Frederick National Laboratory
for Cancer Research

sponsored by the National Cancer Institute

Single Cell RNA-Seq at CCR-SF: Sample prep and best practices using 10X Genomics

Monika Mehta
R&D Scientist, CCR Sequencing Facility

Nov 14th, 2017

DEPARTMENT OF HEALTH AND HUMAN SERVICES • National Institutes of Health • National Cancer Institute

Frederick National Laboratory is a Federally Funded Research and Development Center operated by Leidos Biomedical Research, Inc., for the National Cancer Institute

CCR Sequencing Facility

Core facility dedicated to providing Next-Generation Sequencing (NGS) services to Center for Cancer Research investigators, operating one Illumina HiSeq 4000, two HiSeq 2500, three NextSeq 500, one MiSeq, one PacBio RSII / Sequel, 10X Chromium and BioNano Irys system.

DNA/RNA-Protein Interaction

Transcript Profiling and Discovery

Illumina

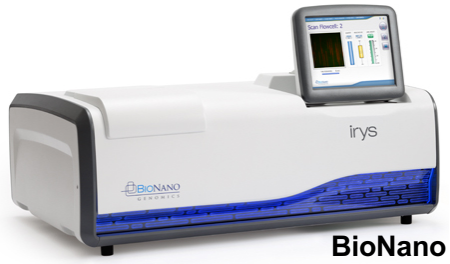
Genome-wide Methylation study

Targeted Exome sequencing

Whole Genome Structural Variants

Single Cell Gene Expression Profiles

10X Chromium



BioNano Irys

Next-Generation Mapping

De Novo Whole Genome Sequencing

Targeted Sequencing

PacBio RS II

Base Modifications

DNA Structure Variation analysis

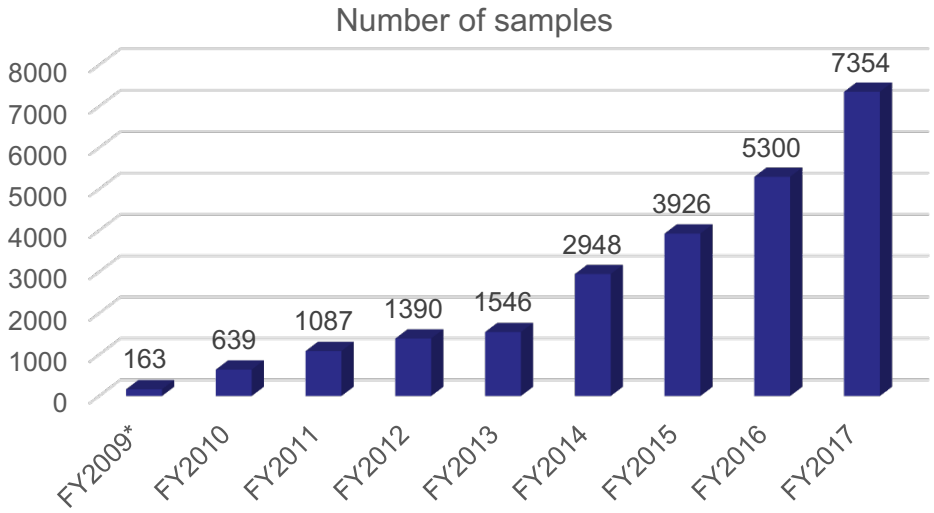
PacBio Sequel

Transcript Isoform Detection

CCR Sequencing Facility

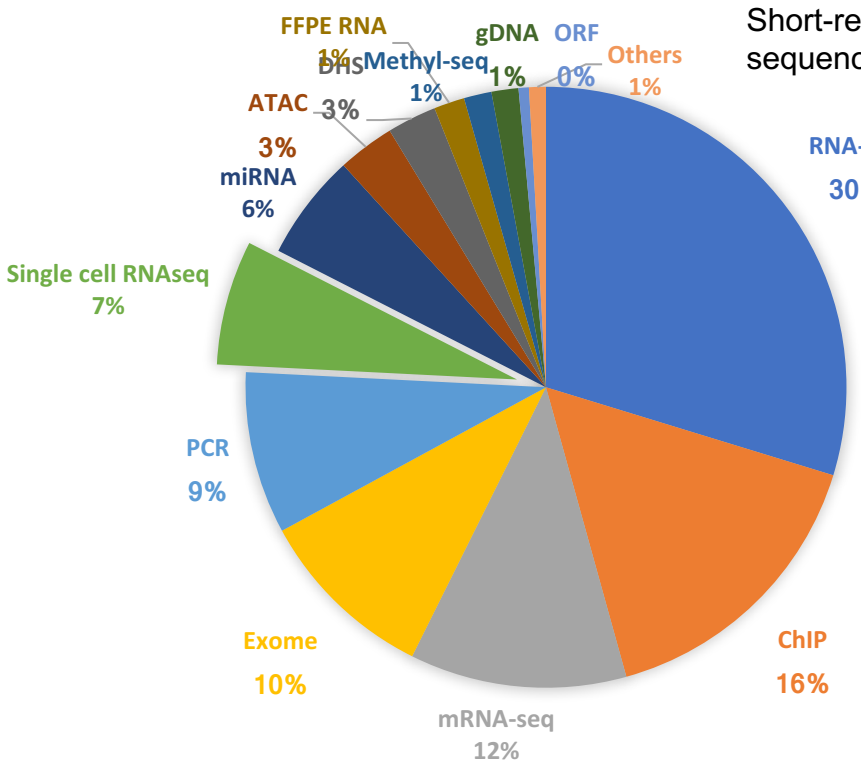
FY2017 Statistics:

- 120 NCI investigators
- More than 7300 samples.
- Delivered over 45 trillion bases



Sample types – FY2017

Short-read sequencing



Single cell RNA-Seq:

Fluidigm C1

10X Chromium

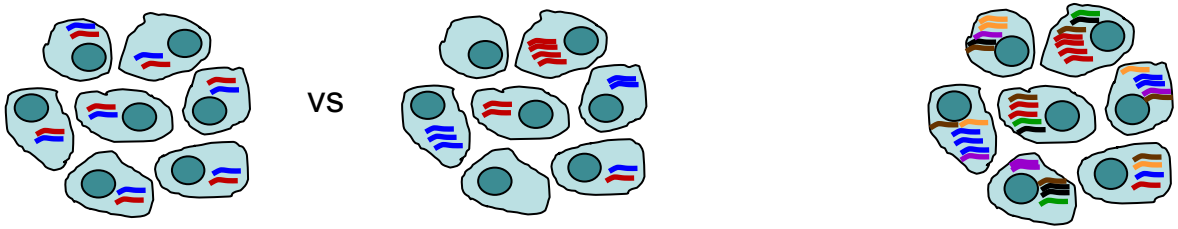
DropSeq

Others

FFPE Exome	MNase
RIP-Seq	10X gDNA
RiboSeq	shRNA
Drip	sgRNA
HiC	VirScan

Single-cell mRNA-seq

- Cell-to-cell variation in tissues and populations – normal, and important for biological functions, under normal and diseased states.
- Frequently manifested through changes in gene expression.
- Single-cell studies: Help to uncover the variation masked by the average of heterogeneous subpopulations, identification of rare populations.



.. for hundreds of genes

- Crucial in understanding cancer cells, stem cells, immune cells, cell lineages etc.

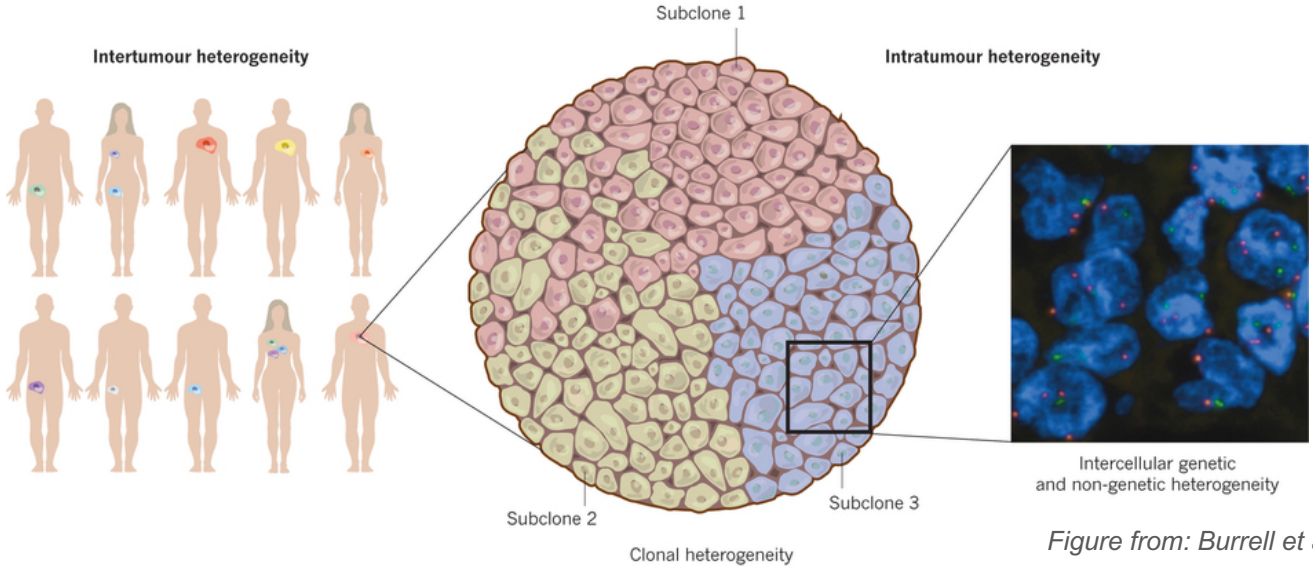
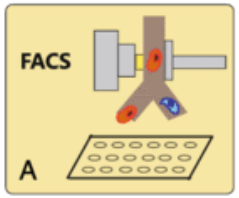
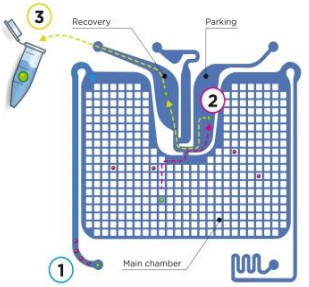


Figure from: Burrell et al., 2013. Nature, 501: 338

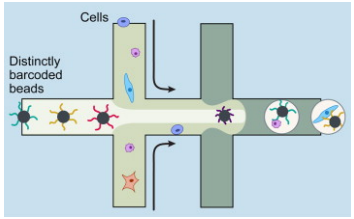
Single-cell mRNA-seq: Gene expression profiling at single cell level

A. Single cell-per-well protocols (low-throughput, full-length transcript analysis possible)
   A. FACS-based methods
   B. Fluidigm C1
   C. DEPArray





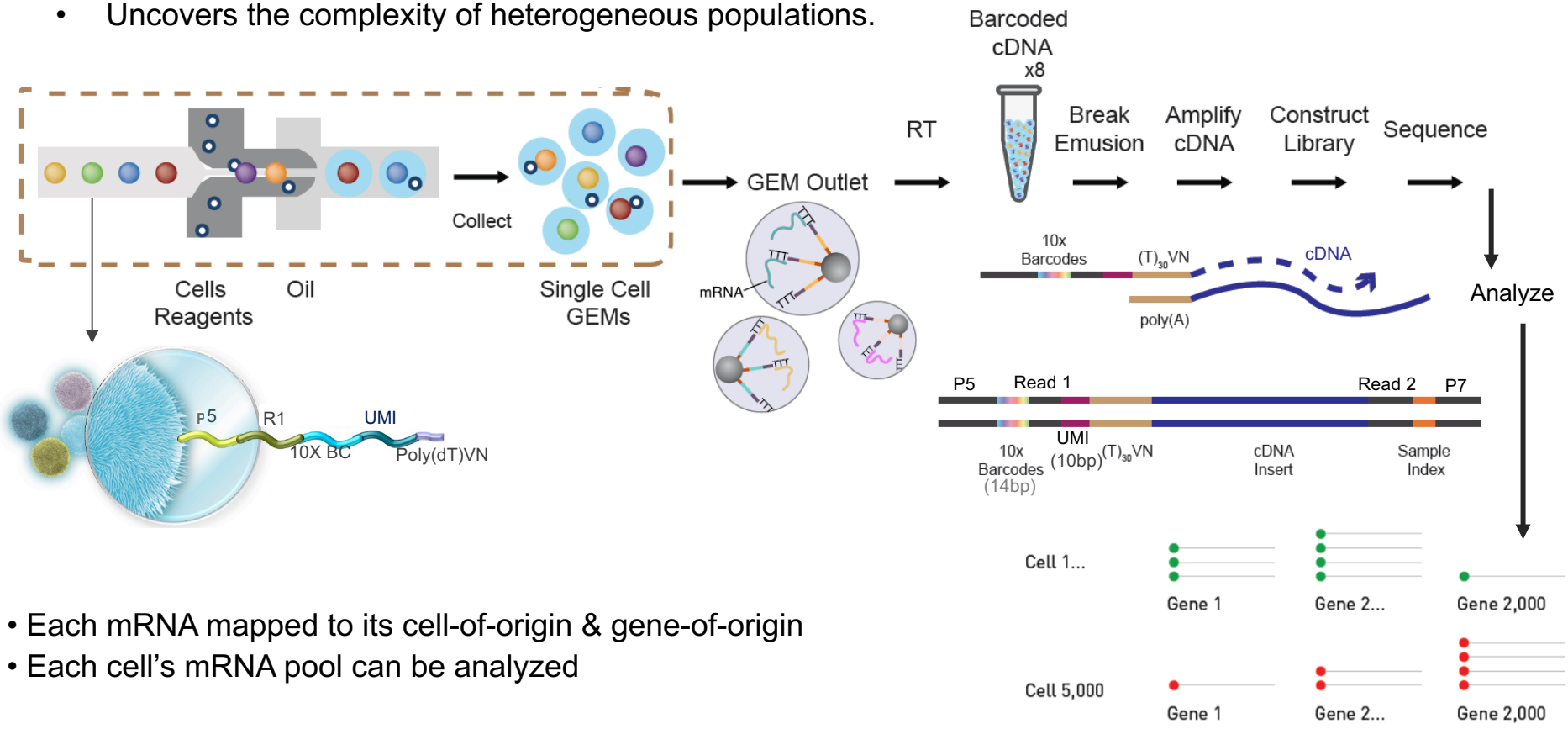
B. Droplet-based massively parallel protocols (high-throughput, expression analysis only)
   A. DropSeq
   B. 10X Genomics Chromium
   C. ddSeq
   D. BD Rhapsody



SF: lib prep/sequencing services for all approaches, end-to-end (sample to sequence) service provided for **10XChromium Single-cell RNAseq.**

Single cell 3’ mRNA-Seq using 10X Chromium

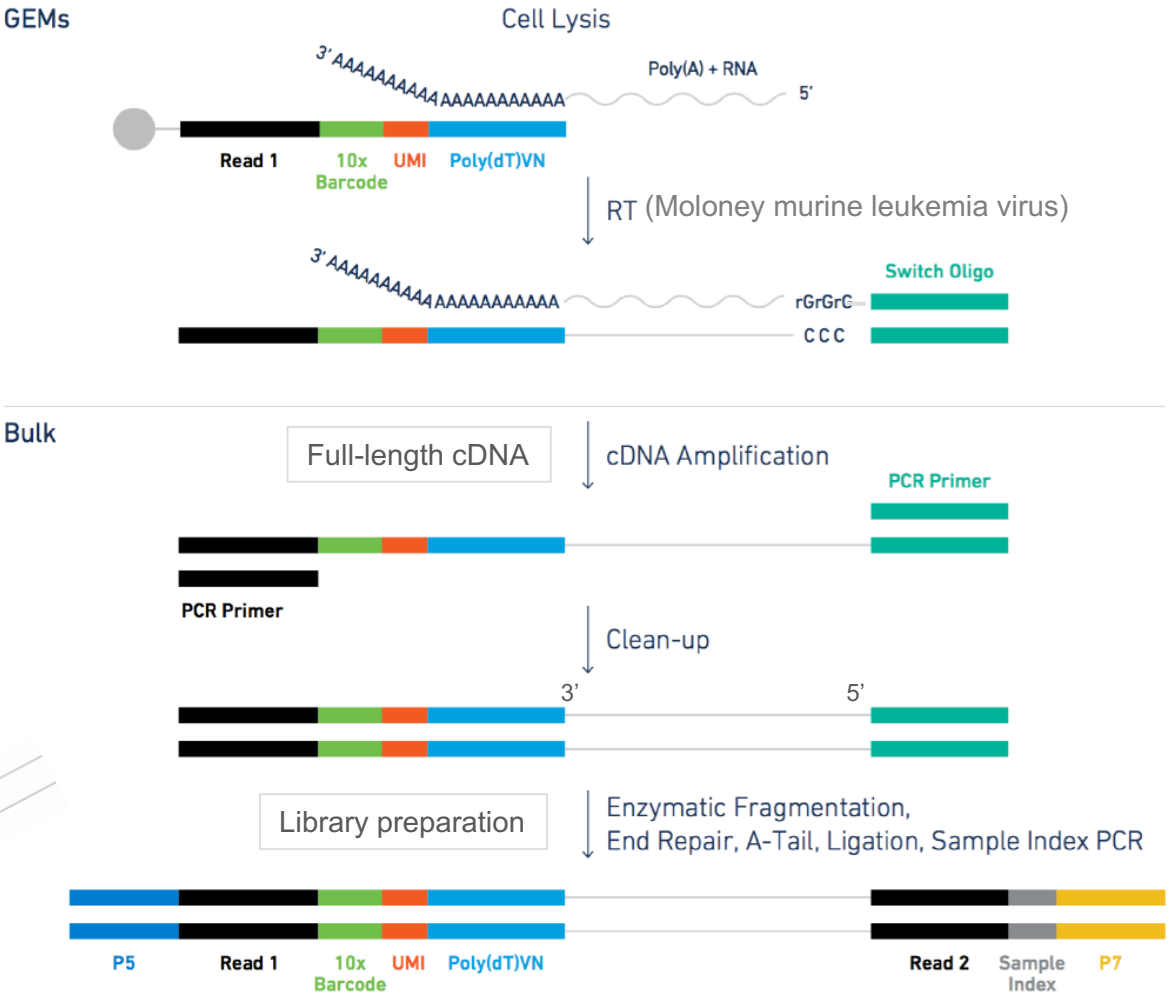
- Based on droplet microfluidics technologies such as Drop-Seq and inDrop
- Combines microfluidics with molecular barcoding to enable high-throughput single-cell RNA sequencing
- Transcriptional profiling of thousands of individual cells by partitioning into nanoliter-scale Gel Bead-In-Emulsions (GEMs)
- Uncovers the complexity of heterogeneous populations.



- Each mRNA mapped to its cell-of-origin & gene-of-origin
- Each cell’s mRNA pool can be analyzed

http://www.10xgenomics.com/single-cell/

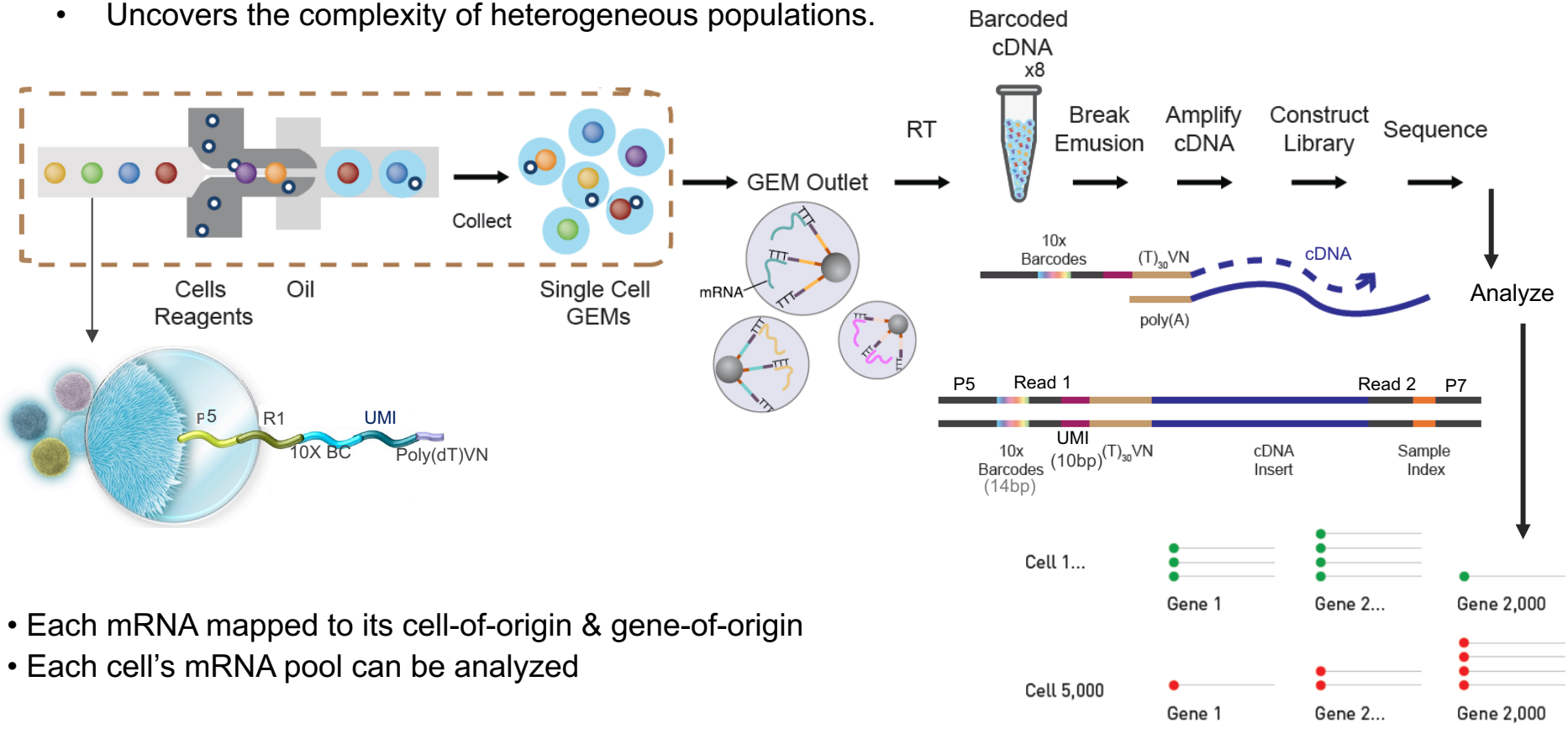
Generation of full-length cDNA using template switching and 3’ sequencing



Sequencing adaptors, Read 1 and Read 2, Barcodes and Sample Index

Single cell 3’ mRNA-Seq using 10X Chromium

- Based on droplet microfluidics technologies such as Drop-Seq and inDrop
- Combines microfluidics with molecular barcoding to enable high-throughput single-cell RNA sequencing
- Transcriptional profiling of thousands of individual cells by partitioning into nanoliter-scale Gel Bead-In-Emulsions (GEMs)
- Uncovers the complexity of heterogeneous populations.



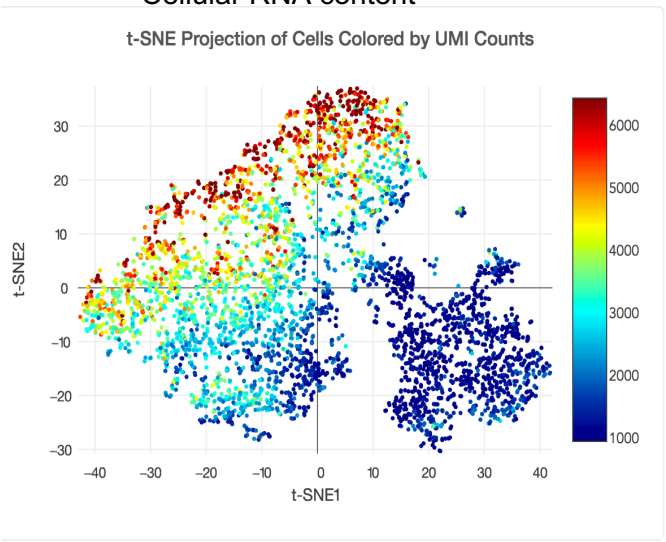
- Each mRNA mapped to its cell-of-origin & gene-of-origin
- Each cell’s mRNA pool can be analyzed

http://www.10xgenomics.com/single-cell/

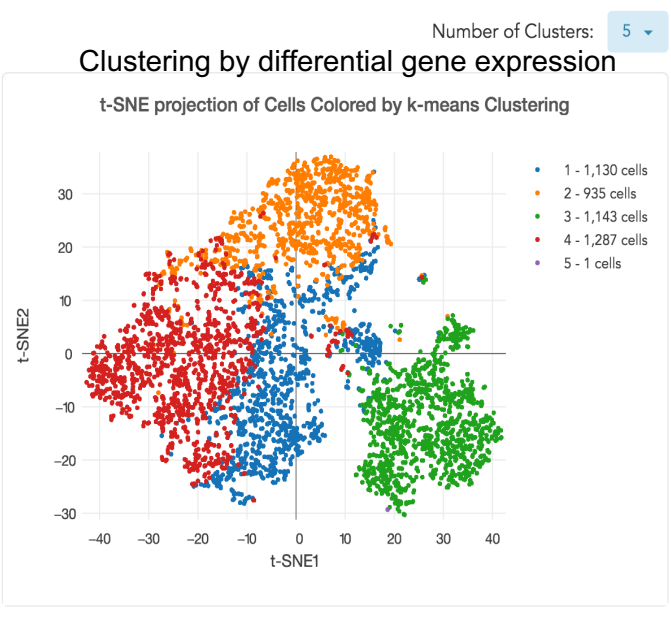
Single cell 3’ mRNA-Seq using 10X Chromium

Examples of clusters generated by 10X Chromium

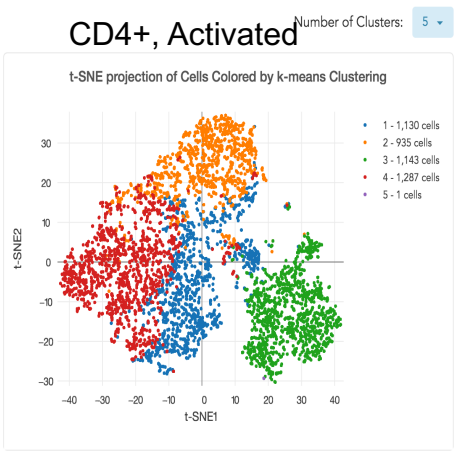
Cellular RNA content



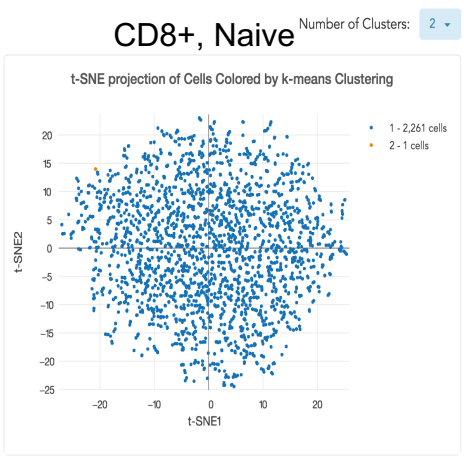
Clustering by differential gene expression



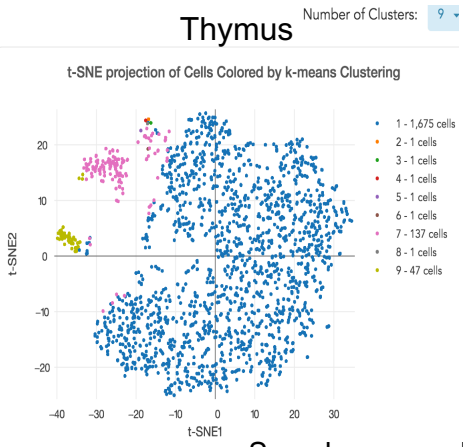
CD4+, Activated



CD8+, Naive



Thymus



Sample source: Dr. Remy Bosselut’s group

Single cell 3’ mRNA-Seq using 10X Chromium

Top differentially expressed genes per cell cluster (UMI counts/cell)

Gene name	Cluster 1	Cluster 2	Cluster 3
Abc1	28.52	0.03	14.7
Xyz2	4.56	8.33	30.85
Fgh3	8.94	17.44	1.27

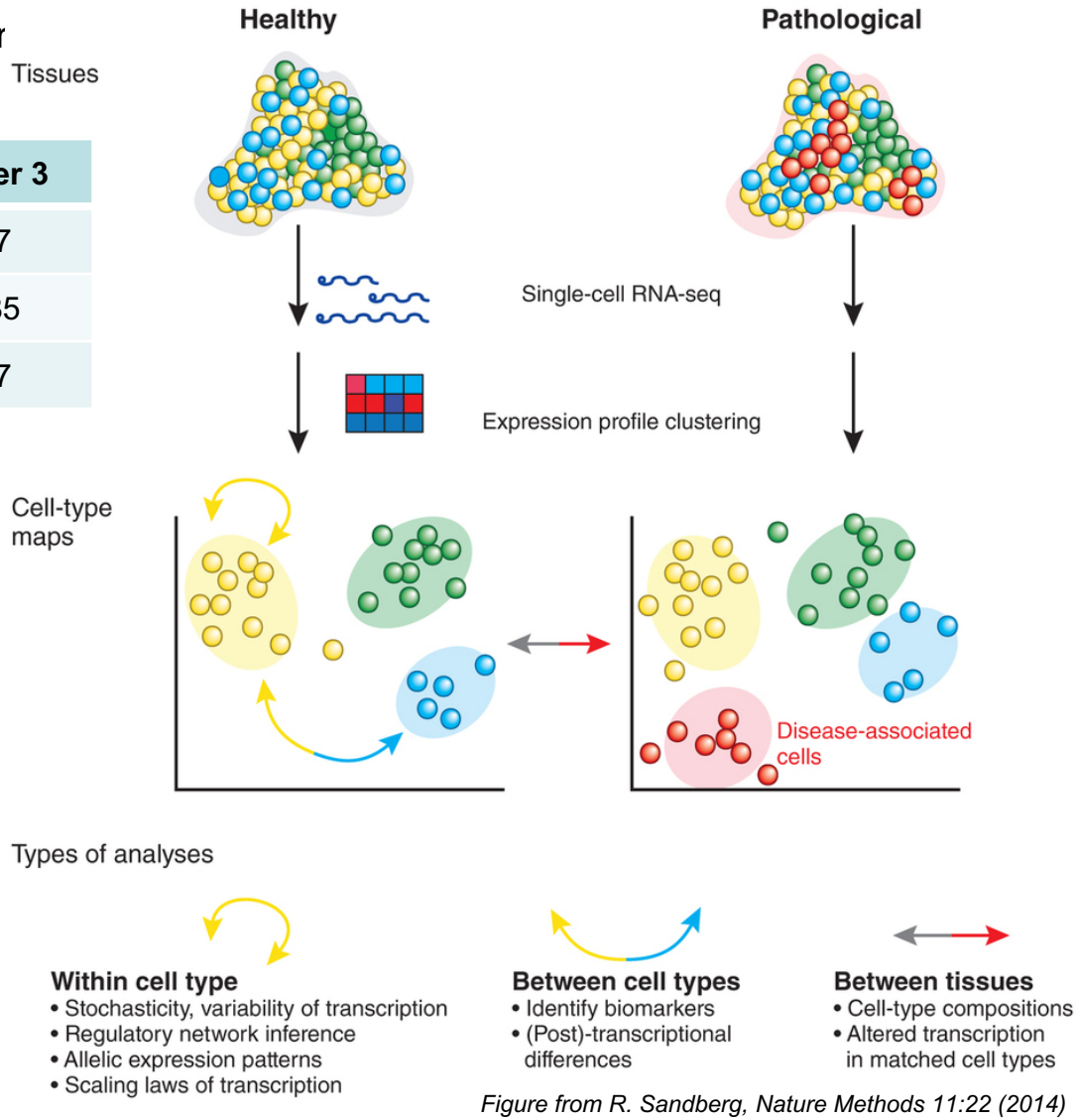


Figure from R. Sandberg, Nature Methods 11:22 (2014)

10X Chromium single-cell projects at CCR-SF

10X Chromium single cell projects at SF	
Number of labs	13
Total number of samples processed	123
Data delivered	101
Total number of cells sequenced	145,982
Median number of cells / sample	1020
Median genes / cell	1,475
Median transcripts / cell	3,650
Total genes detected / sample (Median)	14,914

- Data analyzed by CCR-SF IFX team

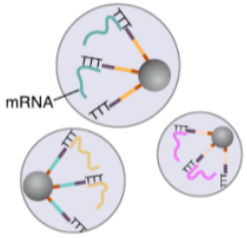
Different cell types processed at CCR-SF:

1. Human and mouse cell lines (HeLa cells and MEFs)
2. B cells (naïve and memory B cells)
3. Cells from the dermis layer of skin
4. T cells (naïve and activated, CD4+, CD8+, & thymus cells)
5. Stem cells (bone marrow cells)
6. Cutaneous MAIT (mucosal-associated invariant T) cells
7. CAR-T cells (cultured, activated)
8. Epithelial cells trypsinized from mouse skin
9. Tumor cells (Prostrate cancer) from organoid cultures
10. Cells from patient urine samples
11. Hepatic cancer patient-derived frozen tumor cells
12. Thymic epithelial cells from WT and mutant mice
13. Neuronal cells (cultured)
14. Single nucleus sequencing from frozen post-mortem brain tissues.

How many cells?

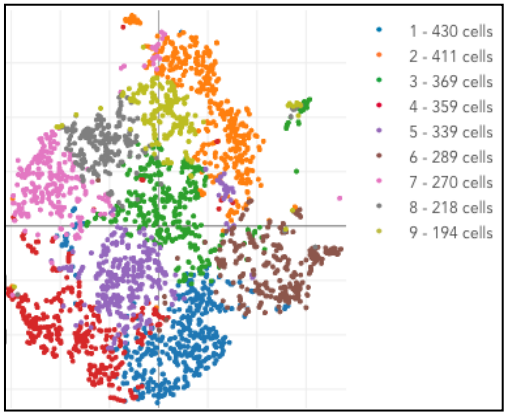
~ 10-15% transcripts captured per cell

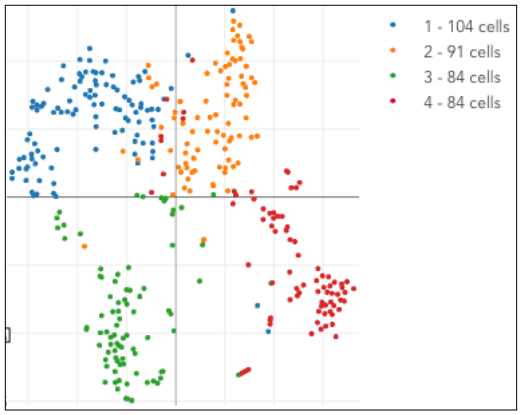
- PolyA capture
- RT conversion within the cell lysate



More cells – higher likelihood of more representative data

More cells – more confidence in heterogeneity analysis

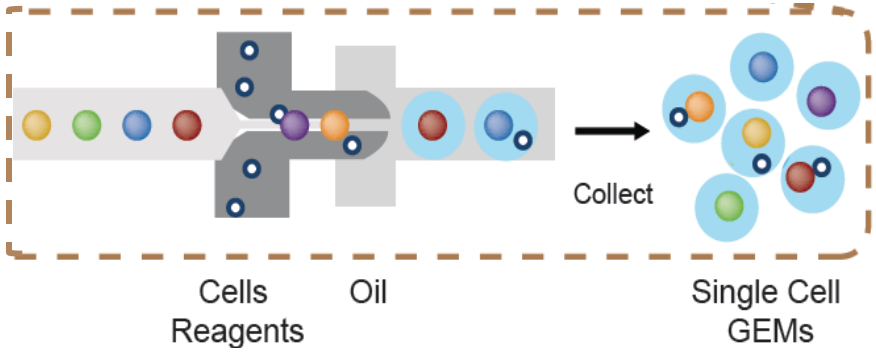




More cells – more probability of multiplets

10X Chromium:

- ~65% cell processing efficiency
- 500-10,000 cells per channel, ~0.4-7.6% doublet rate



How many cells?

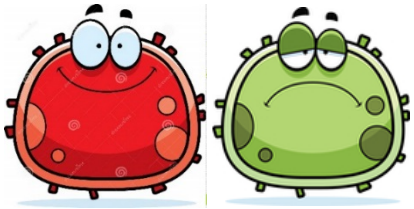
Multiplet Rate (%)	# of Cells Loaded	# of Cells Recovered*
~0.4%	~870	~500
~0.8%	~1700	~1000
~1.6%	~3500	~2000
~2.3%	~5300	~3000
~3.1%	~7000	~4000
~3.9%	~8700	~5000
~4.6%	~10500	~6000
~5.4%	~12200	~7000
~6.1%	~14000	~8000
~6.9%	~15700	~9000
~7.6%	~17400	~10000

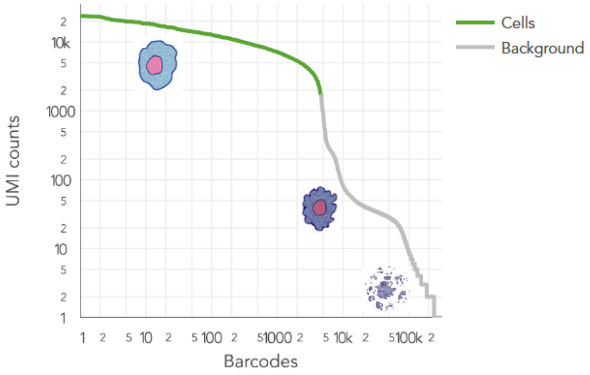
* For good quality samples

- Accurate cell count
- Single cell suspensions – no aggregates
- Enough cells for washes – removal of debris, nucleic acids, RT inhibitors

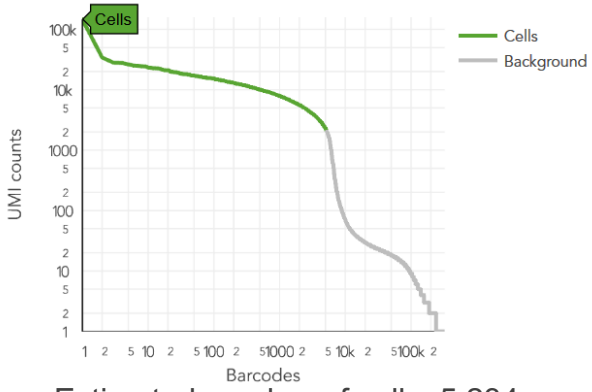
Quality matters!

- Low viability / dead cells
  - mRNA degraded, excluded from analysis – data from fewer cells
  - mRNA enters other GEMs – incorrect data
  - Unhealthy/dying cells may interfere with subpopulation analysis
- Ideal viability > 90%; >70% is good
- Fresh cells > Frozen; some cell types more susceptible to degradation
- Removal of dead cells, if enough number of cells available

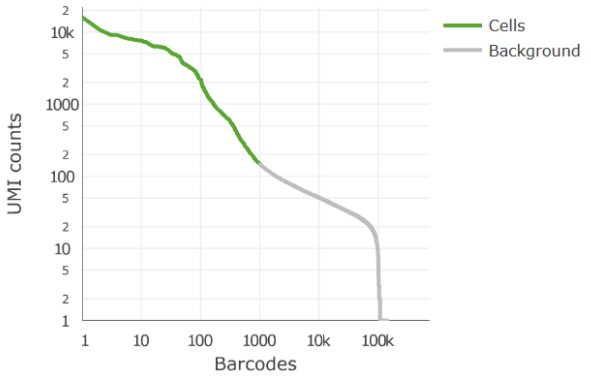




Estimated number of cells: 4,601
Estimated viability: ~87%



Estimated number of cells: 5,234
Estimated viability: ~90%



Estimated number of cells: 1,011
Estimated viability: ~20%

- Tips: Gentle handling of the cells; low speed spins, wide bore tips, maintain cells in ice, shorter processing time, no RT inhibitors

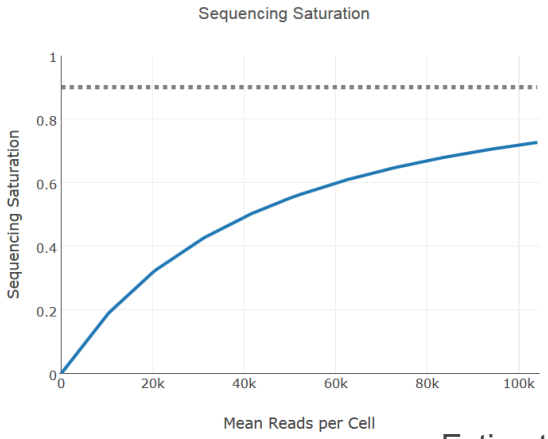
Sequencing Depth

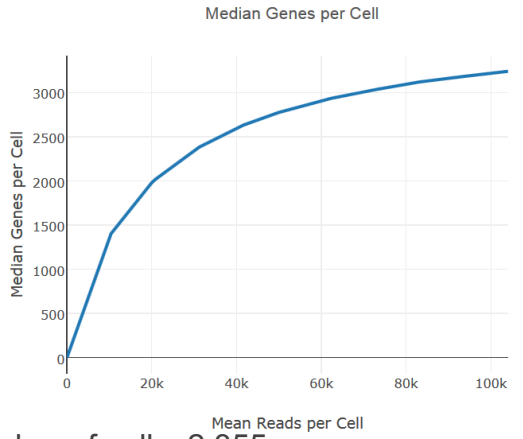
Depends on

- Cell type: More RNA vs less RNA per cell
- Cell number: More cells, more reads
- Cell quality: More healthy cells, more reads

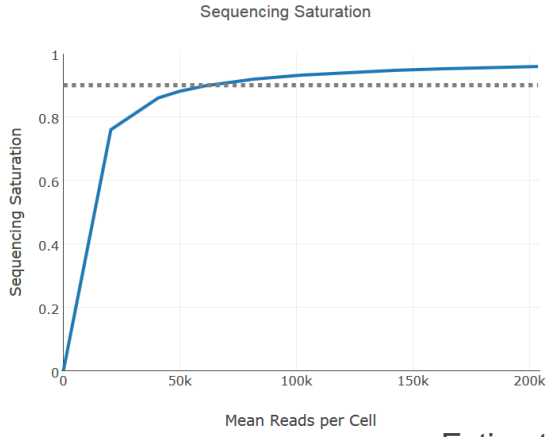
Typical: 100,000 reads per cell

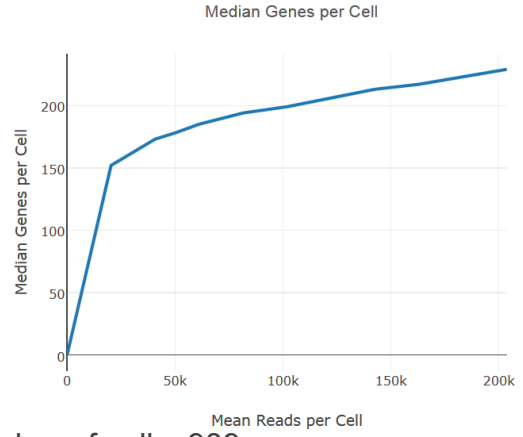
- 2-3 samples per NextSeq lane (SR)





Estimated number of cells: 2,355
Median genes per cell: 3,241





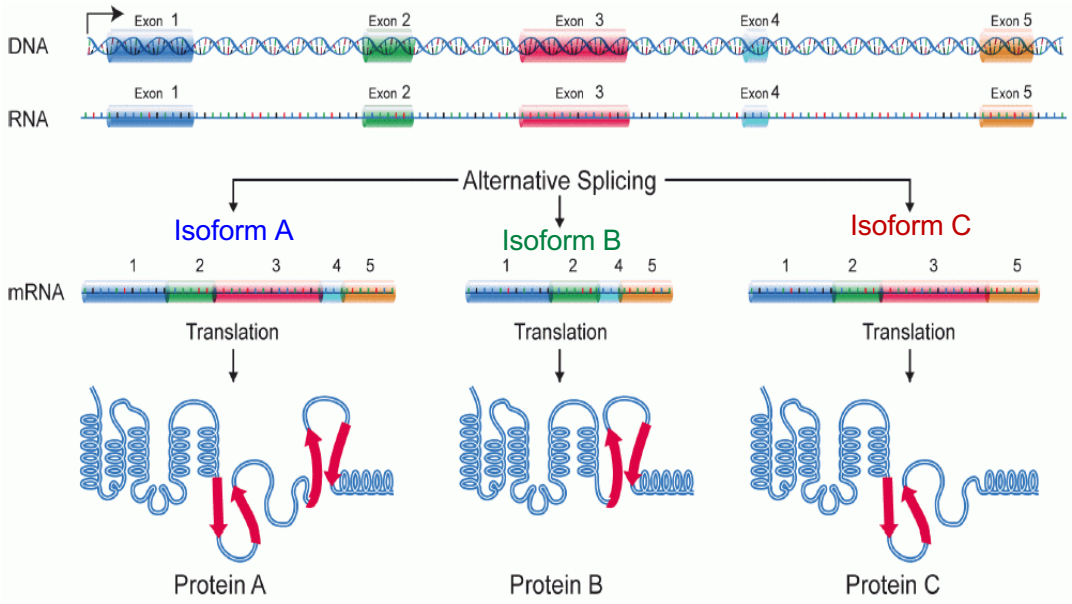
Estimated number of cells: 989
Median genes per cell: 484

Tips

- Cells should have high viability
- Remove dead/unhealthy cells, if possible
- Cell number: enough to enable at least a couple of washes and cell counting before loading
- Cell handling: as gentle as possible. Use low speed spins, wide bore tips
- Minimize cell stress: maintain cells in ice
- Processing time: as short as possible. Capture as quickly after extraction as possible.
- If delays are inevitable, keep cells in the medium they are happy in
- Fresh cells better than frozen.
- Single cell suspensions should be free from cell aggregates and debris
- No RT inhibitors in the cell suspensions
- Accurate cell count

Coming soon – Full-length RNA-seq from single cells

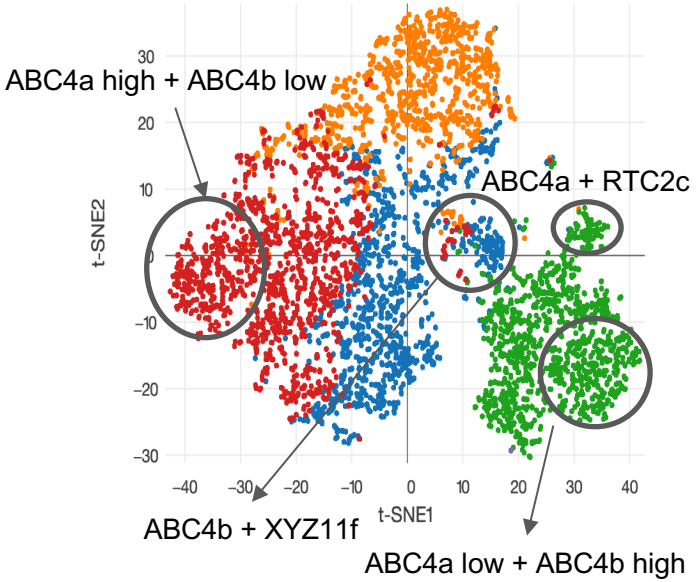
Alternate Splicing: Increases informational diversity, regulates various cellular processes, commonly dysregulated in cancer, thought to result in functional diversity of most cancer-relevant genes.



Subpopulation 1 (20% cells): *ABCa* + *ABCb*

Subpopulation 2 (50% cells): *ABCa* + *ABCc*

Subpopulation 3 (30% cells): *ABCb*

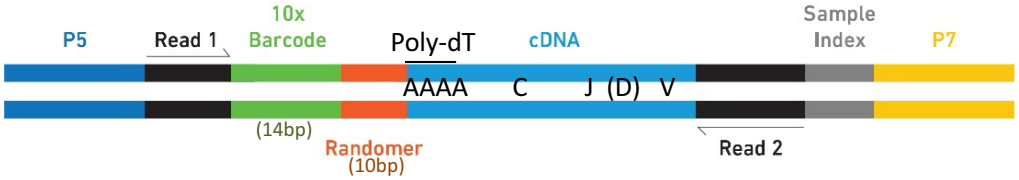


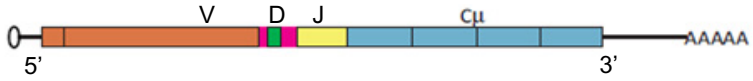
- Using long-read sequencing technology to obtain full-length sequence and isoform information from single cells

Coming soon – Analysis of V(D)J rearrangement in B- and T-cells

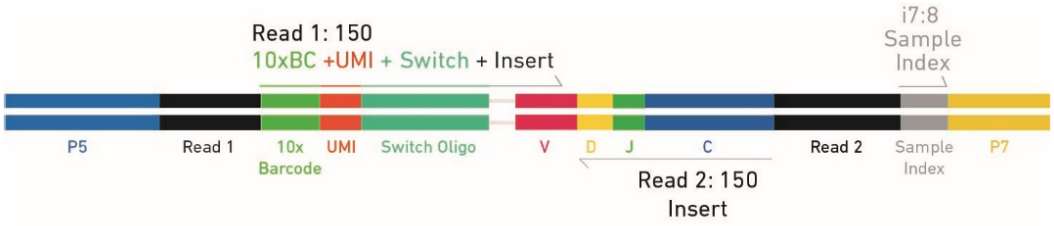
- Single cell sequencing to reveal the true clonality and diversity of the immune repertoire
- Simultaneous analysis of a single sample for cellular heterogeneity and phenotype, as well as T-cell receptor and B cell immunoglobulin repertoires.

3' Gene Expression Library Structure:

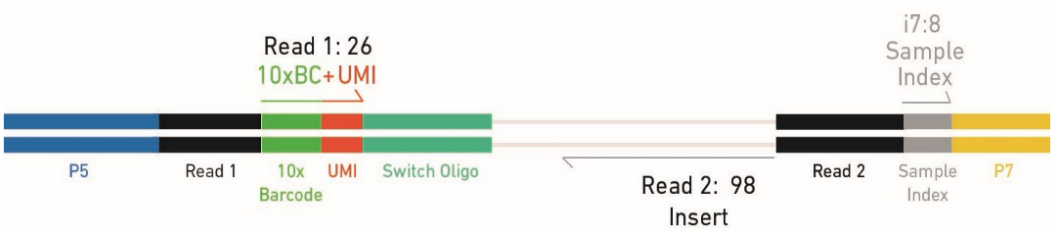




V(D)J Enriched Library Structure:

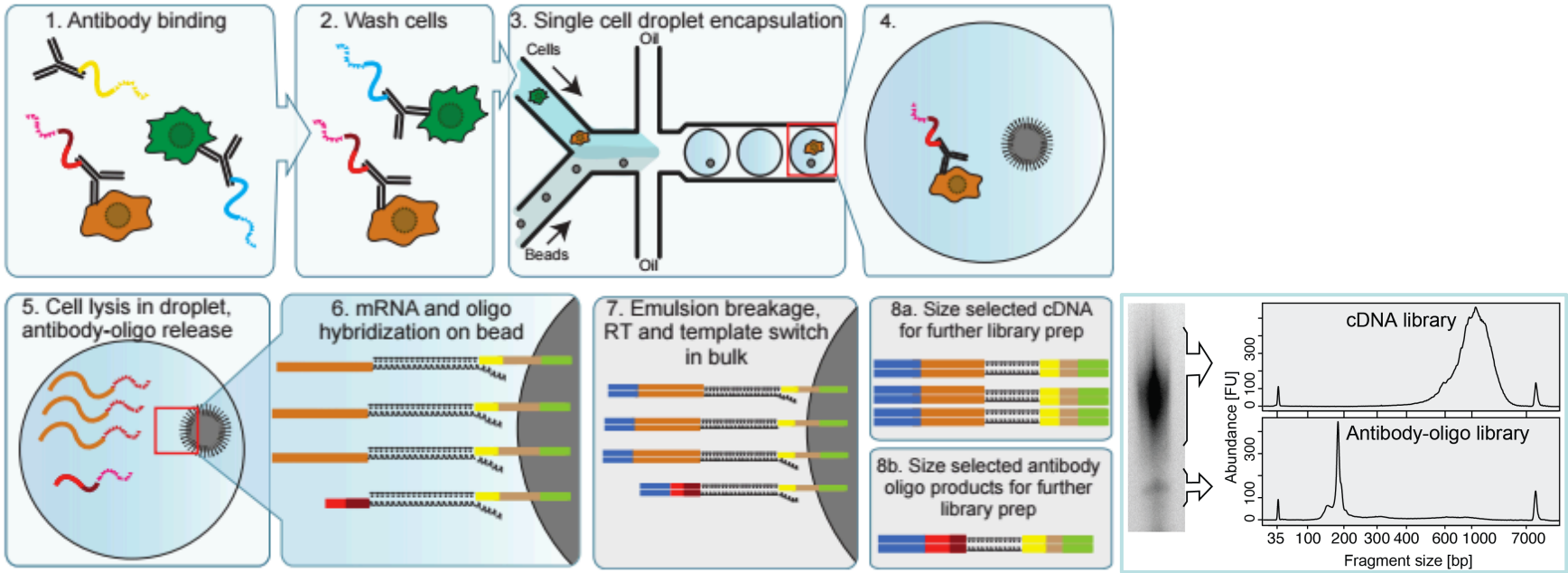


5' Gene Expression Library Structure:



Coming soon – CITE-seq (Cellular Indexing of Transcriptomes and Epitopes by sequencing)

- Couples measurement of surface protein markers on thousands of single cells with simultaneous mRNA sequencing - NYGC
- Correlation between mRNA and protein levels
- Clustering of cells with similar transcriptional profiles

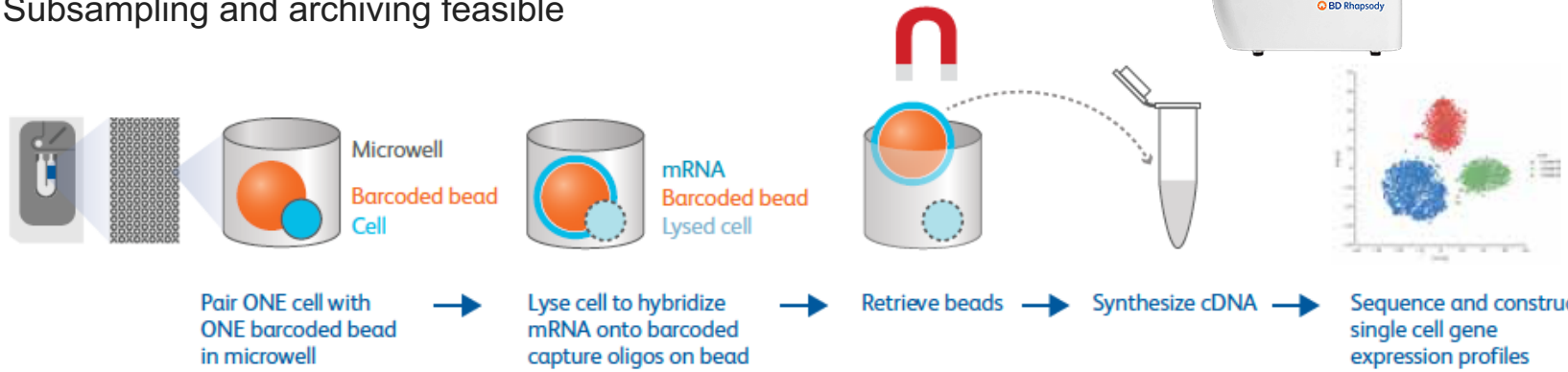


Stoeckius et al, 2017, Nature Methods **14**: 865

Coming soon – targeted RNAseq from single cells

Targeted RNAseq from single cells using BD Rhapsody system

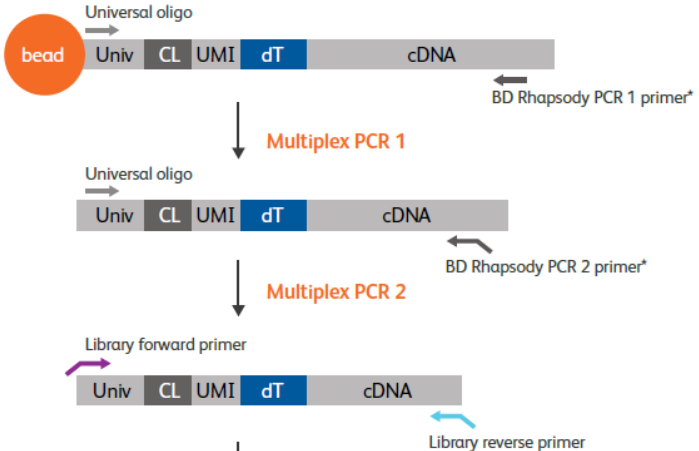
- Analysis of low abundance transcripts
- Less sequencing required, cheaper
- Subsampling and archiving feasible



- Panels – immune cell panels available, custom option in future.

Whole Genome Sequencing from single cells

Amplification of cDNA using targeted primer panels



Acknowledgements

Sequencing Facility
Bao Tran
Jyoti Shetty
Castle Raley
Yuliya Kriga
Yelena Levin
Sujatha Gowda
Oksana German
Tatyana Smirnova
Heidi Smith
Christopher Hautman

SF Bioinformatics
Yongmei Zhao
Keyur Talsania
Ashley Walton
Tsai-Wei Shen
Sucheta Godbole
Cris Vera

Leidos Biomedical
Dwight Nissley
Daniel Soppet
Robert Stephens
Xiaolin Wu

Scientific support from:
10X Genomics
Illumina

NCI / NIH
David Goldstein
Mariam Malik
Paul Meltzer
Javed Khan
Remy Bosselut
Avinash Bhandoola
Cheryl Winkler
Xin Wang
Francis McMahon
Yasmine Belkaid
Ramiro Iglesias-Bartolome
KathyKelly
Jairaj Acharya
Chris Nagao
Tae-Wook Chun
Jeffrey Kopp

Thank you!

Contact us
https://ostr.cancer.gov/resources/fnl-cores/sequencing-facility/
sftechdev@mail.nih.gov

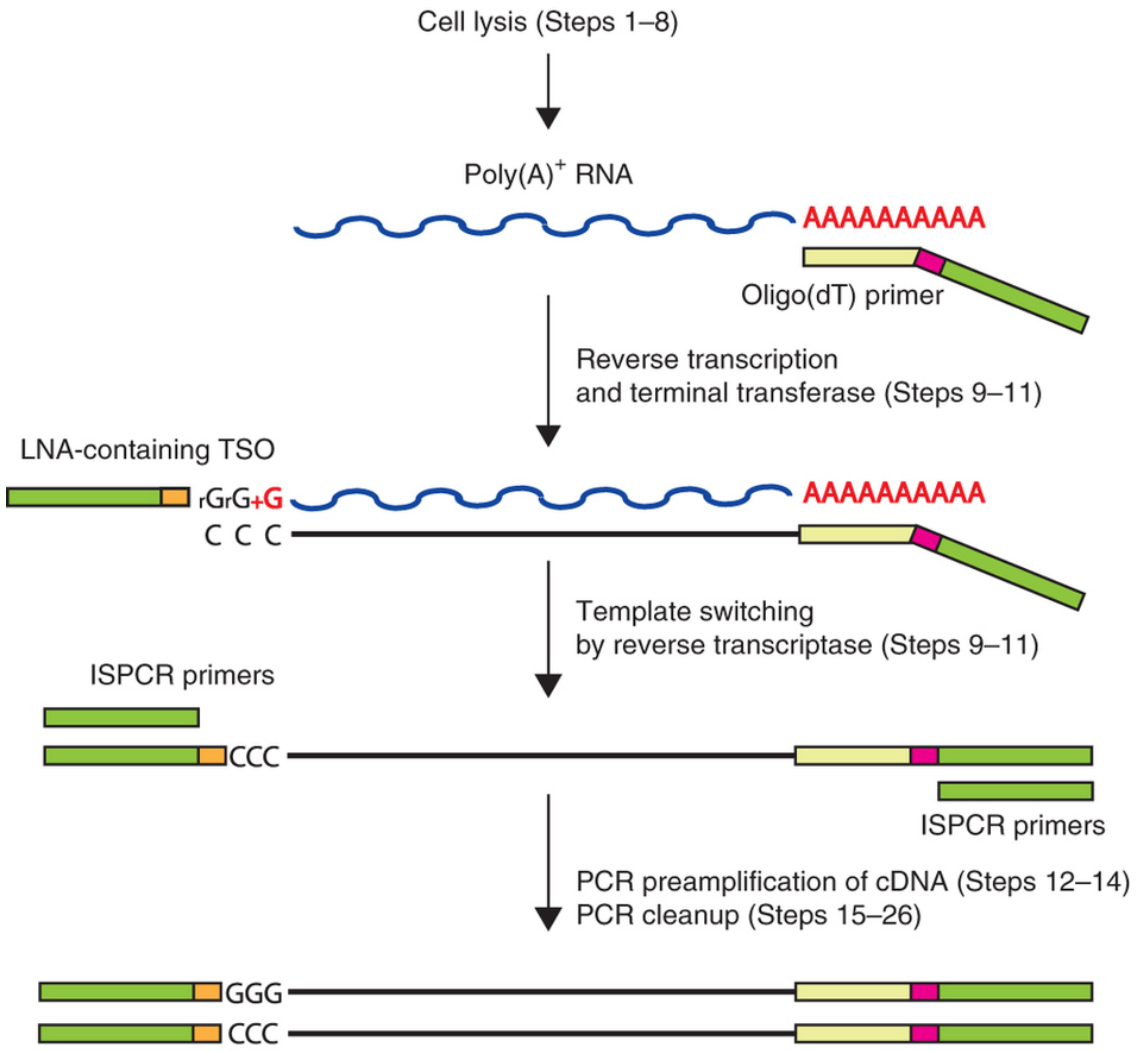
Bao Tran
Laboratory Director
tranb2@mail.nih.gov

Jyoti Shetty
Illumina Lab Manager
shettyju@mail.nih.gov

Castle Raley
PacBio Lab Manager
raleyjc@mail.nih.gov

Mehta Monika
R&D Scientist
monika.mehta@nih.gov

Yongmei Zhao
Bioinformatics Manager
yongmei.zhao@mail.nih.gov


Cell lysis (Steps 1–8)
Poly(A)+ RNA
AAAAAAAAAA
Oligo(dT) primer
Reverse transcription
and terminal transferase (Steps 9–11)
LNA-containing TSO
rGrG+G
C C C
AAAAAAAAAA
Template switching
by reverse transcriptase (Steps 9–11)
ISPCR primers
CCC
ISPCR primers
PCR preamplification of cDNA (Steps 12–14)
PCR cleanup (Steps 15–26)
GGG
CCC

Short-read Sequencing : Sequencing by Synthesis

Illumina Sequencers (MiSeq, NextSeq, HiSeq)

- Max read length: 300bp
- Variety of sequencing applications: genome assembly, transcriptome analysis, SNP detection, DNA methylation analysis, metagenomic studies.
- High level multiplexing of samples possible.
- Big advantage: scale.

- Limitations of short read sequencing – genome assembly, isoform analysis, structural variation, haplotypes and SNP phasing.

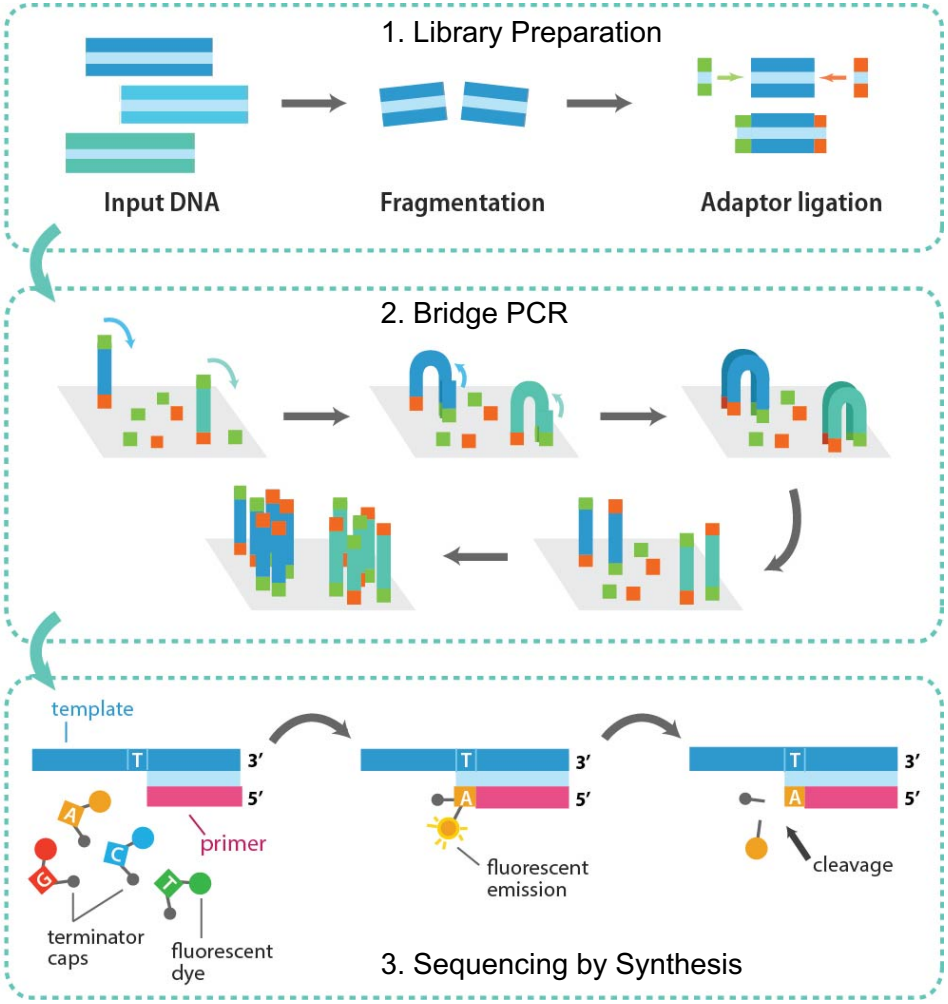
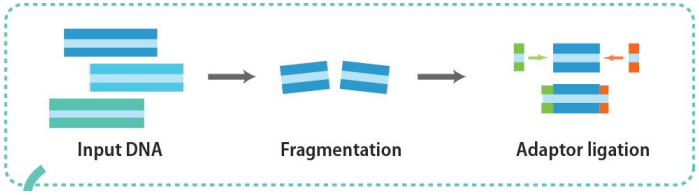


Figure from: https://www.abmgood.com/Next-Generation-Sequencing-Service.html

Library preparation for Short-read Sequencing



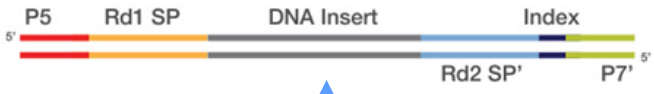
Single cells

Total RNA

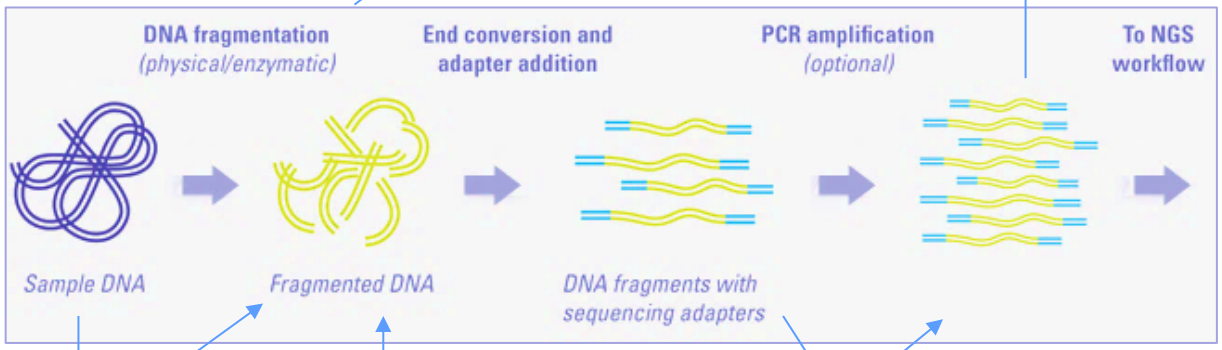
Lysates → mRNA enrichment rRNA depletion ← **FFPE RNA**

cDNA synthesis
(**Gene expression, whole transcriptome, miRNA, lncRNA**..)

Bisulfite conversion
(Methyl-Seq)



Genomic DNA



Barcoded amplicons
(SV analysis by 10XGenomics)

ChIP DNA

Target selection
(**Exome-Seq** etc)