Basic Single Cell & Spatial Transcriptomics

Jenny Drnevich, PhD

Assistant Director, HPCBio

Roy J. Carver Biotechnology Center

June 24, 2025

Learning objectives

- To elucidate the differences between bulk, single cell and spatial RNA-Seq.
- To get a brief overview of the various single cell and spatial platforms
- To learn the details and vocabulary of 10x Genomic's sequencing methods for single cell and Visium
- To review the standard steps of quantification and analysis of 10x data, alternative methods and limitations of all methods.
- To present pros/cons of manual vs. automatic cell type calling
- To discuss the coming development of spatial data methods.

What is transcriptomics?

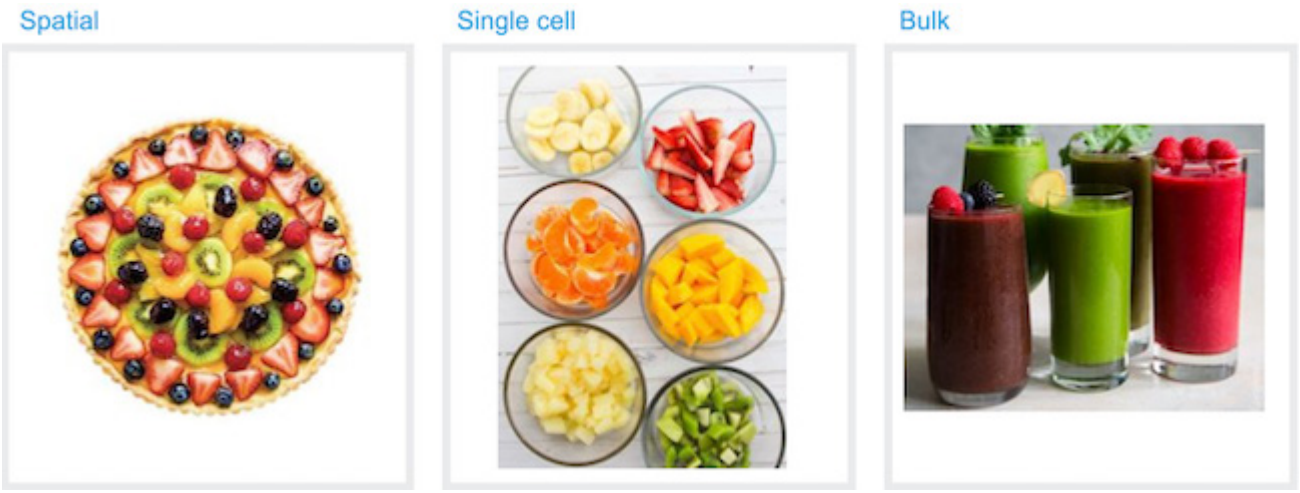
- The study of the "transcriptome" or the transcribed portion of the genome, **RNA**
- Most often this focuses on **mRNAs**, which are translated into proteins
- But also can include all other species of RNAs: **rRNAs, miRNAs, lncRNAs**, etc.

Bulk vs. single cell vs. spatial

Spatial - the actual arrangement of cells in a tissue

Single cell - dissociated cells that can be separated into types

Bulk - whole tissue extractions averaged over all cells



Hernadez, A. DNA Services, Applications [Internet]. Urbana (IL): University of Illinois, Carver Biotechnology Center; [cited 2024 Jun 6]. Available from: https://biotech.illinois.edu/htdna/applications

Single cell methods overview (see review)

Droplet based

- CEL-seq
- Drop-Seq
- Smart-Seq2
- 10x Genomics

Spatial methods overview (see review)

Sequencing based

- 10x Genomics Visium, Visium HD
- Slide-seq (now Takara/Curio), Stereo-seq, Light-seq

Probe based - NanoString GeoMx

Imaging based

- NanoString CosMx
- MERFISH
- STARmap
- 10x Genomics Xenium

Brand new June 2026: Illumina StrataMap

https://www.illumina.com/products/by-brand/stratamap.html

- 750 mm^2 capture area
- 1 um resolution
- poly(A) capture

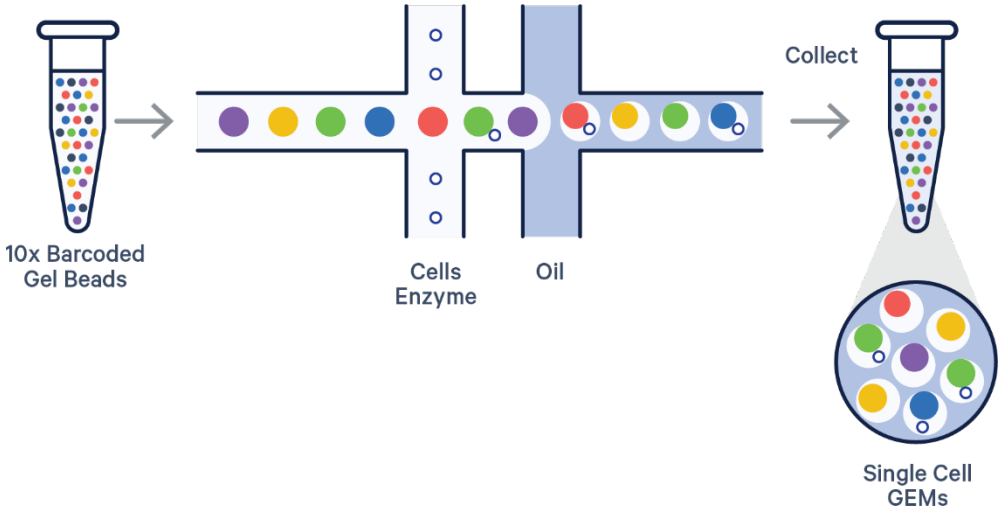
Short reads vs. long reads

Illumina's short reads (<150 nt) allowed all the sequencing-based technologies to be developed and allows excellent gene-level quantification

Pac Bio's HiFi long reads allow full-length (~10 kb) transcript sequencing plus Kinnex kit improves through-put; have worked with 10x Genomics to make the **10x single cell** libraries able to be sequenced by Pac Bio.

10x Genomics Chromium single cell

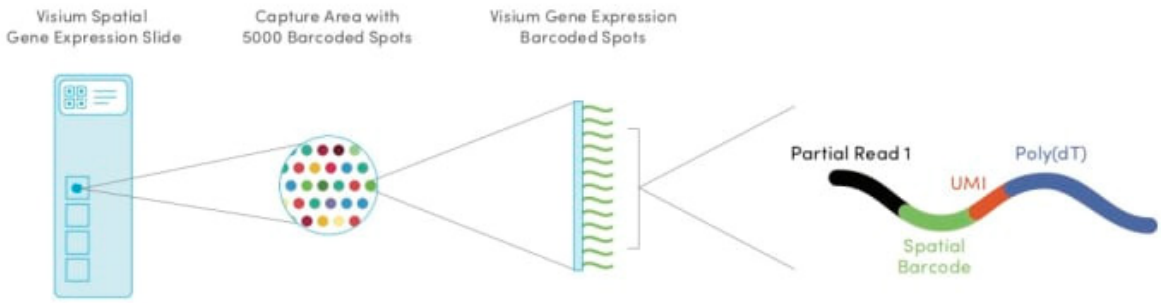
- Special Gel Beads flow in a lane past a channel with cells + enzyme cocktail, then past an oil channel creating nanodroplets.
- Most nanodroplets do not contain any cell; up to ~7% can contain 2+ cells
- Cell lysis and capturing of RNAs occurs within each droplet
- The output of each main lane becomes **one sample**.



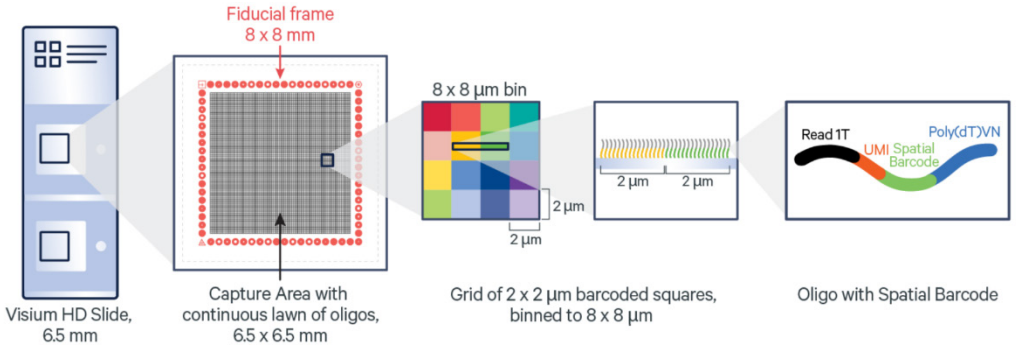
Ortolano, N. The neXt generation of single cell RNA-seq: An introduction to GEM-X technology [Internet]. 10x Genomics, Inc. 2024 Mar 11 [cited 2024 Jun 6]. Available from https://www.10xgenomics.com/blog/the-next-generation-of-single-cell-rna-seq-an-introduction-to-gem-x-technology. Used with permission of 10x Genomics, Inc.

10x Genomics Visium and Visium HD spatial

- Special slides have 2-4 capture areas compose of separated 55 µm "spots" or continuous 2 µm "squares"
- Tissue sections are imaged & placed on capture areas, then permeabilized to release RNAs
- RNAs migrate down to spots/squares below.
- The output of each capture area becomes **one sample**.



Habern, O. Answering Your Questions About the Visium Spatial Gene Expression Solution [Internet]. 10x Genomics, Inc. 2020 Jan 13 [cited 2024 Jun 6]. Available from https://www.10xgenomics.com/blog/answering-your-questions-about-the-visium-spatial-gene-expression-solution. Used with permission of 10x Genomics, Inc.



Habern, O. Your introduction to Visium HD: Spatial biology in high definition [Internet]. 10x Genomics, Inc. 2024 Apr 19 [cited 2024 Jun 6]. Available from https://www.10xgenomics.com/blog/your-introduction-to-visium-hd-spatial-biology-in-high-definition. Used with permission of 10x Genomics, Inc.

10x Genomics terminology (sc and spatial)

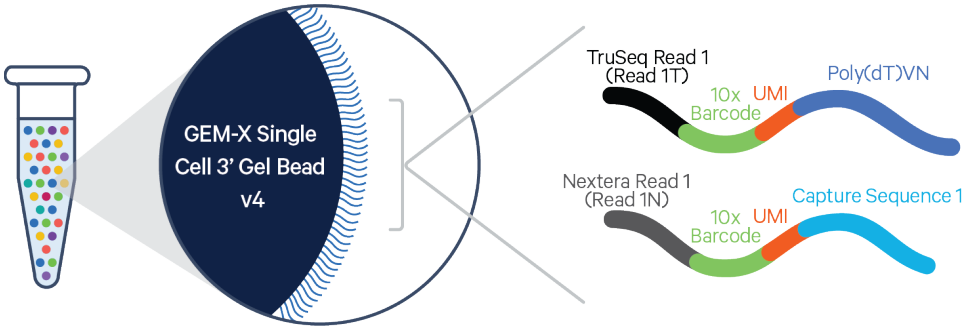
GEMs - Gel Beads-in-emulsions. The gel bead is covered in a lawn of oligonucleotides

Spot/Square - Physical location on a Visium slide covered in a lawn of oligonucleotides

Barcode - Part of the oligo sequence repeated within a GEM bead or slide spot; used to ID which transcripts were in the same GEM or spot/square.

UMI - Unique molecular identifier; UMIs differ between every oligo to uniquely label each original transcript. Used to collapse PCR duplicates.

Capture sequence - Part of the oligo used to capture the species of interest. Usually Poly(dT) for mRNAs.



Ortolano, N. The neXt generation of single cell RNA-seq: An introduction to GEM-X technology [Internet]. 10x Genomics, Inc. 2024 Mar 11 [cited 2024 Jun 6]. Available from https://www.10xgenomics.com/blog/the-next-generation-of-single-cell-rna-seq-an-introduction-to-gem-x-technology. Used with permission of 10x Genomics, Inc.

Anatomy of 10x-Illumina library construct

P5 & P7 - Illumina adapters to attach to the flow cell.

i5 & i7 indexes - Used to mark sequences from the same sample (i.e. each Chromium lane or Visium capture area).

Read 1 - Barcode + UMI

Read 2 - Sequence of the captured transcript

From Chromium Next GEM Single Cell 3' v3.1: Dual Index Libraries Technical Note (CG000325 Rev B), 10x Genomics, Inc. (2021, March 31). Used with permission of 10x Genomics, Inc.

Overview of sequencing outputs

- Multiple samples usually combined together and sequenced in 1+ NovaSeq lane
- Each sequencing lane produces one set of 4 fastq files
  1. R1 - 10x barcode + UMI
  2. R2 - actual transcript sequence
  3. I1 - sample i5 index
  4. I2 - sample i7 index
- Depending on the sequencing center, they may further **demultiplex** into one set of 4 fastq files per sample
  - I1 and I2 no longer needed at this point

Next steps after sequencing - quantification

1. Transcript **read** sequences need to be identified (i.e. which gene each came from)
2. **Reads** grouped by barcode (i.e., GEM/spot/square)
3. **Reads** collapsed to **UMI counts** (discard PCR duplicates)
4. Call cells/spots/squares
    1. SC - Use **UMI counts** to see which GEMs likely contained a cell
    2. Visium - visually find tissue-covered spots/squares
5. Output **UMI counts** per gene per cell/spot/square

Alignment and quantification – 10x pipelines

- Cell Ranger
  - Set of pipelines to process single cell gene expression and all other types of Chromium libraries; aligns to genome using STAR
  - System requirements: 16 cores, 128 GB RAM, 1 TB disk space, 64-bit Linux OS
  - Alternatively, upload to 10x Genomics Cloud Atlas for free*
    - Web page point-and-click interface for most Cell Ranger pipelines
    - *limited number of analyses and downloads; deleted after 6 months
    - Faster run time than most compute clusters (not including up/download)

- Space Ranger
  - Set of pipelines to process Visium and Visium HD libraries; aligns to genome using STAR
  - System requirements: 32 cores, 128 GB RAM, 1 TB disk space, 64-bit Linux OS
  - Support for analyzing Visium HD in 10x Genomics Cloud Atlas coming in 2024
  - Often may need to use Loupe Browser to manually align images and select tissue-covered spots

Both available on IGB's Biocluster

Alignment and quantification – alternative pipelines for single cell data

- Alevin-fry (new simpleaf in rust)
  - Pseudo-align to the transcriptome; runs 30X+ faster
- Kallisto/bustools (new kb-python)
  - Also a pseudo-aligner to transcriptome; runs ~40x faster
  - More focused on downstream trajectory analyses?
- STARsolo
  - Output almost identical to `cellranger count` except no secondary analyses
  - 10X faster than `cellranger count --nosecondary`

Brüning et al. 2022 compare all 3 to Cell Ranger; biggest difference was faster run times

Older versions on IGB's Biocluster

Output from cellranger count

```
├── analysis/
│   ├── clustering/
│   ├── diffexp/
│   ├── pca/
│   ├── tsne/
│   └── umap/
├── cloupe.cloupe                      <- file for Loupe Browser
├── filtered_feature_bc_matrix/         <- filtered UMI counts to read into R
│   ├── barcodes.tsv.gz
│   ├── features.tsv.gz
│   └── matrix.mtx.gz
├── filtered_feature_bc_matrix.h5      <- Filtered UMI counts compressed to Hdf5
├── metrics_summary.csv
├── molecule_info.h5
├── raw_feature_bc_matrix/              <- raw UMI counts if want to call cells on own
│   ├── barcodes.tsv.gz
│   ├── features.tsv.gz
│   └── matrix.mtx.gz
├── raw_feature_bc_matrix.h5
└── web_summary.html                   <- Main summary file
```

Other cellranger xxxxx outputs differ slightly

- 10x Genomics has developed many different products within their Chromium line.
- The Cell Ranger software has many different subcommands, and newer products will use `cellranger multi` or `cellranger vdj` instead.
- The same main files are output, but the web_summary.html changes with not only the subcommand but even within a subcommand between different Cell Ranger versions.
- See Cell Ranger outputs overview for the current output descriptions.

Output from Space Ranger (Visium)

- Also variable depending on Space Ranger version used and which Visium assay used. See output overview page
- Will still get the important `web_summary.html` and `cloupe.cloupe` files.

Loupe Browser

- The last part of 10x's "end-to-end" product offerings
- Free, GUI-interface software to explore Chromium and Visium results in cloupe.cloupe files create by *ranger pipelines; many tutorials
- Always do **first QC check** of the cloupe.cloupe and the web_summary.html
  - Alerts or warnings?
  - Number of cells/spots and number of genes detected?
  - % of reads mapped to genome?
  - First look at clustering of cells/spots
- Can do limited additional analyses:
  - Additional cell filtering based on QC metrics
  - Re-do clustering/tSNE/UMAP
  - Differential expression testing between defined groups
  - Can compare > 1 sample if have run *ranger aggr pipeline
- Cons:
  - "Point-and-click" less reproducible
  - Sub-optimal normalization, although can import other results via .csv files or LoupeR.

Why additional downstream analysis is needed

- Within one sample, no normalization done
- When aggregating >1 sample, only normalization is to down-sample to smallest library size
- Cell calling is liberal by design to not miss cells with naturally low RNA abundance
- QC thresholds for dead/dying cells can vary greatly between tissue and cell types so can't be automated
- Proper statistical methods for > 2 samples not available in Loupe
- Loupe figures not high enough quality for publication

Analysis software options (free)

R-based

- Seurat software suite (will do in lab)
- Bioconductor
  - 347 single cell and 86 spatial packages
  - Amezquita et .al's Orchestrating Single-Cell Analysis with Bioconductor book
  - Crowell and Dong et al.'s Orchestrating Spatial Transcriptomics Analysis with Bioconductor book
  - Righelli et al.'s ISMB 2023 workshop (also now bioc-scrnaseq/)

Python-based

- scverse
  - 76+ total packages (scanpy, Squidpy, scVelo, Squidpy, SpatialData,etc.)
  - 23 tutorials

Warning about cloud or AI based analyses

- Many companies have integrated, “easy to use point-and-click” analyses, including AI-assistance.
- This often involves uploading your data to the cloud or otherwise giving companies access to your data – HIPAA compliance?
- Carefully read the End User Agreement on how they are allowed to use your data!!

Break!