

RNA-Seq

Lutz Froenicke

DNA Technologies and Expression Analysis Core

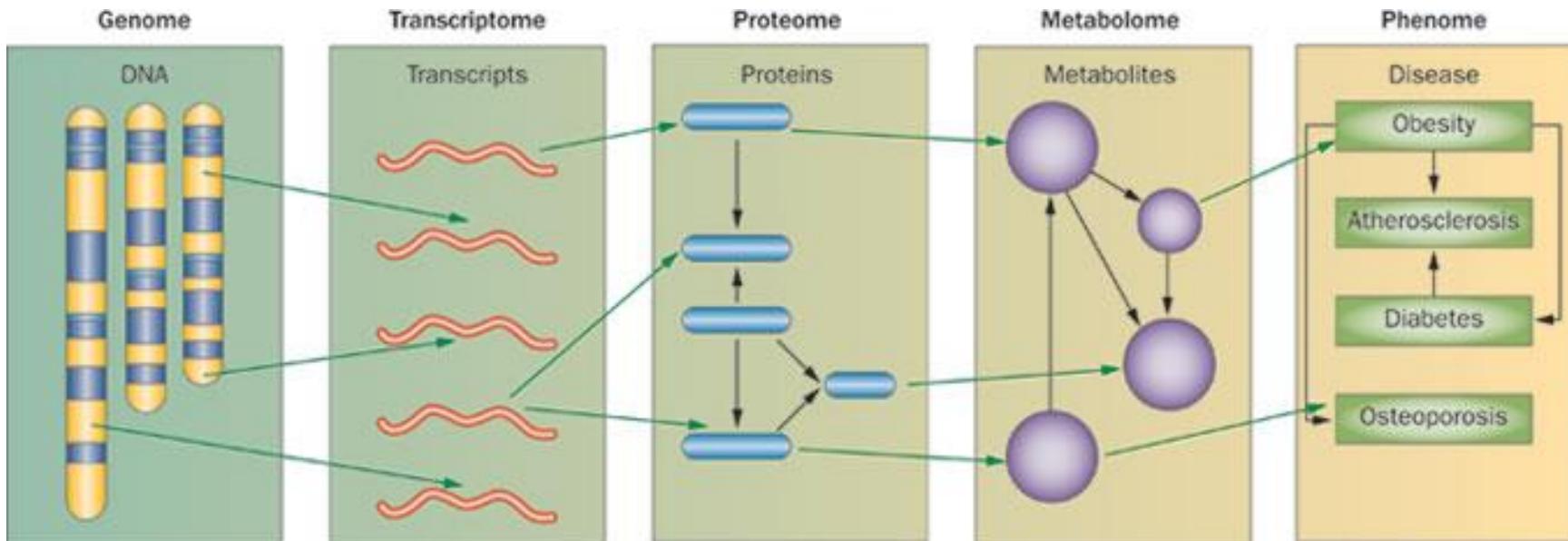
UC Davis Genome Center



Outline

- Who are we and what are we doing?
- Overview HTS sequencing technologies
- How does Illumina sequencing work?
Sequencing library and run QC
- How does RNA-seq work?
- PacBio and Nanopore Sequencing
- Some cutting edge protocols & applications

The UCD GENOME CENTER



DNA Tech & Expression Analysis Proteomics Core Metabolomics Core

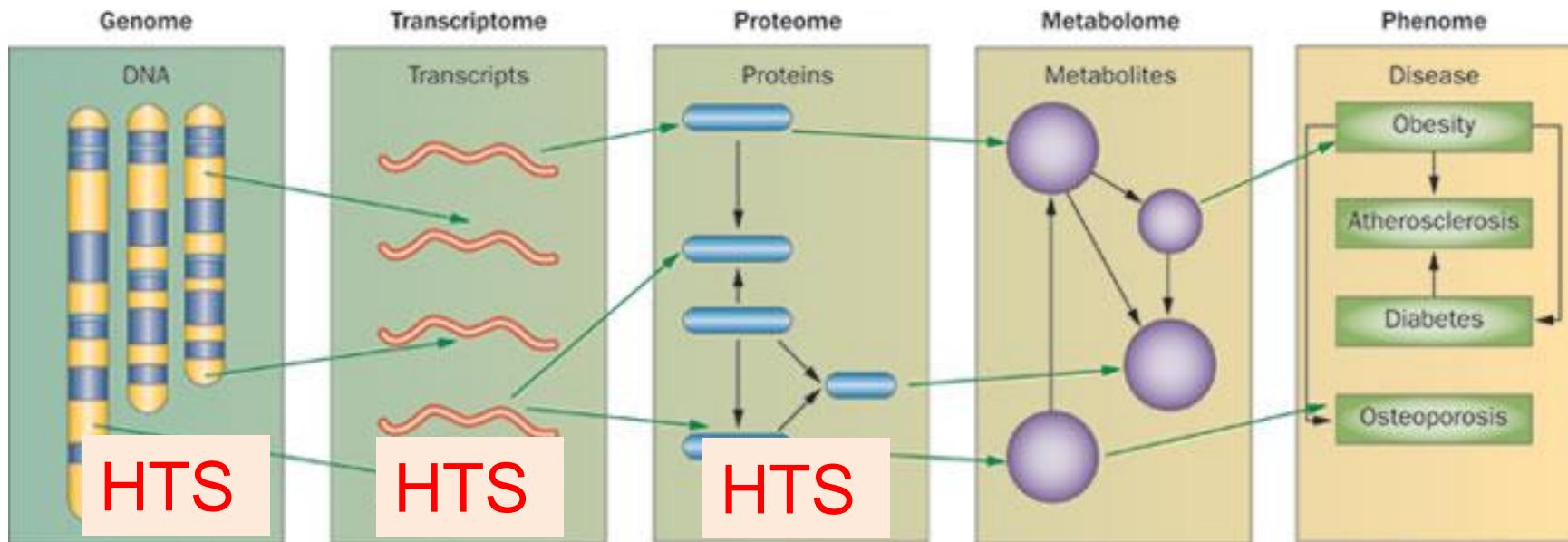
“DNA makes RNA and RNA makes protein”
the Central Dogma of Molecular Biology; simplified from Francis Crick
1958

nature
REVIEWS **CARDIOLOGY**

MacLellan, W. R. et al. (2012) Systems-based approaches to cardiovascular disease
Nat. Rev. Cardiol. doi:10.1038/nrccardio.2011.208

“DNA makes RNA and RNA makes protein”

the Central Dogma of Molecular Biology; simplified from Francis Crick 1958



DNA Tech & Expression Analysis Proteomics Core Metabolomics Core

The UCD GENOME CENTER

MacLellan, W. R. et al. (2012) Systems-based approaches to cardiovascular disease
Nat. Rev. Cardiol. doi:10.1038/nrcardio.2011.208

nature
REVIEWS
CARDIOLOGY



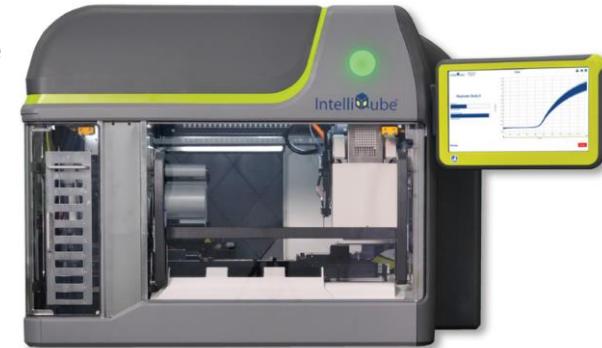
Genome Center Core service lab - working with UCD, UC, and outside clients; ~ ten staff members

Pioneering technologies - Illumina Genome Analyzer (2007), PacBio RS (2012), Fluidigm C1 (2014), 10XGenomics (2016), ONT PromethION (2018)



We offer

- High Throughput Sequencing in all shapes and sizes: Illumina, PacBio (two Sequel II), Nanopore & custom sequencing projects
- **Single-cell RNA-seq, sc-MultiOme** - 10X Genomics, Parse Biosciences
- Optical Genome Mapping (Bionano Saphyr)
- Free consultations -- in collaboration with the Bioinformatics Core
- Sequencing library preparation workshops (not in pandemics)
- High throughput COVID testing (in pandemics; > 600,000 tests)



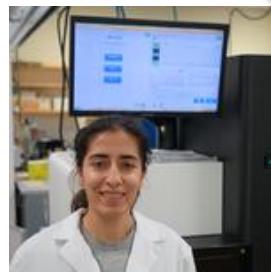
Integrated RT-PCR systems for HT COVID testing

The DNA Tech Core Team

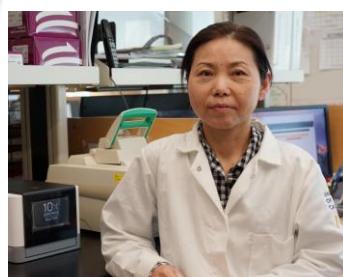
Single-cell



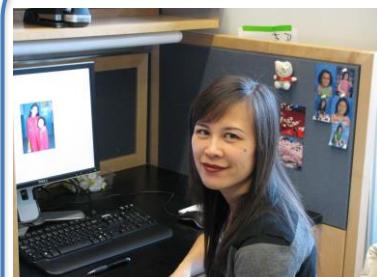
Emily



Claire



Hong



Oanh



Ruta



Siranoosh



Aaron



Vanessa



Noravit



Mohan

Illumina

Testing

PacBio

Nanopore /
HMW DNA

<https://dnatech.genomecenter.ucdavis.edu/faqs/>

Do you offer DNA isolations and RNA isolations as a service?
 How to prepare samples for multiplexed amplicon sequencing on Illumina systems?
 How do I ship RNA samples? How do I ship RNA samples if the transport will take a long time?
 How should I prepare and sequence samples for ChIP-seq?
 Do you have recommendations for the isolation of plant total RNA samples?
 How to prepare samples for multiplexed amplicon sequencing on the PacBio Sequel?
 How do I prepare DNA samples for RR-Seq (reduced representation sequencing)
 Which protocols or kits do you recommend for RNA isolations from human and animal samples? How many cells will I need?

Search FAQs

Search

01 General Information (20)

How do I get started working with the Core facilities? How do I set up an account with the Genome Center?

How do I set up an account in the PPMS ordering and instrument reservation system?

How do I make an appointment for a consultation?

How and when will I be invoiced?

How do I submit samples or libraries?

What is a Purchase Order? How do I create a PO?

How Do I Contact You?

Who do I ask for administrative/billing questions?

How do I subscribe to your newsletter or listserv?

Can I use Core facility equipment?

What Information Do You Require About My Project?

Where are you located?

How do I get into the Genome Center Building?

Do you archive submitted samples? Do you return samples?

When can I visit you / reach you?

Do you provide Bioinformatics help?

How do I acknowledge your services?

Do you ask for co-authorships?

Do you support pilot projects with Seed Grants?

What happened to the BGI@UC DAVIS ?

02 Prices or Recharge Rates (4)

What are the prices associated with genotyping and sequencing?

How and when will I be invoiced?

Which recharge-rate scale (price scale) will apply to my project?

What is a Purchase Order? How do I create a PO?

03 Sample Preparation & Sample Requirements (18)

How should I QC my genomic DNA samples before sequencing?

Which DNA isolation protocols do you recommend for Illumina sequencing?

How to submit samples for Labchip GX RNA-QC and fragment analysis?

What are the sample requirements for DNA and RNA samples or for sequencing libraries ?

What type of samples are recommended for the isolation of HMW-DNA? (for Long-Read Sequencing)

What type of samples are recommended for RNA isolations for gene annotations?

Can you run samples with less than the recommended input material?

Can I submit samples of lower integrity than recommended?

How should I purify my samples? How should I remove DNA or RNA contamination?

How to purify DNA samples for long-read sequencing (PacBio, Nanopore)? How to remove polysaccharides?

Do you offer DNA isolations and RNA isolations as a service?

How to prepare samples for multiplexed amplicon sequencing on Illumina systems?

How do I ship RNA samples? How do I ship RNA samples if the transport will take a long time?

How should I prepare and sequence samples for ChIP-seq?

Recent Posts

Holiday Schedule

Core operations resume in Phase 2 of the COVID response

DNA Technologies Core has to ramp down lab work Adjusting DNA Tech Core operation to the COVID-19 guidelines

Join us for the PacBio Day Symposium — February 26th

Latest Tweets

How the US lets hot school days sabotage learning. Both, a hotter environment and lack of air conditioning are impa... <https://t.co/4IuLyUKyRb>, Jun 10

This looks like a breakthrough in whole-genome-amplification (WGA). A very nifty modification of the MDA protocol. <https://t.co/RTzAQhPyZ>, Jun 10

Horizontal gene transfer in fishes: <https://t.co/qD9GYXePib> Turbid water caused by spawning herrings: <https://t.co/2uaWMFj6>, Jun 9

Genome-sequence based phylogeny of angiosperms. "synteny-based species tree shows high resolution and overall stron... <https://t.co/sI6TodoolL>, Jun 9

Single nuclei RNA sequencing reveals two halves of the hippocampus have different gene activity <https://t.co/UplPgxOc4w>, Jun 2

Important Information

About Us / Services

Illumina Library Construction Services

Illumina Library Sequencing Services

PacBio Sequel II Library Prep & Sequencing

Getting Started Guide

Sample and Library Requirements

Sample Submission, Scheduling, Shipping

Subscribe To Our Newsletter

Archives

December 2020

June 2020

March 2020

February 2020

September 2019

July 2019

March 2019

December 2018

October 2018

September 2018

August 2018

June 2018

04 Library Preparation and QC (12)

How do I size select libraries for the HiSeq 4000 with beads?

Do you recommend PCR-free sequencing library preparations?

What are the sample requirements for DNA and RNA samples or for sequencing libraries ?

When do you recommend 3'-Tag RNA-seq?

How to remove primer contamination from sequencing libraries? (free primers)

My libraries show peaks larger than expected. Can I still sequence these PCR-bubbles?

Which indexing scheme should I use for Illumina sequencing to prevent index hopping? (UDI adapter)

How should I prepare and sequence samples for ChIP-seq?

How many indexes are available for my libraries?

How do I pool sequencing libraries? Can you pool them for me?

Which strand is sequenced for my strand-specific RNA-seq data?

My PCR-free libraries do not look as expected on the Bioanalyzer. How should I QC PCR-free libraries?

05 Sequencing (17)

How should I submit the barcode sequence information? In which direction will they be sequenced?

Can I submit a library I made?

What are UMLs and why are they used in high-throughput sequencing?

Where can I find a tutorial on Illumina and NGS sequencing?

Which read numbers/yields can I expect from Illumina sequencing?

My libraries show peaks larger than expected. Can I still sequence these PCR-bubbles?

Which indexing scheme should I use for Illumina sequencing to prevent index hopping? (UDI adapter)

How should I prepare and sequence samples for ChIP-seq?

How should I sequence ATAC-seq libraries?

What type of library services are available through the Core?

In which form will I receive the data?

When can I expect to get my data?

Can I reserve a spot in the queue?

Do you store samples and sequencing libraries?

Can I use custom sequencing primers? What melting temperatures should these have?

Do you offer 16S sequencing and 18S sequencing?

Sanger DNA sequencing and other services at UC Davis

06 Sequencing Data (13)

What data will I receive for Illumina sequencing? Demultiplexing, Trimming, Filtering

When should I trim my Illumina reads and how should I do it?

Should I remove PCR duplicates from my RNA-seq data?

Why does FASTQC show unexpectedly high sequence duplication levels (PCR-duplicates)?

Where can I find the UMLs in the Tag-Seq data? When and how should I trim my Tag-Seq data? What is the low complexity stretch in the Tag-Seq data?

Which data will I receive from the PacBio Sequel II sequencer? Will they have quality scores?

How do I download my sequencing data?

Do you de-multiplex the sequencing data?

My FASTQ file contains some "N"s. Is there a problem with my data?

Where and how can I get my data?

Do you archive the sequencing data?

How should the miRNA/small-RNA data be trimmed?

Which strand is sequenced for my strand-specific RNA-seq data?

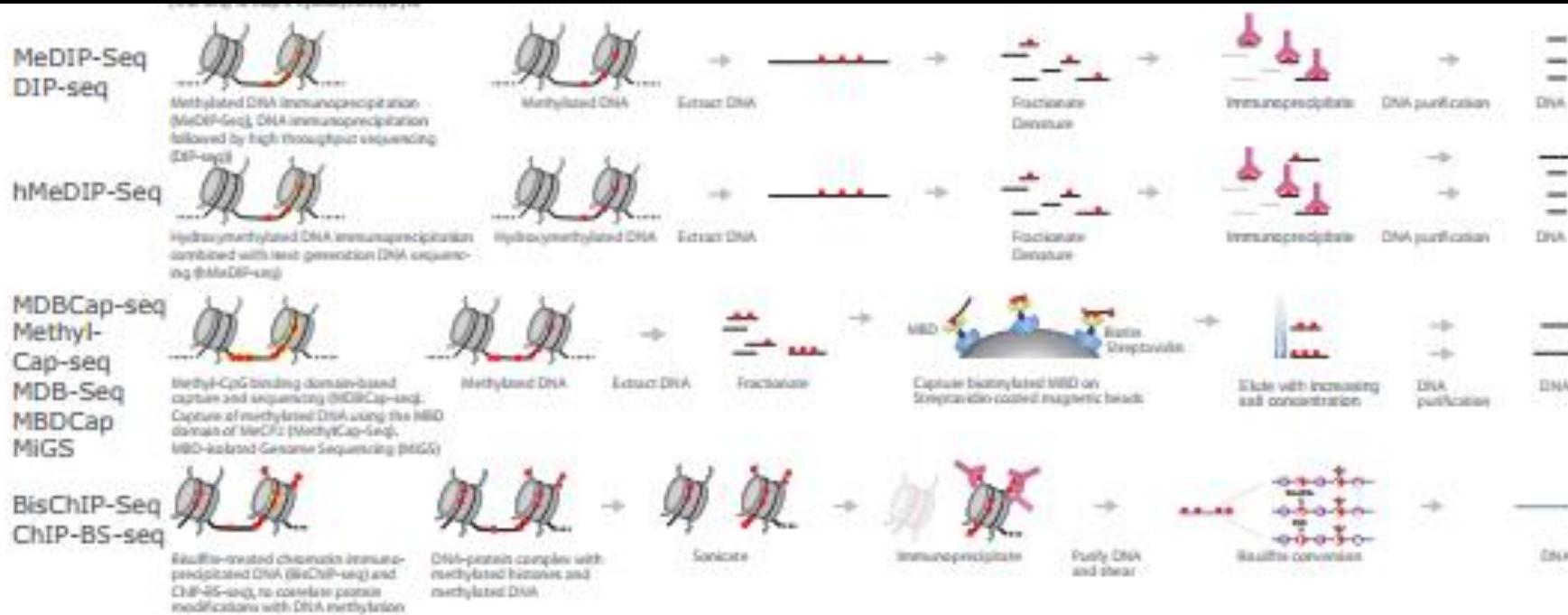
Complementary Approaches

Illumina	PacBio	PromethION Nanopore
Still-imaging of clusters (~1000 clonal molecules)	Movie recordings fluorescence of single molecules	Recording of electric current through a pore
Short reads - 2x300 bp Miseq	Up to 70 kb, N50 25 kb	Up to 70 kb, N50 25 kb
Repeats are mostly not analyzable	spans retro elements	spans retro elements
High output - up to 2.4 Tb per lane	up to 100 Gb per SMRT-cell, up to 20 Gb HiFi data per cell	Up to 100 Gb per flowcell
High accuracy (< 0.5 %)	Raw data error rate 15 % CCS data < 0.1%	Raw data error rate 2-10 %
Considerable base composition bias	No base composition bias	Some systematic errors
Very affordable	Costs 5 to 10 times higher	Costs same or 2x higher
<i>De novo</i> assemblies of thousands of scaffolds	“Near perfect” genome assemblies; lowest error rate	“Near perfect” genome assemblies with compl. data; highest contiguity

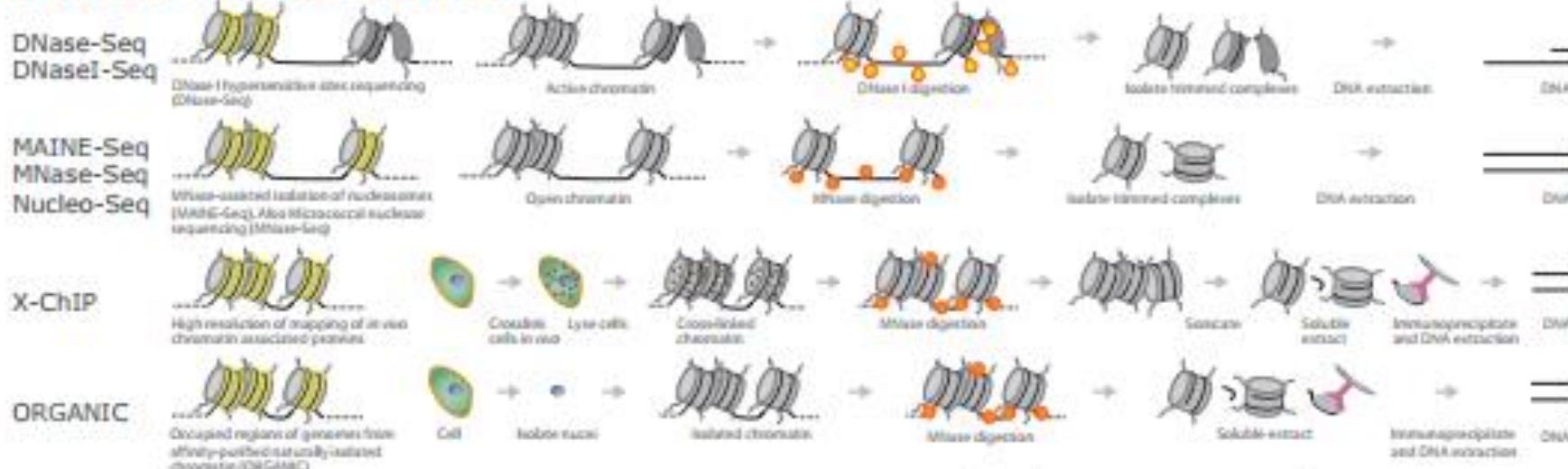
High Throughput Short Read Sequencing: Illumina



- Whole genome sequencing & Exome sequencing:
Variant detection (small variants SNPs and indels)
Copy number variation (CNVs; prenatal diagnostics)
- Genotyping by sequencing
- Genome assemblies: small genomes
- Metagenomics
- RNA-seq: gene expression, transcript expression
- Small RNA-seq
- Single-cell RNA-seq
- Epigenetics: Methyl-Seq:
- ChIP-Seq (detecting molecular interactions)
- 3D Organization of the nucleus (Hi-C)



DNA-Protein Interactions



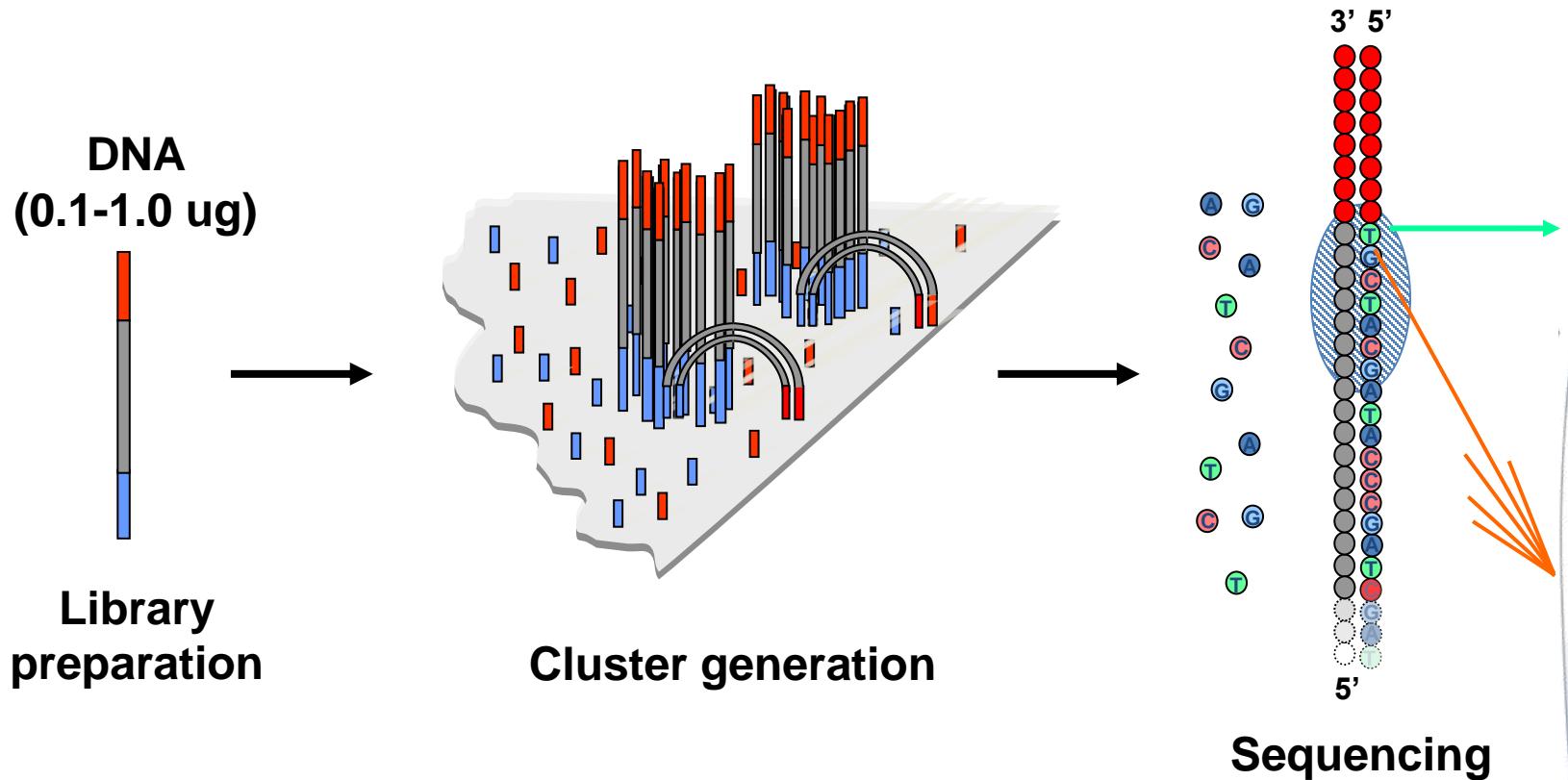
Long Read Sequencing: PacBio and Nanopore



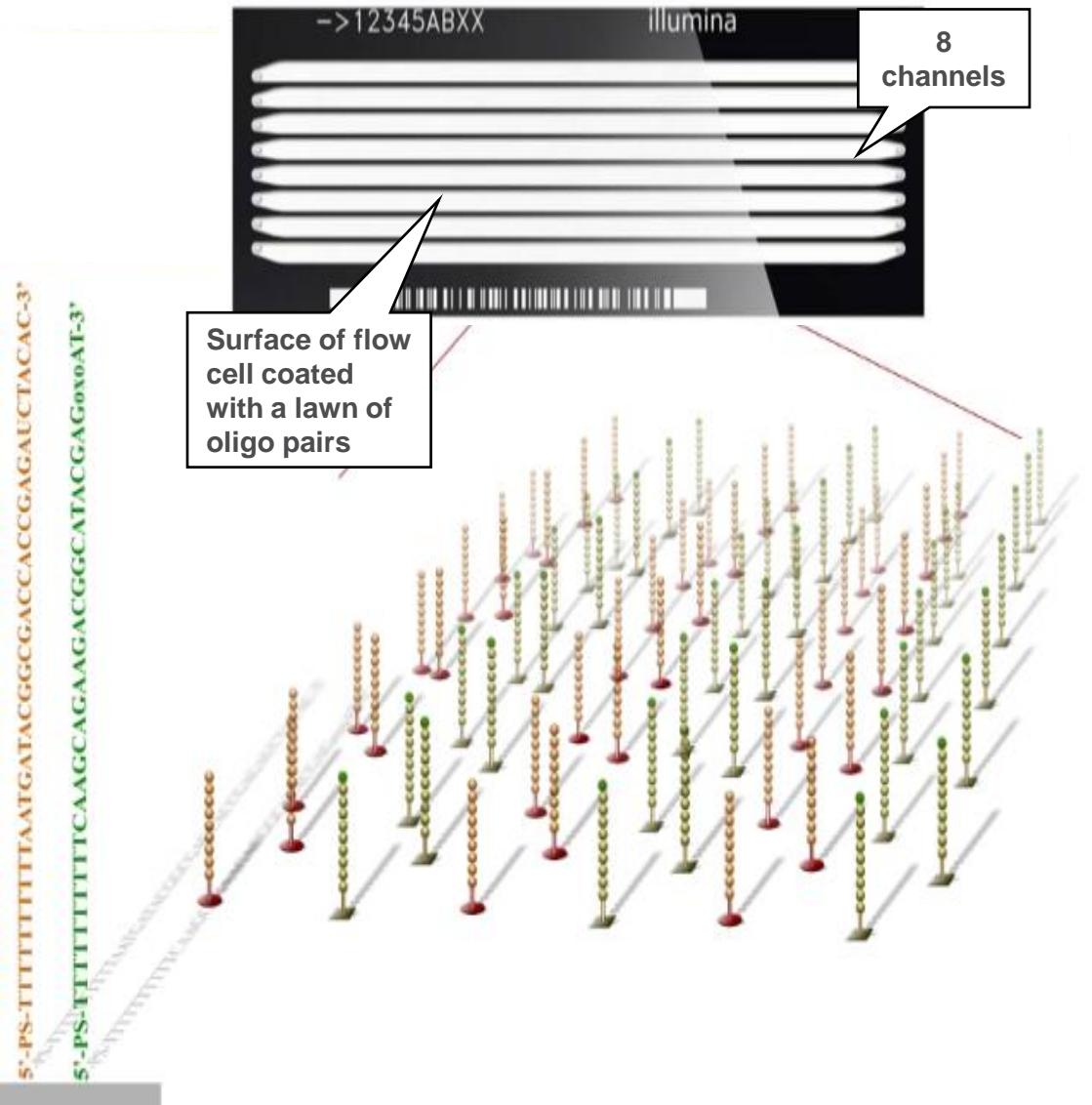
- Whole genome sequencing: Highest quality genome assemblies, Structural variant detection
- RNA-sequencing
 - full transcript data, Iso-form detection and quantification
 - Direct RNA-seq identifies base modifications (Nanopore)
- Metagenomics
- Epigenetics (Nanopore: any modified bases, PacBio bacteria)

Illumina Sequencing Technology

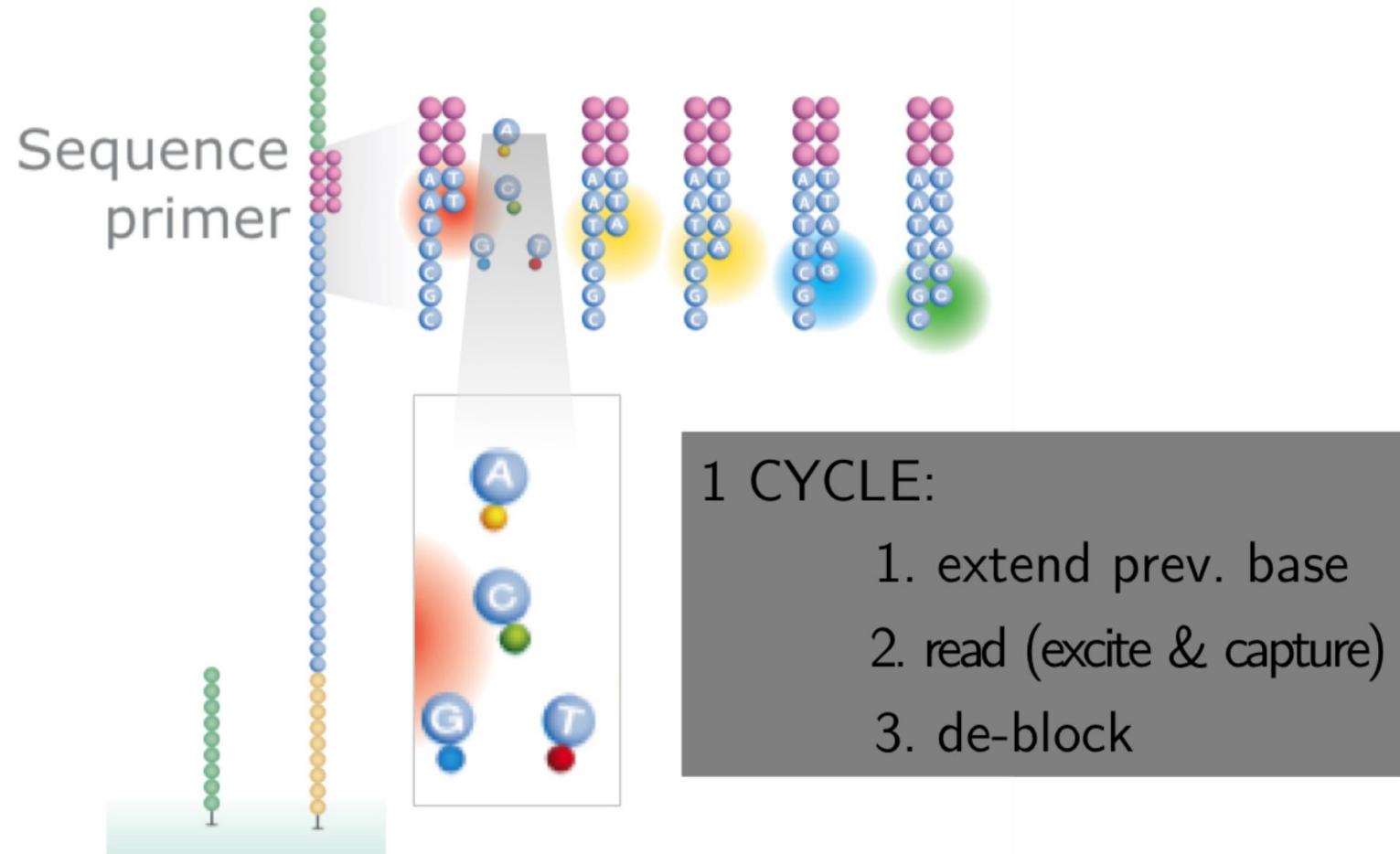
Sequencing By Synthesis (SBS) Technology



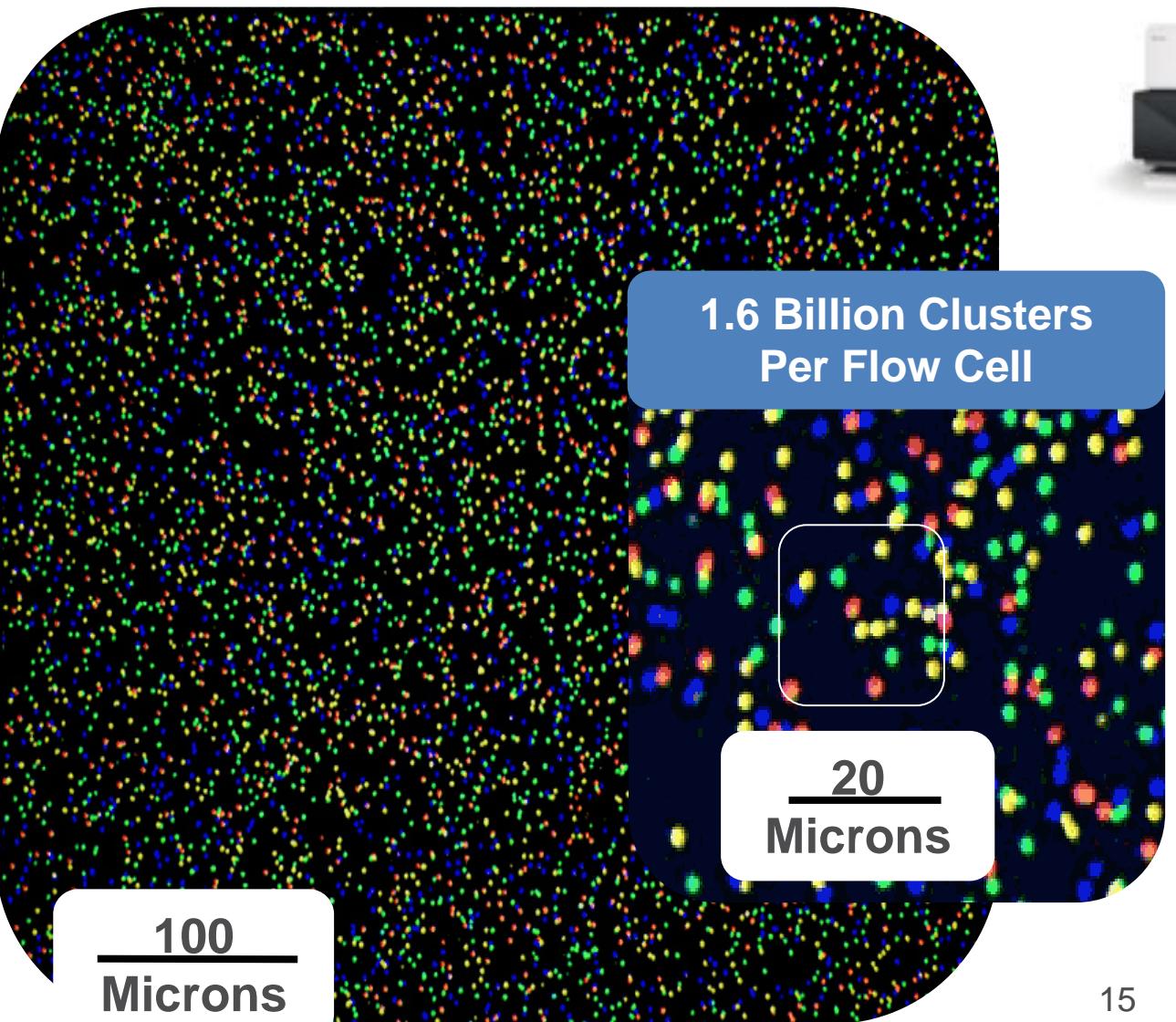
TruSeq Chemistry: Flow Cell



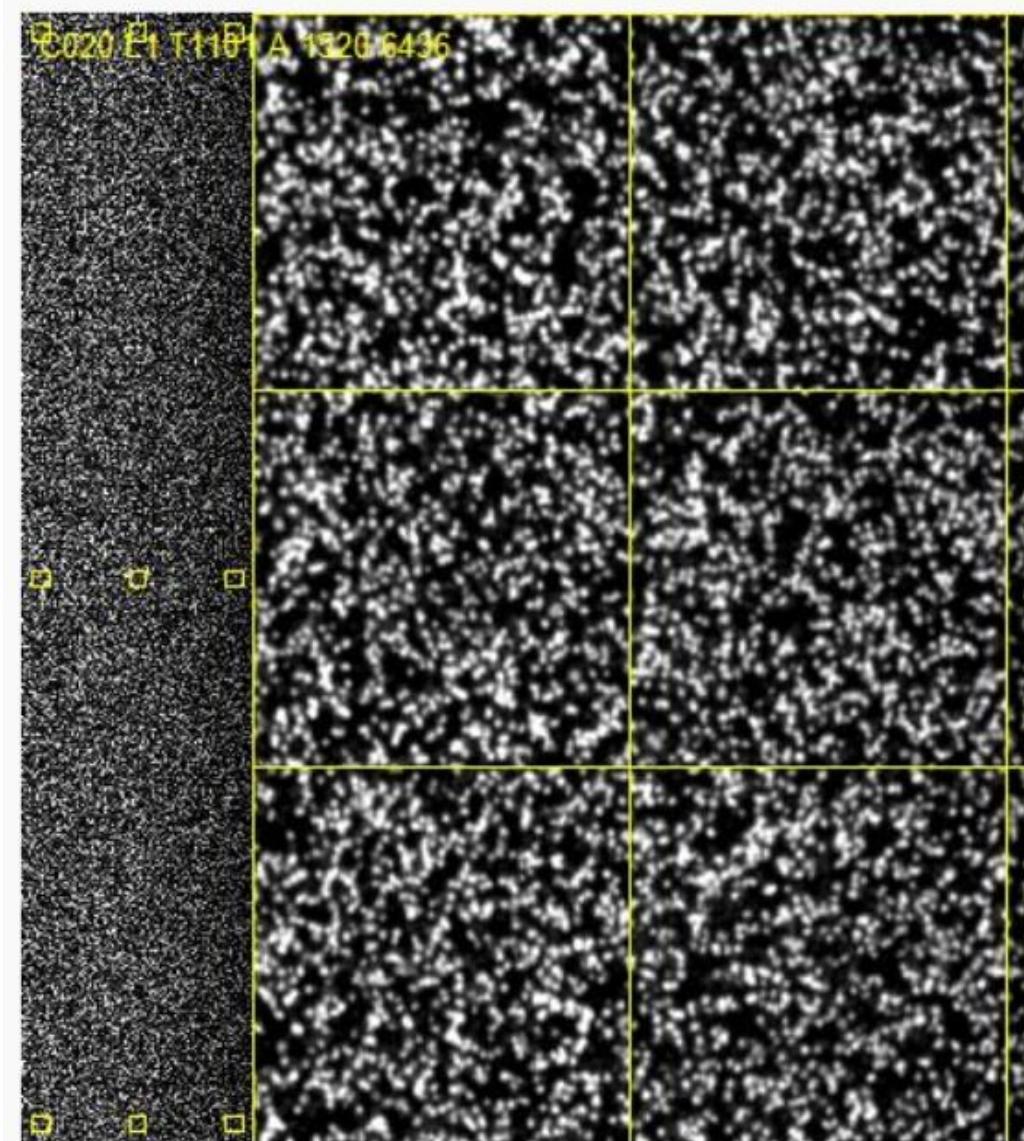
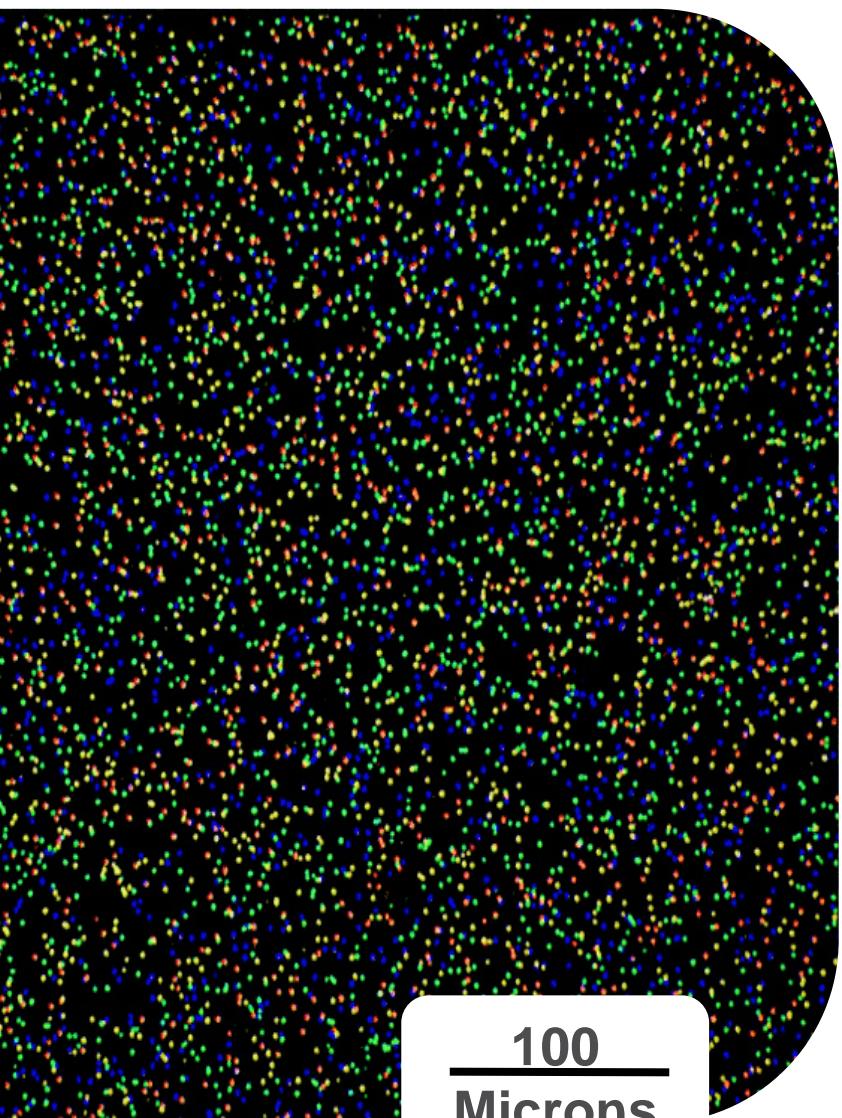
Illumina's sequencing is based on **fluorophore-labelled dNTPs** with **reversible terminator elements** that will become incorporated and excited by a laser one at a time.



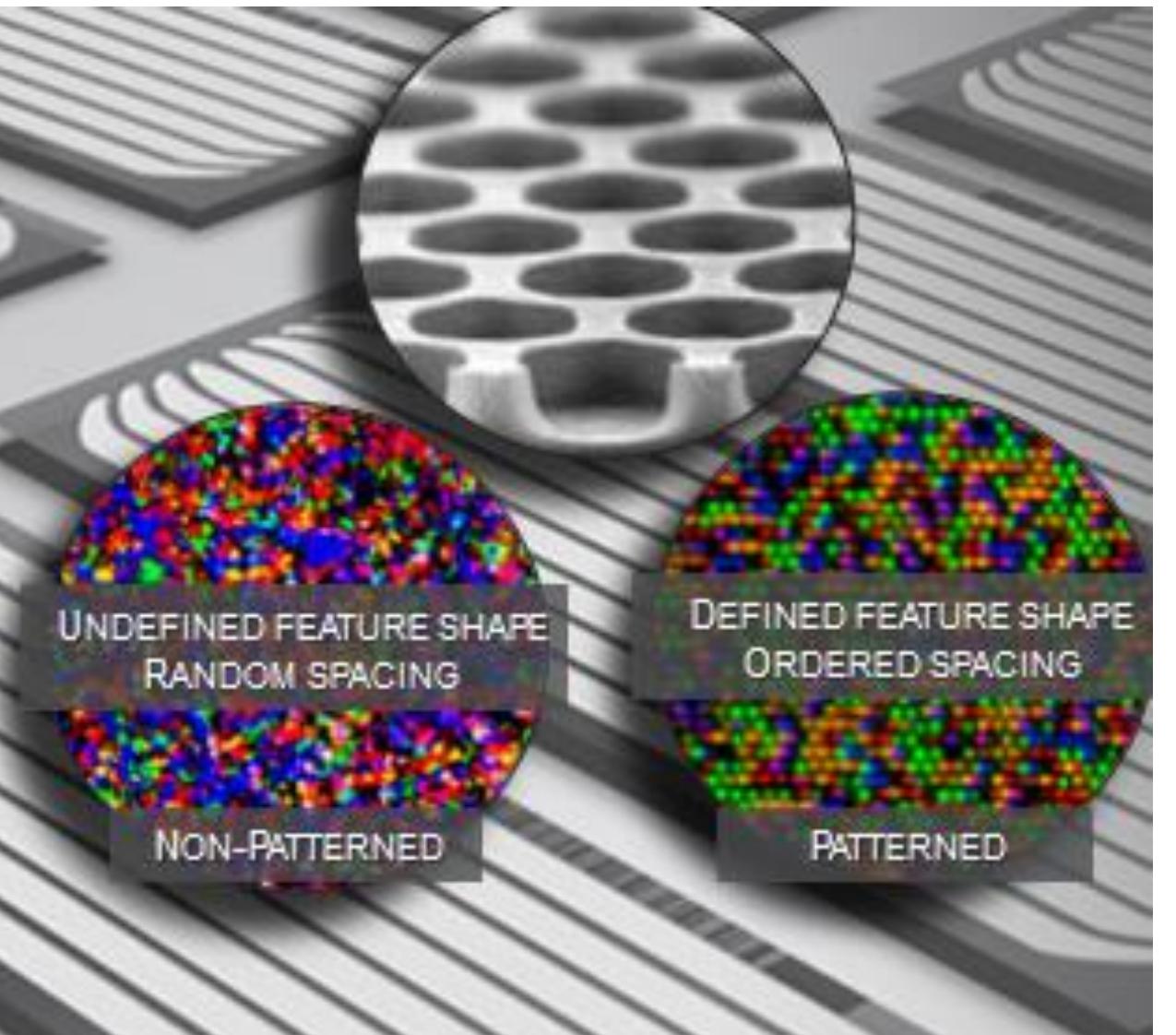
False colored and merged four channel flowcell images



B&W imaging

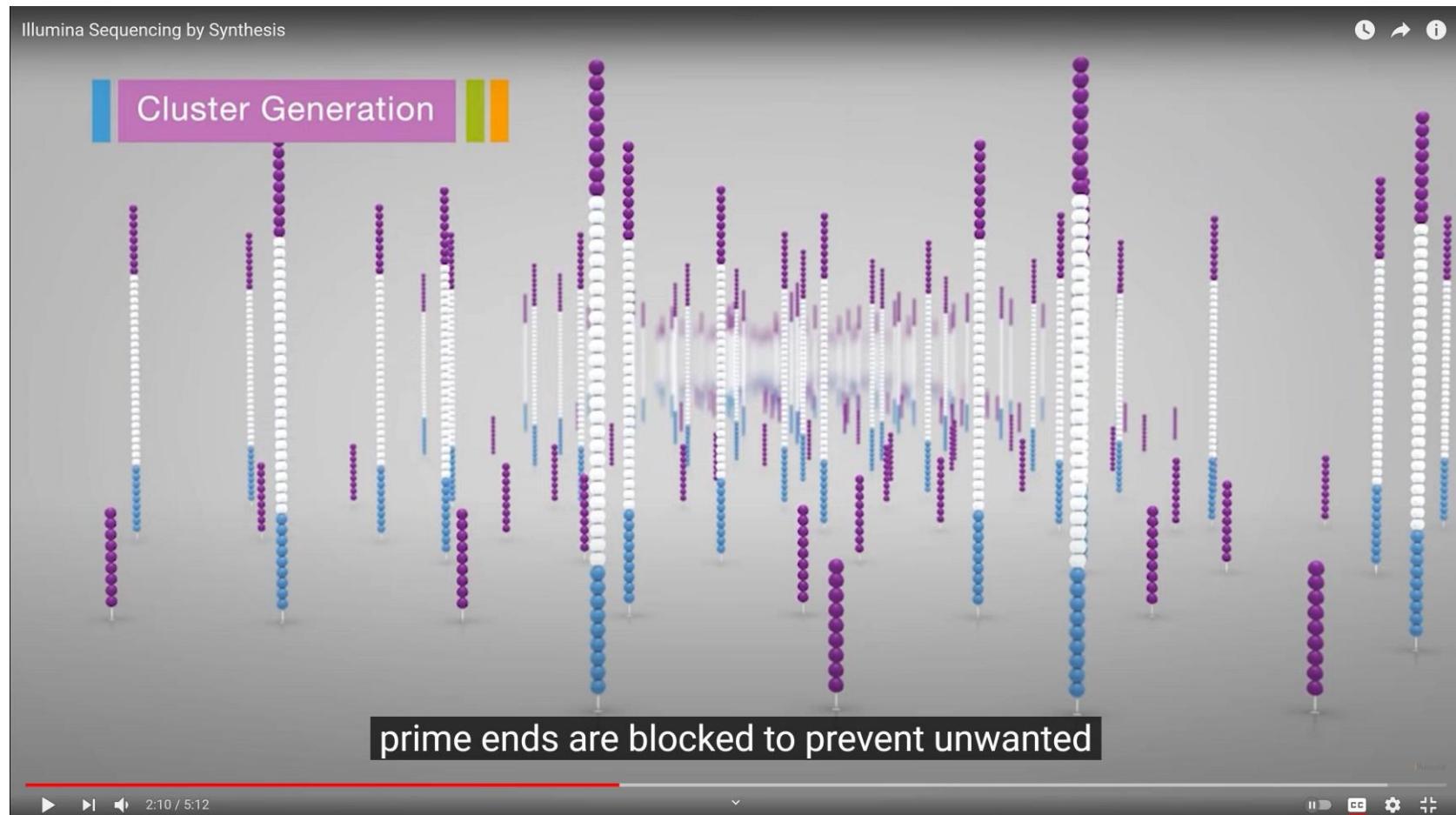


Patterned Flowcell



Highly recommended: Illumina SBS video

<https://www.youtube.com/watch?v=fCd6B5HRaZ8>



What will go wrong ?

- cluster identification
- bubbles
- synthesis errors:

ClusterCluster
ClustsrCluster
ClusterCluster
ClusterCluster
CllsterCluster

What will go wrong ?

- synthesis errors:

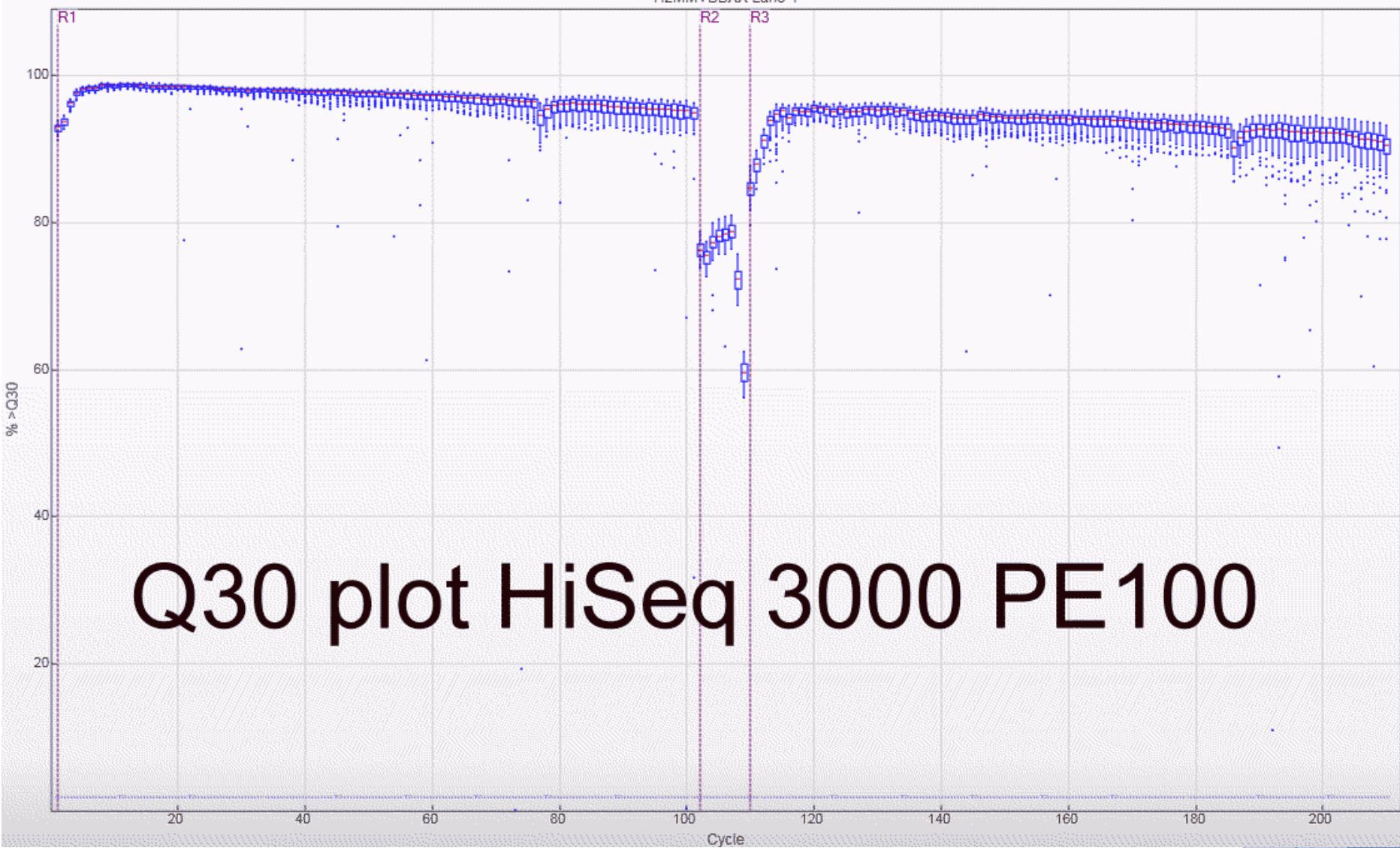
ClusterCluster
ClustsrCluster
ClusterCluster
ClusterCluster
CllsterCluster

CllsterClusterC
ClusterCluster
ClusterCluster
CllusterCluste
ClusterCluster

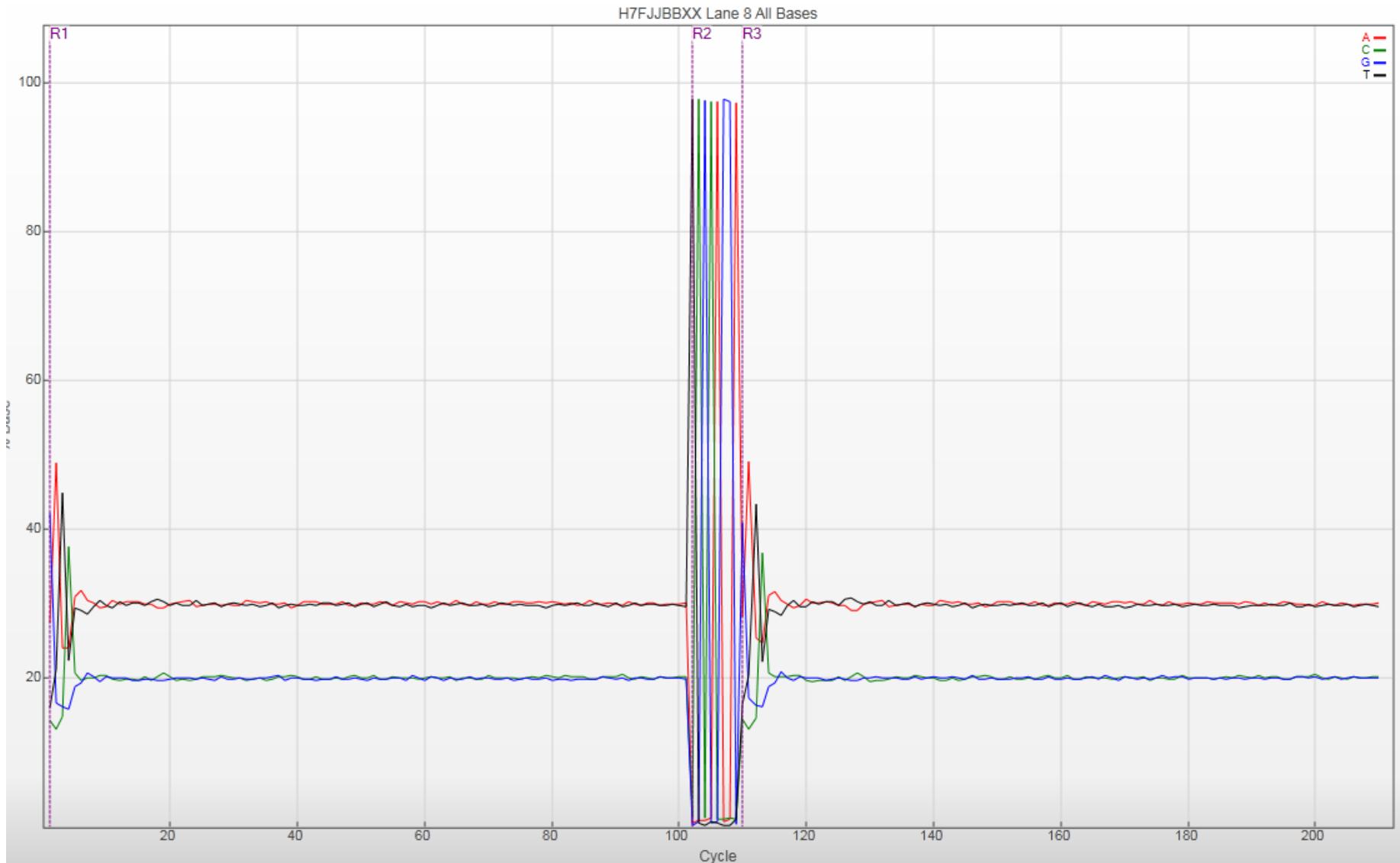
Phasing & Pre-Phasing
problems

The first lines of your data

Illumina SAV viewer

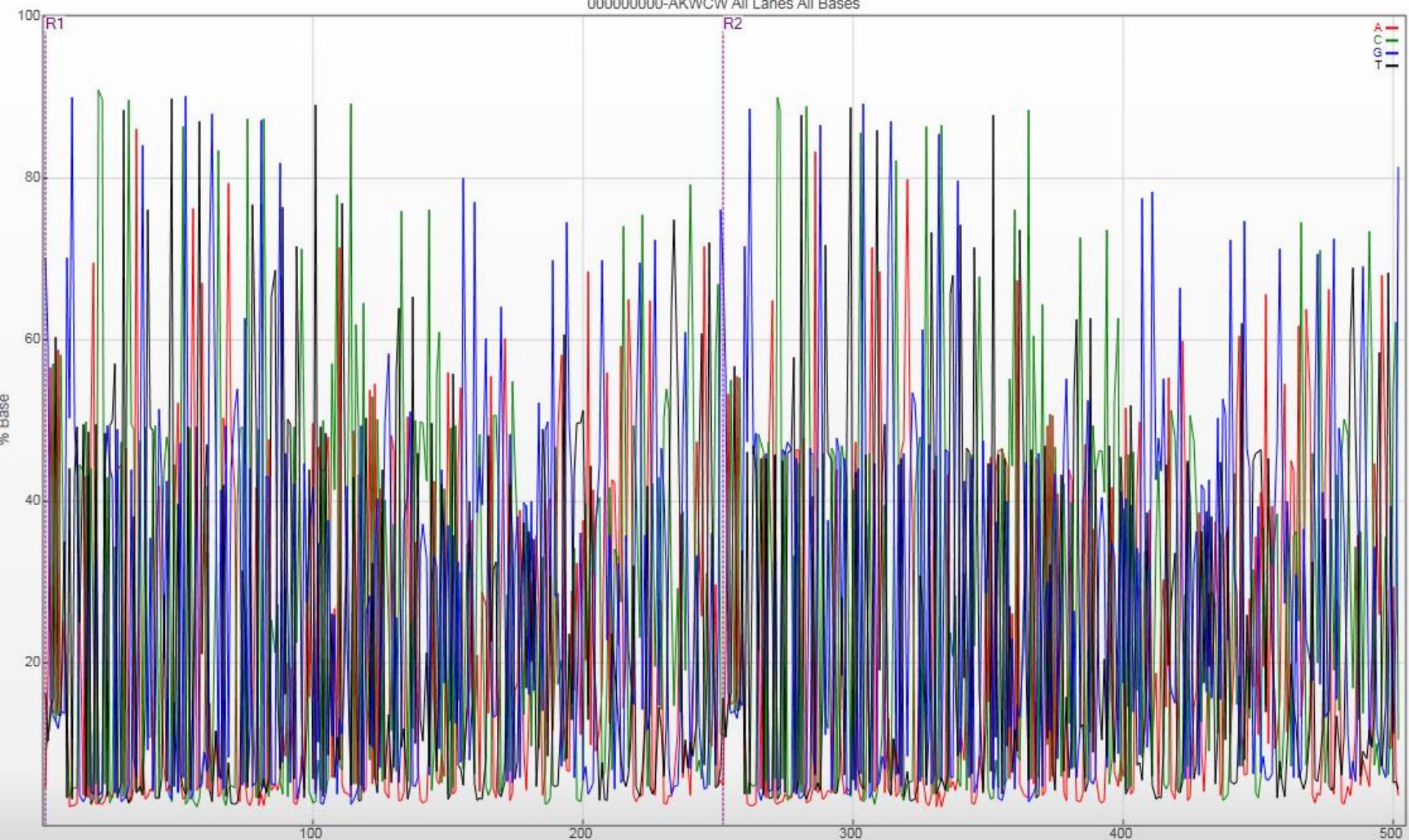


base composition



amplicon

000000000-AKWCW All Lanes All Bases



RNA-seq history

First publication in 2006: 454 sequencing; BLAST to EST data

BMC Genomics 

Methodology article Open Access

Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach

Matthew N Bainbridge¹, René L Warren¹, Martin Hirst¹, Tammy Romanuik¹, Thomas Zeng¹, Anne Go¹, Allen Delaney¹, Malachi Griffith¹, Matthew Hickenbotham², Vincent Magrini², Elaine R Mardis², Marianne D Sadar¹, Asim S Siddiqui¹, Marco A Marra¹ and Steven JM Jones*¹

Address: ¹British Columbia Cancer Agency (BCCA) Genome Sciences Centre, Vancouver, British Columbia, Canada and ²Washington University School of Medicine, Genome Sequencing Center, St. Louis, Missouri 63108, USA

Email: Matthew N Bainbridge - matthewb@bcgsc.ca; René L Warren - rwarren@bcgsc.ca; Martin Hirst - mhirst@bcgsc.ca; Tammy Romanuik - tromanui@bcgsc.ca; Thomas Zeng - tzeng@bcgsc.ca; Anne Go - ago@bcgsc.ca; Allen Delaney - adelaney@bcgsc.ca; Malachi Griffith - malachig@bcgsc.ca; Matthew Hickenbotham - mhickenb@watson.wustl.edu; Vincent Magrini - magarini@watson.wustl.edu; Elaine R Mardis - emardis@watson.wustl.edu; Marianne D Sadar - msadar@bcgsc.ca; Asim S Siddiqui - asims@bcgsc.ca; Marco A Marra - mmarra@bcgsc.ca; Steven JM Jones* - sjones@bcgsc.ca

* Corresponding author

Published: 29 September 2006 Received: 20 April 2006
BMC Genomics 2006, 7:246 doi:10.1186/1471-2164-7-246 Accepted: 29 September 2006

RNA-seq history

First publication in 2006.
DGE papers in 2008

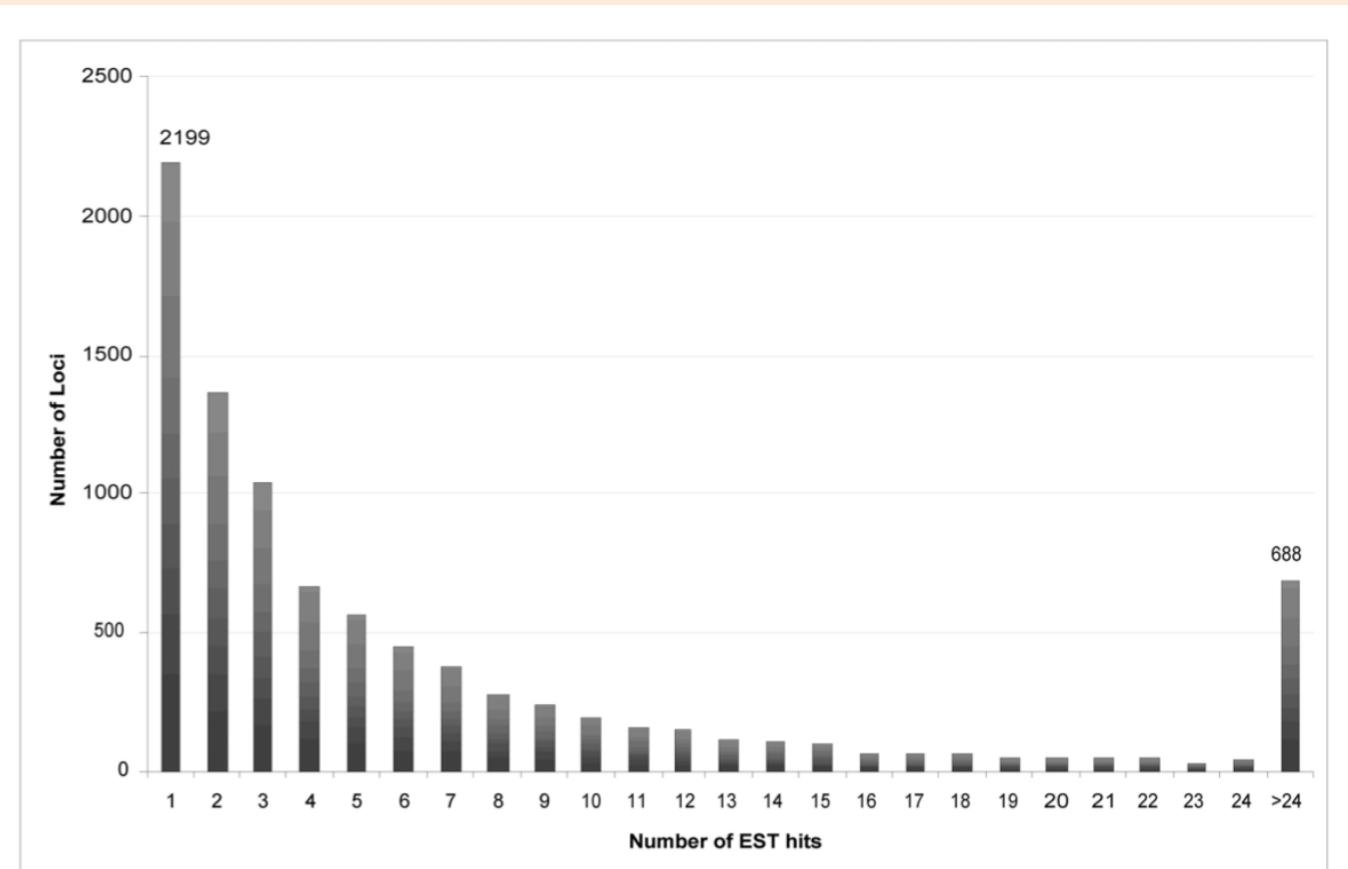


Figure I
A histogram showing the number of gene loci hit by a given number of ESTs.

Why do RNA-seq

- Functional studies:
 - Drug treated vs untreated cell line
 - Wild type vs knock out
 - Life stages, development
- Splice variant analysis
- Transcriptome assembly
- Novel genes
- SNP finding/typing

Challenges

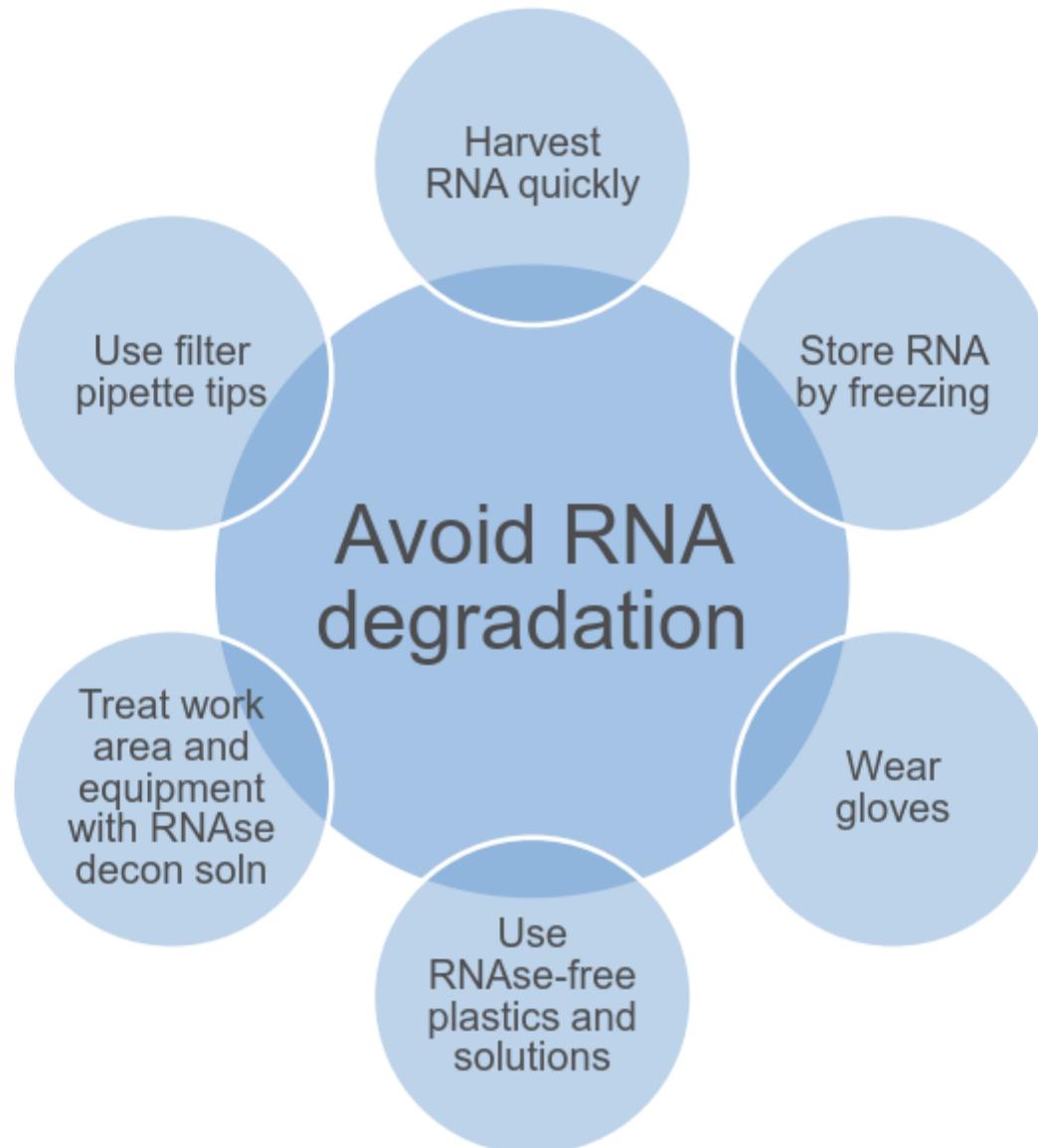
- Sampling, storage, and RNA isolation – Purity? Quantity? Quality?
- RNA is fragile and must be properly handled
- mRNA abundances vary by orders of magnitude
- Organellar RNA & other RNA-species can block overall signal
- RNA population turns over quickly within a cell

RNA is not that fragile



Actually: Avoid DEPC-treated reagents -- remnants can inhibit enzymes

RNA Handling Best Practices



- 18S (2500b) , 28S (4000b)

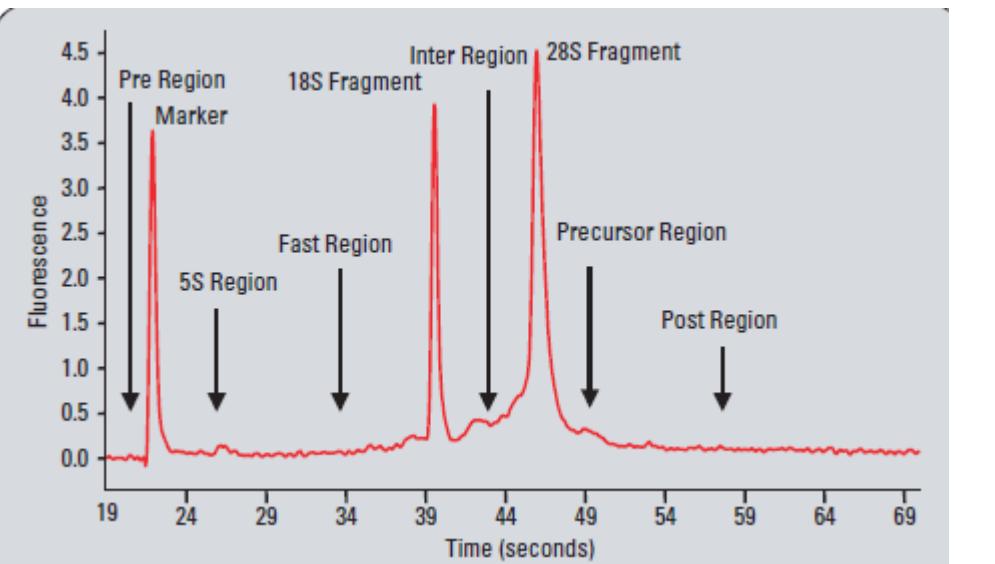
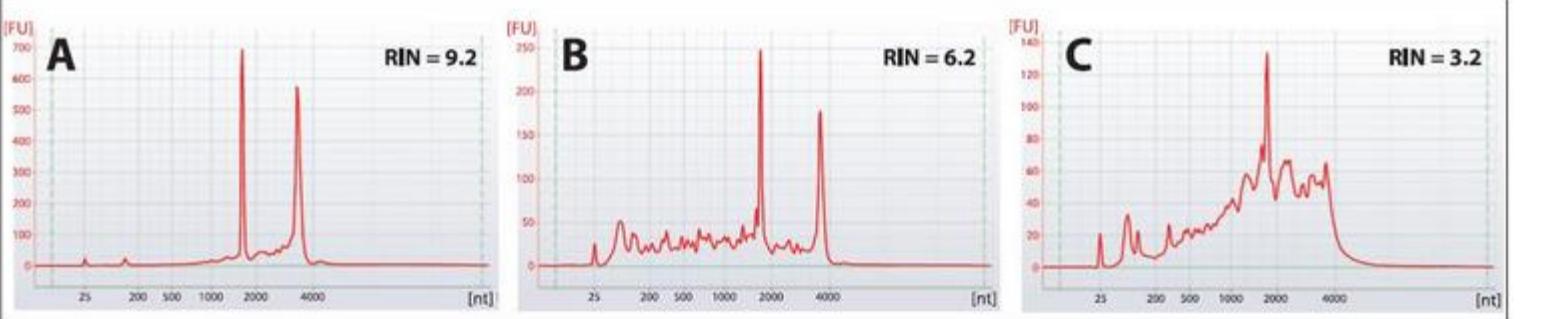
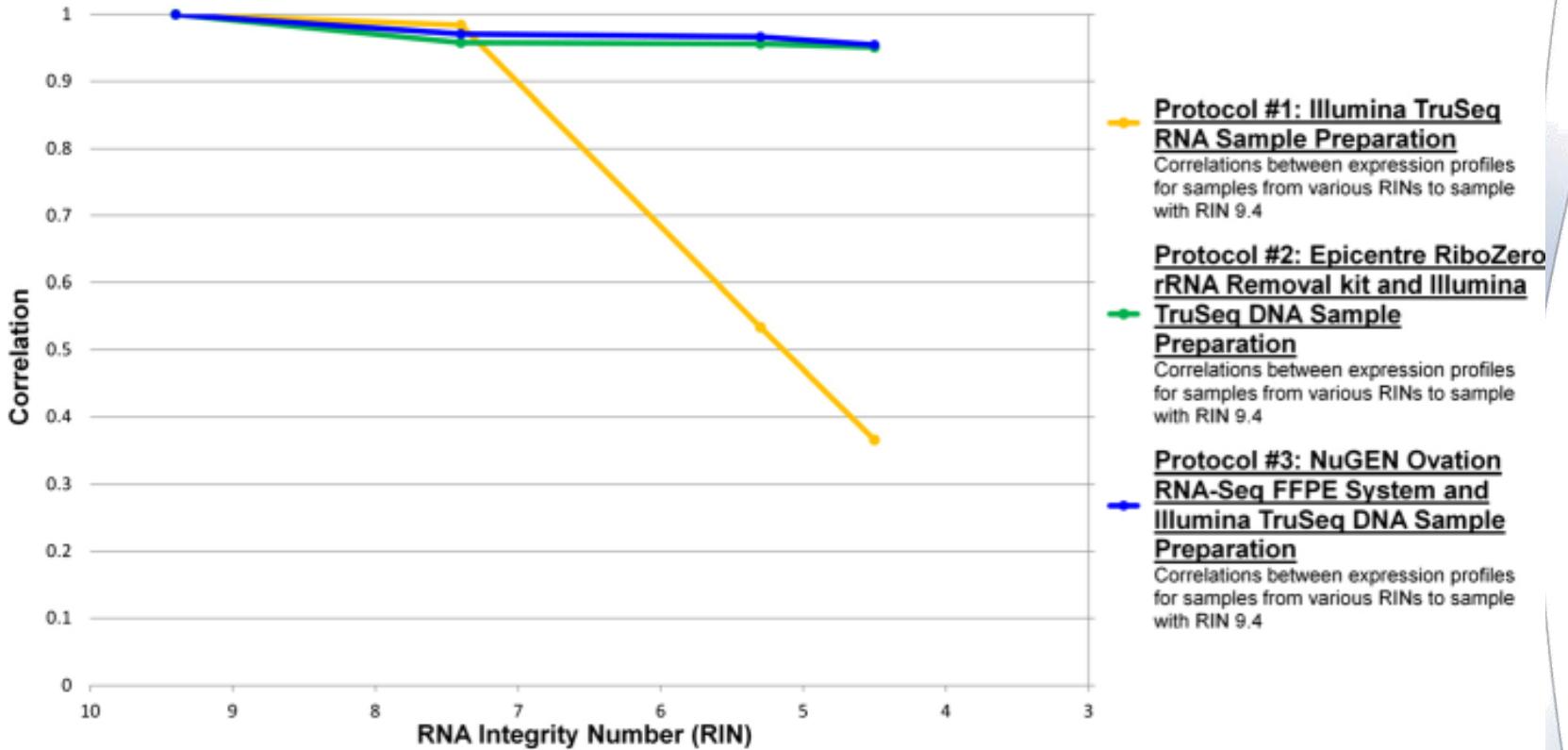


Figure 2.1 Example Agilent Bioanalyzer Electropherograms from three different total RNAs of varying integrity. Panel [A] represents a highly intact total RNA (RIN = 9.2), panel [B] represents a moderately intact total RNA (RIN = 6.2), and panel [C] represents a degraded total RNA sample (RIN = 3.2).



RNA integrity <> reproducibility



Chen et al. 2014

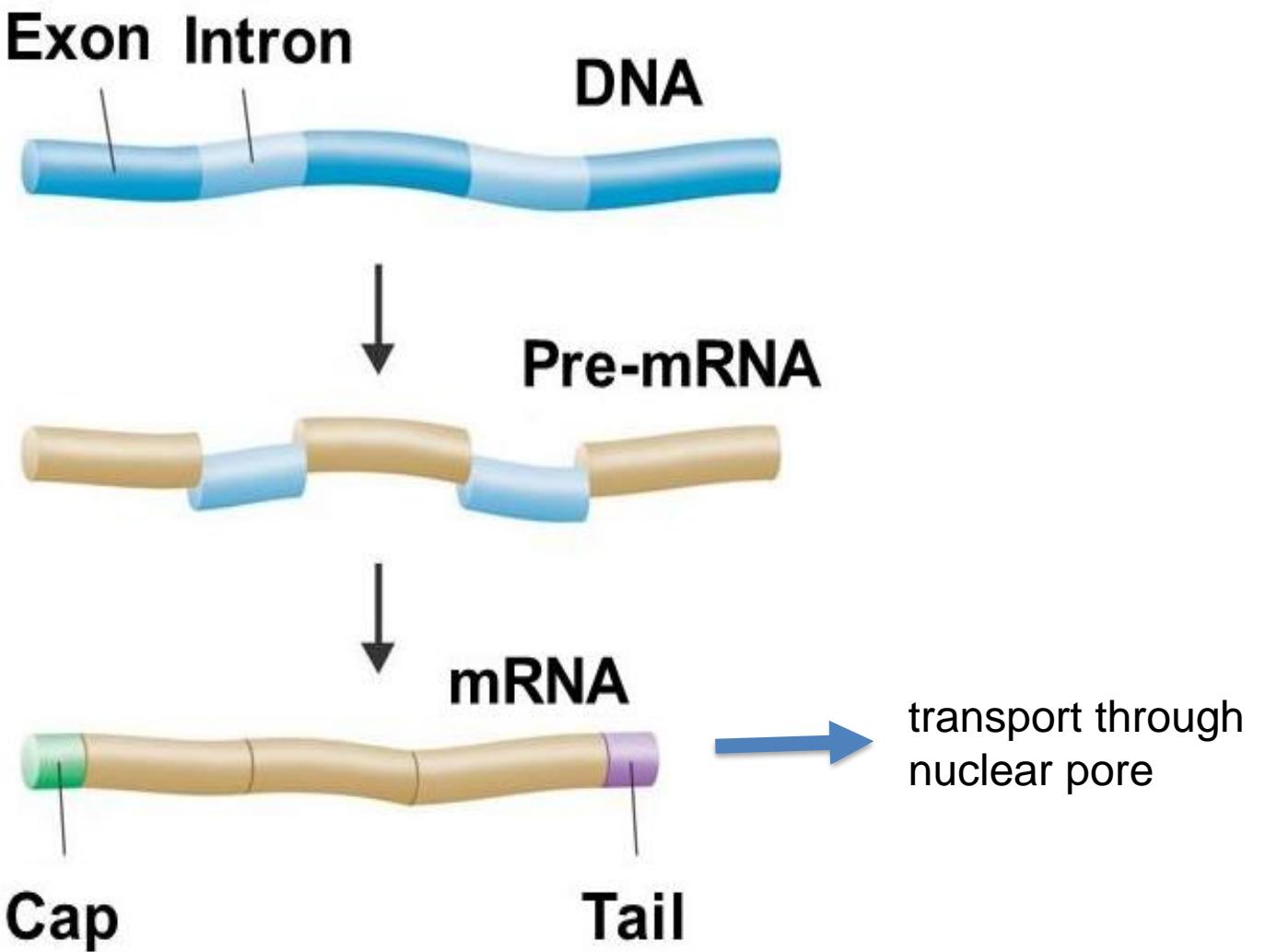
Quantitation & QC methods

- Intercalating dye methods (PicoGreen, Qubit, etc.):
Specific to dsDNA, accurate at low levels of DNA
Great for pooling of indexed libraries to be sequenced in one lane
Requires standard curve generation, many accurate pipetting steps
- Bioanalyzer:
Quantitation is good for rough estimate
Invaluable for library QC
High-sensitivity DNA chip allows quantitation of low DNA levels
- qPCR
Most accurate quantitation method
More labor-intensive
Must be compared to a control

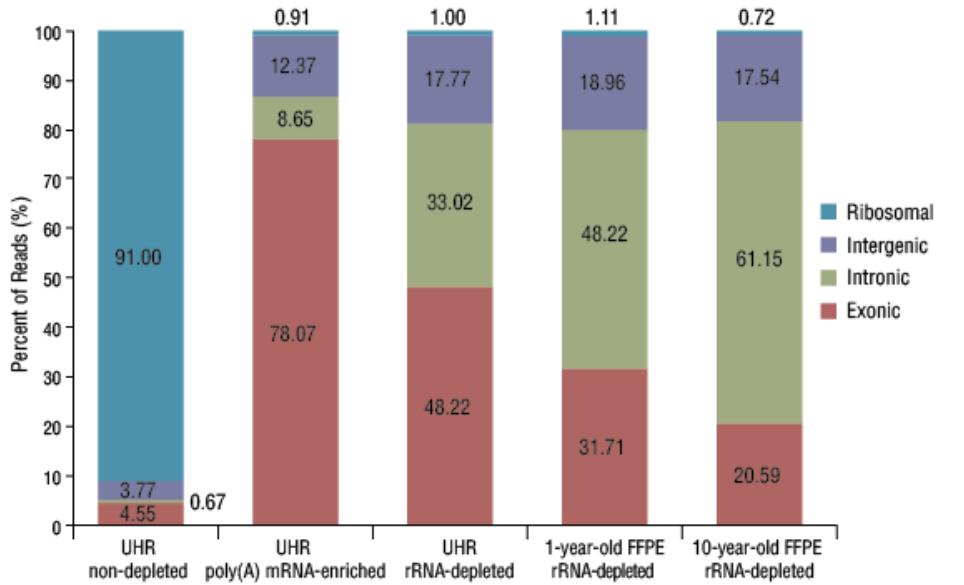
Recommended RNA inputs

Library prep kit	Starting material
mRNA (TruSeq)	100 ng – 1 µg total RNA
Small RNA (TruSeq)	100 ng -1 µg total RNA
Ribo depletion (Epicentre)	500 ng – 2 µg total RNA
SMARTer™ Ultra Low RNA (Takara)	Single-cell – 10 ng
Ovation RNA seq V2, Single Cell RNA seq (NuGen)	10 ng – 100 ng

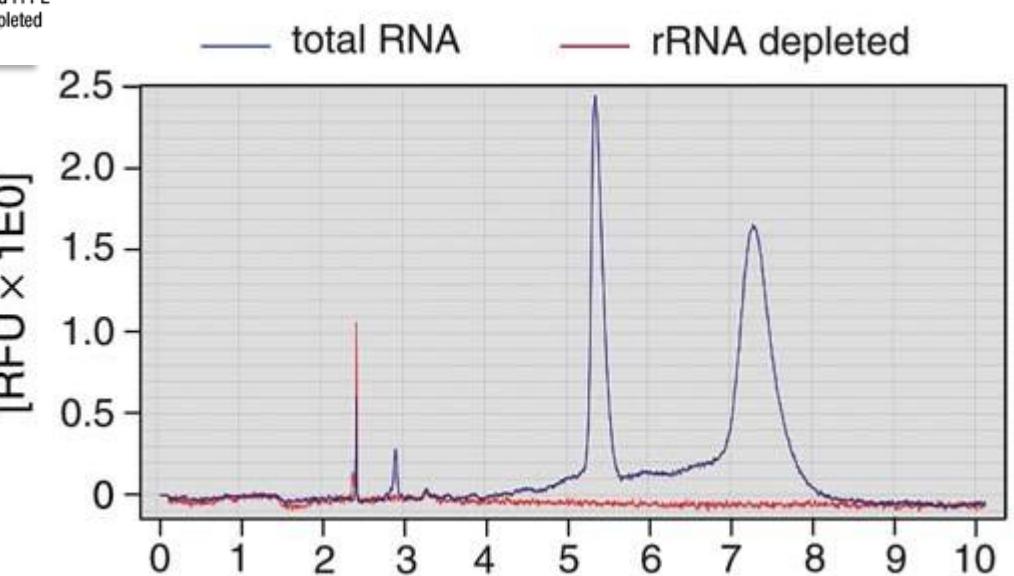
transcription and processing in nucleus



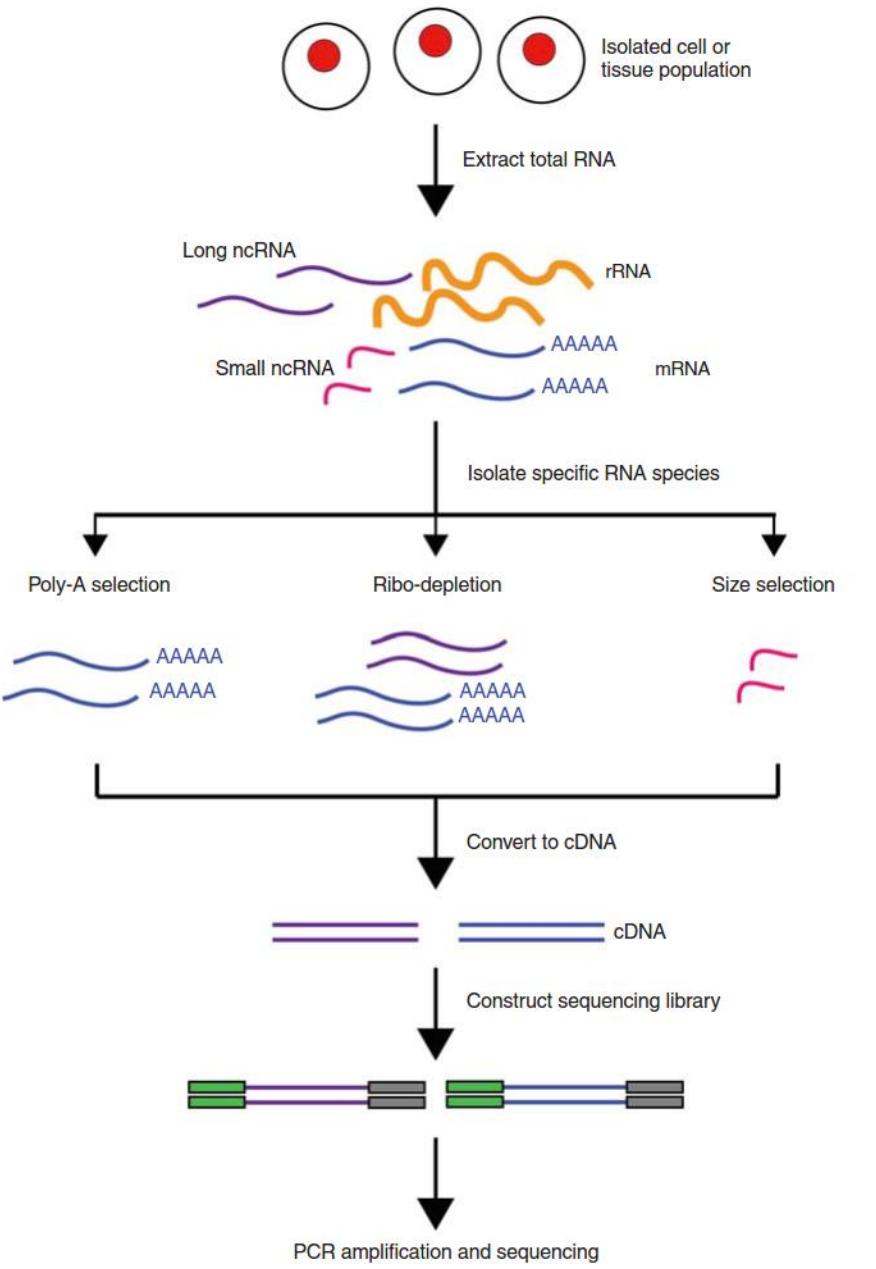
mRNA makes up only about 2% of a total RNA sample



- more than 90% rRNA content
- multiple other non-coding RNA species



Bioanalyzer trace before and after ribo-depletion



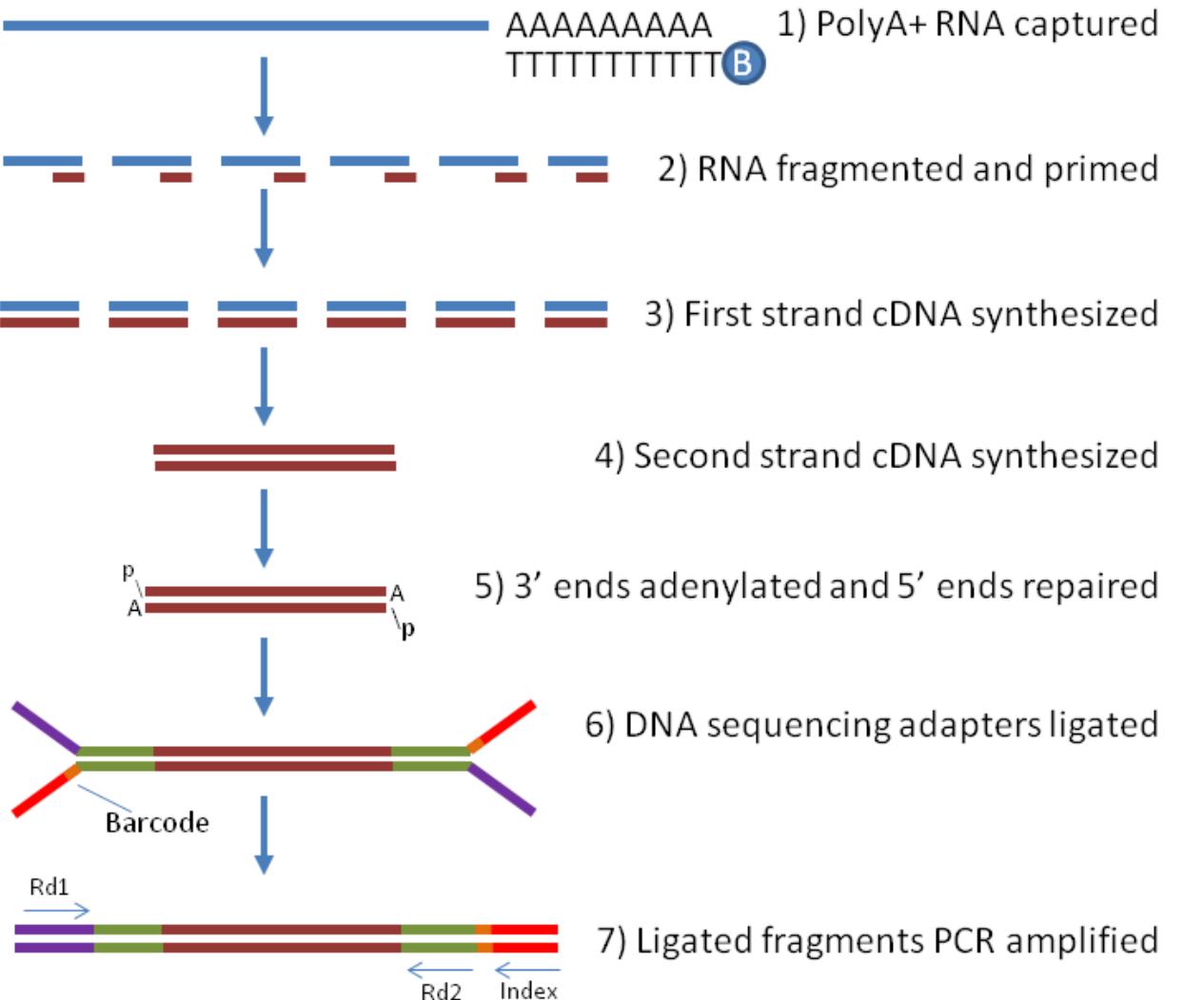
RNA-seq?

Sorry – Illumina and
PacBio are only
sequencing DNA.

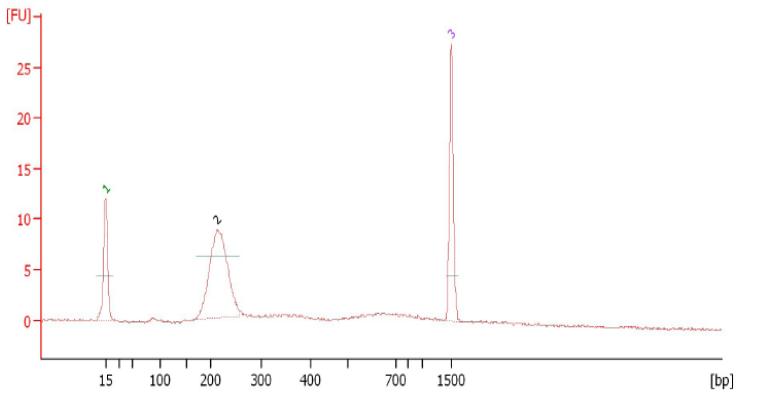
RNA-Seq library prep procedure

1. RNA-sample QC, quantification, and normalization
2. Removal of ribosomal RNA sequences:
via positive or negative selection: Poly-A enrichment or ribo-depletion
3. Fragment RNA:
heating in Mg++ containing buffer – chemical fragmentation has little bias
4. First-strand synthesis:
random hexamer primed reverse transcription
5. RNase-H digestion:
 - creates nicks in RNA strand; the nicks prime 2nd-strand synthesis
 - dUTP incorporated into 2nd strand only
6. A-tailing and adapter ligation exactly as for DNA-Seq libraries
7. PCR amplification of only the first strand to achieve strand-specific libraries - archeal polymerases will not use dUTP containing DNA as template

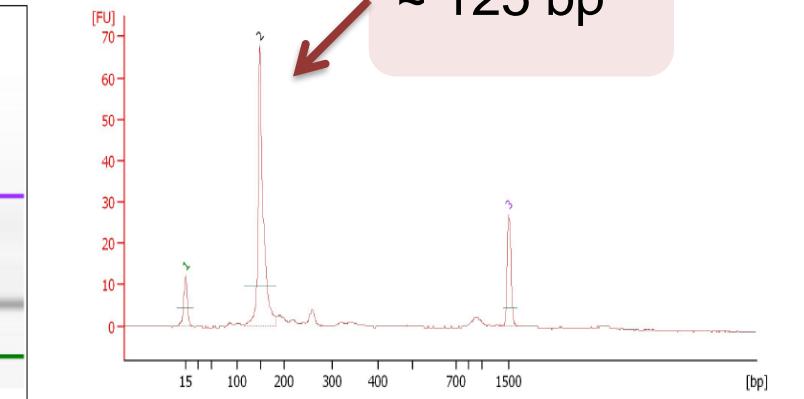
Conventional RNA-Seq library preparation w. Poly-A capture



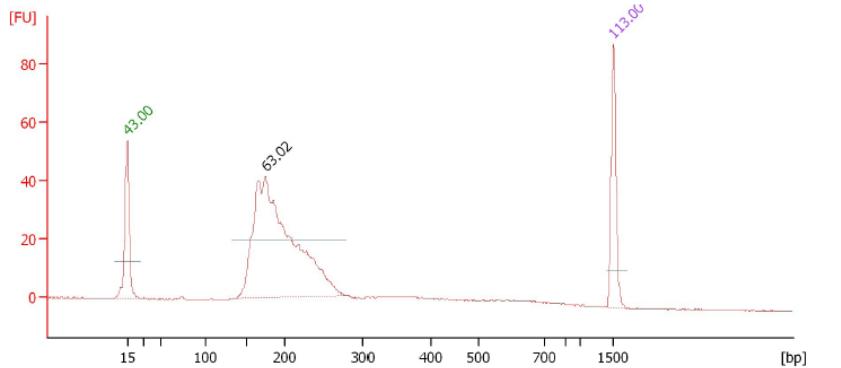
Library QC by Bioanalyzer



Beautiful

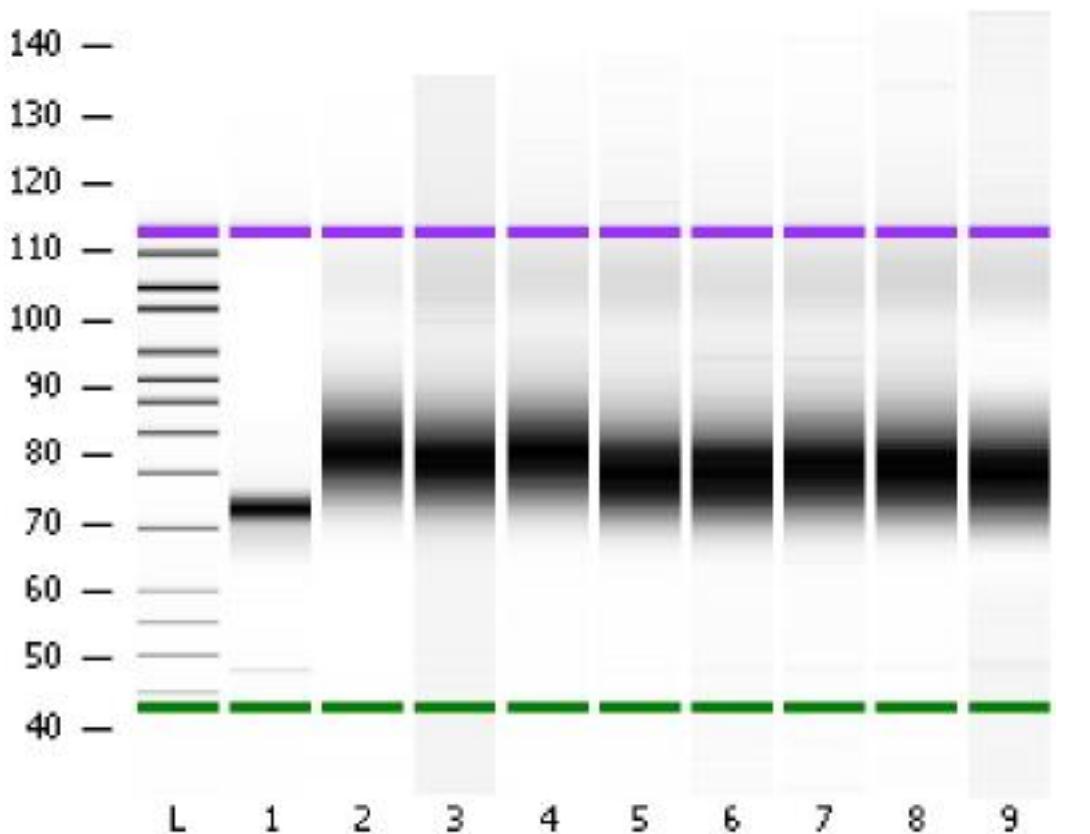


00% Adapters

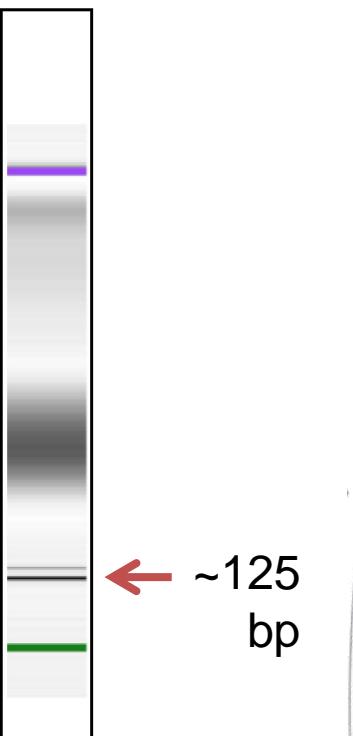


Beautiful

Library QC



Examples for successful libraries



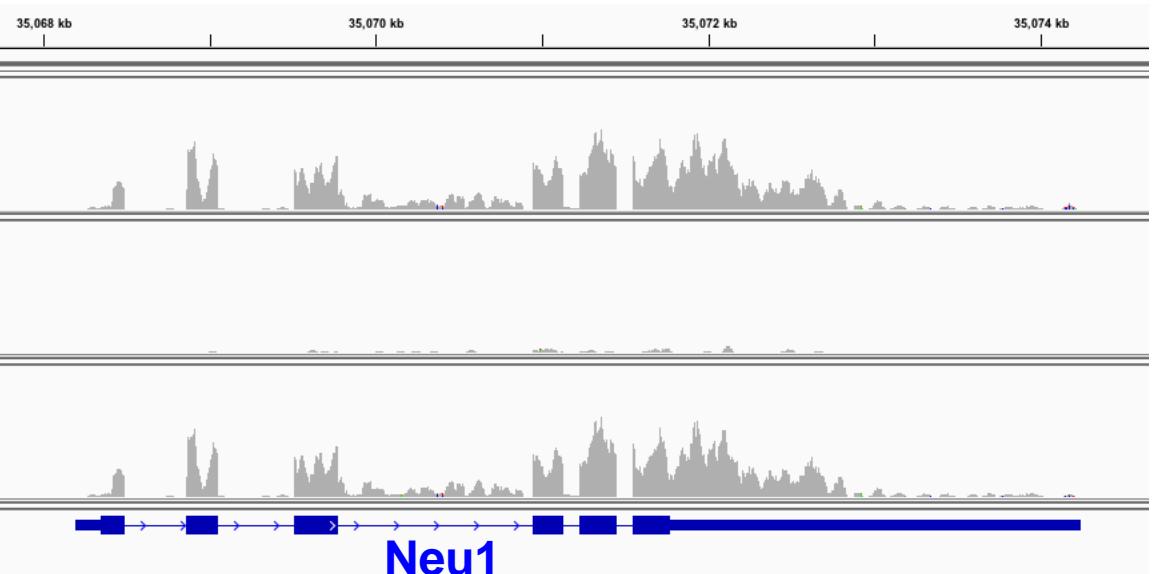
Adapter
contamination
at ~125 bp

Considerations in choosing an RNA-Seq method

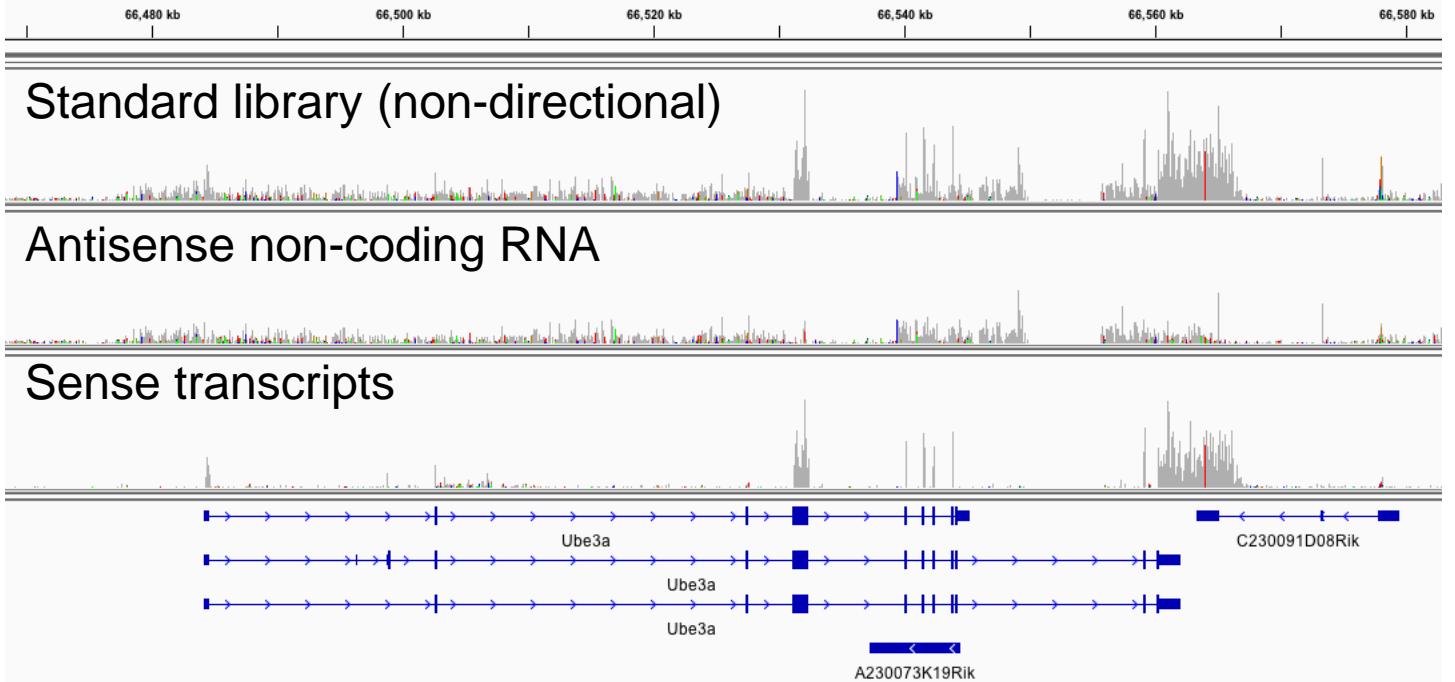
- Transcript type:
 - mRNA, extent of degradation
 - small/micro RNA
- Strandedness:
 - un-directional ds cDNA library
 - directional library
- Input RNA amount:
 - 0.1-4ug original total RNA
 - linear amplification from 0.5-10ng RNA
- Complexity:
 - original abundance
 - cDNA normalization for uniformity
- Boundary of transcripts:
 - identify 5' and/or 3' ends
 - poly-adenylation sites
 - Degradation, cleavage sites

strand-specific information

Standard library
(non-directional)



Strand-specific RNA-seq

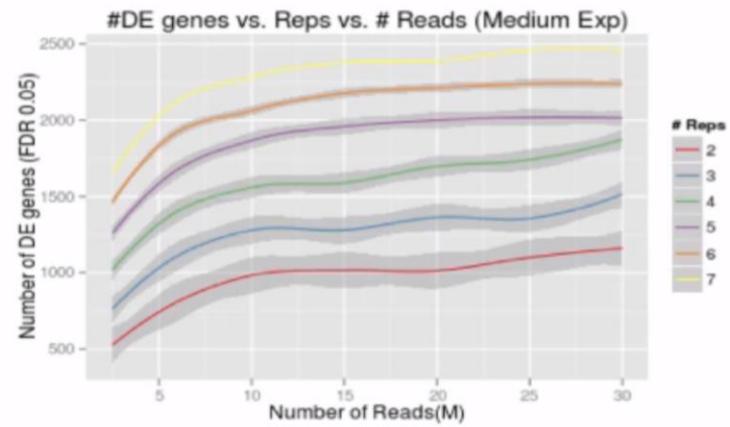
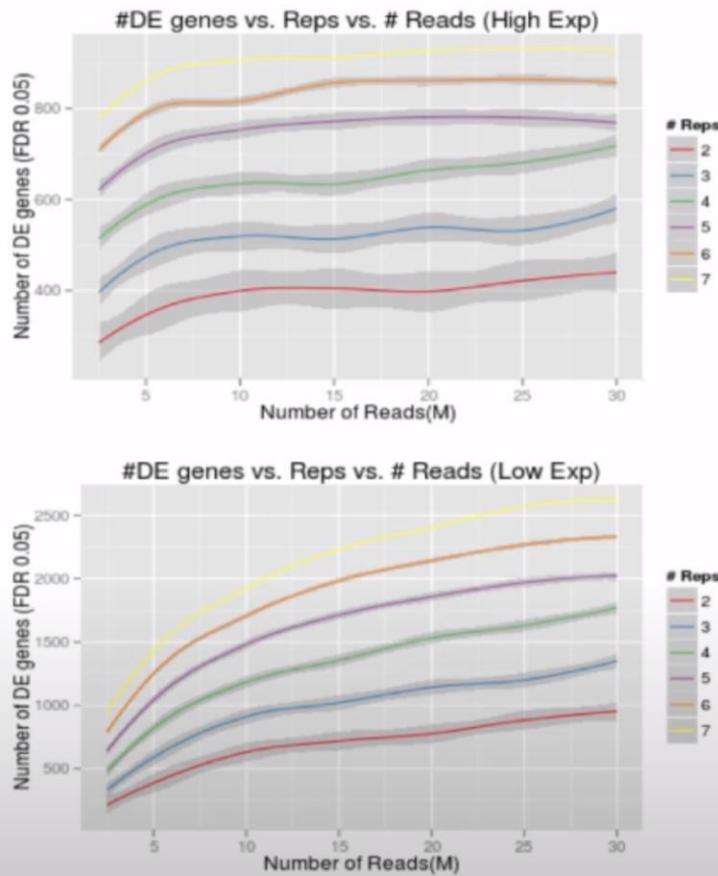


- Informative for non-coding RNAs and antisense transcripts
- Essential when NOT using polyA selection (mRNA)
- No disadvantage to preserving strand specificity

RNA-seq for DGE

- Differential Gene Expression (DGE)
 - 50 bp single end reads
 - 25 million reads per sample (eukaryotes)
 - 10 mill. reads > 80% of annotated genes
 - 25 mill. . reads > 90% of annotated genes
 - 5 million reads per sample (bacteria)

Experimental Design



For high expressers: Increasing sequencing depth has little effect on increasing number of DE genes detected, while biological replicates are clearly more beneficial.

For low expressers: Both sequencing depth and biological replicates increases power to detect DE genes.

Liu et al. (2014) RNA-Seq differential expression studies: more sequence or more replication?, Bioinformatics, 30(3):1-4

RNA-seq reproducibility

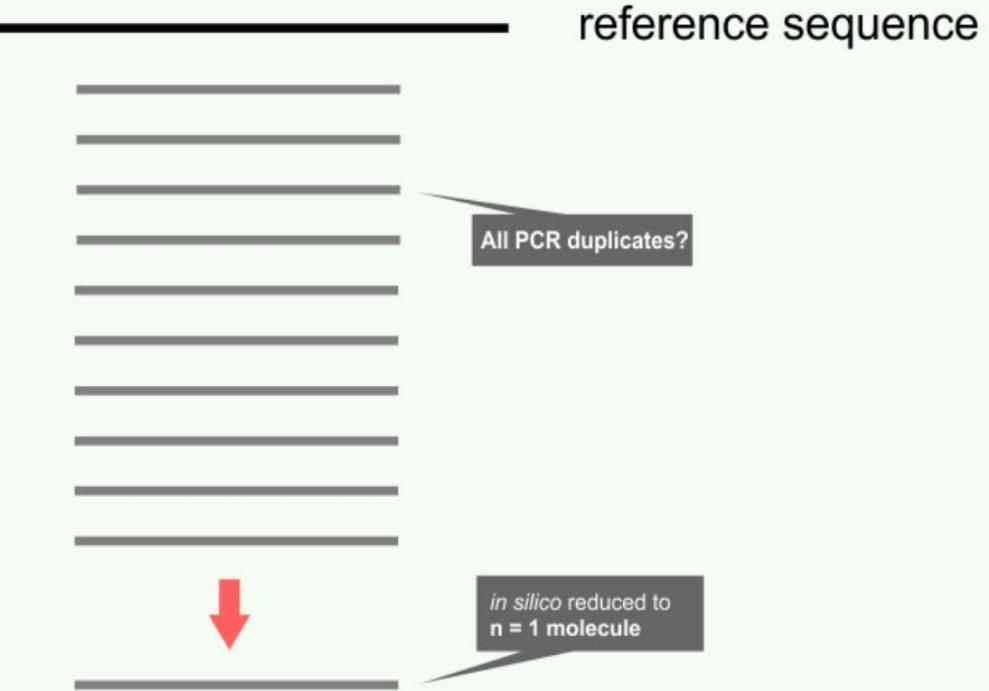
- Two big studies multi-center studies (2014)
- High reproducibility of data given:
 - same library prep kits, same protocols
 - same RNA-samples
 - RNA isolation protocols have to be identical
 - robotic library preps?

UMIs – Unique Molecular Identifiers

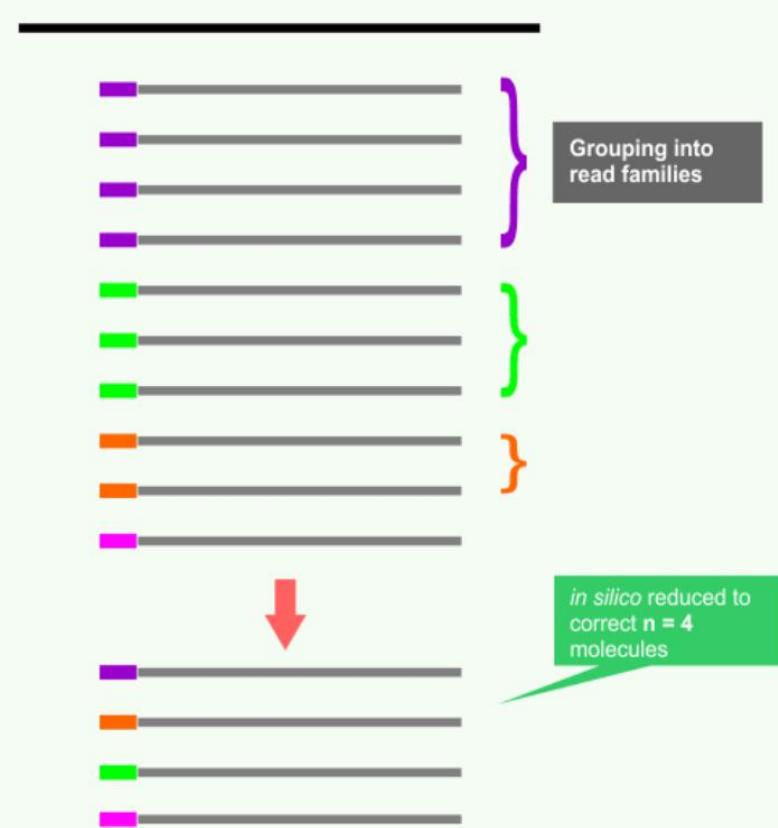
Molecular indexing for precision counts

UMI application in **quantitative studies** (e.g. RNA-seq, scRNA-seq, miRNA-Seq, ChIP-seq).

PCR duplicate removal without UMIs

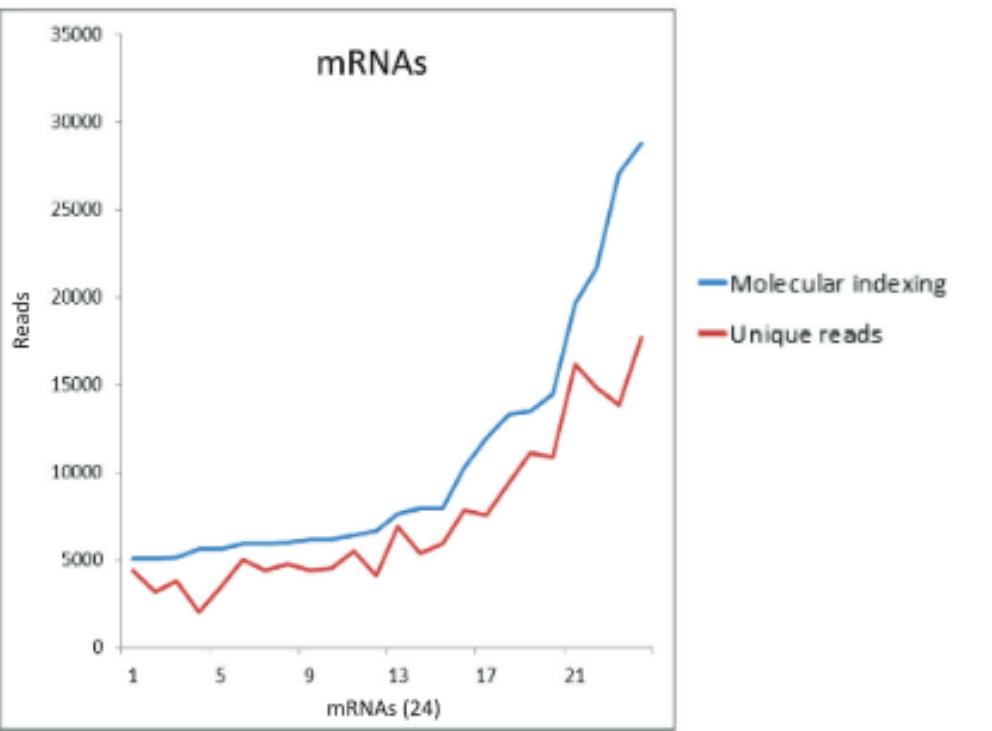


PCR duplicate removal with UMIs



Molecular indexing – for precision counts

B

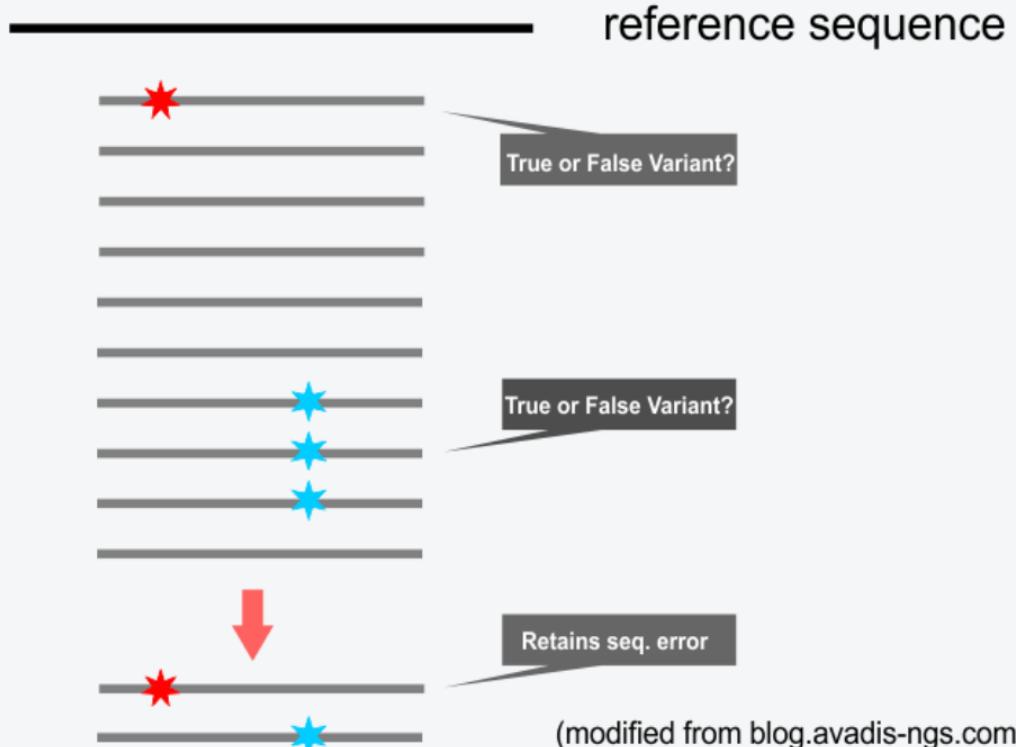


UMIs – Unique Molecular Identifiers

Molecular indexing for low abundance variants

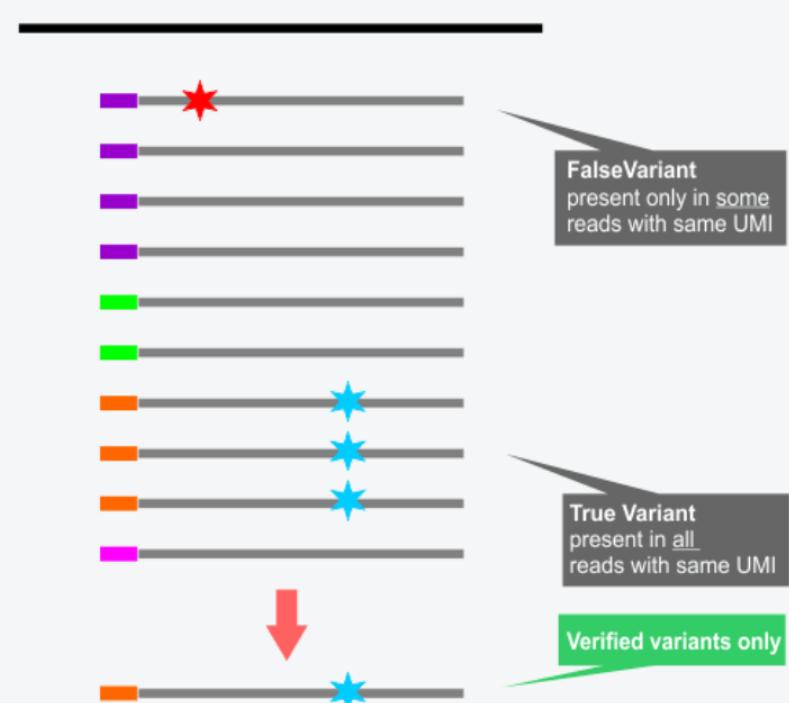
UMI application in deep sequencing **genomic variation** studies (e.g. WGS, exome capture, cfDNA)

Variant calling without UMIs



(modified from blog.avadis/ngs.com)

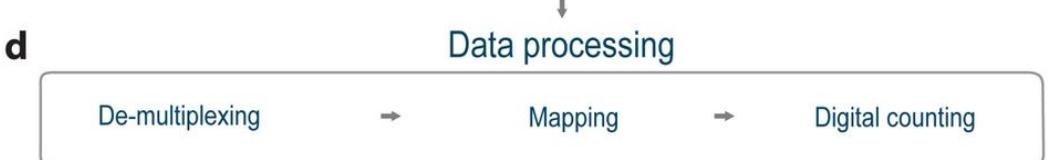
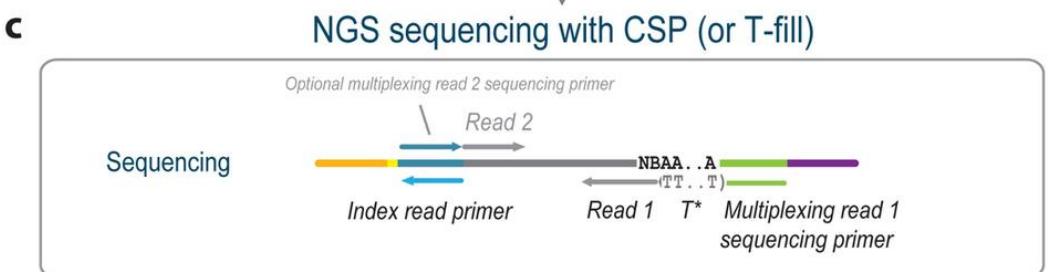
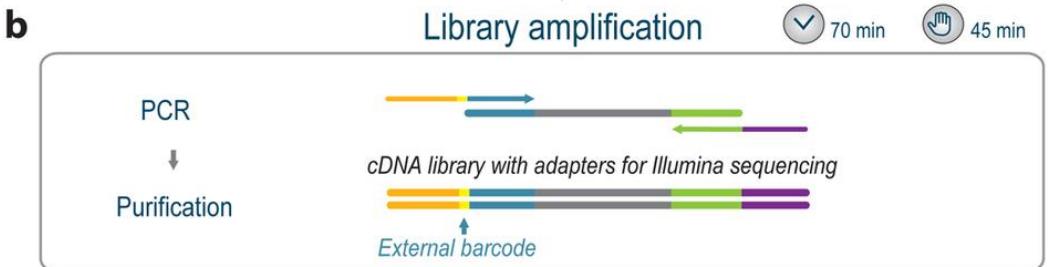
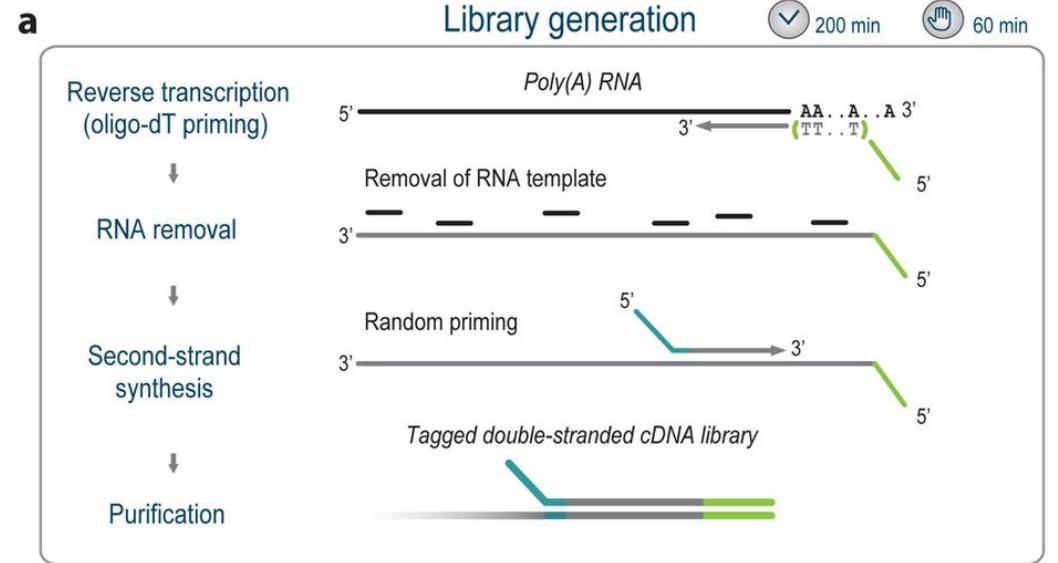
Variant calling with UMIs



3'-Tag-Seq

- In contrast to full length RNA-seq
- Sequencing 1/10 for the average transcript
- Less dependent on RNA integrity
- Microarray-like data
- Options:
 - **BRAD-Seq : 3' Digital Gene Expression**
 - **Lexogen Quant-Seq**

Lexogen Quant-Seq



- we include UMIs

Other RNA-seq objectives

- Transcriptome assembly:
 - 300 bp paired end **plus**
 - 100 bp paired end
- Long non coding RNA studies:
 - 100 bp paired end
 - 60-100 million reads
- Splice variant studies:
 - 100 bp paired end
 - 60-100 million reads

RNA-seq targeted sequencing:

- Capture-seq (Mercer et al. 2014)
- Nimblegen and Illumina
- Low quality DNA (FFPE)
- Lower read numbers 10 million reads
- Targeting lowly expressed genes.

Typical RNA-seq drawbacks

- Very much averaged data:
Data from mixed cell types & mixed cell cycle stages
- Hundreds of differentially expressed genes
(which changes started the cascade?)

higher resolution desired

→ beyond steady-state RNA-seq

mechanisms influencing the mRNA steady-state

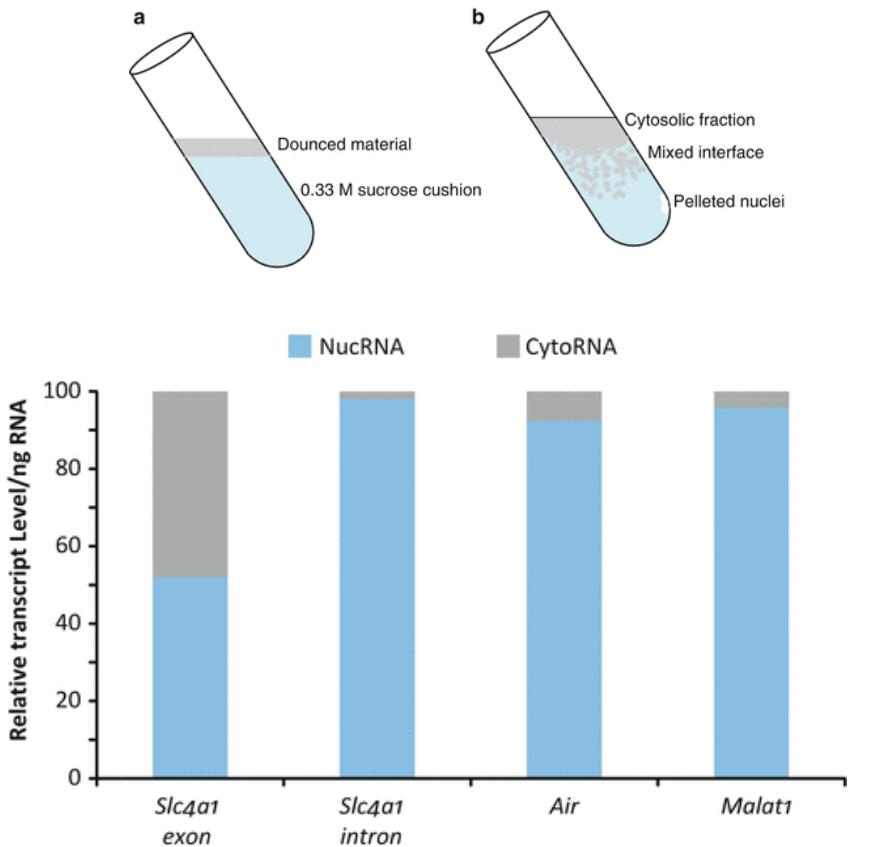
- Transcription rates
- Transport rates
- miRNAs and siRNAs influence both translation and degradation
- RNA modifications (e.g. methylated RNA bases, m⁶A, m⁵C, pseudouridine, ...)
- RNA degradation pathways
- (differential translation into proteins)

beyond steady-state RNA-seq

- GRO-Seq; PRO-Seq; nuclear RNA-Seq:
what is currently transcribed
- Ribosomal Profiling:
what is currently translated
- Degradome Sequencing:
what is ... ?

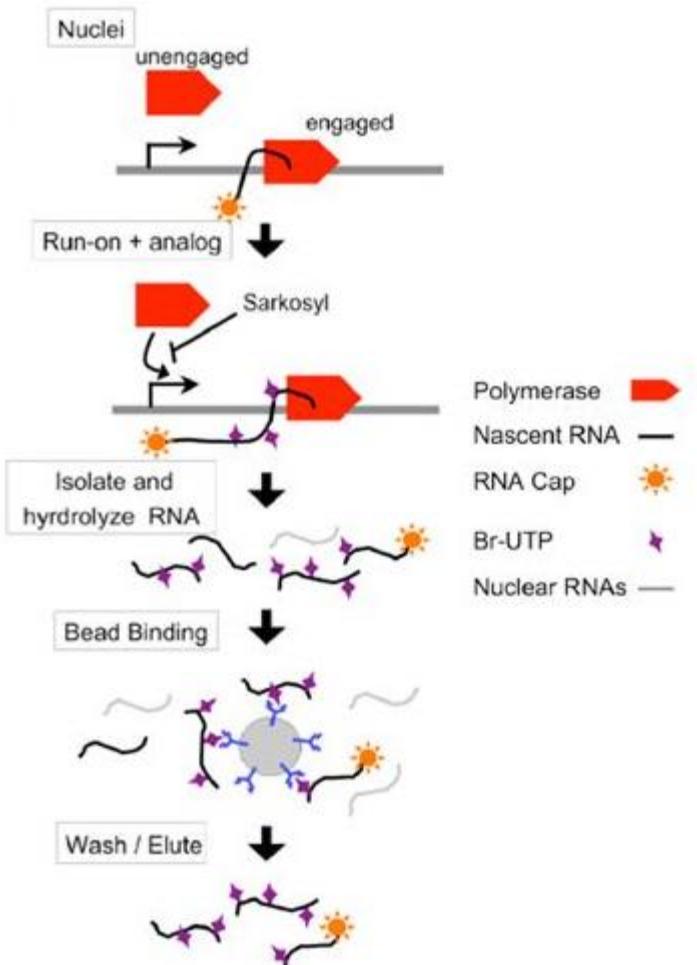
nucRNA-seq

- Fractioning of nuclei and cytosol
- Studying active transcription



Dhaliwal et al. 2016

GRO-Seq



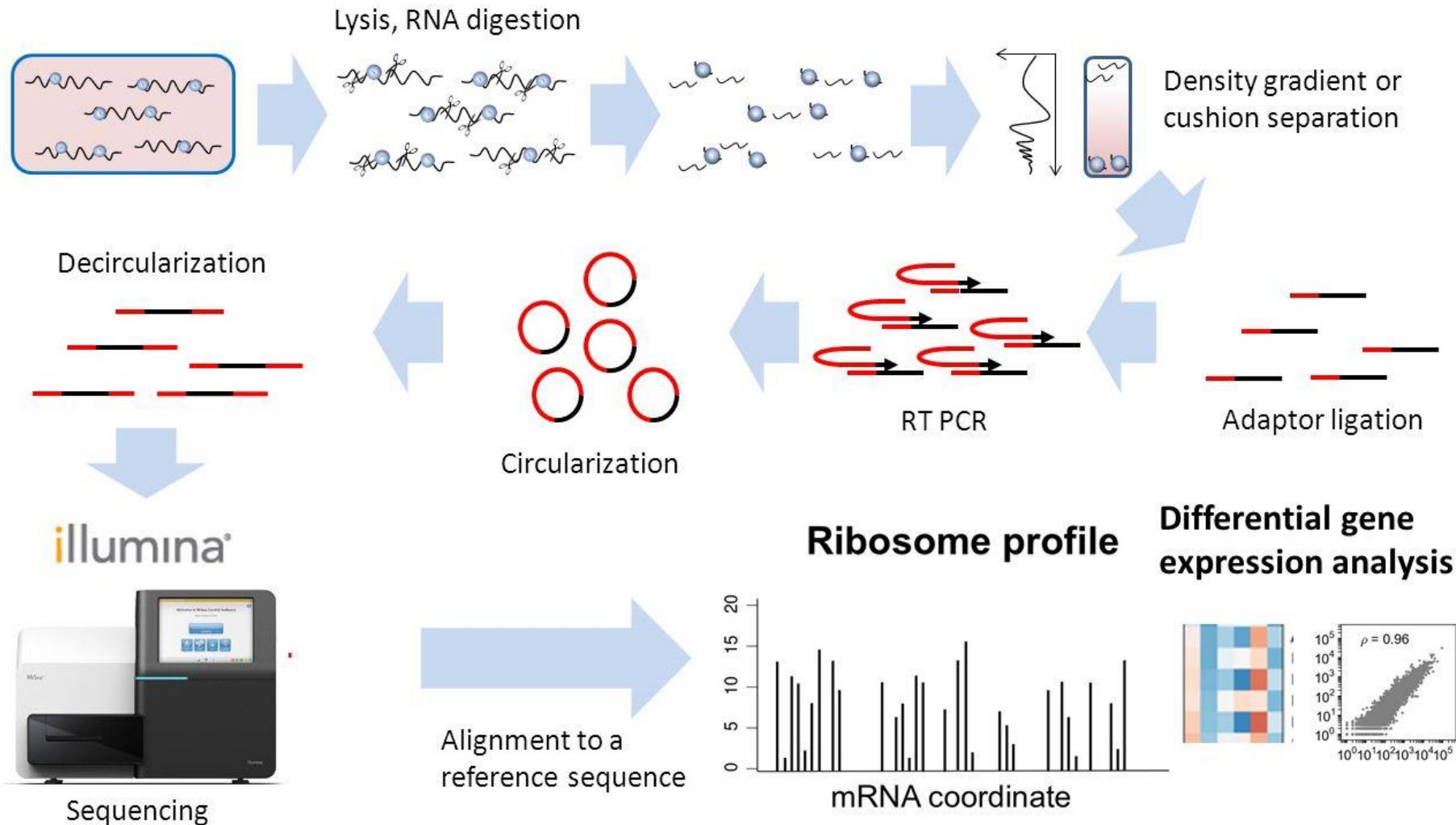
- Global Run-On – sequencing
- pulse-chase experiments (Br-UTP)
- uses isolated nuclei
- sarcosyl prevents binding of polymerase (only transcription in progress will be seq.)
- measures active transcription rather than steady state
- Maps position and orientation
- Earliest changes identify primary targets
- Detection of novel transcripts including non-coding and enhancer RNAs

Core et al, *Science*, 2008

2008: GRO - without the seq

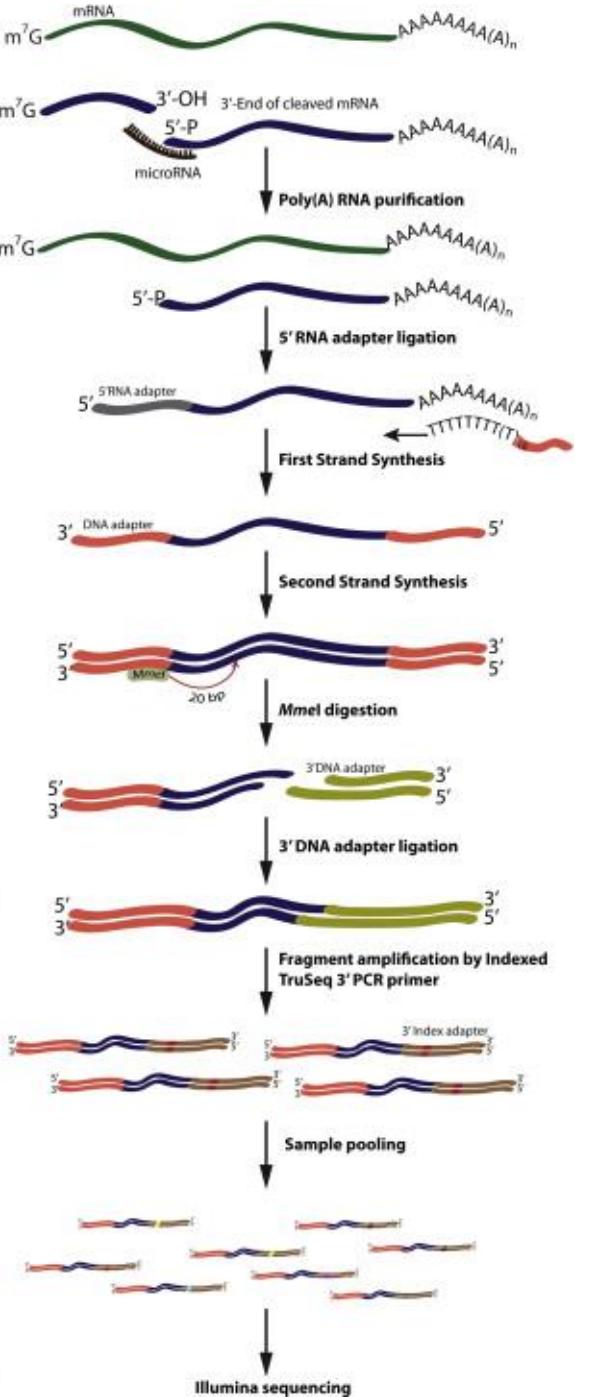
Ribosomal profiling (ribo-seq)

Ingolia et al (2009) Science 324: 218-23



Degradome Sequencing

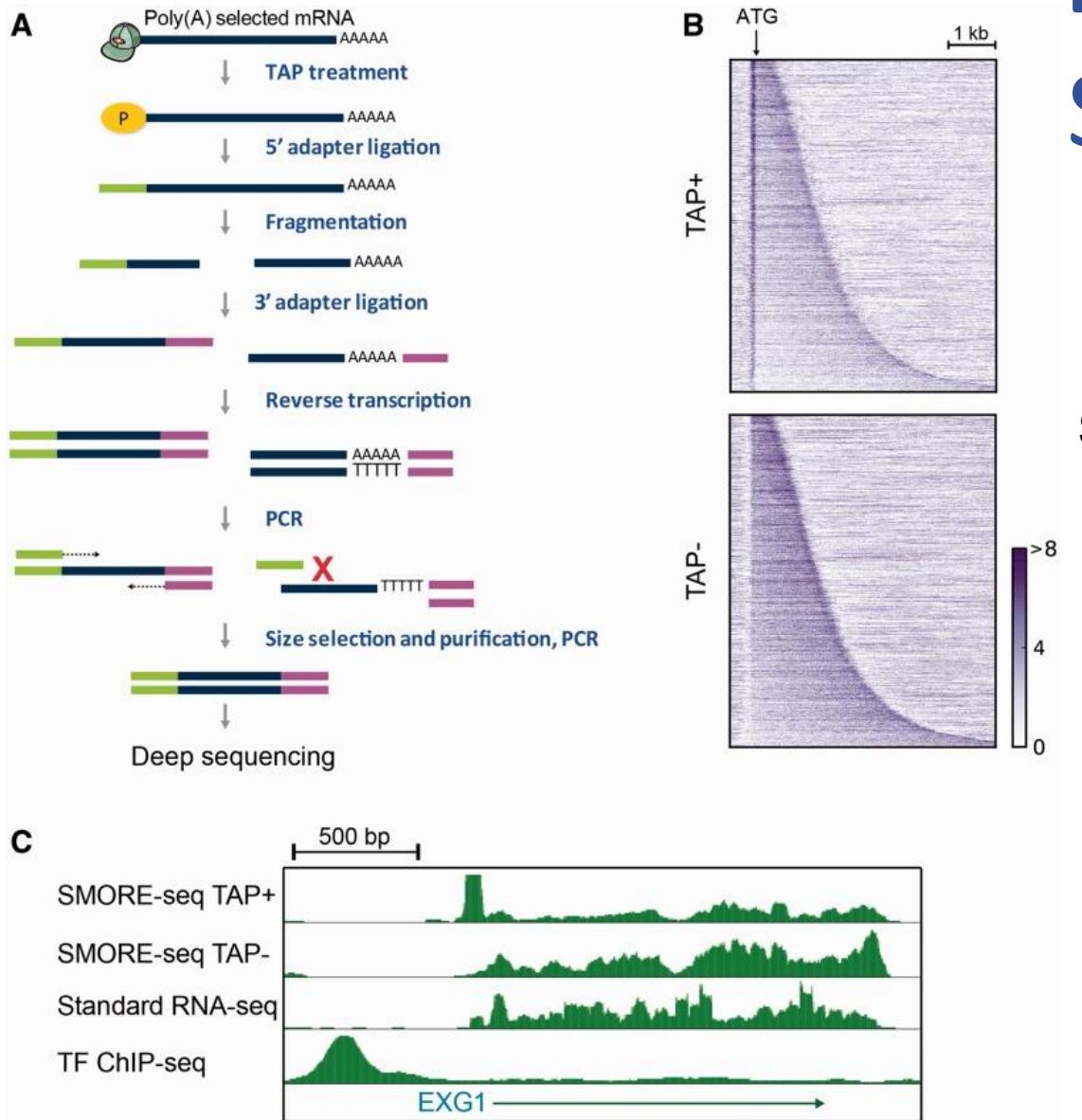
Day 1



PARE-Seq
(Parallel Analysis of RNA Ends)

Zhai et al . 2013

Degradome Sequencing



Park et al . 2014

RNA velocity

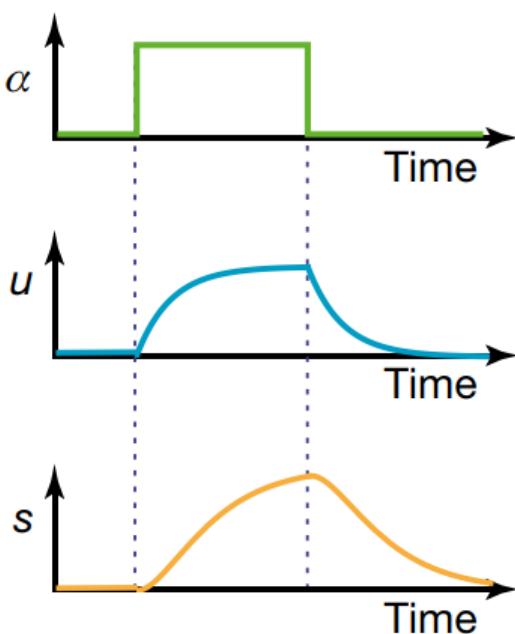
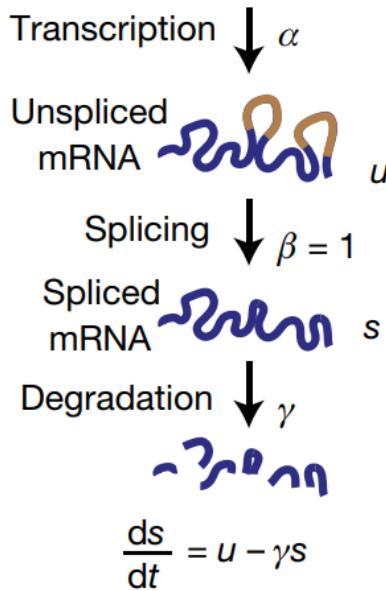
(2018)

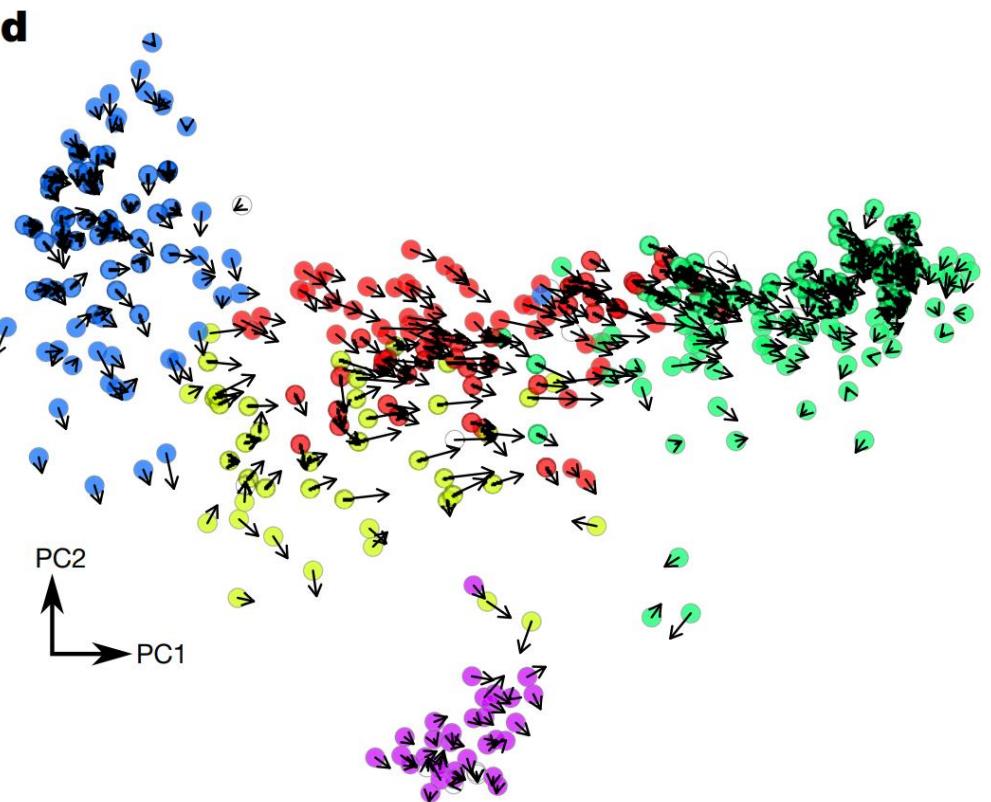
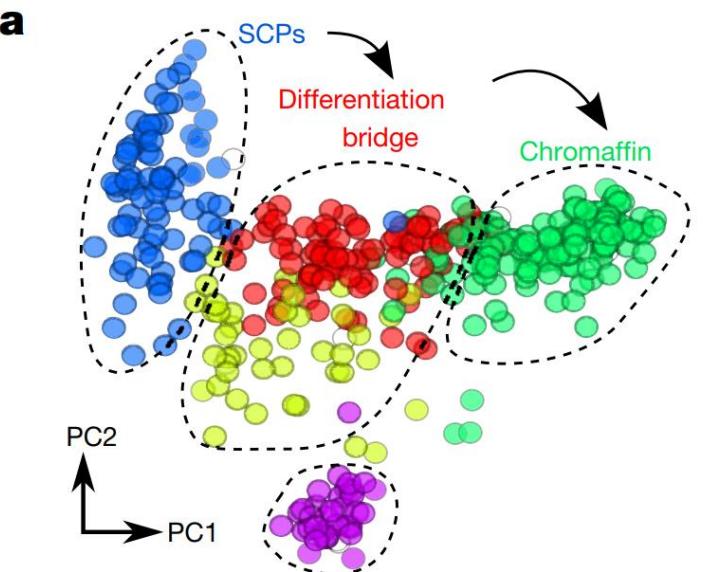
LETTER

<https://doi.org/10.1038/s41586-018-0414-6>

RNA velocity of single cells

Gioele La Manno^{1,2}, Ruslan Soldatov³, Amit Zeisel^{1,2}, Emelie Braun^{1,2}, Hannah Hochgerner^{1,2}, Viktor Petukhov^{3,4}, Katja Lidschreiber⁵, Maria E. Kastriti⁶, Peter Lönnerberg^{1,2}, Alessandro Furlan¹, Jean Fan³, Lars E. Borm^{1,2}, Zehua Liu³, David van Bruggen¹, Jimin Guo³, Xiaoling He⁷, Roger Barker⁷, Erik Sundström⁸, Gonçalo Castelo-Branco¹, Patrick Cramer^{5,9}, Igor Adameyko⁶, Sten Linnarsson^{1,2*} & Peter V. Kharchenko^{3,10*}





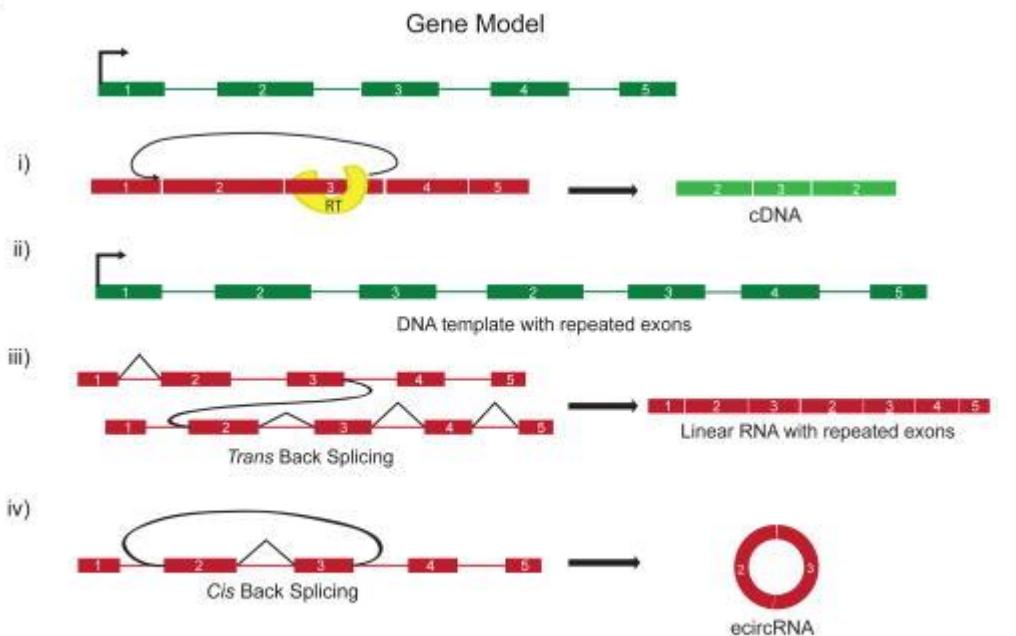
Circular RNA (circRNA)

- Evolutionary conserved
- Eukaryotes
- Spliced (back-spliced)
- Some tissues contain more circRNA than mRNA
- Sequencing after exonuclease digestion (RNase R)
- Interpretation of ribo-depletion RNA-seq data ????

Role of circRNAs ?

Back-splicing and other mechanisms

A



- miRNA sponge
- protein expression regulators:
mRNA traps
(blocking translation)
- Interactions with RNA binding proteins

B

Jeck and Sharpless, 2014



PACIFIC
BIOSCIENCES™

<http://pacificbiosciences.com>

THIRD GENERATION DNA SEQUENCING

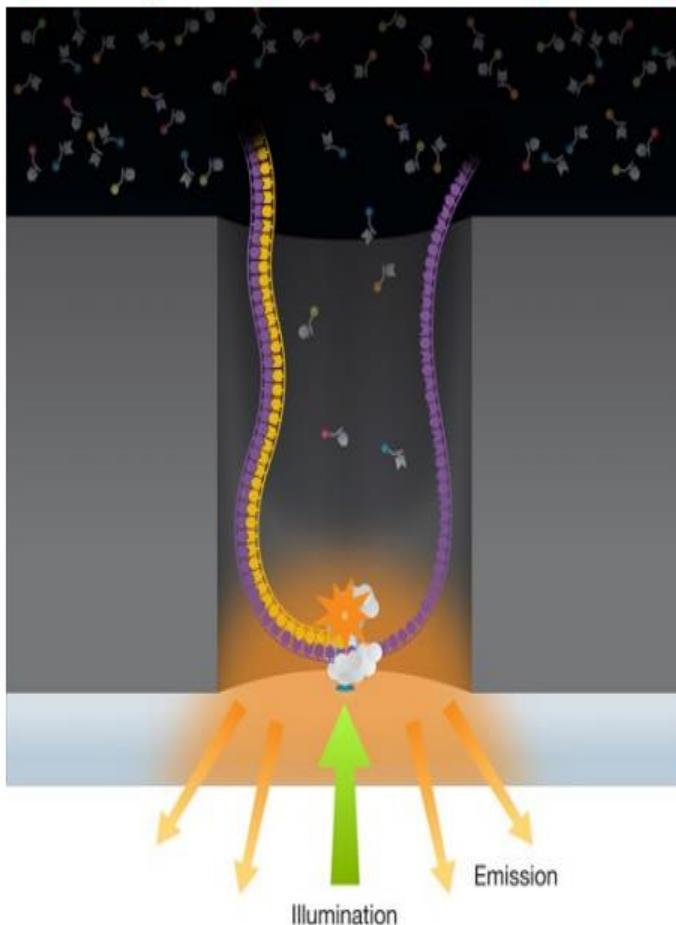


Single Molecule Real Time (SMRT™) sequencing
Sequencing of single DNA molecule by single
polymerase
Very long reads: average reads over 8 kb, up to 30 kb
High error rate (~13%).
Complementary to short accurate reads of Illumina

Third Generation Sequencing : Single Molecule Sequencing

Pacific Biosciences

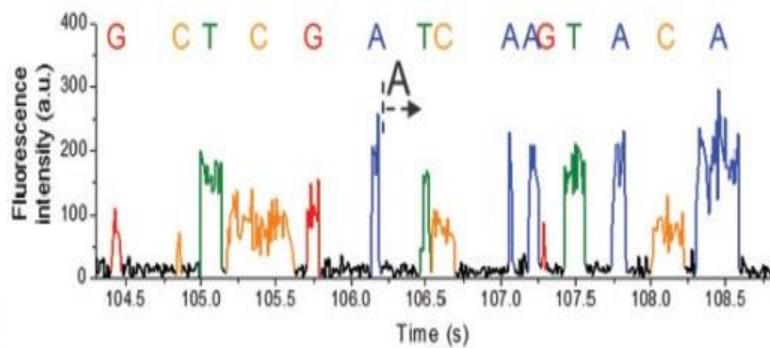
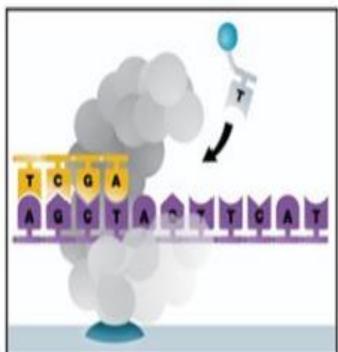
70 nm aperture
“Zero Mode
Waveguide”



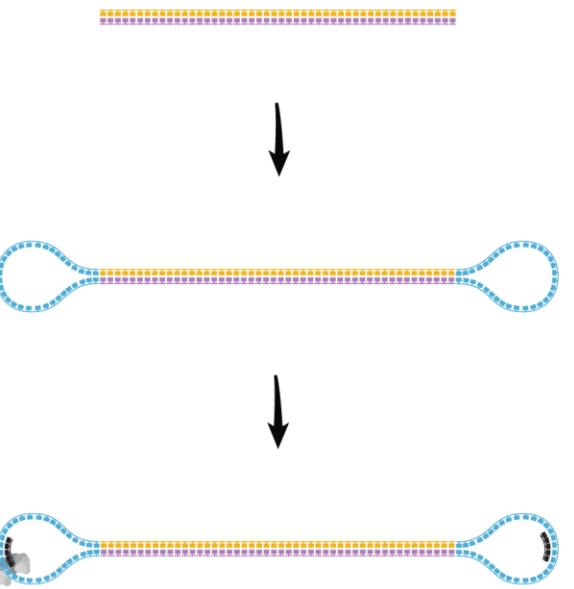
4 nucleotides with different fluorescent dye simultaneous present

2-3 nucleotides/sec
2-3 Kb (up to 50) read length
6 TB data in 30 minutes

laser damages polymerase



Start with high-quality double stranded DNA



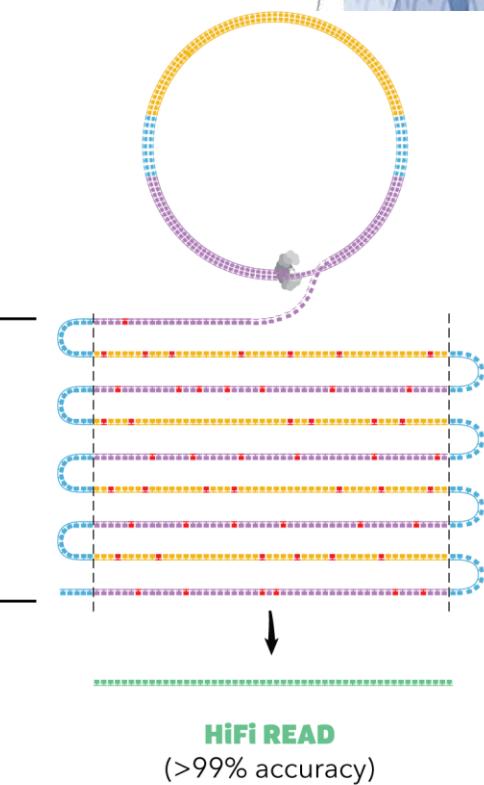
Ligate SMRTbell adapters and size select

Circularized DNA is sequenced in repeated passes

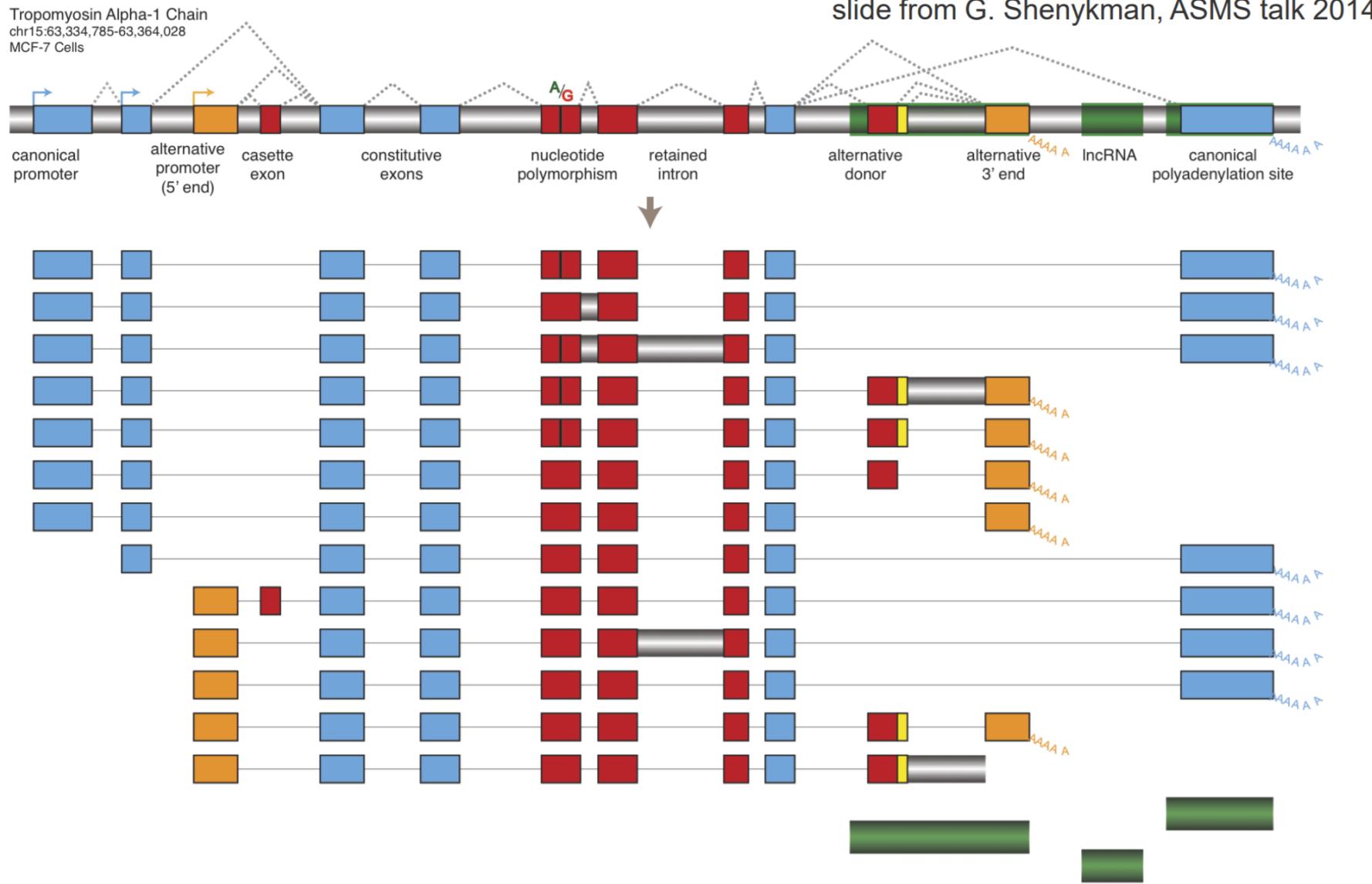
The polymerase reads are trimmed of adapters to yield subreads

Anneal primers and bind DNA polymerase

Consensus is called from subreads



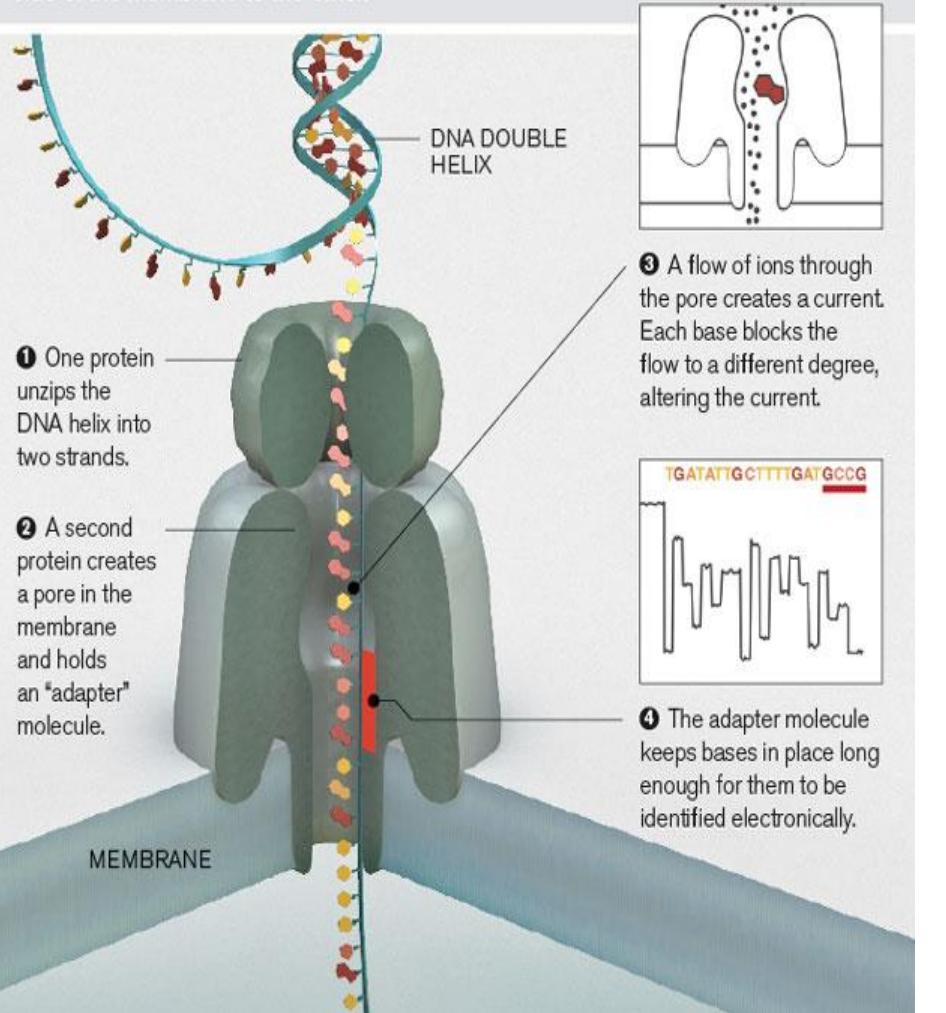
A Single Gene Locus → Many Transcripts



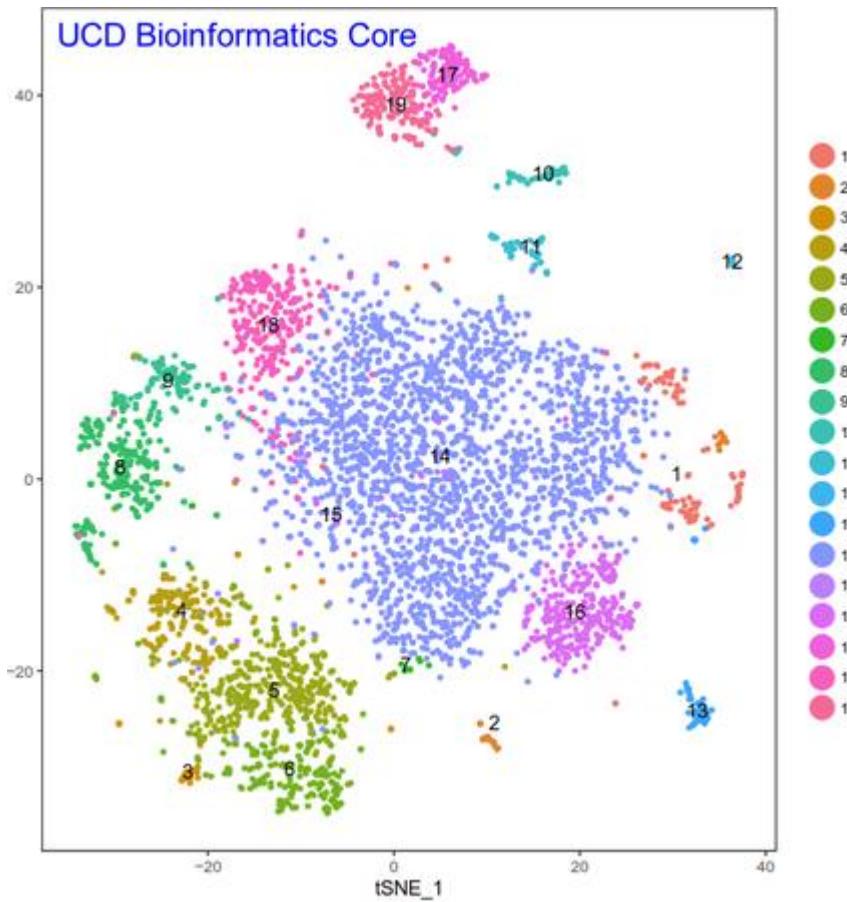
Iso-Seq Pacbio

- Sequence full length transcripts
→ no assembly
- High accuracy (except very long transcripts)
- More than 95% of genes show alternate splicing
- On average more than 5 isoforms/gene
- Precise delineation of transcript isoforms (PCR artifacts? chimeras?)

DNA can be sequenced by threading it through a microscopic pore in a membrane. Bases are identified by the way they affect ions flowing through the pore from one side of the membrane to the other.

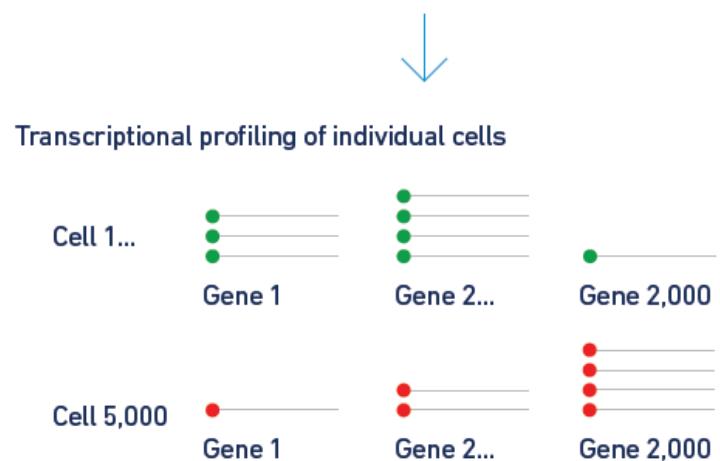
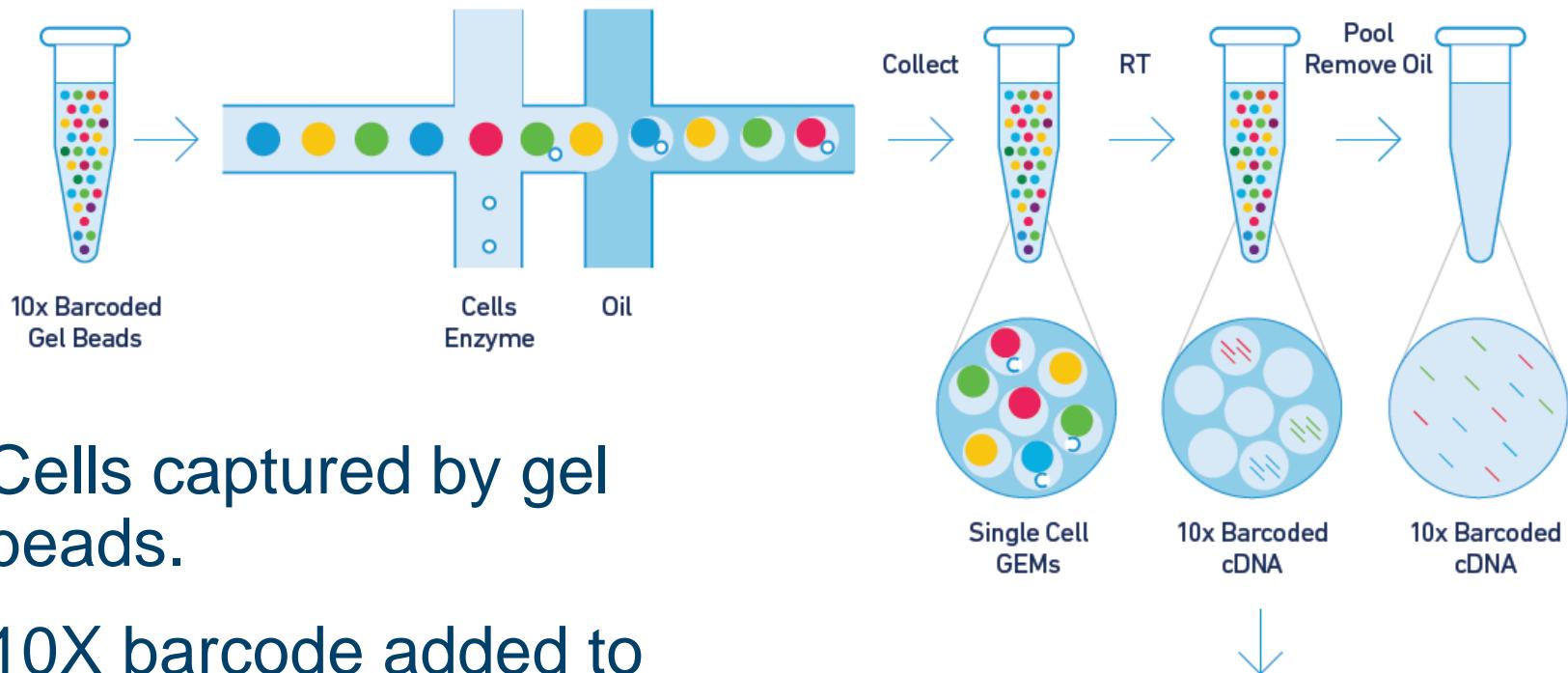


scRNA-seq (single cells)



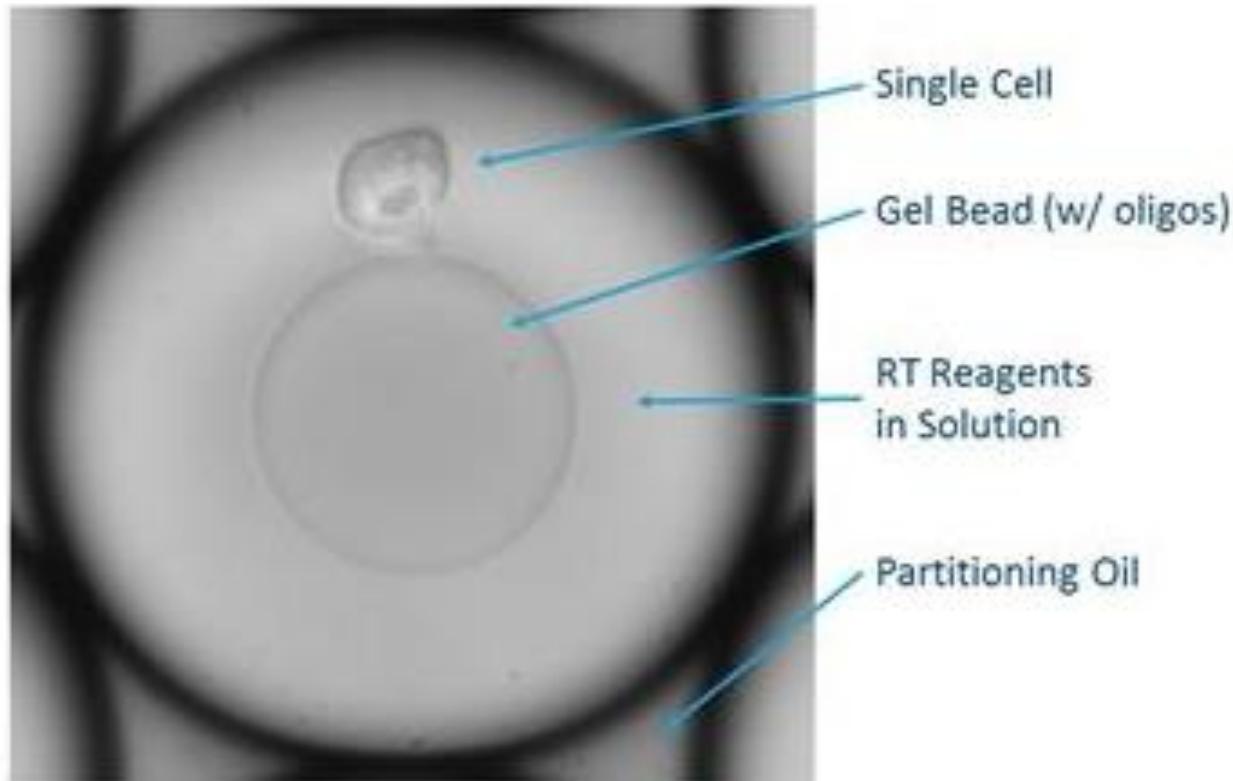
- Gene expression profiling of individual cells.
- Resulting data can distinguish cell types and cell cycle stages - no longer a mix
- Allows the analysis of low abundance cell types

cDNA preparation



Cell partitioning into GEMs

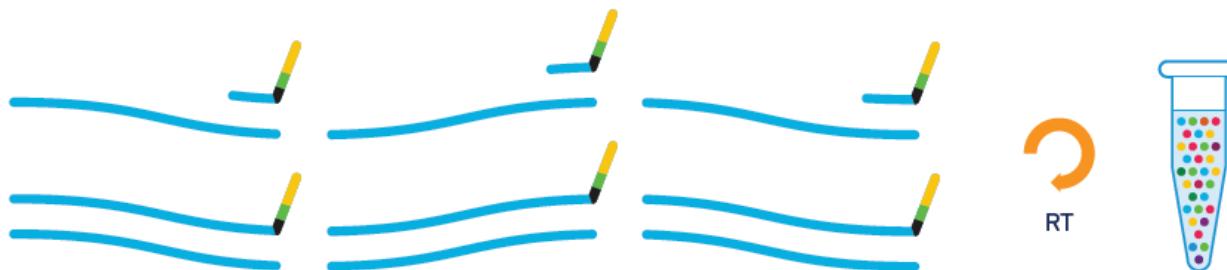
- GEMs Gel Bead-In EMulsions



Credit: 10X Genomics

Library preparation

1 Molecular Barcoding in GEMs



Credit: 10X Genomics

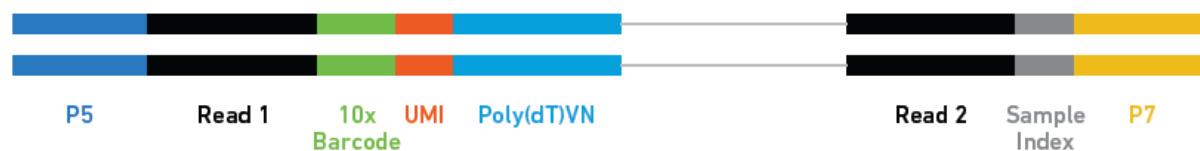
2 Pool, Library Prep



3 Sequence and Analyze

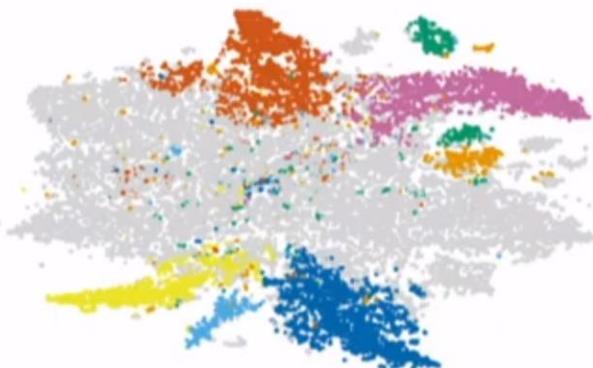


Final Library Construct

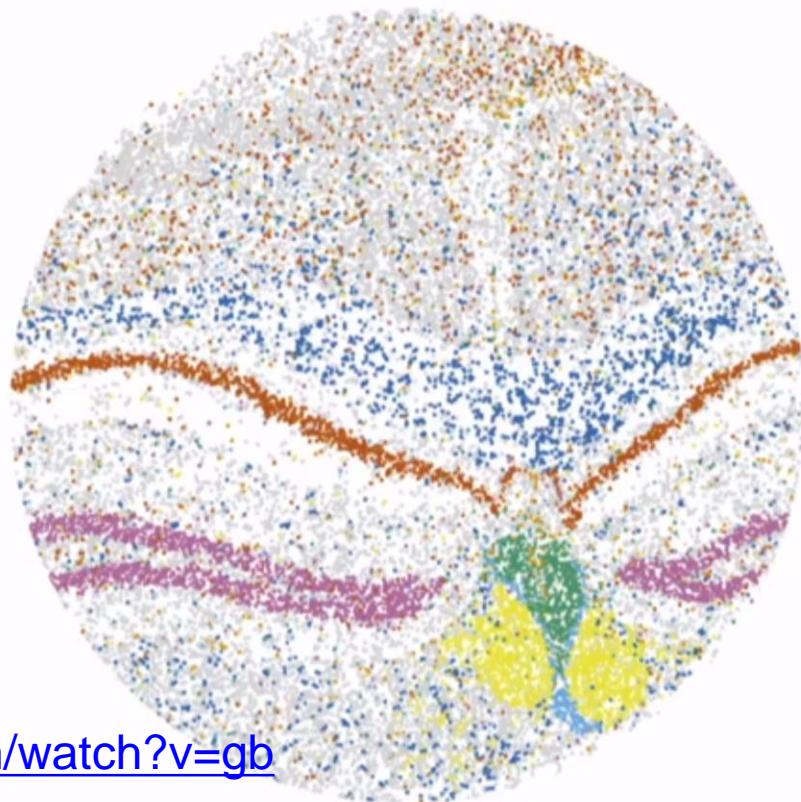


Spatial Transcriptomics (10XGenomics Visium; Slide-Seq)

4 Map gene expression into space

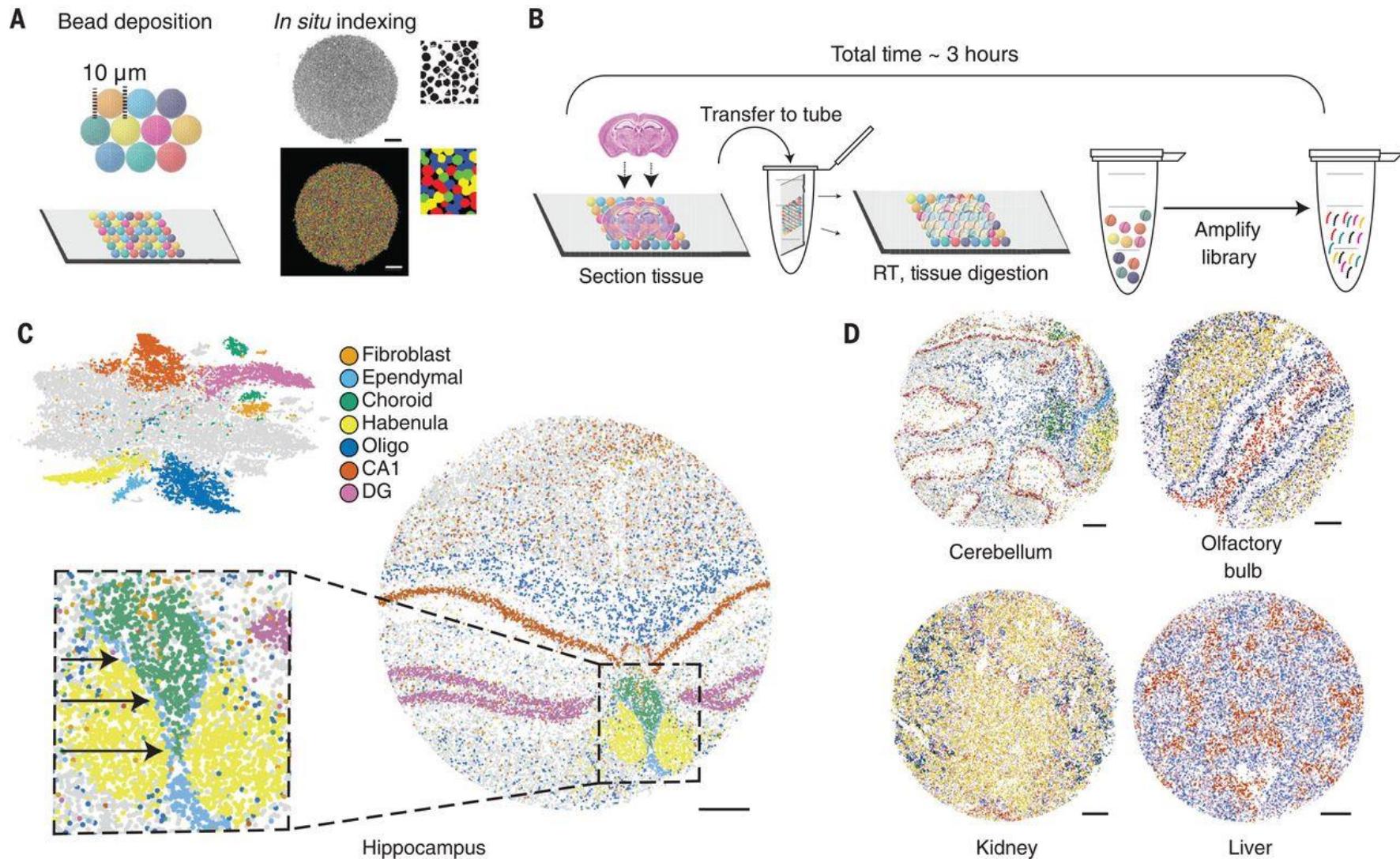


- Fibroblast
- Ependymal
- Choroid
- Habenula
- Oligo
- CA1
- DG



Evan Macosko

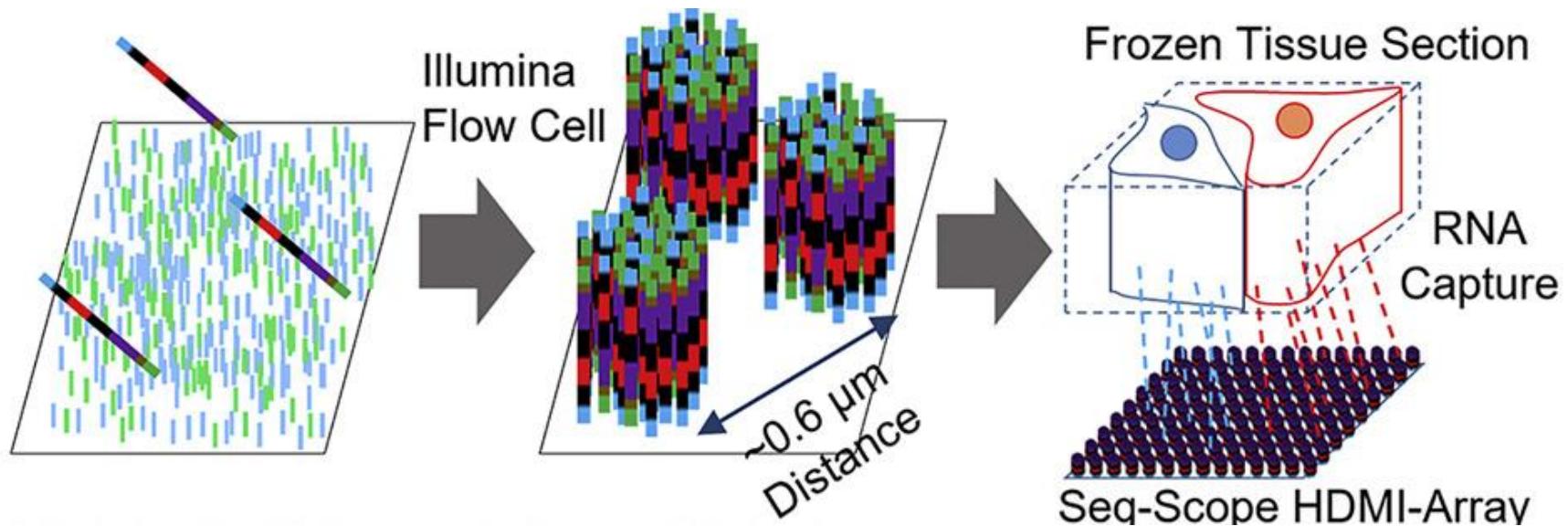
<https://www.youtube.com/watch?v=gbOvgwIQPo8&t=2783s>



Microscopic examination of spatial transcriptome using Seq-Scope

Cho et al. 2011

- Seq-Scope: two rounds of sequencing
- 1st: generation of UMI harboring clusters (HDMI array) on Miseq flowcell → UMI localization
- Mounting of tissue section, fixation, digestion, release of RNA, cDNA-synthesis, denaturation
- 2nd: Sequencing round: → transcript and UMI counts

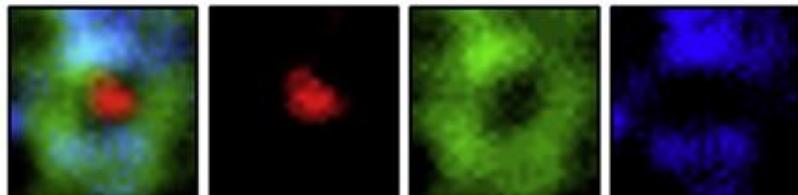


Microscopic examination of spatial transcriptome using **Seq-Scope**

Cho et al. 2021

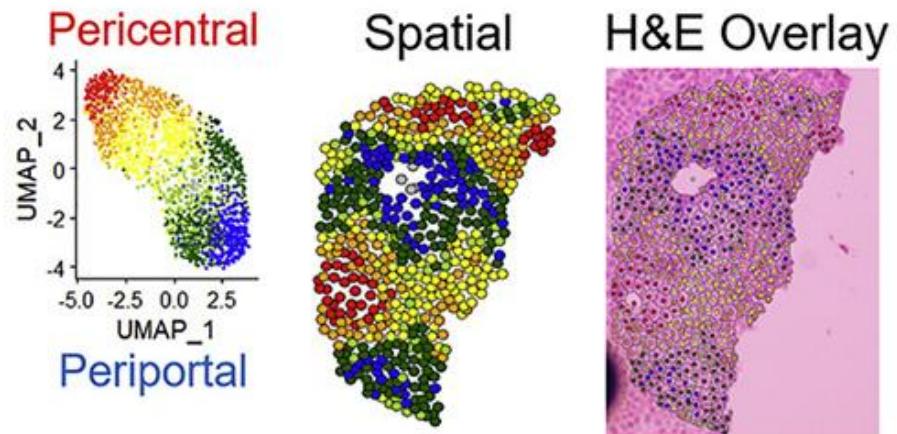
- Seq-Scope: two rounds of sequencing
- 1st: generation of UMI harboring clusters (HDMI array) on Miseq flowcell
- Mounting of tissue section, fixation, digestion, release of RNA, cDNA-synthesis, denaturation
- 2nd: Sequencing round: transcript and UMI counts

Subcellular Transcriptome



Nucleus/Cytoplasm/Mitochondria

Single Cell Transcriptome (Liver)



Seq-Scope

Microscopic examination of spatial transcriptome using **Seq-Scope** Cho *et al.* 2021

- HDMI array generation

- Initial HDMI-TruEcoRI “seed” library

5'-CAAGCAGAAAGACGGCATACGAGAT TCTTTCCCTACACCGACGCTTCCGATCT HNNBNBNBNBNBNBNNNN CCCGTTGCACATGTCGGCGTCATA GAATT CCGAGTCCAG GTGTAGATCTCGGTGGTCGCCGTATCATT-3'
 P7 sequence TruSeq Read 1 HDMI HDMI Read 1B EcoRI P5 sequence

- Sequencing by Synthesis in MiSeq instrument (SBS, single-end)

P7 sequence TruSeq Read 1 HDMI HDMI Read 1B EcoRI P5 sequence
 ~~~CAAGCAGAAGACGGCATACGAGAT TCTTCCCTACAGACGCTTCCGATCT NNVNVNVNVNVNVNVN VN NNNN CCCGTCGAACATGTCTGGCGTCATA GAATTG CCGAGTCCAG GTGTAGATCTCGTGGTCGCCGTATCATT-3'  
 < < < < < < < < GGGCAAGCTGTACAGACCGCAGTAT CTTAAG GCGTCAGGTC-5'

- EcoRI Cut

P7 sequence TruSeq Read 1 HDMI HDMI Read 1B EcoRI P5 sequence  
~~~~~~CAAGCAGAAGACGGCATACGAGAT~~ TCTTTCCTACACGACGCTTCCGATCT **NNVN**NNNNNNNNNNNNN CCCGTCGAACATGTCTGGCGTCATA **GAA**TTC CGCAGTCCAG GTGTAGATCTGGTGGTCGCCGTATCATT-3'  
 CTAGA **NNBN**NNNNNNNNNNNNN GGGCAAGCTTGACAGACGCAGTAT CTTA**A** GCGTCAGGTC-5'

- NaOH wash

P7 sequence TruSeq Read 1 HDMI HDMI Read 1
~~CAAGCAGAAAGACGGCATACGAGAT TCTTCCCTACACGACGTCTCCGATCT NNVNNNNNNNNNNNNNN CCGTTCGCAACATGTCGGCGTCATAG-3

- UMI-oligo annealing and Phusion extension

- NaOH wash & completion of HDMI-array

P7 sequence TruSeq Read 1 HDMI HDMI Read 1B UMI Oligo-dT
 CAAGCAGAAAGACGGCATACGAGAT TCTTCCCTACGACGCTTCCGATCT NNVNVNVNVNVNVN>NNNNN CCCGTGCAACATGTCTGGCGTAG NNNNNNNNN TTTTTTTTTTTTTTTTTTTTTTTTTTT-3'

Future's so bright





Thank you!