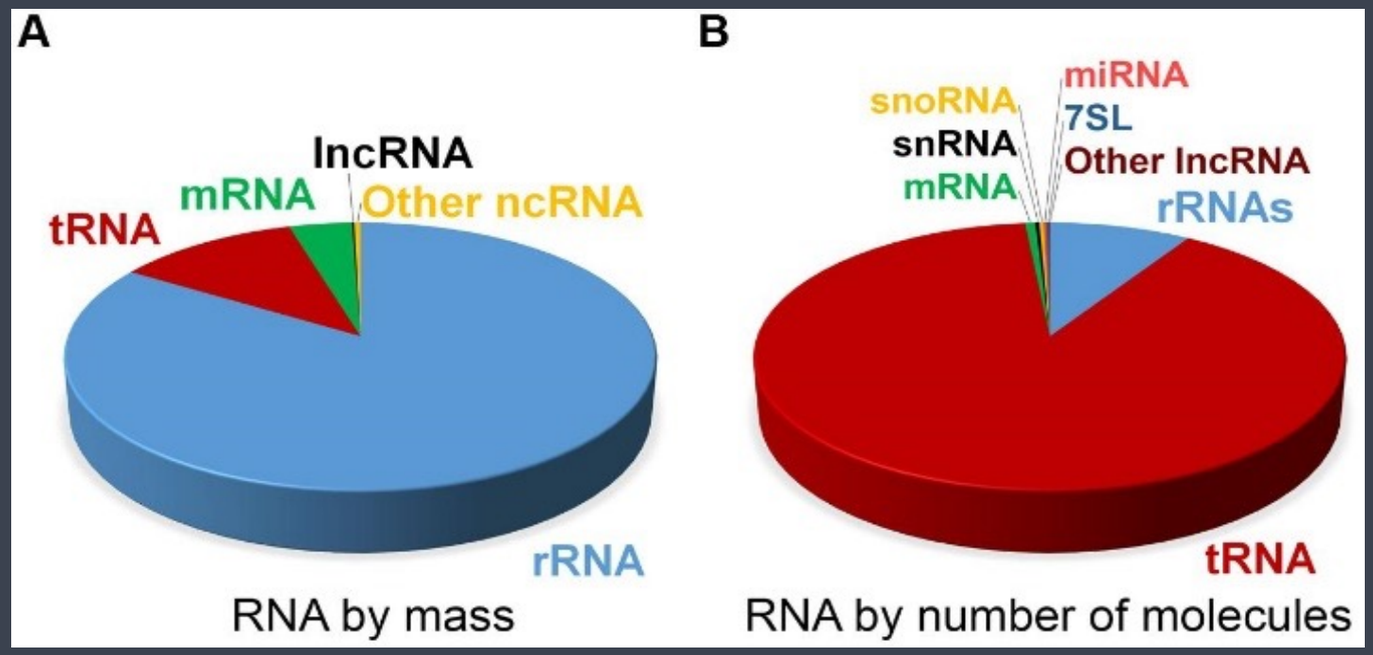
Practical strategies for sequencing

Sarah Boswell, PhD

Radhika Khetani, PhD

Poly(A) enrichment or ribosomal RNA depletion?

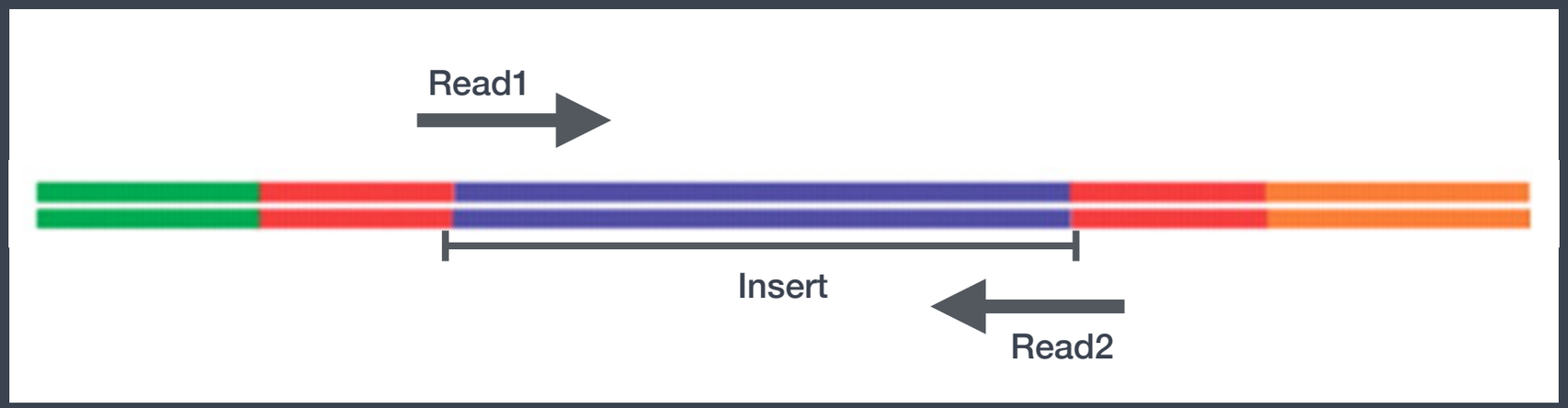


Depends on which RNA entities you are interested in…

- For differential gene expression, it is best to enrich for Poly(A)+
  - EXCEPTION – If you are aiming to obtain information about long non-coding RNAs, then do a ribosomal RNA depletion.

Single-end or Paired-end data?

Depends on your goals, paired-end reads are better for reads that map to multiple locations, for assemblies and for splice isoform differentiation.



- ✓ SE - Single end dataset => Only Read1
- ✓ PE - Paired-end dataset => Read1 + Read2
  - can be 2 separate FASTQ files or just one with interleaved pairs

Single-end or Paired-end data?

Depends on your goals, paired-end reads are better for reads that map to multiple locations, for assemblies and for splice isoform differentiation.

- ✦ For differential gene expression, which one you pick depends on-
  - If you are specifically interested in isoform-level differences
  - The abundance of paralogous genes in your system of interest
  - Your budget, paired-end data is usually 2x more expensive

How much sequencing data to collect?

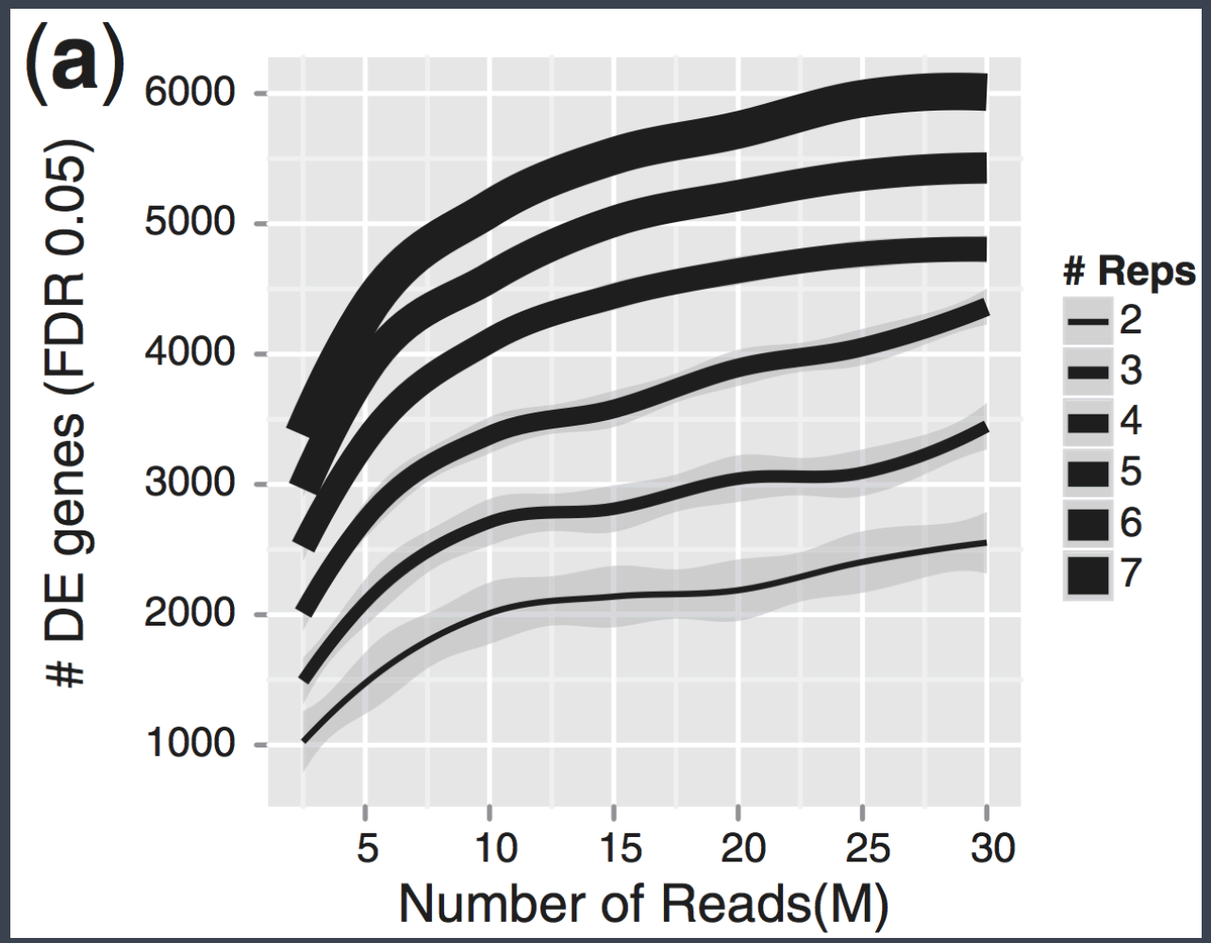
For human samples - minimum 20 million reads/sample

- Modify that number based on the size of your transcriptome (species)
- Increase if looking for low-expression genes
- Increase if looking for splice-isoforms (and get PE data)
- Increase if performing rRNA depletion instead of PolyA selection
- Modify based on number of replicates.

  More replicates >> More reads!!

How much sequencing data to collect?

More replicates >> More reads!!



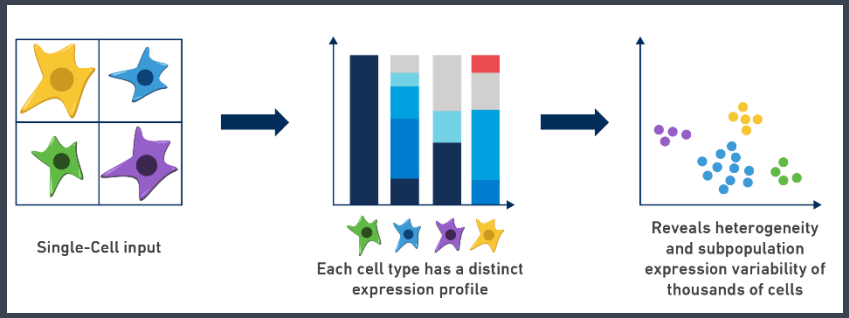
Liu, Y., et al., Bioinformatics (2014) 30(3): 301–304

Sequencing Strategies for scRNA-seq

- Number of samples you can multiplex and number of cells needed to sequence depends on your experimental question.
- Standard is 20,000-60,000 reads per cell
- One NextSeq run will get ~450 M reads total.
  - If you are sequencing 3,000 cells per sample that is 3-8 samples per sequencing run.

Sequencing Strategies for scRNA-seq

- Required number of cells increases with complexity of the sample.
  - Goal of 50-100 cells per expected cell type



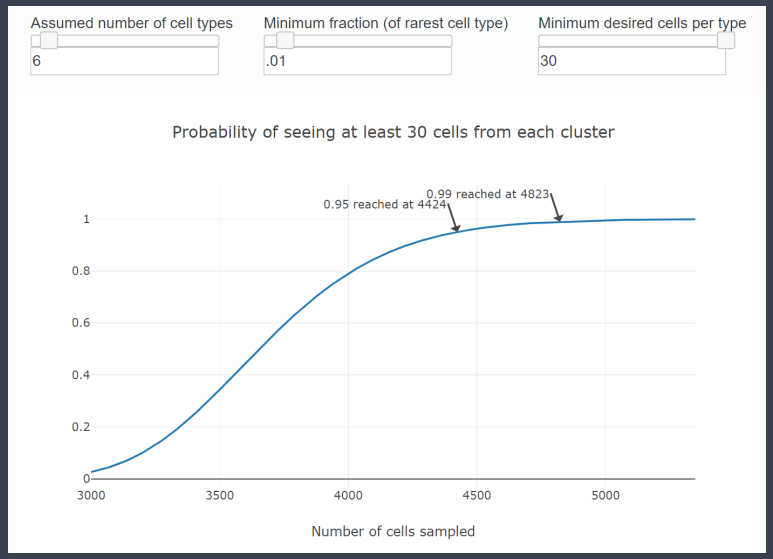
- As the number of genes involved in the biology decrease then the read requirements increase.
  - Cells that have fewer differences are harder to tell apart.

Hague et al. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications, Genome Med 2017 (DOI: 10.1186/s13073-017-0467-4)

JA. Griffiths et al. Using single-cell genomics to understand developmental processes and cell fate decisions, MSB (2018) (DOI 10.15252/msb.20178046)

https://satijalab.org/howmanycells

- To identify a population you should aim for 50-100 cells of that type.



- If your looking for a cell that is 1% of your population need to sequence 5,000-10,000 cells.

Pilot your scRNA-seq Experiment

- For scRNA-seq a pilot project is important for ensuring the correct sequencing strategy.
- Duplicate samples from condition A & B.
- Pilot experiment will inform you about:
  - Sample prep quality
  - Number of cells
  - Reads per cell

Sequencing Strategies for scRNA-seq

- Cell-type classification requires lower read depth only 10,000-40,000 reads per cell.
- Suggest starting with ~35,000 reads per cell.
- The more cells you encapsulate the more sequencing needed to read out the cells.
- You can always re-sequence samples.

Hague et al. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications, Genome Med 2017 (DOI: 10.1186/s13073-017-0467-4)

JA. Griffiths et al. Using single-cell genomics to understand developmental processes and cell fate decisions, MSB (2018) (DOI 10.15252/msb.20178046)