



Single cell RNA-seq with differentiating skeletal muscle cell line

UCI COSMOS Cluster 6

Liz Rebboah

Mortazavi Lab

Why single cell genomics?



Shai Bel
@Mucus_Man

"Single cell RNA seq of (insert organ name here) reveals heterogeneous cell population" Nature 2019 (every week)

1:26 PM · Jul 14, 2019 · Twitter for Android

363 Retweets 2.1K Likes

Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors

Alexandra-Chloé Villani^{1,2,*†}, Rahul Satija^{1,3,4,*}, Gary Reynolds⁵, Siranush Sarkizova¹, Karthik Shekhar¹, James Fletcher⁵, ...

* See all authors and affiliations

Science 21 Apr 2017;
Vol. 356, Issue 6335, eaah4573
DOI: 10.1126/science.aah4573

Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq

Barbara Treutlein, Qian Yi Lee, J. Gray Camp, Moritz Mall, Winston Koh, Seyed Ali Mohammad Shariati, Sopehak Sim, Norma F. Neff, Jan M. Skotheim, Marius Wernig & Stephen R. Quake

Nature 534, 391-395 (2016) | Cite this article

Single-nucleus RNA-seq of differentiating human myoblasts reveals the extent of fate heterogeneity



Weihua Zeng, Shan Jiang, Xiangduo Kong, Nicole El-Ali, Alexander R. Ball, Jr, Christopher I-Hsing Ma, Naohiro Hashimoto, Kyoko Yokomori & Ali Mortazavi

Nucleic Acids Research, Volume 44, Issue 21, December 2016, Page e158,
<https://doi.org/10.1093/nar/gkw739>

Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma

Anoop P. Patel^{1,2,3,4}, Itay Tirosh^{5,6}, John J. Trombetta³, Alex K. Shalek³, Shawn M. Gillespie^{2,3,4}, Hiroaki Wakimoto¹, Daniel...
+ See all authors and affiliations

Science 20 Jun 2014;
Vol. 344, Issue 6190, pp. 1396-1401
DOI: 10.1126/science.1254257

Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer

WooSung Chung, Hye Hyeon Eum, Hae-Ock Lee, Kyung-Min Lee, Han-Byeo Lee, Kyu-Tae Kim, Han Suk Ryu, Sangmin Kim, Jeong Eon Lee, Yeon Hee Park, Zhengyan Kan, Wonsik Han & Woong-Yang Park

Nature Communications 8, Article number: 15081 (2017) | Cite this article

Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq

Diego Adhemar Jaitin,^{1,2} Assaf Weinreb,^{1,2} Ido Amit,¹ David Lare-Astiaso,¹ Hadass Keren-Shaul,¹ Eyal David,¹ Tomer Meir Salama,³ Amos Tanay,⁴ Alexander van Oudenaarden,¹ and Ido Amit^{1,*†}

¹Department of Immunology, Weizmann Institute, Rehovot 76100, Israel

²Hubrecht Institute-KNAW, Royal Netherlands Academy of Arts and Sciences, 3584 CT Utrecht, the Netherlands

³Flow Cytometry Unit, Department of Biological Services, Weizmann Institute of Science, Rehovot 76100, Israel

⁴Department of Computer Science and Applied Mathematics, Department of Biological Regulation, Weizmann Institute, Rehovot 76100, Israel

*Co-first author

†Lead Contact

*Correspondence: ido.amit@weizmann.ac.il
<http://dx.doi.org/10.1016/j.cell.2016.11.039>

Why single cell genomics?

- Individual cells have distinct functions in complex biological systems
 - Tissues are made up of many subtypes of cells
 - Immune system
 - Brain
 - Tumors
 - Individual cell types are dynamic
 - Differentiation
 - Reprogramming (iPSCs)
- We can identify and characterize individual cell types by their gene expression

Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors

Alexandra-Chloé Villani^{1,2,*†}, Rahul Satija^{1,3,4,*}, Gary Reynolds⁵, Sianush Sarkizova¹, Karthik Shekhar¹, James Fletcher⁵, ...

* See all authors and affiliations

Science 21 Apr 2017;
Vol. 356, Issue 6335, eaah4573
DOI: 10.1126/science.aah4573

Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq

Barbara Treutlein, Qian Yi Lee, J. Gray Camp, Moritz Mall, Winston Koh, Seyed Ali Mohammad Shariati, Sopheak Sim, Norma F. Neff, Jan M. Skotheim, Marius Wernig & Stephen R. Quake

Nature 534, 391-395 (2016) | Cite this article

Single-nucleus RNA-seq of differentiating human myoblasts reveals the extent of fate heterogeneity



Weihua Zeng, Shan Jiang, Xiangduo Kong, Nicole El-Ali, Alexander R. Ball, Jr, Christopher I-Hsing Ma, Naohiro Hashimoto, Kyoko Yokomori & Ali Mortazavi

Nucleic Acids Research, Volume 44, Issue 21, December 2016, Page e158,
<https://doi.org/10.1093/nar/gkw739>

Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma

Anoop P. Patel^{1,2,3,4}, Itay Tirosh^{5,6}, John J. Trombetta³, Alex K. Shalek³, Shawn M. Gillespie^{2,3,4}, Hiroaki Wakimoto¹, Daniel...
+ See all authors and affiliations

Science 20 Jun 2014;
Vol. 344, Issue 6190, pp. 1396-1401
DOI: 10.1126/science.1254257

Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer

WooSung Chung, Hye Hyeon Eun, Hae-Ock Lee, Kyung-Min Lee, Han-Byeo Lee, Kyu-Tae Kim, Han Suk Ryu, Sangmin Kim, Jeong Eon Lee, Yeon Hee Park, Zhengyan Kan, Wonshik Han & Woong-Yang Park

Nature Communications 8, Article number: 15081 (2017) | Cite this article

Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq

Diego Adhemar Jaitin,^{1,2} Assaf Weinert,^{1,2} Ido Amit,¹ David Lare-Astiaso,¹ Hadass Keren-Shaul,¹ Eyal David,¹ Tomer Meir Salama,³ Amos Tanay,⁴ Alexander van Oudenaarden,¹ and Ido Amit^{1,*†}

¹Department of Immunology, Weizmann Institute, Rehovot 76100, Israel

²Hubrecht Institute-KNAW, Royal Netherlands Academy of Arts and Sciences, 3584 CT Utrecht, the Netherlands

³Flow Cytometry Unit, Department of Biological Services, Weizmann Institute of Science, Rehovot 76100, Israel

⁴Department of Computer Science and Applied Mathematics, Department of Biological Regulation, Weizmann Institute, Rehovot 76100, Israel

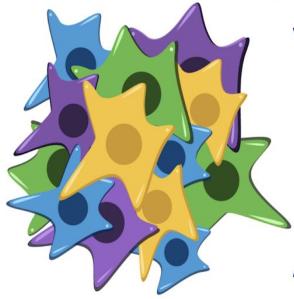
*Co-first author

†Lead Contact

*Correspondence: ido.amit@weizmann.ac.il
<http://dx.doi.org/10.1016/j.cell.2016.11.039>

Comparison of RNA-seq techniques

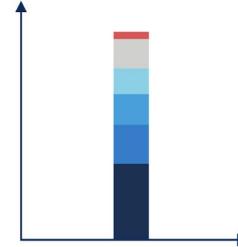
Tissue



Bulk Analysis



Bulk RNA input



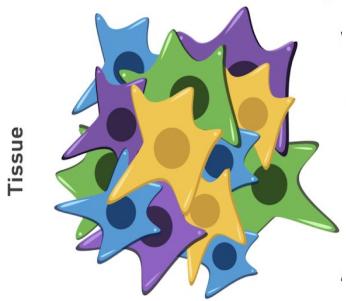
Average gene expression
from all cells



?

Cellular heterogeneity
masked

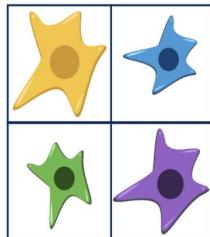
Comparison of RNA-seq techniques



Bulk Analysis



Bulk RNA input

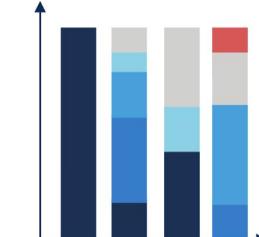


Single-Cell Analysis

Average gene expression from all cells



Single-Cell input

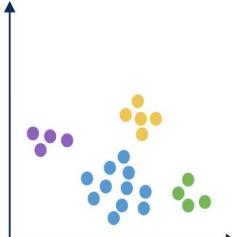


Each cell type has a distinct expression profile



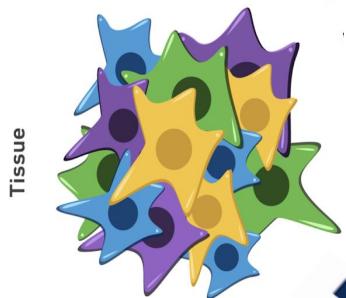
Cellular heterogeneity masked

Reveals heterogeneity and subpopulation expression variability of thousands of cells



?

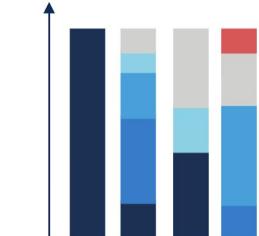
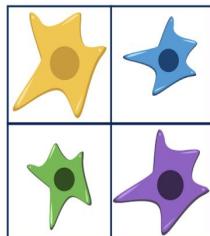
Comparison of RNA-seq techniques



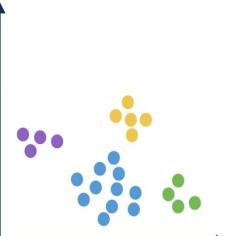
Bulk Analysis



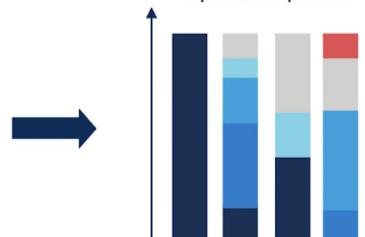
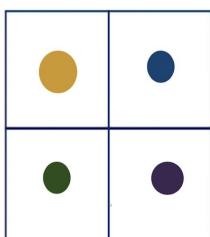
Average gene expression
from all cells



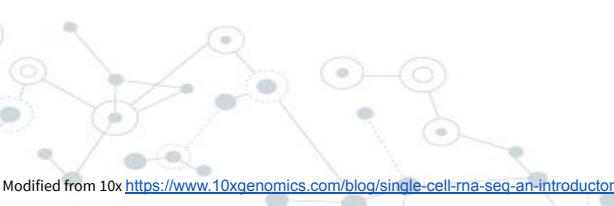
Cellular heterogeneity
masked



Single-Nucleus Analysis



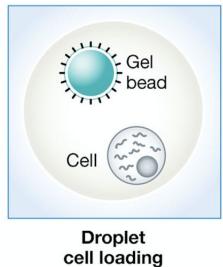
Reveals heterogeneity and subpopulation expression variability of thousands of cells



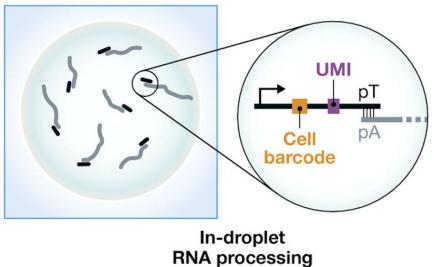
Single cell isolation is either droplet-based or plate-based

DROPLET-BASED METHODS

e.g. Drop-seq
10X Chromium



Gel bead dissolution
Cell lysis



+ Low cost per cell (< \$0.01)

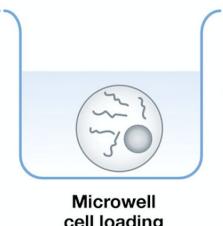
+ Extremely high cell throughput (>10⁴ cells per experiment)
- Smaller cell libraries (~10⁴ molecules per cell)

Read 1
• Cell ID
• UMI

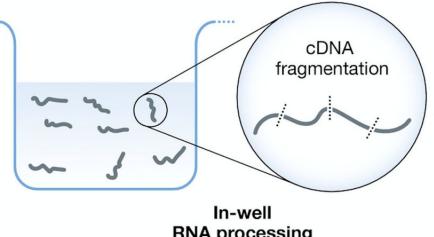
Read 2
• Gene 3' sequence

PLATE-BASED METHODS

e.g. Smart-Seq2
MARS-seq



Cell lysis
Reverse transcription



+ High read-depth per cell (>10⁶ reads per cell)
+ Reads may be generated across whole transcript length

- Moderate cell throughput (10²–10³ cells per experiment)

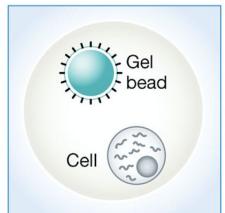
Both reads
• Cell ID
• Gene sequence

Example	Method	Equipment
10X Chromium	Droplet	Chromium Controller
Bio-Rad ddSeq	Droplet	ddSeq Single-Cell Isolator
Smart-seq2	Plate	FACS machine or Fluidigm C1 (or none)
Split-seq	Plate	None

The list of single cell methods is always growing...

DROPLET-BASED METHODS

e.g. Drop-seq
10X Chromium



Gel bead dissolution
Cell lysis

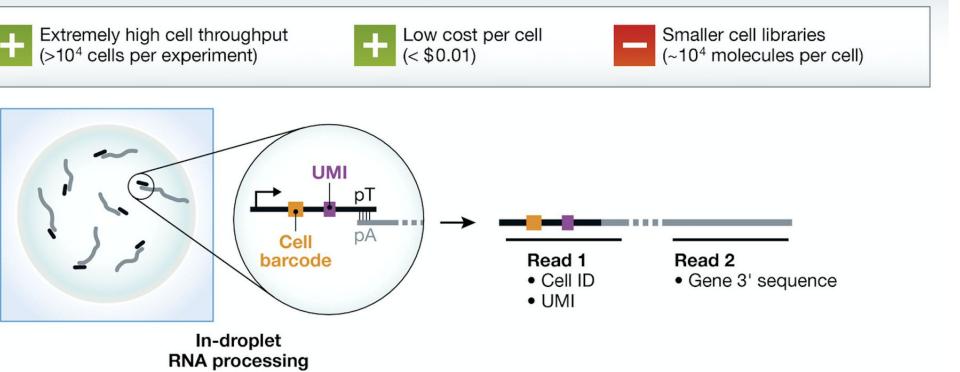
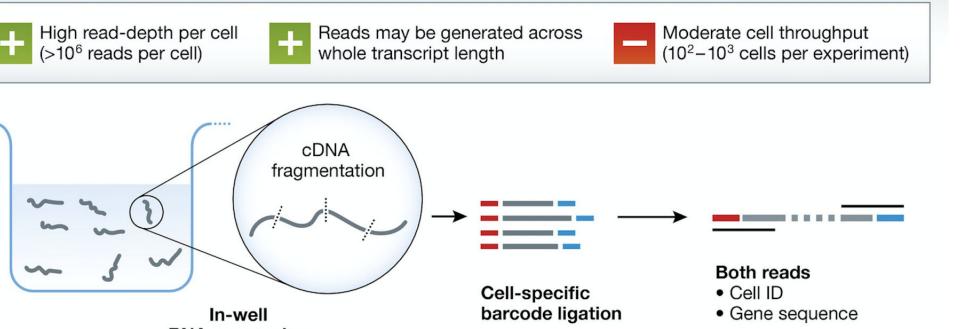


PLATE-BASED METHODS

e.g. Smart-Seq2
MARS-seq



Cell lysis
Reverse transcription



Method	Reference	Sequencing Mode	Early Estimate	Late Estimate
Tang method	[2]	Short Reads	2008	2009
CyTOF	[3]	Short Reads	2011	2012
STRT-seq / C1	[4]	Short Reads	2011	2012
SMART-seq	[5]	Short Reads	2012	2013
CEL-seq	[6]	Short Reads	2012	2013
Quartz-Seq	[7]	Short Reads	2012	2013
PMA / SMA	[8]	Short Reads	2012	2013
scBS-seq	[9]	Short Reads	2013	2014
AbPair	[10]	Short Reads	2014	2014
MARS-seq	[11]	Short Reads	2014	2015
DR-seq	[12]	Short Reads	2014	2015
G&T-Seq	[13]	Short Reads	2014	2015
SCTG	[14]	Short Reads	2014	2015
SiDR-seq	[15]	Short Reads	2014	2015
sci-ATAC-seq	[16]	Short Reads	2014	2015
Hi-SCL	[17]	Short Reads	2015	2015
SUPER-seq	[18]	Short Reads	2015	2015
Drop-Chip	[19]	Short Reads	2015	2015
CytoSeq	[20]	Short Reads	2015	2016
inDrop	[21]	Short Reads	2015	2016
sc-GEM	[22]	Short Reads	2015	2016
scTrio-seq	[23]	Short Reads	2015	2016
scMAT-seq	[24]	Short Reads	2015	2016
PLAYR	[25]	Short Reads	2015	2016
Genshaft-etal-2016	[26]	Short Reads	2015	2016
Dermants-etal-2016	[27]	Short Reads	2015	2016
CRISP-seq	[28]	Short Reads	2015	2016
scGESTALT	[29]	Short Reads	2015	2016
CEL-Seq2 / C1	[30]	Short Reads	2015	2016
STRT-seq[2]	[31]	Short Reads	2016	2017
RNAseq @10xgenomics	[32]	Short Reads	2016	2017
RNAseq / Gene Expression @nanostringtech	[33]	Short Reads	2016	2017
scTargeted Gene Expression @fluidigm	[34]	Short Reads	2016	2017
scTCR Wafergen	[35]	Short Reads	2016	2017
CROP-seq	[36]	Short Reads	2016	2017
SIC-seq	[37]	Short Reads	2016	2017
mcSCRIB-seq	[38]	Short Reads	2016	2017
Patch-seq	[39]	Short Reads	2016	2017
Geo-seq	[40]	Short Reads	2016	2017
scNOME-seq	[41]	Short Reads	2016	2017
scCOOL-seq	[42]	Short Reads	2016	2017
CUT&Run	[43]	Short Reads	2016	2017
MATQ-seq	[44]	Short Reads	2016	2017
Quartz-Seq2	[45]	Short Reads	2017	2018
Seq-Well	[46]	Short Reads	2017	2018
DroNc-Seq	[47]	Short Reads	2017	2018
sci-RNA-seq	[48]	Short Reads	2017	2018
scATAC @10xgenomics	[49]	Short Reads	2017	2018
scVDJ @ 10xgenomics	[50]	Short Reads	2017	2018
scNMT triple omics	[51]	Short Reads	2017	2018
SPLIT-seq Split Biosciences	[52]	Short Reads	2017	2018
CITE-seq	[53]	Short Reads	2017	2018
scMNase-seq	[54]	Short Reads	2017	2018
Chaligne-etal-2018	[55]	Short Reads	2017	2018
LINNAEUS	[56]	Short Reads	2017	2018
TracerSeq	[57]	Short Reads	2017	2018
CellTag	[58]	Short Reads	2017	2018
ScarTrace	[59]	Short Reads	2017	2018
scRNA-Seq Dolomite Bio	[60]	Short Reads	2017	2018
Trac-looping	[61]	Short Reads	2017	2018

How to isolate 1 cell per well?

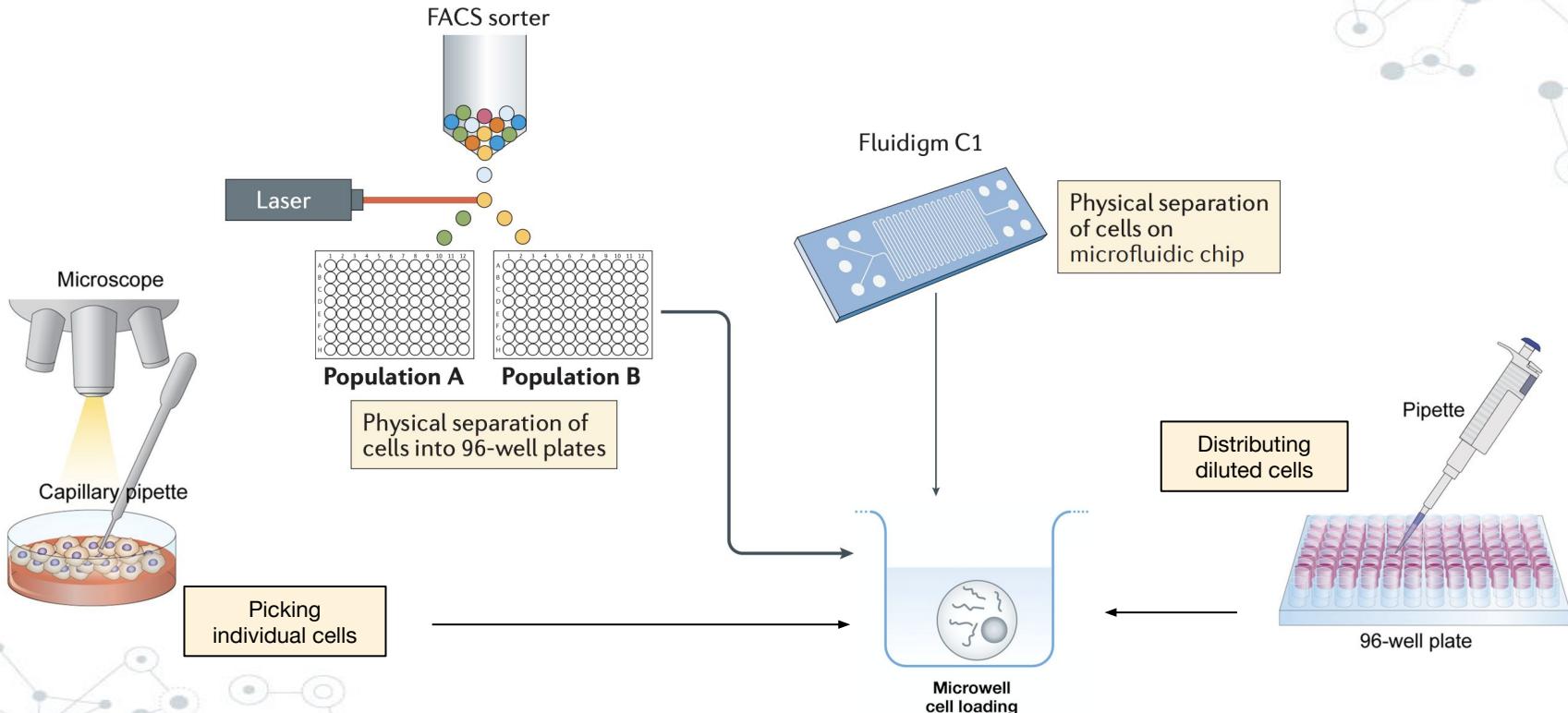
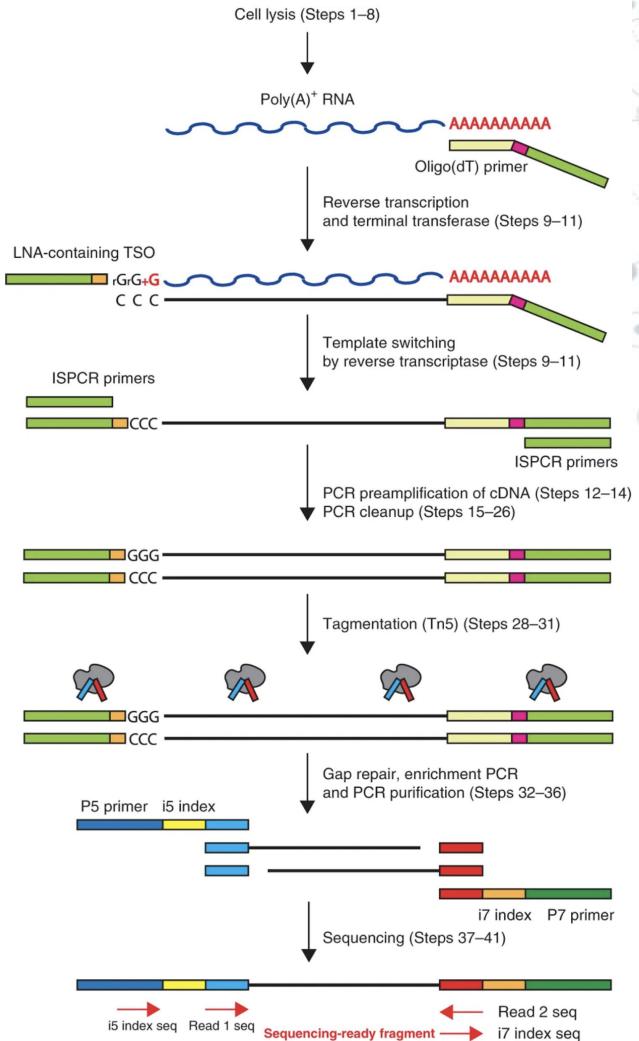


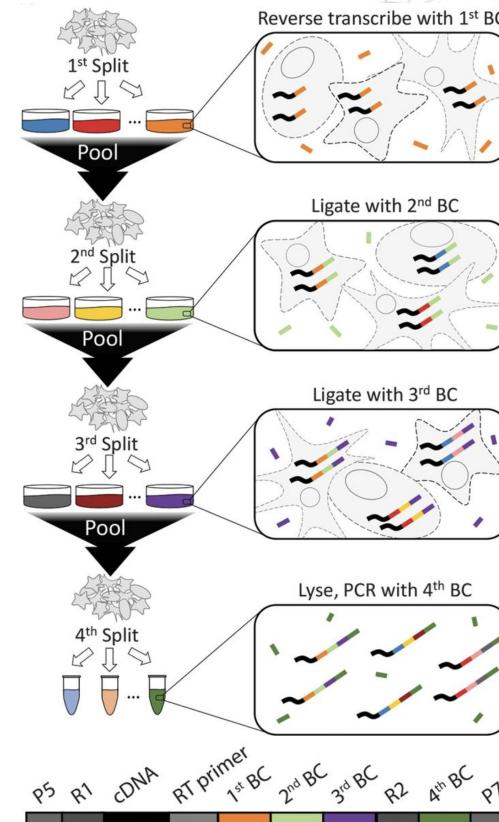
Plate-based methods capture full-length mRNA

- **Switching Mechanism at 5' of RNA Template**
 - Special reverse transcriptase adds 3 untemplated Cs to 3' end of cDNA strand
 - Template-switching oligo (TSO) has locked nucleic acids (LNA) that base pair to Cs and anchor the oligo
 - RT switches seamlessly to TSO from RNA strand
 - Resulting cDNA contains **complete** 5' end
- Amplifies the entire full-length transcript
- Optimized for small amounts of RNA

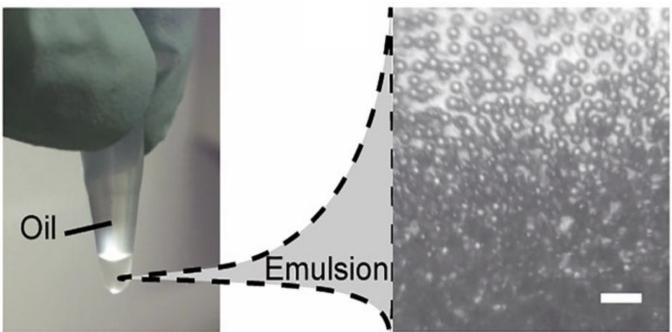
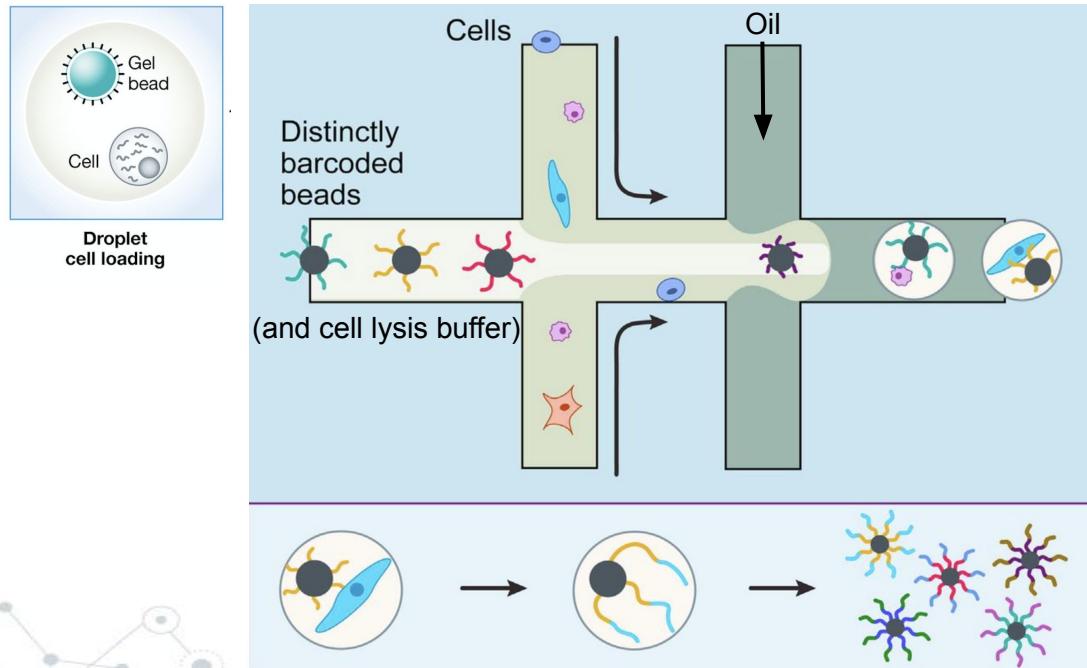


Split-pool plate-based approach utilizes combinatorial barcoding to uniquely label cells

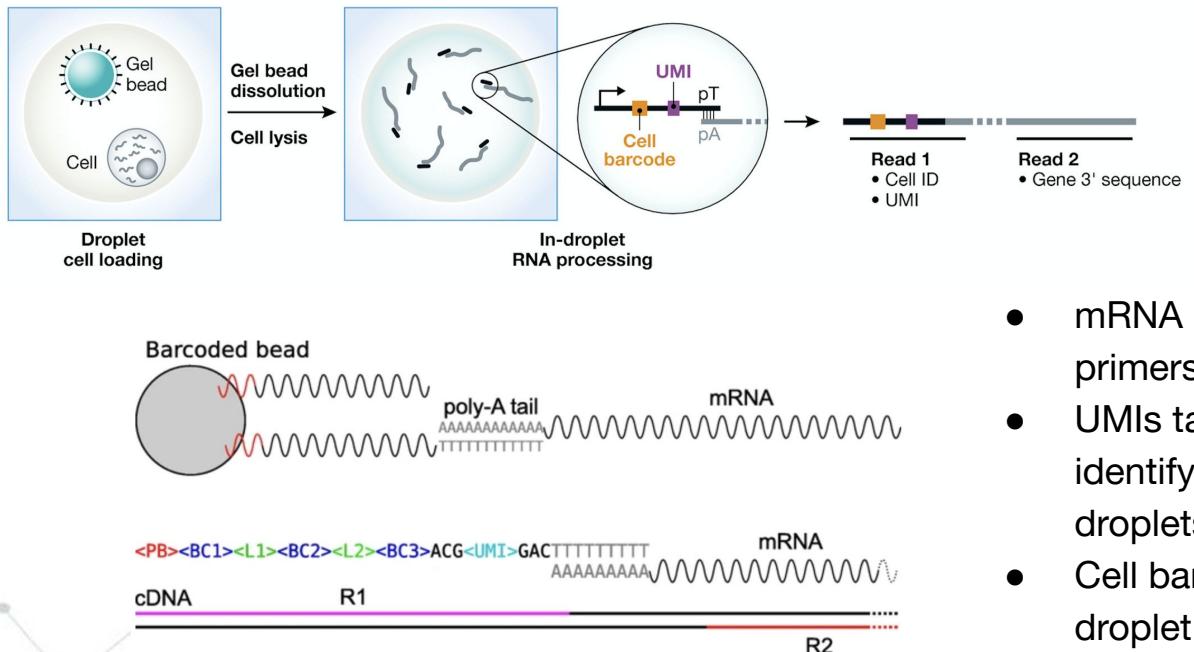
- Pros
 - No special equipment
 - Fixed cells are their own “container” - not microwells or droplets
 - No restrictions on cell size
 - Captures full-length transcripts
 - Captures up to 100,000 cells in parallel across 48 samples
- Cons
 - 15% retention from starting number of cells
 - Kits are still expensive



DNA-barcoded microparticles (“beads”) combine with cells in water-based droplets, separated from each other in an oil emulsion

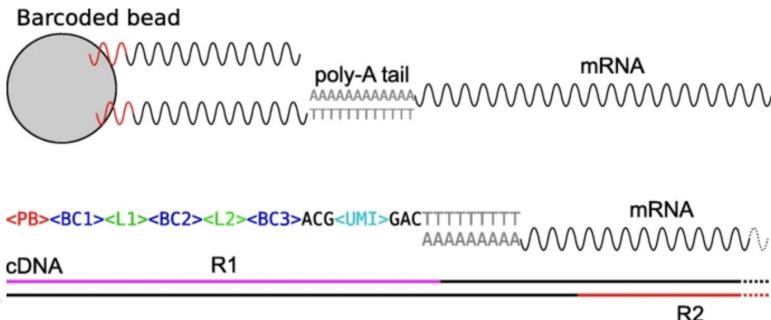
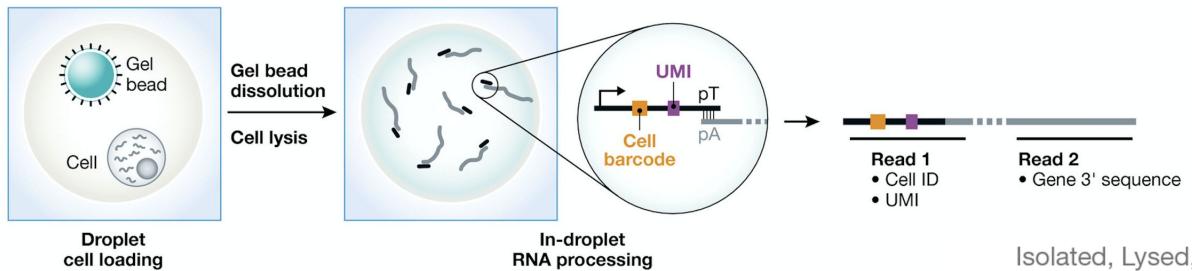


Cell lysis and reverse transcription occurs within droplets, adding a unique molecular identifier (UMI) and cell barcode to each mRNA

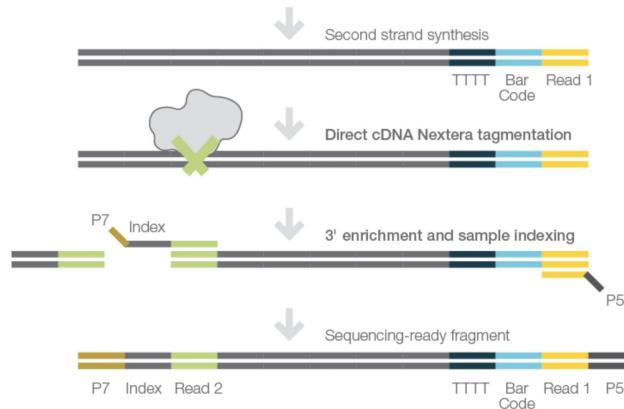


- mRNA poly-A tail hybridizes to bead primers
- UMIs tag each individual mRNA, used to identify PCR duplicates, cell doublets, and droplets without cells
- Cell barcodes tag all mRNAs within a droplet with the same barcode

After mRNA is barcoded and reverse transcribed, the emulsion is broken and cDNA is fragmented and PCR amplified

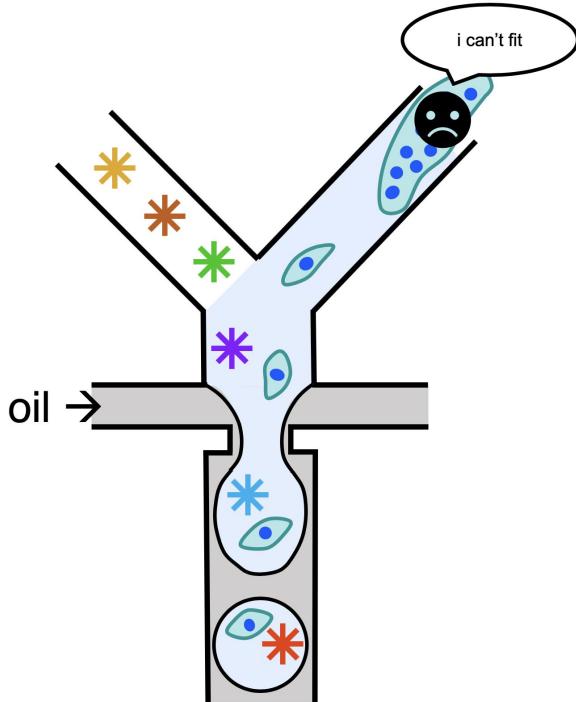


Isolated, Lysed, and Barcoded Single Cells



Technical challenges

- Preparing a single cell suspension from tissue is difficult
 - Need enough high-quality starting material
 - Lots of trial-and-error
 - The longer the protocol, the more stress on cells
- Cell size matters
 - Large cells do not fit through microfluidics channels (neurons, oligodendrocytes, myocytes)
 - Single-nucleus RNA-seq gets around this problem, but nuclei isolation from tissue is still difficult
- Most methods require some kind of specialized equipment

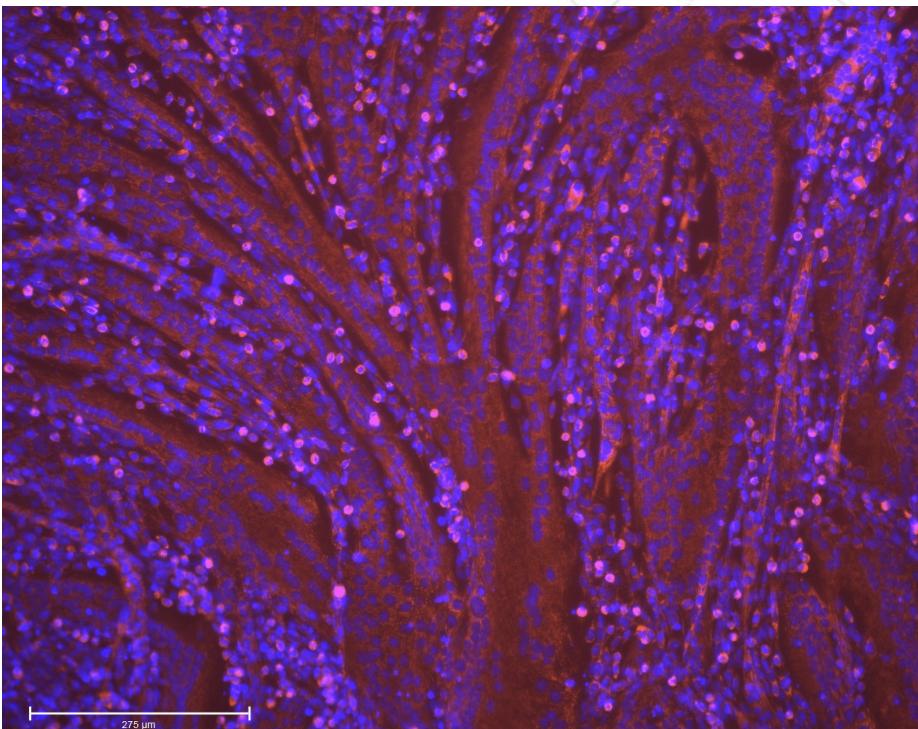
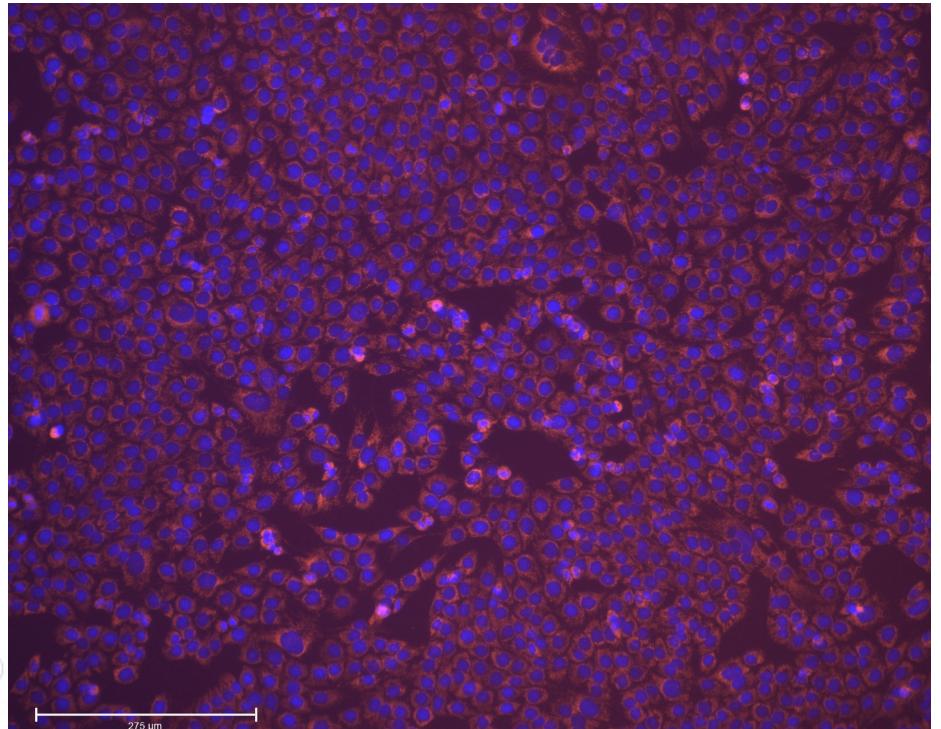


Skeletal muscle differentiation timeline

The diagram illustrates the progression of skeletal muscle differentiation through seven stages, each accompanied by a corresponding timeline and descriptive text.

0 hours	6 hours	14 hours	24 hours	48 hours	72 hours	4-7 days
Cell division	Fusion begins when cells are confluent	Cells begin to align	Cells elongate and fuse	Many cell fusions, some mature myotubes	Elongated, multi-nucleated myotubes	Contractions and striations

Multinucleated cells are too large to pass through microfluidics channels



Bio-Rad ddSeq scRNA-seq library prep workflow

Day 1

- 1 Prepare, Count, and Assess Viability of Single-Cell Suspension
Hands-on: 45 minutes
Total: 45 minutes
Reagents: Bovine Serum Albumin (BSA), Phosphate-Buffered Solution (PBS)
- 2 Prepare Cell and Barcode Suspension Mixes
Hands-on: 15 minutes
Total: 15 minutes
Reagents: Cell Suspension Buffer, DTT, RNA Stabilizer, RT Enzyme (RTE), Enhancer Enzyme, Barcode Buffer, 3' Barcode Mix
- 3 Isolate Single Cells
Hands-on: 30 minutes
Total: 35 minutes
Reagents: Encapsulation Oil, ddSEQ Priming Solution
- 4 Reverse Transcribe Samples
Hands-on: 5 minutes
Total: 1 hour 40 minutes
- 5 Break Emulsion
Hands-on: 5 minutes
Total: 5 minutes
Reagents: Droplet Disruptor, Nuclease-free water
- 6 Clean Up First Strand Synthesis
Hands-on: 30 minutes
Total: 1 hour
Reagents: EtOH, Resuspension Buffer (RSB), Purification Beads (SBP)
- 7 Synthesize Second Strand cDNA
Hands-on: 10 minutes
Total: 2 hours
Reagents: Second Strand Buffer (SSB), Second Strand Enzyme (SSE)

Overnight Incubation
Safe Stopping Point

Bio-Rad ddSeq scRNA-seq library prep workflow

Day 1

1

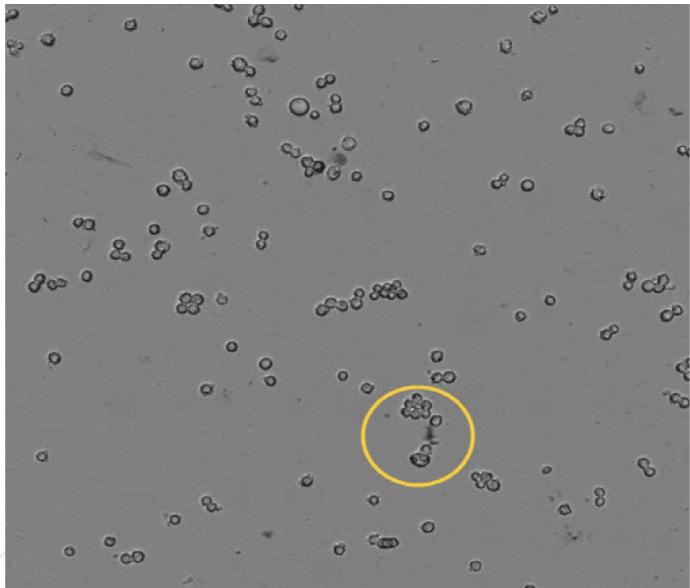
Prepare, Count, and Assess Viability of Single-Cell Suspension

Hands-on: 45 minutes

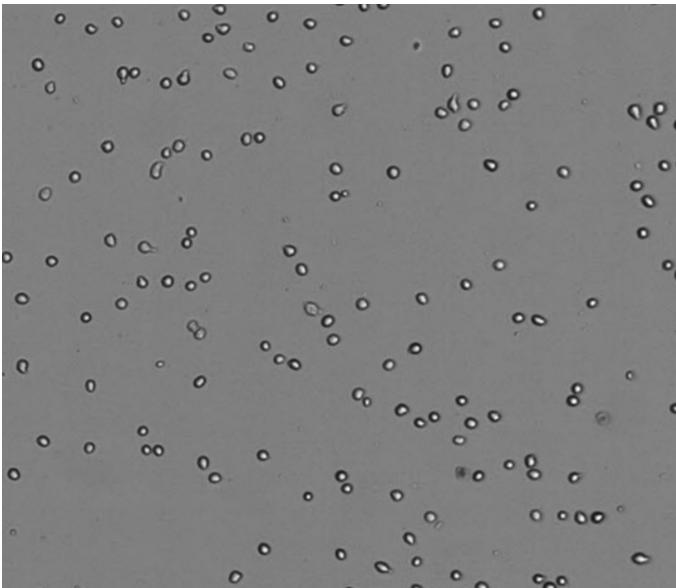
Total: 45 minutes

Reagents: Bovine Serum Albumin (BSA), Phosphate-Buffered Solution (PBS)

Cell suspension with doublets



Single cell suspension



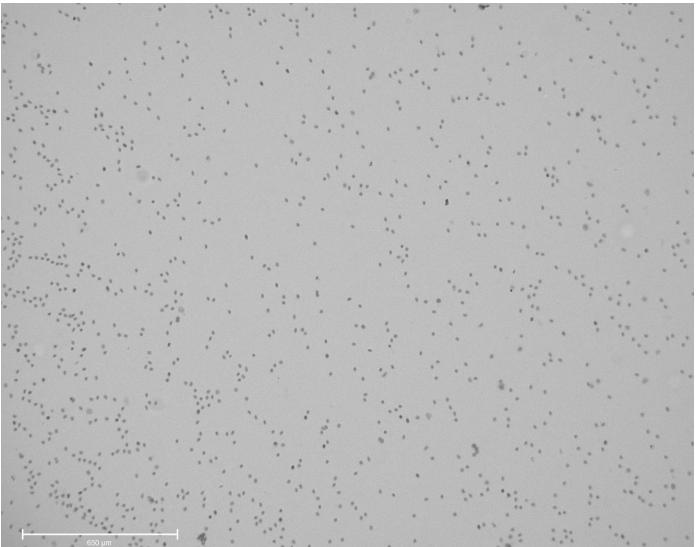
Bio-Rad ddSeq scRNA-seq library prep workflow

Day 1

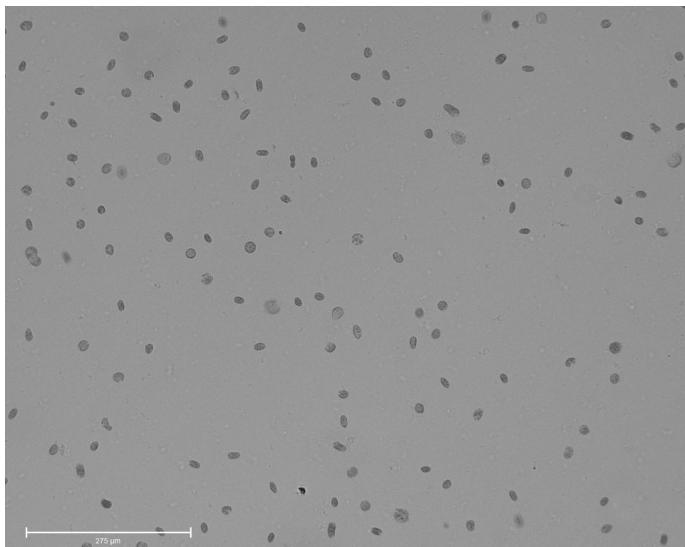
1

Prepare, Count, and Assess Viability of Single-Cell Suspension
Hands-on: 45 minutes
Total: 45 minutes
Reagents: Bovine Serum Albumin (BSA), Phosphate-Buffered Solution (PBS)

Single nucleus suspension



4x

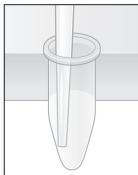
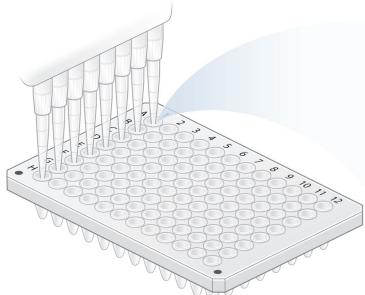
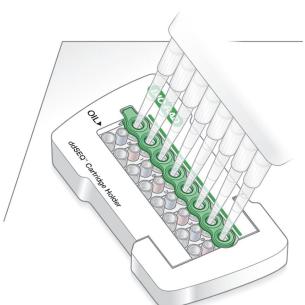


10x

Bio-Rad ddSeq scRNA-seq library prep workflow

Day 1

- 1 Prepare, Count, and Assess Viability of Single-Cell Suspension
Hands-on: 45 minutes
Total: 45 minutes
Reagents: Bovine Serum Albumin (BSA), Phosphate-Buffered Solution (PBS)
- 2 Prepare Cell and Barcode Suspension Mixes
Hands-on: 15 minutes
Total: 15 minutes
Reagents: Cell Suspension Buffer, DTT, RNA Stabilizer, RT Enzyme (RTE), Enhancer Enzyme, Barcode Buffer, 3' Barcode Mix
- 3 Isolate Single Cells
Hands-on: 30 minutes
Total: 35 minutes
Reagents: Encapsulation Oil, ddSEQ Priming Solution

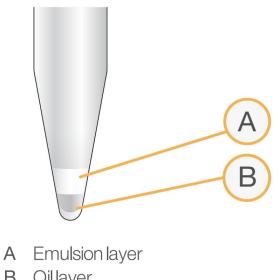


A Aqueous layer
B Oil layer
C Oil + air bubbles

Bio-Rad ddSeq scRNA-seq library prep workflow

Day 1

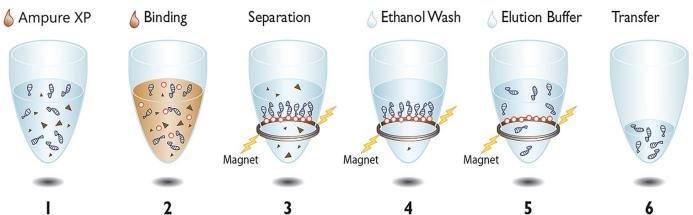
- 1 Prepare, Count, and Assess Viability of Single-Cell Suspension
Hands-on: 45 minutes
Total: 45 minutes
Reagents: Bovine Serum Albumin (BSA), Phosphate-Buffered Solution (PBS)
- 2 Prepare Cell and Barcode Suspension Mixes
Hands-on: 15 minutes
Total: 15 minutes
Reagents: Cell Suspension Buffer, DTT, RNA Stabilizer, RT Enzyme (RTE), Enhancer Enzyme, Barcode Buffer, 3' Barcode Mix
- 3 Isolate Single Cells
Hands-on: 30 minutes
Total: 35 minutes
Reagents: Encapsulation Oil, ddSEQ Priming Solution
- 4 Reverse Transcribe Samples
Hands-on: 5 minutes
Total: 1 hour 40 minutes
- 5 Break Emulsion
Hands-on: 5 minutes
Total: 5 minutes
Reagents: Droplet Disruptor, Nuclease-free water



Bio-Rad ddSeq scRNA-seq library prep workflow

Day 1

- 1 Prepare, Count, and Assess Viability of Single-Cell Suspension
Hands-on: 45 minutes
Total: 45 minutes
Reagents: Bovine Serum Albumin (BSA), Phosphate-Buffered Solution (PBS)
- 2 Prepare Cell and Barcode Suspension Mixes
Hands-on: 15 minutes
Total: 15 minutes
Reagents: Cell Suspension Buffer, DTT, RNA Stabilizer, RT Enzyme (RTE), Enhancer Enzyme, Barcode Buffer, 3' Barcode Mix
- 3 Isolate Single Cells
Hands-on: 30 minutes
Total: 35 minutes
Reagents: Encapsulation Oil, ddSEQ Priming Solution
- 4 Reverse Transcribe Samples
Hands-on: 5 minutes
Total: 1 hour 40 minutes
- 5 Break Emulsion
Hands-on: 5 minutes
Total: 5 minutes
Reagents: Droplet Disruptor, Nuclease-free water
- 6 Clean Up First Strand Synthesis
Hands-on: 30 minutes
Total: 1 hour
Reagents: EtOH, Resuspension Buffer (RSB), Purification Beads (SBP)

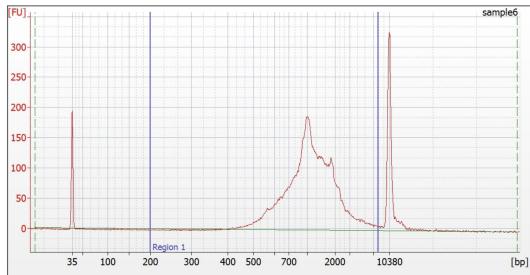


Bio-Rad ddSeq scRNA-seq library prep workflow

Day 1

- 1 Prepare, Count, and Assess Viability of Single-Cell Suspension
Hands-on: 45 minutes
Total: 45 minutes
Reagents: Bovine Serum Albumin (BSA), Phosphate-Buffered Solution (PBS)
- 2 Prepare Cell and Barcode Suspension Mixes
Hands-on: 15 minutes
Total: 15 minutes
Reagents: Cell Suspension Buffer, DTT, RNA Stabilizer, RT Enzyme (RTE), Enhancer Enzyme, Barcode Buffer, 3' Barcode Mix
- 3 Isolate Single Cells
Hands-on: 30 minutes
Total: 35 minutes
Reagents: Encapsulation Oil, ddSEQ Priming Solution
- 4 Reverse Transcribe Samples
Hands-on: 5 minutes
Total: 1 hour 40 minutes
- 5 Break Emulsion
Hands-on: 5 minutes
Total: 5 minutes
Reagents: Droplet Disruptor, Nuclease-free water
- 6 Clean Up First Strand Synthesis
Hands-on: 30 minutes
Total: 1 hour
Reagents: EtOH, Resuspension Buffer (RSB), Purification Beads (SBP)
- 7 Synthesize Second Strand cDNA
Hands-on: 10 minutes
Total: 2 hours
Reagents: Second Strand Buffer (SSB), Second Strand Enzyme (SSE)

Overnight Incubation
Safe Stopping Point



Bioanalyzer profile of cDNA pre-fragmentation, fragment size distribution is 400-8000bp

Bio-Rad ddSeq scRNA-seq library prep workflow

Day 2

8 Clean Up cDNA

Hands-on: 30 minutes
Total: 40 minutes
Reagents: EtOH, Purification Beads (SBP), Resuspension Buffer (RSB)

Pre-PCR Post-PCR

Safe Stopping Point

9 Tagment cDNA

Hands-on: 5 minutes
Total: 15 minutes
Reagents: Tagment Buffer (TCB), Tagment Enzyme (TCE), Tagment Stop Buffer (TSB)

10 Amplify Tagmented cDNA

Hands-on: 15 minutes
Total: 1 hour and 5 minutes
Reagents: Tagmentation PCR Mix (TPM), Tagment PCR Adapter (TPP1)

Safe Stopping Point

11 Clean Up Libraries

Hands-on: 30 minutes
Total: 1 hour 30 minutes
Reagents: EtOH, Purification Beads (SBP), Resuspension Buffer (RSB)

Safe Stopping Point

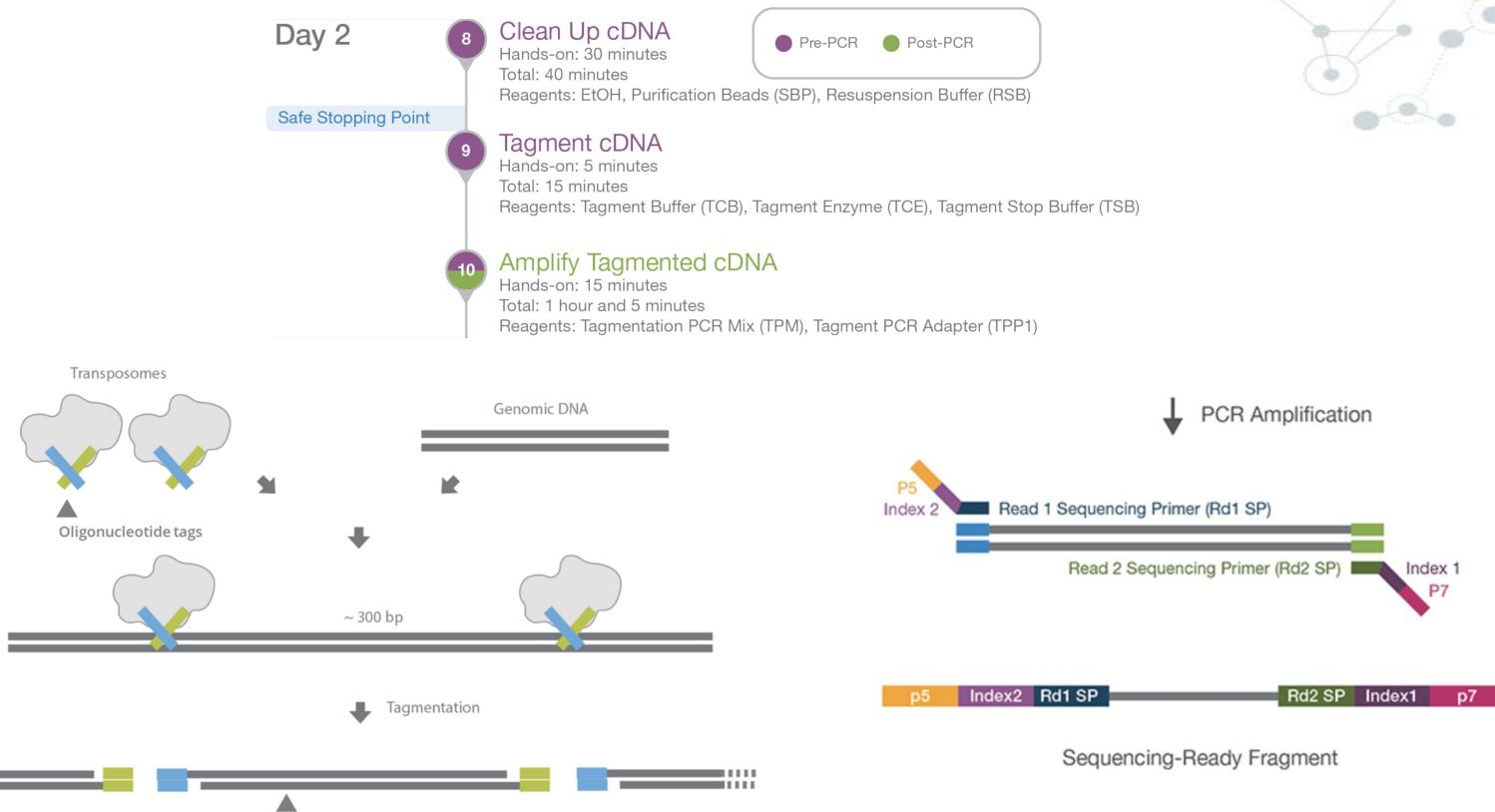
12 Check Libraries

Hands-on: 10 minutes
Total: 1 hour

13 Prepare for Sequencing

Hands-on: 15 minutes
Total: 30 minutes
Reagents: Sequencing Primer (SP), HT1

Bio-Rad ddSeq scRNA-seq library prep workflow



Adapted from Illumina

Bio-Rad ddSeq scRNA-seq library prep workflow

Day 2

8 Clean Up cDNA

Hands-on: 30 minutes
Total: 40 minutes
Reagents: EtOH, Purification Beads (SBP), Resuspension Buffer (RSB)

Pre-PCR Post-PCR

Safe Stopping Point

9 Tagment cDNA

Hands-on: 5 minutes
Total: 15 minutes
Reagents: Tagment Buffer (TCB), Tagment Enzyme (TCE), Tagment Stop Buffer (TSB)

10 Amplify Tagmented cDNA

Hands-on: 15 minutes
Total: 1 hour and 5 minutes
Reagents: Tagmentation PCR Mix (TPM), Tagment PCR Adapter (TPP1)

Safe Stopping Point

11 Clean Up Libraries

Hands-on: 30 minutes
Total: 1 hour 30 minutes
Reagents: EtOH, Purification Beads (SBP), Resuspension Buffer (RSB)

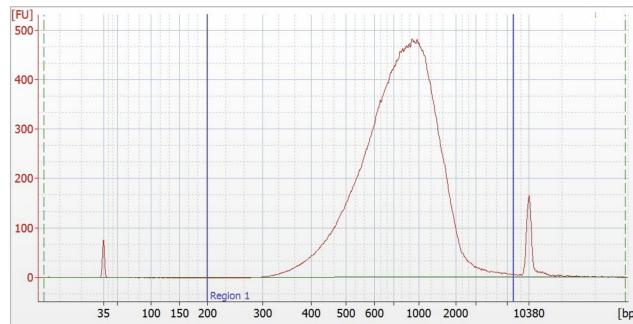
Safe Stopping Point

12 Check Libraries

Hands-on: 10 minutes
Total: 1 hour

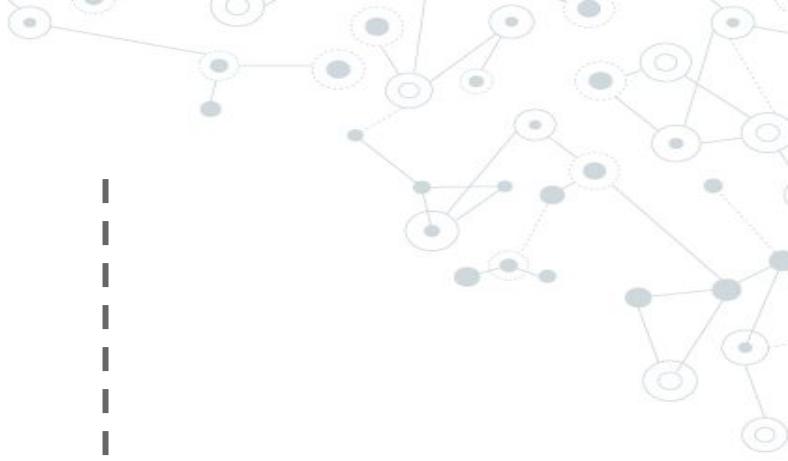
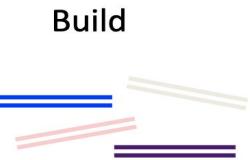
13 Prepare for Sequencing

Hands-on: 15 minutes
Total: 30 minutes
Reagents: Sequencing Primer (SP), HT1



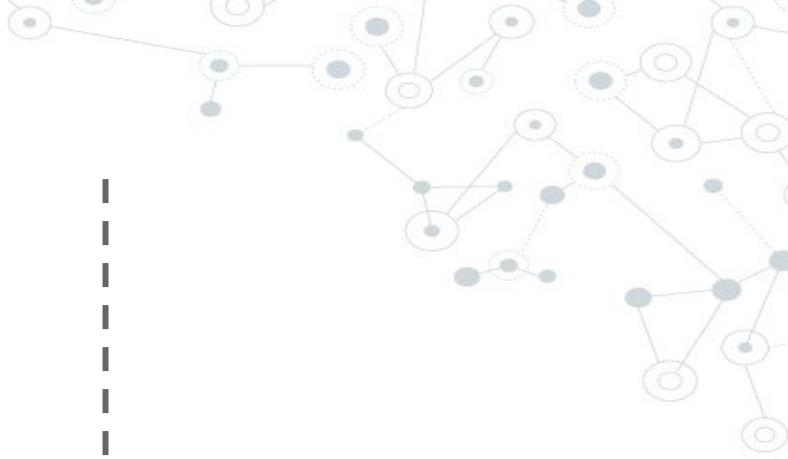
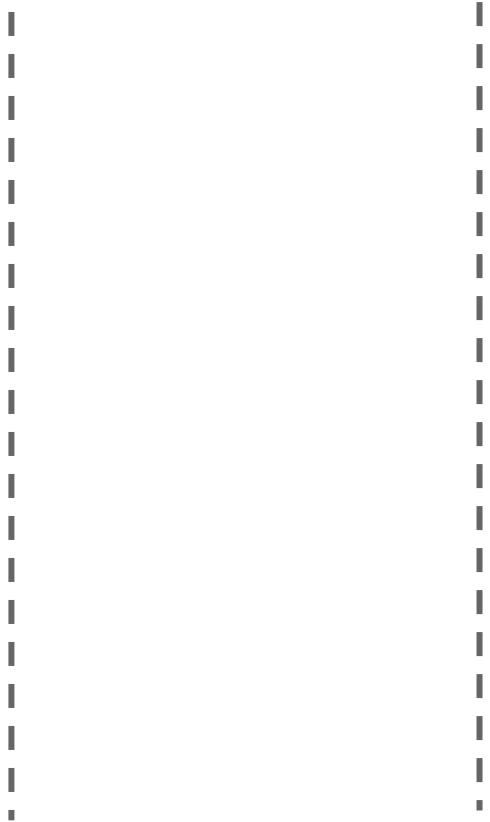
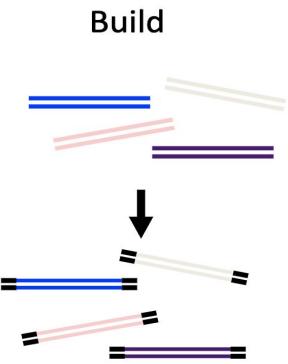
Bioanalyzer profile of final library, fragment size distribution is 300-1000bp

Sequencing recap



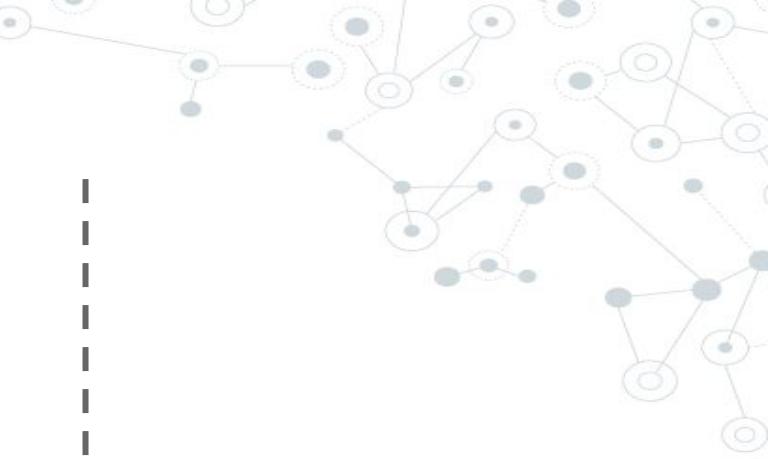
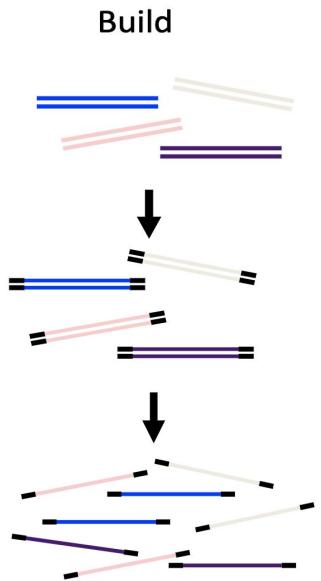
Courtesy of Prof. Katrine Whiteson

Sequencing recap

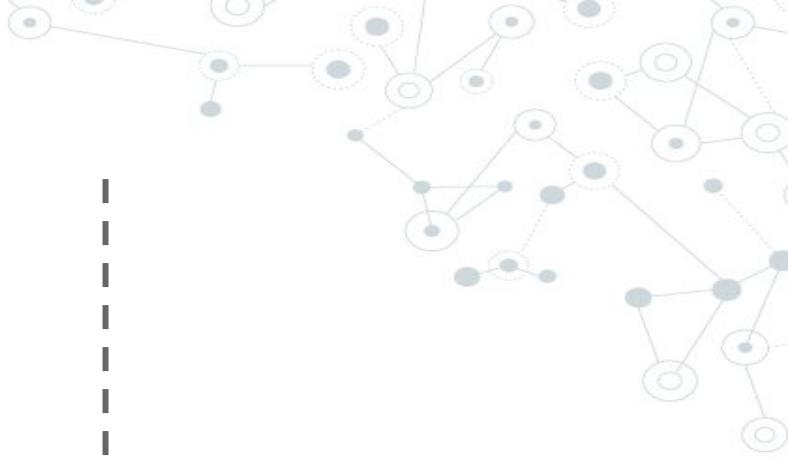
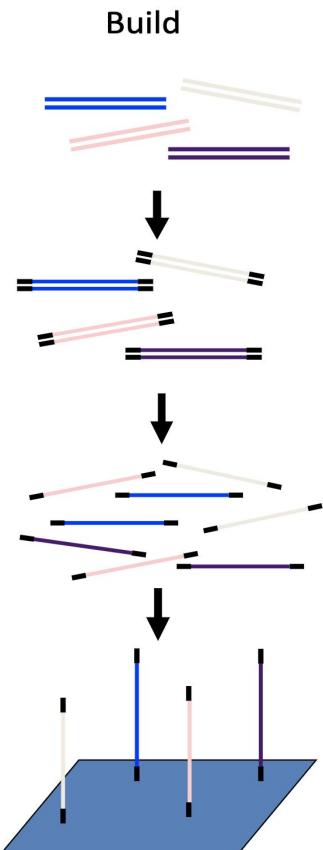


Courtesy of Prof. Katrine Whiteson

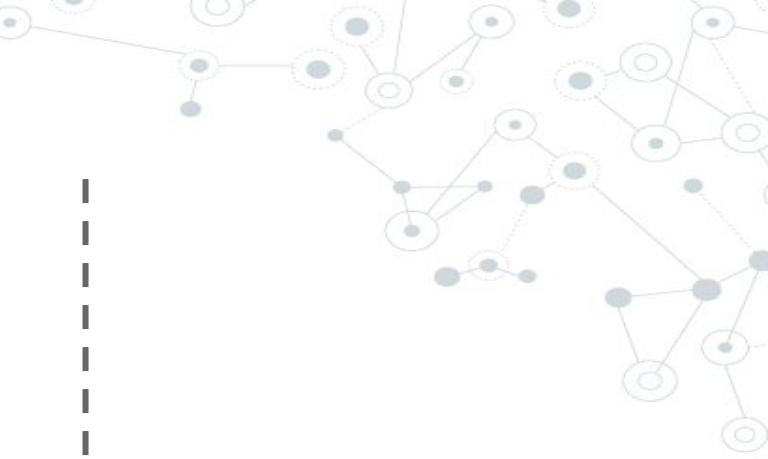
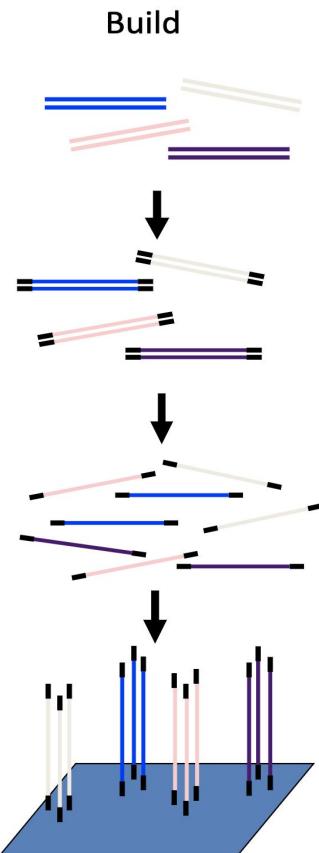
Sequencing recap



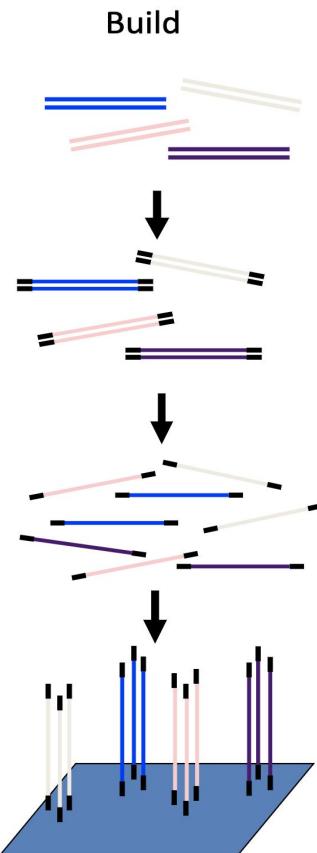
Sequencing recap



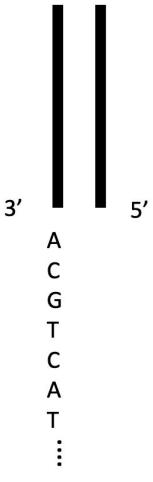
Sequencing recap



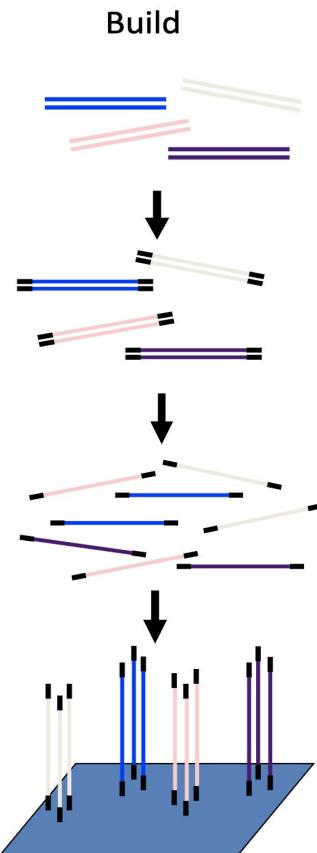
Sequencing recap



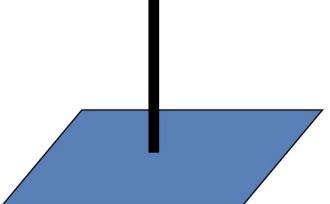
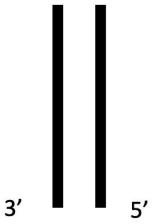
Sequence



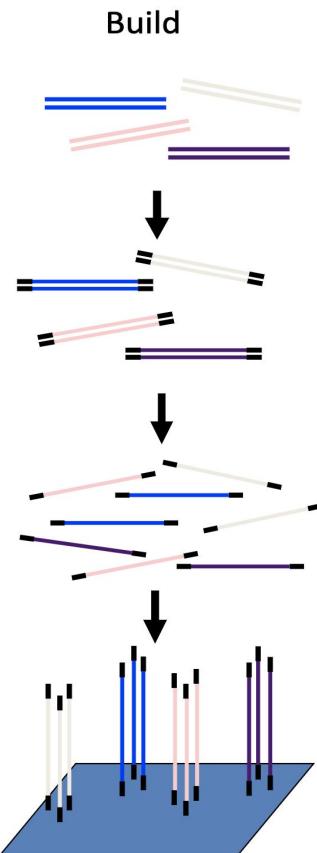
Sequencing recap



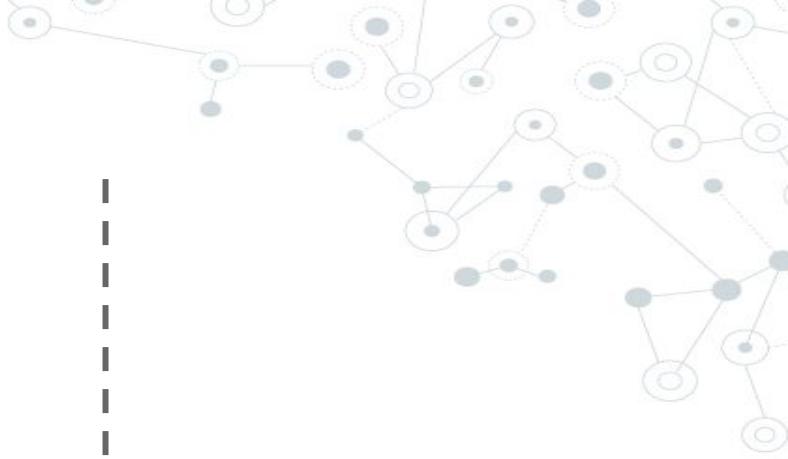
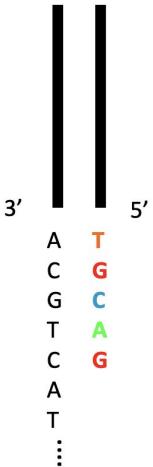
Sequence



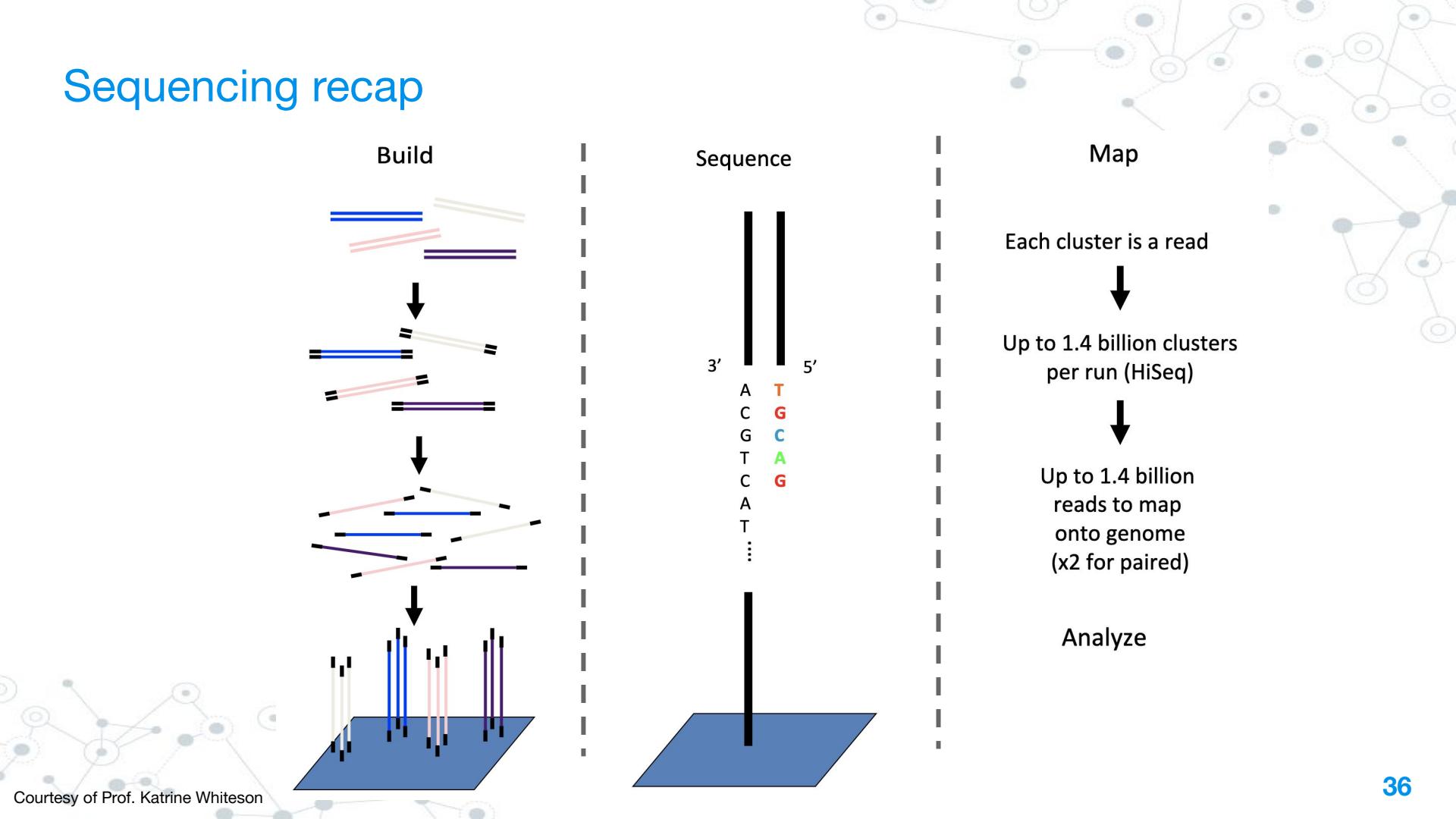
Sequencing recap



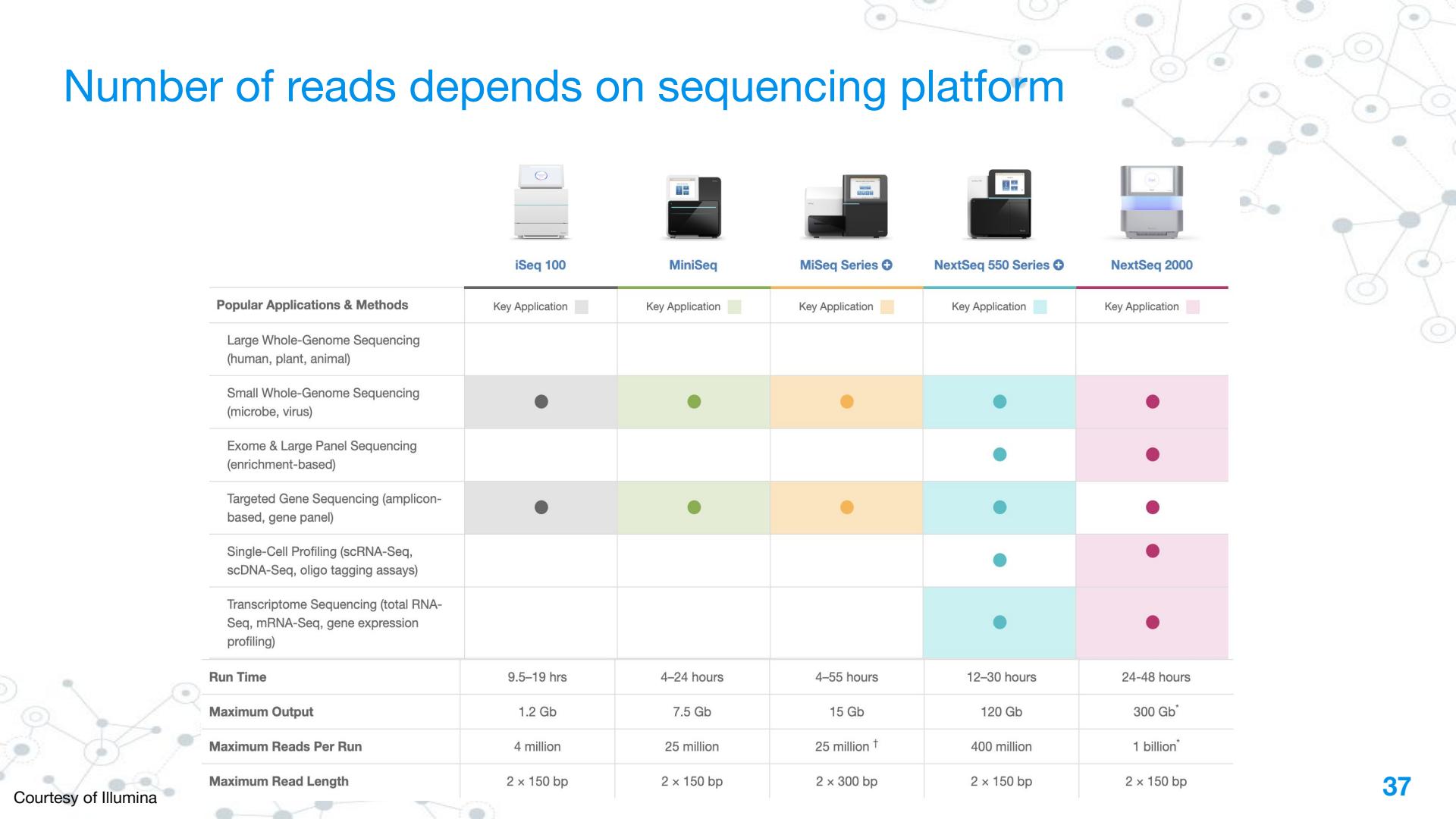
Sequence



Sequencing recap



Number of reads depends on sequencing platform



	iSeq 100	MiniSeq	MiSeq Series	NextSeq 550 Series	NextSeq 2000
Popular Applications & Methods	Key Application	Key Application	Key Application	Key Application	Key Application
Large Whole-Genome Sequencing (human, plant, animal)					
Small Whole-Genome Sequencing (microbe, virus)	●	●	●	●	●
Exome & Large Panel Sequencing (enrichment-based)				●	●
Targeted Gene Sequencing (amplicon-based, gene panel)	●	●	●	●	●
Single-Cell Profiling (scRNA-Seq, scDNA-Seq, oligo tagging assays)				●	●
Transcriptome Sequencing (total RNA-Seq, mRNA-Seq, gene expression profiling)				●	●
Run Time	9.5-19 hrs	4-24 hours	4-55 hours	12-30 hours	24-48 hours
Maximum Output	1.2 Gb	7.5 Gb	15 Gb	120 Gb	300 Gb*
Maximum Reads Per Run	4 million	25 million	25 million †	400 million	1 billion*
Maximum Read Length	2 × 150 bp	2 × 150 bp	2 × 300 bp	2 × 150 bp	2 × 150 bp

Base calling is error-prone regardless of sequencing platform, so each base is assigned a quality score in the raw data

Fastq files are text files where each read has 4 lines

- 1 - Sequencing information, begins with @
- 2- Sequence itself
- 3- Line delimiter
- 4- ASCII quality score

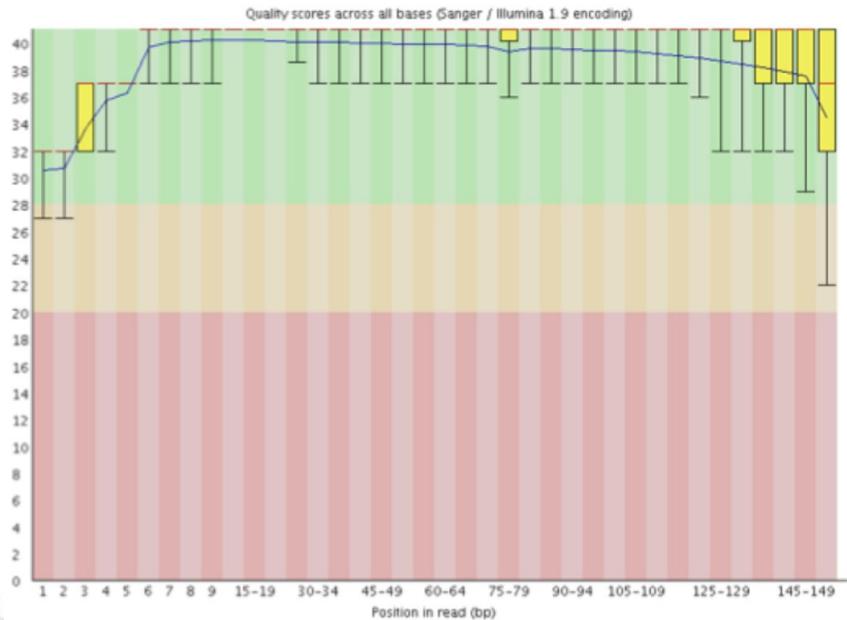
```
1009 $ zcat ER1_R1.fastq.gz | head -n 8
@NS500169:594:H7C5VBGXC:1:11101:22214:1018 1:N:0:CTCTCTAC
GCAGANTGATGCCATCGCATTGCTCTAGCTACCTCTGAGCTGAATGTCGACGTCCCCTGACTTATAATAT
+
AAAAAA#EEEEEEEEE#EEEEEEEEE#EEEEEEEEE#EEEEEEEEE#EEEEEEEEE#EEEEEEEEEAE<66///6
@NS500169:594:H7C5VBGXC:1:11101:22709:1018 1:N:0:CTCTCTAC
GCAGCNTGTTAGCCATCGCATTGCTCATCACCTCTGAGCTGAATCATCAACGCCGTTATCGACTTTTTTTT
+
AAAAAA#EEEEEEEEE#EEEEEEEEE#EEEEEEEEE#EEEEEEEEE#EEEEEEEEE#EEEEEEEEE<<</>/A/6
```

The byte representing quality runs from 0x21 (lowest quality; '!' in ASCII) to 0x7e (highest quality; '~' in ASCII). Here are the quality value characters in left-to-right increasing order of quality ([ASCII](#)):

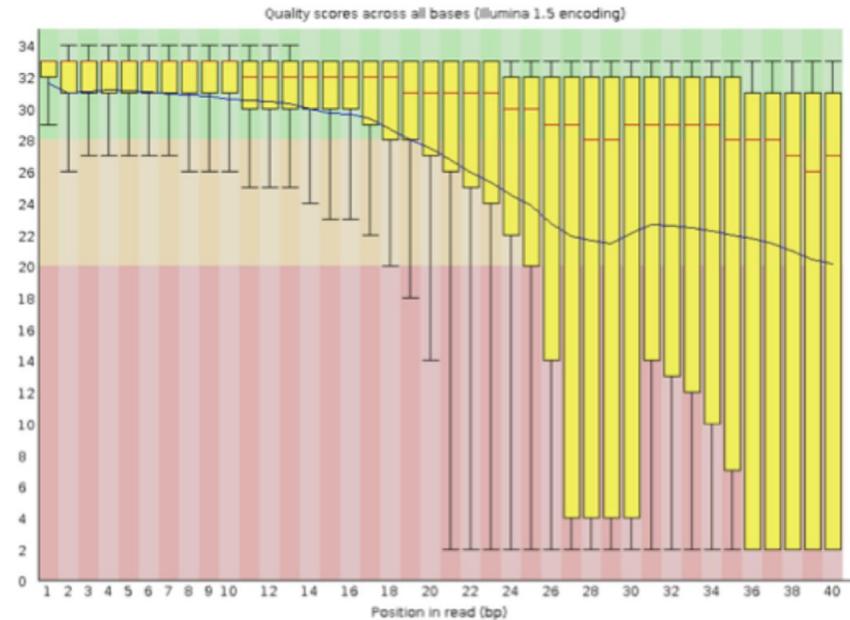
```
! #$%& ' ()*+,--./0123456789 : ;<=>?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\ ]^_`abcdefghijklmnopqrstuvwxyz{| }~
```

QC read quality

Good data



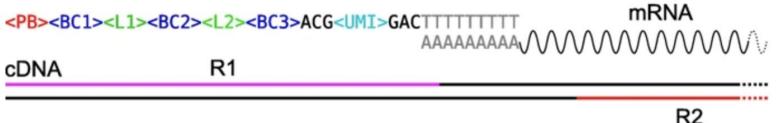
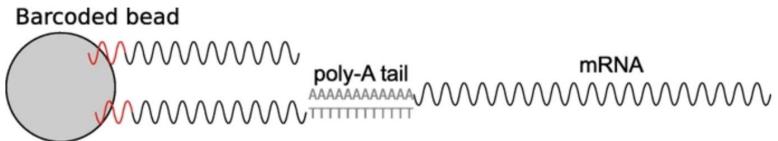
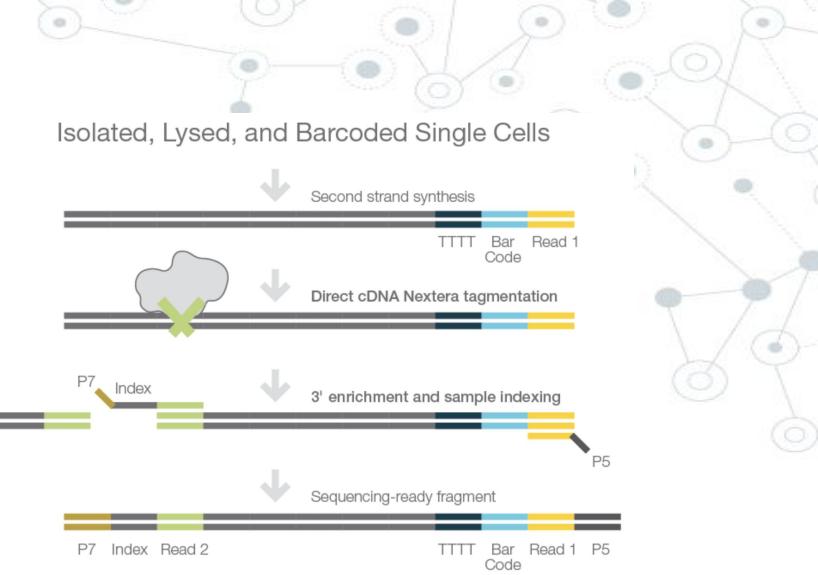
Bad data



Demultiplexing

Multiple barcodes/adapters/indices
“known sequences” are attached to the ends of fragments going into the sequencer

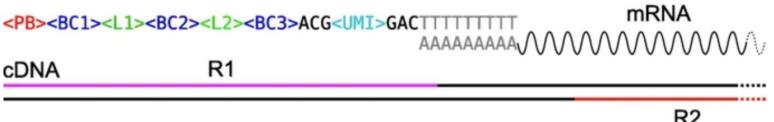
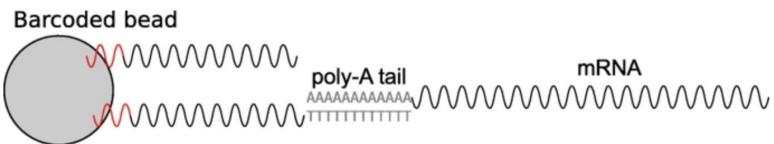
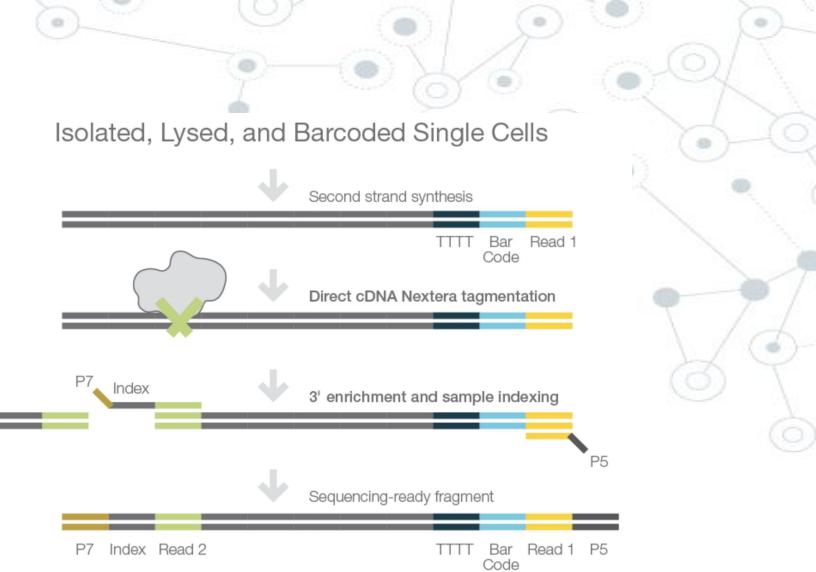
1. P5/P7 primers to attach to flow cell oligo lawn
2. Sample index: one per sample
3. Cell barcode: one per cell
4. UMI: one per molecule



Demultiplexing round 1

Multiple barcodes/adapters/indices
“known sequences” are attached to the ends of fragments going into the sequencer

1. P5/P7 primers to attach to flow cell oligo lawn
2. **Sample index: one per sample**
3. Cell barcode: one per cell
4. UMI: one per molecule



Demultiplexing round 2

Multiple barcodes/adapters/indices

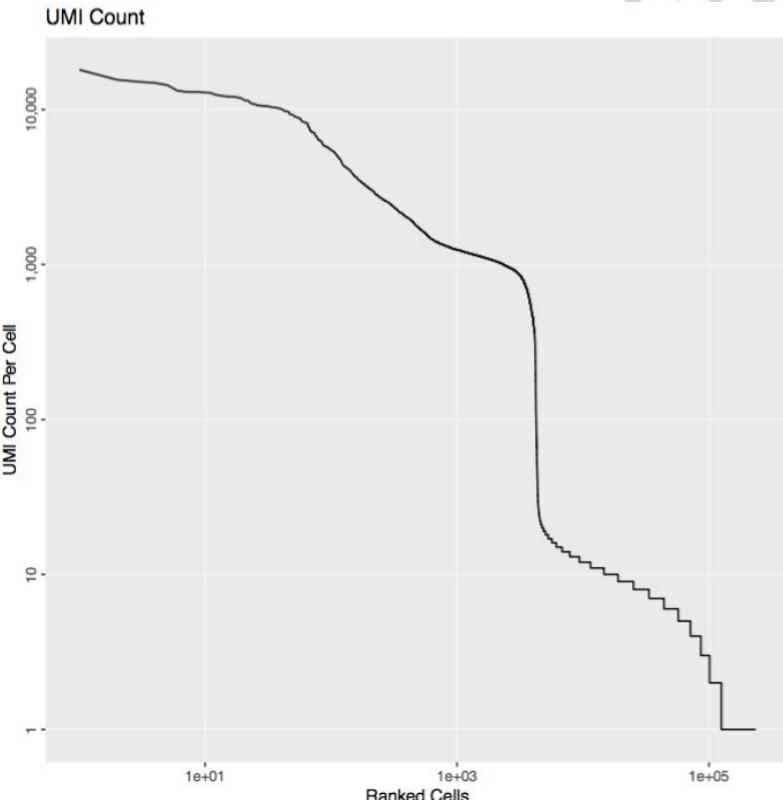
“known sequences” are attached to the ends of fragments going into the sequencer

1. P5/P7 primers to attach to flow cell oligo lawn
2. Sample index: one per sample
3. **Cell barcode: one per cell**
4. UMI: one per molecule

```
1074 $ cat cell1.fastq | head -n 4
@NS500169:594:H7C5VBGXC:1:11101:13683:1029:AAGCCAGGATTGGGTGCT:GGCCCCGCG
ATTTAAAAAATCATTGAAAATGTGATCCACAGCCTAGACATTCTGTCCCTAGCCTTAGACAGACAAGCTGCTCAG
+
///AAEE//EEA//EE//EEEE<E//6/<EAE/EE//E6//EEEEEE<///A///<AEE/A/EE/6AE/6
(base)
Mon Jan 27 19:54:10 [1.79 4.56 5.60] erebboah@hpc-login-1-3:/dfs3/pub/erebboah/data/MB1
1075 $ cat cell2.fastq | head -n 4
@NS500169:594:H7C5VBGXC:1:11101:21510:1028:GAGCTTCAGACTGGGATC:TAGAATAT
CAGGTGACCTATGCTATAGCAGTCTCCTTGAAGTCTGGATAAAATAGTGTCACCTCCCTGTCTCAAATCCAATA
+
A/AAAEE/EEE6AAEE//EE/E6EE<EEEE/EE//AEEE///EE/AA/6EEE6/AA<<<A/A<</6AA<A//A6
```

UMI cutoff QC - “knee plot”

Past the inflection point “knee”, UMIs per cell (barcode) decreases, therefore those barcodes failed in capture and are considered too noisy for further analysis - likely empty droplets without any cell



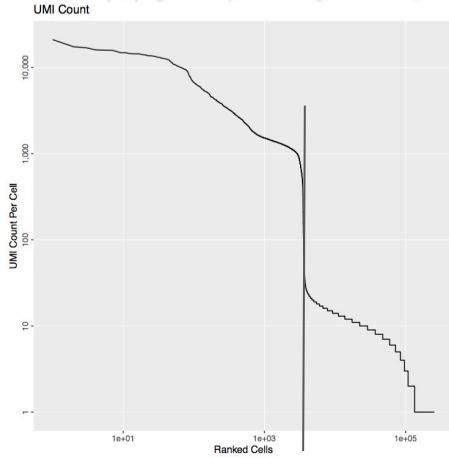
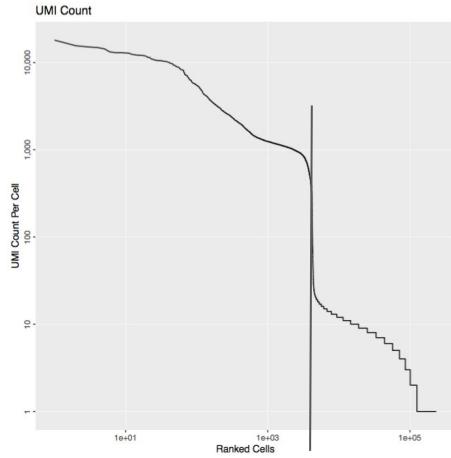
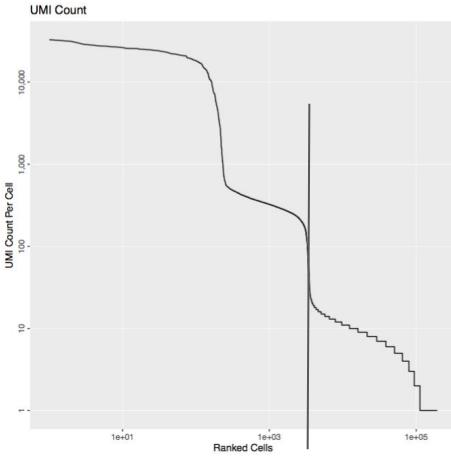
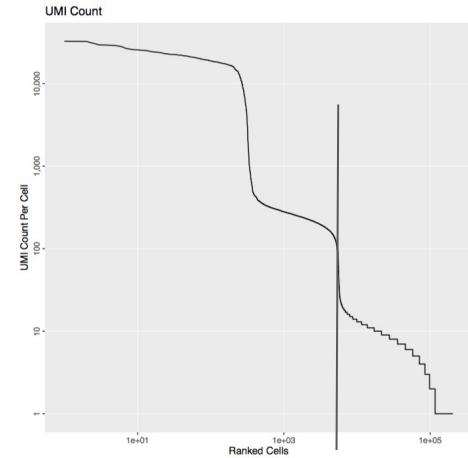
UMI cutoff QC - get rid of noise

MB1

MB2

MT1

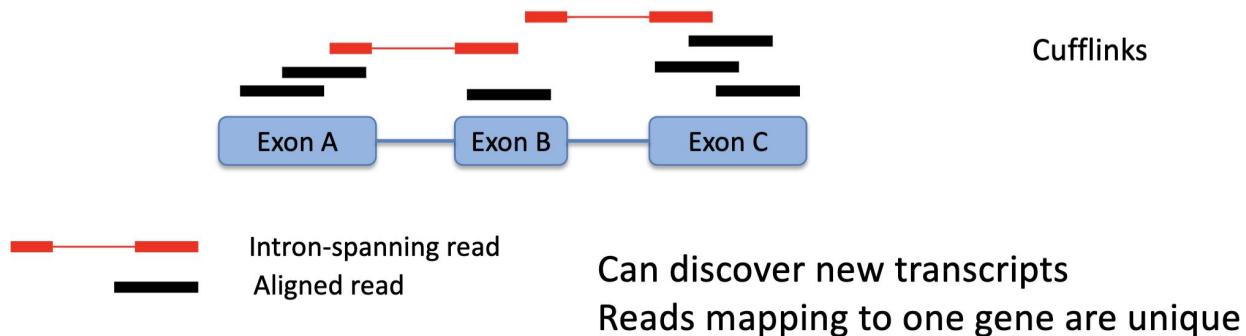
MT2



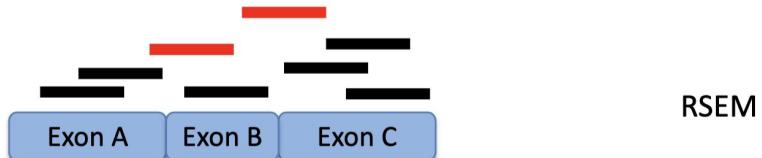
Library	UMI cutoff	Total cells
MB1	200	5,385
MB2	200	3,299
MT1	200	4,180
MT2	200	3,546

Mapping to the reference genome or transcriptome

Genome

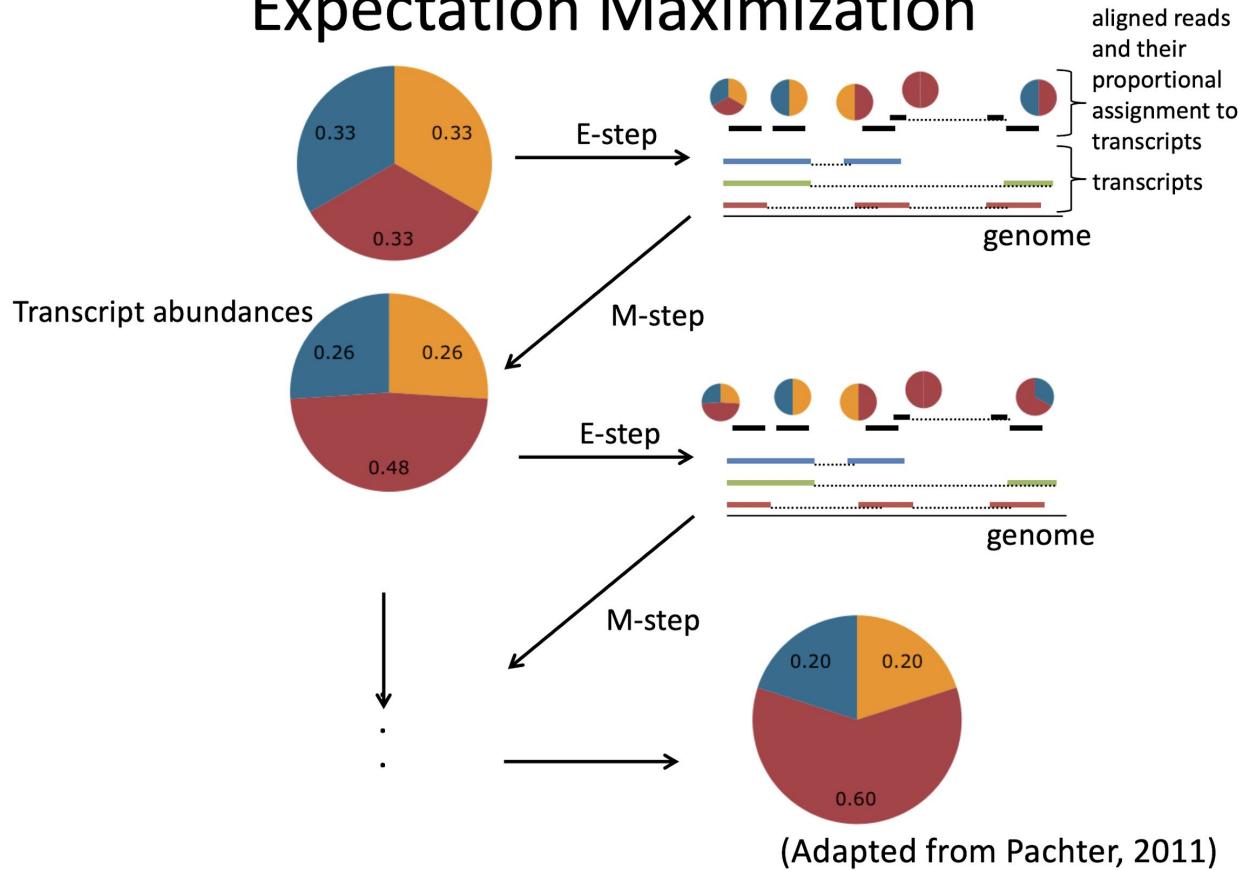


Transcriptome



Must supply a list of known transcripts (no discovery)
Reads mapping to more than one isoform of a transcript are now multireads

Quantitation using Batch Expectation Maximization

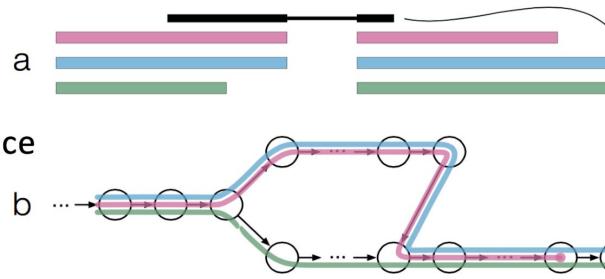


Some problems with Batch EM

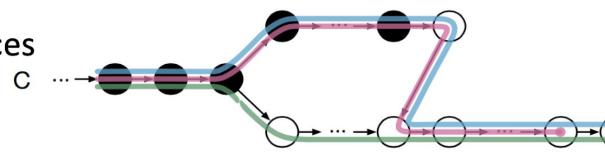
- All reads are mapped before EM.
- Due to multi-mapping, read alignments generate very large files.
- Computationally intensive to read the alignments from file and expensive memory-wise to store them.
- Batch EM algorithm is relatively slow due to multiple iterations performed by the algorithm.
- Cufflinks and RSEM uses batch EM.

Kallisto: streaming EM quantification on a de Bruijn graph

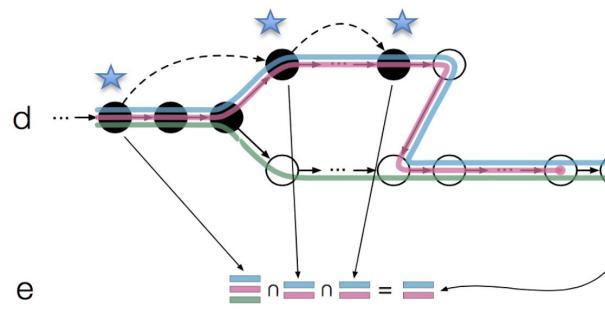
Build a k-mer index on the reference transcriptome graph



A read would match a set of vertices (black)



We can use the informative ones (star)



Calculate which transcripts are supported by the read

Mapping QC - done internally by mapping package used

Started job on	Jan 24 01:03:10
Started mapping on	Jan 24 01:12:36
Finished on	Jan 24 01:17:32
Mapping speed, Million of reads per hour	66.09
Number of input reads	5433855
Average input read length	21
UNIQUE READS:	
Uniquely mapped reads number	1590406
Uniquely mapped reads %	29.27%
Average mapped length	21.26
Number of splices: Total	0
Number of splices: Annotated (sjdb)	0
Number of splices: GT/AG	0
Number of splices: GC/AG	0
Number of splices: AT/AC	0
Number of splices: Non-canonical	0
Mismatch rate per base, %	1.06%
Deletion rate per base	0.00%
Deletion average length	1.00
Insertion rate per base	0.00%
Insertion average length	1.00
MULTI-MAPPING READS:	
Number of reads mapped to multiple loci	2102131
% of reads mapped to multiple loci	38.69%
Number of reads mapped to too many loci	322235
% of reads mapped to too many loci	5.93%
UNMAPPED READS:	
% of reads unmapped: too many mismatches	19.08%
% of reads unmapped: too short	1.94%
% of reads unmapped: other	5.09%
CHIMERIC READS:	
Number of chimeric reads	0
% of chimeric reads	0.00%

Counts matrix

<https://tinyurl.com/sbtgb5a>

Library	Total cells/nuclei
MB1 (cells)	5,385
MB2 (cells)	3,299
MT1 (nuclei)	4,180
MT2 (nuclei)	3,546

	MB1.cell1	MB1.cell2	MB1.cell3	MB1.cell4	MB1.cell5	MB1.cell6	MB1.cell7	MB1.cell8	MB1.cell9	MB1.cell10
Galnt1	5	1	3	2	0	3	5	1.00	1	4
Myf5	2	0	0	0	0	1	1	1.00	0	1
Mkrn2	0	0	0	0	0	0	0	0.00	0	0
Pparg	0	1	0	1	0	0	0	0.00	1	1
Raf1	3	0	4	2	1	2	0	4.00	1	0
Sept1	0	0	0	0	0	0	0	0.00	0	0
Pdgfb	0	0	0	0	0	0	0	1.00	0	0
Acvr1l	0	0	0	0	0	0	0	0.00	0	0
Grasp	0	1	0	0	0	0	1	0.00	0	0
Acvr1b	0	0	0	0	0	0	0	0.00	0	1
Tom1l2	0	1	1	1	0	0	0	0.00	0	0
Gpa33	0	0	0	0	0	0	0	0.00	0	0
Zfp385a	0	0	0	0	0	0	0	0.00	0	1
Itga5	0	0	0	1	0	0	0	1.00	0	0
Wdr77	2	3	1	4	3	0	6	0.00	4	3
Atp5f1	21	34	12	9	7	12	10	12.74	11	14
Sox9	0	12	1	0	0	0	1	0.00	1	1
Hnrnpd	3	0	1	1	2	1	1	0.00	1	0



https://satijalab.org/seurat/v3.1/pbmc3k_tutorial.html

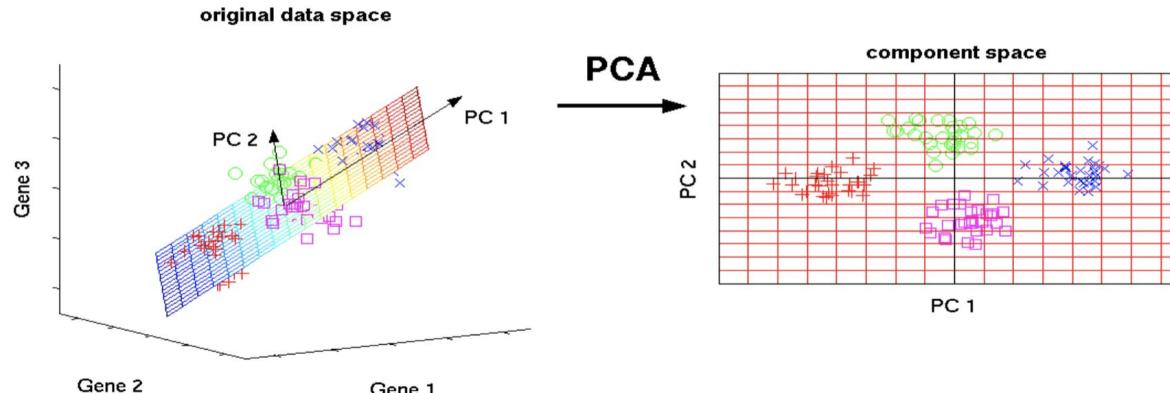
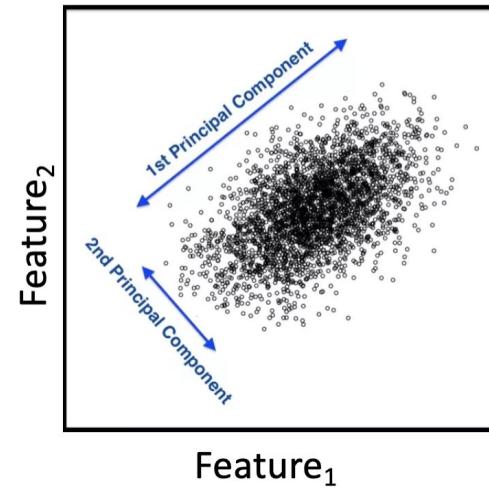
Principal Component Analysis

- Sometimes, it's a linear combination of features that best captures a relationship, e.g.

$$Y_1 = a_{11} \text{ Feature}_1 + a_{12} \text{ Feature}_2$$

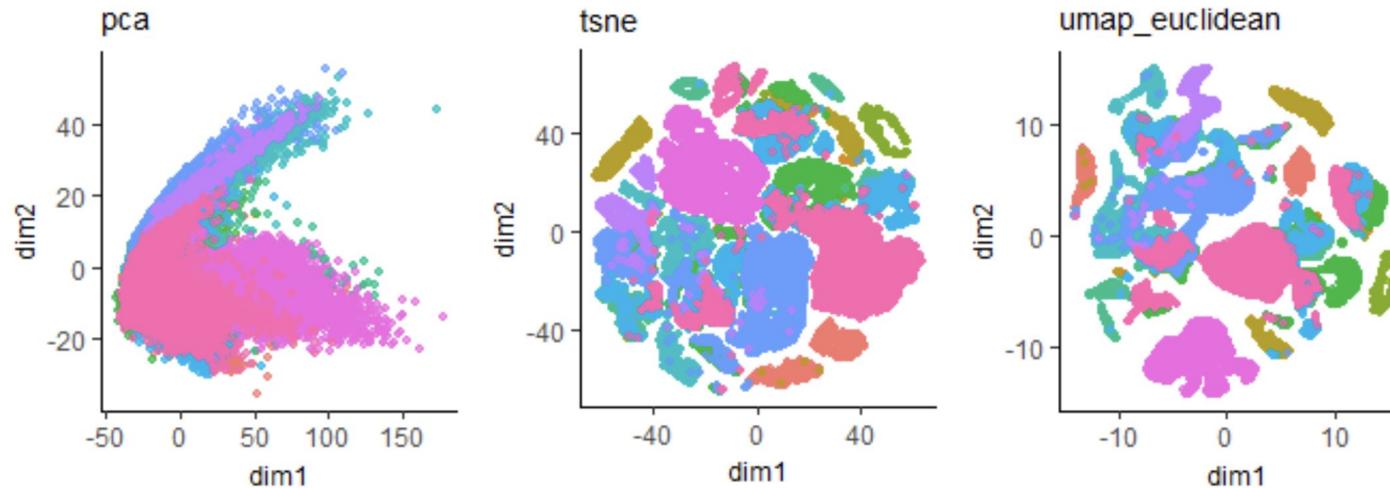
$$Y_2 = a_{21} \text{ Feature}_1 + a_{22} \text{ Feature}_2$$

- PCA is linear transformation into a new coordinate system, where orthogonal PCs maximize the variance in successive order, i.e. PC1 accounts for most of the variance, then PC2, etc...
- Instead of many features, we only need to worry about fewer PCs → dimensionality reduction



Courtesy of Prof. Ali Mortazavi

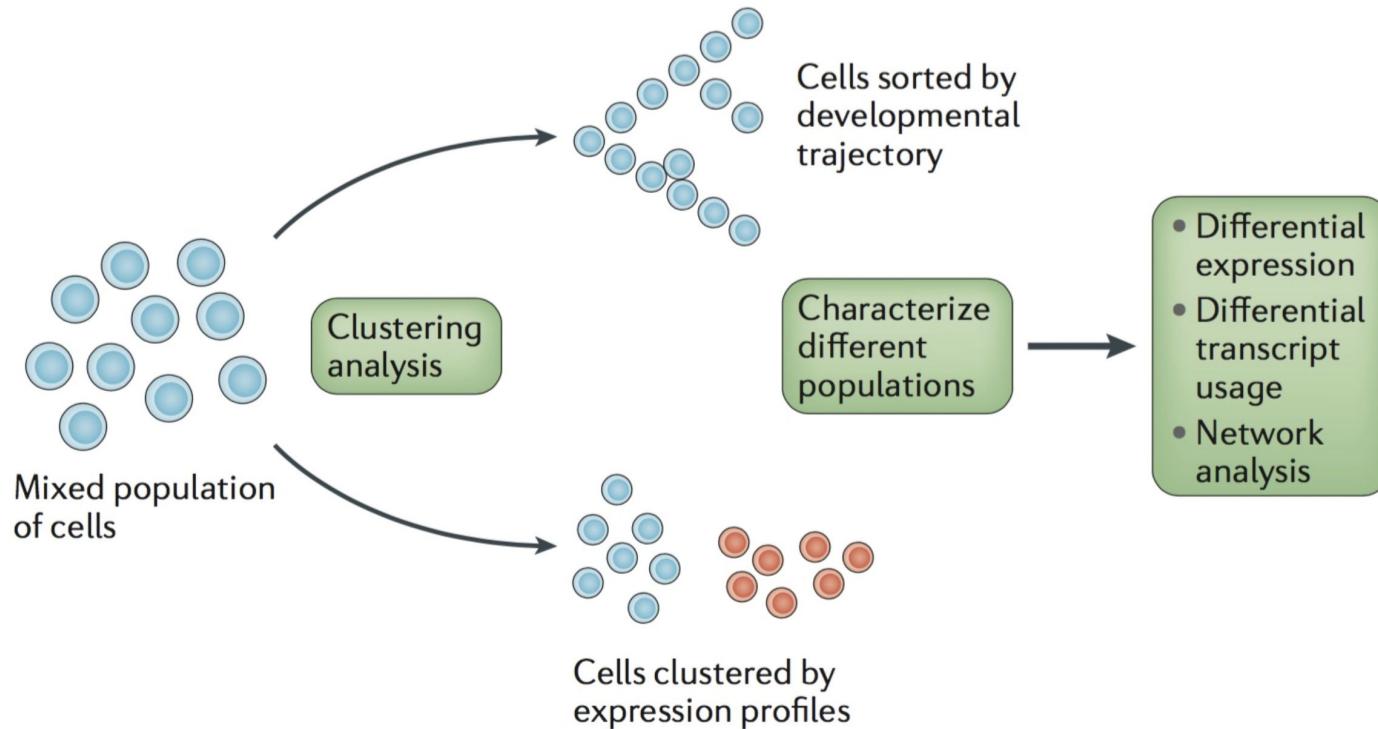
Non-linear Dimensionality Reduction Techniques can reveal additional data structure



tissue

- Bladder • Kidney • Lung • Marrow • Spleen • Tongue
- Heart • Liver • Mammary • Muscle • Thymus • Trachea

Finding markers to separate subgroups of cells



Differential expression analysis

- **Goal:**

To determine the set of differentially expressed genes, transcripts between two conditions or time points in a time-course experiment.
- **Approach:**
 1. Determine gene, transcript abundances.
 2. Test for differentially expressed genes and transcripts e.g. test for statistical significance of differences between the expressions of genes from different conditions, time points.

Several databases are designed for data-mining



Molecular Function
Biological Process
Cellular Component



Functional motifs

Example proteins



More proteins

IPR000174 Interleukin-8 receptor



IPR000276 Rhodopsin-like GPCR superfamily



IPR001277 C-X-C chemokine receptor, type 4



IPR001355 Interleukin 8A receptor



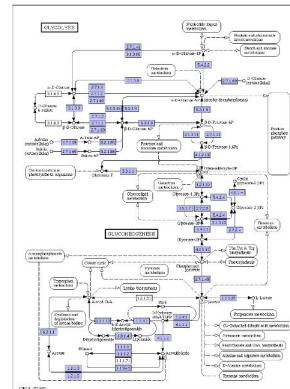
ModBase



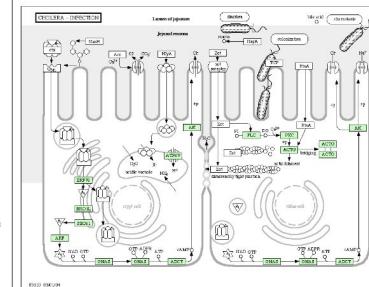
PDB Chain



Metabolic pathways



KEGG orthologues



Slides courtesy of Ana Conesa

GO: Gene Ontology

- ✓ Project developed by the **Gene Ontology Consortium**
- ✓ Provides a **controlled vocabulary** to describe gene and gene product attributes in **any organism**
- ✓ Lastest version more than 43,000 terms
- ✓ Includes both the development of the **Ontology** and the maintenance of a **Database** of annotations

The three categories of GO

Molecular Function (10,000+)

the tasks performed by individual gene products; examples are *transcription factor* and *DNA helicase*

Biological Process (29,000+)

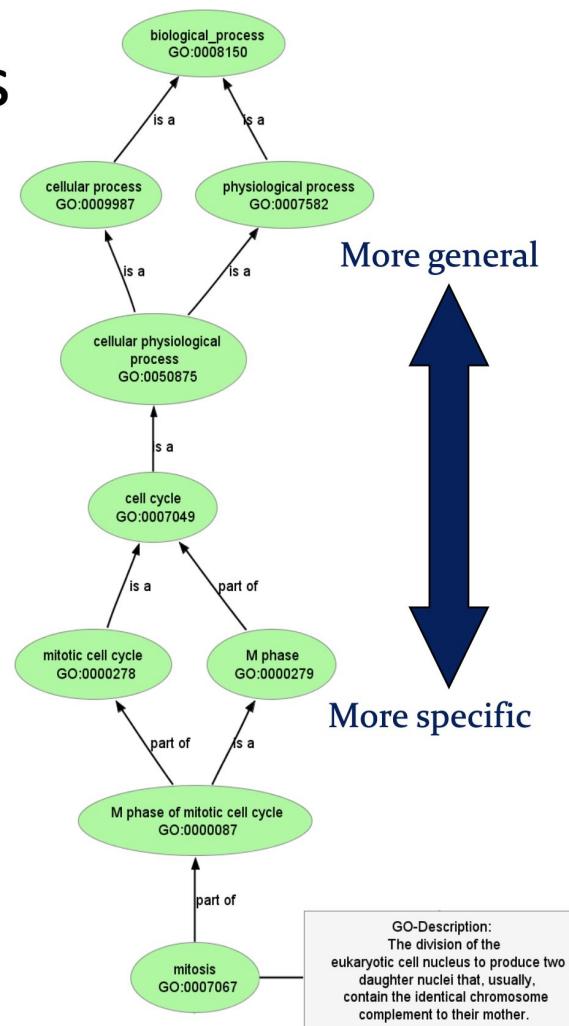
broad biological goals, such as *mitosis* or *purine metabolism*, that are accomplished by ordered assemblies of molecular functions

Cellular Component (4,000+)

subcellular structures, locations, and macromolecular complexes; examples include *nucleus*, *telomere*, and *origin recognition complex*

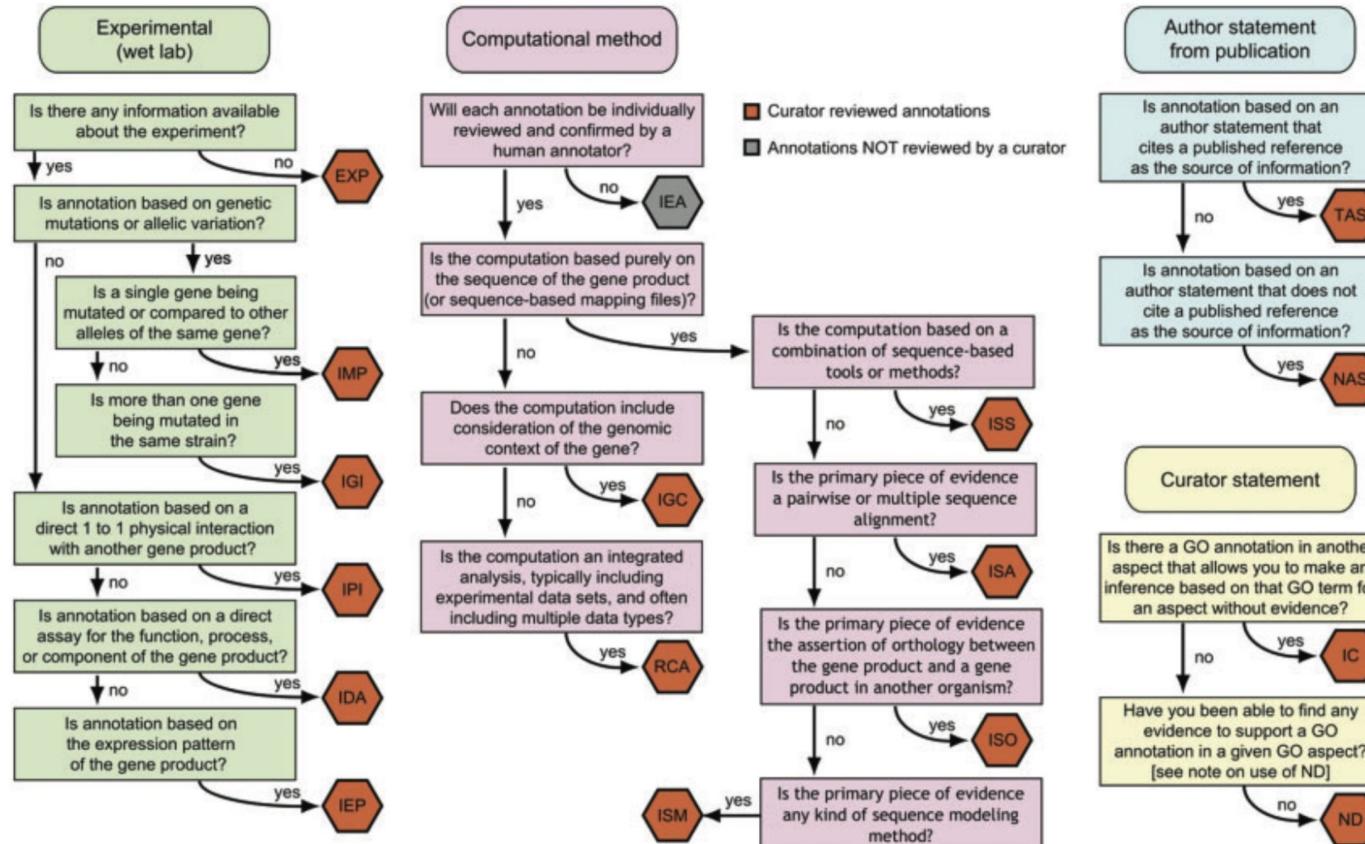
The GO hierarchical levels

- ✓ Annotations are given to the most specific (low) level
- ✓ True path rule: annotation at a given term implies annotation to all its parent terms



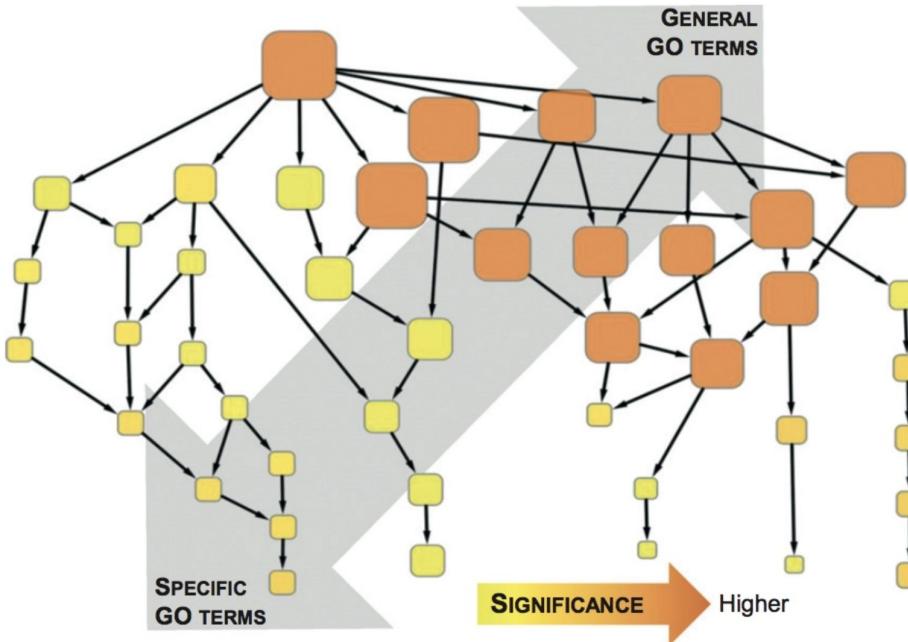
Slides courtesy of Ana Conesa

Annotations are given an evidence code



(du Plessis, 2011)

Enrichment analyses of GO terms favor more general terms



(Primmer, 2013)