Cancer Center
NCI-DESIGNATED COMPREHENSIVE CANCER CENTER

# Single-cell RNA-seq Preparation and Best Practices

## General considerations to increase the likelihood of high quality data

### Consult Analyst
Engage with an analyst before beginning the experiment.

Discuss goals available samples, batches, etc.

### Define Experimental Variables
Define the experimental question(s) and design experiments to properly test while avoiding confounders. Confirm your experimental groups (e.g. positive controls, genetic change.)

### Prevent Batch Effects
Be careful not to introduce technical variables that cannot be controlled for (e.g. don't process control and experimental groups on separate days.)

### Randomize Sample Order
Minimize batch biases in large experiments by alternating processing of experimental and control group samples.

### Conduct Pilot Studies
Perform preliminary sequencing to confirm quality and positive/ negative controls.

Pilot studies can save money!

### Consult Analyst
Review raw data and results; Discuss data management plan.

### Consult Data Generating Facility
Ensure compliance with sample input policies (e.g., buffers, concentrations) and understand the deliverables from the data generating facility.
Use Genomics.Core@CUAnschutz.edu for the Genomics Shared Resource.

### Ensure Sufficient Sample Size
Verify adequate sample numbers for analysis.

### Avoid Sample Pooling
Pooling erases sample identity and introduces noise. Confounders are impossible to adjust for computationally.

### Collect Metadata
Gather important sample information (e.g. sex, time point, batch.)

### Quality Control
Conduct quality checks throughout data generation. Discuss go/no-go decisions.

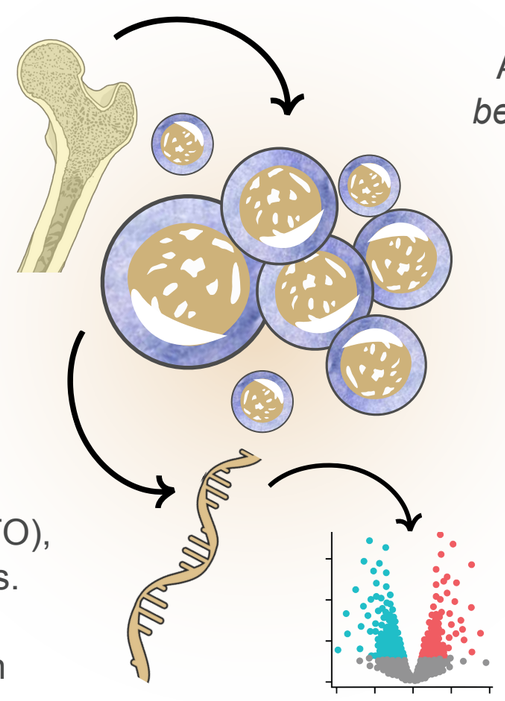
## scRNA-seq captures and sequences the transcriptome of individual cells

Used to explore transcriptional changes between and within cell populations, either at baseline or after treatment, gene KO, etc.

### Define the objective
Outline the biological question, cells of interest, and the primary readouts before preparing samples.

### Plan cell numbers & enrich strategically
Adjust the input cell number to the rarest population desired. Use positive or negative selection (e.g. FACS/microfluidics, bead- or column-based kits) to enrich target populations.

### Consider reducing batch effects (and cost) with sample multiplexing
Current options include on-chip multiplexing (OCM - recommended), fixed RNA-profiling (Flex), hashtag oligos (HTO), cell multiplexing oligos (CMO), and genotype-based methods. Confirm the approach is compatible with the cell barcode chemistry, and understand that some methods can result in experimental failures.



### Sample viability drives data quality
Aim for >80-90% viability per sample and optimize each step *before* sequencing. Cell membrane integrity is critical to produce robust and interpretable results.

### Deplete dead/dying cells and remove debris
Use positive or negative selection (e.g. FACS/microfluidics, EasySep/Annexin V) to remove dead cells.
Clear debris and cell aggregates with extra large volume washes and cell strainer/filters (e.g. MACS Smart Strainer.)

### Calculate cell concentrations accurately
Avoid skewed target recovery by calculating stock concentrations with the *total* cell count (not just viable cells.)
Count the final suspension 2-3x;
individual counts can vary as much as 25%.

| Library Type | 3'/5' scRNA-seq | Fixed (Flex) scRNA-seq | Long read scRNA-seq |
|---|---|---|---|
| **Protocol** | Droplet-based \| UMIs (10x Genomics) | Droplet-based \| UMIs (10x Genomics) | Plate-based \| full-length transcript (SmartSeq) |
| **Gene Features** | Whole-transcriptome (3' or 5' mRNA) | Probe-based (option to include custom probes) | Whole-transcriptome (includes non-coding molecules) |
| **Cells Captured** | <20k cells / sample | <20k cells / sample | 100-1k cells / sample |
| **Sequencing Depth** More reads may be needed for lowly expressed genes or RNA-rich samples (e.g. cell lines) | 50-100k reads / cell | 10-40k reads / cell | >1M reads / cell |
| **Multimodal Options** | ATAC, CITE/protein, TCR/BCR, CRISPR | CITE/protein, CRISPR | Generally not compatible |
| **Common Analyses (beyond clustering)** ≥2 biological replicates are recommended (more for *in vivo*) | Differential expression<br>Differential abundance<br>Trajectory inference<br>RNA velocity<br>VDJ clonotypes<br>CNV/CNA inference<br>Cell-cell communication | Differential expression<br>Differential abundance<br>Trajectory inference<br>CNV/CNA inference<br>Cell-cell communication | Alternative splicing & isoform detection<br>Differential expression<br>Differential abundance<br>Trajectory inference<br>CNV/CNA inference<br>SNP/variant calling<br>Cell-cell communication |

### Before sequencing
- [ ] Discussed experiment with an analyst & sequencing facility
- [ ] Defined target cells & questions
- [ ] Finalized sample multiplexing stragey
- [ ] Budgeted for throughput and sequencing depth
- [ ] Tissue dissociation is optimized
- [ ] Samples measure >80-90% viable
- [ ] Depleted dead / dying cells
- [ ] Removed debris / cell aggregates
- [ ] Calculated 2-3 cell counts that are <25% variable

### After sequencing
- [ ] Remove empty droplets / barcodes
- [ ] Identify doublets / multiplets
- [ ] Filter low quality cells and samples
- [ ] Measure ambient RNA (correct if needed)

*Some analyses will require experience with a coding environment (R/Python)*

## Contact the Biostatistics & Bioinformatics Shared Resource (BBSR) for cancer-related analysis

James.Costello@CUAnschutz.edu, Bioinformatics@CUAnschutz.edu

## Contact CUBE for all non-cancer-related projects

scRNA-seq v2.2