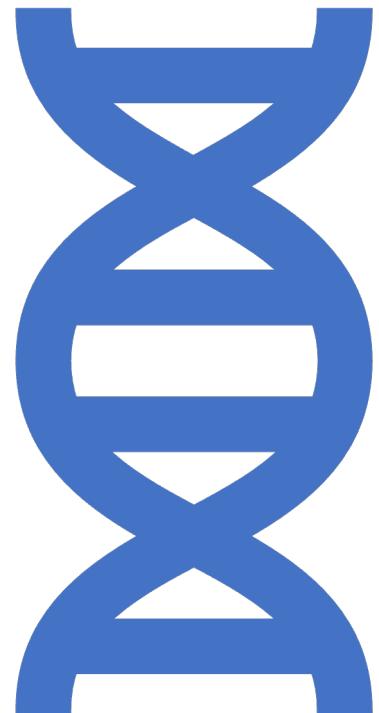


Practical Advice for Next Generation Sequencing

Anoja Perera (agp@stowers.org)

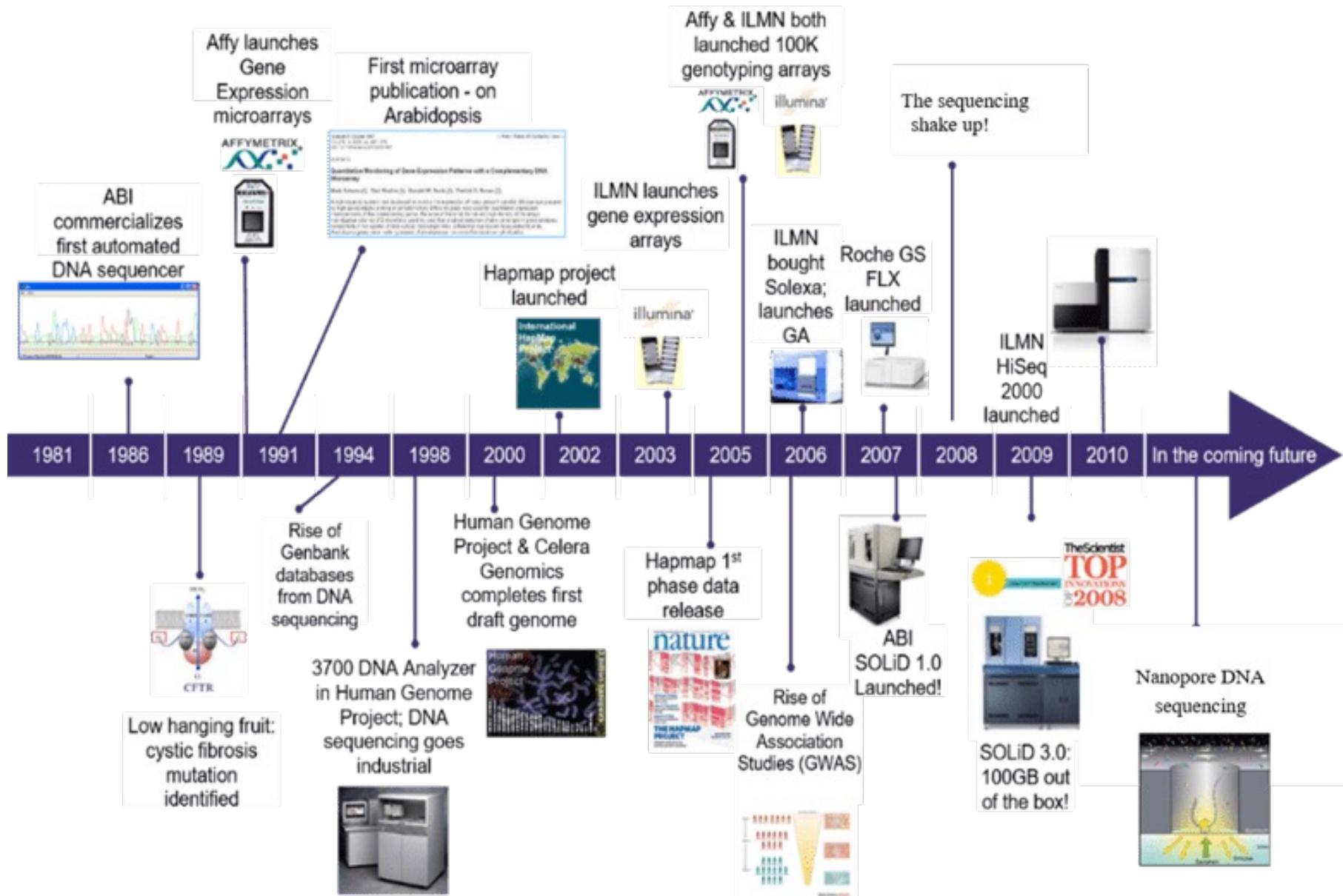
Director, Sequencing and Discovery Genomics
Stowers Institute for Medical Research



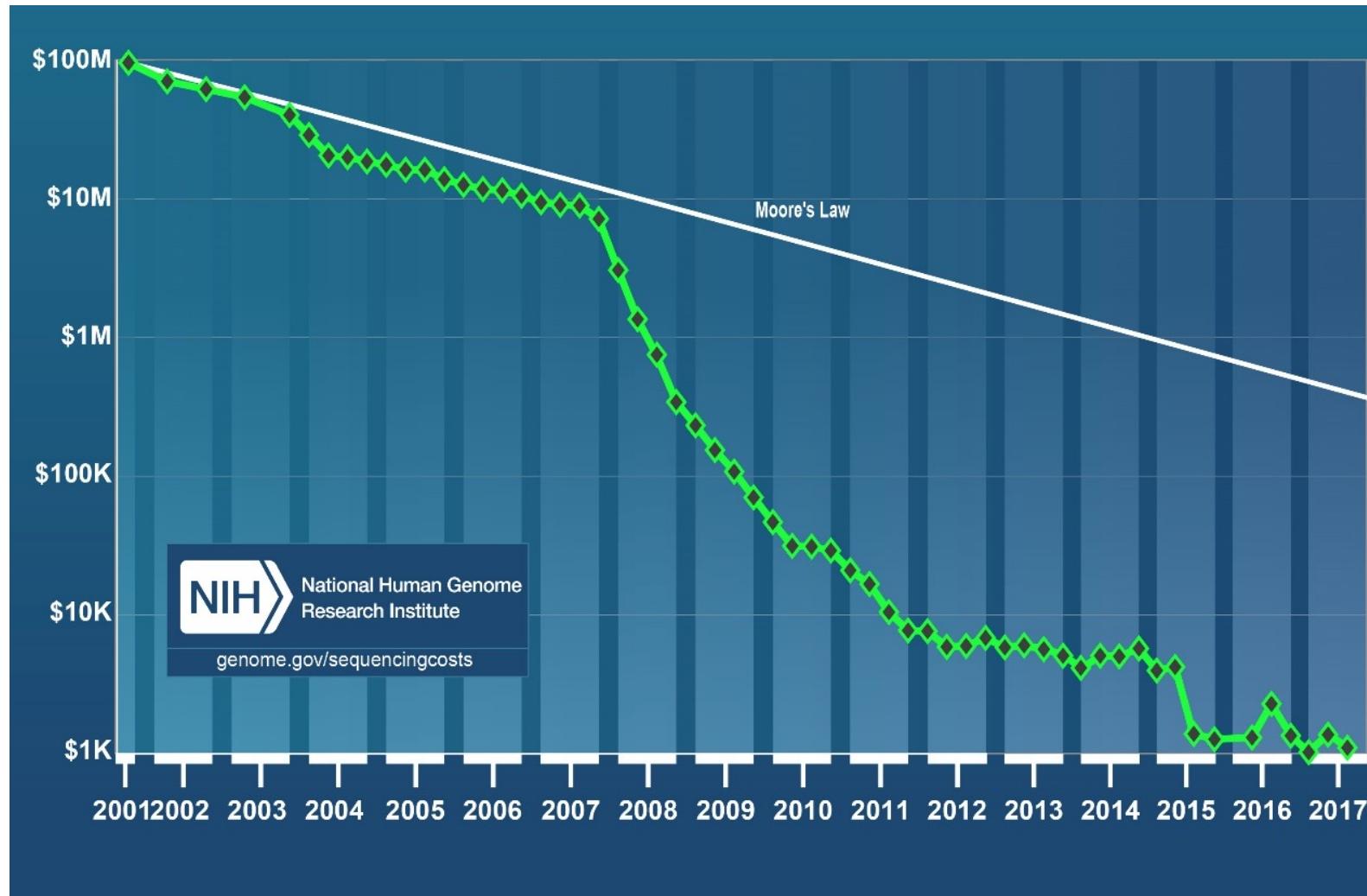
Agenda

- Introduction
- Long-Read Sequencing
- Short-Read Sequencing
- Data Quality Assessment
- Single Cell Sequencing
- Spatial Transcriptomics

A Quick History

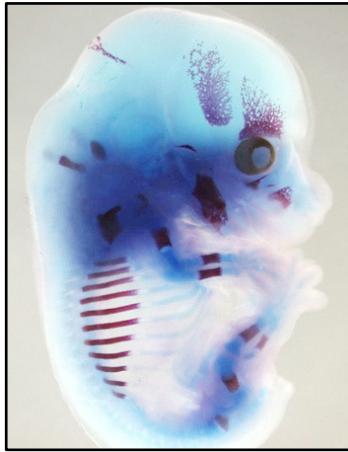


Cost per Human Genome



Source: National Human Genome Research Institute

Examples of Research at Stowers Institute

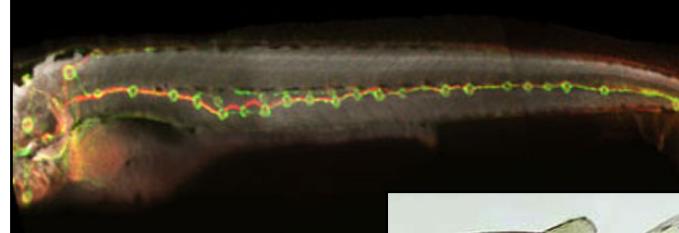


Craniofacial development in mouse embryos

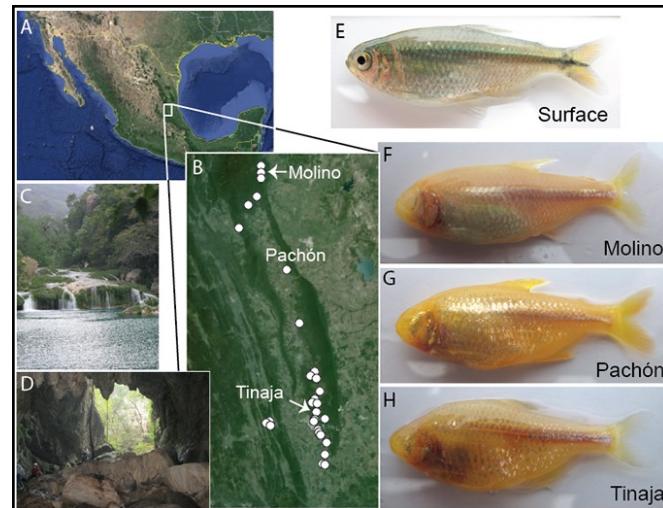


Stem cell regeneration in planaria

Zebrafish sensory system similar to human inner ear



Metabolism in different species of cavefish



Pros and Cons of Long-Read data

Pros

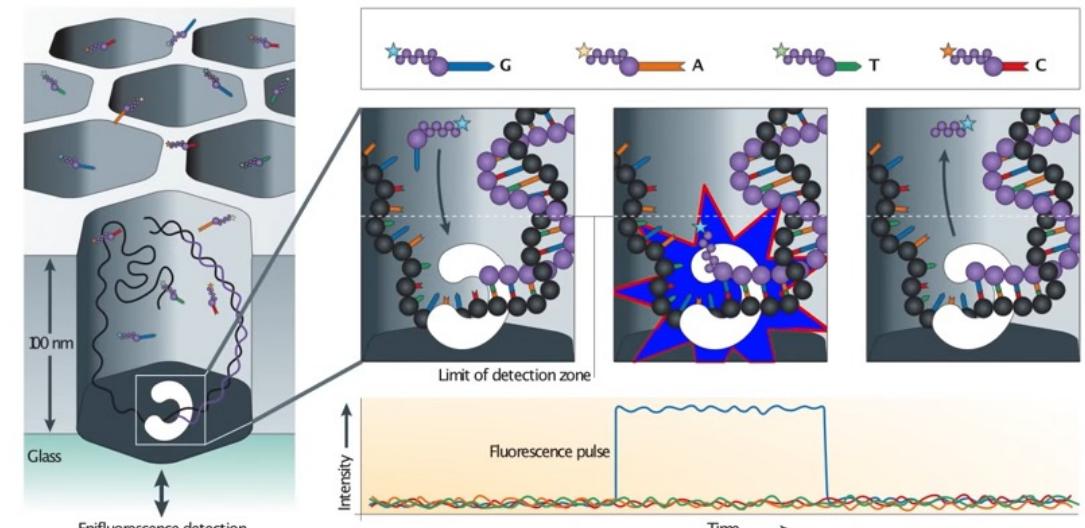
- Direct sequencing of libraries without amplification
- Determination of base modifications such as methylated bases
- Aids in *de novo* assembly by spanning low complexity and repetitive regions
- Assists in determining haplotypes, structural variations, indels etc.
- Can obtain full length transcript information and hence aid isoform discovery (Iso-Seq/MAS-Seq)

Cons

- Costly
- Limited supported applications
- Slow

PacBio Sequencing

Pacbio Revio – Outsource (~\$1300/SMRT Cell)



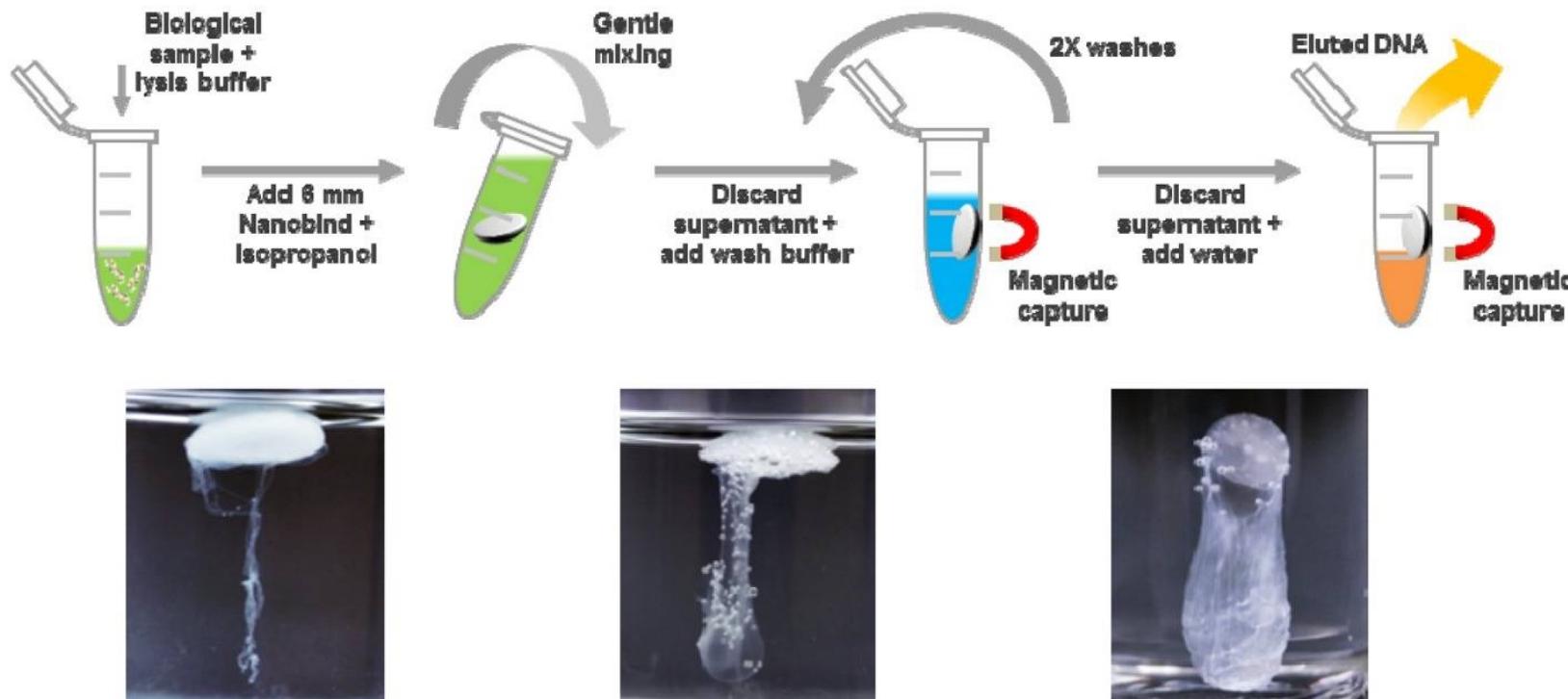
Nature Reviews Genetics 11, 31–46 (2010)



Library	Run time ¹	HiFi yield ^{2,3}		Base quality ²
		1 Revio SMRT Cell	4 Revio SMRT Cells	
15–20 kb	24 hours	90 Gb	360 Gb	90% Q30+

[PacBio Sequencing – How it Works - Bing video](#)

Key to long read sequencing: HMW DNA!



PacBio and NEB Monarch Kits work great!

Pacbio Library Prep

Overview of the SMRTbell Express 2.0 Large-insert Library Workflow

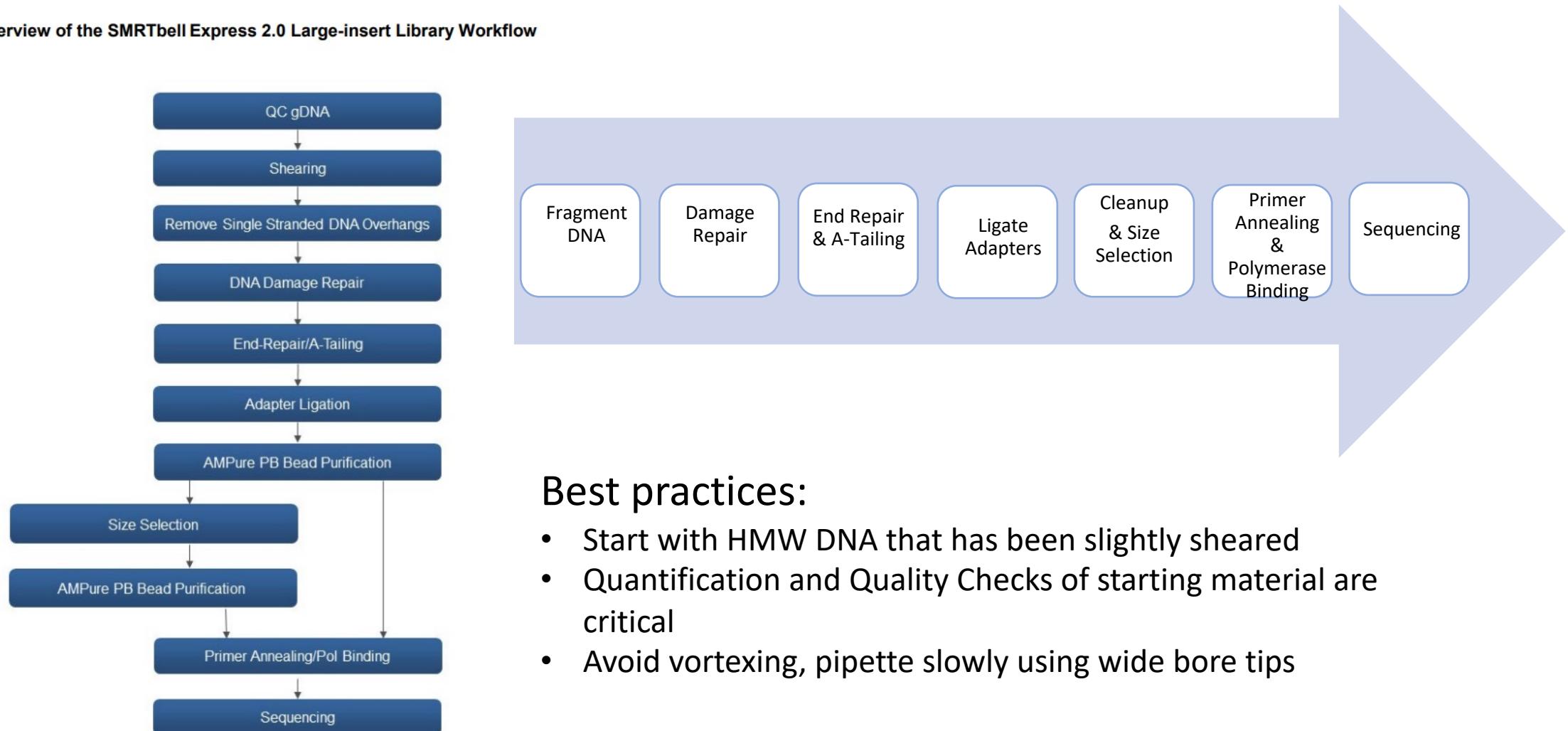
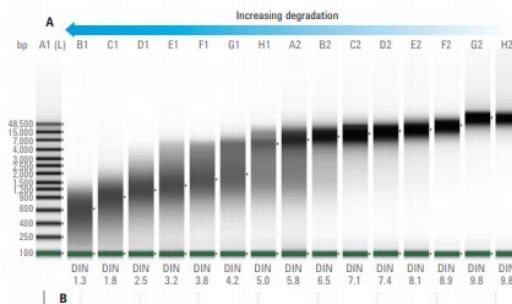


Figure 1: Workflow for Preparing Large-insert Libraries Using the SMRTbell Express Template Preparation Kit 2.0.

DNA Quality and Quantity Assessment

- Check **quality** on a pulse field gel, Agilent Tape Station or Advanced Analytical FEMTO



Agilent Tape Station



Advanced Analytical Femto

- Check for **impurities** via Nanodrop. 260 nm: 280 nm of ~ 1.8 is accepted as “pure” DNA; a ratio of ~ 2.0 is accepted as “pure” RNA. If the ratio is lower in either case, it may indicate the presence of protein, phenol or other contaminants that absorb strongly at or near 280 nm.
- Check for **quantity** via Invitrogen Qubit Fluorometer



Thermo Nanodrop



Invitrogen Qubit Fluorometer

Physical / Mechanical Shearing

Method	Time	DNA Input	Fragment Length
Covaris Ultrasonicator	~ 1 min	Up to 5 µg	100 bp – 5 kb
Covaris g-Tube	10 min	~ 10 µg	6 – 20 kb
Needle Shearing	20 min	2 – 10 µg	> 30 kb
Diagenode Megaruptor	10 – 20 min	Up to 8 µg	2 – 75 kb



Oxford Nanopore Sequencing (ONT)

Additional Pros:

- No limitations in length of molecules that can be sequenced (>2 Mb)
- Capable of measuring DNA, RNA, protein, and modifications
- Can be used for unique applications such as pathogen detection, plasmid sequencing, full length amplicon sequencing
- Active community improving and developing protocols
- Some platforms (Fongle, MinION) are portable!

Additional Cons:

- Error rate is higher than other methods
- Protocols are often not standardized
- Getting consistent output per flowcell is impossible!

Oxford
Nanopore
Sequencing

ONT sequencing platforms



Flongle



MinION
Mk1B



GridION X5



PromethION



MinION
Mk1C



PromethION 24/48

<https://binged.it/3EG0dB7>

Oxford Nanopore Sequencing in the Field



Sequencing microbial DNA in Antarctica



NASA astronaut sequencing bacteria in space



Sequencing Ebola virus in Guinea



Sequencing Zika virus in Brazil
Ricardo Filho/Zika project

Short Reads: Illumina Sequencing

Pros and Cons of Short-Read data

Pros

- Cost per base is significantly lower than long-reads methods
- Great for counting applications such as Chromatin Immunoprecipitation Sequencing (ChIP-Seq) or gene expression studies such as RNA-Seq
- Due to accuracy, works well for SNP detection
- Fast data generation
- More established pipelines for data analysis
- Many supported applications

Cons

- PCR bias introduced by clonal amplification
- Difficulty sequencing high GC regions
- Difficulty in resolving substitution, deletions, duplications, haplotypes, palindromic and repetitive regions

Short Reads Sequencing: Illumina Capacity



MiSeq Series



NextSeq 550 Series



NextSeq 1000 & 2000



Production-Scale Sequencers



NovaSeq 6000 Series



NovaSeq X Series

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

Short Reads Sequencing: New Sequencers



- Flexibility**
1–4 flow cells
16 lanes
Unparalleled operational efficiency
- Speed**
Daily Sequencing
Industry leading run times across applications
- Power**
3.2 Billion Reads*
480 Gb
More data per day than any other benchtop sequencer
- Accuracy**
80–90% bases \geq Q30
Novel 4 color, Rapid SBS chemistry



ELEMENT AVITI

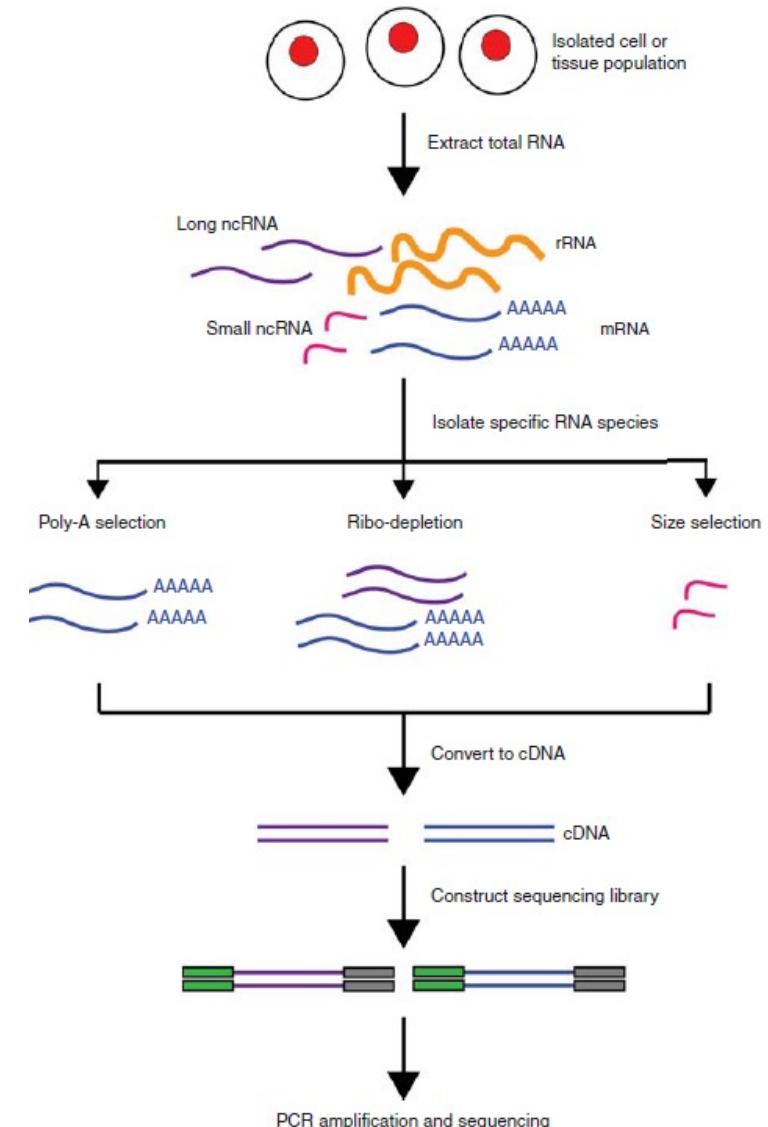
Singular G4

- Cost per base is significantly lower than illumina desktop instruments
- Multiple flowcells can be run at a time
- Flexible output
- Higher accuracy than illumina
- Faster data generation than illumina desktop instruments

Short Reads: Library Construction

RNA Library Construction

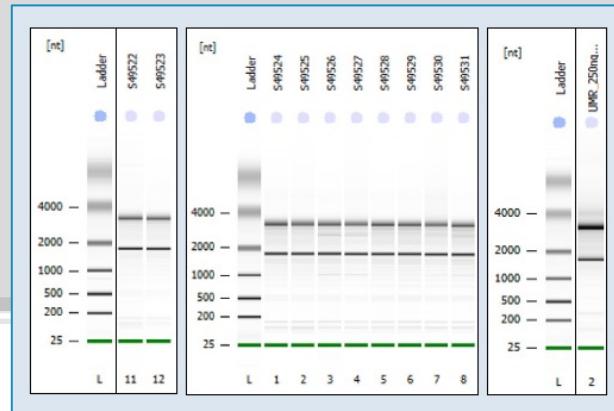
- Depending on the type of RNA interested, different protocols available
- Proper RNA extraction is critical especially if interested in small RNA. (Zymo Direct-zol)
- Quantifications and Quality checks are important to determine library construction method and to get good quality data
- Always have replicates! At least three biological replicates.
- Many RNA kits in the market and some kits work better than others depending on what is needed.
 - We use NEB Ultra II kits to miniaturized reactions to reduce costs.
 - Takara SMARTer works well when starting with few cells or few picograms of material
- For degraded RNA, avoid poly-A selection
- Most ribodepletion kits support Human, Mouse and Rat only. Companies like Qiagen have more additional depletion kits.
- Paired end sequencing is recommended but might not be necessary depending on biological question.



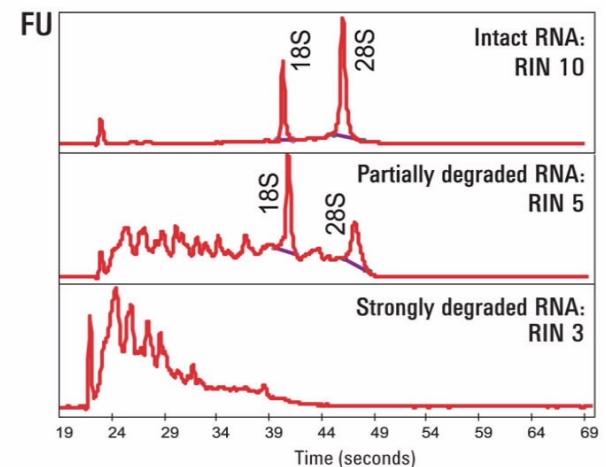
RNA: Starting Material is Important!

- RNase degradation of RNA samples is a common reason for failed experiments.
- Decreasing ratio of ribosomal bands
- Additional peaks below the ribosomal bands
- Decrease in overall RNA signal
- Shift towards shorter fragments
- The Bioanalyzer system provides a RIN (RNA Integrity Number) value, an objective metric of total RNA quality ranging from 10 (highly intact RNA) to 1 (completely degraded RNA).

Sample ID	RIN
S49522	9.1
S49523	8.9
S49524	8.4
S49525	8.4
S49526	8.3
S49527	8.6
S49528	8.3
S49529	7.9
S49530	8.4
S49531	8.0
SUMR	9.2



Agilent Bioanalyzer System



How Much to Sequence?

COVER ME!

What kind of coverage should I expect from my NGS project?

250.8x coverage (130 million reads per sample)

Flowcell

Technology

Reads per Lane

32.5 million

Read Length

Number of lanes

Samples per lanes

Sample

Organism

Sequence Type

Transcriptome Size

77.8 million

That Enough?

For Transcriptome data, we recommend:

15x coverage.

Listen, we all love science, but this might be a bit overkill.

Genome sizes taken from a number of online resources.

Transcriptome sizes found by looking at summary data from [UCSC RefGene Tables](#).

Flowcell read numbers estimated from Illumina and in-house sources.

Amount of sequencing needed will depend on many things; organism, application, starting material, instrument used etc.

<http://metalhelix.github.io/coverme/>

Library QC

- Quality of libraries is a good indication of how your data will look.
- qPCR will help you quantify fragments that are “sequence-able”
- Quantity assessment via Invitrogen Qubit Fluorometer
- Agilent Bioanalyzer to determine quality; size of fragments, presence of adapter dimer, overamplification etc.

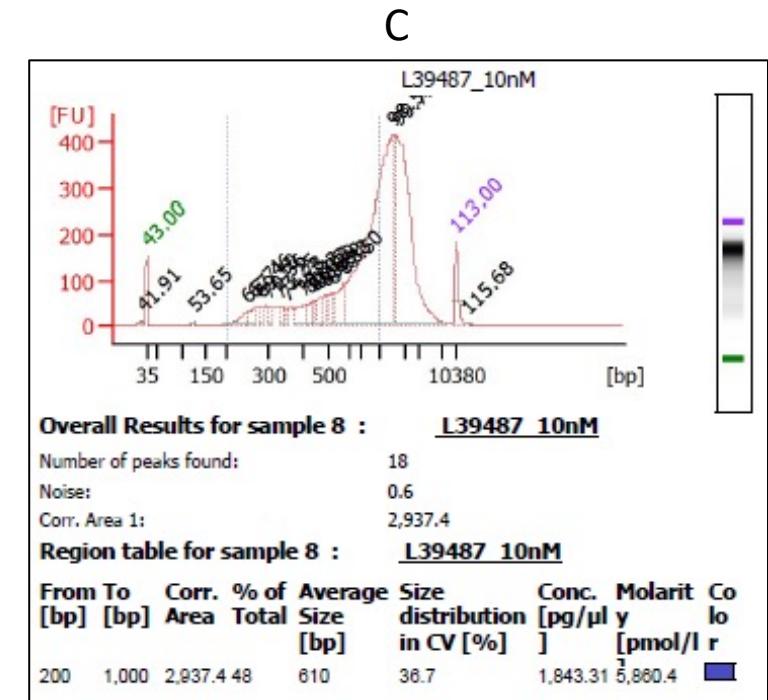
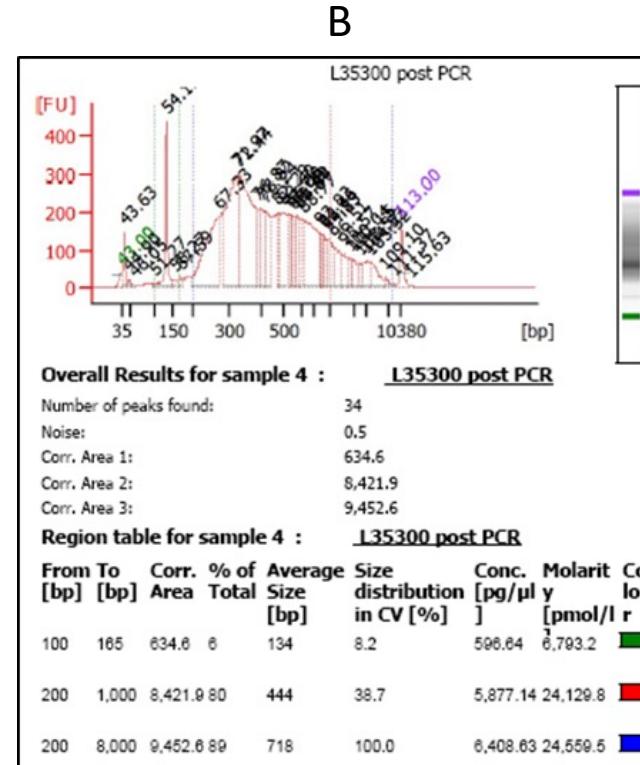
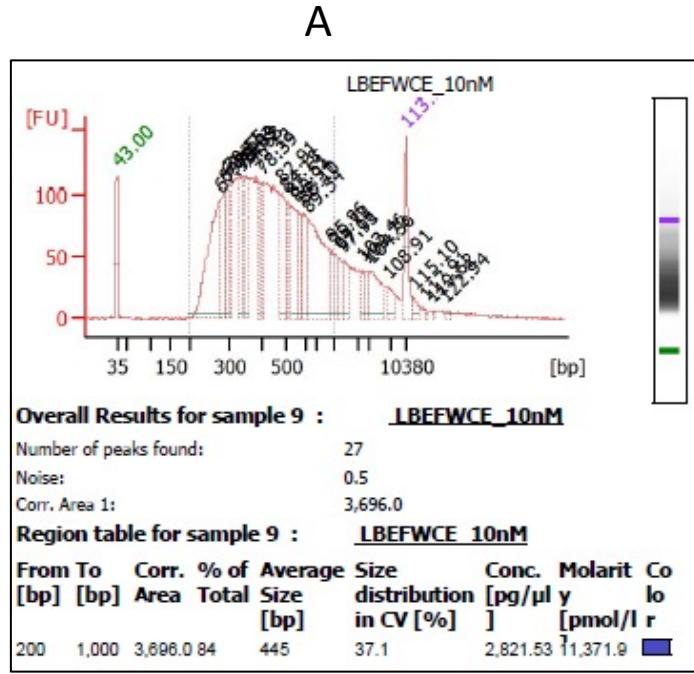


Invitrogen Qubit Fluorometer



Agilent Bioanalyzer System

Good vs. Ugly Libraries



Normal:
Technical
Control

Adapter Dimer

Larger Fragments,
Unsuitable for short read
Sequencing

Library Size Selection



Sage Science's Blue Pippin



Beckman Coulter's SPRI Select beads

- Two popular methods of size selection:
 - Sage Sciences' automated instruments
 - Bead based size selection
- Size selection is done to:
 - get rid of adapter sequences
 - get fragments in the proper size range for sequencing

Post Sequencing Quality Checks

Primary Analysis and QC checks

- Read counts per index/library
- Index accuracy
- Alignment
- Contaminations
- Quality of reads
- Contamination
- Other (RNA-specific checks)

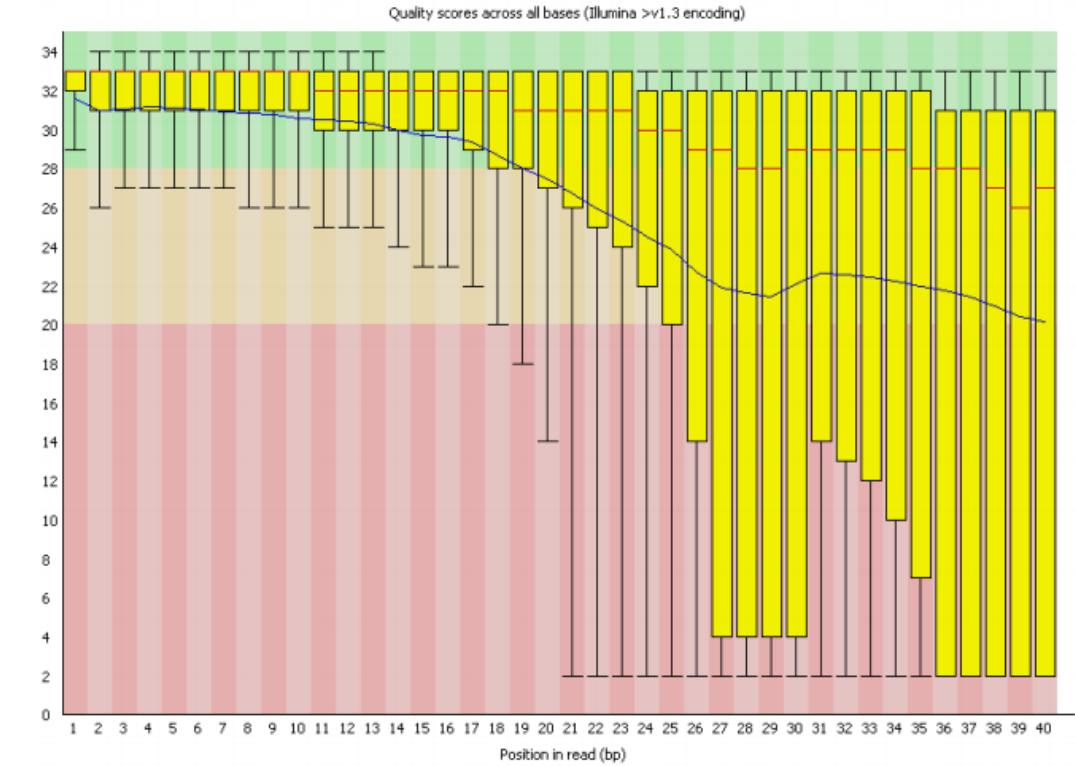
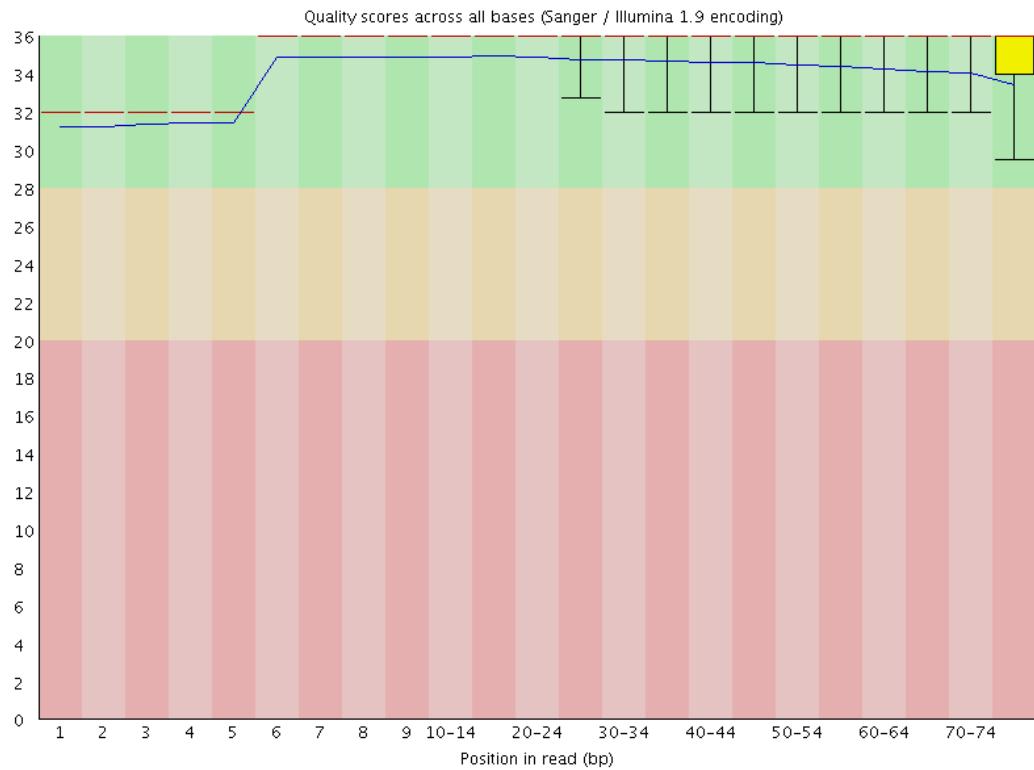
Read Counts and Alignments

Order Type	Lane	Read No	Sample ID	Sample Name	Cluster Cou...	% Align PF	Index Sequence
poly-A Stranded RNA-Seq	1	1	S39858	V6.5_ES_епiLC_24hr_rep1	20600098	84.75	GATCAG
poly-A Stranded RNA-Seq	1	1	S39857	V6.5_ES_2i_rep2	21669575	85.55	ACAGTG
poly-A Stranded RNA-Seq	1	1	S39861	V6.5_ES_епiLC_48hr_rep2	22241206	84.82	ATTOCT
poly-A Stranded RNA-Seq	1	1	S39856	V6.5_ES_2i_rep1	16414815	85.72	TTAGGC
poly-A Stranded RNA-Seq	1	1	S39860	V6.5_ES_епiLC_48hr_rep1	21155517	84.71	CCTAGG
poly-A Stranded RNA-Seq	1	1	S39859	V6.5_ES_епiLC_24hr_rep2	21083786	84.79	CTTGTA

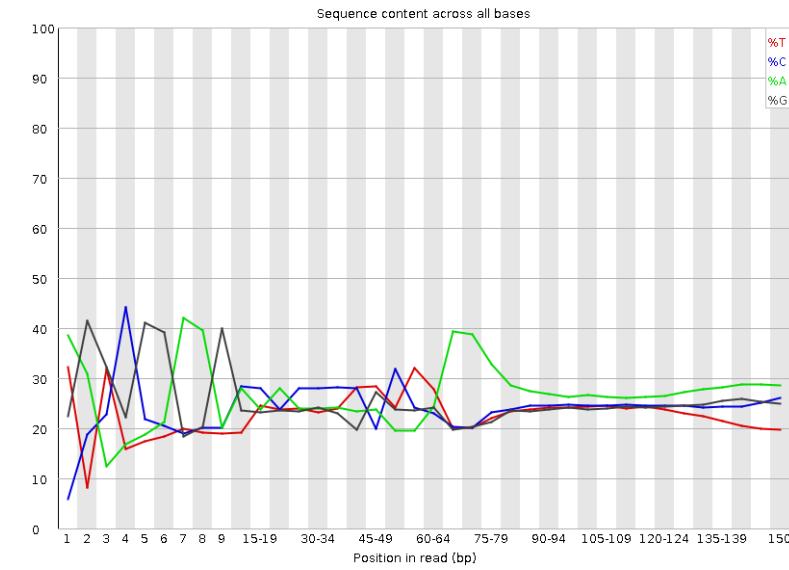
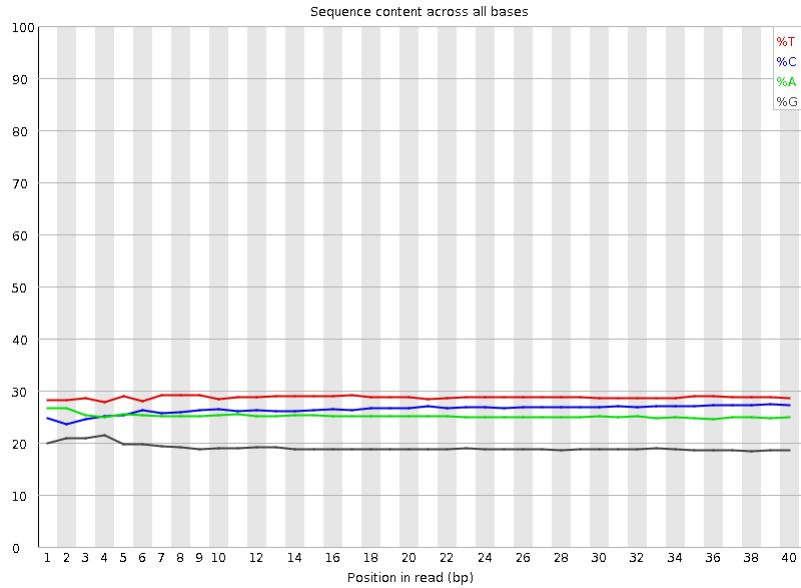
Read Counts and Alignments

Order Type	Lane	Read No	Sample ID	Sample Name	Cluster Cou...	% Align PF	Index Sequence
poly-A Stranded RNA-Seq	1	1	S39858	V6.5_ES_епiLC_24hr_rep1	20600098	84.75	GATCAG
poly-A Stranded RNA-Seq	1	1	S39857	V6.5_ES_2i_rep2	21669575	85.55	ACAGTG
poly-A Stranded RNA-Seq	1	1	S39861	V6.5_ES_епiLC_48hr_rep2	22341206	84.82	ATTCCCT
poly-A Stranded RNA-Seq	1	1	S39856	V6.5_ES_2i_rep1	16414815	85.72	TAGGC
poly-A Stranded RNA-Seq	1	1	S39860	V6.5_ES_епiLC_48hr_rep1	21799517	84.71	CGTACG
poly-A Stranded RNA-Seq	1	1	S39859	V6.5_ES_епiLC_24hr_rep2	21083786	84.79	CTTGTA

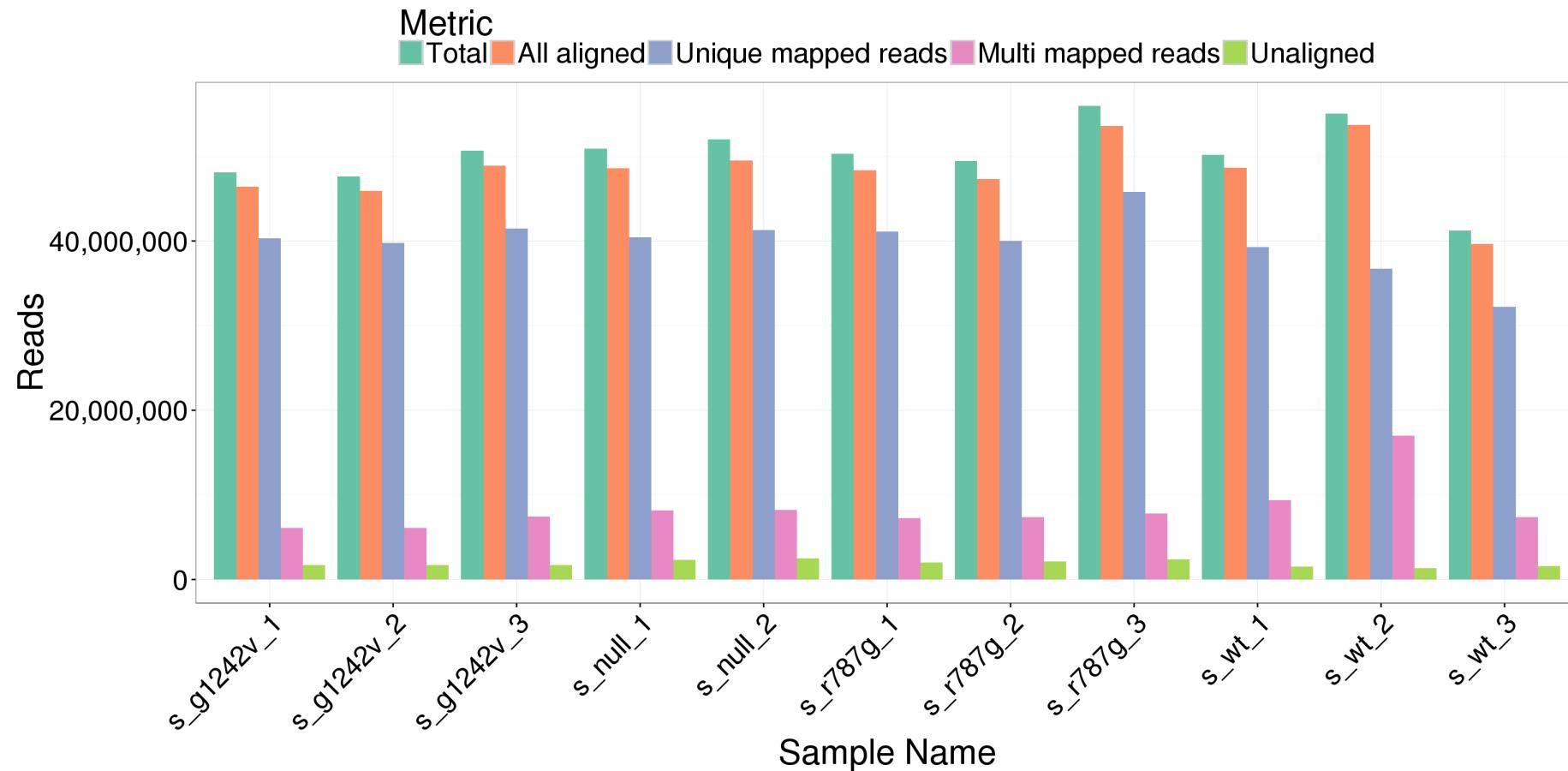
Per Base Sequence Quality



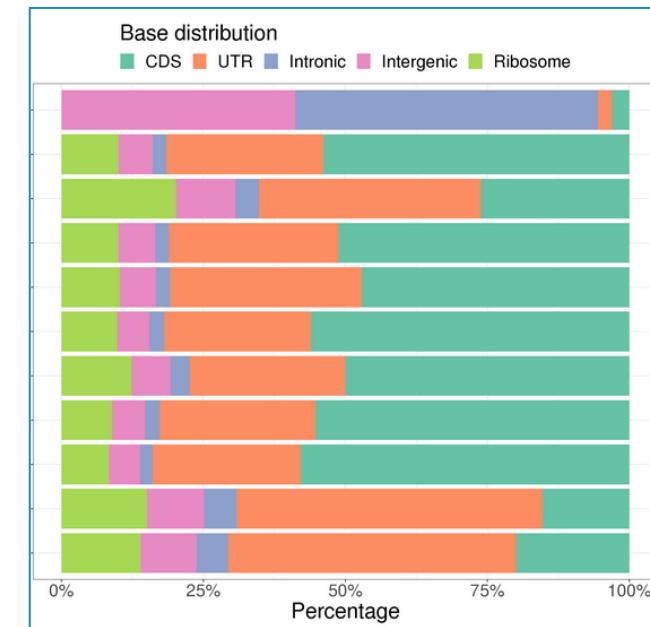
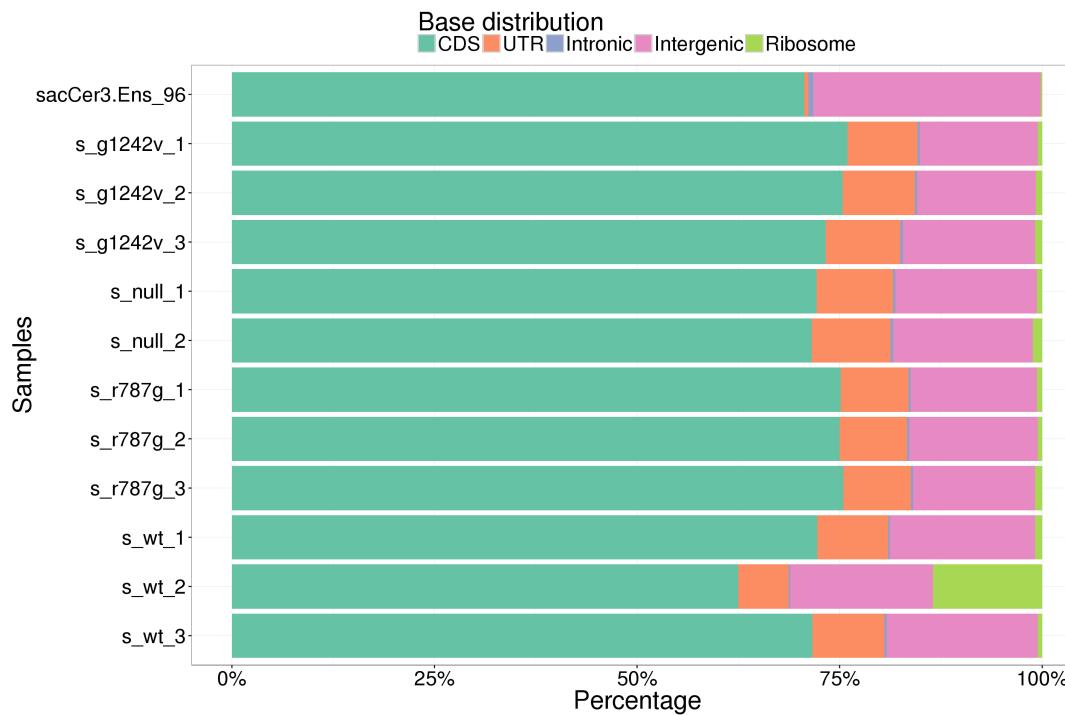
Per Base Sequence Content



Internal Pipeline: Secundo



Internal Pipeline: Secundo



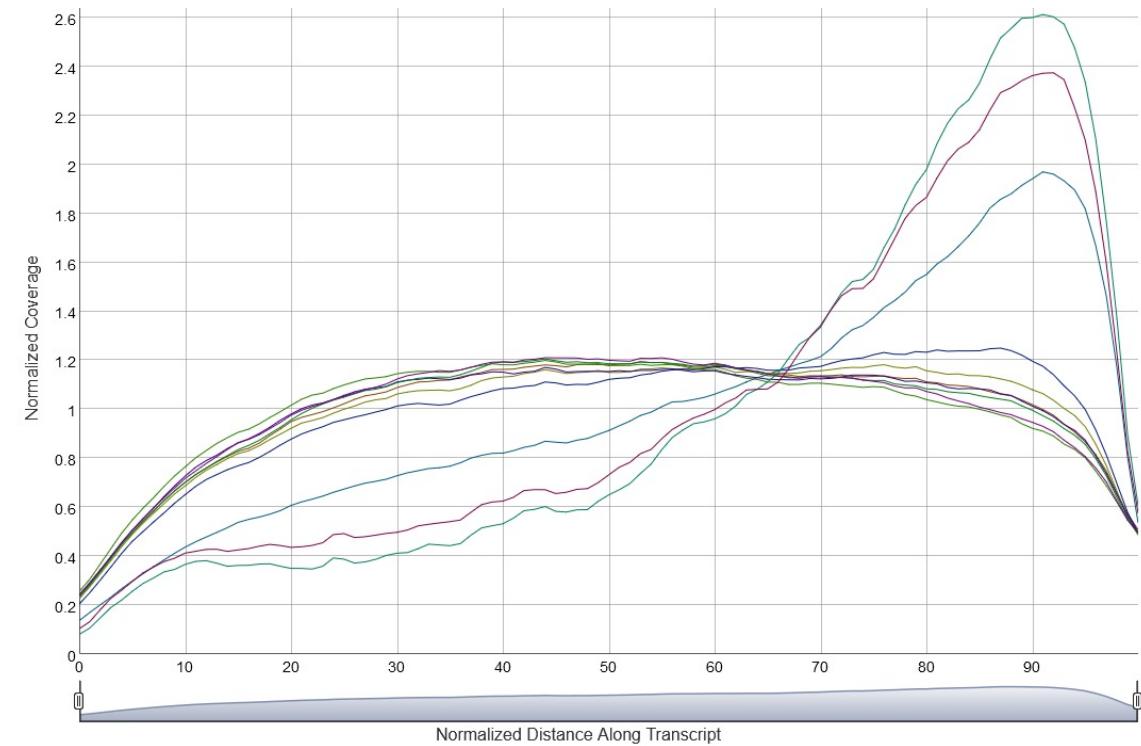
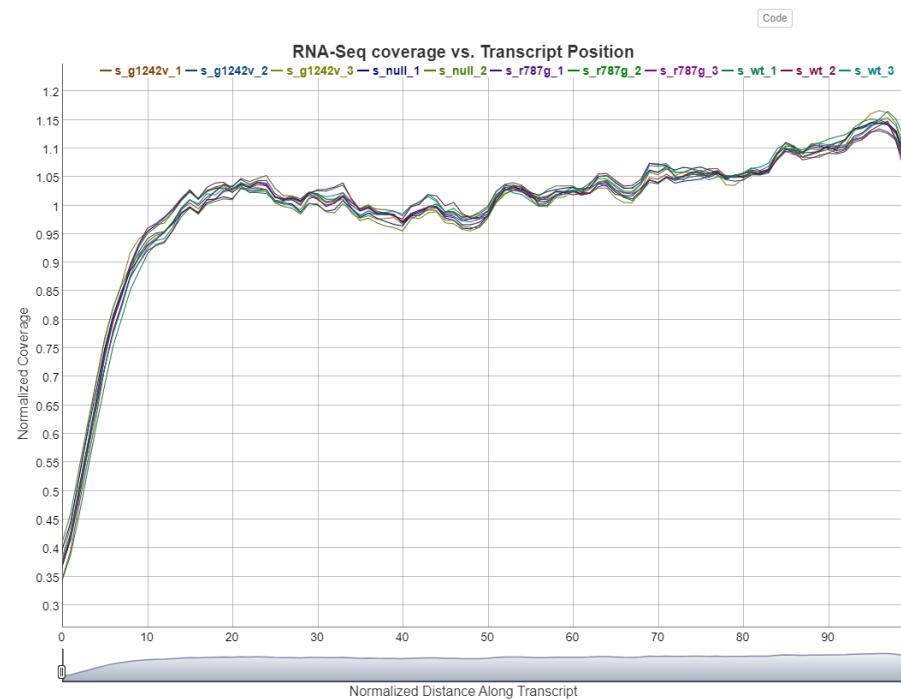
Internal Pipeline: Secundo

Picard coverage plot – 5' to 3' tx coverage

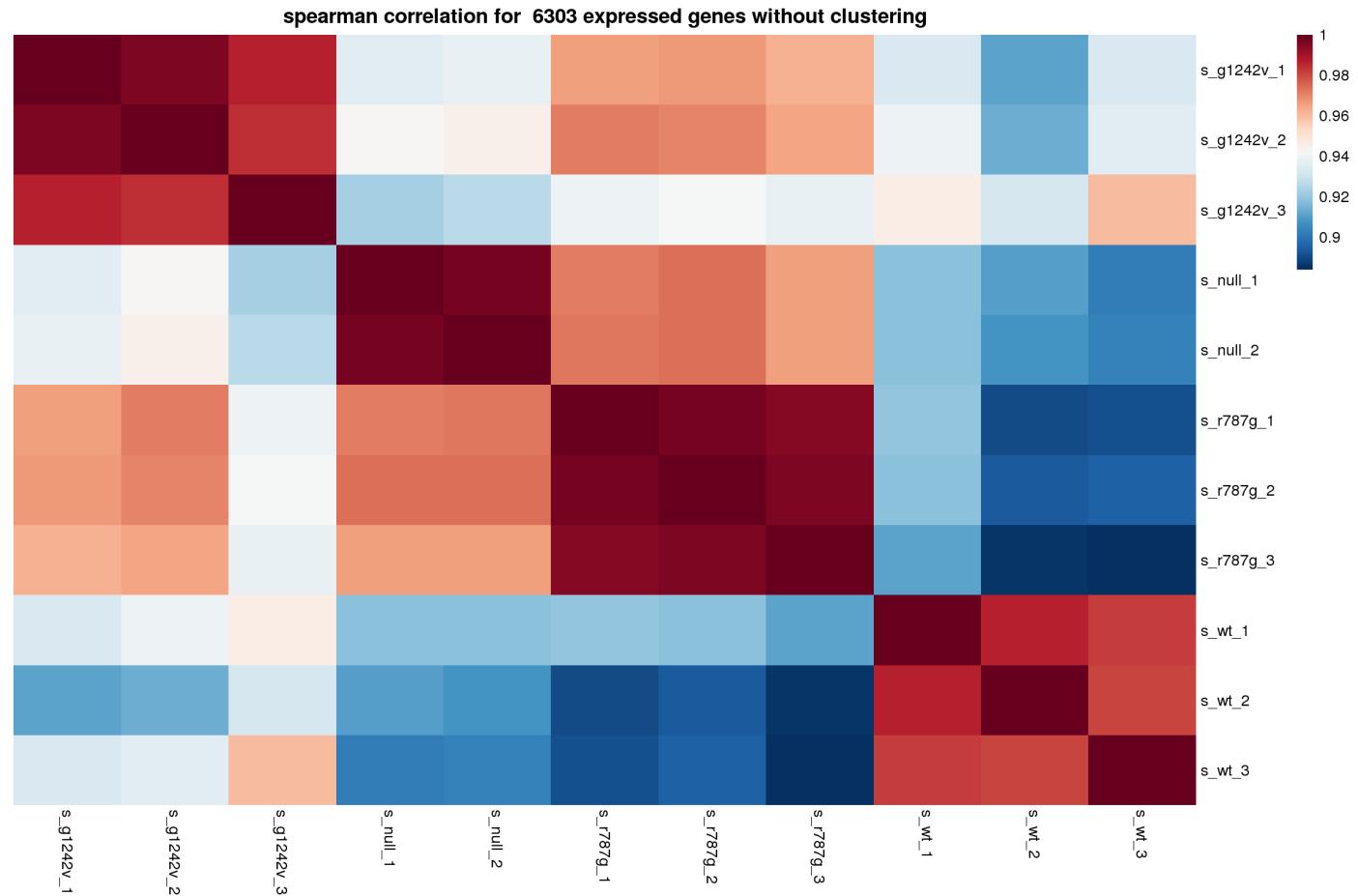
To measure the evenness of coverage, the average normalized coverage is calculated using Picard for the top 1000 expressed transcripts.

The plot shows the coverage vs. position on the transcript, with all transcripts binned to a length of 100.

According to Illumina guidance, the sample is considered acceptable if 3' bias is below 2.



Internal Pipeline: Secundo



Current and Future Trends in ‘Omics

Single Cell RNA-Seq, Things to Consider

- There are multiple products available in the market. Some require instruments and some do not.
- 10X Genomics is currently the leader in single cell market and they have several instruments depending on throughput and applications.
- Some protocols require single cell suspensions, and some require nuclei suspensions.
- Tissue dissociation can be tricky, and quality and quantity checks are a must.
- Critical to pay attention to the buffers that vendors recommend for resuspending cells.
- Sometimes require cell straining and/or cell sorting.
- Amount of sequencing needed will depend on number of cells/nuclei getting captured.
- Vendors have specific guidelines and it is important to know these.

10X Genomics



Single Cell Gene Expression

Measure gene activity on a cell-by-cell basis and characterize cell populations, cell types, and more.

Single Cell Gene Expression Flex

Expand sample access with our most sensitive whole transcriptome profiling, and run them on your own schedule.

Single Cell Multiome ATAC + Gene Expression

Measure gene expression and open chromatin simultaneously from the same cell, across thousands of cells.

Single Cell Immune Profiling

Measure the activity of immune cells and their targets.

Spatial Gene Expression

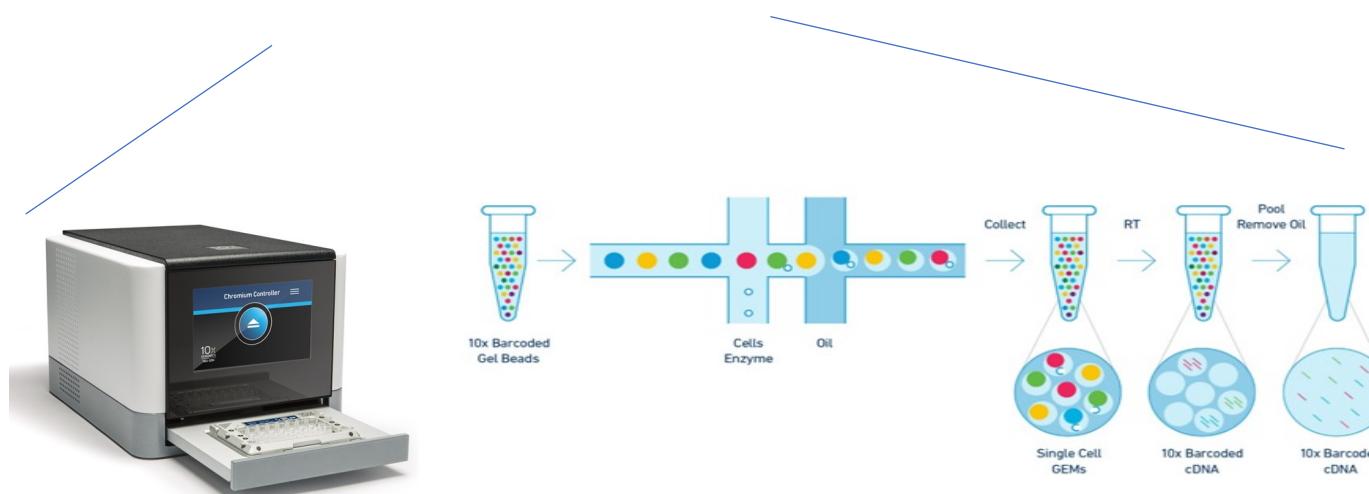
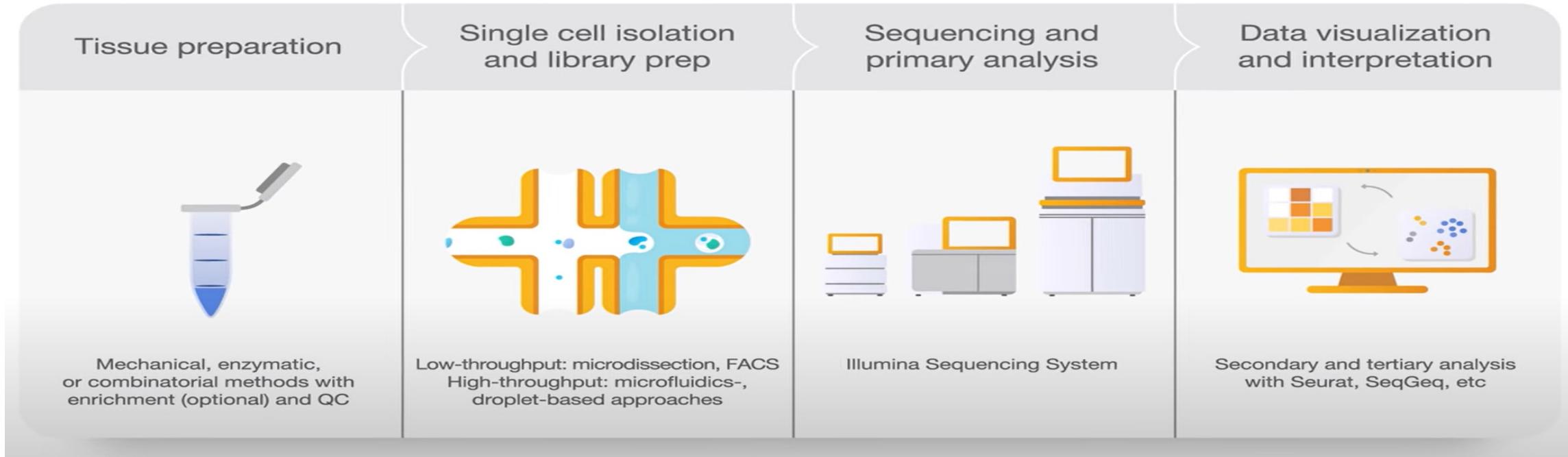
Measure spatial gene expression patterns across a tissue sample.

In Situ Gene Expression

Map gene expression at subcellular resolution with a diverse menu of validated, biologically relevant panels.

[Single Cell Sequencing \(sharepoint.com\)](#)

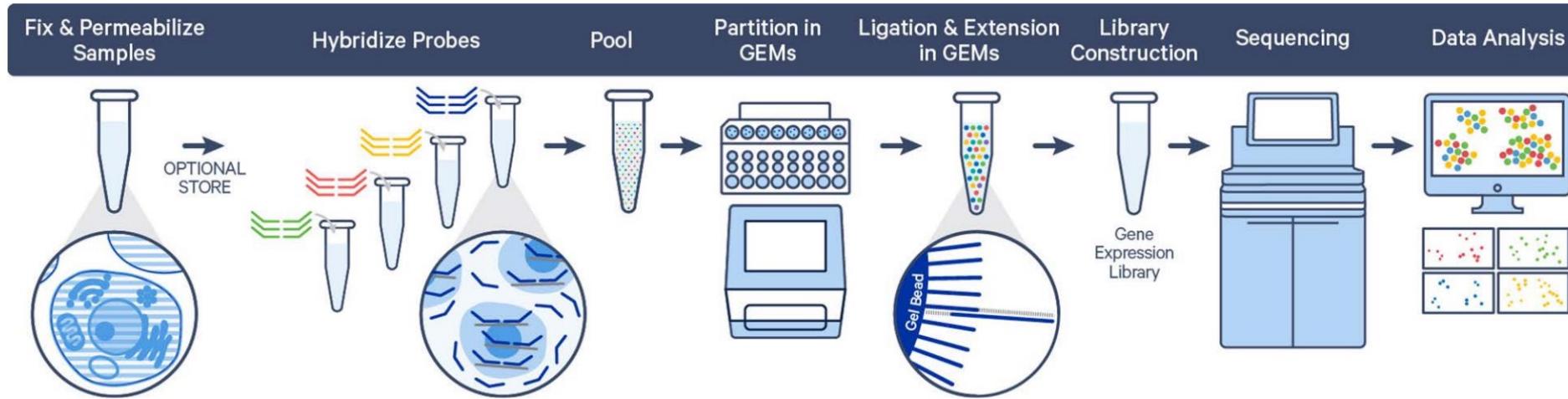
10X Genomics Single Cell RNA-seq



10X Genomics

10X Genomics Update: Fixed RNA Profiling (Human/Mouse)

Fixed RNA Profiling Workflow



- End-to-end workflow, from sample prep to software for data analysis and visualization
- Ability to fix and store cells/nuclei, minimize cell death & ease of shipping
- Ability to process when time permits
- Probe based
- Detect low expressed genes and isoforms
- Ability to multiplex

Chromium Single Cell Multiome ATAC + Gene Expression

System features

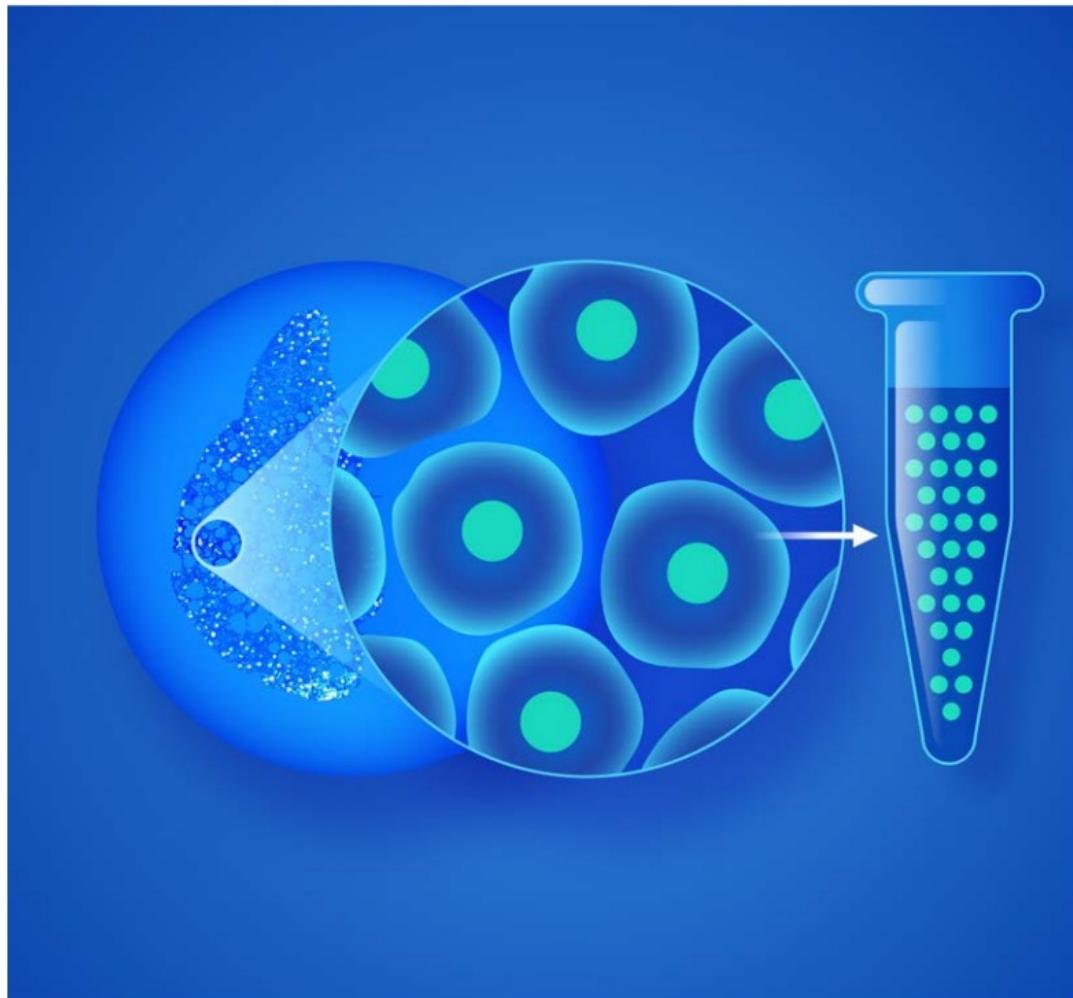
- High sensitivity; similar performance to standalone single cell ATAC or gene expression assays performed on nuclei
- Efficiently partition 500-10,000 nuclei per channel, for up to 80,000 nuclei per run
- Recover up to 65% of loaded nuclei
- Low microfluidic multiplet rate (<1% per 1000 nuclei)
- Demonstrated with cell lines, primary cell, cryopreserved samples, fresh and flash-frozen tissue
- Easy-to-use software for data analysis and visualization



Multiple Labs

Chromium Nuclei Isolation Kit

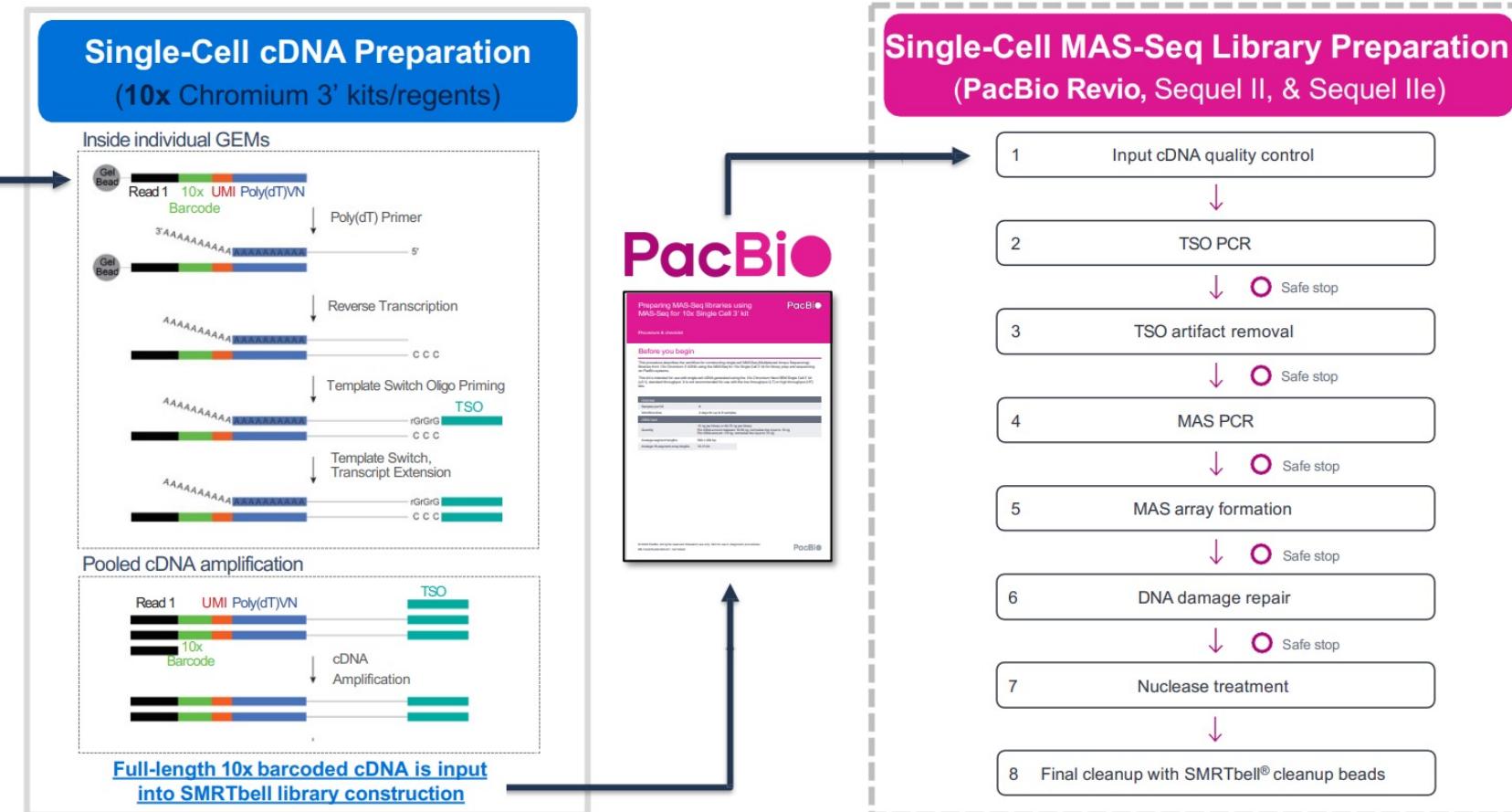
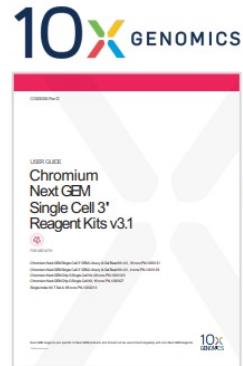
Product Features



- Compatible with most frozen mammalian tissues
- Nuclei isolation starting from 3 to 50 mg of tissue
- Capacity to process up to 8 samples at a time in one hour
- Optimized for use with the following 10x Genomics assays:
 - Single Cell ATAC
 - Single Cell Multiome ATAC + Gene Expression
 - Single Cell Gene Expression (3' and 5')
 - Fixed RNA Profiling

10X-Pacbio MAS-Seq

Options creating PacBio long-read libraries from 10x cDNA



Single Cell Transcriptomics without 10X Genomics

- Tube based scRNA-Seq
 - Fluent PIP-Seq [PIPseq™ T2 3' Single Cell RNA Kit - Fluent BioSciences](#)
 - CS Genetics [csgenetics](#)
- Honeycomb HIVE [Honeycomb Biotechnologies, Inc. – Any cell. Any where. Any time.](#)
- Combinatory barcoding methods
 - Parse Biosciences <https://www.parsebiosciences.com/>
 - Scale Biosciences [Scale Biosciences | Single Cell Library Preparation Kits](#)

Spatial Transcriptomics, A hot New Topic

Spatial Transcriptomics

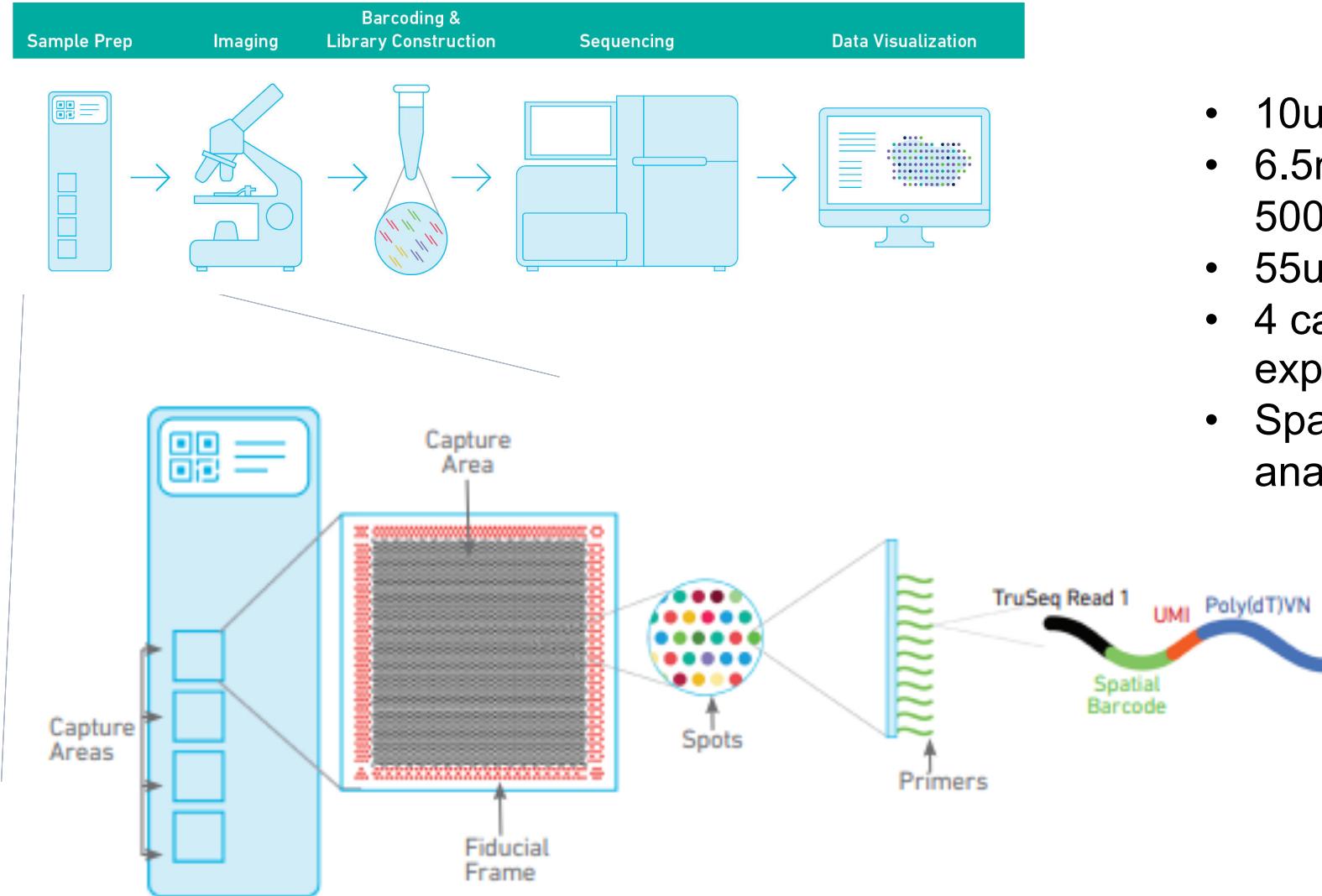
Sequencing based:

- 10X Genomics (Visium)
- SLIDE-Seq/Curio Biosciences (Fei Cen, Broad Institute)
- Stereo-Seq/STORM-Seq (Complete Genomics)

Imaging based:

- 10X Genomics Xenium
- Nanostring CosMx
- Resolve
- Vizgen
- Rebus
- Akoya...

Sequencing Based: 10X Genomics Visium Spatial Gene Expression



- 10um thick fresh-frozen tissue section
- 6.5mm x 6.5mm “Capture Areas” contain 5000 barcoded “spots”
- 55um spots, 100um center-to-center
- 4 capture areas per slide for gene expression analysis
- Space Ranger software tool for data analysis

10X Genomics

Sequencing Based: 10X Genomics Visium

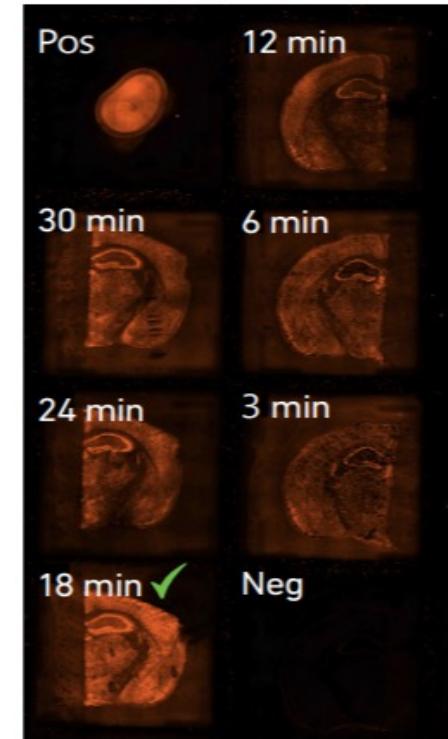
- ~3 day workflow
- Workflow requires:
 - H&E tissue staining
 - brightfield imaging
 - a full day of “Tissue Optimization” for each new tissue type to determine optimal permeabilization time (requires fluorescent imaging)
 - gentle handling of glass slide
 - qPCR to determine number of cDNA PCR cycles for each project
- Image is captured of the actual tissue section being sequenced

H&E stained tissue sections on slide

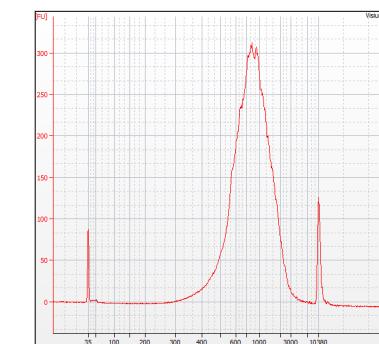


Tissue Optimization Example – permeabilization time course (min)

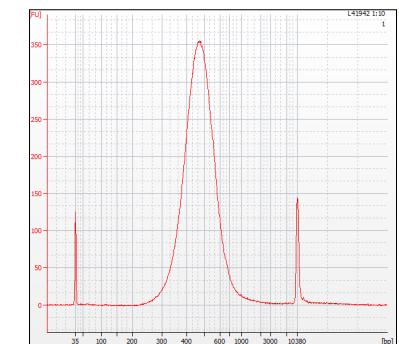
Mouse Brain



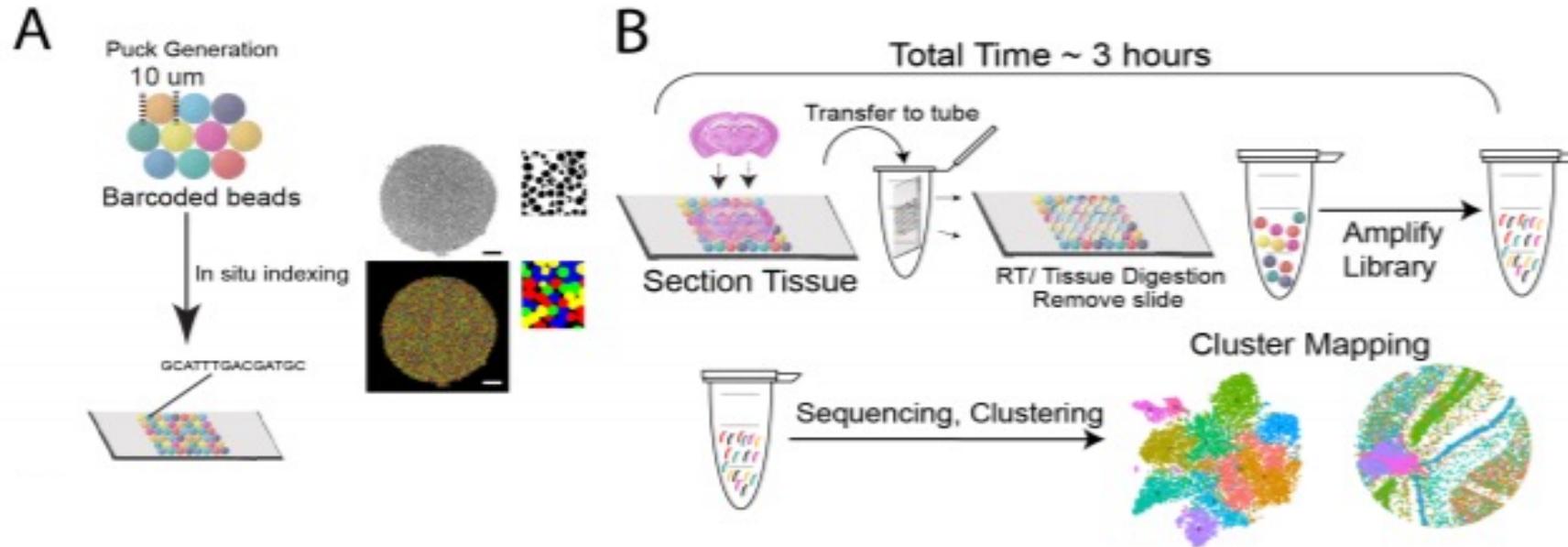
Visium cDNA



Visium Library



Sequencing Based: Slide-seq (now commercialized by Curio Biosciences)

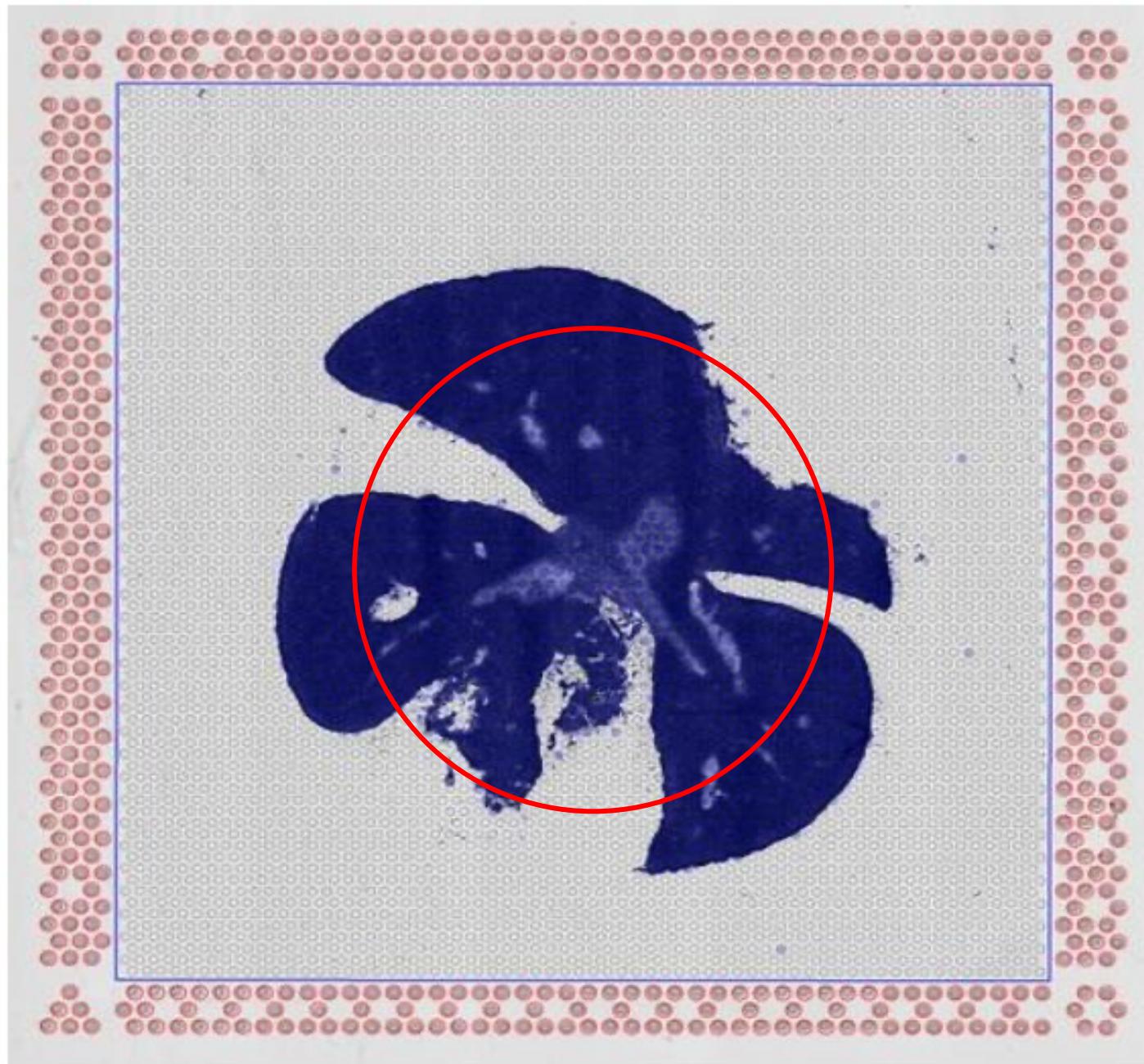


- 10 μm thick fresh-frozen tissue section
- 3mm densely packed circular “puck” containing ~70,000 10 μm spatially barcoded beads
- The puck of barcoded beads first needs to be sequenced via a monobase-sequencing strategy to determine the barcode coordinates
- Slide-seq by Rodrigues et al. (*Science* 2019)
- Slide-seqv2 by Stickels et al. (*Nature* 2020)

Rodrigues et al., published in *Science*

Visium vs. Slide-seq

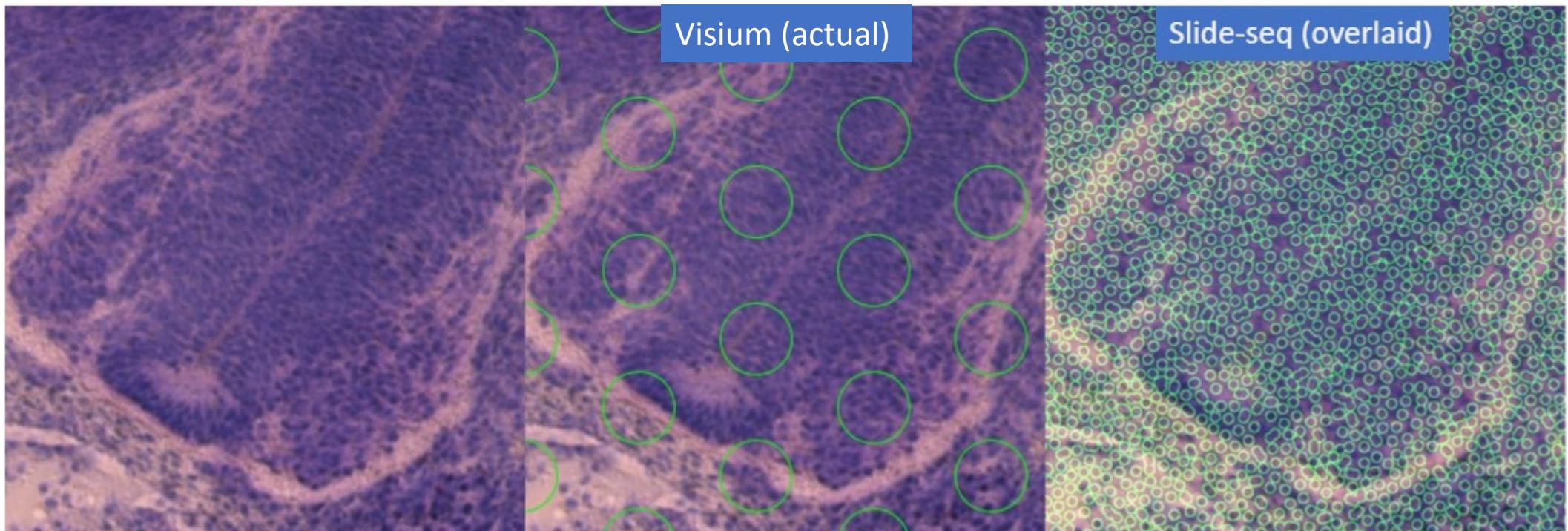
- Visium Capture Area: 6.5mm square
 - Mouse fetal liver tissue covered ~1400/5000 spots
- Slide-seq Puck: 3mm circle
 - Mouse fetal liver tissue was slightly larger than the puck



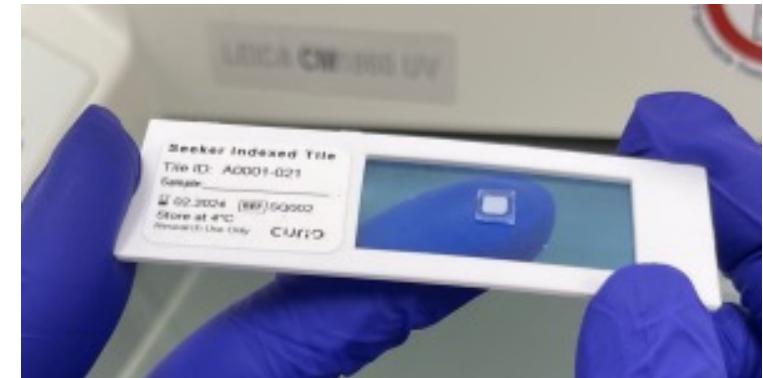
Visium Spots vs. Slide-seq Beads

- Visium spots are sparse
 - 55um diameter
 - 100um center to center
 - Regularly spaced
 - ~70% of tissue is not covered by bead
- Slide-seq beads are dense
 - 10um beads, touching
 - Patchy, irregularly distributed

*Unlike Visium, do not acquire H&E on tissue



Curio Seeker Kit: User Manual for Fresh Frozen Tissues (Early Access)



- ✓ Unbiased (oligo(dT) capture)
- ✓ 10um bead resolution
- ✓ Fresh-frozen Tissue
- ✓ Affordable
- ✓ Commercialized – Kits
- ✓ Technical Support
- ✓ Flexibility
- ✓ Fast

Tissue Sectioning with Histology

- Work with Histology to optimize tissue sectioning prior to any spatial transcriptomics project
- 10um thick tissue
- Fresh frozen tissue
- Embedded in Optimal Cutting Temperature compound (OCT)
- Cryosectioned
- Placed on Visium slide or Curio tile
- Curio tiles come in two sizes; 3mm or 1cm



Arrangement of Planaria Tissue



Probe and Imaging based: Xenium by 10X Genomics

<https://youtu.be/8YF2ZMaV4V0>

- If you are working with Human and Mouse samples, then many of the vendors have good instruments to fit your needs with pre-designed panels
- If custom assays are needed, considerations should be given to assay design time, cost, assay size etc.
- At Stowers, we decided on Xenium due to the wide variety of organisms we work with

Available Pre-designed Panels

All Human Mouse

Human Breast
Breast glandular cells, epithelial cells, adipocytes, tumor, etc.
Genes on panel 280 Add on genes up to 100
[Explore interactive dataset >](#)
[Download gene list ↴](#)

Mouse Brain
Astrocytes, interneurons, microglia, oligodendrocytes, etc.
Genes on panel 247 Add on genes up to 100
[Explore interactive dataset >](#)
[Download gene list ↴](#)

Human Brain
Astrocytes, microglia, glioblastoma, Alzheimer's disease, Parkinson's disease, etc.
Genes on panel 266 Add on genes up to 100
[View dataset >](#)
[Download gene list ↴](#)

[Xenium Platform Page - 10x Genomics](#)

Spatial Transcriptomics, Things to Consider

- Tissue sectioning is an art! Some tissue is very difficult to work with.
- Do you really need single cell or sub cellular resolution?
- Do you need whole transcriptome expression or just few genes?
- The product to use will depend on many factors such as organism, number of assays to run, coverage needed etc.
- Methods are expensive!

Summary

- Good starting material, regardless of application, is critical! Spend time here to avoid working with poor quality data after.
- Do not skip quality and quantity assessments.
- Talk to experts and include bioinformaticians before beginning your experiments.
- If your institute cannot support the type of applications you need, consider outsourcing. Many academic institutes accept outside samples.
- Feel free to reach out: agp@stowers.org