Genetics, Biomedical Computing, and Genomics Monthly Meeting – New Technologies at Genomics Research and Technology Hub

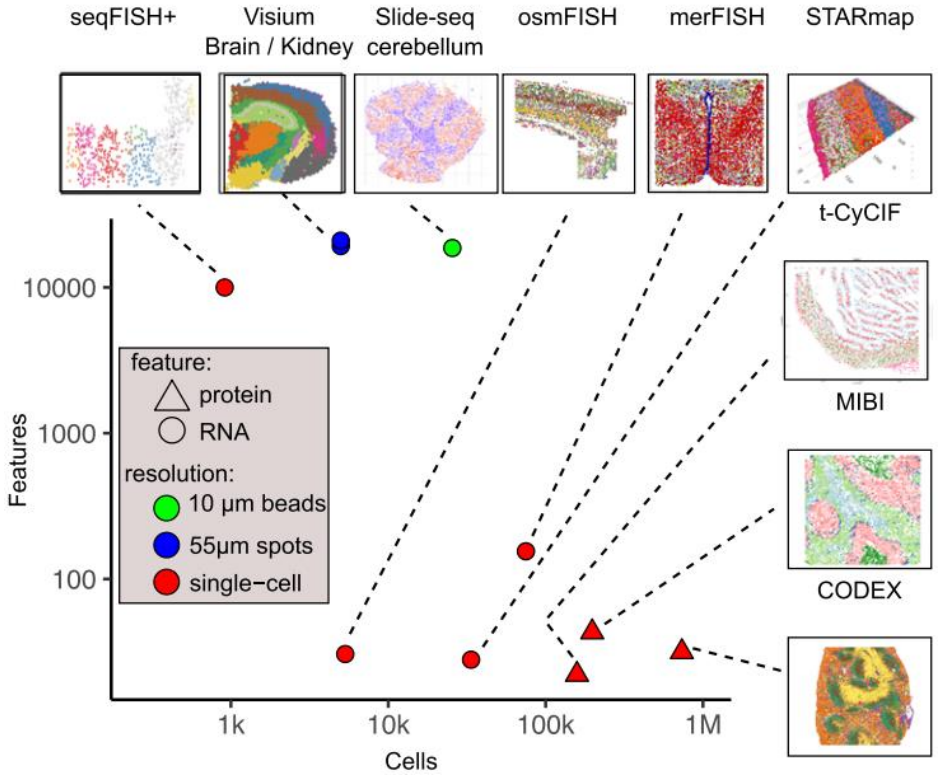
New Analyses in Spatial Transcriptomics (ST)

Jenny Wu, Ph.D.

Feb 6, 2024

New Analyses in Spatial Transcriptomics (ST)

- Analysis workflow for Imaging based and sequencing based spatial omics data
- Cell type deconvolution for sequencing based spatial platforms
- Cellular segmentation for imaging based spatial platforms
- Downstream Analysis: spatial DE, cell-cell communication in space



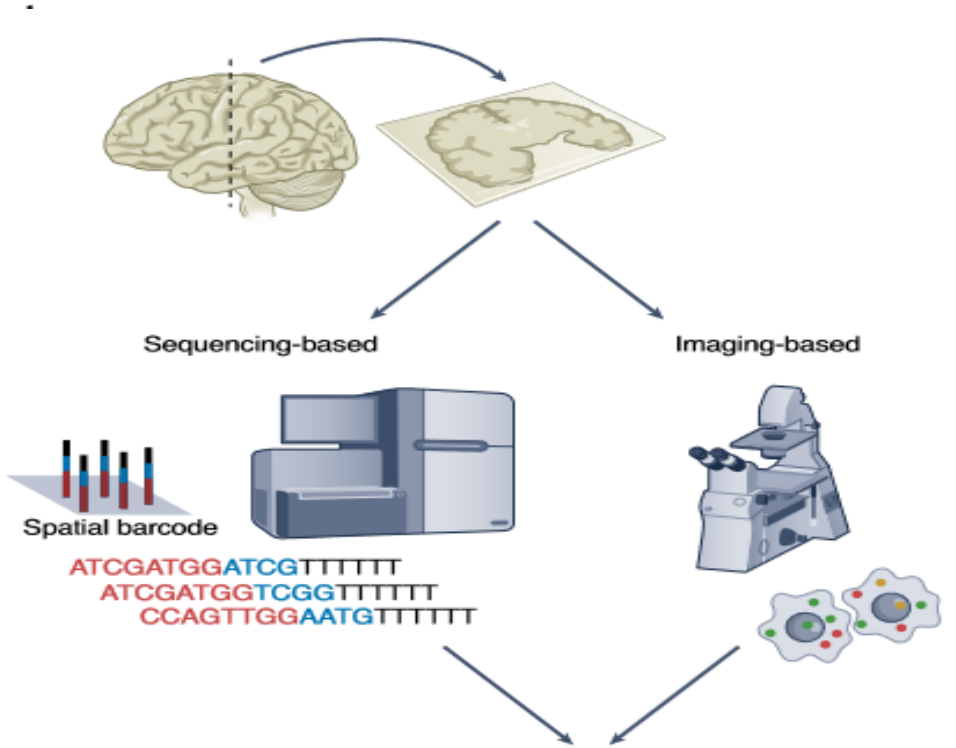
Next spatial data analysis workshop is on Feb 27!

Software tools are being containerized and available on HPC3 biojhub

Dries *et al*, 2021

Two Classes of ST Technology

- Sequencing Based (SST) gene detection
  - 10x Visium, Slide-seq etc
  - Whole transcriptome but not true single cell resolution
- Imaging based (IST) gene detection
  - 10x Xenium, MERFISH etc.
  - Subcellular resolution but limited gene throughput



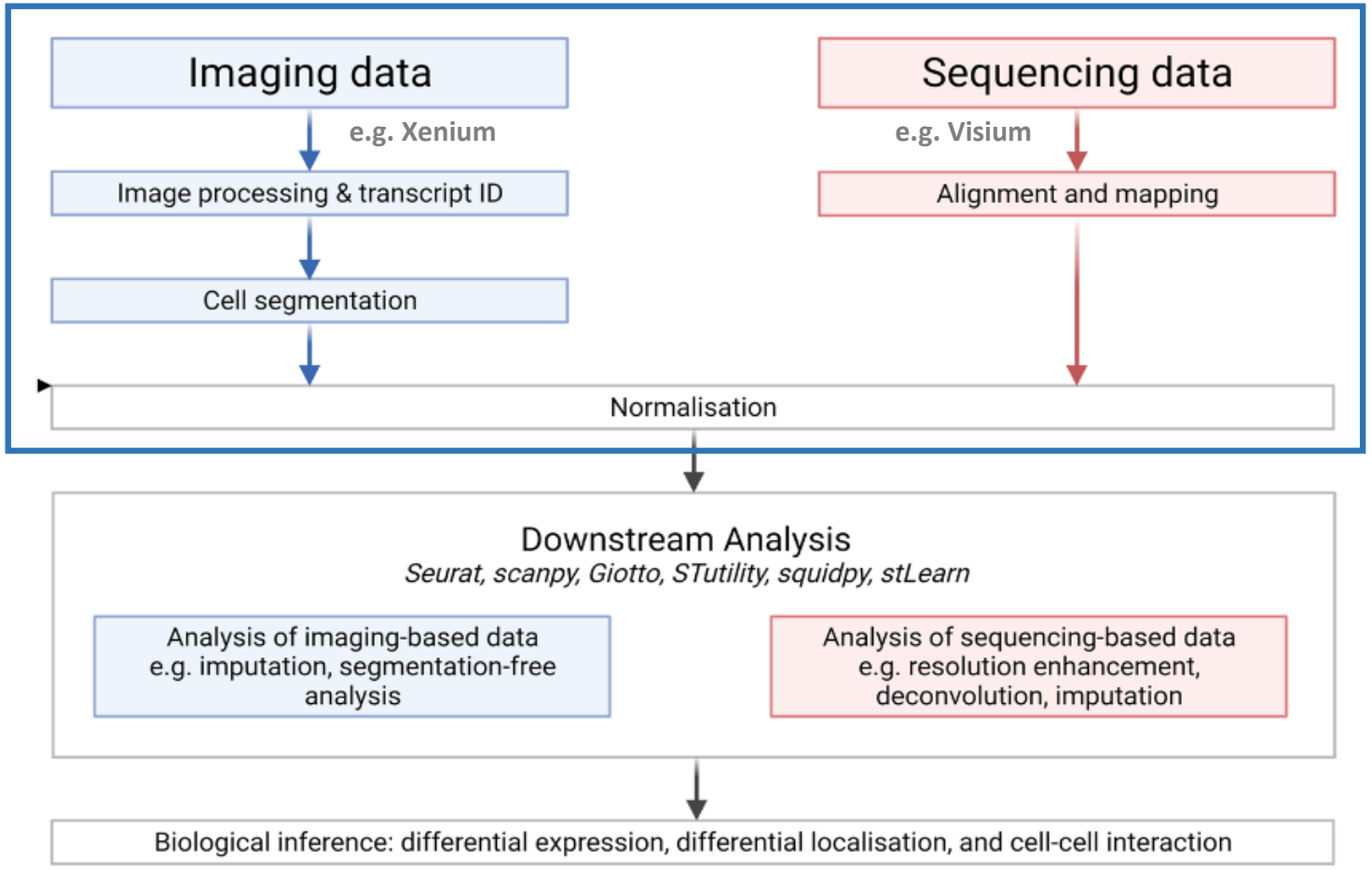
	Gene 1	Gene 2	Gene *N*
Spot 1	10	5	6
Spot 2	5	21	20
...	...	...	...
Spot *N*	8	1	2

	x	*y*
Spot 1	141.2	511.4
Spot 2	514.9	219.9
...	...	...
Spot *N*	8	127.4

Tian *et al*, 2023

General Data Analysis Workflow

- Pre-processing is handled by proprietary software such as *Space Ranger or* on Xenium Analyzer instrument.
- Downstream analysis may be performed with a wide range of transcriptomics analysis packages: *Seurat, Scanpy, Giotto* etc.

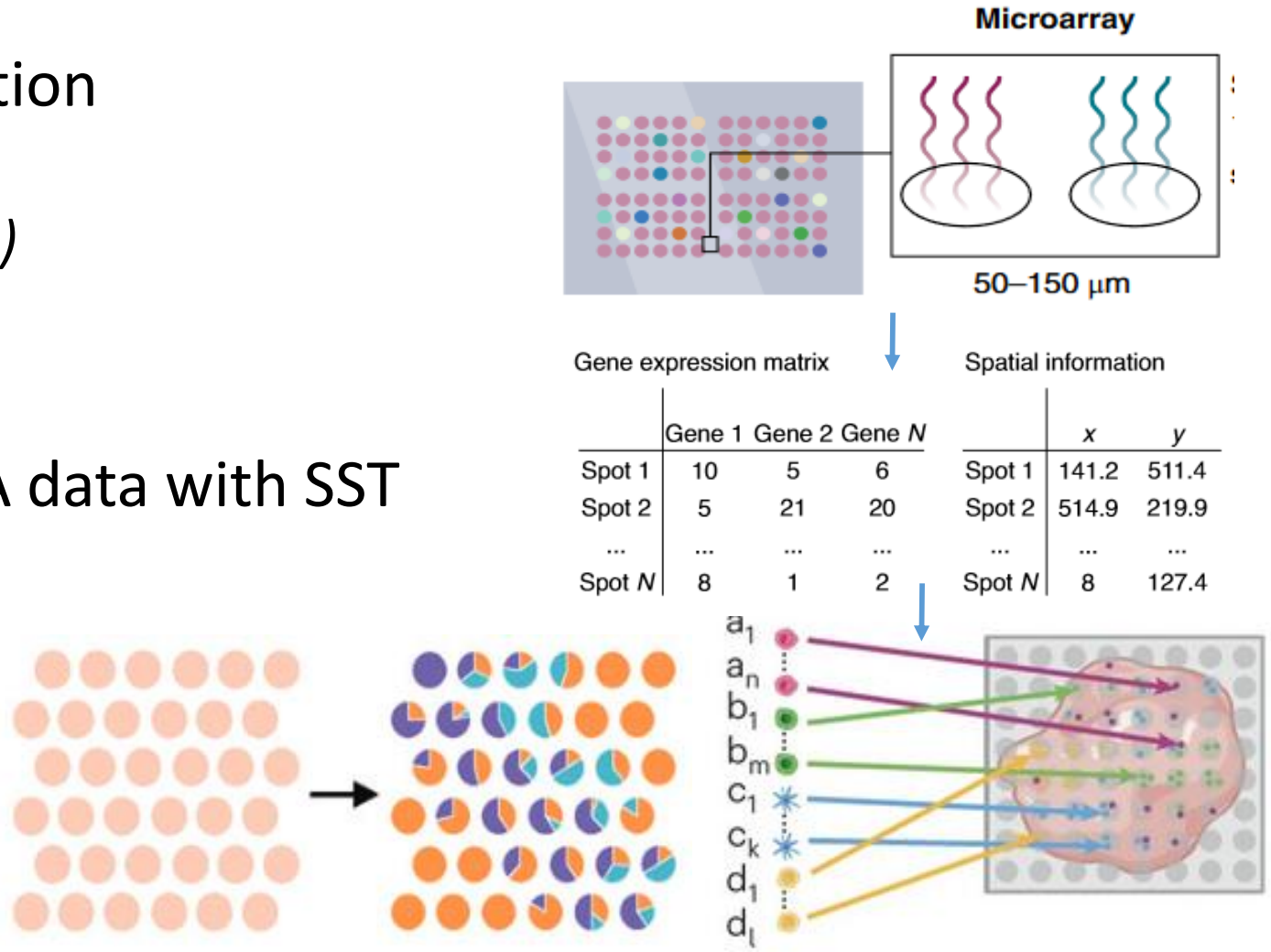


Williams *et al*, 2022

Sequencing Based ST Data Analysis: Visium

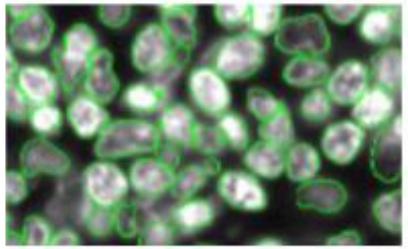
- Cell Type Deconvolution
  - *RCTD*
  - *SpatialDWLS (Giotto)*
  - *BayesSpace*

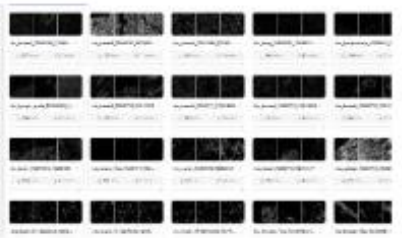
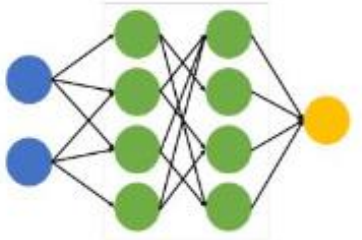
- Alignment of scRNA data with SST Data
  - *CytoSpace*
  - *Tangram*
  - *Cell2location*



Imaging Based ST: Cell Segmentation

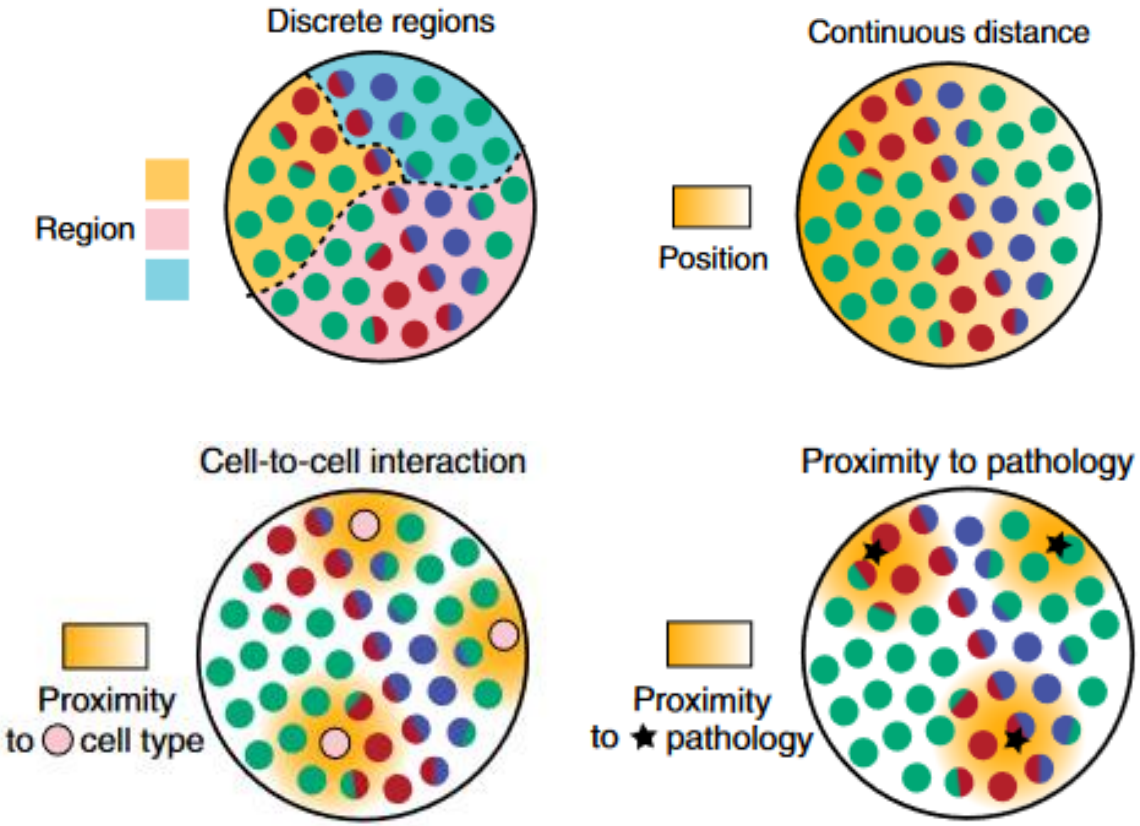
- Many of the downstream analyses and interpretations of the spatially resolved data depend on the ability to resolve individual cells
  - *Cellpose*, *Baysor*
  - *Stardist*, *deepcell*
  - Spot-based Spatial cell-type Analysis by Multidimensional mRNA density estimation (*SSAM*)
  - *DL* can be computational intensive



Downstream Analysis: Spatial DE

- Cell type specific differential gene expression (*C-SIDE*)
  - Takes into account cell types
  - Accounting for localization of other cell types
  - Can incorporate covariates and replicates
  - Can detect DE due to pathology, anatomical regions, cell-cell interactions etc

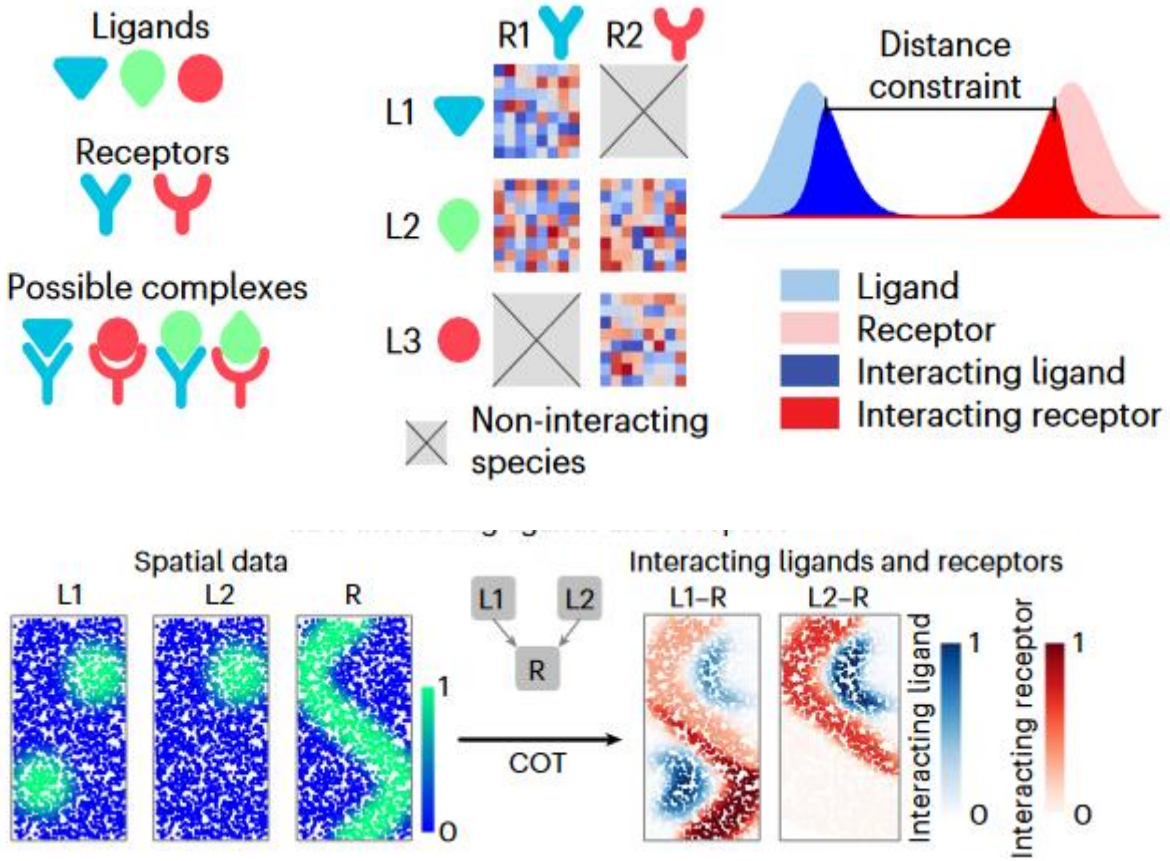


Implemented in spacexr package

Cable *et al*, 2022

Downstream Analysis: Cell-Cell Communication

- Spatial CellChat (v2)
- COMMOT: collective optimal transport based
- SpaTalk : knowledge-graph-based cell-cell communication inference



Cang *et al*, 2023

Challenges in ST Data Analysis

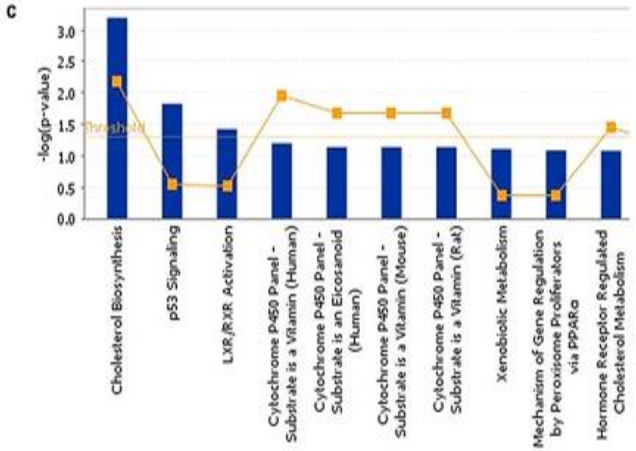
- A larger variety of file formats and data structures due to heterogeneity of methodologies
- No standardized tissue and quality control measurements
- Huge data volume requires more powerful computers
- Computation tools are evolving rapidly

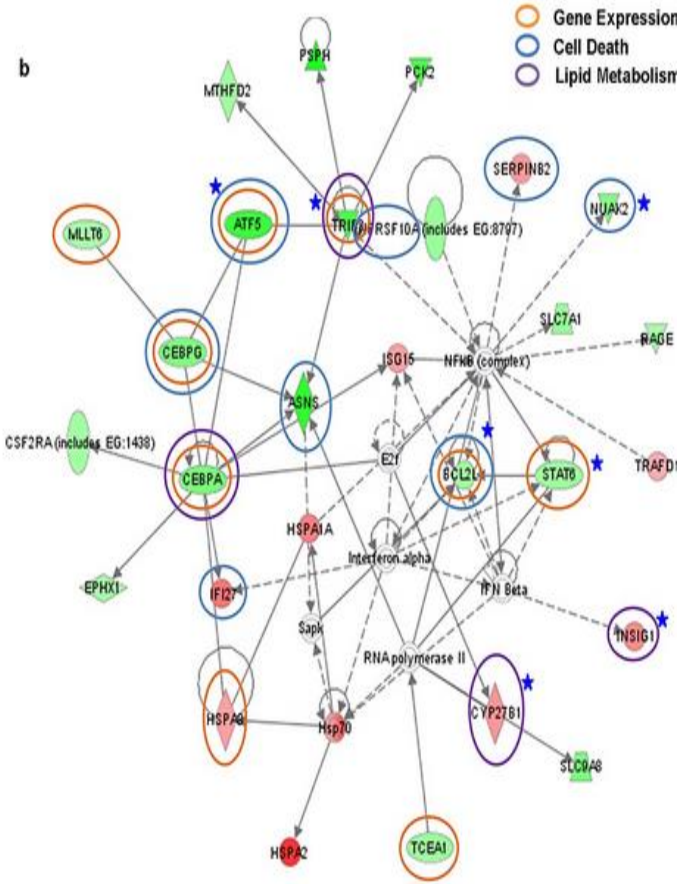
Pathway and Network Analysis: Ingenuity IPA

- IPA: Pathway and network analysis and biological interpretation of any kind of omics data
- Powerful knowledge base with 8M literature based data points

a

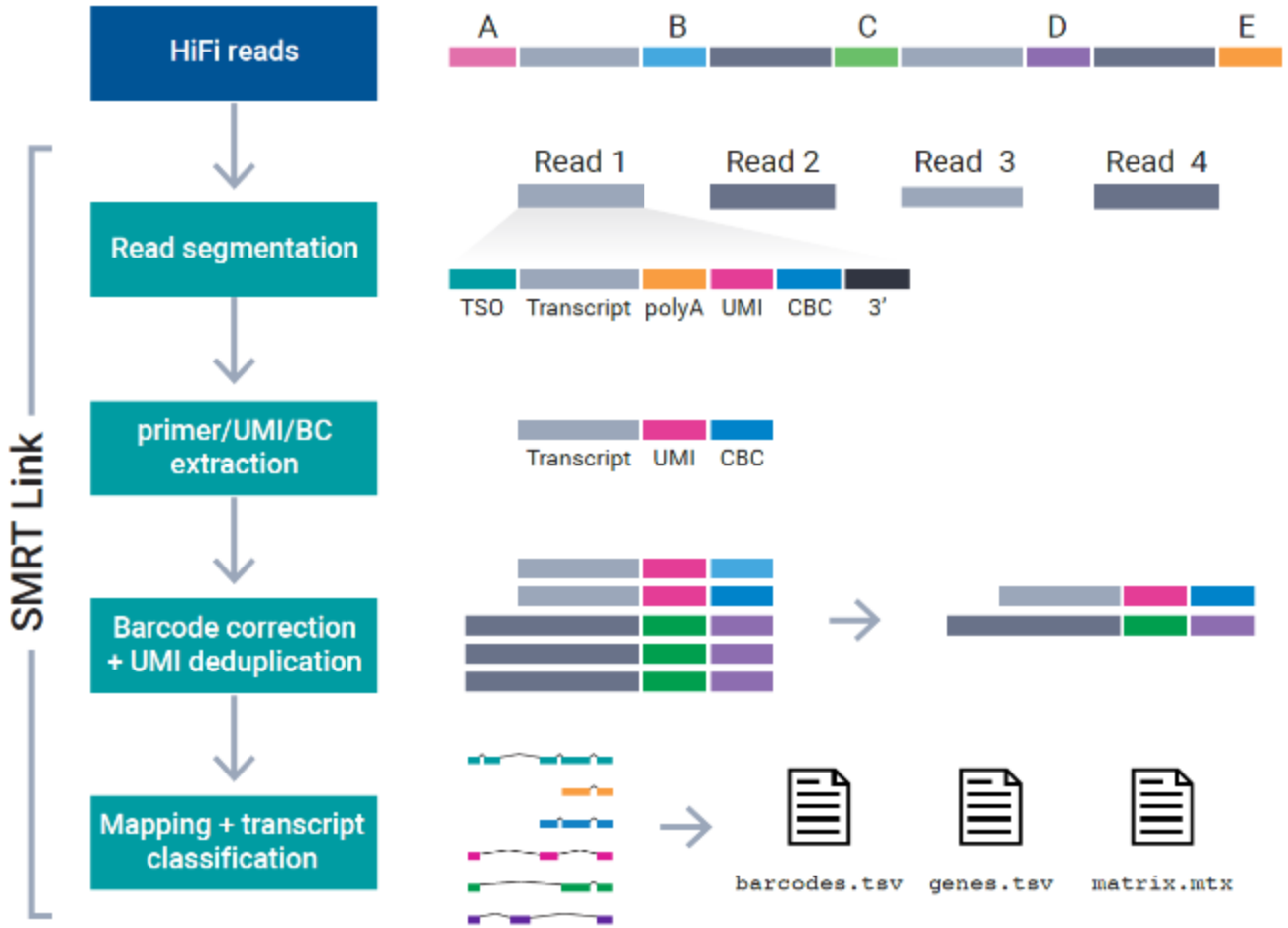
ID	Associated Network Functions	Score
1	Gene Expression, Cell Death, Lipid Metabolism	54
2	Cancer, Renal and Urological Disease, Amino Acid Metabolism	34
3	Cell-To-Cell Signaling and Interaction, Cellular Growth and Proliferation, Hematological System Development and Function	23
4	Gene Expression, Cellular Function and Maintenance, Skeletal and Muscular System Development and Function	23
5	Gene Expression, Cell Cycle, Cellular Growth and Proliferation	23





MAS-Seq bioinformatics workflow overview

The **SMRT Link Read Segmentation and Single-cell Iso-Seq** workflow processes HiFi reads generated from a MAS-Seq library to produce gene- and isoform-level count matrices that are compatible with tertiary single-cell analysis tools



Read segmentation

- HiFi reads are segmented into individual segmented reads (S-reads) that represent the original cDNA sequences.

Primer/UMI/BC extraction

- Primers and polyA tails are removed, but also used to orient the read into 5' – 3' orientation. Single-cell barcode and UMI information are extracted.

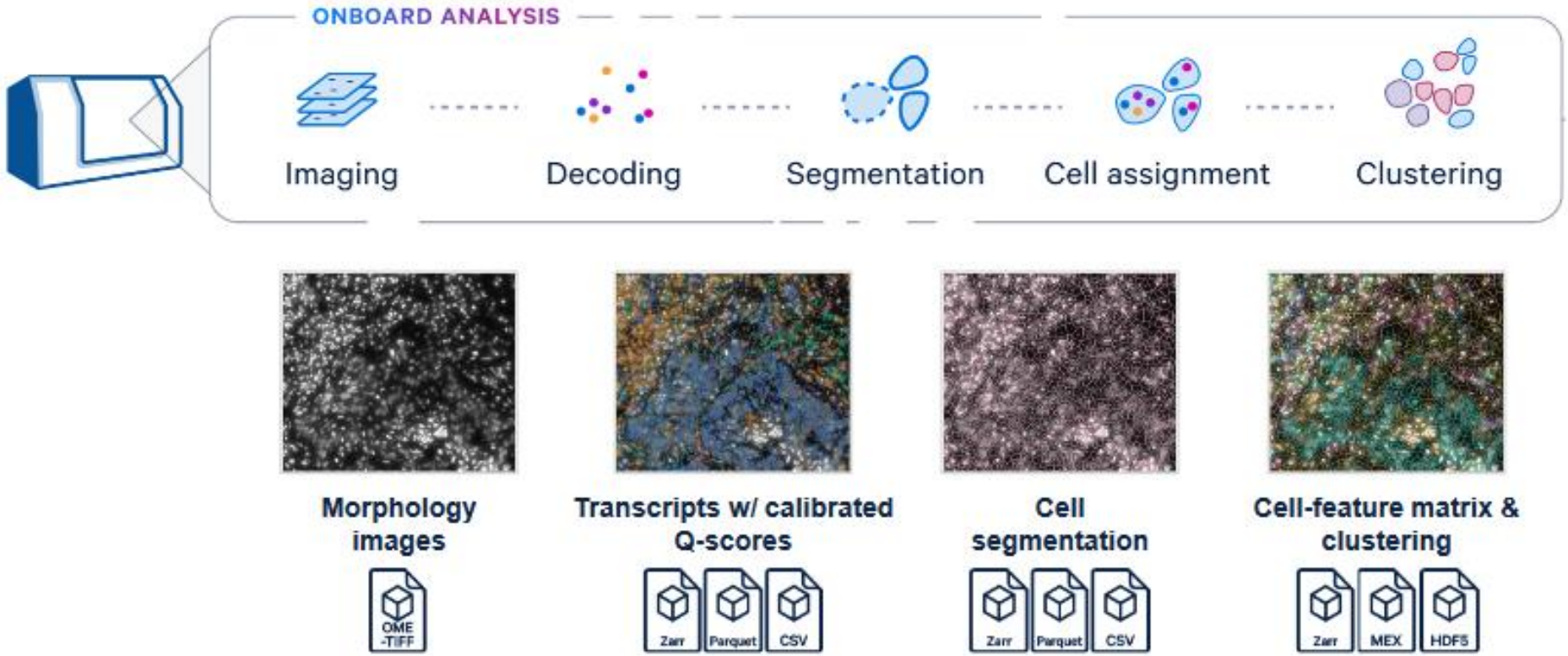
Barcode correction & UMI deduplication

- Cell barcodes are corrected given an expected barcode list. Real cells — cell barcodes that represent encapsulated single cells (as opposed to ambient RNA) are also identified at this step. Reads are then deduplicated based on cell barcodes and UMIs.

Mapping and transcript classification

- Deduplicated reads are mapped to the reference genome and classified against a transcript annotation (e.g., GENCODE). Finally, a gene- and isoform-level single-cell matrix is output for tertiary analysis.

Xenium Onboard Analysis Output Formats



Xenium – Multi-modal Segmentation Built to Enable Broad Coverage