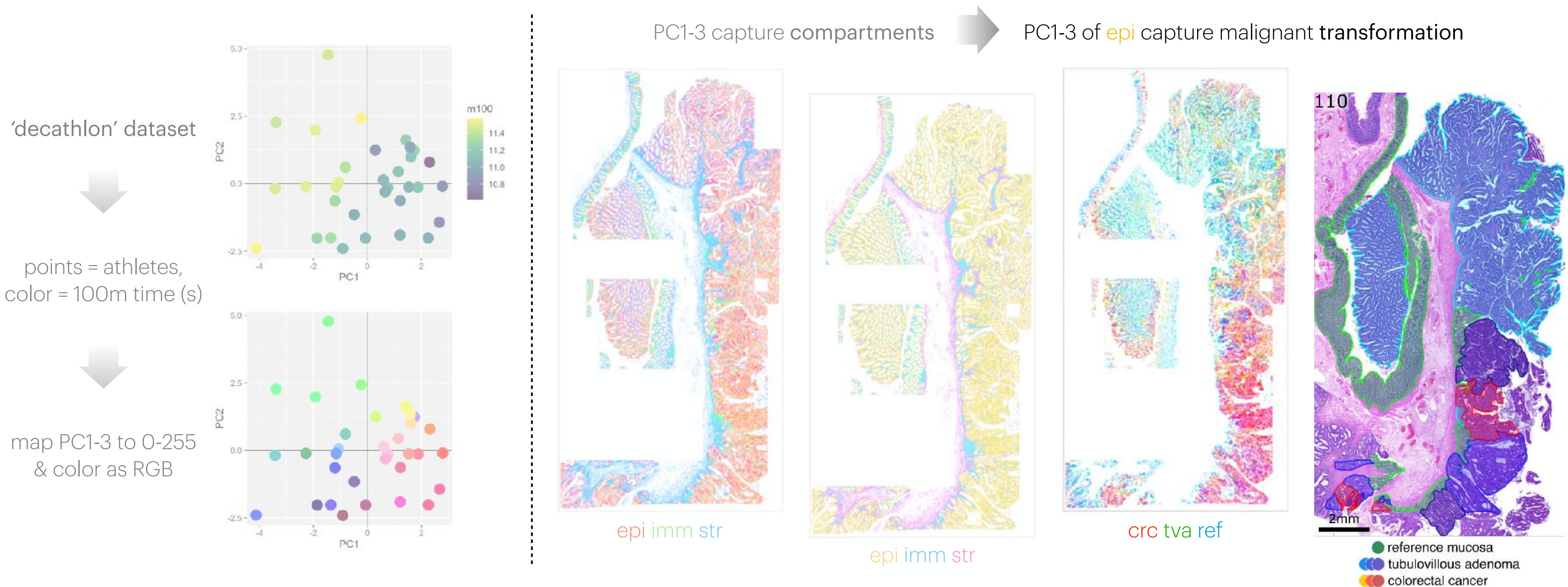
- ▶ **stem-/fetal-like endpoints align with malignancy**



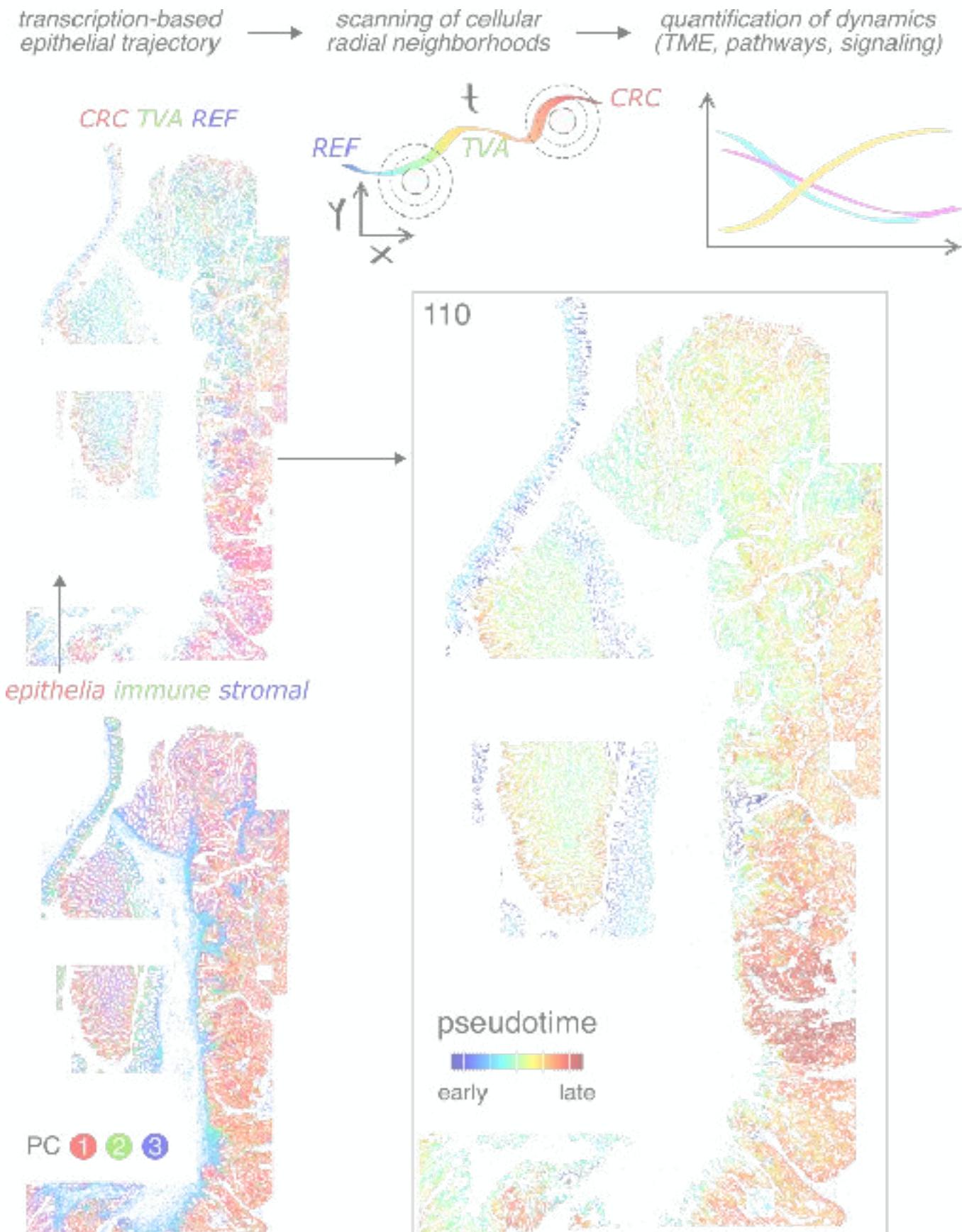
comparing neighborhoods: epi become dense & homogeneous



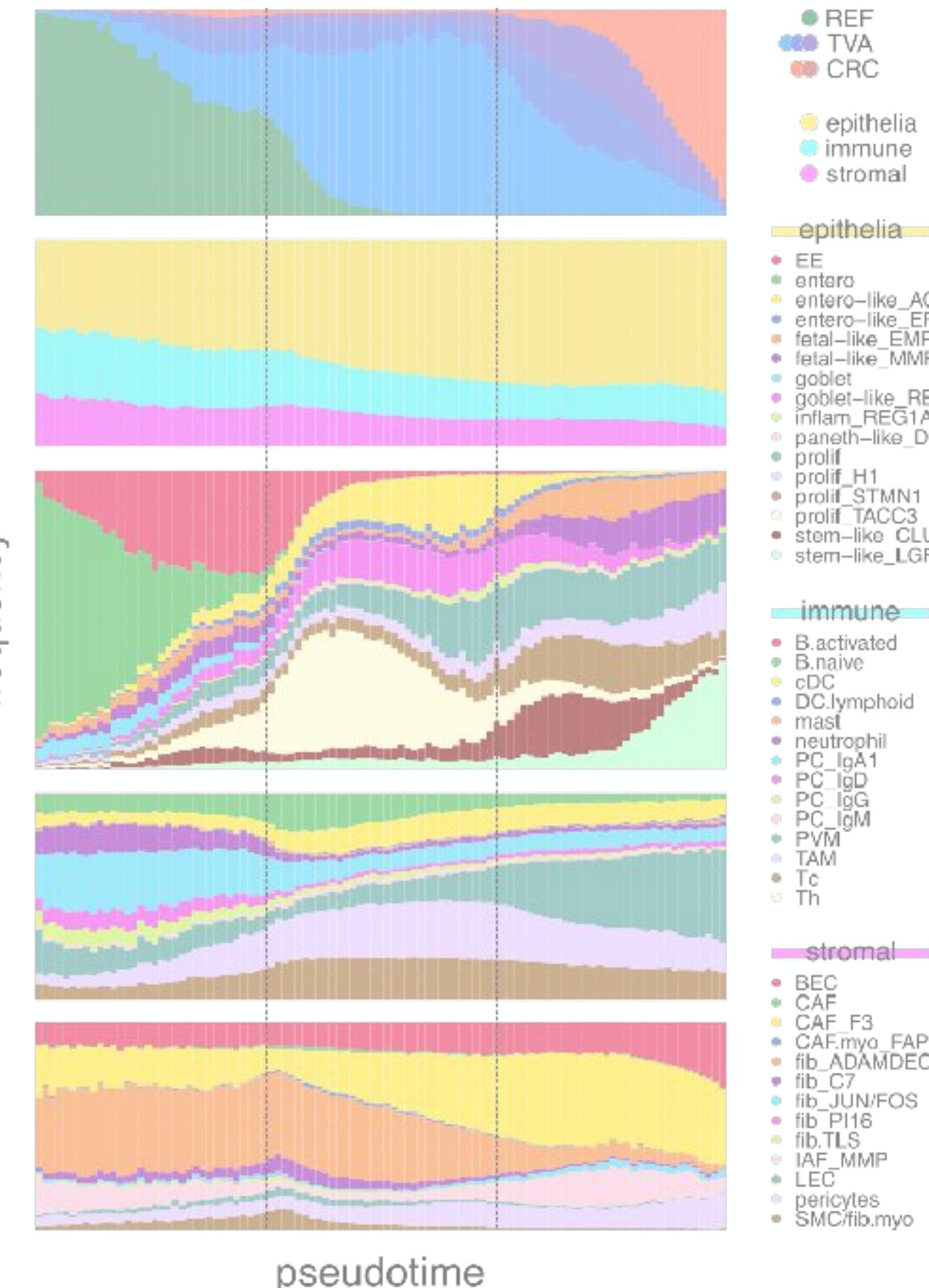
trajectory of malignant transformation through space



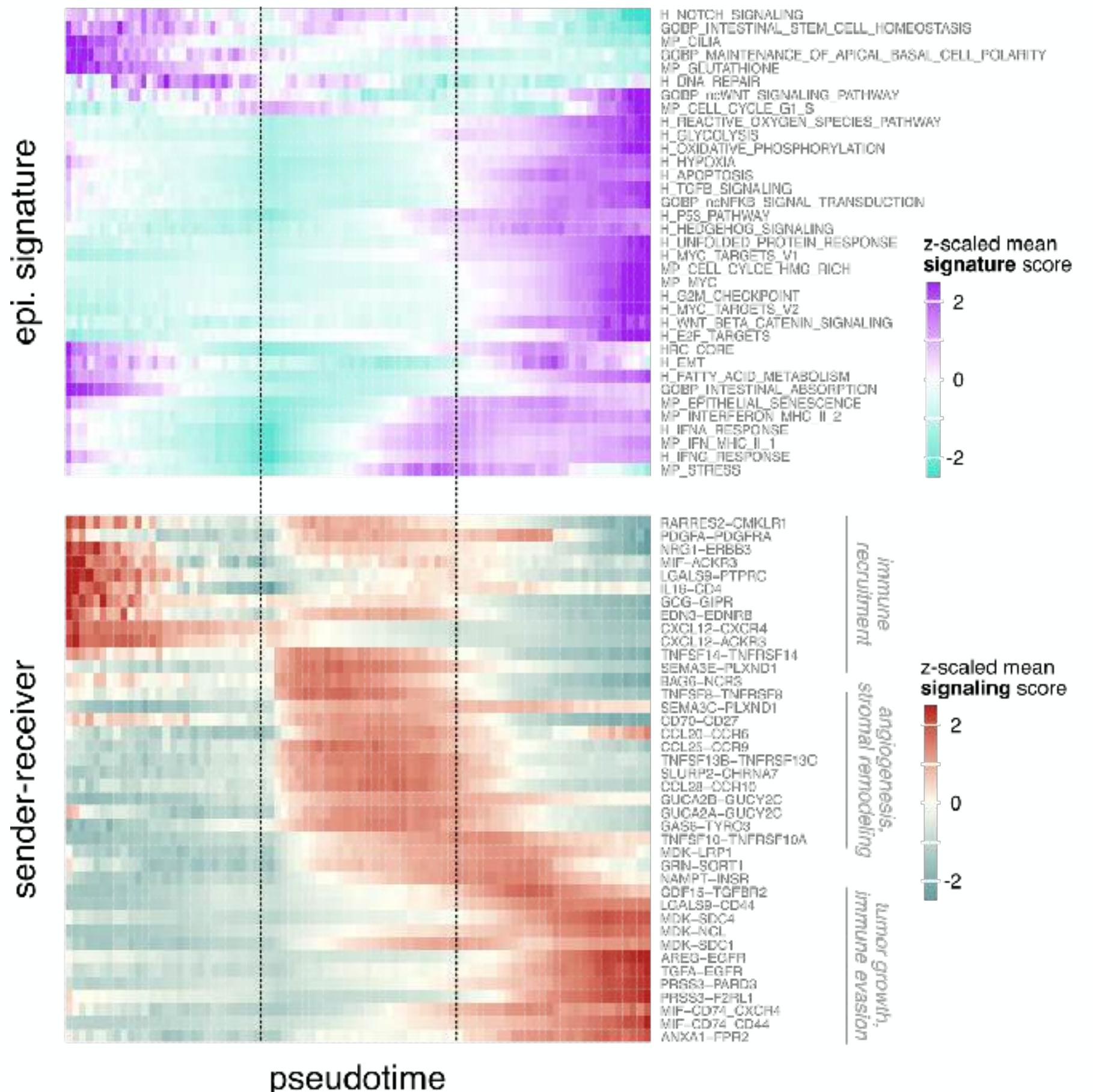
dynamics of malignant transformation through space



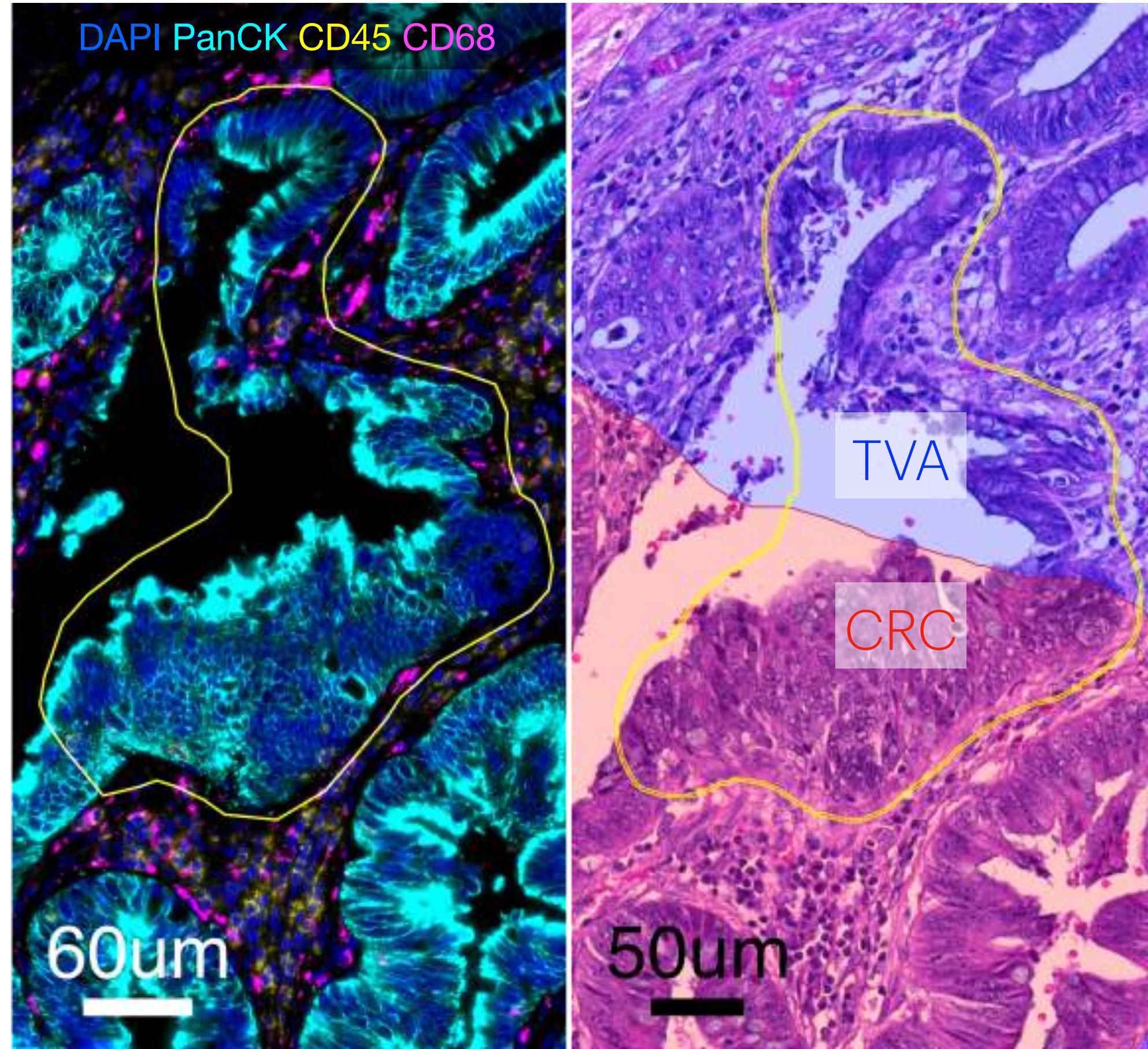
→ how do neighbors change?
dynamics of composition



→ how does function change?
dynamics of pathways & signaling



transition crypts: zooming in on transformation



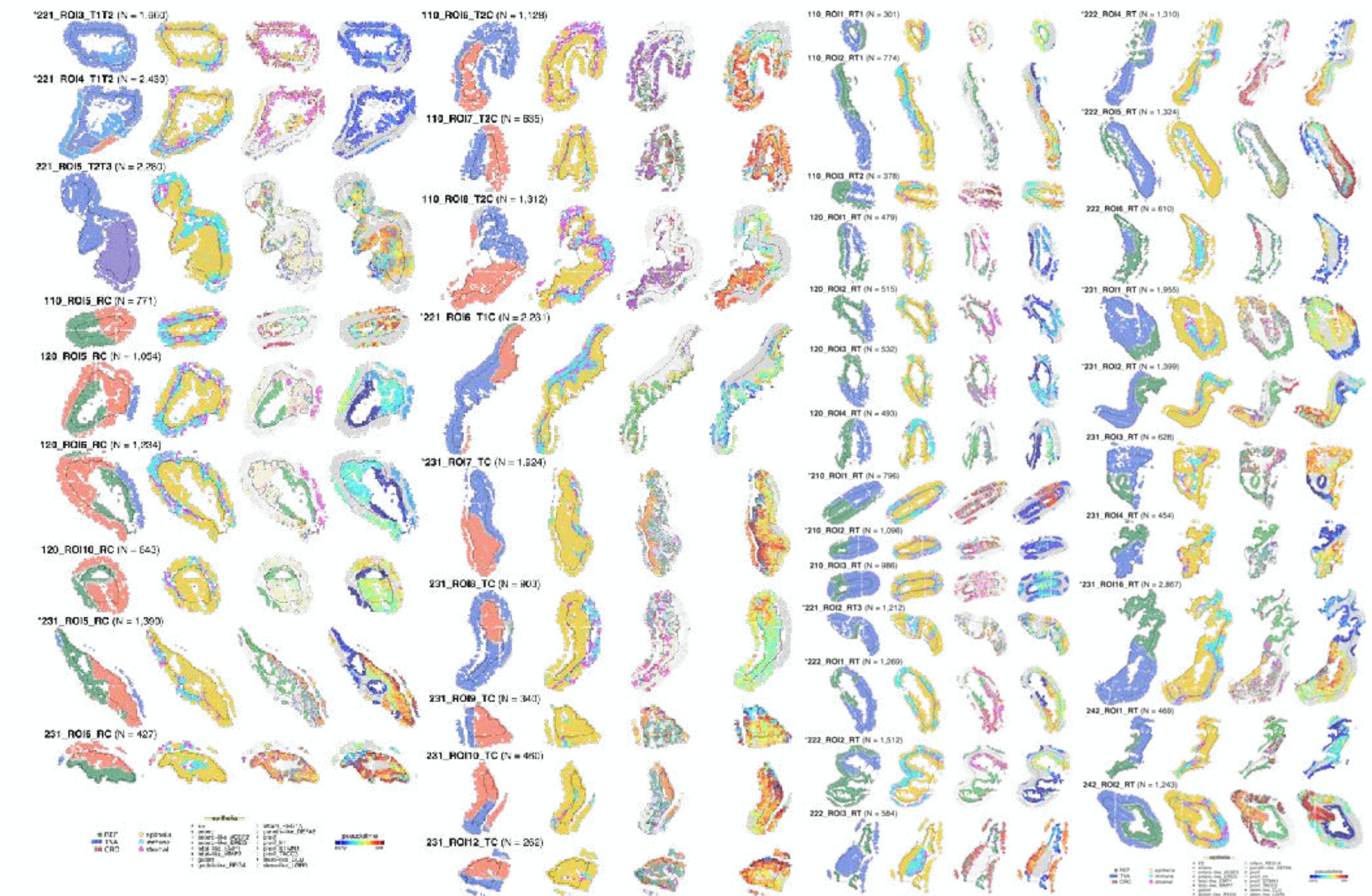
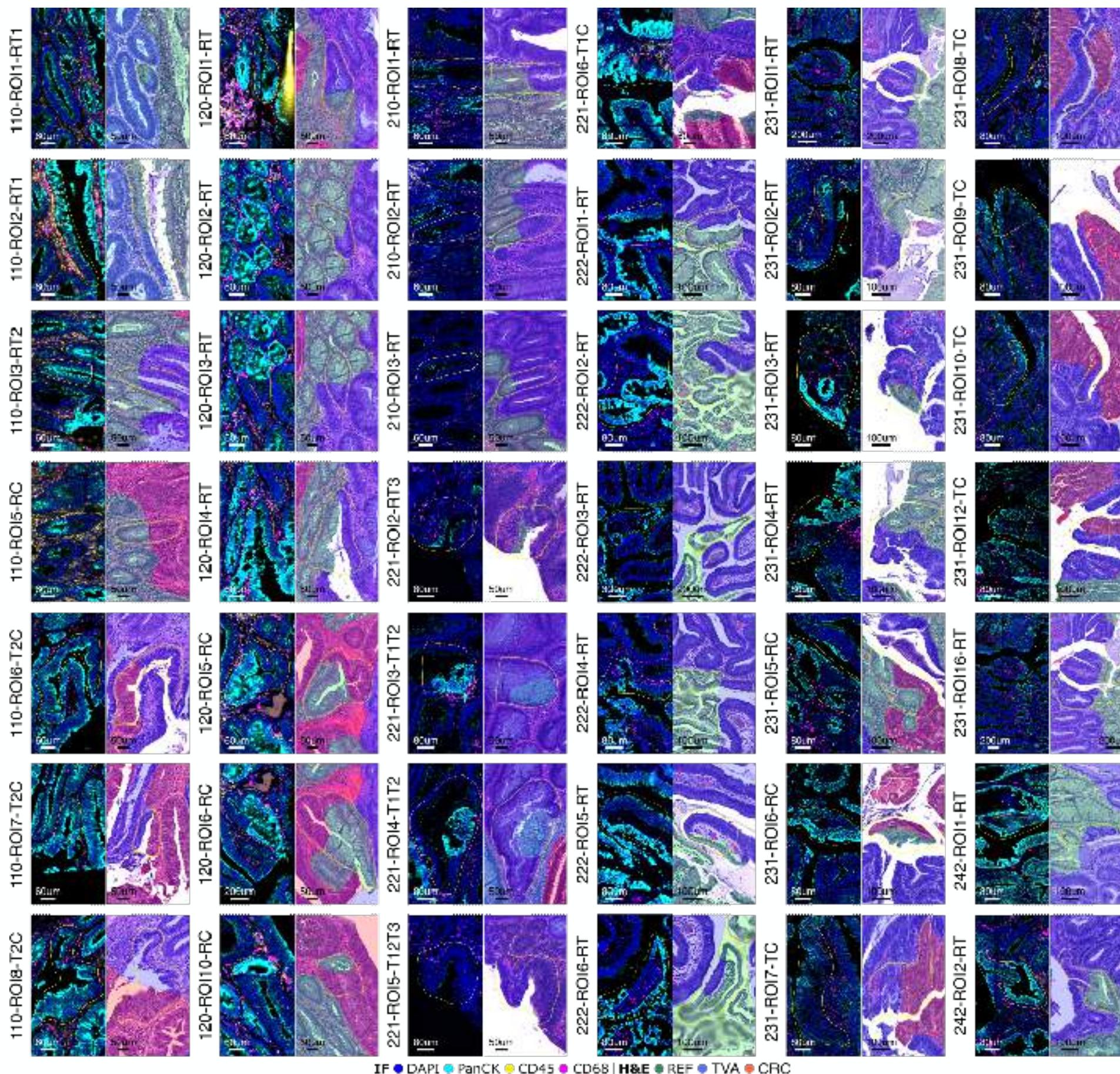
- ◆ single colonic crypt w/
divided histopathological
& transcriptional makeup
- ◆ from rare crypt **fusion**, or an
abrupt **transformation** event

transition crypts: zooming in on transformation

42 transition crypts

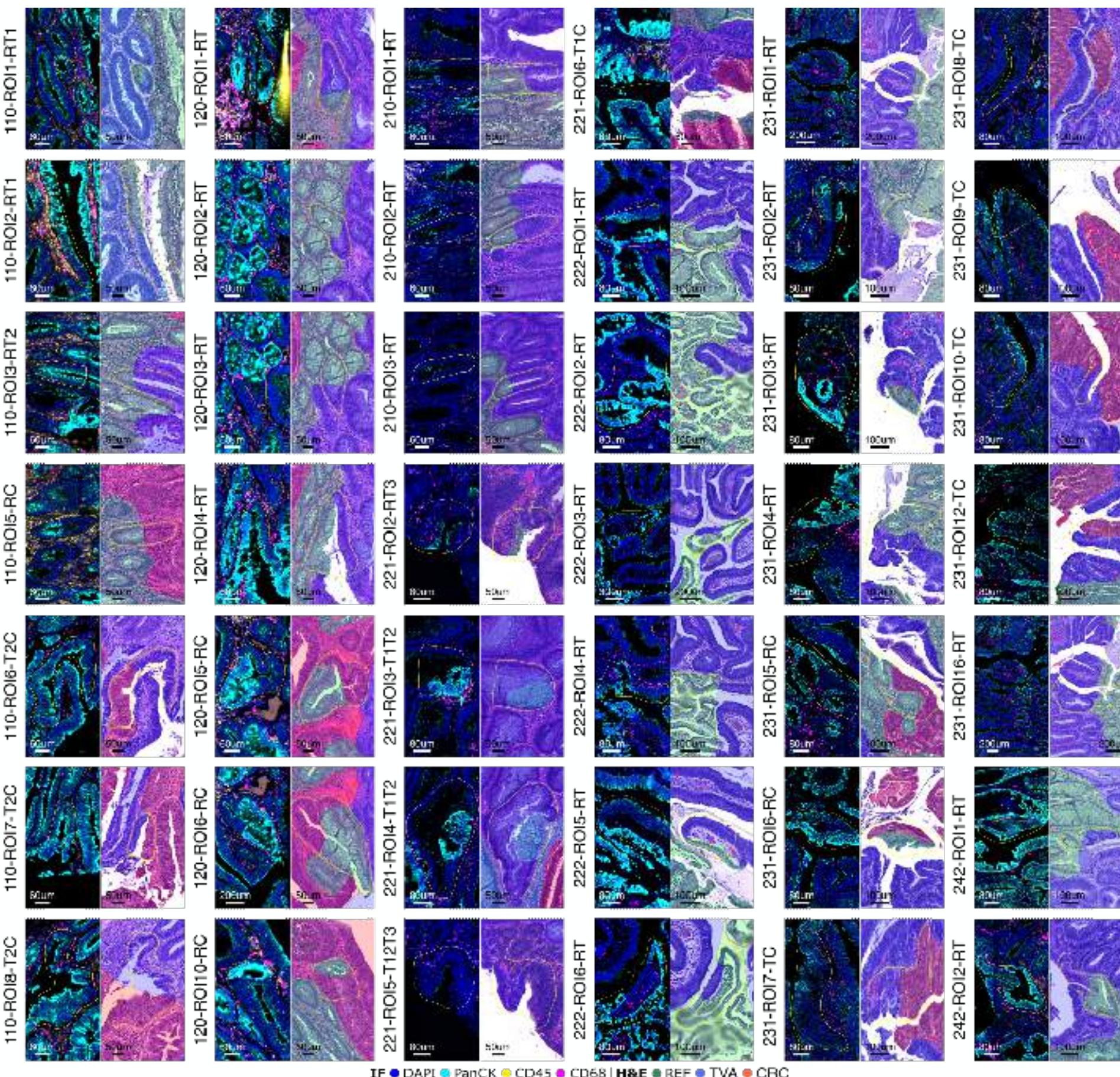
(24 RT, 6 RC, 9 TC, 3 TT)

selection + 50um expansion colored by (fltr)
region, compartment, subpopulation, pseudotime



transition crypts: zooming in on transformation

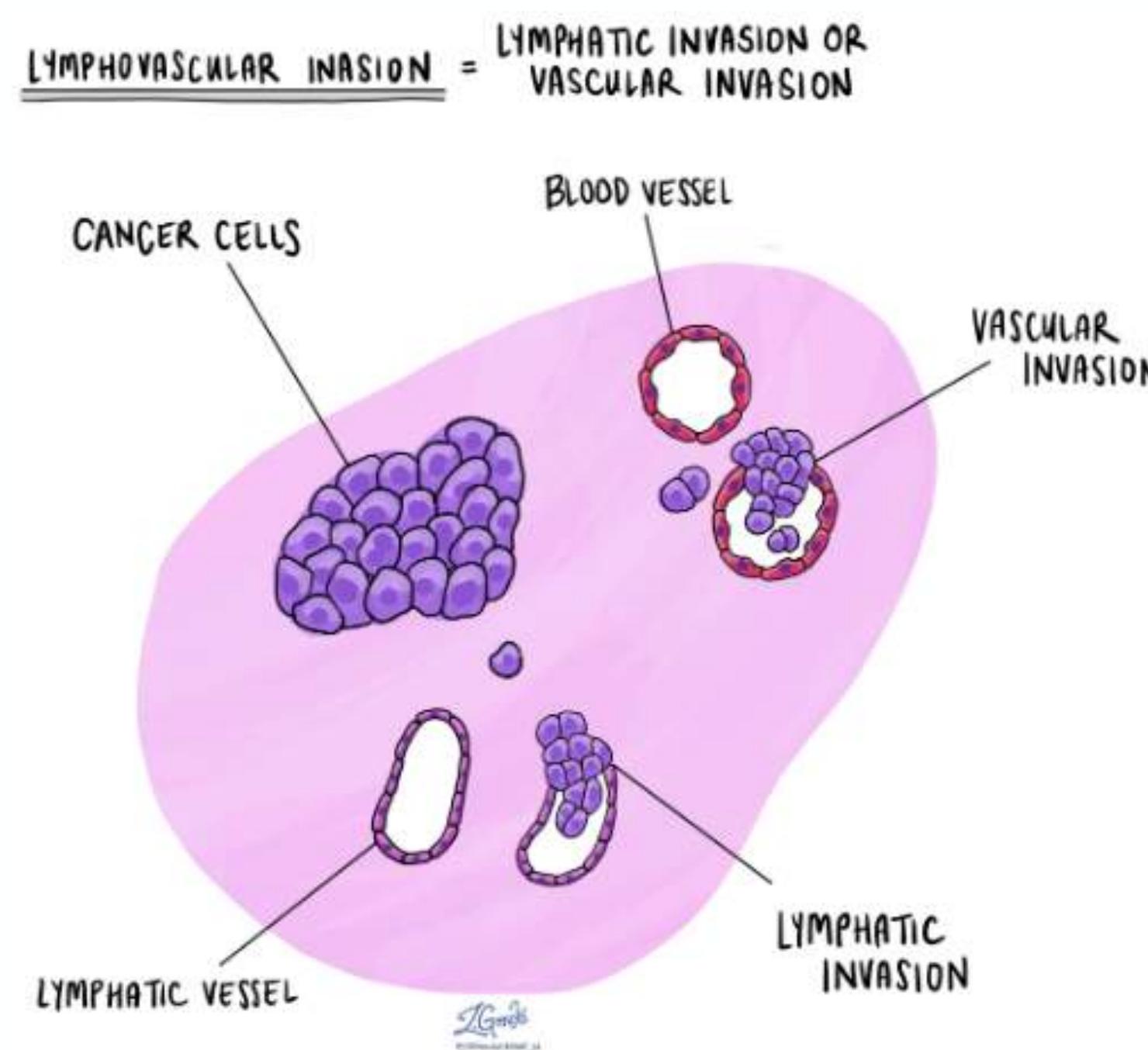
42 transition crypts
(24 RT, 6 RC, 9 TC, 3 TT)



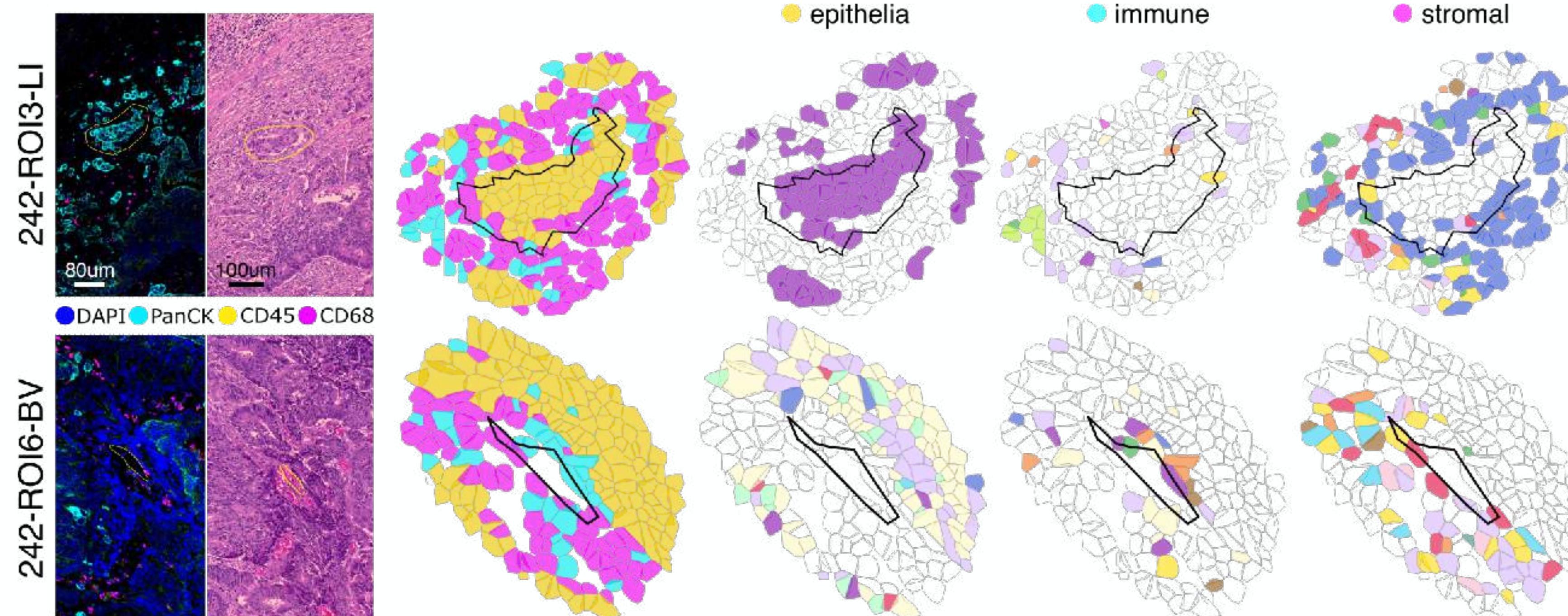
selection + 50um expansion colored by (fltr)
region, compartment, subpopulation, pseudotime



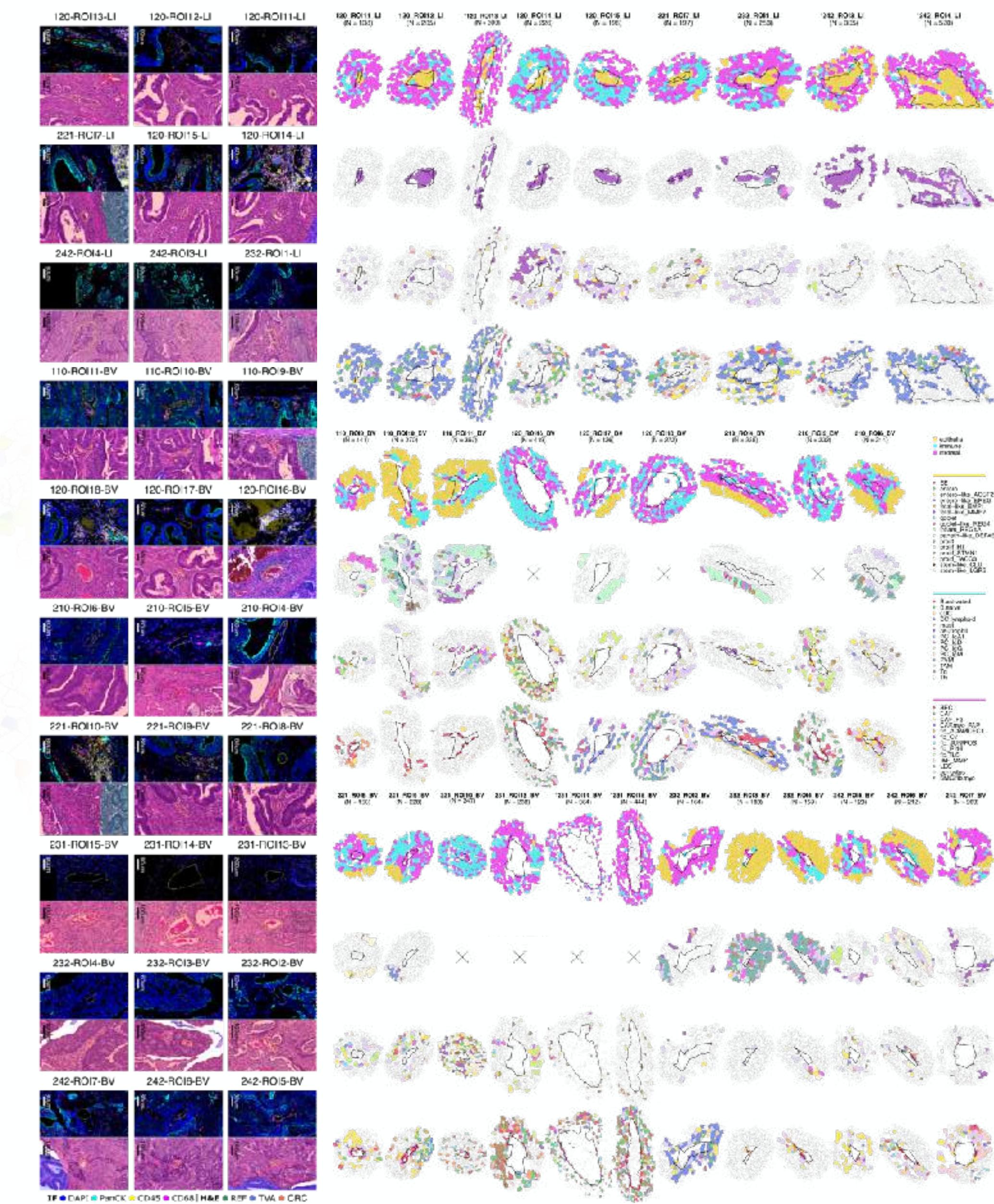
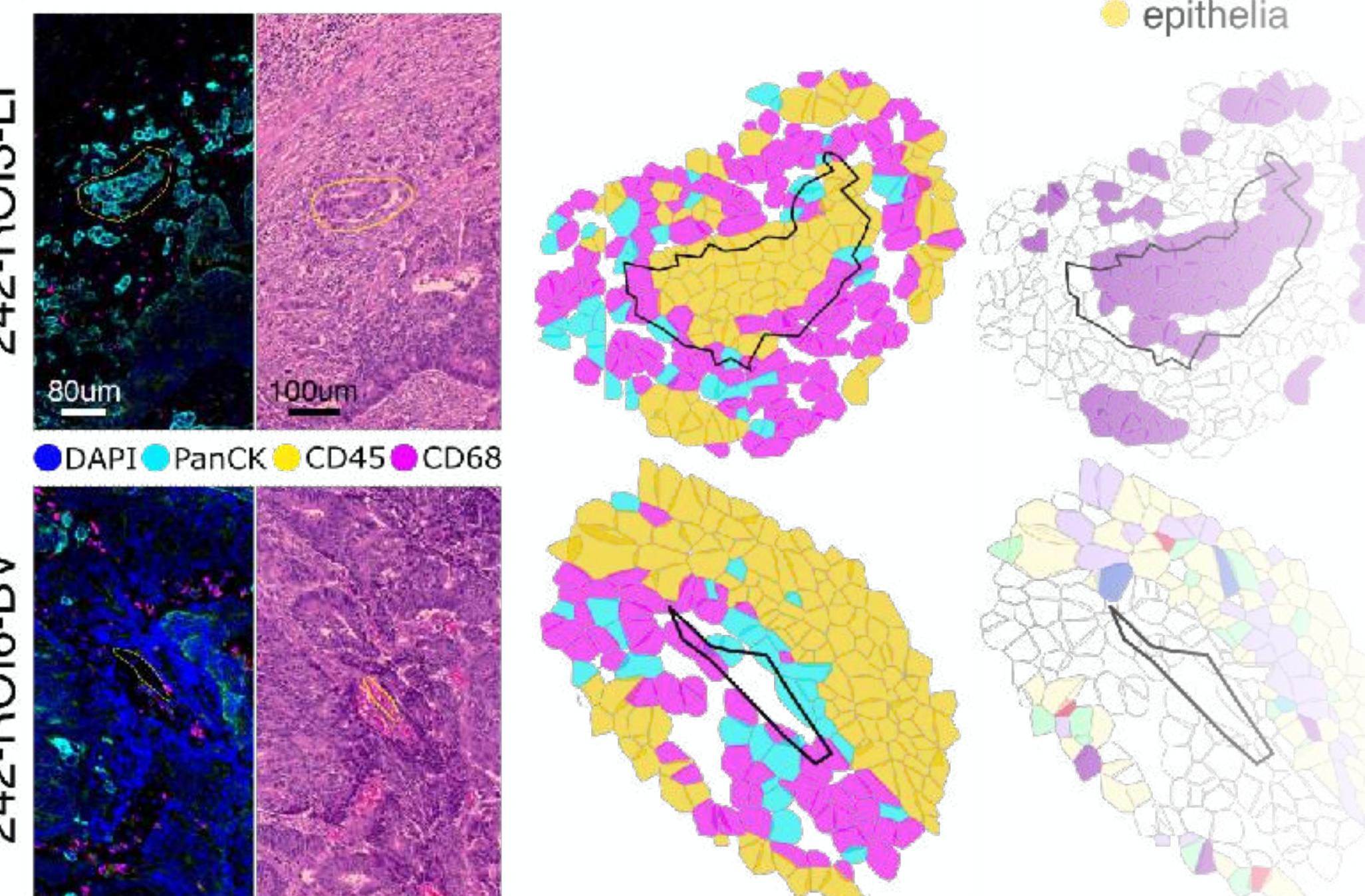
lymphovascular invasions: from growth to metastatic spread



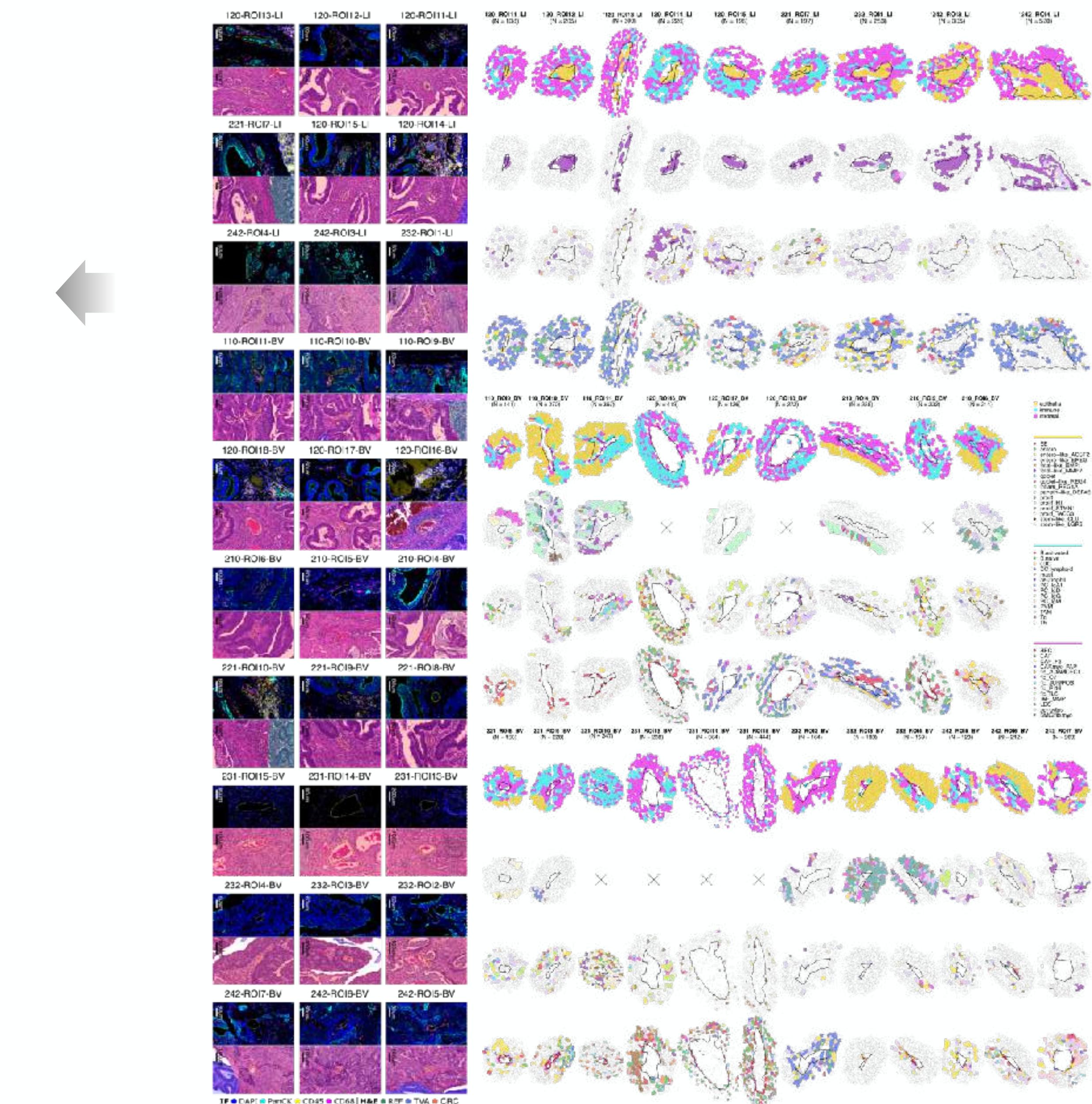
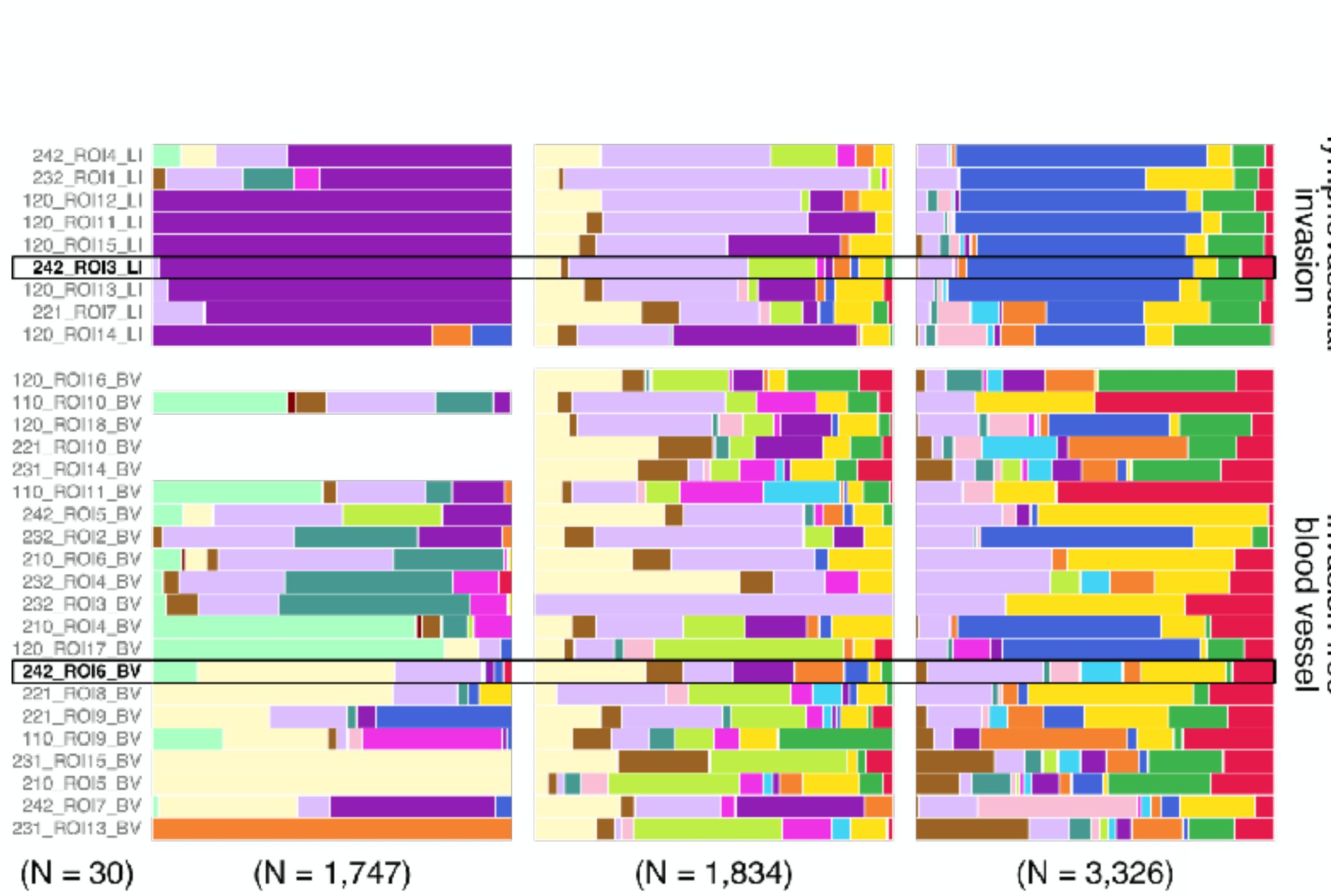
lymphovascular invasions: from growth to metastatic spread



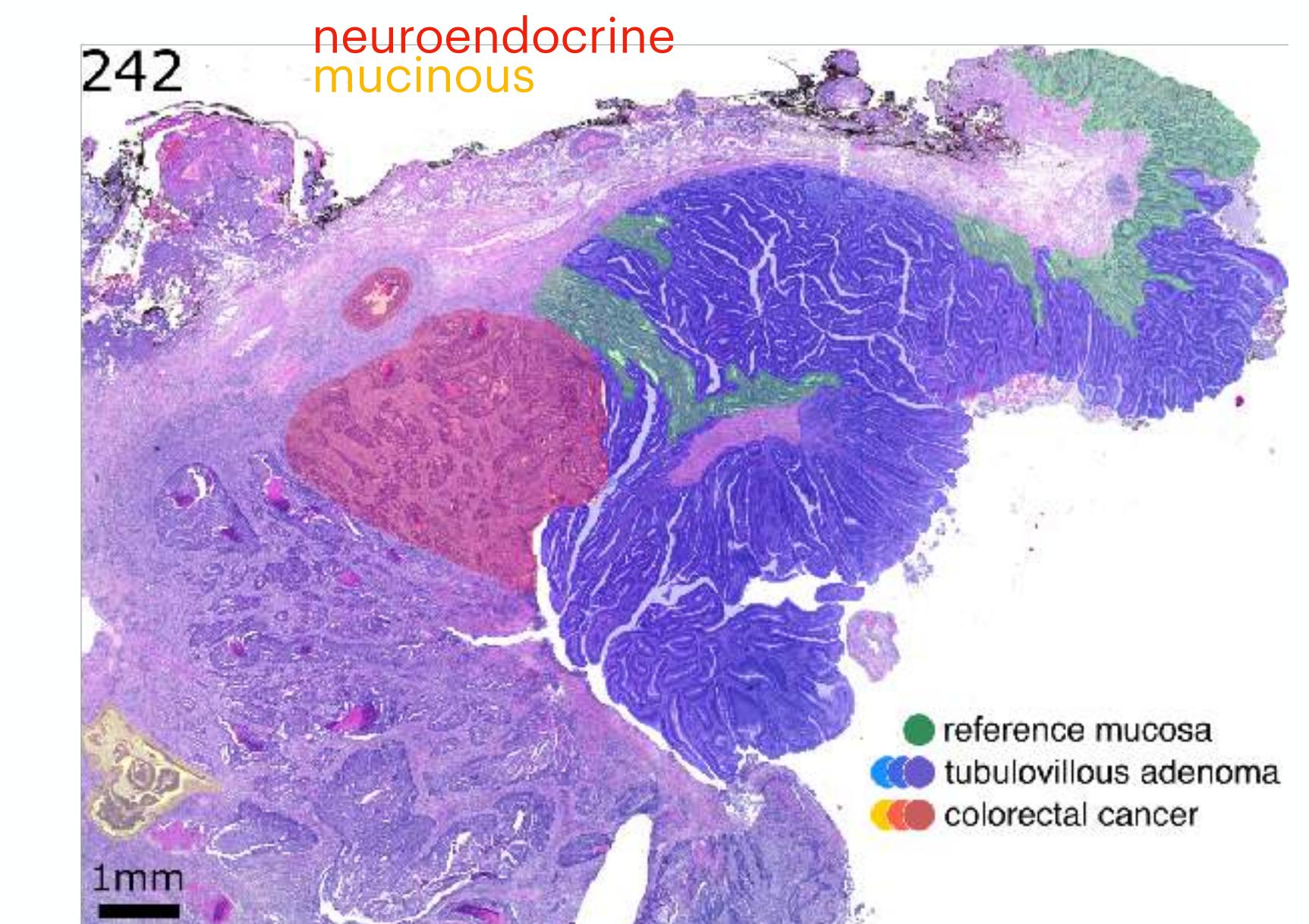
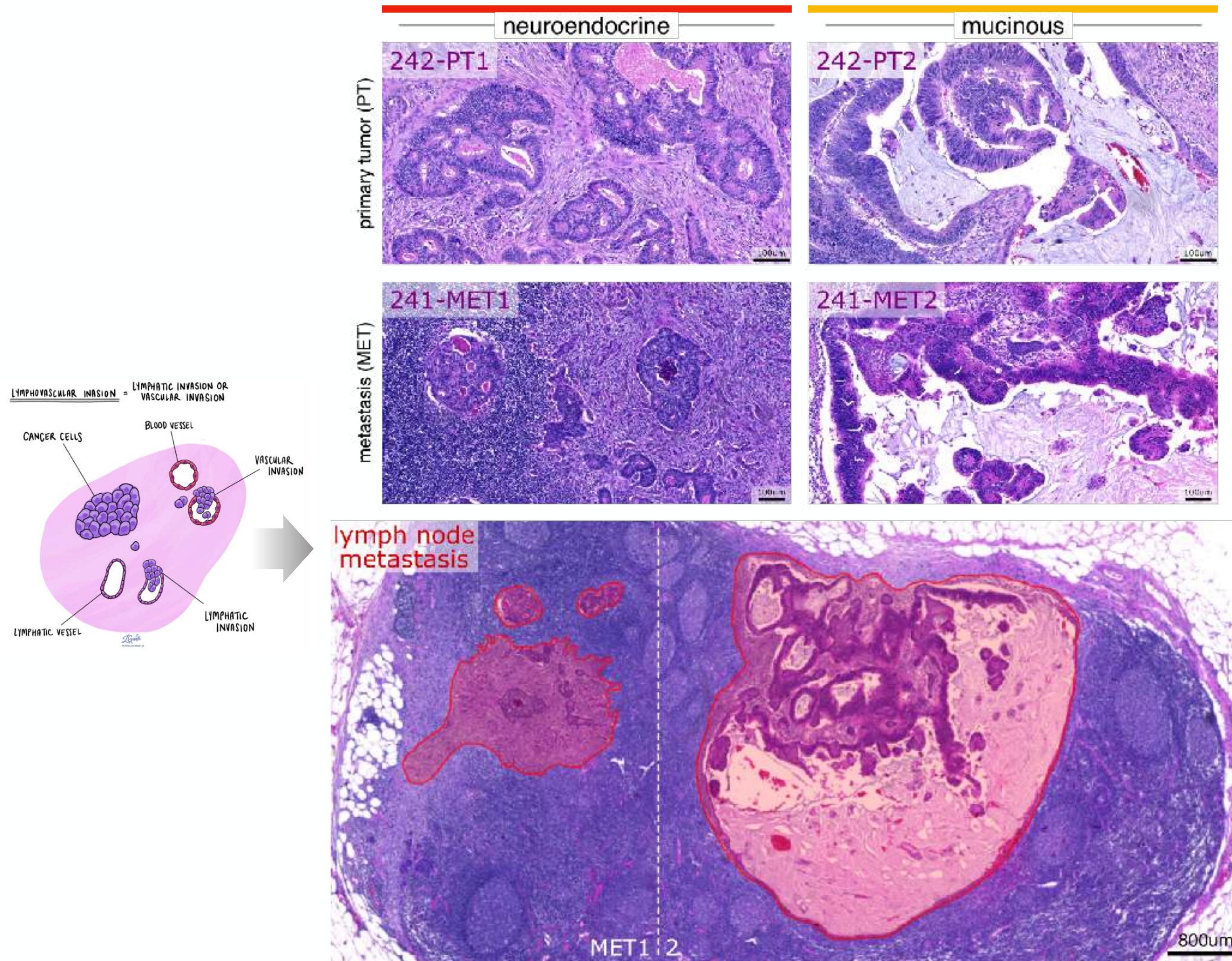
Lymphovascular invasions: from growth to metastatic spread



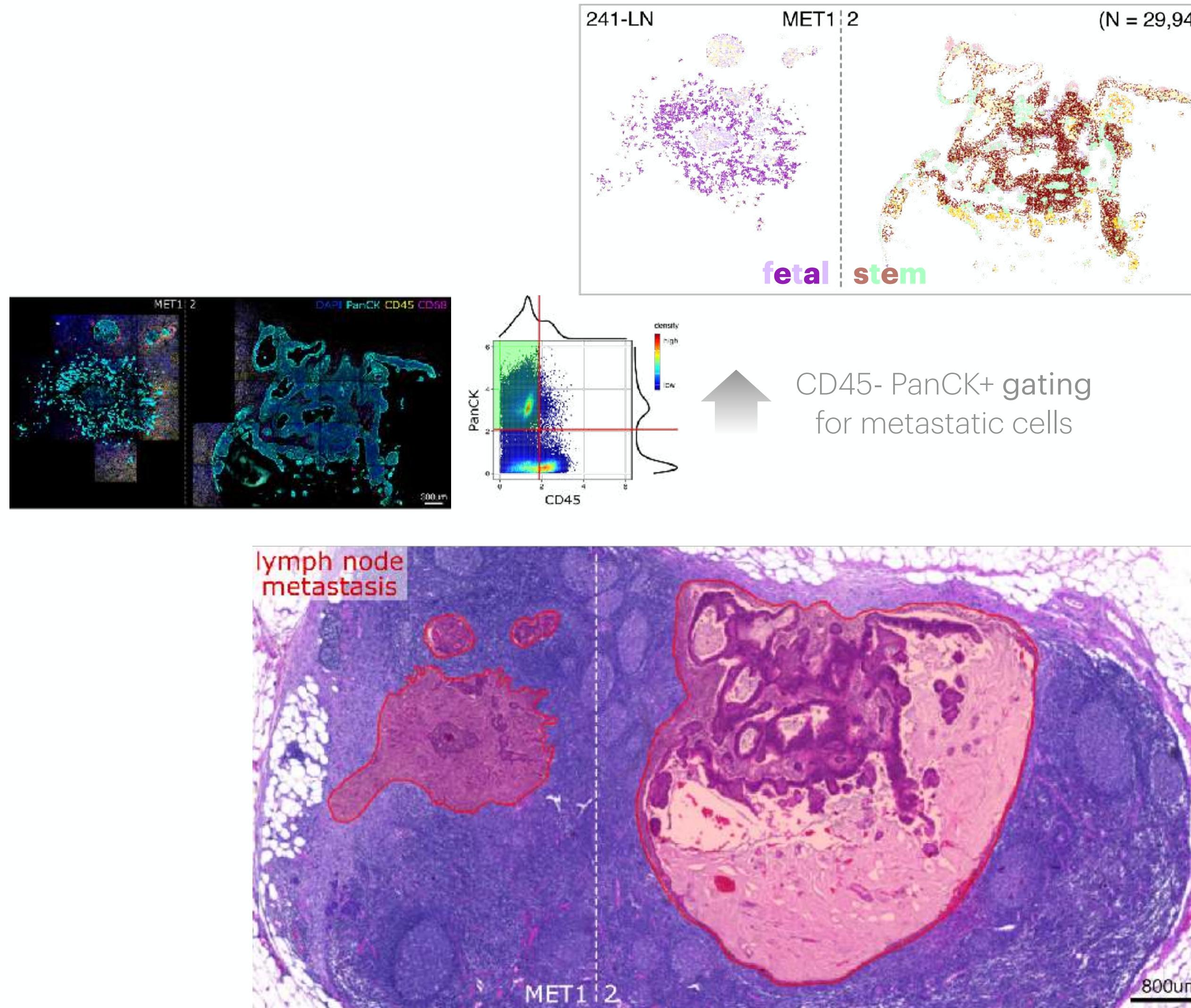
lymphovascular invasions: from growth to metastatic spread



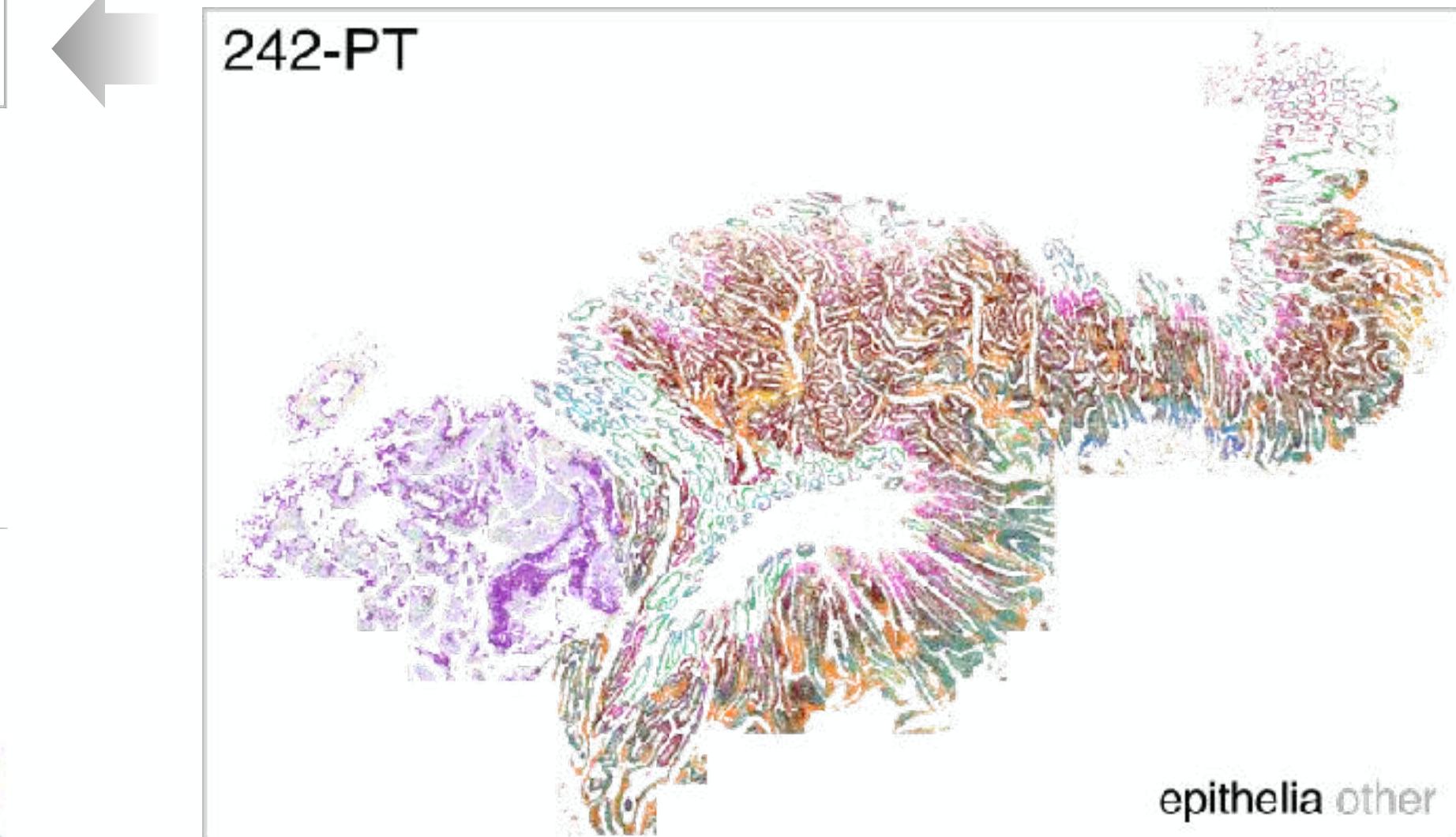
one section, two lymph node metastases



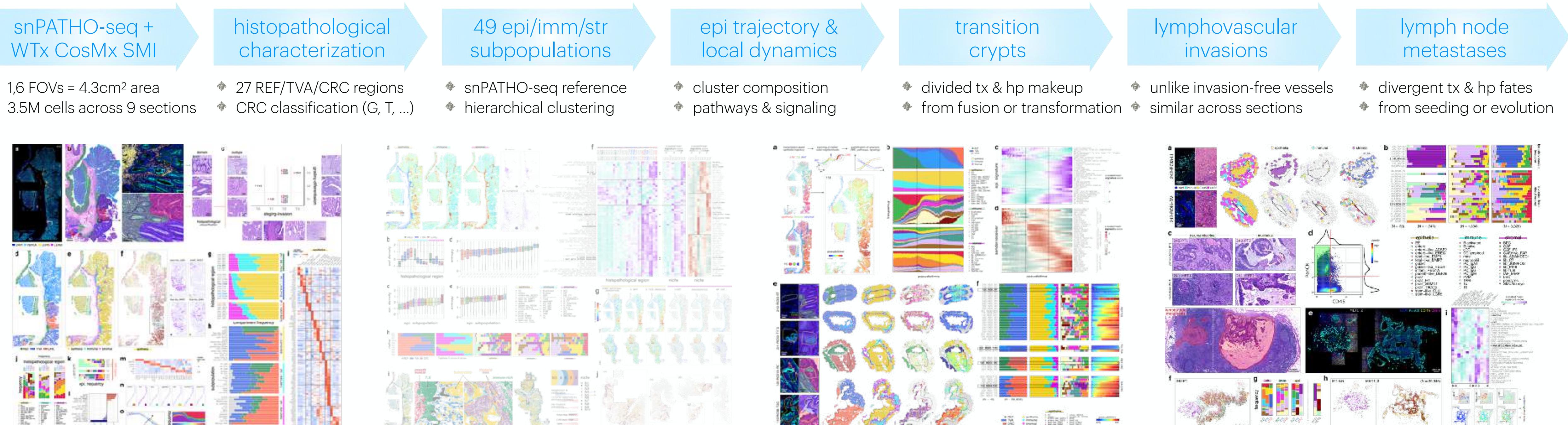
one section, two lymph node metastases



label transfer
using primary
reference profiles



summary (& my take home message)



...there are a *lot* of
comp **methods**

...**simple** questions
want **simple** answers

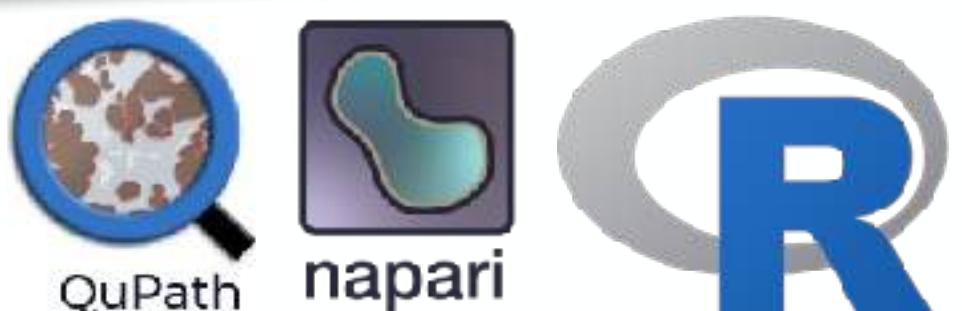
...knowing your way around
the **basics** goes a long way

data & code availability

Data availability

All data have been deposited on Zenodo at the following DOIs. H&E stains are available at [10.5281/zenodo.15548055](https://doi.org/10.5281/zenodo.15548055). Flat files of the CosMx SMI data (FOV placement, count matrices, cell metadata, segmentation boundaries) are available at [10.5281/zenodo.15550908](https://doi.org/10.5281/zenodo.15550908). Napari inputs (Zarr stores of IF markers, FOV and segmentation labels, single-molecule RNA targets), including histopathological domain and ROI annotations, are available at [10.5281/zenodo.15556499](https://doi.org/10.5281/zenodo.15556499) (11), [10.5281/zenodo.15556609](https://doi.org/10.5281/zenodo.15556609) (12), [10.5281/zenodo.15556207](https://doi.org/10.5281/zenodo.15556207) (21), [10.5281/zenodo.15555405](https://doi.org/10.5281/zenodo.15555405) (22), [10.5281/zenodo.15552301](https://doi.org/10.5281/zenodo.15552301) (23), and [10.5281/zenodo.15551924](https://doi.org/10.5281/zenodo.15551924) (24); one file batch per slide (due to file size limitations, some files for sections 11 and 12 are placed at [10.5281/zenodo.15585754](https://doi.org/10.5281/zenodo.15585754)).

We also make available *SingleCellExperiment* (.rds) and *AnnData* objects (.h5ad; written using *zellkonverter*^[67]), as well as language-agnostic *alabaster* file artifacts (written using *alabaster.sce*^[68]) that, for each section, gather processing and analysis results for the CosMx SMI data; these are available at [10.5281/zenodo.15574384](https://doi.org/10.5281/zenodo.15574384). Also included are (un)filtered *cellranger* outputs for the snPATHO-seq data, as well as the low- and high-resolution reference profiles derived thereof, which were used for clustering of the CosMx SMI data.



data & code availability

Data availability

All data have been deposited on Zenodo at the following DOIs. H&E stains are available at [10.5281/zenodo.15548055](https://doi.org/10.5281/zenodo.15548055). Flat files of the CosMx SMI data (FOV placement, count matrices, cell metadata, segmentation boundaries) are available at [10.5281/zenodo.15550908](https://doi.org/10.5281/zenodo.15550908). Napari inputs (Zarr stores of IF markers, FOV and segmentation labels, single-molecule RNA targets), including histopathological domain and ROI annotations, are available at [10.5281/zenodo.15556499](https://doi.org/10.5281/zenodo.15556499) (11), [10.5281/zenodo.15556609](https://doi.org/10.5281/zenodo.15556609) (12), [10.5281/zenodo.15556207](https://doi.org/10.5281/zenodo.15556207) (21), [10.5281/zenodo.15555405](https://doi.org/10.5281/zenodo.15555405) (22), [10.5281/zenodo.15552301](https://doi.org/10.5281/zenodo.15552301) (23), and [10.5281/zenodo.15551924](https://doi.org/10.5281/zenodo.15551924) (24); one file batch per slide (due to file size limitations, some files for sections 11 and 12 are placed at [10.5281/zenodo.15585754](https://doi.org/10.5281/zenodo.15585754)).

We also make available *SingleCellExperiment* (.rds) and *AnnData* objects (.h5ad; written using *zellkonverter*^[67]), as well as language-agnostic *alabaster* file artifacts (written using *alabaster.sce*^[68]) that, for each section, gather processing and analysis results for the CosMx SMI data; these are available at [10.5281/zenodo.15574384](https://doi.org/10.5281/zenodo.15574384). Also included are (un)filtered *cellranger* outputs for the snPATHO-seq data, as well as the low- and high-resolution reference profiles derived thereof, which were used for clustering of the CosMx SMI data.

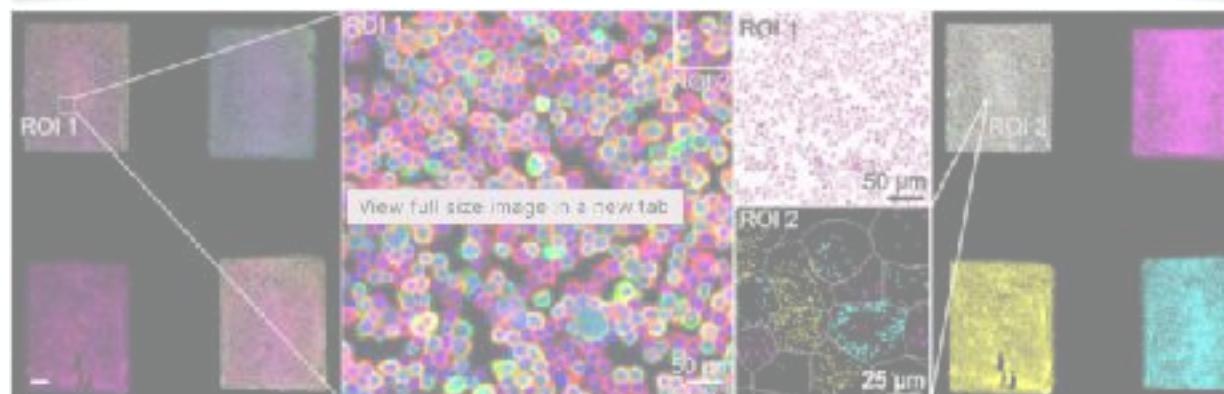


SnakeMake

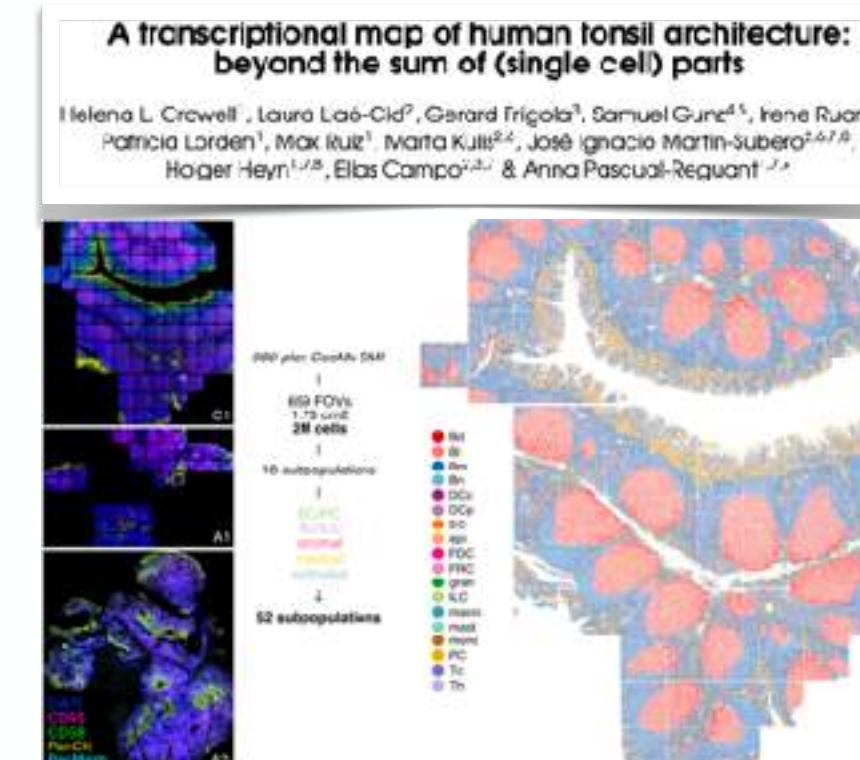
HelenaLC add link to preprint	d75d0cd · 3 days ago	169 Commits
code 100+ scripts	rmv unused scripts	2 weeks ago
data flat files	add empty dirs	2 weeks ago
imgs Napari	add images	last month
logs console capture	add logs	6 months ago
meta signatures, annotations etc.	add CARD8 arcodes	last month
outs intermediate results	add outs	6 months ago
plts 200+ multi-plot PDFs	add plts	6 months ago
.Rprofile	update .Rprofile	5 months ago
.gitignore	ignore rds & gz	2 weeks ago
README.md	add link to preprint	3 days ago
Snakefile	typo fix	2 weeks ago
inf.txt session information	add session info output	2 weeks ago
clustering		
• 03-ist.R :	InSituType clustering supervised by reference profiles extracted from (pooled) snPATHO-seq data on adjacent sections	
• 04-sub.R :	subset cells into epi(theta), imm(unnel), (str)mal	
• 05-jst.R :	InSituType subclustering (analogous to above), using distinct references profiles for each subset	
• 06-lab.R :	used to relabel 03-ist.R outputs (automated; using meta/lab/lv1.json), and 05-jst.R outputs (manual; using meta/lab/lv2,₁.json)	
PREPRINT		
steps		
data setup		
• this repository is file size-limit; any data beyond	03-raw.R	
• let <slide> denote a slide identifier (run 1: 11,12; run 2: 13,14)	read flat files as <i>SingleCellExperiment</i>	
◦ Zarr stores of IF stains should be at imgs/	◦ slash non-RNA targets as .altExps/	
◦ corresponding flat files should be at data/	◦ write out as .h5-backed object	
• in addition, snPATHO-seq un- and filtered barcodes should be at data/ref/raw and /-filt, respectively	03-ref.R	
• Gut Cell Atlas reference data should be at data	◦ exclude cells too close to any FOV b	
(it can be retrieved from https://www.gutcellatlas.org/)	◦ exclude cells with low counts, low coverage, and high negative probe or false positive	
downstream analyses		
• 03-prn.R and 05-rep.R : PCA using feature subset according to 03-ist.R and 05-jst.R, respectively	03-sig.R	genomic signature scoring
• 04-cc.R : cell-cell communication analysis	04-ccc.R	cell-cell communication analysis
• 06-trj.R : epithelial trajectory inference	06-trj.R	epithelial trajectory inference
• 06-ctx.R : spatial context/niche analysis	06-ctx.R	spatial context/niche analysis
• 06-cty.R : niches blinded to epithelia	06-cty.R	niches blinded to epithelia
visualization		
notes		
• Software versions used throughout this study are captured in the session information here	03-pcl.R	majority of scripts serve the purpose of collecting specific subsets of results, and visualizing them
• A complete list of package dependencies (as well as installation commands) are provided in code/00	◦ read polygon data from .csv	• 10-plt-<by1>-<by2>-<out1>,<out2>,<plt>.R
• .Rprofile is used for handling command line	◦ filter for cells passing QC	◦ pool outputs <out1>/<out2> according to <by1>/<by2>
• logs/ capture .Rout files from R CMD BATCH	◦ write to .parquet	◦ generates plts/<out1>,<out2>,<plt>.pdf (depending on <by1>/<by2>, name may also include subset (sub) or section (sid) identifier)
• Intermediate results are written to outs/ (as .r	03-roi.R	◦ sid = one section
• visualizations are written to plts/ (as .pdf)	◦ retrieve Napari-based shape annotations	◦ all_sid = all colon sections
	◦ histopathological regions (REV)	◦ all_sid_all_sub = also include LN (241)
	◦ invasion-free blood vessels (EV)	◦ stash annotations as cell metadata
	◦ lymphovascular invasions (LI)	
	◦ align coordinates with CosMx data	
	◦ stash annotations as cell metadata	

research at CNAG: lots of tech, lots of data

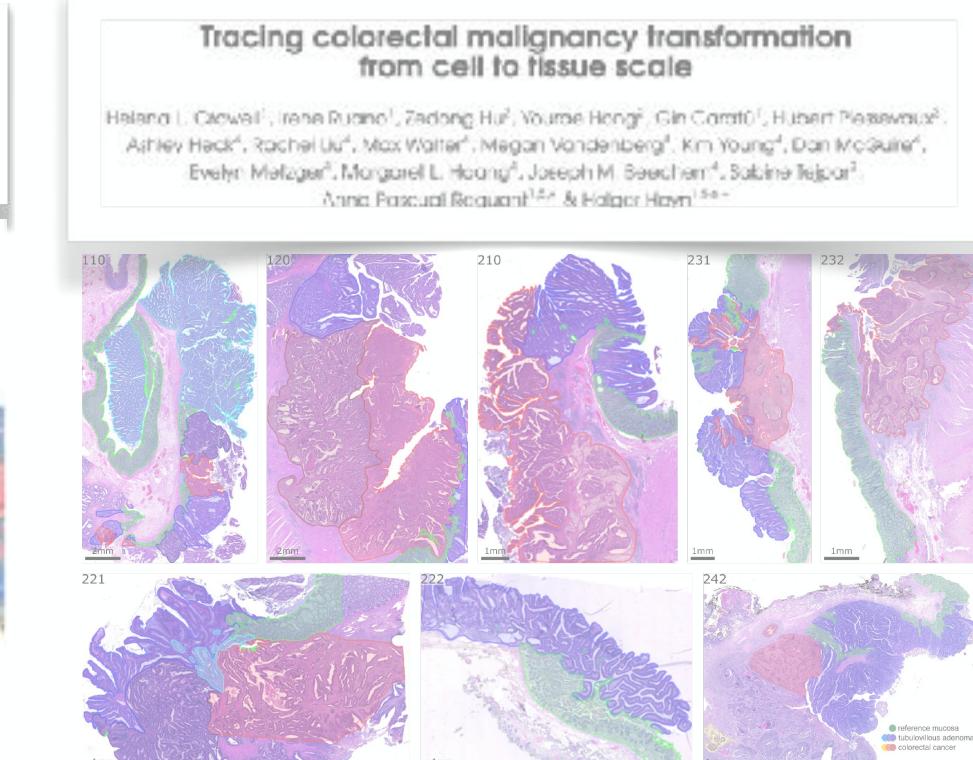
STAMP: Single-cell transcriptomics analysis and multimodal profiling through imaging
Emanuele Pitino^{1,14} · Anna Pascual-Reguant^{1,2,14} · Felipe Segato-Dezem^{3,4,14} · ... · Jasmine T. Plummer^{3,4,9,12} · Halger Heyn^{1,2,13} · Luciano G. Martelotto^{5,6,15} · ... Show more



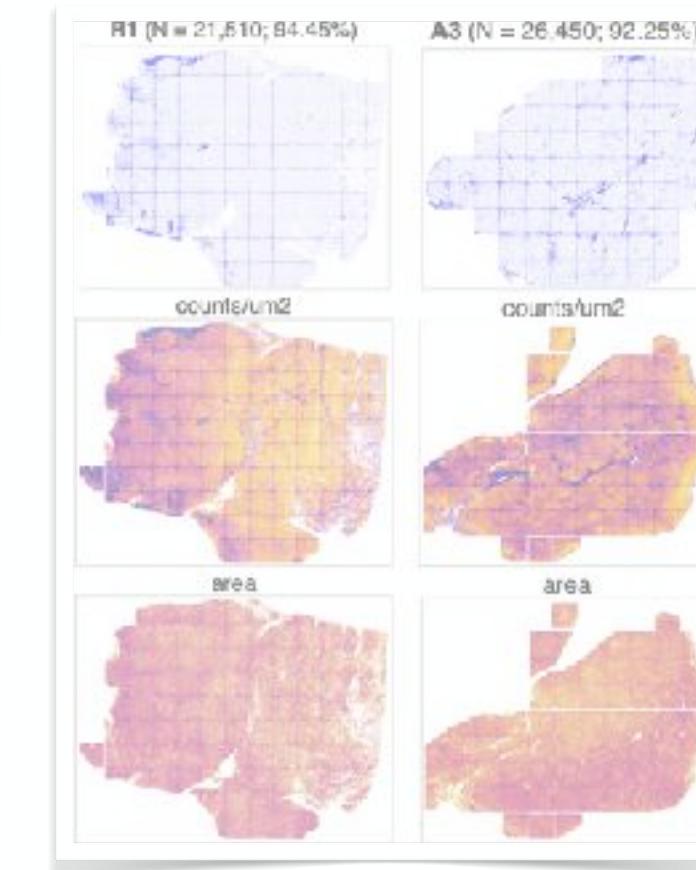
imaging instead of sequencing



1k-plex CosMx on human tonsil



19k-plex CosMx on colorectal cancer



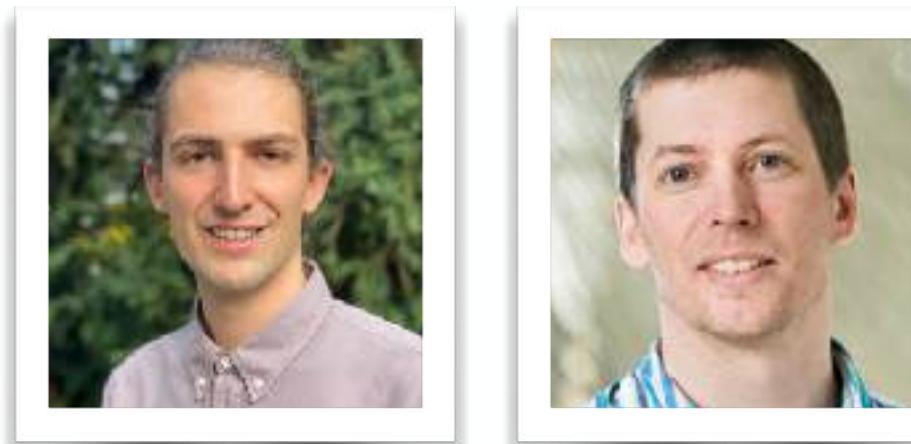
6k-plex CosMx on TMA of lymphoma



single-cell
FACS
scRNA-seq
scATAC-seq
snRNA-seq
snPATHO-seq
10x Multiome
BCR- & TCR-seq
Chromium Connect

spatial
CosMx & Xenium
Visium & HDST
Stereo-sea

structure-based analysis of ST data with **sosta**

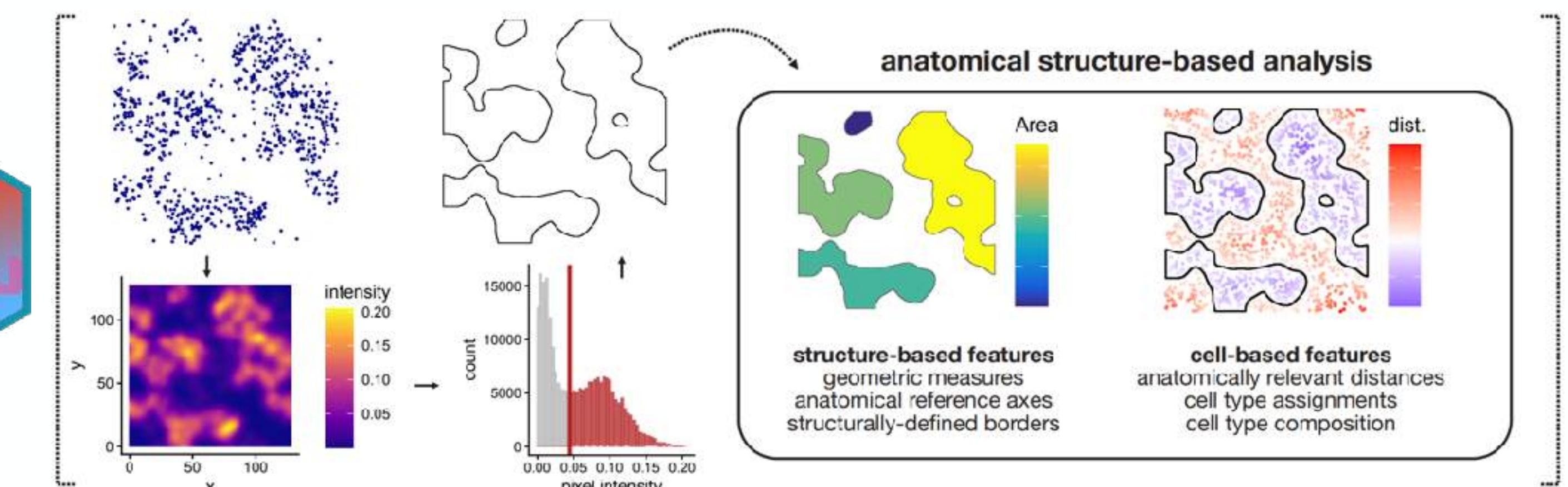
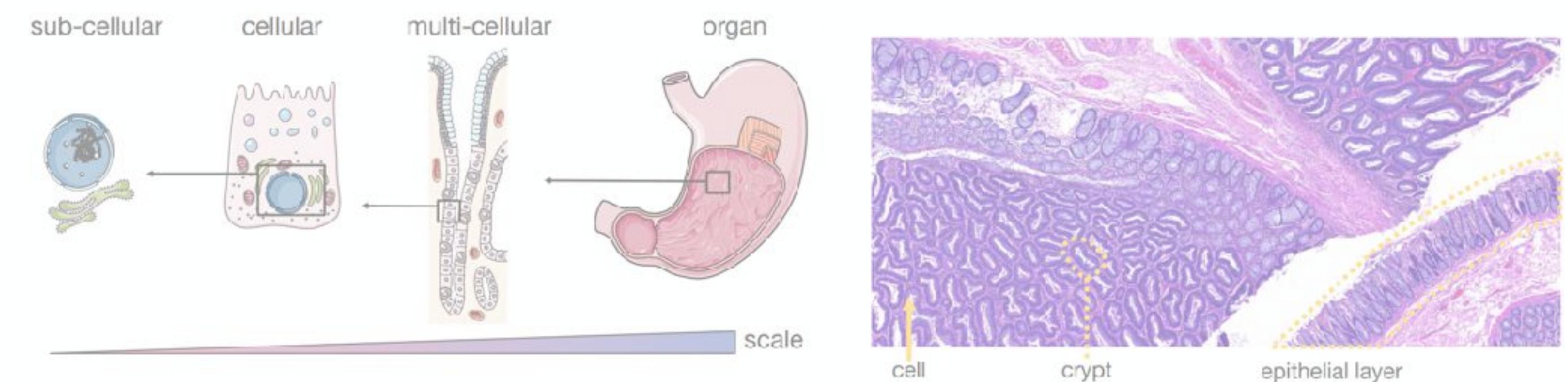


Samuel Gunz
PhD student

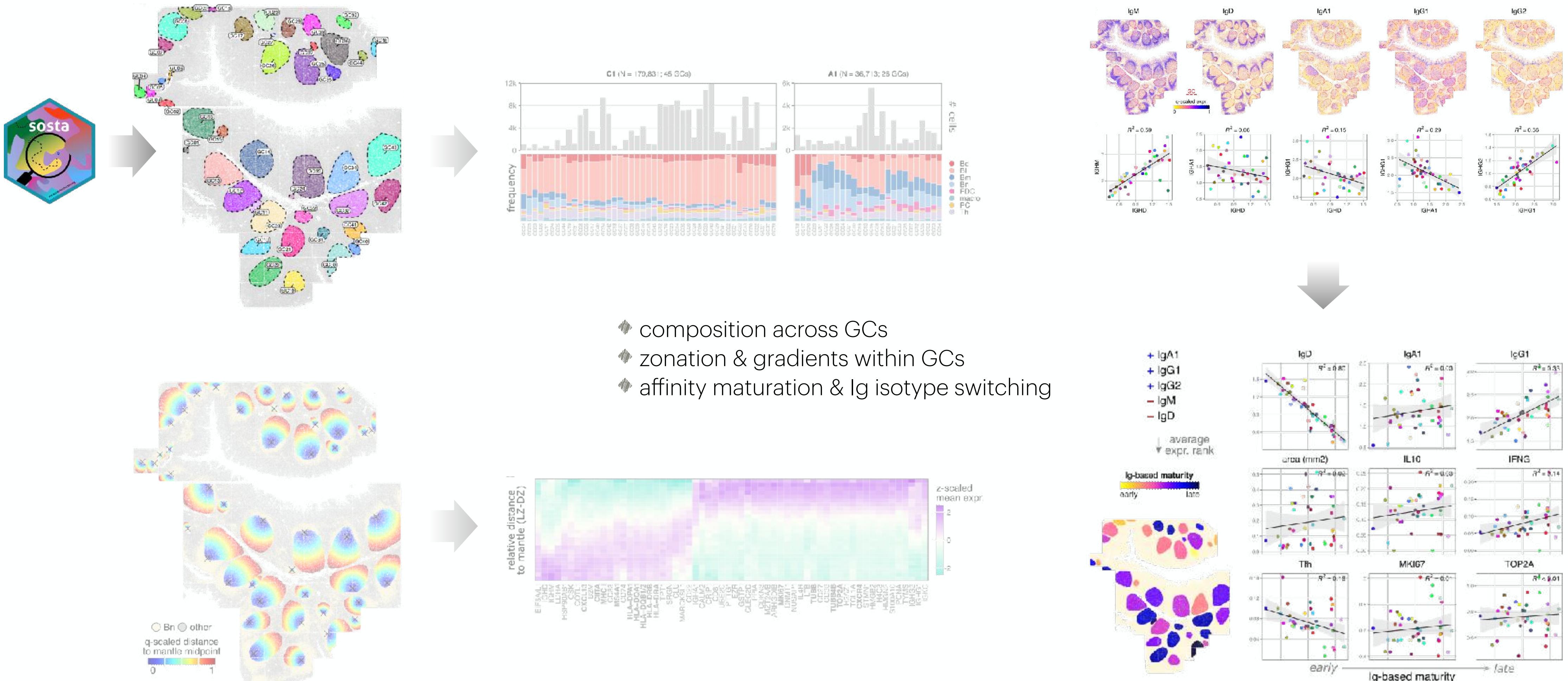
Mark Robinson
PI @ UZH, Switzerland

Analysis of anatomical multi-cellular structures from spatial omics data using sosta

Samuel Gunz, Helena L. Crowell, Mark D. Robinson
[doi: https://doi.org/10.1101/2025.10.13.682065](https://doi.org/10.1101/2025.10.13.682065)

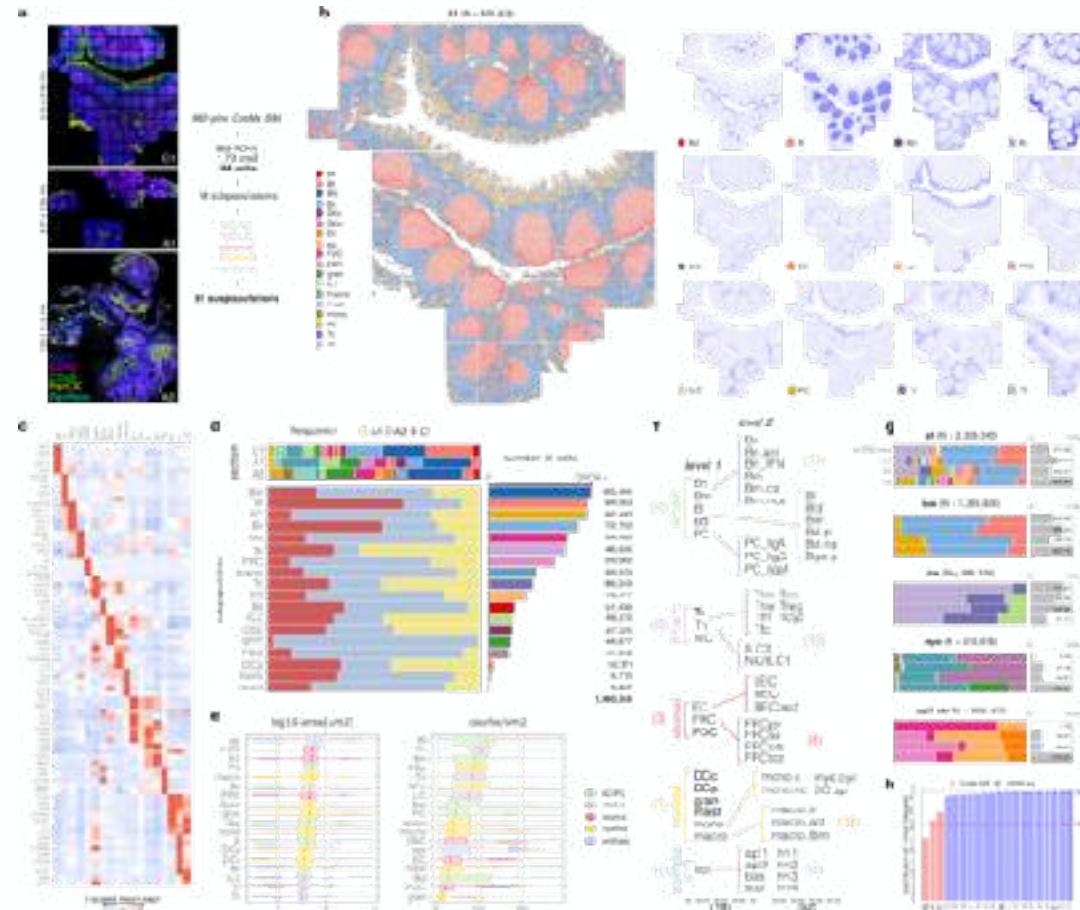


unit-based analysis of germinal centers (GCs)

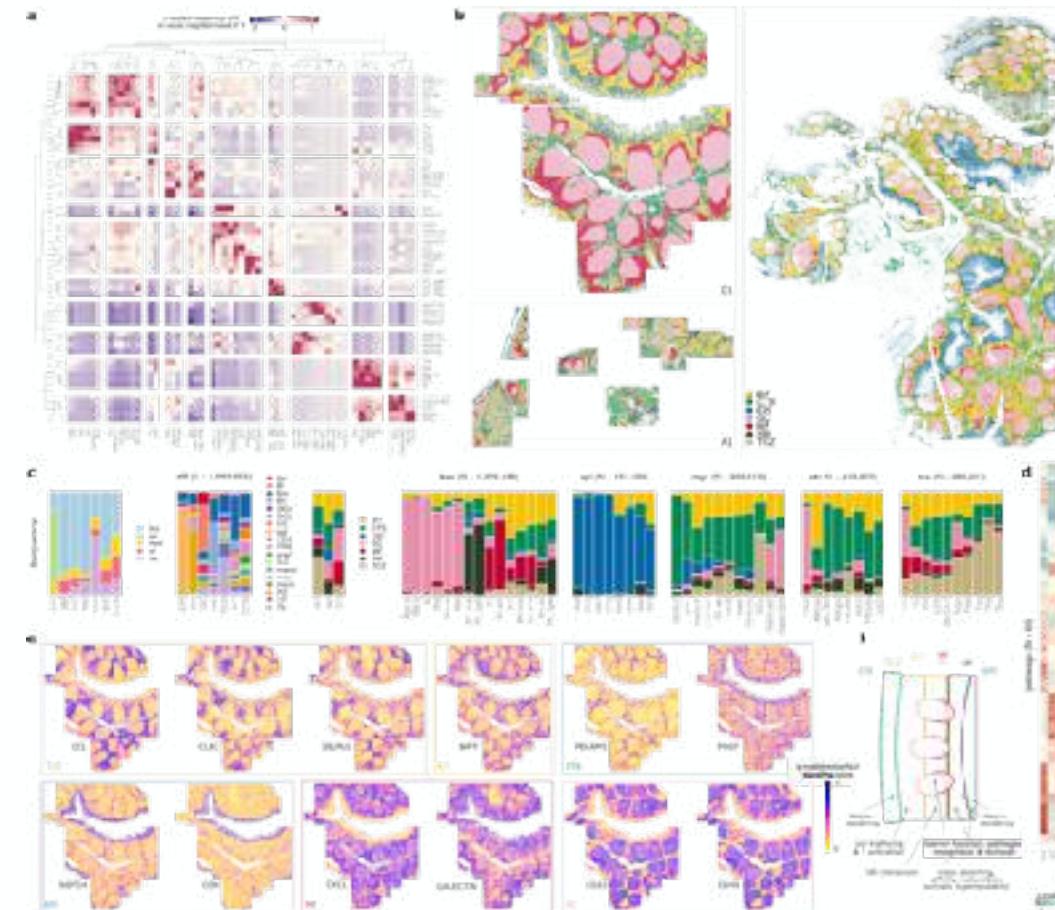


a transcriptional map of tonsil architecture

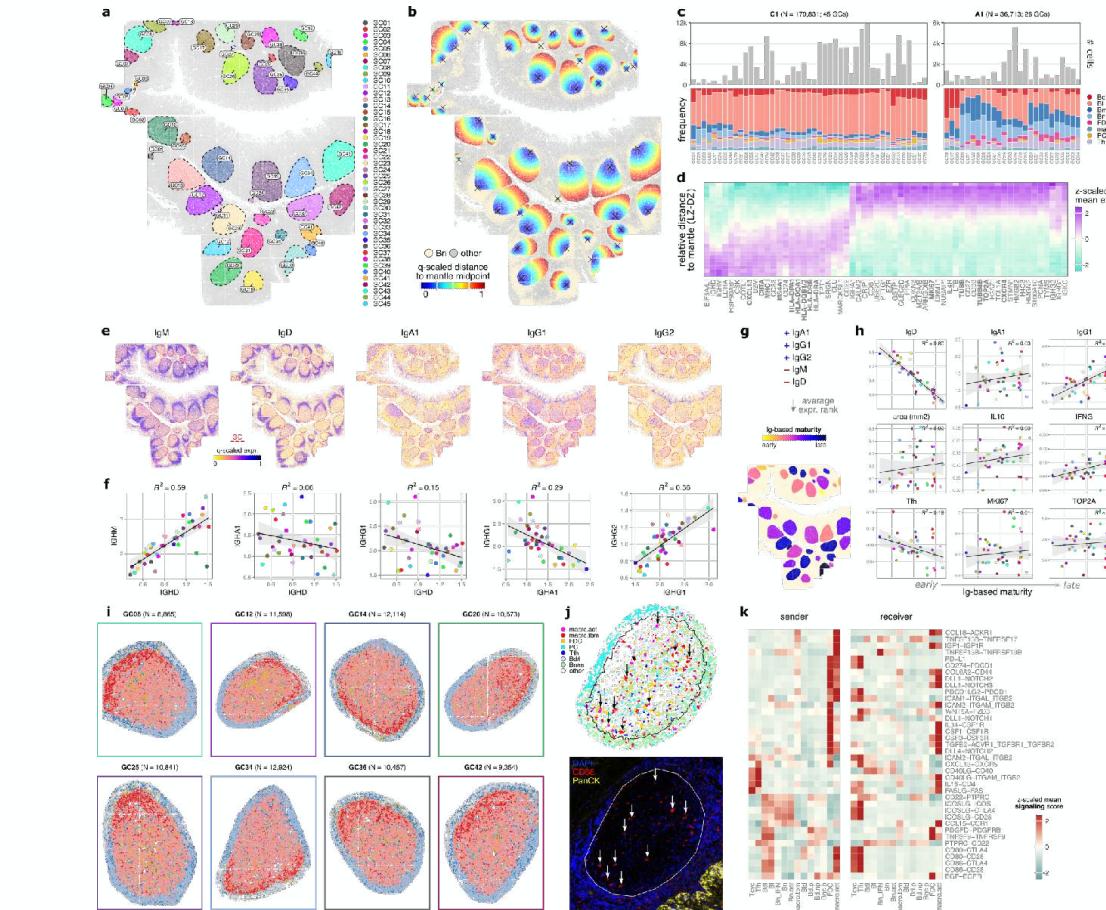
52 subpopulations across
2M cells from 3 sections



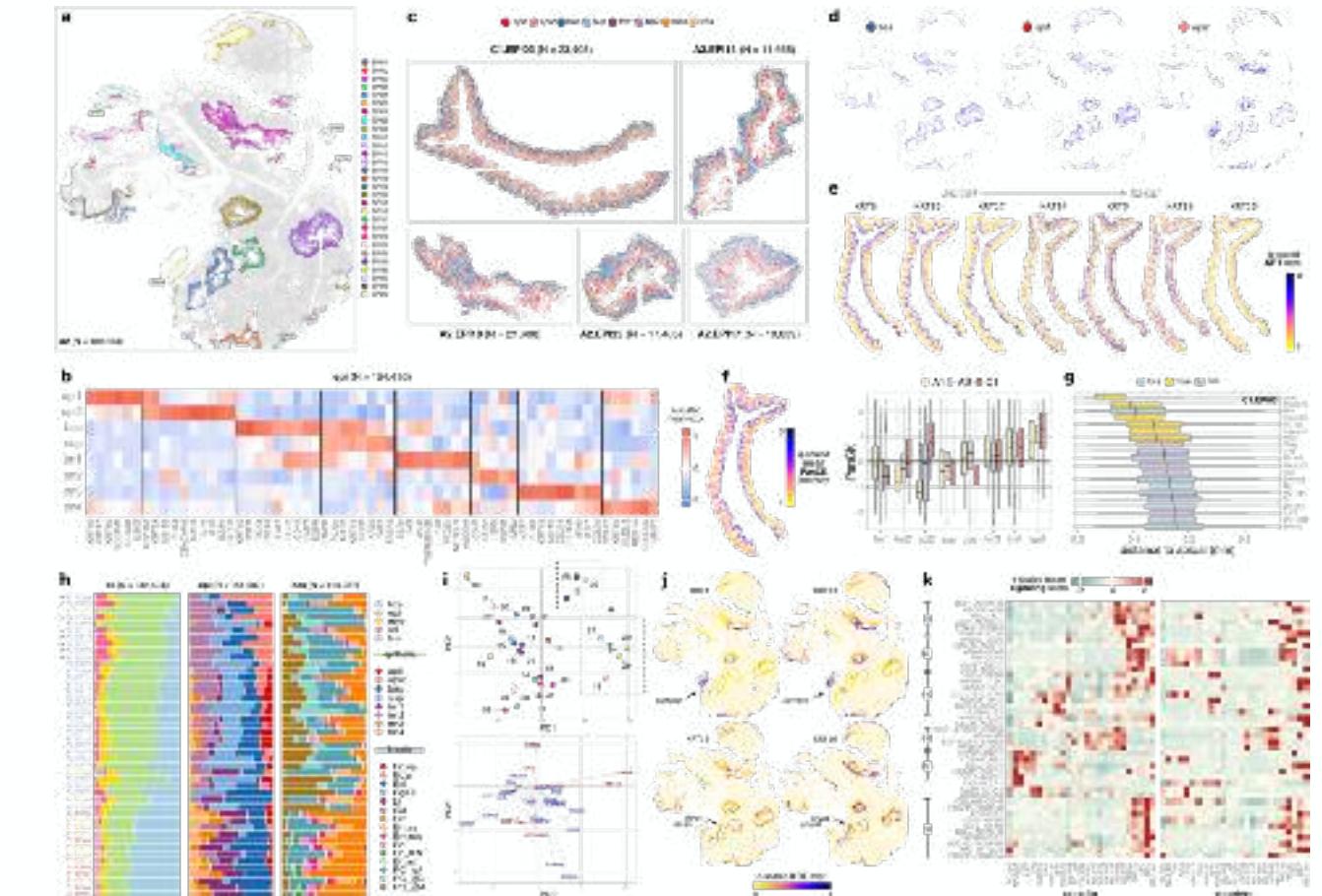
7 niches functionally
compartmentalize tonsil



GCs (composition, zonation,
maturation, efferocytosis)



apical/basal & crypt/surface
epithelium (KRTs, microenv)

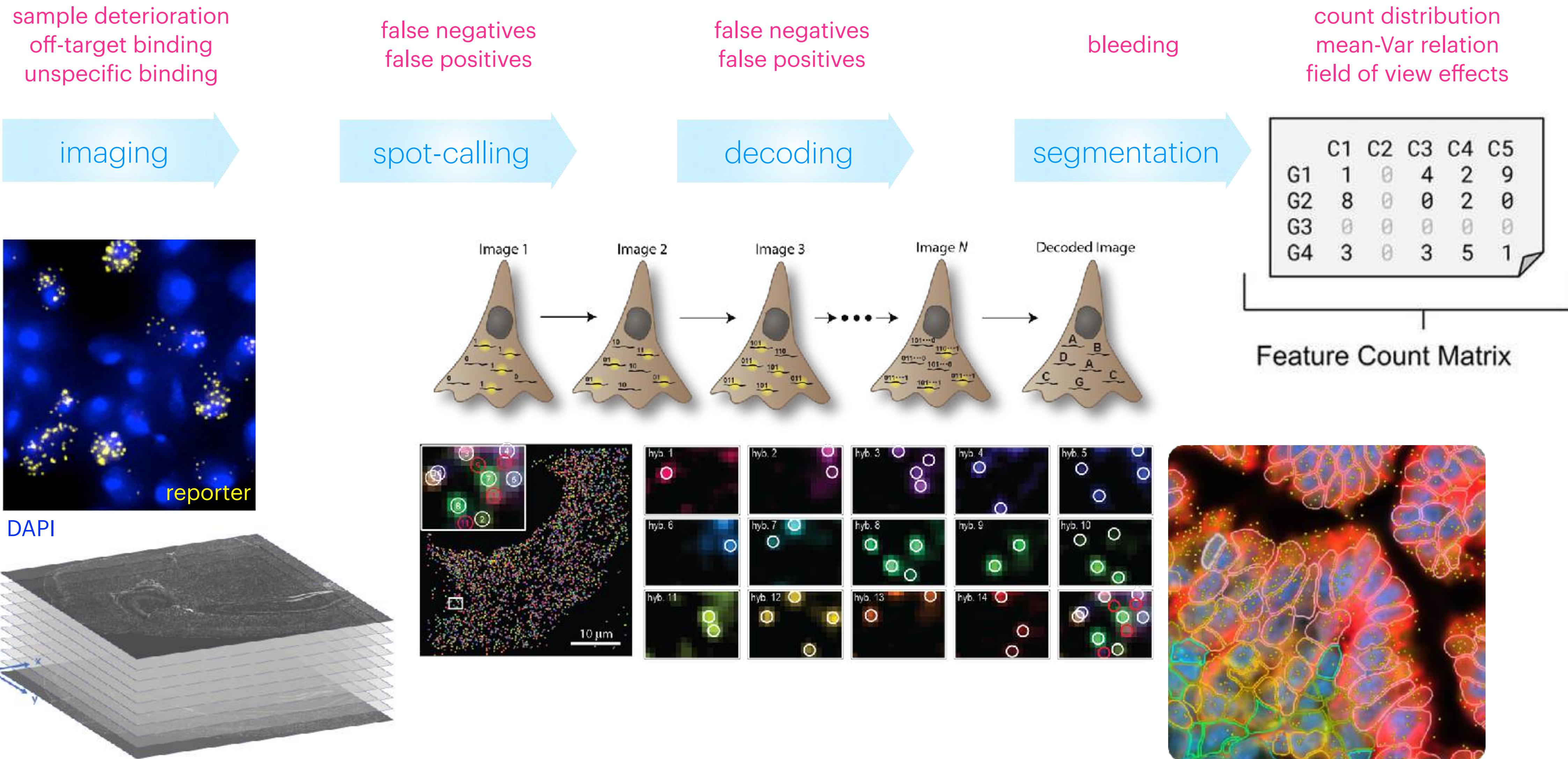


...decent code/pipelines
can be transferable

...but different questions
want different analyses

...knowing the biology
really (really!) helps

but counts from imaging != sequencing



OSTA: Orchestrating Spatial (Transcript)Omics Analysis with Bioconductor

- Ilaria Billalò, University of Padova, Italy
- Pelying Cal, University of Zurich, Zurich, Switzerland
- Leonardo Collado-Torres, Lieber Institute for Brain Development, Baltimore, MD, United States
- Helena L. Crowell, National Center for Genomic Analysis, Barcelona, Spain
- Estella Yixing Dong, Lausanne University Hospital (CHUV) and University of Lausanne, Lausanne, Switzerland
- Martin Emons, University of Zurich, Zurich, Switzerland
- Shila Ghazanfar, University of Sydney, Sydney, Australia
- Raphael Gottardo, Lausanne University Hospital (CHUV) and University of Lausanne, Lausanne, Switzerland
- Samuel Gunz, University of Zurich, Zurich, Switzerland
- Boyi Guo, Johns Hopkins University, Baltimore, MD, United States
- Stephanie C. Hicks, Johns Hopkins University, Baltimore, MD, United States
- Artur Manukyan, Max Delbrück Center for Molecular Medicine, Berlin, Germany
- Dario Righelli, University of Naples Federico II, Italy
- Keri Martirowich, Lieber Institute for Brain Development, Baltimore, MD, United States
- Kristen Maynard, Lieber Institute for Brain Development, Baltimore, MD, United States
- Pratibha Panwar, University of Sydney, Sydney, Australia
- Ellis Patrick, University of Sydney, Australia
- Shreya Rac, University of Sydney, Australia
- Davide Rissò, University of Padova, Italy
- Mark D. Robinson, University of Zurich, Zurich, Switzerland
- Callum Sargeant, Walter and Eliza Hall Institute, Melbourne, Australia
- Jieran Sun, Lausanne University Hospital (CHUV) and University of Lausanne, Lausanne, Switzerland
- Michael Totty, Johns Hopkins University, Baltimore, MD, United States
- Lukas M. Weber, Boston University, Boston, MA, United States
- Vince Carey, Harvard Medical School, Boston, MA, United States
- Lori Ann (Shepherd) Kern, Roswell Park Comprehensive Cancer Center, Buffalo, NY, United States
- Alexandru Mahmoud, Harvard Medical School, Boston, MA, United States



1	seq - Visium H...	#62	estefac...	-	Todo
2	img-workflow...	#93	drisso...	-	Todo
3	img-segmentat...	#99	-	-	Todo
4	img - dimensio...	#55	Pratib...	-	In Progress
5	cis - spatially v...	#56	Imweb...	-	In Progress
6	seq - quality co...	#57	boyigu...	-	In Progress
7	seq - normaliz...	#58	Imweber...	-	In Progress
8	seq deconvolu...	#63	ostolac...	-	In Progress
9	img - normaliz...	#68	drighel...	-	In Progress
10	img - sub-cellu...	#73	Pratib...	-	In Progress
11	img - cell-cell ...	#53	Helena...	-	In Review
12	cis - feature se...	#54	Helena...	-	In Review
13	img-clustering...	#66	Jieran-S...	-	In Review
14	Permission to r...	#67	markro...	-	1st draft
15	aligned Xenium...	#75	Helena...	-	1st draft
16	Workflow chap...	#42	estella...	-	1st draft
17	Workflow chap...	#43	estella...	-	1st draft
18	bkg - data type...	#51	Helena...	-	1st draft
19	img - quality co...	#52	drighel...	-	1st draft
20	seq - clusterin...	#59	Jieran-S...	-	1st draft
21	bkg - spatial-p...	#60	markro...	-	1st draft
22	seq / img - Visiu...	#61	-	-	1st draft
23	cis - registration	#64	Artur...	-	1st draft
24	cis-spat-stat #65		mjemo...	-	1st draft
25	bkg-interoperab...	#109	Helena...	-	1st draft

Welcome

Preamble

Introduction

- 1 Introduction
- 2 Spatial omics
- 3 Infrastructure

Sequencing-based platforms

- 4 Introduction
- 5 Load data
- 6 Quality control
- 7 Normalization, feature selection, and dimensionality reduction
- 8 Clustering
- 9 Spot deconvolution
- 10 Spatial co-localization
- 11 Marker genes
- 12 Workflow: Human DLPFC
- 13 Workflow: spatialLBD

Imaging-based platforms

- 14 Introduction
- 15 Quality control
- 16 Dimensionality reduction
- 17 Clustering
- 18 Neighborhood analysis
- 19 Cell communication
- 20 Sub-cellular analysis
- 21 Xenium workflow
- 22 CosMx workflow

Technology-agnostic/cross-platform/downstream analyses

- 23 Introduction
- 24 Spatially variable genes
- 25 Gene/feature-set signatures
- 26 Spatial Statistics
- 27 Spatial registration
- 28 Multiple-sample differential expression
- 29 Workflow: Xenium + Visium

Appendices

- A Acknowledgments
- B Related resources



lmweber.org/OSTA



OSTA: Orchestrating Spatial (Transcript)Omics Analysis with Bioconductor



OSTA: Orchestrating Spatial (Transcript)Omics Analysis with Bioconductor

